Chaînes de Markov et martingales

Pierron Théo

ENS Ker Lann

Table des matières

Chapitre 1

Conditionnement

1.1 Un exemple

Soient $(\Omega, \mathscr{F}, \mathbb{P})$ un espace de probabilité et $X : \Omega \to \{x_1, \ldots, x_m\}$, $Y : \Omega \to \{y_1, \ldots, y_n\}$ deux variables aléatoires discrètes.

La formule :

$$\begin{cases} \mathscr{F} & \to & [0,1] \\ A & \mapsto & \mathbb{P}(A \mid Y = y_j) = \frac{\mathbb{P}(A \cap \{Y = y_j\})}{\mathbb{P}(Y = y_j)} \end{cases}$$

définit une nouvelle probabilité sur $(\Omega, \mathscr{F})$, notée $\mathbb{P}^{Y=y_j}$.

Définition 1.1 L'espérance conditionnelle de X sachant que $Y = y_j$ est l'espérance de X sous la loi $\mathbb{P}^{Y=y_j}$.

$$\mathbb{E}(X \mid Y = y_j) = \sum_{i=0}^{m} x_i \mathbb{P}(X = x_i | Y = y_j) = u_j$$

Définition 1.2 L'espérance conditionnelle de X sachant Y est la variable aléatoire :

$$U(\omega) \mapsto u_j \text{ si } Y(\omega) = y_j$$

Exemple 1.1 $\Omega = [\![1, 6]\!]$, $\mathscr{F} = \mathcal{P}(\Omega)$ et $\mathbb{P}$ uniforme.

Si on prend $X = \mathrm{Id}$ et $Y = 1_{\{2,4,6\}}$, on a :

$$U(\omega) = \mathbb{E}(X \mid Y)(\omega) = \begin{cases} 4 & \text{si } \omega \text{ est pair} \\ 3 & \text{sinon} \end{cases}$$

Notons $\mathcal{G} = \sigma(Y)$. Elle est formée des unions d'atomes $G_j = \{Y = y_j\}$.

U est constante sur chacun des G_j, donc $\mathcal{G}$-mesurable.

Par ailleurs,

$$\begin{aligned}\mathbb{E}(U1_{G_j}) = u_j\mathbb{P}(G_j) &= \sum_{i=1}^{n} x_i\mathbb{P}(X = x_i \mid Y = y_j)\mathbb{P}(Y = y_j)\\ &= \sum_{i=1}^{n} x_i\mathbb{P}(X = x_i \text{ et } Y = y_j) = \mathbb{E}(X1_{G_j})\end{aligned}$$

Pour tout $G \in \mathcal{G}$, $\mathbb{E}(U1_G) = \mathbb{E}(X1_G)$.

1.2 Espérance conditionnelle – Définition

Par la suite, on fixe un espace de probabilité $(\Omega, \mathscr{F}, \mathbb{P})$.

<u>THÉORÈME 1.1</u> *Soient $\mathcal{G}$ une sous-tribu de $\mathscr{F}$ et $X \in L^1(\Omega, \mathscr{F}, \mathbb{P})$. Il existe une unique variable $\mathbb{E}(X \mid \mathcal{G}) \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ qui vérifie :*

$$\forall G \in \mathcal{G}, \mathbb{E}(\mathbb{E}(X|\mathcal{G})1_G) = \mathbb{E}(X1_G)$$

Remarque 1.1

- *Cette dernière propriété est équivalente à : pour toute variable aléatoire Z $\mathcal{G}$-mesurable bornée, $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})Z) = \mathbb{E}(XZ)$.*
- *De plus, si $X \geqslant 0$ p.s., $\mathbb{E}(X|\mathcal{G}) \geqslant 0$ p.s.*
- *L'unicité est dans L^1 !*
- *$\mathbb{E}(X \mid \mathcal{G})$ est $\mathcal{G}$-mesurable.*

Démonstration.

! : Soient Y et Y' deux versions de $\mathbb{E}(X \mid \mathcal{G})$.
Alors Y et Y' sont $\mathcal{G}$-mesurables donc $G = \{Y > Y'\} \in \mathcal{G}$.
On a aussi $\mathbb{E}(Y1_G) = \mathbb{E}(Y'1_G)$, donc $\mathbb{E}((Y - Y')1_{Y-Y'>0}) = 0$ et $Y \leqslant Y'$ p.s.
Par symétrie, $Y = Y'$.

$\exists$: Dans le cas où $X \in L^2(\Omega, \mathscr{F}, \mathbb{P})$.
Cet espace est un espace de Hilbert dont $L^2(\Omega, \mathcal{G}, \mathbb{P})$ est un sev complet.
Soit Y la projection orthogonale de X. On a :

$$\begin{cases} \mathbb{E}((Y - X)^2) = \inf\limits_{Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})} \mathbb{E}((X - Z)^2) \\ \forall Z \in L^2(\Omega, \}, \mathbb{P}), \mathbb{E}((X - Y)Z) = 0 \end{cases}$$

La variable aléatoire Y convient.
<u>Passage de L^2 à L^1</u> :

Lemme 1.1.1 De positivité
Soit U une variable aléatoire positive bornée. Alors $\mathbb{E}(U \mid \mathcal{G}) \geqslant 0$ p.s.

Démonstration. Soit W une version de $\mathbb{E}(U \mid \mathcal{G})$.
Par l'absurde, si $\mathbb{P}(W < 0) > 0$ alors $\exists n \in \mathbb{N}$ tel que $\mathbb{P}\underbrace{\left(W \leqslant -\frac{1}{n}\right)}_{G} > 0$.

On a alors

$$0 \leqslant \mathbb{E}(U1_G) = \mathbb{E}(W1_G) \leqslant -\frac{1}{n}\mathbb{P}(G) < 0$$

Contradiction. Donc $\mathbb{P}(W \geqslant 0) = 1$. ∎

Si $X \in L^1$, il est loisible de supposer $X \geqslant 0$ p.s.et il existe une suite croissante de variables aléatoires X_n bornées (donc dans L^2) qui converge vers X sans L^1. Alors si on pose $Y_n = \mathbb{E}(X_n \mid \mathcal{G})$, on a Y_n croissante positive (lemme 1.1.1)
On pose $Y = \limsup\limits_{n\to+\infty} Y_n$. Y est $\mathcal{G}$-mesurable et par convergence monotone, pour tout $G \in \mathcal{G}$, $\mathbb{E}(X_n 1_G) = \mathbb{E}(Y_n 1_G)$. En passant à la limite, $\mathbb{E}(X1_G) = \mathbb{E}(Y1_G)$. ∎

Exemple 1.2 Soit $(\Omega, \mathscr{F}, \mathbb{P}) = (]0,1[, \mathscr{B}(]0,1[), \lambda)$.
Soit $f \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ et $\mathcal{G} = \sigma(]\frac{i-1}{n}, \frac{i}{n}], i \in [\![1, n]\!])$.

$$\mathbb{E}(f \mid \mathcal{G}) = \sum_{i=1}^{n} f_i 1_{]\frac{i-1}{n}, \frac{i}{n}]}$$

avec $f_i = n\int_{\frac{i-1}{n}}^{\frac{i}{n}} f \,\mathrm{d}\lambda$.
$\mathbb{E}(f \mid \mathcal{G})$ est bien $\mathcal{G}$-mesurable et par ailleurs,

$$\begin{aligned}\mathbb{E}(f1_{]\frac{j-1}{n}, \frac{j}{n}]}) &= E\left(\sum_{i=1}^{n} f_i 1_{]\frac{i-1}{n}, \frac{i}{n}]} 1_{]\frac{j-1}{n}, \frac{j}{n}]}\right) \\ &= \mathbb{E}(f_j 1_{]\frac{j-1}{n}, \frac{j}{n}]})\end{aligned}$$

<u>Théorème 1.2</u> *Si X est un variable aléatoire positive alors la formule $\mathbb{E}(X \mid \mathcal{G}) = \lim\limits_{n\to+\infty} \mathbb{E}(X \wedge n \mid \mathcal{G})$ définit une variable $\mathcal{G}$-mesurable caractérisée par :*

$$\forall Z \geqslant 0 \quad \mathcal{G}\text{-mesurable, } \mathbb{E}(XZ) = \mathbb{E}(\mathbb{E}(X \mid \mathcal{G})Z)$$

1.3 Espérance conditionnelle – Propriétés

1.3.1 Premières propriétés

Proposition 1.1 Soit $X \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ et $\mathcal{G}$ une sous-tribu de $\mathscr{F}$.

1. Si X est $\mathcal{G}$-mesurable alors $\mathbb{E}(X \mid \mathcal{G}) = X$ p.s.
2. $X \mapsto \mathbb{E}(X \mid \mathcal{G})$ est linéaire
3. $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})) = \mathbb{E}(X)$
4. $|\mathbb{E}(X \mid \mathcal{G})| \leqslant \mathbb{E}(|X| \mid \mathcal{G})$ p.s.
5. Si $X \geqslant X'$ p.s., $\mathbb{E}(X \mid \mathcal{G}) \geqslant \mathbb{E}(X' \mid \mathcal{G})$ p.s.

Démonstration.

1. Unicité
2. Si $U = \mathbb{E}(X \mid \mathcal{G})$ et $V = \mathbb{E}(Y \mid \mathcal{G})$, alors $\alpha U + \beta V$ est une version de $\mathbb{E}(\alpha X + \beta Y \mid \mathcal{G})$.
3. Définition avec $1_G = 1_\Omega = 1$.
4. $|\mathbb{E}(X \mid \mathcal{G})| = |\mathbb{E}(X^+ \mid \mathcal{G}) - \mathbb{E}(X^- \mid \mathcal{G})| \leqslant \mathbb{E}(X^+ \mid \mathcal{G}) + \mathbb{E}(X^- \mid \mathcal{G}) = \mathbb{E}(|X| \mid \mathcal{G})$.
5. Linéarité+lemme 1.1.1 ∎

Remarque 1.2
- $\|\mathbb{E}(X \mid \mathcal{G})\|_1 \leqslant \|X\|_1$.
- *Plus clairement, l'opérateur* $X \mapsto \mathbb{E}(X \mid \mathcal{G})$ *linéaire positif de norme* 1.

1.3.2 Extensions des théorèmes usuels

Proposition 1.2 Soit $X \in L^1(\Omega, \mathscr{F}, \mathbb{P})$, $\mathcal{G}$ une sous-tribu de $\mathscr{F}$.

1. Si $(X_n)_n$ est une suite positive croissante qui tend vers X p.s. alors $\mathbb{E}(X_n \mid \mathcal{G})$ croît vers $\mathbb{E}(X \mid \mathcal{G})$ p.s.
2. Si $(X_n)_n$ est une suite positive, $\mathbb{E}(\liminf X_n \mid \mathcal{G}) \leqslant \liminf \mathbb{E}(X_n \mid \mathcal{G})$ p.s.
3. Si $|X_n(\omega)| \leqslant V(\omega)$ avec $\mathbb{E}(V) < \infty$ et si X_n converge p.s. vers X alors $\mathbb{E}(X_n \mid \mathcal{G})$ converge p.s. vers $\mathbb{E}(X \mid \mathcal{G})$.
4. Si $c : \mathbb{R} \to \mathbb{R}$ est convexe et $\mathbb{E}(|c(X)|) < \infty$, alors $\mathbb{E}(c(X) \mid \mathcal{G}) \geqslant c(\mathbb{E}(X \mid \mathcal{G}))$ p.s.

Démonstration.

1. Voir passage $L^2 \to L^1$.
2. Exo
3. Exo

4. On utilise $c(x) = \sup_{n\in\mathbb{N}}(a_n x + b_n)$. Soit $n \in \mathbb{N}$.
   On a $c(X) \geqslant a_n X + b_n$ donc $\mathbb{E}(c(X) \mid \mathcal{G}) \geqslant a_n \mathbb{E}(X \mid \mathcal{G}) + b_n$ p.s.
   Comme on a un sup sur un ensemble dénombrable, on peut passer au sup et on a :
   $$\mathbb{E}(c(X) \mid \mathcal{G}) \geqslant c(\mathbb{E}(X \mid \mathcal{G}))\text{p.s.}$$ ■

Remarque 1.3 *On a ainsi, pour tout $p \geqslant 1$, $\|\mathbb{E}(X \mid \mathcal{G})\|_p \leqslant \|X\|_p$.*

1.3.3 Propriétés spécifiques

Proposition 1.3 Soient X une variable aléatoire et Y une variable $\mathcal{G}$-mesurable.
$$\mathbb{E}(XY \mid \mathcal{G}) = Y\mathbb{E}(X \mid \mathcal{G}) \text{ p.s.}$$
dès lors que les espérances sont bien définie.

Démonstration. $Y\mathbb{E}(X \mid \mathcal{G})$ est $\mathcal{G}$-mesurable comme produit de deux variables $\mathcal{G}$-mesurables.

Si $G \in \mathcal{G}$, on a
$$\mathbb{E}(Y\mathbb{E}(X \mid \mathcal{G})1_G) = \mathbb{E}(Y1_G\mathbb{E}(X \mid \mathcal{G})) = \mathbb{E}(XY1_G)$$
Par unicité de l'espérance conditionnelle, on a le résultat. ■

Proposition 1.4 Soient $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ deux sous-tribus et $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Alors
$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}) = \mathbb{E}(X \mid \mathcal{H}) \text{ p.s.}$$

Démonstration. $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H})$ est $\mathcal{H}$-mesurable. Soit $H \in \mathcal{H}$.
$$\mathbb{E}(\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H})1_H) = \mathbb{E}(\mathbb{E}(\mathbb{E}(X1_H \mid \mathcal{G}) \mid \mathcal{H})) = \mathbb{E}(X1_H)$$ ■

Proposition 1.5 Deux tribus $\mathcal{G}$ et $\mathcal{H}$ sont indépendantes ssi pour tout $G \in \mathcal{G}$, $\mathbb{E}(1_G \mid \mathcal{H}) = \mathbb{E}(1_G) = \mathbb{P}(G)$ p.s.

Remarque 1.4 *X et Y sont indépendantes ssi pour tout g mesurables*
$$\mathbb{E}(g(X) \mid Y) = \mathbb{E}(g(X)) \text{ p.s.}$$

Démonstration.

$\Rightarrow$ Si $H \in \mathcal{H}$ et $G \in \mathcal{G}$. On a
$$\mathbb{E}(1_G 1_H) = \mathbb{E}(1_G)\mathbb{E}(1_H) = \mathbb{E}(\mathbb{E}(1_G)1_H)$$
donc $\mathbb{E}(1_G) = \mathbb{E}(1_G \mid \mathcal{H})$ par unicité.

$\Leftarrow$ Si $H \in \mathcal{H}$ et $G \in \mathcal{G}$, on a

$$\mathbb{P}(H \cap G) = \mathbb{E}(1_H 1_G) = \mathbb{E}(\mathbb{E}(1_G)1_H) = \mathbb{P}(G)\mathbb{P}(H)$$

d'où $H \perp\!\!\!\perp G$. ■

Proposition 1.6 Soient X, Y deux variables aléatoires à valeurs dans E, F, $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. On suppose que $X \perp\!\!\!\perp \mathcal{G}$ et que Y est $\mathcal{G}$-mesurable.

Alors pour tout $f : E \times F \to \mathbb{R}$ mesurable,

$$\mathbb{E}(f(X,Y) \mid \mathcal{G}) = \int_E f(x,Y)\mathbb{P}_X(\mathrm{d}x) = \Phi(Y)$$

Démonstration. $\Phi(Y)$ est $\mathcal{G}$-mesurable. Soit Z $\mathcal{G}$-mesurable bornée, on a :

$$\begin{aligned}\mathbb{E}(f(X,Y)Z) &= \int f(x,y)z\mathbb{P}_{(X,Y,Z)}(\mathrm{d}x, \mathrm{d}y, \mathrm{d}z) \\ &= \int\int f(x,y)\mathbb{P}_X(\mathrm{d}x)z\mathbb{P}_{Y,Z}(\mathrm{d}y, \mathrm{d}z) \\ &= E(\Phi(Y)Z)\end{aligned}$$

■

1.4 Calcul d'espérance conditionnelle

1.4.1 Cas des variables discrètes

Soit $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ et $Y : \Omega \to \{y_1, \ldots, y_n\}$ une variable discrète.

$$\mathbb{E}(X \mid Y) = \mathbb{E}(X \mid \sigma(Y))$$

$\sigma(Y)$ est engendrée par les atomes $G_i = \{Y = y_i\}$.

Proposition 1.7 On a

$$\mathbb{E}(X \mid Y) = \sum_{i=1}^{\infty} \frac{\mathbb{E}(X 1_{G_i})}{\mathbb{P}(G_i)} 1_{G_i}$$

Démonstration. Le membre de droite est $\sigma(Y)$-mesurable. De plus, on a

$$\mathbb{E}\left(\sum_{i=1}^{\infty} \frac{\mathbb{E}(X 1_{G_i})}{\mathbb{P}(G_i)} 1_{G_i} 1_{G_j}\right) = \mathbb{E}(X 1_{G_j})$$

■

1.4.2 Cas des variables à densité

Proposition 1.8 Soit (X, Y) de densité $p(x, y)$ par rapport à la mesure de Lebesgue. Alors pour tout fonction mesurable h,

$$\mathbb{E}(h(X) \mid Y) = \varphi(Y)$$

avec

$$\varphi(u) = \begin{cases} \frac{\int h(x)p(x,u)\,\mathrm{d}x}{\int \rho(x,u)\,\mathrm{d}x} & \text{si } p(x,u)\,\mathrm{d}x > 0 \\ 0 & \text{sinon} \end{cases}$$

(densité conditionnelle de X par rapport à Y).

Démonstration. $\varphi(Y)$ est Y-mesurable.

Soit g mesurable bornée.

$$\begin{aligned} \mathbb{E}(h(x)g(y)) &= \int h(x)g(y)p(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int \frac{\int h(x)p(x,y)\,\mathrm{d}x}{q(y)} g(y)q(y)\,\mathrm{d}y \\ &= \int \varphi(y)g(y)q(y)\,\mathrm{d}y = \mathbb{E}(\phi(Y)g(Y)) \end{aligned}$$

avec $q(y) = \displaystyle\int p(x, y)\,\mathrm{d}y$. ■

Exemple 1.3 Soit (X, Y) de densité $\mathrm{e}^{-Y} 1_{0<X<Y}$.

$$\mathbb{E}(h(x) \mid Y) = \frac{\int h(x)\mathrm{e}^{-Y} 1_{0<x<Y}\,\mathrm{d}x}{\int \mathrm{e}^{-Y} 1_{0<x<Y}\,\mathrm{d}x} = \frac{1}{Y}\int_0^Y h(x)\,\mathrm{d}x$$

1.4.3 Variables gaussiennes

Rappel : Un vecteur est dit gaussien ssi toute combinaison linéaire de ses composantes est une variable aléatoire gaussienne.

Proposition 1.9 Soit $(X_1, \dots, X_n, Y_1, \dots, Y_m)$ un vecteur gaussien centré. Alors $(X_1, \dots, X_n) \perp\!\!\!\perp (Y_1, \dots Y_m)$ ssi $\mathrm{Cov}(X_i, Y_j) = 0$.

Proposition 1.10 Soit $(X, Y_1, \dots, Y_n)$ un vecteur gaussien centré. Alors $\mathbb{E}(X \mid \sigma(Y_1, \dots, Y_n))$ est la projection orthogonale de X sur l'espace vectoriel engendré par les Y_i. Il existe donc λ_i tel que $\mathbb{E}(X \mid \sigma(Y_1, \dots, Y_n)) = \sum_i \lambda_i Y_i = \widehat{X}$. De plus, pour tout h mesurable,

$$\mathbb{E}(h(X) \mid \sigma(Y_1, \dots, Y_n)) = \int h(x) q_{\widehat{X},\sigma^2}(x)\,\mathrm{d}x$$

où $q_{\widehat{X},\sigma^2}$ est la densité d'une gaussienne $\mathcal{N}(\widehat{X},\sigma^2)$:

$$q_{\widehat{X},\sigma^2}(u) = \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{(u-\widehat{X})^2}{2\sigma^2}}$$

$(\sigma^2 = \mathbb{E}((X-\widehat{X})^2))$.

Démonstration. Si $X \in \sigma(Y_1,\dots,Y_n)$ le résultat est trivial :

$$\mathbb{E}(X \mid \sigma(Y_1,\dots,Y_n)) = X \text{ p.s.}$$

Sinon, soit $\widehat{X} = \sum_i \lambda_i Y_i$ la projection orthogonale sur Vect $\{Y_1,\dots,Y_n\}$. $\mathbb{E}((X-\widehat{X})Y_i) = 0$ donc en particulier, $(X-\widehat{X},Y_1,\dots,Y_n)$ est un vecteur gaussien centré et d'après la proposition précédente, $X-\widehat{X}$ est indépendant de $(Y_1,\cdots,Y_n)$. D'où

$$\mathbb{E}(X \mid \sigma(Y_1,\dots,Y_n)) = \mathbb{E}(X-\widehat{X} \mid Y_1,\dots,Y_n) + \widehat{X} = \mathbb{E}(X-\widehat{X}) + \widehat{X} = \widehat{X} \quad \blacksquare$$

Exemple 1.4 Soit (X,Y) un vecteur gaussien centré. On projette sur $\mathbb{R}Y$ donc il existe $\lambda \in \mathbb{R}$ tel que :

$$\mathbb{E}(X \mid Y) = \lambda Y \text{ p.s.}$$

On passe à l'espérance :

$$\mathbb{E}(XY) = \mathbb{E}(\mathbb{E}(X \mid Y)Y) = \lambda\mathbb{E}(Y^2)$$

Donc $\lambda = \frac{\mathbb{E}(XY)}{\mathbb{E}(Y^2)}$.

1.5 Notion de loi conditionnelle

1.5.1 Définitions

Définition 1.3 Soit $(E,\mathcal{E})$ et $(F,\mathcal{F})$ deux espaces mesurables. On appelle probabilité de transition une application $\nu : E \times \mathcal{F} \to [0,1]$ telle que pour tout $x \in E$, $A \mapsto \nu(x,A)$ est une probabilité sur $\mathcal{F}$ et pour tout $A \in \mathcal{F}$, $x \mapsto \nu(x,A)$ est mesurable.

Exemple 1.5 Si λ est une mesure σ-finie sur $\mathcal{F}$ et $f : E \times F \to \mathbb{R}$ mesurable telle que $\int f(x,y)\lambda(\mathrm{d}y) = 1$ alors $\nu(x,A) = \int_A f(x,y)\lambda(\mathrm{d}y)$ est une probabilité de transition.

Proposition 1.11 Soit ν une probabilité de transition sur $(E, \mathcal{E})$, $(F, \mathcal{F}$.

Si h est mesurable positive (bornée) sur $(F, \mathcal{F})$ alors $\varphi(x) = \int h(y)\nu(x, \mathrm{d}y)$ est mesurable.

Si λ est une mesure de probabilité sur $(E, \mathcal{E})$, alors $\mu(A) = \int \nu(x, A)\lambda(\mathrm{d}x)$ est une mesure sur $(F, \mathcal{F})$.

Définition 1.4 Soient X, Y deux variables aléatoires sur $(E, \mathcal{E})$ et $(F, \mathcal{F})$. On appelle loi conditionnelle de Y suivant X toute probabilité de transition ν sur $E \times \mathcal{F}$ telle que pour tout h mesurable positive sur $(F, \mathcal{F})$, on ait :

$$\varphi(X) = \mathbb{E}(h(Y) \mid X) = \int h(y)\nu(X, \mathrm{d}y)$$

Remarque 1.5 D'après la proposition précédente, φ est mesurable donc le membre de droite est bien mesurable par rapport à $\sigma(X)$.

Par définition, on pose alors $\mathbb{P}(Y \in A \mid X) = \nu(X, A)$ (pour $A \in \mathcal{F}$.

1.5.2 Existence et unicité

Unicité : Si ν et ν' sont deux lois conditionnelles de Y sachant X alors pour tout $A \in \mathcal{F}$, $\nu(X, A) = \nu'(X, A) = \mathbb{P}(Y \in A \mid X)$ p.s.ie pour tout $A \in \mathcal{F}$, $\nu(x, A) = \nu'(x, A)$ $\mathbb{P}_X$-p.s.

Si les mesures de probabilité sur $(F, \mathcal{F})$ sont caractérisées par leurs valeurs sur un nombre dénombrables d'ensembles (c'est le cas dans $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ alors $\mathbb{P}_X$-p.s.pour tout $A \in \mathcal{F}$, $\nu(x, A) = \nu'(x, A)$.

Existence :

Théorème 1.3 *Si $(E, \mathcal{E})$ et $(F, \mathcal{F})$ sont des espaces polonais (métriques, complets, séparables) munis de leur tribu borélienne, il existe une loi conditionnelle de Y sachant X.*

1.5.3 En pratique

Cas discret

Si X est une variable discrète,

$$\nu(x, A) = \begin{cases} \mathbb{P}(Y \in A \mid X = x) & \text{si } \mathbb{P}(X = x) > 0 \\ \delta_{y_0}(A) & \text{pour } y_0 \text{ fixé dans } F \end{cases}$$

Cas des variables à densité

Si (X,Y) a pour densité $p(x,y)$ alors on note $q(x)=\int p(x,y)\,\mathrm{d}y$ et on a :

$$\nu(x,A)=\begin{cases}\frac{1}{q(x)}\int_A p(x,y)\,\mathrm{d}t & \text{si } q(x)>0\\ \delta_0(A) & \text{sinon}\end{cases}$$

Exemple 1.6 $p(x,y)=e^{-y}1_{0\leqslant x\leqslant y}$. On cherche la loi de X sachant Y.

$$\nu(y,A)=\frac{1}{\int p(x,y)\,\mathrm{d}x}\int_A p(x,y)\,\mathrm{d}x=\frac{\mathrm{e}^{-y}}{y\mathrm{e}^{-y}}\int_{A\cap[0,y]}1\,\mathrm{d}x)=\frac{1}{y}\lambda(A\cap[0,y])$$

C'est donc une loi uniforme sur $[0,y]$.

Cas gaussien

Si $(Y,X_1,\dots,X_n)$ est un vecteur gaussen centré, on a

$$\nu((x_1,\dots,x_n),A)=\int_A q_{\sum\lambda_i x_i,\sigma^2}(u)\,\mathrm{d}u$$

où $q_{\sum\lambda_i x_i,\sigma^2}$ est la loi gaussienne et $\sum_i \lambda_i x_i$ le projeté orthogonal sur l'espace Vect $\{X_i\}$.

Chapitre 2

Martingales et temps d'arrêt

2.1 Définitions et exemples

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité.

Définition 2.1 Une filtration de $(\Omega, \mathcal{F}, \mathbb{P})$ est une suite croissante de sous-tribus de $\mathcal{F}$. On dit que $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$ est un espace de probabilité filtré.

Exemple 2.1 Soit $(X_n)_n$ une suite de variables aléatoires. La suite $\mathcal{F}_n^X = \sigma(X_0, \ldots, X_n)$ est une filtration. On l'appelle filtration canonique associée à $(X_n)_n$.

Si $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1[, \mathscr{B}([0,1[), \lambda)$ alors la suite $\mathcal{G}_n = \sigma([\frac{i-1}{2^n}, \frac{i}{2^n}[, i \in [\![1, 2^n]\!])$ définit une filtration appellée filtration dyadique de $[0,1[$.

Définition 2.2 Soit $(X_n)_n$ une suite de variables aléatoires et $(\mathcal{F}_n)_n$ une filtration. On dit que $(X_n)_n$ est adaptée à $(\mathcal{F}_n)_n$ ssi X_n est $\mathcal{F}_n$-mesurable.

On dit que $(X_n)_n$ est prévisible ssi X_n est $\mathcal{F}_{n-1}$-mesurable.

Dans la suite, on fixe un espace $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$ filtré.

Définition 2.3 Soit $(X_n)_n$ une suite de variables adaptée à $(\mathcal{F}_n)_n$ et telle que $\mathbb{E}(|X_n|) < \infty$.

On dit que $(X_n)_n$ est une martingale ssi pour tout $n \in \mathbb{N}$, $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n$ p.s.

On dit que $(X_n)_n$ est une sous-martingale ssi pour tout $n \in \mathbb{N}$, $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) \geqslant X_n$ p.s.

On dit que $(X_n)_n$ est une sur-martingale ssi pour tout $n \in \mathbb{N}$, $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) \leqslant X_n$ p.s.

Proposition 2.1 Si $(X_n)_n$ est une martingale, on a en fait pour tout $m \geqslant n$, $\mathbb{E}(X_m \mid \mathcal{F}_n) = X_n$ p.s.et pour tout $n \in \mathbb{N}$, $\mathbb{E}(X_n) = \mathbb{E}(X_0)$.

Démonstration. Le cas $n = m$ est clair, $m = n + 1$ découle de la définition.

Enfin, si $m \geqslant n+2$,

$$\mathbb{E}(X_m \mid \mathcal{F}_n) = \mathbb{E}(\mathbb{E}(X_m \mid \mathcal{F}_{m-1}) \mid \mathcal{F}_n) = \mathbb{E}(X_{m-1} \mid \mathcal{F}_n) \qquad \blacksquare$$

Exemple 2.2

- Soit X une valeur aléatoire telle que $\mathbb{E}(|X|) < \infty$ et $(\mathcal{F}_n)_n$ une filtration. Alors, $X_n = \mathbb{E}(X \mid \mathcal{F}_n)$ est une martingale dite fermée.
  X_n est intégrable car X l'est. $(X_n)_n$ est adaptée et :
  $$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = \mathbb{E}(\mathbb{E}(X \mid \mathcal{F}_{n+1}) \mid \mathcal{F}_n) = \mathbb{E}(X \mid \mathcal{F}_n) = X_n$$
  A-t-on $X_n \to X$?
- Soit $x \in \mathbb{R}^n$, Y_i iid à valeurs dans $\mathbb{R}^n$. On pose $X_0 = x$ et $X_n = x + \sum_{i=1}^{n} Y_i$. On a :
  $$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = \mathbb{E}(X_n + Y_{n+1} \mid \mathcal{F}_n) = X_n + \mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) = X_n + \mathbb{E}(Y_{n+1})$$
  Suivant le signe de $\mathbb{E}(Y_0)$, on a une martingale, sous-martingale ou sur-martingale.
- Soit G un groupe, μ une mesure de probabilité sur G telle que $\int_G g \,\mathrm{d}\mu = e_G$. $X_n = eg_1 \dots g_n$ où g_i iid de loi μ.
  $$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = \mathbb{E}(eg_1 \dots g_{n+1} \mid \mathcal{F}_n) = X_n \underbrace{\mathbb{E}(g_{n+1} \mid \mathcal{F}_n)}_{=e} = X_n$$

2.1.1 Transformées de martingales

Proposition 2.2 Soit $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction convexe et $(X_n)_n$ une suite de variables aléatoires adaptée telles que $\mathbb{E}(|\varphi(X_n)| < \infty$.

- Si $(X_n)_n$ est une martingale, alors $\varphi(X_n)$ est une sous-martingale.
- Si $(X_n)_n$ est une sous-martingale et φ croissante, alors $\varphi(X_n)$ est une sous-martingale.

Démonstration. On a par Jensen $\mathbb{E}(\varphi(X_{n+1} \mid F_n)) \geqslant \varphi(\mathbb{E}(X_{n+1} \mid F_n)) = \varphi(X_n)$ si $(X_n)_n$ est une martingale. De même, $\varphi(\mathbb{E}(X_{n+1} \mid F_n)) \geqslant \varphi(X_n)$. $\blacksquare$

Proposition 2.3 Soit (X_n) une suite adaptée de variables L^1 et $(H_n)_n$ une suite prévisible bornée. On définit $(H \cdot X)_0 = 0$ et pour tout $n > 0$, $(H \cdot X)_n = H_1(X_1 - X_0) + H_2(X_2 - X_1) + \ldots + H_n(X_n - X_{n-1})$.

(On peut voir H comme une mise et $X_n - X_{n-1}$ comme un gain.)

Alors :

- Si $(X_n)_n$ est une martingale, $(H \cdot X)_n$ aussi.

- Si $(H_n)_n$ est positive, et $(X_n)_n$une sur/sous-martingale alors $(H \cdot X_n)_n$ est une sur/sous-martingale.

Démonstration. Comme $(H_n)_n$ est bornée, $(H\cdot X)_n$ est L^2. Elle est aussi bien adaptée. On a :

$$\begin{aligned}\mathbb{E}((H \cdot X)_{n+1} - (H \cdot X)_n \mid F_n) &= \mathbb{E}(H_{n+1}(X_{n+1} - X_n)) \mid F_n) \\ &= H_{n+1}\mathbb{E}(X_{n+2} - X_n \mid F_n) = 0\end{aligned}$$ ■

2.2 Temps d'arrêt est théorèmes

2.2.1 Notion de temps d'arrêt

<u>Définition 2.4</u> Une variable aléatoire T à valeurs dans $\overline{\mathbb{N}}$ est un temps d'arrêt par rapport à la filtration $(F_n)_n$ ssi $\{T = n\} \in F_n$ et $T_\infty = \Omega \backslash \bigcup_{n\in\mathbb{N}} \{T = n\}$.

On note de plus $F_\infty = \sigma\left(\bigcup_{n\in\mathbb{N}} F_n\right)$.

Exemple 2.3

- Si $N \in \mathbb{N}$ est fixé alors $T = N$ est un temps d'arrêt. En effet, $\{T \leqslant n\} = \varnothing$ si $n < N$ et Ω sinon, qui appartient bien à F_n.
- Si $A \in \mathscr{B}(\mathbb{R})$, $T = \inf\{n, X_n \in A\}$ est un temps d'arrêt. $\{T = n\} = \{X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A\} \in F_n$. Attention, en général, $L = \sup\{n, X_n \in A\}$ n'est pas un temps d'arrêt (même si $(X_n)_n$ est finie).

Proposition 2.4 Si $S, T, (T_k)_k$ sont des temps d'arrêt, alors $S \wedge T$, $S \vee T$, $\inf T_k$, $\sup T_k$, $\liminf T_k$ et $\limsup T_k$ sont des temps d'arrêt.

Démonstration. Soit $n \in \mathbb{N}$, $\{S \wedge T \leqslant n\} = \{S \leqslant n\} \cup \{T \leqslant n\} \in F_n$ et $\{S \vee T \leqslant\} = \{S \leqslant n\} \cap \{T \leqslant n\} \in F_n$.

De même $\{\liminf T_k \leqslant n\} = \bigcup_{n\geqslant 1}\bigcap_{k\geqslant n} \{T_k \leqslant n\} \in F_n$. ■

<u>Définition 2.5</u> Soit T un temps d'arrêt. On définit la tribu arrêtée au temps T par

$$F_T = \{A \in F, A \cap \{T = n\} \in F_n \forall n\}$$

Proposition 2.5 Si $S \leqslant T$ sont deux temps d'arrêt, alors $F_S \subset F_T$.

Démonstration. Soit $A \in F_S$. $A \cap \{S = k\} \in F_k$ pour tout k.

$$A \cap \{T = n\} = A \cap \bigcup_{k=0}^{n} \{S = k, T = n\} = \bigcup_{k=0}^{n} (A \cap \{S = k\} \cap \{T = n\}) \in F_n \quad \blacksquare$$

Proposition 2.6 Soit $(Y_n)_n$ une suite adaptée et T un temps d'arrêt. Alors

$$1_{T<\infty} Y_T = \begin{cases} Y_n & \text{si } T = n \\ 0 & \text{sinon} \end{cases}$$

est F_T-mesurable.

Remarque 2.1 Si $T < \infty$ p.s. alors on peut écrire $Y_T(\omega) = Y_n(\omega)$ si $T(\omega) = n$ donc $Y_T(\omega) = Y_{T(\omega)}(\omega)$.

Démonstration. Soit $B \in \mathscr{B}(\mathbb{R})$, $n \in \mathbb{N}$.

$$F_n \ni \{Y_n \in B\} \cap \{T = n\} = \{1_{T<\infty} Y_T \in B\} \cap \{T = n\} \in F_n \quad \blacksquare$$

Remarque 2.2 Si T est un temps d'arrêt et $n \in \mathbb{N}$ alors $T \wedge n$ est un temps d'arrêt est $Y_{T\wedge n}$ est $F_{T\wedge n}$-mesurable donc F_n-mesurable.

2.2.2 Théorème d'arrêt

<u>Théorème 2.1</u> *Soit $(X_n)_n$ une sur-martingale et soit T un temps d'arrêt. Alors la suite $X^T = (X_{T\wedge n})_n$ est une sur-martingale, en particulier, $\mathbb{E}(X_{T\wedge n}) \leqslant \mathbb{E}(X_0)$ pour tout n.*

Si $(X_n)_n$ est une martingale alors X^T aussi et $\mathbb{E}(X_{T\wedge n}) = \mathbb{E}(X_0)$ pour tout n.

Corollaire 2.1 *Si $(X_n)_n$ est une (sur)martingale et si T est bornée, alors $\mathbb{E}(X_T)(\leqslant) = \mathbb{E}(X_0)$.*

Démonstration. Si T est bornée alors $T \leqslant N$ pour $N \in \mathbb{N}$. Donc $X_T = X_{T\wedge N}$ et $\mathbb{E}(X_T) = \mathbb{E}(X_0)$.

Pour $n \geqslant 1$, $H_n = 1_{T\geqslant n} = 1 - 1_{T<n}$. C'est une suite prévisible et bornée. On a $X_{T\wedge n} = X_0 + (H \cdot X)_n$.

Le résultat découle donc du fait que $H \cdot X$ est une (sur)-martingale. $\blacksquare$

Exemple 2.4 Marche aléatoire $X_0 = 0$, $X_n = \sum_{i=1}^{n} Y_i$ avec $Y_i \sim B(\pm 1, \frac{1}{2})$ iid. $T = \inf\{n, X_n = 1\}$ est un temps d'arrêt fini p.s.

$\mathbb{E}(X_T) = 1$ et $\mathbb{E}(X_0) = 0$

2.3 Théorème d'arrêt – le retour (il n'était pas parti longtemps)

Il est légitime de demander quand on a $\mathbb{E}(X_T) = \mathbb{E}(X_0)$.

<u>THÉORÈME 2.2</u> *Soit $(X_n)_n$ une sur-martingale (resp. martingale) et soit T un temps d'arrêt. Sous chacune des conditions suivantes (resp. sauf la dernière), la variable X_T est intégrable et $\mathbb{E}(X_T) \leqslant \mathbb{E}(X_0)$ (resp. $=$).*

(i) *T est bornée*

(ii) *X_n est bornée indépendemment de n et $T < \infty$ p.s.*

(iii) *$\mathbb{E}(T) < \infty$ et il existe $k > 0$ tel que $|X_{n+1}(\omega) - X_n(\omega)| \leqslant k$ pour (presque) tout n, ω.*

(iv) *$X_n \geqslant 0$ pour tout n et T fini p.s.*

Démonstration.

(i) Déjà fait

(ii) Le temps $T \wedge n$ est borné donc d'après (i), $\mathbb{E}(X_{T\wedge n}) = \mathbb{E}(X_0)$ et on utilise la théorème de convergence dominée. En effet, $T < \infty$ p.s., $X_{T\wedge n} \to X_T$ p.s. et $|X|$ majoré.

(iii) On écrit $X_{T\wedge n} - X_0 = \displaystyle\sum_{k=0}^{T\wedge n-1} (X_{i+1} - X_i)$ donc $|X_{T\wedge n} - X_0| \leqslant (T \wedge n)k \leqslant Tk$.

Donc $\mathbb{E}|T_{T\wedge n} - X_0| \leqslant \mathbb{E}(T)k$.

(iv) $X_n \geqslant 0$ pour tout n et $T < \infty$ p.s.

Par Fatou, $\mathbb{E}(\liminf X_{T\wedge n}) \leqslant \liminf \mathbb{E}(T \wedge n) = \mathbb{E}(X_0)$. Ainsi, $\mathbb{E}(X_T) \leqslant \mathbb{E}(X_0)$. ∎

<u>THÉORÈME 2.3</u> *Soit X_n une martingale uniformément intégrable et T un temps d'arrêt alors $X_T = \mathbb{E}(X_\infty \mid F_T)$. En particulier, $\mathbb{E}(X_T) = \mathbb{E}(X_\infty) = \mathbb{E}(X_0)$.*

De plus, si $S \leqslant T$ est un temps d'arrêt, alors $X_S = \mathbb{E}(X_T \mid F_S)$.

Ici $X_T = \displaystyle\sum_{n=0}^{\infty} X_n 1_{T=n} + X_\infty 1_{T=\infty}$.

Démonstration. On vérifie que $X_t \in L^1$. On a :

$$\begin{aligned}
\mathbb{E}(|X_T|) &= \mathbb{E}\left(\sum_{n=0}^{\infty}|X_n|1_{T=n} + |X_\infty 1_{T=\infty}\right) \\
&= \sum_{n=0}^{\infty}\mathbb{E}(1_{T=n}|\mathbb{E}(X_\infty \mid F_n)|) + \mathbb{E}(|X_\infty 1_{T=\infty}) \\
&\leqslant \sum_{n=0}^{\infty}\mathbb{E}(1_{T=n}\mathbb{E}(|X_\infty| \mid F_n)) + \mathbb{E}(|X_\infty 1_{T=\infty}) \\
&= \sum_{n=0}^{\infty}\mathbb{E}(\mathbb{E}(|X_\infty|1_{T=n} \mid F_n)) + \mathbb{E}(|X_\infty 1_{T=\infty}) \\
&= \sum_{n=0}^{\infty}\mathbb{E}(|X_\infty|1_{T=n} + \mathbb{E}(|X_\infty|1_{T=\infty} = \mathbb{E}(|X_\infty|) < \infty
\end{aligned}$$

On vérifie maintenant que $X_T = \mathbb{E}(X_\infty \mid F_T)$. Soit $A \in F_T$. On a

$$\mathbb{E}(X_T 1_A) = \sum_{n=0}^{n=\infty}\mathbb{E}(X_n 1_{A\cap\{T=n\}}) = \sum_{n=0}^{n=\infty}\mathbb{E}(X_\infty 1_{A\cap\{T=n\}}) = \mathbb{E}(X_\infty 1_A)$$

Ensuite, si $S \leqslant T$,

$$X_S = \mathbb{E}(X_\infty \mid F_S) = \mathbb{E}(\mathbb{E}(X_\infty \mid F_T) \mid F_S) = \mathbb{E}(X_T \mid F_S)$$ ■

Exemple 2.5 $S_0 = k$, $S_n = \sum_{i=1}^{n} X_i$ avec X_i iid $\sim B(\pm 1, \frac{1}{2})$. On pose $T = \inf\{n, S_n = 0 \text{ ou } S_n = m\}$.

La martingale $(X_{T\wedge n})_n$ est bornée donc uniformément intégrable et on a $\mathbb{E}(S_T) = \mathbb{E}(S_0)$. Or $\mathbb{E}(S_T) = k$ donc $\mathbb{P}(S_T = m) = \frac{k}{m}$ et $\mathbb{P}(S_T = 0) = 1 - \frac{k}{m}$. On a aussi $\mathbb{E}(T) = k(m-k)$.

Si $X_i \sim B(\pm 1, p)$. $Z_n = (\frac{1-p}{p})^{S_n}$ est une martingale et $Z_{n\wedge T}$ est bornée donc uniformément intégrable.

On a $\mathbb{E}(Z_T) = \mathbb{E}(Z_0)$ d'où $\mathbb{P}(S_T = 0) + (\frac{1-p}{p})^m \mathbb{P}(S_T = m) = (\frac{1-p}{p})^k$ et $\mathbb{P}(S_T = 0) + \mathbb{P}(S_T = m) = 1$. On en déduit $\mathbb{P}(S_T = 0) = 1 - \mathbb{P}(S_T = m)$ et $\mathbb{P}(S_T = m) = \frac{(\frac{1-p}{p})^k - 1}{(\frac{1-p}{p})^m - 1}$.

Chapitre 3

Convergence de martingales

3.1 Convergence presque sûre

Soit $(x_n)_n \in \mathbb{R}^{\mathbb{N}}$, $a < b \in \mathbb{Q}$. On définit

$$\begin{aligned} S_1(x) &= \inf\{n \geqslant 0, x_n \leqslant a\} \\ T_1(x) &= \inf\{n \geqslant S_1, x_n \geqslant b\} \\ S_{k+1}(x) &= \inf\{n \geqslant T_k, x_n \leqslant a\} \\ T_{k+1}(x) &= \inf\{n \geqslant S_k, x_n \geqslant b\} \end{aligned}$$

On définit aussi $N_n([a,b],x) = \sum_{k=1}^{n} 1_{T_k \leqslant n}$ et $N_\infty([a,b],x) = \sum_{k=1}^{\infty} 1_{T_k < \infty} = \sup\{k, T_k < \infty\}$.

Lemme 3.0.1
x converge dans $\overline{\mathbb{R}}$ ssi pour tout a, b, $N_\infty([a,b],x) < \infty$.

Démonstration. x diverge ssi $\liminf x_n < \limsup x_n$ ssi il existe $(a,b) \in \mathbb{Q}^2$ tel que $\liminf x_n \leqslant a < b \leqslant \limsup x_n$ ssi il existe $a, b \in \mathbb{Q}$ tel que $N_\infty([a,b],x) = \infty$. ∎

3.1.1 Le cas positif

Proposition 3.1 Soit $(X_n)_n$ une martingale positive et $(a,b) \in \mathbb{Q}^2$.
Alors $\mathbb{P}(N_\infty([a,b],X) \geqslant k) \leqslant (\frac{a}{b})^k$.
En particulier, $N_\infty([a,b],X)$ est fini p.s.

Démonstration. On remarque que $\{N_\infty([a,b],X) \geqslant k\} = \{T_k < \infty\}$.
Ensuite, comme $\{S_k < \infty\} \subset \{T_{k-1} < \infty\}$ donc il suffit de montrer que

$$\mathbb{P}(T_k < \infty) \leqslant \frac{a}{b}\mathbb{P}(S_k < \infty)$$

$T_k \wedge n$ et $S_k \wedge n$ sont des temps d'arrêt bornés donc par le théorème d'arrêt, $\mathbb{E}(X_{T_k \wedge n}) = \mathbb{E}(X_{S_k \wedge n}) = \mathbb{E}(X_0)$. Ainsi,

$$\mathbb{E}(X_{T_k} 1_{T_k \leqslant n}) + \mathbb{E}(X_n 1_{T_k > n}) = \mathbb{E}(X_{S_k} 1_{S_k \leqslant n}) + \mathbb{E}(X_n 1_{S_k > n})$$

D'où

$$b\mathbb{P}(T_k \leqslant n) \leqslant \mathbb{E}(X_{T_k} 1_{T_k \leqslant n}) + \mathbb{E}(X_n 1_{S_k \leqslant n < T_k}) = \mathbb{E}(X_{S_k} 1_{S_k \leqslant n}) \leqslant a\mathbb{P}(S_k \leqslant n)$$

D'où le résultat.

En particulier, $\sum\limits_{k=1}^{\infty} \mathbb{P}(T_k < n) < \infty$ donc par Borel-Cantelli,

$$\mathbb{P}(T_k < \infty \text{ infiniment souvent}) = 0$$

donc X ne traverse plus $[a, b]$ à partir d'un certain rang. ∎

Corollaire 3.1 *Soit $(X_n)_n$ une martingale positive. Alors X converge p.s. vers une variable X_∞ intégrable.*

Démonstration. La convergence découle de la proposition et du lemme précédents. On a

$$\mathbb{E}(X_\infty) \leqslant \liminf \mathbb{E}(X_n) = E(X_0) < \infty$$ ∎

Exemple 3.1

- Soit $S_n = \sum X_i$ avec $X_i \sim B(\pm 1, p)$ iid. $Y_n = (\frac{1-p}{p})^{S_n}$ est une martingale positive donc Y_n converge p.s.
- $X_n = \prod\limits_{i=1}^{n} X_i$ avec $Y_i \geqslant \delta > 0$, $\mathbb{E}(Y_1) = 1$. X_n converge p.s.
- Urne de Pòlya : $X_n = \frac{S_n}{n+2}$ est la proportion de boules rouges est une martingale positive.

3.1.2 Cas borné dans L^1

Lemme 3.0.2 des montées de Doob
Soit $(X_n)_n$ une sous-martingale et $a < b \in \mathbb{Q}$. Pour tout $n \geqslant 0$, on a

$$(b - a)\mathbb{E}(N_n([a, b], X)) \leqslant \mathbb{E}((X_n - a)^+ - (X_0 - a)^+)$$

Démonstration. Posons $Y_n = (X_n - a)^+$. C'est l'image de X_n par l'application convexe $x \mapsto (x - a)^+$ donc c'est une sous-martingale.

Posons $H_n = \sum_{k=1}^{\infty} 1_{S_k < n \leqslant T_k}$. C'est une suite prévisible car $\{S_k < n \leqslant T_k\} = \{S_k \leqslant n-1\} \setminus \{T_k \leqslant n-1\}$. On a

$$(H \cdot Y)_n = \sum_{k=1}^{N_n} (Y_{T_k} - Y_{S_k}) + \underbrace{1_{S_{N_n+1} < n}(Y_n - Y_{S_{N_n+1}})}_{\geqslant 0}$$

puisque sur $\{S_{N_n+1} < n\}$ on a $Y_{S_{N_n+1}} = 0$ donc $Y_n \geqslant 0$. On a donc

$$(HY)_n \geqslant \sum_{i=1}^{N_n} (Y_{T_k} - Y_{S_k}) \leqslant N_n(b-a)$$

En particulier, $\mathbb{E}((HY)_n) \geqslant (b-a)\mathbb{E}(N_n)$.

La suite $K_n = 1 - H_n$ est prévisible bornée et $\mathbb{E}((KY)_n) \geqslant \mathbb{E}((KY)_0) \neq 0$. Finalement

$$(b-a)\mathbb{E}(N_n) \leqslant \mathbb{E}((HY)_n) \leqslant \mathbb{E}((H+K)Y) = \mathbb{E}(Y_n - Y_0)$$

ie $(b-a)\mathbb{E}(N_n) \leqslant \mathbb{E}(Y_n - Y_0)$. ■

Remarque 3.1 Si (X_n) est une sur-martingale, on a un résultat similaire.

<u>Théorème 3.1</u> *Soit $(X_n)_n$ une sous-martingale telle que $\sup_n \mathbb{E}((X_n)^+) < \infty$ ie $(X_n)_n$ est bornée dans L^1. Alors (X_n) converge p.s. vers une variable $X_\infty \in L^1$.*

Démonstration. Soit $a < b \in \mathbb{Q}$.

$$(b-a)\mathbb{E}(N_n([a,b],X)) \leqslant \mathbb{E}((X_n - a)^+) \leqslant |a| + \mathbb{E}((X_n)^+) \leqslant |a| + \sup_n \mathbb{E}((X_n)^+)$$

Quand $n \to +\infty$, on trouve $(b-a)\mathbb{E}(N_\infty([a,b],X)) < \infty$. En particulier, $N_\infty([a,b],X)$ est fini p.s.

On peut permuter p.s. et $\forall a, b$ donc p.s., $\forall a < b$, $N_\infty([a,b],X) < \infty$ ie X_n converge p.s.

De plus, par Fatou, $\mathbb{E}(|X_\infty|) \leqslant \liminf \mathbb{E}(|X_n|) \leqslant \sup_n \mathbb{E}(|X_n|) < \infty$. ■

Remarque 3.2 On a utilisé que $\sup_n \mathbb{E}(X_n^+) < \infty$ ssi $\sup_n \mathbb{E}(|X_n|) < \infty$. Montrons-le : on a

$$E(X_0) \leqslant \mathbb{E}(X_n) = \mathbb{E}(X_n^+) - E(X_n^-)$$

Donc $E(X_n^-) \leqslant \mathbb{E}(X_n^+) - \mathbb{E}(X_0)$ et

$$\sup_n \mathbb{E}(X_n^-) \leqslant \sup_n \mathbb{E}(X_n^+) - \mathbb{E}(X_0)$$

COROLLAIRE 3.2 *Si X est une sur-martingale positive, alors elle converge p.s. et sa limite est L^1. De plus, $\mathbb{E}(X_\infty \mid F_n) \leqslant X_n$ p.s.*

Démonstration. Il suffit d'appliquer le théorème précédente à $-X_n$. ∎

Exemple 3.2 $S_n = \sum_{i=1}^{n} X_i$ avec $X_i \sim B(\pm 1, \frac{1}{2})$ iid avec $S_0 = 1$.

Soit $T = \inf\{n, S_n = 0\}$. La suite $(S_{T\wedge n})_n$ est une martingale positive qui converge donc p.s. vers S_∞.

Sur $\{T = \infty\}$, on a $|S_{T\wedge(n+1)} - S_{T\wedge n}| = |S_{n+1} - S_n| = 1$ donc $\mathbb{P}(T = \infty) = 0$ car $(S_{T\wedge n})_n$ converge.

On vérifie $S_\infty = 0$ et $0 = \mathbb{E}(S_T) \neq \mathbb{E}(S_0) = 1$ donc, en particulier, $S_{T\wedge n}$ ne converge pas dans L^1.

3.2 Convergence dans L^1

3.2.1 Martingales fermées

<u>THÉORÈME 3.2</u> *Soit $(X_n)_n$ une martingale. Les conditions suivantes sont équivalentes :*

(i) *X converge p.s. dans L^1 vers $X_\infty \in L^1$*

(ii) *il existe $Z \in L^1$ tel que $X_n = \mathbb{E}(Z \mid F_n)$*

Si c'est le cas, on peut prendre $Z = X_0$ et on dit que la martingale est fermée.

Démonstration.

(i) ⇒ (ii) On sait que pour tout $m > n$, $\mathbb{E}(X_m \mid F_n) = X_n$ p.s.
Par ailleurs, $Y \mapsto \mathbb{E}(Y \mid F_n)$ est une contraction de L^1 donc quand $m \to +\infty$, on a $\mathbb{E}(X_\infty \mid F_n) = X_n$ p.s.

(ii) ⇒ (i) S'il existe $Z \in L^1$ tel que $X_n = \mathbb{E}(Z \mid F_n)$ alors $(X_n)_n$ est bornée dans L^1 puisque $\sup \mathbb{E}(|X_n|) \leqslant \mathbb{E}(|Z|)$. D'après le cours, X_n converge p.s. vers $X_\infty \in L^1$.
Si Z est bornée, la convergence L^1 découle du théorème de convergence dominée.
Sinon, soit $\varepsilon > 0$ et M assez grand pour que $\mathbb{E}(|Z - Z1_{|Z|\leqslant M}|) \leqslant \varepsilon$. Pour $n \in \mathbb{N}$,

$$\begin{aligned}\mathbb{E}(|X_n - \mathbb{E}(Z1_{|Z|\leqslant M} \mid F_n)|) &= \mathbb{E}(|\mathbb{E}(Z - Z1_{|Z|\leqslant M} \mid F_n)|) \\ &\leqslant \mathbb{E}(|Z - Z1_{|Z|\leqslant M}|) \leqslant \varepsilon\end{aligned}$$

D'après le cas borné, $\mathbb{E}(Z1_{|Z|\leqslant M} \mid F_n)$ converge dans L^1 donc pour tout m, n assez grands,

$$\mathbb{E}(|\mathbb{E}(Z1_{|Z|\leqslant M} \mid F_n) - \mathbb{E}(Z1_{|Z|\leqslant M} \mid F_n)|) \leqslant \varepsilon$$

En conbinant ces deux résultats, on trouve pour tout m, n assez grands, $\mathbb{E}(|X_m - X_n|) \leqslant 3\varepsilon$, donc (X_n) est de Cauchy dans L^1. ■

Corollaire 3.3 *Soit $Z \in L^1$ et $X_n = \mathbb{E}(Z \mid F_n)$. Alots $(X_n)_n$ converge p.s. et dans L^1. De plus, $X_\infty = \mathbb{E}(Z \mid F_\infty)$ où $F_\infty = \sigma\left(\bigcup_{n=0}^{\infty} F_n\right)$.*

Démonstration. X_∞ est F_∞-mesurable. En effet, pour tout n, X_n est F_∞-mesurable donc X_∞ aussi.

Si $A \in F_n$, $\mathbb{E}(Z1_A) = \mathbb{E}(X_n 1_A) = \mathbb{E}(X_\infty 1_A)$. Par classe monotone, on déduit que pour tout $A \in F_\infty$, $\mathbb{E}(Z1_A) = \mathbb{E}(X_\infty 1_A)$. ■

Exemple 3.3 On se place dans $([0,1], \mathscr{B}([0,1]), \lambda)$, $G_n = \sigma([\frac{i-1}{2^n}, \frac{i}{2^n}[, i \in [\![1, 2^n]\!])$.

Si $f \in L^1$ alors

$$f_n = \mathbb{E}(f \mid G_n) = \frac{1}{2^n}\sum_i \int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} f 1_{[\frac{i-1}{2^n}, \frac{i}{2^n}[} \,\mathrm{d}\lambda$$

qui converge p.s. et dans L^1 (étagée) vers

$$f_\infty = \mathbb{E}(f \mid F_\infty) = \mathbb{E}(f \mid \mathscr{B}([0,1])) = f$$

car $f \in L^1$.

Comment montrer qu'une martingale est fermée ?

3.2.2 Uniforme intégrabilité

Définition 3.1 Une famille $(X_i)_i$ de variables aléatoires L^1 est dite uniformément intégrable ssi $\sup_{i \in I} \mathbb{E}(|X_i| 1_{|X_i| > N}) \to 0$ quand $N \to +\infty$.

Remarque 3.3 Si $(X_i)_i$ est uniformément intégrable, alors $(X_i)_i$ est bornée dans L^1. En effet, il suffit de prendre N assez grand pour que le sup soit inférieur à 1 et on a alors $\sup \mathbb{E}(|X_i|) \leqslant N$.

Exemple 3.4
- Si $X \in L^1$, $\{X\}$ est uniformément intégrable.
- Si $X_1, \ldots, X_n \in L^1$ alors $(X_1, \ldots, X_n)$ est uniformément intégrable.
- Les variables dominées par $Z \in L^1$ ie $\{X, |X| \leqslant Z \text{ p.s.}\}$ sont uniformément intégrables.

- Une famille bornée dans L^p avec $p > 1$ est uniformément intégrable. En effet,

$$\begin{aligned}\mathbb{E}(|X_i|1_{|X_i|>N}) = \mathbb{E}\left(\frac{|X_i|}{|X_i|^p}|X_i|^p 1_{|X_i|>N}\right) &\leqslant N^{1-p}\mathbb{E}(|X_i|^p) \\ &\leqslant N^{1-p}\sup\mathbb{E}(|X_i|^p)\end{aligned}$$

Plus concrètement,

Proposition 3.2 Soit $(X_i)_i$ une famille bornée dans L^1. X est uniformément intégrable ssi

$$\forall\varepsilon > 0, \exists\delta > 0, (A \in F \text{ et } \mathbb{P}(A) < \delta) \Rightarrow (\mathbb{E}(|X_i|1_A) \leqslant \varepsilon\forall i \in I)$$

Démonstration.

$\Rightarrow$ Soit $\varepsilon > 0$, et N assez grand pour que $\sup\mathbb{E}(|X_i|1_{|X_i|\geqslant N}) < \frac{\varepsilon}{2}$.
Si on pose $\delta = \frac{\varepsilon}{2N}$, on a, si $\mathbb{P}(A) < \delta$:

$$\mathbb{E}(|X_i|1_A) = \mathbb{E}(|X_i|1_{A\cap\{|X_i|>N\}}) + \mathbb{E}(|X_i|1_{A\cap\{|X_i|\leqslant N\}}) \leqslant \frac{\varepsilon}{2} + N\mathbb{P}(A) \leqslant \varepsilon$$

$\Leftarrow$ Par l'inégalité de Markov, $\mathbb{P}(|X_i| > N) \leqslant \frac{\mathbb{E}(|X_i|)}{N} \leqslant \frac{\sup\mathbb{E}(|X_i|)}{N}$.
Soit $\varepsilon > 0$ et δ associé. Si N est assez grand pour que $\frac{\sup\mathbb{E}(|X_i|)}{N} < \delta$ alors pour tout i, $\mathbb{E}(|X_i|1_{|X_i|>N}) \leqslant \varepsilon$ (prendre $A = \{|X_i| > N\}$). Donc X est uniformément intégrable. ∎

Corollaire 3.4 *Soient $X \in L^1(\Omega, F, \mathbb{P})$, et I l'ensemble des sous-tribus de F, $X_G = \mathbb{E}(X \mid G)$.*
Alors $(X_G)_{G\in I}$ est uniformément intégrable.

Démonstration. On va utiliser la caractérisation précédente.

Soit $\varepsilon > 0$, comme X est uniformément intégrable, il existe $\delta > 0$ tel que pour tout $A \in F$ de probabilité au plus δ, on ait $\mathbb{E}(|X|1_A) \leqslant \varepsilon$. Par Markov, on a $\mathbb{P}(|\mathbb{E}(X \mid G)| \geqslant N) \leqslant \frac{\mathbb{E}(|X|)}{N}$.

Soit N assez grand pour que $\frac{\mathbb{E}(|X|)}{N} < \delta$. On a

$$\begin{aligned}\mathbb{E}(|\mathbb{E}(X \mid G)|1_{|E(X|G)|>N}) &\leqslant \mathbb{E}(\mathbb{E}(|X| \mid G)1_{|\mathbb{E}(X|G)|>N}) \\ &= \mathbb{E}(|X|1_{|\mathbb{E}(X|G)|>N}) \leqslant \varepsilon\end{aligned}$$

en prenant $A = \{|\mathbb{E}(X \mid G)| > N\}$.

Cette majoration est indépendante de G donc on peut passer au sup et on a le résultat. ∎

THÉORÈME 3.3 *Soit $(X_n)_n$ une suite de variables qui converge en probabilité vers X_∞. $(X_n)_n$ est uniformément intégrable ssi elle converge dans L^1 vers X_∞.*

Remarque 3.4 C'est mieux que le théorème de convergence dominée car on a une équivalence.

Démonstration.

$\Leftarrow$ Si $(X_n)_n$ converge dans L^1 alors X est bornée dans L^1 et si $\varepsilon > 0$, il existe N assez grand de sorte que pour tout $n \geqslant N$, $\mathbb{E}(|X_n - X_N|) \leqslant \frac{\varepsilon}{2}$. La famille $(X_0, \ldots, X_N)$ est uniformément intégrable donc il existe $\delta > 0$ tel que pour tout $i \in [\![0, N]\!]$, $\mathbb{E}(|X_i|1_A) \leqslant \frac{\varepsilon}{2}$ si $\mathbb{P}(A) < \delta$.
On a alors pour tout $n > N$,

$$\mathbb{E}(|X_n|1_A) = \mathbb{E}(|X_n - X_N|1_A) + \mathbb{E}(|X_n|1_A) \leqslant \varepsilon$$

$\Rightarrow$ Si X est uniformément intégrable, alors $(X_m - X_n)_{m,n}$ l'est aussi. Soit $\varepsilon > 0$. Pour N assez grand, $\mathbb{E}(|X_m - X_n|1_{|X_m - X_n| > N}) \leqslant \varepsilon$.

$$\begin{aligned}\mathbb{E}(|X_m - X_n|)) &= \mathbb{E}(|X_m - X_n|1_{|X_m - X_n| < \varepsilon}) \\ &+ \mathbb{E}(|X_m - X_n|1_{\varepsilon \leqslant |X_m - X_n| \leqslant N}) \\ &+ \mathbb{E}(|X_m - X_n|1_{|X_m - X_n| > N}) \\ &\leqslant \varepsilon + N\mathbb{P}(|X_m - X_n| \geqslant \varepsilon) + \varepsilon\end{aligned}$$

Or

$$\mathbb{P}(|X_m - X_n| \geqslant \varepsilon) \leqslant \mathbb{P}\left(|X_m - X_\infty| \leqslant \frac{\varepsilon}{2}\right) + \mathbb{P}\left(|X_n - X_\infty| \geqslant \frac{\varepsilon}{2}\right) \to 0$$

Donc $\limsup \mathbb{E}(|X_m - X_n|) \leqslant 2\varepsilon$ donc X est de Cauchy donc converge dans L^1. ■

THÉORÈME 3.4 *Soit $(X_n)_n$ une martingale. Il y a équivalence entre*

(i) *X converge p.s. et dans L^1 vers une limite L^1*

(ii) *X est fermée*

(iii) *X est uniformément intégrable*

3.3 Convergence L^p avec $p > 1$

3.3.1 Inégalités maximales et convergence L^p

Lemme 3.4.1
Soit $(X_n)_n$ une sous-martingale et $S \leqslant T$ des temps d'arrêt bornés. Alors $\mathbb{E}(X_S) \leqslant \mathbb{E}(X_T)$.

Remarque 3.5 On a déjà vu le cas $S = 0$.

Démonstration. On pose $H_n = 1_{S<n\leqslant T} = 1_{S\leqslant n-1} - 1_{T\leqslant n-1} \in F_{n-1}$.
Si N est tel que $S \leqslant T \leqslant N$, on a

$$(HX)_N = X_T - X_S \text{ et } \mathbb{E}((HX)_N) \geqslant \mathbb{E}((HX)_0) = 0$$

■

Théorème 3.5 Inégalité maximale de Doob *Soit X_n une sous-martingale. Pour tout $a > 0$, on a :*

$$a\mathbb{P}(\sup_{k\in[\![0,n]\!]} X_k \geqslant a) \leqslant \mathbb{E}(X_n 1_{\sup_{k\in[\![0,n]\!]} X_k \geqslant a}) \leqslant \mathbb{E}(|X_n|)$$

Démonstration. Soit $T = \inf\{n, X_n \geqslant a\}$. On pose $A = \{\sup_{k\leqslant n} X_k \geqslant a\} = \{T \leqslant n\}$.
D'après le lemme précédent, comme n et $T \wedge n$ sont bornées, $\mathbb{E}(X_{T\wedge n}) \leqslant \mathbb{E}(X_n)$. On a

$$X_{T\wedge n} = X_T 1_{T\leqslant n} + X_n 1_{T>n} \geqslant a1_A + X_n 1_{A^c}$$

Donc $\mathbb{E}(X_n) \geqslant \mathbb{E}(X_{T\wedge n}) \geqslant a\mathbb{P}(A) + \mathbb{E}(X_n 1_{A^c})$ id $\mathbb{E}(X_n(1 - 1_{A_c})) \geqslant a\mathbb{P}(A)$.
D'où $a\mathbb{P}(\sup X_k \geqslant a) \leqslant \mathbb{E}(X_n 1_{\sup X_k \geqslant a})$. ■

Théorème 3.6 Inégalité maximale L^p *Soit $p > 1$ et X_n une sous-matingale positive. On pose $\widetilde{X}_n = \sup_{k\in[\![0,n} X_k$. Alors pour tout $n \geqslant 1$,*

$$\mathbb{E}(\widetilde{X}_n^p) \leqslant \left(\frac{p}{p-1}\right)^p \mathbb{E}(X_n^p)$$

En particulier, si Y_n est une martingale et $Y_n^* = \sup_{k\in[\![0,n]\!]} |Y_k|$ alors

$$\mathbb{E}((Y_n^*)^p) \leqslant \left(\frac{p}{p-1}\right)^p \mathbb{E}(|Y_n|^p)$$

Démonstration.

- Le second point est une conséquence du premier ($X_n = |Y_n|$)
- On peut supposer que pour tout n, $\mathbb{E}(X_n^p) < \infty$, sinon on se retrouve avec $\infty \leqslant \infty$
- Par Jensen, $\mathbb{E}(X_k^p) = \mathbb{E}(\mathbb{E}(X_n \mid F_k)^p) \leqslant \mathbb{E}(\mathbb{E}(X_n^p \mid F_k)) = \mathbb{E}(X_n^p)$ pour $k \leqslant n$ et $\mathbb{E}(\widetilde{X}_n^p) < \infty$.
- D'après l'inégalité maximale précédente, si $a > 0$,

$$a\mathbb{P}(\widetilde{X}_n \geqslant a) \leqslant \mathbb{E}(X_n 1_{\widetilde{X}_n \geqslant a})$$

Donc $a^{p-1}\mathbb{P}(\widetilde{X}_n \geqslant a) \leqslant a^{p-2}\mathbb{E}(X_n 1_{\widetilde{X}_n \geqslant a})$. On intègre en a de 0 à ∞.

- Le premier bout donne

$$\begin{aligned}\int_0^{\infty} a^{p-1}\mathbb{P}(\widetilde{X}_n \geqslant a)\,\mathrm{d}a &= \mathbb{E}\left(\int_0^{\infty} a^{p-1}1_{\widetilde{X}_n\geqslant a}\,\mathrm{d}a\right)\\ &= \mathbb{E}\left(\int_0^{\widetilde{X}_n} a^{p-1}\,\mathrm{d}a\right) = \frac{1}{p}\mathbb{E}(\widetilde{X}_n^p)\end{aligned}$$

- Et le deuxième :

$$\int_0^{\infty} a^{p-2}\mathbb{E}(X_n 1_{\widetilde{X}_n\geqslant a})\,\mathrm{d}a = \frac{1}{p-1}\mathbb{E}(X_n\widetilde{X}_n^{p-1}) \leqslant \frac{1}{p-1}\mathbb{E}(X_n^p)^{\frac{1}{p}}\mathbb{E}(\widetilde{X}_n^p)^{\frac{p-1}{p}}$$

En combinant les trois points précédents, on obtient le résultat. ■

<u>Théorème 3.7</u> Convergence L^p *Soit $(X_n)_n$ une martingale. Supposons qu'il existe $p > 1$ tel que $\sup_n \mathbb{E}(|X_n|^p) < \infty$. Alors X_n converge p.s. et dans L^p vers $X_\infty \in L^p$ avec $\mathbb{E}(|X_\infty|^p) = \sup_n \mathbb{E}(|X_n|^p)$ et*

$$\mathbb{E}((X_n^*)^p) \leqslant \left(\frac{p}{p-1}\right)^p \mathbb{E}(|X_\infty|^p)$$

Démonstration. Si $(X_n)_n$ est bornée dans L^p alors elle est bornée dans L^1 donc on a la convergence p.s. vers $X_\infty \in L^1$.

D'après l'inégalité maximale L^p, on a

$$\mathbb{E}((X_n^*)^p) \leqslant \left(\frac{p}{p-1}\right)^p \mathbb{E}(|X_n|^p) \leqslant \left(\frac{p}{p-1}\right)^p \sup_n \mathbb{E}(|X_n|^p) < \infty$$

En passant à la limite, on a $\mathbb{E}((X_\infty^*)^p) < \infty$.

Pour tout n, $|X_n| \leqslant X_\infty^*$ et $|X_n|^p \leqslant |X_\infty^*|^p \in L^1$ donc par convergence dominée,

$$\mathbb{E}(|X_\infty|^p) \leqslant \mathbb{E}(|X_\infty^*|^p)$$

et on a la convergence dans L^p. Comme $|X_n|^p$ est une sous-martingale, la suite $\mathbb{E}(|X_n|^p)$ est croissante et on a :

$$\mathbb{E}(|X_\infty|^p) = \lim_{n\to+\infty} \mathbb{E}(|X_n^p|) = \sup_n \mathbb{E}(|X_n|^p)$$ ■

3.3.2 Cas des martingale de carré intégrable

Soit (M_n) une martingale de carré intégrable ie $\mathbb{E}(M_n^2) < \infty$ pour tout n. On sait que si $n \geqslant m$, $M_m = \mathbb{E}(M_n \mid F_m)$ et $M_n - M_m \perp L^2(F_m)$.

Ainsi, si $k \leqslant l \leqslant m \leqslant n$, $\langle M_n - M_m, M_l - M_k\rangle = 0$.

On a $M_n = M_0 + \sum_{k=1}^{n}(M_k - M_{k-1})$ qui est une décomposition en termes orthogonaux. Donc (Pythagore) :

$$\mathbb{E}(M_n^2) = \mathbb{E}(M_0^2) + \sum_{k=1}^{n}\mathbb{E}((M_k - M_{k-1})^2)$$

<u>Théorème 3.8</u> *Soit $(M_n)_n$ une martingale L^2. Alors (M_n) est bornée dans L^2 ssi $\sum_{k=1}^{\infty}\mathbb{E}((M_k - M_{k-1})^2) < \infty$. Auquel cas, (M_n) converge p.s. et dans L^2 vers $M_\infty \in L^2$.*

Démonstration. C'est la formule d'avant. La suite, c'est le cours d'avant. ∎

3.3.3 Décomposition de Doob(-Meyer)

<u>Théorème 3.9</u> *Soit (X_n) une suite de variables adaptée et L^1. Alors X_n admet la décomposition $X_n = X_0 + M_n + A_n$ où M_n est une martingale nulle en zéro et A_n est une suite prévisible nulle en zéro.*

Cette décomposition est unique au sens où si $X_n = X_0 + \widetilde{M}_n + \widetilde{A}_n$, alors $\mathbb{P}(M_n = \widetilde{M}_n, A_n = \widetilde{A}_n, \forall n) = 1$.

La suite $(X_n)_n$ est une sous-martingale ssi $(A_n)_n$ est croissante, au sens où $\mathbb{P}(A_n \leqslant A_{n+1} \forall n) = 1$.

Démonstration. Si (X_n) admet une telle décomposition, on a $\mathbb{E}(X_n - X_{n-1} \mid F_{n-1}) = A_n - A_{n-1}$ ie on a $A_n = \sum_{k=1}^{n}\mathbb{E}(X_k - X_{k-1} \mid F_{k-1})$.

En définissant A_n par la formule précédente et $M_n = X_n - X_0 - A_n$, on trouve bien que A_n est prévisible et M_n est une martingale nulles toutes deux en 0. ∎

Définition 3.2 Soit M_n une martingale L^2 nulle en 0 pour simplifier. Alors (M_n^2) est une sous-martingale, qui admet donc une décomposition de Doob : $M_n^2 = A_n + N_n$ avec A prévisible croissante et N une martingale.

La suite A est appelée compensateur de M. On la note souvent $\langle M_n \rangle_n$.

Proposition 3.3 Comme $\mathbb{E}(M_n^2) = \mathbb{E}(\langle M \rangle_n)$, M_n est bornée dans L^2 ssi $\mathbb{E}(\langle M \rangle_\infty) < \infty$ où $\langle M \rangle_\infty = \lim \langle M \rangle_n$ (limite croissante).

<u>Théorème 3.10</u> *Soit M_n une martingale L^2.*
- *$\lim_{n\to+\infty} M_n(\omega)$ existe dès lors que $\langle M \rangle_\infty(\omega) < \infty$*
- *Supposons que les accroissements de M_n sont bornés, c'est-à-dire que $|M_n(\omega) - M_{n-1}(\omega)| \leqslant K$ pour (presque) tout n et ω. Alors $\langle M \rangle_\infty$ exists dès lors que $\lim_{n\to+\infty} M_n(\omega)$ existe.*

Démonstration.

- $A_n = \langle M \rangle_n$ est prévisible donc pour tout $k > 0$, $S_k = \inf\{n, A_{n+1} > k\}$ est un temps d'arrêt. De plus, $A_n^{S_k} = A_{S_k \wedge n}$ est prévisible. En effet, si $B \in \mathscr{B}(\mathbb{R})$,

$$\{A_{S_k \wedge n} \in B\} = \underbrace{\bigcup_{r=0}^{n-1} \{S_k = r \text{ et } A_r \in B\}}_{\in F_{n-1}} \cup \underbrace{\{A_n \in B \text{ et } (S_k \leqslant n-1)^c\}}_{\in F_n}$$

  Comme $(M^2)^{S_k} - A^{S_k} = (M^2 - A)^{S_k}$ est une martingale, on a $A^{S_k} = \langle M^{S_k} \rangle$. Par ailleurs, A^{S_k} est borné par k donc $\langle M^{S_k} \rangle \leqslant k$ et M^{S_k} est bornée dans L^2 donc converge p.s. et dans L^2.
  En outre, $\{A_\infty = \infty\} = \bigcup_k \{S_k = +\infty\}$ donc si $\omega \in \{A_\infty = \infty\}$, on a $S_k(\omega) = +\infty$ pour un certain k et $M_n^{S_k(\omega)} = M_n(\omega)$.
- On raisonne par l'absurde : si $\mathbb{P}(A_\infty = \infty \text{ et } \sup_n |M_n| < \infty) > 0$. Alors il existe $c > 0$ tel que $\mathbb{P}(A_\infty = \infty \text{ et } T(c) = \infty) > 0$ où $T(c) = \inf\{n, |M_n| > c\}$. Le temps $T(c)$ est un temps d'arrêt. Par le théorème d'arrêt (appliqué à $N = M^2 - A$ en $\tau = T(c) \wedge n$), $\mathbb{E}(M^2_{T(x)\wedge n}) = \mathbb{E}(A_{T(c)\wedge n})$.
  Par ailleurs, $M^2_{T(c)\wedge n} \leqslant c + K$. L'égalité précédente donne $\mathbb{E}(T(c) \wedge n) \leqslant c + L$ pour tout n et avec $n \to +\infty$, $\mathbb{E}(A_{T(c)}) \leqslant c + K$. Ceci est incompatible avec $\mathbb{P}(A_{T(c)} = +\infty) > 0$. On conclut donc que $\mathbb{P}(A_\infty = \infty) \text{ et } \sup_n |M_n| < \infty) = 0$. ■

3.3.4 Théorème classiques pour les martingales

Lemme 3.10.1 Césaro
Soit b_n une suite croissante de réels positifs qui tend vers $+\infty$ et v_n une suite réelle convergeant vers v_∞. Alors

$$\frac{1}{b_n} \sum_{k=1}^{n} (b_k - b_{k-1}) v_k \to v_\infty$$

Lemme 3.10.2 Kronecker
Soit b_n une suite croissante positive qui diverge vers $+\infty$ et x_n une suite réelle. Alors si $S_n = x_1 + \ldots + x_n$, on a

$$\sum_{k=1}^{n} \frac{x_k}{b_k} \text{ cv } \Rightarrow \frac{S_n}{b_n} \to 0$$

Démonstration. On pose $u_n = \sum_{k=1}^{n} \frac{x_k}{b_k}$ et on suppose $u_n \to u_\infty \in \mathbb{R}$. On remarque que $u_n - u_{n-1} = \frac{x_n}{b_n}$ et

$$S_n = \sum_{k=1}^{n} b_k(u_k - u_{k-1}) = u_n b_n - \sum_{k=1}^{n} u_k(b_k - b_{k-1})$$

Donc $\frac{S_n}{b_n} = u_n - \frac{1}{b_n}\sum_{k=1}^{n} u_k(b_k - b_{k-1}) \to u_\infty - u_\infty = 0$ par Césaro. ■

<u>Théorème 3.11</u> *Soit $(M_n)_n$ ine martingale de carré intégrable nulle en zéro. Alors sur l'ensemble $\{\langle M\rangle_\infty = \infty\}$ on a $\frac{M_n}{\langle M\rangle_n} \to 0$ p.s.*

Démonstration. Soit $A = \langle M\rangle$ le compensateur de M. La suite $H_n = (1 + A_n)^{-1}$ est prévisible et bornée par 1.

$W_n = \sum_k \frac{M_k - M_{k-1}}{1 + A_k} = (HM)_n$ est une martingale et

$$\begin{aligned}
\langle W\rangle_n - \langle W\rangle_{n-1} &= \mathbb{E}(\frac{(M_n - M_{n-1})^2}{1 + A_n} \mid F_{n-1}) \\
&= \frac{1}{1 + A_n}\mathbb{E}((M_n - M_{n-1})^2 \mid F_{n-1}) \\
&= \frac{1}{1 + A_n}\mathbb{E}(M_n^2 - M_{n-1}^2 \mid F_{n-1}) \\
&= \frac{A_n - A_{n-1}}{1 + A_n} \leqslant \frac{A_{n-1}}{1 + A_{n-1}} - \frac{A_n}{1 + A_n}
\end{aligned}$$

En sommant, on trouve $\langle W\rangle_n \leqslant 1$ p.s. Donc W_n converge p.s. et dans L^2. Par Kronecker, on a $\frac{M_n}{1+A_n} \to 0$ donc $\frac{M_n}{A_n}$ aussi. ■

<u>Théorème 3.12</u> Admis *Sous les mêmes hypothèses sur $\{\langle M\rangle_\infty = \infty\}$, on a $\frac{M_n}{\sqrt{\langle M\rangle_n}} \to \mathcal{N}(0,1)$.*

3.4 Martingales rétrogrades

<u>Définition 3.3</u> Une filtration rétrograde est une famille (F_n) indexée par les entiers négatifs de sous-tribus de F et telle que pour tout $n \leqslant m$, $F_n \subset F_m$. On notera $F_{-\infty} = \bigcap_{n\leqslant 0} F_n$.

<u>Définition 3.4</u> Une suite $(X_n)_{n\in -\mathbb{N}}$ de variables adaptées intégrables est une martingale (resp. sous/sur) ssi pour tout $n \leqslant m$, $\mathbb{E}(X_m \mid F_n) = X_n$ (resp. $\geqslant, \leqslant$).

<u>THÉORÈME 3.13</u> *Soit* $(X_n)_{n\in-\mathbb{N}}$ *une sur-martingale rétrograde bornée dans* L^1 *telle que* $\sup_n \mathbb{E}(|X_n|) < \infty$.

Alors X_n *est uniformément intégrable et converge p.s. et* L^1 *en* $-\infty$ *vers* X_∞ *tel que* $\mathbb{E}(X_n \mid F_{-\infty}) \leqslant X_\infty$.

COROLLAIRE 3.5 *Soit* $(X_n)_{n\in-\mathbb{N}}$ *une martingale rétrograde. Alors* $(X_n)_n$ *converge p.s. et dans* L^1 *vers une variable* X_∞.

Démonstration. $X_n = \mathbb{E}(X_0 \mid F_n)$ pour tout $n \leqslant 0$. De plus, on applique le théorème, puisqu'on a $\sup \mathbb{E}(|X_n|) \leqslant \mathbb{E}(|X_0|)$. ∎

COROLLAIRE 3.6 *Soit* $Z \in L^1$ *et* (G_n) *une suite décroissante de tribus. On pose* $G_\infty = \bigcap_{n\geqslant 0} G_n$. *Alors* $\mathbb{E}(Z \mid G_n) \to \mathbb{E}(Z \mid G_\infty)$ *p.s. et dans* L^1.

Démonstration. Pour $n \in \mathbb{N}$, on pose $X_{-n} = \mathbb{E}(Z \mid G_n)$ et $F_n = G_n$. Alors X_n est une martingale rétrograde pour F_n et le théorème précédent conclut. ∎

Exemple 3.5 On peut redémontrer la loi des grands nombres.

$\mathbb{E}(X_1 \mid S_n)$ est S_n-mesurable donc il existe g mesurable telle que $\mathbb{E}(X_1 \mid S_n) = g(S_n)$. Comme les lois $L(X_1, S_n)$ et $L(X_k, S_n)$ sont égales pour $k \in [\![1, n]\!]$ donc pour tout h mesurable bornée,

$$\mathbb{E}(X, h(S_n)) = \mathbb{E}(X_k h(S_n)) = \mathbb{E}(g(S_n)h(S_n))$$

donc $\mathbb{E}(X_k \mid S_n) = g(S_n)$.

En sommant $S_n = \sum_{k=1}^{n} \mathbb{E}(X_k \mid S_n) = ng(S_n)$ ie $\mathbb{E}(X_1 \mid S_n) = g(S_n) = \frac{S_n}{n}$.

Par ailleurs, $\mathbb{E}(X_1 \mid S_n) = \mathbb{E}(X_1 \mid S_n, X_{n+1}, \ldots)$ car les X_i sont iid.

On pose $G_n = \sigma(S_n, X_{n+1}, \ldots) = \sigma(S_n, S_{n+1}, \ldots)$. C'est une suite décroissante de tribus et $g_\infty = \bigcap_n \sigma(S_n, S_{n+1}, \ldots)$ est triviale (loi 0-1 de Kolmogorov)

Donc

$$\frac{S_n}{n} = \mathbb{E}(X_1 \mid G_n) \to \mathbb{E}(X_1 \mid G_\infty) = \mathbb{E}(X_1)$$

dans L^1 et p.s.

Chapitre 4

Application des martingales

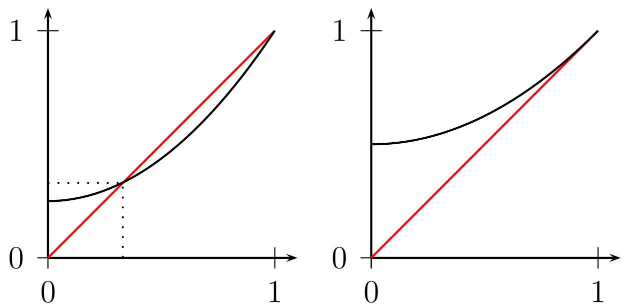
4.1 Galton-Watson

$Z_{n+1} = \sum_{k=1}^{Z_n} X_{n+1,k}$ et $Z_0 = 1$ avec $(X_{j,k})_{j,k}$ iid de loi $\sum_{k=0}^{\infty} p_k \delta_k$.

Si $m = \mathbb{E}(X_{1,1})$, $\mathbb{E}(Z_n) = m^n$. D'où trois cas :

- $m < 1 : \mathbb{E}(Z_n) \to 0$
- $m > 1 : \mathbb{E}(Z_n) \to +\infty$
- $m = 1 : \mathbb{E}(Z_n) = 1$

On considère $T = \inf\{n, Z_n = 0\}$. $\mathbb{P}(T < \infty)$ est le plus petit point fixe de g dans $]0, 1[$.



$Y_n := \frac{Z_n}{m^n}$ est une martingale positive dpnc converge p.s. vers X_∞.

Autrement dit, $Z_n \sim m^n Y_\infty$ sur $\{\omega, T(\omega) = \infty\}$. Précisons dans le cas critique. On va montrer que

- $n\mathbb{P}(Z_n > 0) \to \frac{2}{\sigma^2}$
- $\mathbb{E}(\frac{Z_n}{n} \mid Z_n > 0) \to \frac{\sigma^2}{2}$
- $\frac{Z_n}{n} \mid Z_n > 0$ converge en loi vers une exponentielle $\frac{2}{\sigma^2}$.

Démonstration.

- Soit $g=\sum_{k=0}^{\infty}p_ks^k$. Par Taylor à l'ordre 2 en 1, on a

$$g(s)=s+\frac{\sigma^2}{2}(1-s)^2+o((1-s)^2)$$

On a alors

$$\begin{aligned}\frac{1}{1-g(s)}-\frac{1}{1-s}&=\frac{\frac{\sigma^2}{2}(1-s)^2+o((1-s)^2)}{(1-g(s))(1-s)}\\&=\frac{\frac{\sigma^2}{2}+o(1)}{1-\frac{\sigma^2}{2}(1-s)+o(1-s)}=\frac{\sigma^2}{2}+o(1)\end{aligned}$$

Par ailleurs, $g_n=g\circ\ldots\circ g$ tend uniformément vers 1 en ∞ sur $[0,1]$.

$$\frac{1}{n}\left(\frac{1}{1-g_n(s)}-\frac{1}{1-s}\right)=\frac{1}{n}\sum_{k=1}^{n}\frac{1}{1-g_k(s)}-\frac{1}{1-g_{k-1}(s)}\to\frac{\sigma^2}{2}$$

On a $\mathbb{P}(Z_n>0)=1-g_n(0)$ et $\frac{1}{n}(\frac{1}{1-g_n(0)}-1)\to\frac{\sigma^2}{2}$. Donc $n\mathbb{P}(Z_n>0)\to\frac{2}{\sigma^2}$.

- $\mathbb{E}(\frac{Z_n}{n}\mid Z_n>0)=\frac{\mathbb{E}(Z_n1_{Z_n>0})}{n\mathbb{P}(Z_n>0)}$. Or $Z_n1_{Z_n>0}=Z_n$ donc

$$\mathbb{E}(\frac{Z_n}{n}\mid Z_n>0)=\frac{\mathbb{E}(Z_n)}{n\mathbb{P}(Z_n>0)}=\frac{\mathbb{E}(Z_0)}{n\mathbb{P}(Z_n>0)}\to\frac{\sigma^2}{2}$$

- On a aussi

$$\begin{aligned}&\mathbb{E}(\mathrm{e}^{-t\frac{Z_n}{n}}\mid Z_n>0)=\frac{\mathbb{E}(\mathrm{e}^{-\frac{tZ_n}{n}}1_{Z_n>0})}{\mathbb{P}(Z_n>0)}\\&=\frac{\mathbb{E}(\mathrm{e}^{-tZ_n}n)-\mathbb{E}(\mathrm{e}^{-t\frac{Z_n}{n}}1_{Z_n=0})}{\mathbb{P}(Z_n>0)}=\frac{g_n(\mathrm{e}^{-\frac{t}{n}})-g_n(0)}{1-g_n(0)}\\&=1-\frac{1}{n(1-g_n(0))}\left(\frac{1}{n}\left(\frac{1}{1-g_n(S_n)}-\frac{1}{1-S_n}\right)+\frac{1}{n(1-S_n)}\right)^{-1}\\&\to1-\frac{\sigma^2}{2}\left(\frac{\sigma^2}{2}+\frac{1}{t}\right)^{-1}=\frac{1}{1+\frac{\sigma^2}{2}t}\end{aligned}$$

■

On va s'occuper maintenant du cas sur-critique ($m>1$ et $\sigma^2<\infty$). On montre que $Y_n\to Y_\infty$ p.s. et L^2.

$$\mathbb{E}(Y_n^2)=\frac{\sigma^2}{m^2-m}-\frac{\sigma^2}{m^n(m^2-m)}+1\leqslant c$$

Y_n est donc bornée dans L^2 donc converge p.s. et L^2 vers Y_∞. On a aussi $\mathbb{E}(Y_n) \to 1$ et $\mathbb{E}(Y_n^2) \to \mathbb{E}(Y_\infty^2)$.

Notons $\varphi_\infty(s) = \mathbb{E}(\mathrm{e}^{-sY_\infty})$ et $\varphi_n(s) = \mathbb{E}(\mathrm{e}^{-sY_n})$.

$$\varphi_{n+1}(ms) = \mathbb{E}(\mathrm{e}^{-ms\frac{Z_{n+1}}{m^{n+1}}}) = g(\varphi_n(s))$$

Donc $\varphi_\infty(ms) = g(\varphi_\infty(s))$.

4.2 Wright-Fisher

$X_{n+1}^N \sim B(N, \frac{X_n^N}{N})$, $X_0^N = k$ avec $0 < k < N$.

Si $Y \sim B(n,p)$ alors $\mathbb{E}(Y) = np$. On a $\mathbb{E}(X_{n+1}^N \mid F_n) = X_n^N$.

(X_n^N) est adaptée est intégrable donc X_n^N est une martingale bornée qui converge p.s. et dans L^p vers $X_\infty^N \in \{0, N\}$.

On a $\mathbb{E}(X_\infty^N) = \mathbb{E}(X_0^N) = k$ donc $\mathbb{P}(X_\infty^N = N) = \frac{k}{N}$.

4.3 Inégalité d'Azuma

Proposition 4.1 Soit $0 = X_0, X_1, \dots, X_m$ une martingale telle que $|X_i - X_{i-1}| \leqslant c_i$. Alors $\mathbb{P}(X_m > \lambda) \leqslant \mathrm{e}^{-\frac{\lambda^2}{2\sigma^2}}$ avec $\sigma^2 = \sum_{i=1}^{m} c_i^2$.

En particulier, si $c_i = 1$, on a $\mathbb{P}(X_m > \lambda\sqrt{m}) \leqslant \mathrm{e}^{-\lambda}2$.

Démonstration. Si $Y \in [-1,1]$ et $\mathbb{E}(Y) = 0$, $\mathbb{E}(\mathrm{e}^{uY}) \leqslant \mathrm{ch}(u) \leqslant \mathrm{e}^{\frac{u^2}{2}}$ puisque $h(x) := x\,\mathrm{sh}(u) + \mathrm{ch}(u) \geqslant \mathrm{e}^{ux}$ pour $x \in [-1,1]$.

$$\begin{aligned}\mathbb{E}(\mathrm{e}^{uX_m}) &= \mathbb{E}(\mathrm{e}^{u\sum(X_i - X_{i-1})}) = \mathbb{E}(\mathrm{e}^{u(X_m - X_{m-1})}\mathrm{e}^{uX_{m-1}}) \\ &= \mathbb{E}(\mathbb{E}(\mathrm{e}^{u(X_m - X_{m-1})} \mid F_{m-1})\mathrm{e}^{uX_{m-1}}) \leqslant \mathbb{E}(\mathrm{e}^{\frac{u^2}{2}}\mathrm{e}^{uX_{m-1}}) \leqslant \mathrm{e}^{m\frac{u^2}{2}}\end{aligned}$$

Donc

$$\mathbb{P}(X_n > \lambda) = \mathbb{P}(\mathrm{e}^{uX_m} > \mathrm{e}^{u\lambda}) \leqslant \mathbb{E}(\mathrm{e}^{uX_m})\mathrm{e}^{-\lambda u} \leqslant \mathrm{e}^{-\frac{\lambda^2}{2m}}$$ ■

<u>Théorème 4.1</u> *Soit $f : \mathfrak{S}_n \to \mathbb{R}$ 1-lipschitzienne (pour la distance donnée par $\mathrm{d}(\pi,\rho) = \mathrm{Card}\{i, \pi(i) \neq \rho(i)\}$).*

Soit π une permutation aléatoire de loi uniforme sur $\mathfrak{S}_n$.

$$\mathbb{P}(f(\pi) - \mathbb{E}(f) > \lambda) \leqslant \mathrm{e}^{-\frac{\lambda^2}{8n}}$$

Exemple 4.1

- Avec f la fonction nombre d'inversions, on a $\mathbb{E}(f) = \frac{n(n-1)}{4} \sim \frac{n^2}{4}$. Le théorème affirme alors que f est concentrée dans un intervalle de largeur $O(n^{\frac{3}{2}})$ autour de $\mathbb{E}(f)$.
- Soit E un Banach, $v_1, \ldots, v_n \in E$ unitaires et $X = \left|\sum_{i=1}^{n} \varepsilon_i v_i\right|$ avec $\varepsilon_i \sim B(\pm 1, \frac{1}{2})$ iid.
  Alors $\mathbb{P}(X - \mathbb{E}(X) > \lambda\sqrt{n}) \leqslant \mathrm{e}^{-\frac{\lambda^2}{2}}$.
- Soit $E = [\![1, n]\!]$ et g une fonction de $E \to E$. On pose

$$L(g) = \mathrm{Card}\{y, g(x) = y \text{ n'a pas de solution }\}$$

Si g est choisie uniformément, alors $\mathbb{E}(L(g)) = n(1 - \frac{1}{n})^n \sim \frac{n}{\mathrm{e}}$ et

$$\mathbb{P}(L(g) - \frac{n}{\mathrm{e}} > \sqrt{n}\lambda + 1) \leqslant \mathrm{e}^{\frac{-\lambda^2}{2}}$$

Chapitre 5

Chaînes de Markov

5.1 Probabilités de transition et matrices stochastiques

Définition 5.1 Soit $(E, \mathcal{E})$, $(F, \mathcal{F})$ deux espaces mesurables. On appelle probabilité de transition tout application $\nu : E \times \mathcal{F} \to [0,1]$ tel que pour tout $x \in E$, $\nu(x, \cdot)$ est une probabilité sur $\mathcal{F}$ et pour tout $A \in \mathcal{F}$, $\nu(\cdot, A)$ est mesurable.

Par la suite on prend $E = F$ et $\mathcal{E} = \mathcal{F}$ et on suppose que E est au plus dénombrable.

Définition 5.2 On appelle matrice stochastique une famille de nombres $(Q(x,y))_{x,y \in E}$ tel que $Q(x,y) \in [0,1]$ et pour tout x, $\sum_{y \in E} Q(x,y) = 1$.

Remarque 5.1 Les notions de probabilité de transition et de matrice stochastique sont équivalentes :

- *Si q est donné, $\nu(x,A) = \sum_{y \in A} Q(x,y)$ est une probabilité de transition.*
- *Si ν est donnée, $Q(x,y) = \nu(x, \{y\})$ est une matrice stochastique.*

Définition 5.3 Étant donnée Q une matrice stochastique, on définit Q_n par $Q_0(x,y) = 1_{x=y}$, $Q_1(x,y) = Q(x,y)$ et

$$Q_{n+1}(x,y) = \sum_{z \in E} Q_n(x,z) Q(z,y)$$

Si $f : E \to \mathbb{R}$, on note Qf la fonction $x \mapsto \sum_{y \in E} Q(x,y) f(y)$.

Remarque 5.2 Si E est fini, il est loisible de supposer (aussi nommé wlog[1])

1. without loss of generality

$E = [\![1, n]\!]$. $(Q(x,y))_{x,y}$ est une matrice. La suite Q_n est $(Q^n)_n$ et si $f = (f(1), \dots, f(n))^t$, alors $Qf = Q \times f$.

<u>Définition 5.4</u> Soient Q une matice stochastique sur E et X_n une suite de variables aléatoires à valeurs dans E. On dit que X_n est une chaîne de Markov ssi $L(X_{n+1} \mid X_0 \dots X_n)$ est $Q(X_n, \cdot)$. C'est-à-dire que

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_0 = 0, \dots, X_n = x_n) = Q(x_n, x_{n+1})$$

dès que $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) > 0$.

Remarque 5.3 La définition est celle d'une chaîne de Markov homogène car Q ne dépend pas de n.

5.1.1 Premières propriétés

Proposition 5.1 Une suite de variables aléatoires à valeurs dans E est une chaîne de Markov ssi pour tout n, $x_0, \dots, x_n \in E$,

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

De plus, si $\mathbb{P}(X_0 = x_0) > 0$; alors $\mathbb{P}(X_n = x_n \mid X_0 = x_0) = Q_n(x_0, x_n)$.

Démonstration.

$\Rightarrow$ Par récurrence :

$$\begin{aligned}
&\mathbb{P}(X_0 = x_0, \dots, X_{n+1} = x_{n+1}) \\
&= \mathbb{P}(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) \mathbb{P}(X_0 = x_0, \dots, X_n = x_n) \\
&= Q(x_n, x_{n+1}) \mathbb{P}(X_0 = x_0, \dots, X_n = x_n)
\end{aligned}$$

$\Leftarrow$ On réutilise la formule précédente et on a

$$\frac{\mathbb{P}(X_0 = x_0) \prod_{i=1}^{n} Q(x_i, x_{i+1})}{\mathbb{P}(X_0 = x_0) \prod_{i=1}^{n-1} Q(x_i, x_{i+1})} = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_0 = x_0, \dots, X_n = x_n)$$

Comme $Q_n(x_0, x_n) = \sum_{x_1, \dots, x_{n-1}} Q(x_0, x_1) \dots Q(x_{n-1}, x_n)$, on a

$$\begin{aligned}
\mathbb{P}(X_n = x_n \mid X_0 = x_0) &= \frac{\mathbb{P}(X_0 = x_0, X_n = x_n)}{\mathbb{P}(X_0 = x_0)} \\
&= \sum_{x_1, \dots, x_{n-1}} \frac{\mathbb{P}(X_0 = x_0, \dots, X_n = x_n)}{\mathbb{P}(X_0 = x_0)}
\end{aligned}$$ ■

Proposition 5.2 Soit X_n une chaîne de Markov à valeurs dans E.
- Pour tout $n \geqslant 0$ et $f : E \to \mathbb{R}$,

$$\mathbb{E}(f(X_{n+1}) \mid X_0 \dots X_n) = \mathbb{E}(f(X_{n+1}) \mid X_n) = Qf(X_n)$$

  Pour tout $i_1, \dots, i_k \in [\![0, n-1]\!]$

$$\mathbb{E}(f(X_{n+1}) \mid X_{i_1} \dots X_{i_k}, X_n) = Qf(X_n)$$

- Pour tout $n \geqslant 0$, $y_1, \dots, y_p \in E$,

$$\begin{aligned}\mathbb{P}(X_{n+1} = y_1, \dots, &X_{n+p} = y_p \mid X_0 = x_0, \dots, X_n = x_n) \\ &= Q(x_n, y_1)Q(y_1 y_2) \dots Q(y_{p-1}, y_p)\end{aligned}$$

Démonstration.
- On sait que $L(X_{n+1} \mid X_0 \dots X_n) = Q(X_n, \cdot)$.

$$\mathbb{E}(f(X_{n+1}) \mid X_0 \dots X_n) = \sum_{y \in E} f(y) Q(X_n, y) = Qf(X_n)$$

$$\begin{aligned}&\mathbb{E}(f(X_{n+1}) \mid X_{i_1} \dots X_{i_k} X_n) \\ &= \mathbb{E}(\mathbb{E}(f(X_{n+1}) \mid X_0 \dots X_n) \mid X_{i_1} \dots X_{i_k} X_n) \\ &= \mathbb{E}(Qf(X_n) \mid X_{i_1} \dots X_{i_k} X_n) = Qf(X_n)\end{aligned}$$

- On fait comme avant et tout se simplifie. ∎

Proposition 5.3 Soit X_0 une variable aléatoire dans E de loi μ_0. Soit U_n une suite de variables aléatoires iid à valeurs dans un espace $(F, \mathcal{F})$ indépendantes de X_0. Supposons que X_n est définie par récurrence

$$X_{n+1} = f(X_n, U_{n+1})$$

où $f : E \times F \to \mathbb{R}$ est mesurable.

Alors X_n est une chaîne de Markov de matrice associée

$$Q(x, y) = \mathbb{P}_\mu(f(x, U) = y)$$

où μ est la loi commune des U_i.

Démonstration. On a

$$\begin{aligned}&\mathbb{P}(X_{n+1} = x_{n+1} \mid X_0 = x_0, \dots, X_n = x_n) \\ =&\mathbb{P}(f(X_n, U_{n+1}) = x_{n+1} \mid X_0 = x_0, \dots, X_n = x_n) \\ =&\mathbb{P}(f(x_n, U_{n+1}) = x_{n+1} \mid X_0 = x_0, \dots, X_n = x_n) \\ =&\mathbb{P}(f(x_n, U_{n+1}) = x_{n+1}) = Q(x_n, x_{n+1})\end{aligned}$$

∎

5.1.2 Exemples

1. Soit X_n une suite de variables iid de loi μ. C'est une chaîne de Markov de transition $Q(x,y)=\mu(y)$.
2. Soit Y_n une suite de variables iid de loi μ. On pose $X_0=0$ et $X_n=\sum_{i=1}^{n} Y_i$.
   X_n est une chaîne de Markov de matrice de transition $Q(x,y)=\mu(y-x)$. En effet, $X_{n+1}=f(X_n,Y_{n+1})$ avec $f:(x,y)\mapsto x+y$. On a alors
   $$Q(x,y)=\mathbb{P}(f(x,Y_{n+1})=y)=\mathbb{P}(x+Y_{n+1}=y)=\mu(y-s)$$
3. On prend un graphe $G=(S,A)$ dénombrable. On note A_x l'ensemble des arêtes issues de $x\in S$.
   Posons $Q(x,y)=\frac{1}{\mathrm{Card}(A_x)}$ si y est voisin de x et 0 sinon. La chaîne de Markov de transition Q est appelée marche aléatoire simple sur G.
4. Galton-Watson est une chaîne de Markov $Z_{n+1}=f(Z_n,X_{n+1,\cdot})$ avec $f(x,y)=\sum_{i=1}^{x} y_i$.
   La matrice de transition est la convolée : $Q(x,z)=\mu_Z^{*x}$.
5. Wright-Fisher est aussi une chaîne de Markov de matrice de transition
   $$Q(i,j)=\binom{N}{j}\left(\frac{i}{N}\right)^j\left(1-\frac{i}{N}\right)^{N-j}$$

5.2 Un peu d'algèbre et de géométrie

On suppose E fini. Q est alors une matrice stochastique.

5.2.1 Théorème de Perron-Frobénius

<u>Théorème 5.1</u> *Soit A une matrice stochastique dont tous les coefficients sont strictement positifs. Alors 1 est valeur propre simple de vecteur propre associé $e=(1,\dots,1)^t$.*

Toutes les autres valeurs propres complexes de A sont de module <1.

Démonstration. e est trivialement un vecteur propre associé à 1. Soit λ une valeur propre complexe différente de 1 et $x\neq 0$ un vecteur propre associé.

$Ax=\lambda x$ ssi $\sum_{j=1}^{n} a_{i,j}x_j=\lambda x_i$. En prenant les modules, on obtient

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

Si x_k est estl que $|x_k| = \max_{i \in [\![1,n]\!]} |x_i|$, on a

$$|\lambda||x_k| \leqslant \sum_{j=1}^{n} a_{i,j}|x_k|$$

ie $|\lambda| \leqslant 1$. On a de plus égalité ssi tous les x_i sont égaux ie si x et e sont colinéaires.

Il reste à montrer que 1 est valeur propre simple.

Lemme 5.1.1
Si λ est une valeur propre strictement plus grande que 1. Si l'espace associé esst de dimension 1 engendré par e alors il existe x tel que $Ax = \lambda x + e$.

Démonstration. Quitte à considéter $A - \lambda \operatorname{Id}$, on peut supposer $\lambda = 0$. Il faut montrer que $\operatorname{Vect}\{e\} = \operatorname{Ker}(A) \subset \operatorname{Im}(A)$.

Si ce n'était pas le cas, $\operatorname{Ker}(A)$ et $\operatorname{Im}(A)$ seraient en somme directe et dans une base adaptée, A s'écrirait $\begin{pmatrix} 0 & 0 \\ 0 & A' \end{pmatrix} =: B$ avec A' inversible. Alors $\det(B - X \operatorname{Id})$ montre que 0 est valeur propre simple. Contradiction. ■

S'il existe x tel que $Ax = \lambda x + e$, on trouve que pour tout i, $\sum_{j=1}^{n} a_{i,j}x_j = x_i + 1$. En prenant x_k le plus grand des x_j, on a $x_k + 1 \leqslant x_k$. Contradiction. ■

Corollaire 5.1 *S'il existe r tel que A^r ait tous ses coefficients strictement positifs, le théorème s'applique. Les valeurs propres de A^r sont les λ^r.*

Corollaire 5.2 *Soit A une matrice stochastique dont une puissance a tous ses coefficients positifs. Alors quand $k \to \infty$, A^k tend vers A^∞ de la forme $\begin{pmatrix} \alpha_1 & \cdots & \alpha_n \\ \vdots & & \vdots \\ \alpha_1 & \cdots & \alpha_n \end{pmatrix}$ avec $\alpha_i \geqslant 0$ et $\sum_{i=1}^{n} \alpha_i = 1$.*

Démonstration. Si A est diagonalisable, on écrit $A = PDP^{-1}$ avec $D = \operatorname{diag}(1, \lambda_1, \dots, \lambda_{n-1})$ donc $A^k = P \operatorname{diag}(1, \lambda_i^k) P^{-1}$ d'où quand $k \to \infty$, $A^\infty = P \operatorname{diag}(1, 0, \dots, 0) P^{-1}$.

Sinon, on a $A \sim (D + N)$ (décomposition de Dunford) et on vérifie que $(D + N)^k \to \operatorname{diag}(1, 0, \dots, 0)$. ■

Exemple 5.1 $Q = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$ avec $0 < p < 1$ et $0 < q < 1$.

On a $\mathrm{Sp}(Q) = \{1, 1-p-q\}$ et les vecteurs propres associés sont $(1,1)^t$ et $(p,-q)^t$. On a $Q^n = PDP^{-1} = \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix} + \frac{\lambda^n}{p+q}\begin{pmatrix} p & -p \\ -q & q \end{pmatrix}$. Donc Q^n converge à la vitesse λ^n.

5.2.2 Matrices bistochastiques

Définition 5.5 A est bistochastique ssi A et A^t sont stochastique.

Exemple 5.2 Si $\sigma \in \mathfrak{S}_n$, la matrice de passage entre les bases $(e_1, \dots, e_n)$ et $(e_{\sigma(1)}, \dots, e_{\sigma(n)})$ (matrice de permutation) est bistochastique.

Théorème 5.2 Birckhoff *Une matrice A est bistochastique ssi on peut l'écrire $A = \sum\limits_{k=1}^{p} \lambda_k P_k$ où P_k sont des matrices de permutation, λ_k positifs de somme 1 et $p \leqslant (n-1)^2 + 1$.*

Définition 5.6 Soit E un espace vectoriel réel et X une partie de E. L'enveloppe convexe de X, notée $\mathrm{cv}(x)$ est l'ensemble des points combinaisons linéaires d'éléments de x avec des coefficients de somme 1.

Théorème 5.3 Carathéodory *Si X est une partie non vide et E dimension $n < \infty$, alors pour tout $x \in \mathrm{cv}(X)$ il existe $\lambda_1, \dots, \lambda_p$ positifs de somme 1 et $x_1, \dots, c_p \in X$ tel que $\sum\limits_{i=1}^{p} \lambda_i x_i = x$ et on a $p \leqslant n+1$.*

Démonstration. Par définition, $x = \sum\limits_{i=1}^{p} \lambda_i x_i$. Si $p \leqslant n+1$, c'est fini. Sinon, le système $(x_i - x_1)_{2 \leqslant i \leqslant p}$ est lié donc il existe μ_i tel que $\sum\limits_{i=2}^{p} \mu_i (x_i - x_1) = 0$.

Ceci s'écrit encore $\sum\limits_{i=1}^{p} \mu_i x_i = 0$ en posant $\mu_1 = -\sum\limits_{i=2}^{p} \mu_i$, ie $\sum\limits_{i=1}^{p} \mu_i = 0$.

Posons $\alpha = \min\{\frac{\lambda_i}{\mu_i}, \mu_i > 0\} > 0$. On a alors $\lambda_i - \alpha\mu_i \geqslant 0$ et pour i qui réalise le minimium, $\lambda_i = \alpha\mu_i$.

$x = \sum\limits_{i=1}^{p} (\lambda_i - \alpha\mu_i) x_i$ et il y a $p-1$ termes non nuls dans la somme et on recommence. ∎

Corollaire 5.3 *En dimension finie, l'enveloppe convexe d'un compact est compact. On en déduit que $\mathrm{cv}(X)$ est compact (c'est l'image de $\Delta \times X^{n+1}$ par $(\lambda, x) \to \sum\limits_{i=1}^{n+1} \lambda x_i$ qui est continue.*

Définition 5.7 Soit E un eve de dimension finie $n \geqslant 2$ et X un convexe non vide de E. Un point $x \in X$ est dit extrémal ssi $X \setminus \{x\}$ est convexe.

THÉORÈME 5.4 KREIN-MILMAN *Sous les hypothèses précédentes, tout compact convexe est l'enveloppe convexe de ses points extrémaux.*

On note B_n l'ensemble des matrices bistochastiques et $C_n = \text{Vect}\{B_n\}$.

Lemme 5.4.1
C_n est de dimension au plus $(n-1)^2$.

Démonstration. Si $X \in C_n$, il existe $c \in \mathbb{R}$ tel que $\sum_i X_{i,j} = \sum_j X_{i,j} = c$. L'application φ qui à $X \in C_n$ associe son mineur de taille $n-1 \times n-1$ est linéaire et injective. En effet, si $\varphi(X) = 0$ alors

$$X_{i,n} = c \text{ et } \sum_{k=1}^{n} X_{k,n} = c$$

On a donc $(n-1)c = 0$ et $c = 0$ donc $X = 0$. D'où le résultat. ∎

Lemme 5.4.2
B_n est un polyèdre convexe dont les points extrémaux sont les matrices de permutation.

Démonstration. Soit $(e_{i,j})_{i,j}$ la base canonique et $(e_{i,j}^*)$ la base duale. On a

$$B_n = \left\{ A, \sum_{i=1}^{n} e_{i,j}^* A = \sum_{j=1}^{n} e_{i,j}^* A = 1, e_{i,j}^*(A) \geqslant 0 \right\}$$

On vérifie que les matrices de permutations sont extrémales. Pour prouver la réciproque, on raisonne par récurrence sur la dimension n. ∎

5.3 Propriété de Markov

5.3.1 Chaîne de Markov canonique

Proposition 5.4 Soit E un ensemble au plus dénombrable et Q matrice stochastique. On peut trouver un espace de probabilité $(\widetilde{\Omega}, \widetilde{\mathscr{F}}, \widetilde{\mathbb{P}})$ sur lequel il existe une suite $(X_n^x)_n$ qui est une chaîne de Markov de transition Q issue de X_0^x.

Démonstration. On prend $\widetilde{\Omega} = [0,1[$, $\widetilde{\mathscr{F}} = \mathscr{B}([0,1[)$ et $\widetilde{\mathbb{P}}$ la mesure de Lebesgue.

Soit $\omega \in [0,1[$ et sa décomposition dyadique $\sum_{n=0}^{\infty} \varepsilon_n(\omega)2^{-(n+1)}$.

Les ε_n sont iid $\sim \mathcal{B}(\frac{1}{2})$. Soit $\varphi : \mathbb{N}^2 \to \mathbb{N}$ une injection.

$\eta_{i,j} = \varepsilon_{\varphi(i,j)}$ est encore iid. On pose $U_i = \sum_{k=0}^{\infty} \eta_{i,j} 2^{-(j+1)}$ c'est une suite de variables $\mathcal{U}_{[0,1]}$.

On construit alors (X_n^x) par récurrence. Soit $y_1, \dots, y_n$ une énumération de E. On pose

$$\begin{cases} X_0^x = x \\ X_1^x = y_k \text{ si } \sum_{j=1}^{k-1} Q(x, y_j) < U_1 \leqslant \sum_{j=1}^{k} Q(x, y_j) \\ X_n^x = y_k \text{ si } \sum_{j=1}^{k-1} Q(X_n^x, y_j) < U_{n+1} \leqslant \sum_{j=1}^{k} Q(X_n^x, y_j) \end{cases}$$

On a $\lambda(X_1^x = y) = Q(x,y)$ et comme les U_i sont indépendantes,

$$\begin{aligned} &\lambda(X_{n+1}^x = y_k \mid X_0^x = x_0, \dots, X_n^x = x_n) \\ &= \lambda\left(\sum_{j=1}^{k-1} Q(x_n, y_j) < U_{n+1} \leqslant \sum_{j=1}^{k} Q(x_n, y_j) \mid X_0 = x_0, \dots, X_n^x = x_n\right) \\ &= \lambda\left(\sum_{j=1}^{k-1} Q(x_n, y_j) < U_{n+1} \leqslant \sum_{j=1}^{k} Q(x_n, y_j)\right) \\ &= Q(x_n, y_k) \end{aligned}$$

donc $(X_n^x)_n$ est une chaîne de Markov de transition Q. ■

On prend $\Omega = E^{\mathbb{N}}$, $X_n(\omega) := \omega_n$ et $\mathscr{F}$ la plus petite tribu qui rend mesurable les X_n, ie engendrée par les cylindres $C_n = \{X_0 = x_0, \dots, X_n = x_n\}$.

Lemme 5.4.3
Soit $\psi : (\widetilde{\Omega}, \widetilde{\mathscr{F}}) \to (\Omega, \mathscr{F})$ définis précédemment.

ψ est mesurable ssi $X_n \circ \psi$ l'est pour tout n.

Démonstration.

$\Rightarrow$ Trivial

$\Leftarrow$ $\{A \in \mathscr{F}, \psi^{-1}(A) \in \widetilde{F}\}$ est une tribu qui contient tous les $X_n^{-1}(y)$ donc rend mesurable les X_n. Nécessairement, c'est $\mathcal{F}$. ■

<u>THÉORÈME 5.5</u> *Soit E au plus dénombrable et Q matrice stochastique. Pour tout $x \in E$, il existe une unique probabilité $\mathbb{P}_x$ sur $(\Omega, \mathscr{F})$ telle que, sous $\mathbb{P}_x$, la suite X_n est une chaîne de Markov de transition Q et $\mathbb{P}_x(X_0 = x) = 1$.*

Démonstration. Soit $x \in E$. Par la proposition précédente, il existe $(\widetilde{\Omega}, \widetilde{\mathscr{F}}, \widetilde{\mathbb{P}})$ et (X_n^x) chaîne e Markov de transition Q.

On définit $\mathbb{P}_x$ comme la mesure image de $\widetilde{\mathbb{P}}$ par $\omega \mapsto (X_n^x(\omega)_n$ qui est mesurable par le lemme. On a

$$\mathbb{P}_x(X_0 = x) = \widetilde{\mathbb{P}}(X_0^x = x) = 1$$

et

$$\begin{aligned}
\mathbb{P}_x(X_0 = x_0, \dots, X_n = x_n) &= \widetilde{\mathbb{P}}(X_0^x = x_0, \dots, X_n^x = x_n) \\
&= \widetilde{\mathbb{P}}(X_0^x = x_0)Q(x_0, x_1) \dots Q(x_{n-1}, x_n) \\
&= \mathbb{P}_x(X_0 = x_0)Q(x_0, x_1) \dots Q(x_{n-1}, x_n)
\end{aligned}$$

D'où le résultat.

Pour l'unicité, si $\mathbb{P}'_x$ vérifie les mêmes hypothèses que $\mathbb{P}_x$, alors elles sont égales sur tous les $\{X_0 = x_0, \dots, X_n = x_n\}$ donc partout. ■

Remarque 5.4

- *Pour tout n, x, y, $\mathbb{P}_x(X_n = y) = Q_n(x, y)$*
- *Si μ est une mesure sur E, alors*

  $$\mathbb{P}_\mu = \sum_{x \in E} \mu(x) \mathbb{P}_x$$

  est une mesure de probabilité sur $(\Omega, \mathscr{F})$.
  Sous $\mathbb{P}_\mu$, les coordonées forment une chaîne de Markov de transition Q et de loi initiale $X_0 \sim \mu$.
- *Soit (X'_n) définie sur $(\Omega', \mathscr{F}', \mathbb{P}')$ une chaîne de Markov de transition Q et de loi initiale μ sous $\mathbb{P}'$. Alors pour tout $B \in \mathscr{F}$,*

  $$\mathbb{P}'((X'_n)_n \in B) = \mathbb{P}_\mu(B)$$

5.3.2 Propriété de Markov

En travaillant sur $\Omega = E^{\mathbb{N}}$, on peut introduire les opérateurs de décalage (ou translation, shift) :

$$\theta_k : \begin{cases} \Omega & \to & \Omega \\ \omega & \mapsto & (\omega_{n+k})_n \end{cases}$$

D'après le lemme précédent, ce sont des applications mesurables. On pose $\mathscr{F}_n = \sigma(X_0, \dots, X_n)$. On note $\mathbb{E}_x$ l'espérance sous $\mathbb{P}_x$.

<u>THÉORÈME 5.6</u> PROPRIÉTÉ DE MARKOV FAIBLE *Soient G une application mesurable sur Ω et $x \in E$.*

$$\mathbb{E}_x(G \circ \theta_n \mid \mathscr{F}_n) = \mathbb{E}_{X_n}(G)$$

De manière équivalente, pour tout $F \in \mathscr{F}_n$ mesurable,

$$\mathbb{E}_x(FG \circ \theta_n) = \mathbb{E}_x(F\mathbb{E}_{X_n}(G))$$

Remarque 5.5 *Le théorème se généralise sans problème au cas où la loi initiale est μ, ie*

$$\mathbb{E}_\mu(G \circ \theta_n \mid \mathscr{F}_n) = \mathbb{E}_{X_n}(G)$$

Démonstration. Par un argument de classe monotone, il suffit de traiter le cas

$$F = 1_{\{X_0=0,\dots,X_n=x_n\}} \text{ et } G = 1_{\{X_0=y_0,\dots,X_p=y_p\}}$$

On a

$$\mathbb{E}_y(G) = \mathbb{P}_Y(X_0 = y_0, \dots, X_p = y_p) = 1_{y=y_0}Q(y_0, y_1)\dots Q(y_{p-1}, y_p)$$

$$\begin{aligned}\mathbb{E}_x(FG \circ \theta_n) &= \mathbb{P}_x(X_0 = x_0, \dots, X_n = x_n, X_n = y_0, \dots, X_{n+p} = y_p) \\ &= 1_{x=x_0}Q(x_0, x_1)\dots Q(x_{n-1}, x_n)1_{y=y_0}Q(y_0, y_1)\dots Q(y_{p-1}, y_p)\end{aligned}$$

Enfin

$$\begin{aligned}\mathbb{E}_x(F\mathbb{E}_{X_n}(G)) &= \mathbb{E}_x(1_{X_0=x_0,\dots,X_n=x_n}1_{X_n=y_0}Q(y_0, y_1)\dots Q(y_{p-1}, y_p)) \\ &= 1_{x=x_0}Q(x_0, x_1)\dots Q(x_{n-1}, x_n)_{x_n=y_0}Q(y_0, y_1)\dots Q(y_{p-1}, y_p)\end{aligned}$$

D'où l'égalité. ■

<u>THÉORÈME 5.7</u> PROPRIÉTÉ DE MARKOV FORTE *Soit T un temps d'arrêt de la filtration $(\mathscr{F}_n)$. Soit G une fonction mesurable sur $(\Omega, \mathscr{F})$. Alors*

$$\mathbb{E}_x(1_{T<\infty}G \circ \theta_T \mid \mathcal{F}_T) = 1_{T<\infty}E_{X_T}(G)$$

ie pour tout F fonction $\mathscr{F}_T$ mesurable

$$\mathbb{E}_x(1_{T<\infty}FG \circ \theta_T) = \mathbb{E}_x(1_{T<\infty}F\mathbb{E}_{X_T}(G))$$

Remarque 5.6 *X_T est $\mathscr{F}_T$-mesurable.*

Démonstration. Pour tout n, par Markov faible,

$$\begin{aligned}\mathbb{E}_x(1_{T=n}FG \circ \theta_T) &= \mathbb{E}_x(1_{T=n}FG \circ \theta_n) = \mathbb{E}_x(1_{T=n}F\mathbb{E}_{X_n}(G)) \\ &= \mathbb{E}_x(1_{T=n}F\mathbb{E}_{X_T}(G))\end{aligned}$$

On somme en n et ça marche (Fubini). ■

COROLLAIRE 5.4 *Soit T un temps d'arrêt tel que $\mathbb{P}_x(X_T = y) = 1$ pour un certain $y \in E$ et $\mathbb{P}_x(T < \infty) = 1$.*

Alors, sous $\mathbb{P}_x$, $\theta_{T(\omega)}$ est indépendante de $\mathscr{F}_T$ et a pour loi $\mathbb{P}_y$.

5.4 Applications de la propriété de Markov

Soit $F \subset E$ au plus dénombrable muni d'une matrice de transition Q. Posons $T_F = \inf\{n, X_n \in F\}$ et $h : F \to \mathbb{R}^+$ bornée.

La fonction $g : E \to \mathbb{R}^+$ définie par

$$g(x) = \mathbb{E}_x(h(X_{T_F})1_{T_F<\infty})$$

est solution du problème de Dirichlet

$$\begin{cases} (I-Q)g = 0 & \text{dans } F^c \\ g = h & \text{dans } F \end{cases}$$

Démonstration. Si $x \in F$, $T_F(X_0) = 0$ donc $h(X_{T_F}) = h(x)$.

Donc $g(x)\mathbb{E}_x(h(X_{T_F})1_{T_F<\infty}) = h(x)$.

Si $x \notin F$, $T_F(X_0) = T_F(X_1) \geqslant 1$. On a alors

$$\begin{aligned} g(x) &= \mathbb{E}_x(h(X_{T_F(X_0)})1_{T_F(X_0)<\infty}) \\ &= \mathbb{E}_x(h(X_{1+T_F(X_1)})1_{T_F(X_1)<\infty}) \\ &= \mathbb{E}_x(\mathbb{E}_x(h(X_{1+T_F(X_0)})1_{T_F(X_1)<\infty} \mid F_1)) \\ &= \mathbb{E}_x(\mathbb{E}_{X_1}(h(X_{T_F(X_0)}))1_{T_F(X_0)<\infty}) \\ &= \mathbb{E}_x(g(X_1)) \end{aligned}$$

Or $\mathbb{E}_x(g(X_1)) = \sum_{y\in E}^{g}(y)Q(x,y) = Q_y(x)$ donc $g(x) = Qg(x)$. ∎

Chapitre 6

Classification des états

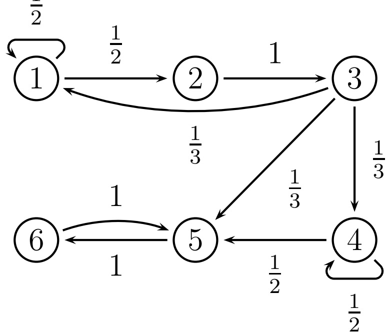
6.1 Dichotomie récurrence/transcience

6.1.1 Motivations

Sur l'espace $E = [\![1, 6]\!]$, on considère la chaîne de Markov de transition

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

Le graphe associé est

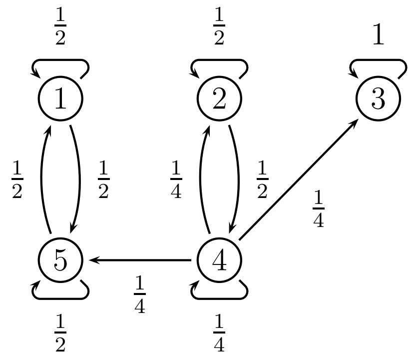


Sous $\mathbb{P}_1$, les états $\{1, 2, 3, 4\}$ semblent visités un nombre fini de fois. Sous $\mathbb{P}_1$, 5 et 6 semblent visités un nombre infini de fois.

Sur $E = [\![1,5]\!]$, on considère la chaîne de Markov de transition

$$Q = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

La graphe associé est



Sous $\mathbb{P}_1$, $\{2,3,4\}$ ne sont jamais visités.

Sous $\mathbb{P}_2$, $\{2,4\}$ sont visités un nombre fini de fois et $\{1,5\}$ ou $\{3\}$ sont visités un nombre infini de fois.

6.1.2 États récurrents et transcients

On se donne un espace au plus dénombrable E muni d'une matrice de transition Q. On travaille avec la chaîne canonique X_n. On pose

$$T_x := \inf\{n, X_n = x\} \text{ et } N_x := \sum_{n=0}^{\infty} 1_{X_n = x}$$

On introduit la fonction de Green (ou fonction potentiel)

$$G(x,y) := \mathbb{E}_x(N_y) = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y) = \sum_{n=0}^{\infty} Q_n(x,y) \in [0, +\infty[$$

Proposition 6.1 Soit $x \in E$. On a l'alternative :

- ou bien $\mathbb{P}_x(T_x < \infty) = 1$ et alors $N_x = \infty$ $\mathbb{P}_x$-p.s. Dans ce cas, on dit que x est récurrent.
- ou bien $\mathbb{P}_x(T_x < \infty) < 1$ et alors $N_x < \infty$ $\mathbb{P}_x$-p.s. Plus précisément $G(x,x) = \mathbb{E}_x(N_x) = \frac{1}{\mathbb{P}_x(T_x = \infty)} < \infty$.
  Dans ce cas, x est dit transitoire ou transcient.

Démonstration. Pour $k \geqslant 1$, on calcule

$$\begin{aligned}\mathbb{P}_x(N_x \geqslant k+1) &= \mathbb{E}_x(1_{N_x \geqslant k+1}) \\ &= \mathbb{E}_x(1_{T_x<\infty} A_{N_x \geqslant k} \circ \theta_{T_x}) \\ &= \mathbb{E}_x(\mathbb{E}_x(1_{T_x<\infty} 1_{N_x \geqslant k} \circ \theta_{T_x} \mid \mathscr{F}_{T_x})) \\ &= \mathbb{E}_x(1_{T_x<\infty} \mathbb{E}_x(1_{N_x \geqslant k} \circ \theta_{T_x} \mid \mathscr{F}_{T_x})) \\ &= \mathbb{E}_x(1_{T_x<\infty} \mathbb{E}_x(1_{N_x \geqslant k}))\end{aligned}$$

Donc $\mathbb{P}_x(N_x \geqslant k+1) = \mathbb{P}_x(T_x < \infty)\mathbb{P}_x(N_x \geqslant k)$.

Or $\mathbb{P}_x(N_x \geqslant 1) = 1$ donc $\mathbb{P}_x(N_x \geqslant k) = \mathbb{P}_x(T_x < \infty)^{k-1}$.

- Si $\mathbb{P}_x(T_x < \infty) = 1$, $\mathbb{P}_x(N_x = \infty) = 1$
- Si $\mathbb{P}_x(T_x < \infty) < 1$ alors

$$\begin{aligned}G(x,x) = \mathbb{E}_x(N_x) &= \sum_{k=1}^{\infty} \mathbb{P}_x(N_x \geqslant k) = \sum_{k=1}^{\infty} \mathbb{P}_x(T_x < \infty)^{k-1} \\ &= \frac{1}{1-\mathbb{P}_x(T_x<\infty)} = \frac{1}{\mathbb{P}_x(T_x=\infty)} < \infty\end{aligned}$$ ∎

Proposition 6.2 Soit $x \in E$. Alors
- x est récurrent ssi $G(x,x) = \infty$
- $G(x,y) = \mathbb{P}_x(T_y < \infty)G(y,y)$ pour tout $x, y \in E$.

Démonstration. Le premier point est évident par ce qu'il y a au-dessus.

$$\begin{aligned}G(x,y) = \mathbb{E}_x(N_y) &= \mathbb{E}_x(1_{T_y<\infty} N_y \circ \theta_{T_y}) \\ &= \mathbb{E}_x(1_{T_y<\infty})\mathbb{E}_y(N_y) = \mathbb{P}_x(T_y < \infty)G(y,y)\end{aligned}$$ ∎

Exemple 6.1 On considère la marche simple sur $\mathbb{Z}$: $X_n = \sum_{i=1}^{n} \varepsilon_i$ où $\varepsilon_i \sim \mathcal{B}(\pm 1, p)$. Quelle est la nature de 0 ?

$$G(0,0) = \sum_{n=0}^{\infty} Q_n(0,0) = \sum_{n=0}^{\infty} Q_{2n}(0,0)$$

On a $Q_{2n}(0,0) = \binom{2n}{n} p^n (1-p)^n \sim \frac{(4p(1-p))^n}{\sqrt{\pi n}}$.

Si $p = \frac{1}{2}$, $Q_{2n}(0,0) \sim \frac{1}{\sqrt{\pi n}}$ et $G(0,0) = \infty$.

Sinon, $4p(1-p) < 1$ donc $G(0,0) < \infty$ ie 0 est transitoire.

Il existe donc une partition de l'espace des états : $E = E_R \sqcup E_T$.

Lemme 6.0.1
Soit $x \in E_R$ et $y \in E$ tel que $G(x,y) > 0$.

Alors $y \in E_R$ et $\mathbb{P}_y(T_x < \infty) = 1$. En particulier, $G(y,x) > 0$.

Démonstration. On montre d'abord que $\mathbb{P}_y(T_x < \infty) = 1$.

$$\begin{aligned} 0 = \mathbb{P}_x(N_x < \infty) &\geqslant \mathbb{P}_x(T_y < \infty \text{ et } T_x \circ \theta_{T_y} = \infty \\ &= \mathbb{E}_x(1_{T_y<\infty} 1_{T_x\circ\theta_{T_y}=\infty}) = \mathbb{P}_x(T_y < \infty)\mathbb{E}_y(1_{T_x=\infty}) \\ &= \mathbb{P}_x(T_y < \infty)\mathbb{P}_y(T_x = \infty) \end{aligned}$$

Le produit est nul donc $\mathbb{P}_x(T_y < \infty) = 0$ ou $\mathbb{P}_y(T_x = \infty) = 0$.

Comme $G(x,y) > 0$, le premier terme est non nul donc $\mathbb{P}_y(T_x = \infty) = 0$ et $\mathbb{P}_y(T_x < \infty) = 1$. En particulier, $G(y,x) > 0$.

Comme $G(x,y) > 0$ et $G(y,x) > 0$, il existe n, m tel que $Q_n(x,y) > 0$ et $Q_m(x,y) > 0$. Soit $k \in \mathbb{N}$.

$$Q_{n+k+m}(y,y) \geqslant Q_m(y,x)Q_k(x,x)Q_n(x,y)$$

D'où

$$\begin{aligned} G(y,y) = \sum_{p=0}^{\infty} Q_p(y,y) &\geqslant \sum_{k=0}^{\infty} Q_{n+k+m}(y,y) \\ &\geqslant Q_m(y,x) \underbrace{\sum_{k=0}^{\infty} Q_k(x,x)}_{=\infty} Q_n(x,y) = \infty \end{aligned}$$

Donc y est récurrent. ∎

<u>Théorème 6.1</u> Classification des états *Il existe une partition de $E_R = \bigsqcup E_{R_i}$ telle que*

- *si $x \in E_{R_i}$ alors $\mathbb{P}_x$-p.s., $N_y = +\infty$ pour $y \in E_{R_i}$ et $N_y = 0$ pour $y \notin E_{R_i}$.*
- *si $x \in E_T$, on note $T_{E_R} = \inf\{n, X_n \in E_R\}$, alors $\mathbb{P}_x$-p.s.,*
  - *si $T_{E_R} = \infty$ alors $N_y < \infty$ pour tout $y \in E$*
  - *si $T_{E_R} < \infty$ et $X_{T_{E_R}} \in E_{R_i}$ alors pour tout $n \geqslant T_{E_R}$, on a $X_n \in E_{R_i}$*

Démonstration.

- On pose $x \sim y$ ssi $G(x,y) > 0$. C'est une relation d'équivalence par les résultats précédents. Les E_{R_i} sont les classes d'équivalence de cette relation.
- Si $x \in E_{R_i}$ alors $G(x,y) = 0$ pour tout y.
  Si $y \in E_{R_j}$ avec $i \neq j$, on ne peut pas changer de classe d'équivalnce. Si $y \in E_T$, par le lemme, $G(x,y) = 0$ et si $y \in E_{R_i}$, par le lemme, $\mathbb{P}_x(T_y < \infty) = 1$. Par ailleurs,

$$\mathbb{P}_x(N_y = \infty) = \mathbb{E}_x(1_{T_y<\infty} 1_{N_y=\infty} \circ \theta_{T_y}) = \mathbb{P}_x(T_y < \infty)\mathbb{P}_y(N_y = \infty) = 1$$

- Si $x \in E_T$ et $T_{E_R} = \infty$, par la relation

  $$\mathbb{P}_x(N_y = +\infty) = \mathbb{P}_x(Y_y < \infty)\mathbb{P}_y(N_y = \infty)$$

  Si $y \in E_T$ alors $\mathbb{P}_y(N_y = \infty) = 0$ donc $\mathbb{P}_x(N_y = \infty) = 0$.
  Si $T_{E_R} < \infty$, par la propriété de Markov fort appliquée à T_{E_R} et le point précédent, on a $X_n \in E_{R_i}$ pour tout $n \geqslant T_{E_R}$. ∎

Définition 6.1 Un point $x \in E$ tel que $\{x\}$ est une classe de récurrence (ie $\mathbb{P}_x(X_1 = x) = Q(x, x) = 1$) est dit absorbant.

Exemple 6.2 Reprenons le dernier graphe. On a $\{1, 2, 3, 4, 5\} = \{2, 4\} \sqcup \{1, 5\} \sqcup \{3\}$. On montre que $1, 3, 5$ sont récurrents.

$$\begin{aligned}\mathbb{P}_1(T_1 = \infty) &= \mathbb{P}_1(X_1 = 5 \text{ et } \forall n, X_n = 5)\\ &\leqslant \mathbb{P}_1(X_1 = 5 \text{ et } \forall n \leqslant N, X_n = 5)\\ &= Q(1, 5)Q(5, 5)^{N-1} \to 0\end{aligned}$$

On fait pareil pour 5. $\{3\}$ est absorbant. Montrons que 2 et 4 sont transitoires.

$$\mathbb{P}_2(T_2 = \infty) \geqslant \mathbb{P}_2(X_1 = 4 \text{ et } X_n = 3) \geqslant Q(2, 4)Q(4, 3) > 0$$

$$\mathbb{P}_4(T_4 = \infty) \geqslant Q(4, 3) > 0$$

Donc $\mathbb{P}_4(T_4 < \infty) < 1$.

Définition 6.2 On dit qu'une chaîne de Markov de transition Q sur E est irréductible ssi $G(x, y) > 0$ pour tout $x, y \in E$.

Corollaire 6.1 *Si $(X_n$ est une chaîne irréductible,*

- *soit tous les états sont récurrents : pour tout $x \in E$,*

  $$\mathbb{P}_x(N_y = \infty, \forall y \in E) = 1$$

- *soit tous les états sont transitoires : pour tout $x \in E$,*

  $$\mathbb{P}_x(N_y < \infty, \forall y \in E) = 1$$

Définition 6.3 Une chaîne irréductible dont tous les états sont récurrents est dite irréductible récurrente.

Corollaire 6.2 *Si l'espace d'états E est fini, une chaîne irréductible est nécessairement récurrente irréductible.*

Démonstration. Supposont qu'un état (et donc tous) est transitoire. On a $N_y < \infty$ pour $y \in E$ donc $\sum_{y \in E} N_y < \infty$. Or on a

$$\sum_{y \in E} N_y = \sum_{y \in E} \sum_{n=0}^{\infty} 1_{X_n = y} = \sum_{n=0}^{\infty} \sum_{y \in E} 1_{X_n = y} = +\infty$$ ∎

Définition 6.4 On appelle période de l'état $x \in E$ et on note d_x le pgcd des n tels que $Q_n(x,x) > 0$.

Lorsque $d_x = 1$, on dit que x est apériodique.

Exemple 6.3 La période de 0 pour la marche simple sur $\mathbb{Z}$ est 2.

Proposition 6.3 Si $x \sim y$ (ie $G(x,y) > 0$) alors $d_x = d_y$.

Démonstration. Il existe $n, m \in \mathbb{N}$ tel que $Q_n(x,y) > 0$ et $Q_m(y,x) > 0$. Alors

$$Q_{n+m+Nk}(x,x) \geqslant Q_n(x,y)(Q_k(y,y))^N Q_m(y,x)$$

Donc $Q_{n+m+Nk}(x,x) > 0$ dès que $Q_k(y,y) > 0$ et dans ce cas, $d_x \mid n+m+Nk$ pour tout $N \geqslant 1$ donc $d_x \mid k$ donc $d_x \mid d_y$. Par symétrie, $d_y \mid d_x$ donc on a l'égalité. ■

Corollaire 6.3 *Si une chaîne X_n à valeurs dans E est irréductible, alors tous les états ont la même période, appelée période de la chaîne. Si cette période vaut 1, on dit que la chaîne est apériodique.*

6.1.3 Absorption dans les classes de récurrence

Si $A \subset E$, on pose $T_A = \inf\{n, X_n \in A\}$ et on définit $\tau_x(A) = \mathbb{E}_x(T_A)$ et $m_x(A) = \mathbb{P}_x(T_A < \infty)$. On veut déterminer $\tau_x(E_{R_i})$ et $m_x(E_{R_i})$ pour $i \in I$.

Si $x \in E_{R_i}$ alors $\tau_x(E_{R_i}) = 0$ et $m_x(E_{R_i}) = 1$ et si $j \neq i$ alors $m_x(E_{R_j}) = 0$ et par convention on pose $\tau_x(E_{R_j}) = 0$. Et si $x \in E_T$?

Proposition 6.4 Les probabilités et temps d'absorption sont les solutions minimales des systèmes linéaires

$$\begin{cases} m_x(E_{R_j}) & = \sum_{y \in E} Q(x,y) m_y(E_{R_j}) \\ \tau_x(E_{R_j}) & = 1 + \sum_{y \in E} Q(x,y) \tau_y(E_{R_j}) \end{cases}$$

De façon plus concise, $m = Qm$ et $\tau = 1 + Q\tau$.

Démonstration. Soit $x \in E_T$, $T_{E_{R_i}}(X_0 \ldots) = 1 + T_{E_{R_i}}(X_1, \ldots)$. On a

$$\begin{aligned} \tau_x(E_{R_i}) &= \mathbb{E}_x(1 + T_{E_{R_i}}(X_1, \ldots)) = 1 + \mathbb{E}_x(T_{E_{R_i}}(X_1, \ldots)) \\ &= 1 + \mathbb{E}_x(\mathbb{E}_x(T_{E_{R_i}}(X_1, \ldots) \mid \mathcal{F}_1)) \\ &= 1 + \mathbb{E}_x(\mathbb{E}_{X_1}(T_{E_{R_i}})) = 1 + \mathbb{E}_x(\tau_{X_1}(E_{R_i})) \\ &= 1 + \sum_{y \in E} \tau_y(E_{R_i}) \mathbb{P}_x(X_1 = y) \\ &= 1 + \sum_{y \in E} \tau_y(E_{R_i}) Q(x,y) \end{aligned}$$

De même,

$$\begin{aligned} m_x(E_{R_i}) &= \mathbb{P}_x(T_{E_{R_i}}(X_0,\dots) < \infty) = \mathbb{P}_x(1 + T_{E_{R_i}}(X_1,\dots) < \infty) \\ &= \mathbb{P}_x(T_{E_{R_i}}(X_1,\dots) < \infty) = \mathbb{E}_x(1_{T_{E_{R_i}}(X_1,\dots)<\infty}) \\ &= \mathbb{E}_x(\mathbb{E}_x(1_{T_{E_{R_i}}(X_1,\dots)<\infty} \mid \mathcal{F}_1)) = \mathbb{E}_x(\mathbb{E}_{X_1}(1_{T_{E_{R_i}}(X_0,\dots)<\infty})) \\ &= \mathbb{E}_x(m_{X_1}(E_{R_i})) = \sum_{y\in E} Q(x,y) m_y(E_{R_i}) \end{aligned}$$ ■

Exemple 6.4 On considère le graphe

$$\begin{tikzcd} 0 \arrow[loop left, "1"] & 1 \arrow[l, "\alpha"] \arrow[r, bend left, "\beta"] & 2 \arrow[l, bend left, "\gamma"] \arrow[r, "\delta"] & 3 \arrow[loop right, "1"] \end{tikzcd}$$

On a $E = \{0,1,2,3\} = E_T \sqcup E_{R_1} \sqcup E_{R_2}$ avec $E_T = \{1,2\}$, $E_{R_1} = \{0\}$ et $E_{R_2} = \{3\}$. Calculons les probabilités d'absorption en $\{0\}$. On note $m_x(0) = m_x(E_{R_1})$ pour faire plus court. On a

$$\begin{cases} m_0(0) &= 1 \\ m_1(0) &= \alpha m_0(0) + \beta m_2(0) \\ m_2(0) &= \gamma m_1(0) + \delta m_3(0) \\ m_3(0) &= 0 \end{cases} \Rightarrow \begin{cases} m_0(0) &= 1 \\ m_1(0) &= \frac{\alpha}{1-\beta\gamma} \\ m_2(0) &= \frac{\alpha\gamma}{1-\beta\gamma} \\ m_3(0) &= 0 \end{cases}$$

Pour les temps d'absorption, on a

$$\begin{cases} \tau_0(0) &= 0 \\ \tau_1(0) &= 1 + \alpha\tau_0(0) + \beta\tau_2(0) \\ \tau_2(0) &= 1 + \gamma\tau_1(0) + \delta\tau_3(0) \\ \tau_3(0) &= 0 \end{cases} \Rightarrow \begin{cases} \tau_0(0) &= 0 \\ \tau_1(0) &= \frac{1+\beta}{1-\beta\gamma} \\ \tau_2(0) &= \frac{1+\gamma}{1-\beta\gamma} \\ \tau_3(0) &= 0 \end{cases}$$

Définition 6.5 On définit τ_x^* par $\tau_x^*(E_{R_i}) = \mathbb{E}_x(T_{E_{R_i}} 1_{T_{E_{R_i}}<\infty})$.

THÉORÈME 6.2 *Si* $x \in E_{R_i}$, $\tau_x^*(E_{R_i}) = 0$. *Sinon*

$$\tau_x^*(E_{R_i}) = \sum_{y\in E} Q(x,y)\mathbb{P}_y(T_{E_{R_i}} < \infty) + \sum_{y\in E} Q(x,y)\tau_y^*(E_{R_i})$$

Démonstration. Si $x \in E_{R_i}$, $T_{E_{R_i}} = 0$ donc $\tau_x^*(E_{R_i}) = 0$. Sinon, on a

$T_{E_{R_i}}(X_0,\dots) = 1 + T_{E_{R_i}}(X_1,\dots)$. On a

$$\begin{aligned}
\tau_x^*(E_{R_i}) &= \mathbb{E}_x(T_{E_{R_i}} 1_{T_{E_{R_i}}<\infty}) \\
&= \mathbb{E}_x(\mathbb{E}_x((1+T_{E_{R_i}}(X_1,\dots))1_{T_{E_{R_i}}<\infty} \mid \mathcal{F}_1 \\
&= \mathbb{E}_x(\mathbb{E}_{X_1}((1+T_{E_{R_i}}(X_0,\dots))1_{T_{E_{R_i}}<\infty})) \\
&= \sum_{y\in E} Q(x,y)\mathbb{E}_y((1+T_{E_{R_i}}(X_0,\dots))1_{T_{E_{R_i}}<\infty}) \\
&= \sum_{y\in E} Q(x,y)\mathbb{E}_y(1_{T_{E_{R_i}}<\infty}) + \sum_{y\in E} Q(x,y)\mathbb{E}_y(T_{E_{R_i}} 1_{T_{E_{R_i}}<\infty}) \\
&= \sum_{y\in E} Q(x,y)\mathbb{P}_y(T_{E_{R_i}}<\infty) + \sum_{y\in E} Q(x,y)\tau_y^*(E_{R_i}) \qquad \blacksquare
\end{aligned}$$

6.2 Mesure invariante

6.2.1 Définitions et exemples

On se donne un espace E au plus dénombrables et Q une matrice stochastique sur E.

<u>Définition 6.6</u> Soit μ une mesure positive sur E non nulle telle que $\mu(x) < \infty$ pour tout $x \in E$. On dit que μ est invariante si $\mu(y) = \sum_{x\in E} Q(x,y)\mu(x)$ ie $\mu Q = \mu$.

Supposons E fini. On considère alors $\frac{\mu}{\mu(E)}$ mesure de probabilité. Soit $f : E \to \mathbb{R}^+$.

$$\begin{aligned}
\mathbb{E}_\mu(f(X_1)) &= \sum_{x\in E}\mu(x)\mathbb{E}_x(f(X_1)) = \sum_{x\in E}\mu(x)\sum_{y\in E} f(y)Q(x,y) \\
&= \sum_{y\in E} f(y)\sum_{x\in E}\mu(x)Q(x,y) = \sum_{y\in E} f(y)\mu(y)
\end{aligned}$$

Si μ est invariante, la loi de X_n sous $\mathbb{P}_\mu$ est constante égale à μ.

Exemple 6.5

- $\pi = (\frac{q}{p+q}, \frac{p}{p+q})$ est invariante pour la chaîne à deux états.
- Pour la marche simple sur $\mathbb{Z}^d$, $Q(x,y) = \mu(y-x)$, la mesure de comptage sur $\mathbb{Z}^d$ est invariante.

<u>Définition 6.7</u> Soit μ une mesure positive sur E finie en tout point. On dit que μ est réversible ssi pour tout $x, y \in E$,

$$\mu(x)Q(x,y) = \mu(y)Q(y,x)$$

Exemple 6.6

- Marche aléatoire sur $\mathbb{Z}$.
  $Q(i,i+1)=p$, $Q(i,i-1)=1-p=:q$.
  Pour tout i, $\mu(i)Q(i,i+1)=\mu(i+1)Q(i+1,i)$ et $\mu(i)=(\frac{p}{q})^i$ convient.
- Soit G un graphe fini, on note A_x l'ensemble des voisins de x et on suppose $|A_x|<\infty$ pour tout x. X_n est la marche aléatoire simple sur G. On a $Q(x,y)=\frac{1}{|A_x|}$. $\mu(x)=|A_x|$ est une mesure réversible.
- Urne d'Ehrenfest : $Q(j,j+1)=1-\frac{j}{d}$ et $Q(j,j-1)=\frac{j}{d}$. $\mu(j)=\binom{d}{j}$ convient.

Proposition 6.5 Si μ est réversible alors μ est invariante.

Démonstration. Soit $x,y\in E^2$.

$$\mu(x)Q(x,y)=\mu(y)Q(y,x)$$

Donc

$$\sum_{x\in Z}\mu(x)Q(x,y)=\mu(y)\sum_{x\in E}Q(y,x)$$

donc $\mu Q(y)=\mu(y)$. ∎

<u>Théorème 6.3</u> *Soit x un état récurrent. La formule*

$$\nu_x(y)=\mathbb{E}_x\left(\sum_{k=0}^{T_x-1}1_{X_k=y}\right)$$

définit une mesure invariante. De plus, $\nu_x(y)>0$ ssi y est dans la classe de récurrence de x.

Démonstration. Si $y\notin\overline{x}$ alors $G(x,y)=0$ donc $\sum_{n\in\mathbb{N}}Q_n(x,y)=0$. Donc

$$0=\sum_{n\in\mathbb{N}}\mathbb{E}_x(1_{X_n=y})=\mathbb{E}_x\left(\sum_{n=0}^{\infty}1_{X_n=y}\right)$$

donc $\nu_x(y)=0$.

On écrit ensuite

$$\begin{aligned}
\nu_x(y) = \mathbb{E}_x(= \left(\sum_{k=1}^{T_x} 1_{X_k=y}\right) &= \mathbb{E}_x\left(\sum_{k=1}^{\infty} 1_{X_k=y}1_{k\leqslant T_x}\right) \\
&= \mathbb{E}_x\left(\sum_{z\in E}\sum_{k=1}^{\infty} 1_{X_k=y}1_{X_{k-1}=z}1_{k\leqslant T_x}\right) \\
&= \mathbb{E}_x\left(\sum_{z\in E}\sum_{k=1}^{\infty} 1_{X_{k-1}=z}1_{k\leqslant Tx}\mathbb{E}_x(1_{X_k=y} \mid \mathcal{F}_{k-1})\right) \\
&= \mathbb{E}_x\left(\sum_{z\in E}\sum_{k=1}^{\infty} 1_{X_{k-1}=z}1_{k\leqslant Tx}Q(z,y)\right) \\
&= \sum_{z\in E} Q(z,y)\mathbb{E}_x\left(\sum_{k=1}^{\infty} 1_{X_{k-1}=z}1_{k\leqslant T_x}\right) \\
&= \sum_{z\in E} Q(z,y)\mathbb{E}_x\left(\sum_{k=1}^{T_x} 1_{X_{k-1}=z}\right) = \sum_{z\in E} Q(z,y)\nu_x(z)
\end{aligned}$$

ie $\nu_x Q^n = \nu_x$.

Par définition $\nu_x(x) = 1$ et

$$\nu_x(x) = \sum_{z\in E} Q_n(z,x)\nu_x(z) = \sum_{z\in E} Q(z,x)\nu_x(z)$$

Si $y \in \overline{x}$, il existe n, m tel que $Q_n(y,x) > 0$ et $Q_m(x,y) > 0$. Alors la relation précédente implique $v_x(y) < \infty$. de même, on trouve

$$\nu_x(y) = \sum_{n=0}^{\infty} Q_n(z,y)\nu_x(z) > 0 \qquad \blacksquare$$

6.2.2 Mesure invariante et récurrence

<u>Théorème 6.4</u> *Soit X_n une chaîne récurrente irréductible (tous les états communiquent et sont tous récurrents). Alors la mesure invariante est unique à scalaire près.*

Démonstration. Soit μ une mesure invariante. On montre par récurrence que

$$\mu(y) \geqslant \mu(x)\mathbb{E}_x\left(\sum_{k=0}^{(T_x-1)\wedge p} 1_{X_k=y}\right)$$

Si $x = y$ alors c'est trivial, de même que si $p = 0$.

On suppose la formule vraie pour $p \geqslant 1$ fixé. On a

$$\begin{aligned}
\mu(y) = \sum_{z\in E} Q(z,y)\mu(z) &\geqslant \sum_{z\in E} Q(z,y)\mu(x)\mathbb{E}_x\left(\sum_{k=0}^{(T_x-1)\wedge p} 1_{X_k=z}\right) \\
&= \mu(x)\sum_{z\in E}\sum_{k=0}^{p}\mathbb{E}_x(1_{X_k=z}1_{T_x-1\geqslant k}1_{X_{k+1}=y}) \\
&= \mu(x)\mathbb{E}_x\left(\sum_{k=0}^{p\wedge(T_x-1)} 1_{X_{k+1}=y}\right) = \mu(x)\mathbb{E}_x\left(\sum_{k=1}^{(p+1)\wedge T_x} 1_{X_k=y}\right)
\end{aligned}$$

Par récurrence, on a la formule. On fait alors tendre p vers ∞ et on a $\mu(y) \geqslant \mu(x)\nu_x(y)$.

Par invariance de μ, on a pour tout n,

$$\mu(s) = \sum_{z\in E}\mu(z)Q_n(z,x) \geqslant \sum_{z\in E}\mu(x)\nu_x(z)Q_n(z,x) = \sum_{s\in Z}\mu(x)\nu_x(x) = 1$$

Dès que $Q_n(z,x) > 0$, on a $\mu(z) = \mu(x)\nu_x(z)$. ∎

<u>Théorème 6.5</u> *Supposons la chaîne irréductible récurrente.*

- *ou bien il existe une (unique) probabilité invariante μ sur E, auquel cas $\mathbb{E}_x(T_x) = \frac{1}{\mu(x)} < \infty$ pour tout $x \in E$.*
- *ou bien toute mesure invariante est de masse infinie et dans ce cas $\mathbb{E}_x(T_x) = \infty$.*

<u>Définition 6.8</u> Dans le premier cas, on dit que la chaîne est récurrente positive. Dans le second, on dit qu'elle est récurrente nulle.

Démonstration. Soit μ l'unique probabilité invariante dans le premier cas. On a $\mu(y) = \frac{\nu_x(y)}{\nu_x(E)}$ et $\mu(x) = \frac{1}{\nu_x(E)}$.

$$\nu_x(E) = \sum_{y\in E}\nu_x(y) = \sum_{y\in E}\mathbb{E}_x\left(\sum_{k=0}^{T_x-1}1_{X_k=y}\right) = \mathbb{E}_x\left(\sum_{k=0}^{T_x-1}\sum_{y\in E}1_{X_k=y}\right) = \mathbb{E}_x(T_x)$$

Le même calcul est valable dans le cas récurrent nul. ∎

Proposition 6.6 Supposons la chaîne irréductible. S'il existe une probabilité invariante, alors la chaîne est récurrente (positive).

Proposition 6.7 Soit X_n une chaîne irréductible. S'il existe une mesure invariante finie alors la chaîne est récurrente (donc récurrente positive).

Démonstration. Soit γ une mesure invariante finie et $y \in E$ tel que $\gamma(y) > 0$. On a vu que

$$G(x, y) = \mathbb{P}_x(T_y < \infty)G(y, y)$$

Donc $\sum_{n\in\mathbb{N}} Q_n(x, y) \leqslant G(y, y)$. On multiplie par $\gamma(x)$ et on somme sur x.

$$\sum_{x\in E}\sum_{n}\gamma(x)Q_n(x, y) \leqslant \sum_{x\in E}\gamma(x)G(y, y)$$

En permutant les sommes :

$$\sum_{n}\gamma Q_n(y) \leqslant \gamma(E)G(y, y)$$

Donc comme $\gamma Q_n(y) = \gamma(y) > 0$, $G(y, y) = +\infty$, ie y est récurrent.

Par irréductibilité, tous les états sont récurrents. ∎

Exemple 6.7 $Q(k, k + 1) = p = 1 - Q(k, k - 1)$ et $Q(0, 1) = 1$. La mesure $\mu(k) = (\frac{p}{q})^{k-1}$ avec $\mu(0) = q$ est réversible donc invariante.

Si $p < q$ μ est finie et la chaîne est récurrente.

Bilan : Il existe une partition $E = E_T \sqcup (E_{R_i})$ avec

$$\begin{cases} x \in E_T \quad \text{ssi} \quad \mathbb{P}_x(T_x < \infty) < 1 \\ x \in E_{R_i} \quad \text{ssi} \quad \mathbb{P}_x(T_x < \infty) = 1 \end{cases}$$

Notions de mesures réversibles $\pi(x)Q(x, y) = \pi(y)Q(y, x)$ et invariantes $\pi Q = \pi$ (inversible $\Rightarrow$ invariant).

Sur chaque classe E_{R_i}, il existe une mesure invariante

$$\nu_x^i(y) = \mathbb{E}_x\left(\sum_{k=0}^{T_x-1} 1_{X_k=y}\right)$$

Toute combinaison linéaire des ν^i définit une mesure invariante sur E.

Si la chaîne est irréductible, il existe une unique mesure invariante (à scalaire près). Dans ce cas, si la mesure est finie, la chaîne est récurrente positive : $\mathbb{E}_x(T_x) < \infty$.

Si toute mesure invariante est infinie, $\mathbb{E}_x(T_x) = \infty$ (chaîne récurrente nulle).

Chapitre 7

Comportement asymptotique

7.1 Convergence à l'équilibre

7.1.1 Cas E fini

On a vu que Q est une matrice stochastique donc toutes les puissances ont des coefficients strictement positifs à partir d'uncertain rang.

Alors $Q^n \to Q^\infty$ où Q^∞ est stochastique et toutes les lignes sont identiques égales à π. On vérifie que $\pi Q = \pi$.

Donc pour tout $x, y \in E$, $Q^n(x,y) \to \pi(y)$ et la convergence est exponentielle : $|Q^n(x,y) - \pi(y)| \leqslant Cste \times \theta^n$.

7.1.2 Le cas général

Soit E une espace au plus dénombrable, Q stochastique sur E, X_n la chaîne canonique associée.

Lemme 7.0.1
Soit (X_n) une chaîne récurrente irréductible. Si la chaîne est apériodique alors pour tout $x, y \in E$, il existe $n_0 \in \mathbb{N}$ tel que $Q^n(x,y) > 0$ pour $n \geqslant n_0$.

Démonstration. Il suffit de traiter le cas $x = y$. Comme $d_x = 1$, il exist eN tel que $Q_N(x,x) > 0$ et $Q_{N+1}(x,x) > 0$. Alors pour $j \in [\![0, N-1]\!]$, on a

$$Q_{N^2+j}(x,x) = Q_{j(N+1)+(N-j)N}(x,x) > 0 \qquad \blacksquare$$

Donc on a une puissance de Q avec des coefficients strictement positifs donc toutes les puissances ultérieures le sont.

<u>Théorème 7.1</u> *Supposons la chaîne récurrence positive, irréductible et apériodique. Si π est l'unique mesure de probabilité invariantes, alors pour tout*

$x \in E$

$$\sum_{y \in E} |Q_n(x,y) - \pi(y)| \to 0$$

Démonstration. Sur $E \times E$, on considère la chaîne de Markov canonique (X_n^1, X_n^2) de matrice de transition $\overline{Q}((x_1,x_2),(y_1,y_2)) = Q(x,y_1)Q(x_2,y_2)$. On vérfife facilement que c'est une matrice stochastique.

On remarque que (X_n^1, X_n^2) est irréductible. En effet, par le lemme précédent, il existe n_1, n_2 tel que $Q_n(x_1,y_1) > 0$ pour $n \geqslant n_1$ et $Q_n(x_2,y_2) > 0$ pour $n \geqslant n_2$.

Pour $N \geqslant n_1 \vee n_2$, on a $\overline{Q_n}((x_1,x_2),(y_1,y_2)) > 0$. La chaîne est récurrente positive. En effet, la mesure $\pi \otimes \pi$ est invariante pour $\overline{Q}$.

$$\begin{aligned}
(\pi \otimes \pi)\overline{Q}(y_1,y_2) &= \sum_{(x_1,x_2)} \pi(x_1)\pi(x_2)\overline{Q}((x_1,x_2),(y_1,y_2)) \\
&= \sum_{x_1,x_2} \pi(x_1)\pi(x_2)Q(x_1,y_1)Q(x_2,y_2) = \pi(y_1)\pi(y_2)
\end{aligned}$$

La chaîne est récurrence positive. On note $\Delta = \{(x,x), x \in E\}$.

$$T = T_\Delta = \inf\{n \geqslant 0, (X_n^1, X_n^2) \in \Delta\} = \inf\{n, X_n^1 = X_n^2\}$$

est fini presque sûrement. On remarque ensuite que

$$\begin{aligned}
\mathbb{P}_x(X_n = y) - \pi(y) &= \overline{\mathbb{P}_{\pi\otimes\delta_x}}(X_n^2 = y) - \overline{\mathbb{P}_{\pi\otimes\delta_x}}(X_n^1 = y) \\
&= \overline{\mathbb{E}_{\pi\otimes\delta_x}}(1_{X_n^2=y} - 1_{X_n^1=y})
\end{aligned}$$

On distingue suivant les valeurs de T (fini p.s. sous $\overline{P}_{\pi\otimes\delta_x}$).

$$\begin{aligned}
\mathbb{P}_x(X_n = y) - \pi(y) &= \overline{\mathbb{E}_{\pi\otimes\delta_x}}(1_{T>n}(1_{X_n^2=y} - 1_{X_n^1=y})) \\
&\quad + \sum_{k=0}^{n} \overline{\mathbb{E}_{\pi\otimes\delta_x}}(1_{T=k}(1_{X_n^2=y} - 1_{X_n^1=y})) \\
&= \overline{\mathbb{E}_{\pi\otimes\delta_x}}(1_{T>n}(1_{X_n^2=y} - 1_{X_n^1=y})) \\
&\quad + \sum_{z \in E}\sum_{k=0}^{n} \overline{\mathbb{E}_{\pi\otimes\delta_x}}(1_{T=k}1_{X_k^1=X_k^2=z}(1_{X_n^2=y} - 1_{X_n^1=y}))
\end{aligned}$$

Or

$$\overline{\mathbb{E}_{\pi\otimes\delta_x}}(1_{T=k}1_{X_k^1=X_k^2=z}1_{X_n^2=y}) = \overline{\mathbb{E}_{\pi\otimes\delta_x}}(1_{T=k}1_{X_k^1=X_k^2=z})Q_{n-k}(z,y)$$

qui est symétrique en X_1, X_2 donc le double somme est nulle. On a donc

$$\mathbb{P}_x(X_n = y) - \pi(y) = \overline{\mathbb{E}_{\pi\otimes\delta_x}}(1_{T>n}(1_{X_n^2=y} - 1_{X_n^1=y})$$

En sommant on obtient

$$\begin{aligned}\sum_{y\in E}|\mathbb{P}_x(X_n=y)-\pi(y)| &= \sum_{y\in E}\overline{\mathbb{E}_{\pi\otimes\delta_x}}(1_{T>n}(1_{X_n^2=y}-1_{X_n^1=y}))\\ &\leqslant \overline{\mathbb{E}_{\pi\otimes\delta_x}}\left(1_{T>n}\sum_{y\in E}|1_{X_n^2=y}-1_{X_n^1=y}|\right)\\ &\leqslant 2\overline{\mathbb{E}_{\pi\otimes\delta_x}}(1_{T>n}) = 2\overline{\mathbb{P}_{\pi\otimes\delta_x}}(T>n)\to 0\end{aligned}$$ ■

Remarque 7.1 *L'apériodicité est essentielle. On prend par exemple* $E=\{0,1\}$, $Q=\begin{pmatrix}0&1\\1&0\end{pmatrix}$. *On a* $Q^{2n}=\mathrm{Id}$ *et* $Q^{2n+1}=Q$ *et on n'a pas* $Q^n(x,y)\to\pi(y)$.

Même remarque pour la chaîne d'Ehrenfest qui est de période 2. La loi $B(d,\frac{1}{2})$ *est invariante, la chaîne est irréductible apériodique et récurrente positive mais on n'a pas* $Q^n\to(B(d,\frac{1}{2}),\ldots,B(d,\frac{1}{2}))^T$.

7.2 Théorème ergodique

<u>Théorème 7.2</u> Ergodique *Soit* X_n *uen chaîne récurrente irréductible et* μ *une mesure invariante. Soit* f *et* g *deux fonctions mesurables positives telles que* $\int f\,\mathrm{d}\mu<\infty$ *et* $0<\int g\,\mathrm{d}\mu<\infty$.

Pour tout $x\in E$, $\mathbb{P}_x$*-p.s., on a*

$$\frac{\sum_{k=0}^{n-1}f(X_k)}{\sum_{k=0}^{n-1}g(X_k)}\to\frac{\int f\,\mathrm{d}\mu}{\int g\,\mathrm{d}\mu}$$

presque sûrement.

Remarque 7.2 *On peut supprimer l'hypothèse* $\int f\,\mathrm{d}\mu<\infty$ *quitte à considérer* f *comme limite croissante des* f_n *avec* $\int f_n\,\mathrm{d}\mu<\infty$.

Corollaire 7.1 LGN pour les chaînes de Markov *Si* X_n *est irréductible récurrente positive et si* π *est l'unique probabilité invariante alors pour tout* $x\in E$, $\mathbb{P}_x$*-p.s., on a :*

$$\frac{1}{n}\sum_{k=0}^{n-1}f(X_k)\to\int f\,\mathrm{d}\pi$$

Démonstration. On pose $T_x^0=0$, $T_x^1=\inf\{n,X_n=x\}$ et $T_x^{n+1}=\inf\{k>T_x^n,X_k=x\}$.

Comme la chaîne est récurrence les T_x^n sont finis $\mathbb{P}_x$-p.s.

On pose aussi $Z_k(f) = \displaystyle\sum_{m=T_x^k}^{T_x^{k+1}-1} f(X_m)$.

Lemme 7.2.1
Les variables $(Z_k(f))_k$ sont iid.

Démonstration. Il suffit de montrer que pour tout $h_0, \dots, h_k$ mesurables bornées,

$$\mathbb{E}_x\left(\prod_{i=0}^{k} h_i(Z_i(f))\right) = \prod_{i=0}^{k} \mathbb{E}_x(h_i(Z_0(f)))$$

On le fait par récurrence sur k, sachant que $k = 0$ est trivial. Si on suppose la relation vraie pour $k-1$ fixé, on sait que $Z_0(f), \dots, Z_{k-1}(f)$ sont F_k-mesurables et θ_{T_k} est indépendante de F_{T_k} et de loi $\mathbb{P}_x$.

$$Z_k(f) = Z_0(f) \circ \theta_{T_k}$$

Par Markov, on a

$$\begin{aligned}
\mathbb{E}_x\left(\prod_{i=0}^{k} h_i(Z_i(f))\right) &= \mathbb{E}_x\left(\prod_{i=0}^{k-1} h_i(Z_i(f)) h_k(Z_0(f) \circ \theta_{T_k})\right) \\
&= \mathbb{E}_x\left(\prod_{i=0}^{k-1} h_i(Z_i(f)) \mathbb{E}_x(h_k(Z_0(f)))\right) \\
&= \prod_{i=0}^{k-1} \mathbb{E}_x(h_i(Z_i(f))) \mathbb{E}_x(h_k(Z_0(f))) = \prod_{i=0}^{k} \mathbb{E}_x(h_i(Z_0(f)))
\end{aligned}$$

■

La chaîne est irréductible récurrente donc toutes les mesures invariantes sont proportionnelles à

$$nu_x(y) = \mathbb{E}_x\left(\sum_{k=0}^{T_x^1-1} 1_{X_k=y}\right)$$

En particulier, la mesure μ s'écrit $\mu(y) = \mu(x)\nu_x(y)$ car $\nu_x(x) = 1$. On observe que

$$\begin{aligned}
\mathbb{E}_x(Z_0(f)) &= \mathbb{E}_x\left(\sum_{k=0}^{T_x^1-1} f(X_k)\right) \\
&= \mathbb{E}_x\left(\sum_{y \in E} \sum_{j=0}^{T_x^1-1} f(y) 1_{X_k=y}\right) \\
&= \sum_{y \in E} f(y)\nu_x(y) = \frac{1}{\mu(x)} \int f \,\mathrm{d}\mu
\end{aligned}$$

D'après la loi forte des grands nombres (usuelle), $\mathbb{P}_x$-p.s.

$$\frac{1}{n}\sum_{k=0}^{n-1} Z_k(f) \to \frac{1}{\mu(x)}\int f\,\mathrm{d}\mu$$

On note $N_x(n) = \sum_{k=0}^{n} 1_{X_k = x}$ de sorte que

$$T_x^{N_x(n)} \leqslant n < T_x^{N_x(n)+1}$$

On a donc

$$\frac{1}{N_x(n)}\sum_{k=0}^{T_x^{N_x(n)}-1} f(X_k) \leqslant \frac{1}{N_x(n)}\sum_{k=0}^{n-1} f(X_k) \leqslant \frac{1}{N_x(n)}\sum_{k=0}^{T_x^{N_x(n)+1}-1} f(X_k)$$

Puis

$$\frac{1}{N_x(n)}\sum_{k=0}^{N_x(n)-1} Z_k(f) \leqslant \frac{1}{N_x(n)}\sum_{k=0}^{n-1} f(X_k) \leqslant \frac{1}{N_x(n)}\sum_{k=0}^{N_x(n)} f(X_k)$$

Par la LGN,

$$\frac{1}{N_x(n)}\sum_{k=0}^{n-1} f(X_k) \to \frac{1}{\mu(x)}\int f\,\mathrm{d}\mu$$

et

$$\frac{1}{N_x(n)}\sum_{k=0}^{n-1} g(X_k) \to \frac{1}{\mu(x)}\int g\,\mathrm{d}\mu$$

En faisant le rapport, on obtient le résultat. ■

