

Probabilité et Martingale

Classe sino-française, USTC

Responsables du cours :

Prof. **Philippe Marchal**¹ et

Prof. **SUN WEN**²

D'après un polycopié de

LI Xiao³ et **CHEN Ruikai**

Février 2022 -Avril 2022

1. www.math.univ-paris13.fr/~marchal/
2. sites.google.com/view/wensun-proba
3. home.ustc.edu.cn/~cnlx

Table des matières

1 Révision de la théorie de probabilité	5
1.1 Convergence de variables aléatoires	5
1.2 Loi des grands nombres	6
1.3 Convergence en loi	7
1.4 Convergence des mesures empiriques	13
1.5 Vecteurs Gaussiens	14
2 Espérance conditionnelle	19
2.1 Cas discret	20
2.2 Cas général	24
2.3 Variables aléatoires positives	26
2.4 Comment calculer les espérances conditionnelles ?	32
2.5 Espérance conditionnelle pour les vecteurs gaussiens	34
3 Martingale	37
3.1 Définitions et exemples	37
3.2 Temps d'arrêt	41
3.3 Convergence presque sûre des martingales	46
3.4 Application : extinction des processus de Galton-Watson critiques	49
3.5 Urne de Pólya	52
3.6 Marches aléatoires et martingales exponentielles	56

3.7	Martingales bornées dans L^p , $p > 1$	58
3.8	Uniforme intégrabilité	60
3.9	Martingales rétrogrades	65
4	Chaînes de Markov	69
4.1	Définition et premières propriétés	69

Chapitre 1

Révision de la théorie de probabilité

1.1 Convergence de variables aléatoires

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ espace de probabilités, X variable aléatoire à valeurs dans un espace métrique (M, d) : c'est une fonction mesurable $(\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (M, d)$. Alors on voit la convergence de variables aléatoires comme des fonctions :

- **La convergence presque sure** : on rappelle la convergence ponctuelle d'une suite de fonctions $(f_n) : \forall x, \mathcal{F}_n(x) \rightarrow f(x)$. Pour les variable aléatoires, une version plus faible de la convergence ponctuelle est la convergence presque sure :

$$X_n \xrightarrow{p.s.} X \quad \text{si} \quad \exists B \subset \mathcal{F}, \mathbb{P}(B) = 0 \quad t.q. \quad \forall \omega \notin B, X_n(\omega) \rightarrow X(\omega).$$

- **La convergence dans L^p** : Si $1 \leq p < \infty$, on dit que (X_n) converge vers X dans L^p , si $\mathbb{E}(|X_n - X|^p) \xrightarrow{n \rightarrow \infty} 0$. Si $p = \infty$, on dit que X_n converge vers X dans L^∞ si $\mathbb{E}(\sup |X_n - X|) \xrightarrow{n \rightarrow \infty} 0$.
- **convergence en probabilité** : On dit que X_n converge X en probabilité si

$$\forall \epsilon > 0, \quad \mathbb{P}(|X_n - X| > \epsilon) \xrightarrow{n \rightarrow \infty} 0.$$

Si on définit la distance $d(X, Y) = \mathbb{E}(|X - Y| \wedge 1)$, où $|X - Y| \wedge 1 = \inf(1, |X - Y|)$.

Alors

$$X_n \xrightarrow{(p)} X \iff d(X_n, X) \xrightarrow{n \rightarrow \infty} 0.$$

Proposition 1.1.1. *Si (X_n) converge vers X p.s. ou dans L^p ($p \geq 1$), alors $X_n \xrightarrow{p} X$. La réciproque est fautive mais si $X_n \xrightarrow{p} X$, il existe $(n_k)_{k \geq 0}$ t.q. $X_{n_k} \xrightarrow{p.s.} X$.*

Contre-exemple : Soient $(Y_n), (Z_n)$ variables aléatoires indépendantes, uniformes sur $[0, 1]$, posons

$$X_n = 1_{\{|Y_n - Z_n| \geq \frac{1}{n}\}},$$

alors X_n converge en probabilité vers $X = 0$:

$$\forall \epsilon > 0, \mathbb{P}(|X_n - X| > \epsilon) = \mathbb{P}(|X_n| > \epsilon) = \mathbb{P}(X_n = 1) = \mathbb{P}(|Y_n - Z_n| \geq \frac{1}{n}) \xrightarrow[n \rightarrow \infty]{} 0.$$

Il n'y a pas convergence presque sûre : notons que

$$\mathbb{P}(|X_n - X| > \epsilon) = \mathbb{P}(|Y_n - Z_n| \geq \frac{1}{n}) \geq \frac{1}{n},$$

et

$$\sum_{n \geq 1} \mathbb{P}(|X_n - X| \geq \epsilon) = \infty$$

Comme les événements $\{|X_n - X| > \epsilon\} = \{X_n = 1\}$ sont indépendants. D'après le lemme de Borel-Cantelli, p.s., il existe une infinité d'entiers n tels que $\{X_n = 1\}$ se produise. Donc p.s., $|X_n - X| = 1$ pour une infinité d'entier n , p.s. (X_n) ne converge pas vers X .

Proposition 1.1.2. *Si $X_n \xrightarrow{p} X$ et si (X_n) est bornée dans L^q , alors $\forall p \in [1, q)$, X_n converge vers X dans L^p .*

1.2 Loi des grands nombres

Loi de rien et du tout Soient (X_n) suite de variables aléatoires indépendantes, (\mathcal{B}_n) la tribu engendrée par les variables aléatoires $(X_k)_{k \geq n}$ (i.e. la plus petite tribu contenant tous les événements mesurables pour les variables aléatoires $X_k, k \geq n$). Alors \mathcal{B} est une famille décroissante de tribus. Soit $\mathcal{B}_\infty = \bigcap_{n \geq 1} \mathcal{B}_n$ la tribu asymptotique, alors pour tout $A \in \mathcal{B}_\infty$, on a $\mathbb{P}(A) = 0$ ou 1.

Proposition 1.2.1. *Soit (X_n) une suite de variables aléatoires indépendantes, de même loi*

$$\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}.$$

Posons $S_0 = 0, S_n = X_1 + \dots + X_n$, alors p.s. $\sup_n S_n = \infty, \inf_n S_n = -\infty$. Et on en déduit que p.s. il existe une infinité de n t.q. $S_n = 0$.

Théorème 1.2.2. Soit (X_n) une suite de variables aléatoires i.d.d. (indépendantes, identiquement distribuées) telle que $\mathbb{E}(|X_1|)$ existe ($X_1 \in L^1$). Alors

$$p.s. \quad \frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow[n \rightarrow \infty]{} \mathbb{E}(X_1).$$

(En fait, on a aussi convergence dans L^1 . On le verra en étudiant les martingales inverses.)

Contre-exemple : Soit (X_n) indépendantes, à valeurs dans \mathbb{Z} :

$$\forall k \geq 0, \quad \mathbb{P}(X_n = k) = \mathbb{P}(X_n = -k) = \frac{c}{k^\alpha}, \quad 1 < \alpha \leq 2.$$

Alors X_n n'est pas dans L^1 et on peut montrer que

$$\mathbb{P}\left(\frac{X_1 + \cdots + X_n}{n^{\alpha-1}} \geq x\right)$$

converge pour tout $x \in \mathbb{R}$, vers $F(x)$ où F est une fonction non triviale. En particulier, $\frac{X_1 + \cdots + X_n}{n}$ ne converge pas vers 0. (Lois stables)

1.3 Convergence en loi

Définition 1.3.1. On dit que (X_n) converge en loi vers X si pour toute fonction continue bornée f , on a $\mathbb{E}(f(X_n)) \rightarrow \mathbb{E}(f(X))$. Notation : $X_n \xrightarrow{(l)} X$ ou $X_n \xrightarrow{(d)} X$.

Remarque 1.3.2. 1. Si (X_n) est à valeurs dans \mathbb{N} (plus généralement, dans un ensemble discret), alors

$$X_n \xrightarrow{(l)} X \iff \forall k \in \mathbb{N}, \mathbb{P}(X_n = k) \rightarrow \mathbb{P}(X = k).$$

En effet, il suffit de prendre la fonction $f(x) = 1_{\{x=k\}}$.

2. Si X_n sont des variables aléatoires à valeurs dans \mathbb{R}^d ayant une densité f_n et si $f_n \rightarrow f$ presque partout avec $\int f = 1$, alors $X_n \xrightarrow{(l)} X$ où X est une variable aléatoire de densité f . Réciproquement, une variable aléatoire ayant une densité peut être la limite en loi d'une suite de variables aléatoires qui n'ont pas de densité.

Exemple 1.3.3. Soit $X_n = \frac{1}{n}(\delta_0 + \delta_{\frac{1}{n}} + \cdots + \delta_{\frac{n-1}{n}})$ qui n'a pas de densité, mais $X_n \xrightarrow{(l)} X$ où

X est la loi uniforme sur $[0, 1]$. En effet, pour toute $f \in \mathcal{C}_b$, on a

$$\mathbb{E}(f(X_n)) = \frac{1}{n}(f(0) + f(\frac{1}{n}) + \dots + f(\frac{n-1}{n})) \rightarrow \int_0^1 f(x)dx = \mathbb{E}(f(X)).$$

On peut aussi avoir des variables aléatoires (X_n) ayant une densité et qui converge en loi vers X où X n'a pas de densité.

Exemple 1.3.4. Soit X_n la loi uniforme sur $[0, \frac{1}{n}]$, de densité $f_n(x) = n1_{\{x \in [0, \frac{1}{n}]\}}$, mais $X_n \xrightarrow{(l)} X$ où $X = \delta_0$. En effet, pour toute $f \in \mathcal{C}_b$,

$$\mathbb{E}(f(X_n)) = n \int_0^{\frac{1}{n}} f(x)dx \rightarrow f(0),$$

puisque f est continue en 0.

Proposition 1.3.5. 1. Si $X_n \xrightarrow{(p)} X$, alors $X_n \xrightarrow{(l)} X$.

2. Si $X = \delta_a$, et si $X_n \xrightarrow{(l)} X$, alors $X_n \xrightarrow{(p)} X$.

Cas des fonctions de répartition dans \mathbb{R} : Si X est une variable aléatoire à valeurs dans \mathbb{R} , sa fonction de répartition est la fonction $F_X : x \mapsto \mathbb{P}(X \leq x)$. Si F_X est dérivable, sa dérivée est la densité de X .

Proposition 1.3.6. Soit (X_n) une suite de variables aléatoires à valeurs dans \mathbb{R} , alors $X_n \xrightarrow{(l)} X$ s.s.i. $F_{X_n}(x) \rightarrow F_X(x)$ en tout x où la fonction F_X est continue.

Remarque 1.3.7. Les fonctions de répartition sont croissantes, elle sont discontinues en x s.s.i. on a un atome en X , dans ce cas,

$$\mathbb{P}(X = x) = \lim_{y \rightarrow x^+} F_X(y) - \lim_{y \rightarrow x^-} F_X(y).$$

On peut avoir $X_n \xrightarrow{(l)} X$ mais F_{X_n} ne converge pas vers F_X . Par exemple soit X_n est uniforme sur $[0, \frac{1}{n}]$, alors $X_n \xrightarrow{(l)} X$, où $X = \delta_0$. Mais F_{X_n} ne converge pas vers F_X en $x = 0$.

Cas des fonctions à support compact : Ce sont des fonctions telle que le support $\overline{\{x, f(x) \neq 0\}}$ est un compact.

Proposition 1.3.8. Soient X_n, X v.a. à support dans \mathbb{R}^d . Soit H un ensemble de fonctions continues bornées t.q. $\overline{H} \supset C_C(\mathbb{R}^d)$ (\overline{H} est l'adhérence pour la norme sup). Alors on a l'équivalence entre des trois assertions :

1. $X_n \xrightarrow{(l)} X$.
2. $\forall \varphi \in C_C(\mathbb{R}^d)$, on a $\mathbb{E}(\varphi(X_n)) \rightarrow \mathbb{E}(\varphi(X))$.
3. $\forall \varphi \in H$, on a $\mathbb{E}(\varphi(X_n)) \rightarrow \mathbb{E}(\varphi(X))$.

Contre-exemple à la loi des grands nombres : soit X variable symétrique t.q. X a la même loi que $-X$. Si $\mathbb{E}(X)$ existe, alors $\mathbb{E}(X) = \mathbb{E}(-X) = -\mathbb{E}(X) = 0$. Mais il est possible que $\mathbb{E}(|X|)$ n'existe pas, et dans ce cas, on peut avoir un contre-exemple à la loi des grands nombres :

Exemple 1.3.9. Soit X à valeurs dans \mathbb{Z} , $\mathbb{P}(X = k) = \mathbb{P}(X = -k) = \frac{c}{k^\alpha}$, $1 \leq \alpha < 2$. On peut montrer que si X_1, \dots, X_n sont indépendantes, de même loi que X , alors

$$\frac{X_1 + \dots + X_n}{n^{\alpha-1}} \xrightarrow{(p)} Y$$

où Y est une variable aléatoire non triviale. On dit que Y est stable. On en déduit que $\frac{X_1 + \dots + X_n}{n}$ ne converge pas vers δ_0 .

On remarque que la fonction indicatrice n'est pas continue, mais il y a des approximations de f "par l'extérieur" et "par l'intérieur".

Proposition 1.3.10 (régularité des mesures boréliennes). *On a équivalence entre les assertions suivantes :*

1. X_n converge en loi vers X .
2. Pour tout ouvert O , $\liminf \mathbb{P}(X_n \in O) \geq \mathbb{P}(X \in O)$.
3. Pour tout fermé F , $\limsup \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$.
4. Pour tout borélien B t.q. $\mathbb{P}(X \in \partial B) = 0$, on a $\lim \mathbb{P}(X_n \in B) = \mathbb{P}(X \in B)$.

Remarque 1.3.11. 1. En dimension finie, la notion de convergence est assez simple. En dimension infinie, c'est plus compliqué. En analyse fonctionnelle, on définit souvent la convergence en prenant des fonctions tests. Pour la convergence en loi, on fait la même chose. On prend comme fonctions tests les fonctions continues bornées. Mais les fonctions indicatives ne sont pas continues (sauf dans le cas discret). Donc $\mathbb{P}(X_n \in A)$ ne converge pas vers $\mathbb{P}(X \in A)$ même si X_n converge vers X en loi.

2. Si X est à valeurs dans \mathbb{R}^d et a une densité f , alors pour tout ensemble B , $\mathbb{P}(X \in \partial B) = \int_{\partial B} f(x) dx = 0$. Alors si X_n converge en loi vers X , on a $\lim \mathbb{P}(X_n \in B) = \mathbb{P}(X \in B)$.

$$\begin{array}{ccc}
 (0, \frac{2}{n}) & & (1, \frac{2}{n}) \\
 \boxed{\phantom{\hspace{10em}}} & & \\
 (0, \frac{1}{n}) & & (1, \frac{1}{n})
 \end{array}$$

3. Si X_n converge en loi vers X , si on appelle μ_n la loi de X_n , μ la loi de X . On dit que (μ_n) converge étroitement vers μ (en anglais "weak convergence").

Exemple 1.3.12. Soit X_n la loi uniforme sur un rectangle :

Alors $X_n \xrightarrow{(l)} X$ où X suit la loi uniforme sur $I = [0, 1]$, on a

$$\mathbb{P}(X \in \partial I) = \mathbb{P}(X \in I) = 1, \quad \text{mais } \mathbb{P}(X_n \in I) = 0.$$

Définition 1.3.13. Si X est à valeurs dans \mathbb{R}^d , sa fonction caractéristique est la fonction $\phi_X : \mathbb{R}^d \rightarrow \mathbb{C}$, qui envoie ξ vers $\mathbb{E}(e^{i \cdot \langle \xi, X \rangle})$ (transformée de Fourier).

Noter bien : comme ϕ_X est continue et bornée, si X_n converge en loi vers X , alors pour tout $\xi \in \mathbb{R}^d$, on a $\phi_{X_n}(\xi) \xrightarrow{n \rightarrow \infty} \phi_X(\xi)$. En effet, si on pose $g_\xi(y) = e^{i \cdot \langle \xi, y \rangle}$, alors pour tout $\xi \in \mathbb{R}^d$, g_ξ est une fonction continue et bornée, cela implique

$$\phi_{X_n}(\xi) = \mathbb{E}(g_\xi(X_n)) \xrightarrow{n \rightarrow \infty} \mathbb{E}(g_\xi(X)) = \phi_X(\xi).$$

Théorème 1.3.14 (Théorème de Lévy). *Soit (X_n) une suite de variables aléatoires, alors*

$$X_n \xrightarrow{(l)} X \iff \forall \xi \in \mathbb{R}^d, \phi_{X_n}(\xi) \longrightarrow \phi_X(\xi)$$

Théorème 1.3.15 (Théorème central limite). *Soit (X_n) une suite de variables aléatoires i.i.d. (indépendantes identiquement distribuées) telles que $\mathbb{E}(X_n) = m$, $\text{var}(X_n) = \sigma^2$, alors*

$$\sqrt{n} \left(\frac{X_1 + \dots + X_n}{n} - m \right) \xrightarrow[n \rightarrow \infty]{(l)} \mathcal{N}(0, \sigma^2)$$

Remarque 1.3.16. $X \sim \mathcal{N}(\mu, \sigma^2)$ si X a une densité sur \mathbb{R} : $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. En particulier,

$$\forall a, b \in \mathbb{R}, \mathbb{P}\left(\sqrt{n} \left(\frac{X_1 + \dots + X_n}{n} - m \right) \in [a, b]\right) \longrightarrow \mathbb{P}(X \in [a, b]), \text{ avec } X \sim \mathcal{N}(0, \sigma^2).$$

Démonstration. Posons $Y_n = \frac{X_1 + \dots + X_n}{n}$, $m = \frac{(X_1 - m) + \dots + (X_n - m)}{n}$, on a

$$\begin{aligned}\phi_{Y_n}(\xi) &= \mathbb{E}\left(\exp\left(i\xi\left(\frac{(X_1 - m)}{n} + \dots + \frac{(X_n - m)}{n}\right)\right)\right) \\ &= \prod_{k=1}^n \mathbb{E}\left(\exp\left(i\xi\frac{(X_k - m)}{n}\right)\right) = \mathbb{E}\left(\exp\left(i\xi\frac{(X_1 - m)}{n}\right)\right)^n\end{aligned}$$

Or

$$\phi_{(X_1 - m)}(\xi) = 1 + i\xi\mathbb{E}(X_1 - m) - \frac{1}{2}\xi^2\text{var}(X_1) + o(\xi^2) = 1 - \frac{1}{2}\xi^2\text{var}(X_1) + o(\xi^2)$$

On en déduit que

$$\begin{aligned}\log \phi_{\sqrt{n}Y_n}(\xi) &= n \log\left(\mathbb{E}\left(\exp\left(i\frac{\xi}{\sqrt{n}}(X_1 - m)\right)\right)\right) = n \log\left(1 - \frac{1}{2n}\xi^2\text{var}(X_1) + o\left(\frac{\xi}{n}\right)\right) \\ &= n\left(-\frac{1}{2}\frac{\xi^2}{n}\sigma^2 + o\left(\frac{\xi}{n}\right)\right) = -\frac{1}{2}\xi^2\sigma^2 + o(1)\end{aligned}$$

Donc on a

$$\phi_{\sqrt{n}Y_n}(\xi) \xrightarrow[n \rightarrow \infty]{} \exp\left(-\frac{1}{2}\sigma^2\xi^2\right) = \mathbb{E}(\exp(i\xi X)) = \phi(X), \text{ avec } X \sim \mathcal{N}(0, \sigma^2)$$

Donc $\sqrt{n}Y_n \xrightarrow[l]{(l)} X$, or $\sqrt{n}Y_n = \sqrt{n}\left(\frac{X_1 + \dots + X_n}{n} - m\right) \xrightarrow[l]{(l)} X$. □

Remarque 1.3.17. 1. S'il n'existe pas de variance, on peut trouver des contre-exemples (loi stable). Mais on peut relâcher l'hypothèse d'indépendance par

$$\text{cov}(X_n, X_{n+k}) \leq \exp(-c\epsilon), \quad c > 0.$$

2. Informellement, quand on a beaucoup de petits sources d'aléa, et indépendants, on a à la limite une loi gaussienne. Donc les lois gaussiennes sont très utilisées pour modéliser de nombreux phénomènes en science.

Contre-exemple : une suite de variables aléatoires qui convergent en loi mais pas en probabilité. Soient X_1, \dots, X_n, \dots une suite de variables aléatoires i.i.d. telles que $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$. Alors $\mathbb{E}(X_1) = 0$, $\text{var}(X_1) = \sigma^2 = 1$. Soient

$$S_n = X_1 + \dots + X_n,$$

$$U_n = \frac{S_{2^n}}{\sqrt{2^n}}, \quad V_n = X_{2^{n+1}} + \cdots + X_{2^{n+1}}.$$

Alors par le théorème central limite, on a

$$\frac{S_n}{\sqrt{n}} \xrightarrow{(l)} \mathcal{N}(0, \sigma^2) \implies U_n \xrightarrow{(l)} \mathcal{N}(0, \sigma^2).$$

Par l'absurde, on suppose qu'il existe X telle que (U_n) converge vers X en probabilité. Dans ce cas, on a $\mathbb{P}(|U_n - X| \geq \frac{1}{2}) \rightarrow 0$. Donc pour tout $\epsilon > 0$, il existe N t.q. pour tout $n \geq N$, $\mathbb{P}(|U_n - X| \geq \frac{1}{2}) \leq \epsilon$. Comme

$$\{|U_n - U_{n+1}| \geq 1\} \subset \{|U_n - X| \geq \frac{1}{2}\} \cup \{|U_{n+1} - X| \geq \frac{1}{2}\},$$

par l'inégalité triangulaire,

$$\mathbb{P}(|U_n - U_{n+1}| \geq 1) \leq \mathbb{P}(|U_n - X| \geq \frac{1}{2}) + \mathbb{P}(|U_{n+1} - X| \geq \frac{1}{2}) \leq 2\epsilon.$$

Or

$$U_{n+1} = \frac{(X_1 + \cdots + X_{2^n}) + (X_{2^{n+1}} + \cdots + X_{2^{n+1}})}{\sqrt{2^{n+1}}} = \frac{\sqrt{2^n}U_n + V_n}{\sqrt{2^{n+1}}}$$

Donc

$$U_{n+1} - U_n = U_n \left(\frac{1}{\sqrt{2}} - 1 \right) + \frac{1}{\sqrt{2}} \left(\frac{V_n}{\sqrt{2^n}} \right).$$

Soient

$$A_n = \left(\frac{1}{\sqrt{2}} - 1 \right) U_n, \quad B_n = \frac{1}{\sqrt{2}} \left(\frac{V_n}{\sqrt{2^n}} \right),$$

alors A_n, B_n sont indépendantes, et par le théorème central limite, $A_n \rightarrow \mathcal{N}(0, c\sigma^2)$, et $B_n \rightarrow \mathcal{N}(0, c'\sigma^2)$, avec $c = (\frac{1}{\sqrt{2}} - 1)^2$ et $c' = \frac{1}{2}$. Supposons

$$k = \mathbb{P}(\mathcal{N}(0, c\sigma^2) \geq \frac{1}{2}), \quad k' = \mathbb{P}(\mathcal{N}(0, c'\sigma^2) \geq \frac{1}{2}).$$

Alors $\mathbb{P}(A_n \geq \frac{1}{2}) \rightarrow k$ et $\mathbb{P}(B_n \geq \frac{1}{2}) \rightarrow k'$. Comme

$$\{A_n + B_n \geq 1\} \supset \{A_n \geq \frac{1}{2}\} \cap \{B_n \geq \frac{1}{2}\},$$

on peut prendre n assez grand t.q.

$$\mathbb{P}(|U_n - U_{n+1}| \geq 1) \geq \mathbb{P}(A_n \geq \frac{1}{2})\mathbb{P}(B_n \geq \frac{1}{2}) \geq kk' - \epsilon.$$

Mais on sait que $\mathbb{P}(|U_n - U_{n+1}|) \leq \epsilon$, c'est une contradiction pour ϵ assez petit.

1.4 Convergence des mesures empiriques

Soit X une variable aléatoire dans \mathbb{R}^d , par exemple :

$$X = \begin{cases} \text{opinion d'un électron tiré au hasard} \\ \text{vitesse d'un fluide dans un écoulement turbulent} \\ \text{température en un puits} \end{cases}$$

On appelle un échantillonnage ("sampling" en anglais) une suite de variables aléatoires X_1, \dots, X_n i.i.d. de même loi que X . Et on appelle mesure empirique la mesure de probabilité sur \mathbb{R}^d qui vaut $\mu_n = \frac{1}{n}(\delta_{X_1} + \dots + \delta_{X_n})$.

Théorème 1.4.1. *Avec probabilité 1, (μ_n) converge étroitement vers la loi de X . De manière équivalente, soit Y_n défini par $Y_n = \sigma_{X_{U_n}}$, où U_n est la loi uniforme sur $\{1, \dots, n\}$. Donc la loi de Y_n est μ_n , et alors Y_n converge en loi vers X presque sûrement.*

Démonstration. Soit H un sous-ensemble dénombrable dense de $\mathcal{C}_C(\mathbb{R}^d)$ (fonctions à support compact). Soit $\varphi \in H$, on considère les variables aléatoires $\varphi(X_i)$ qui sont indépendantes, de même loi et intégrables (car bornées). Alors loi des grands nombres montre que, avec probabilité 1, $\frac{\varphi(X_1) + \dots + \varphi(X_n)}{n} \rightarrow \mathbb{E}(\varphi(X_1))$, i.e.

$$\int \varphi(x) d\mu_n(x) \rightarrow \int \varphi(x) dP_X(x) \text{ p.s.}$$

donc il existe un ensemble Ω_φ négligeable tel que sur Ω_φ^C ,

$$\int \varphi(x) d\mu_n(x) \rightarrow \int \varphi(x) dP_X(x).$$

Soit $\bar{\Omega} = \cup_{\varphi \in H} \Omega_\varphi$, comme H est dénombrable, on a $P(\bar{\Omega}) = 0$. Alors sur $\bar{\Omega}^C$, on a

$$\forall \varphi \in H, \int \varphi(x) d\mu_n(x) \rightarrow \int \varphi(x) dP_X(x).$$

On a la convergence avec toutes les fonctions $\varphi \in H$, et comme $\bar{H} = \mathcal{C}_C(\mathbb{R}^d)$. D'après Prop 10.3.3, on a la convergence pour toutes les fonctions continues bornées. Donc (μ_n) converge étroitement vers P_X . \square

1.5 Vecteurs Gaussiens

Covariance si X, Y sont des variables aléatoires dans \mathbb{R} , on a défini la covariance de X, Y :

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y),$$

en particulier, si $X = Y$,

$$\text{cov}(X, X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \text{var}(X).$$

Si (X_1, \dots, X_d) sont des variables aléatoires dans \mathbb{R} , on dit que la matrice de covariance de (X_1, \dots, X_d) la matrice $(M_{ij})_{\substack{1 \leq i \leq d \\ 1 \leq j \leq d}}$, où $M_{ij} = \text{cov}(X_i, X_j)$. Comme M est symétrique, on peut l'associer une forme quadratique qui est positive :

$$(\lambda_1, \lambda_d) \mapsto {}^t \lambda M \lambda = \sum_{i,j} \lambda_i M_{ij} \lambda_j = \text{var}\left(\sum \lambda_i X_i\right) \geq 0.$$

Remarque 1.5.1. Quand les X_i sont dans L^2 , la matrice de covariance existe et par Cauchy-Schwarz, $|\mathbb{E}(X_i X_j)| \leq \sqrt{\mathbb{E}(X_i^2)\mathbb{E}(X_j^2)}$.

Proposition 1.5.2. *Si la matrice de covariance existe, c'est une matrice symétrique positive.*

Proposition 1.5.3. *Si M est une matrice carrée $d \times d$ symétrique positive, alors il existe un vecteur (X_1, \dots, X_d) aléatoire t.q.*

$$\phi_{(X_1, \dots, X_d)}(\xi) = \mathbb{E}(\exp(i \langle \xi, X \rangle)) = \exp\left(-\frac{1}{2} {}^t \xi M \xi\right).$$

Le vecteur s'appelle vecteur gaussien de matrice de covariance M . Notation : $(X_1, \dots, X_d) \sim \mathcal{N}(0, M)$.

Théorème 1.5.4 (TCL vectoriel). *Soit (X_n) des variables aléatoires i.i.d. à valeurs dans \mathbb{R}^d de matrice de covariance M , alors*

$$\sqrt{n} \left(\frac{X_1 + \dots + X_n}{n} - \mathbb{E}(X_1) \right) \xrightarrow{(l)} \mathcal{N}(0, M).$$

Définition 1.5.5. On dit que $X = (X_1, \dots, X_d)$ est un vecteur gaussien si toute combinaison linéaire de ses coordonnées suit une loi gaussienne. En particulier, pour tout i , X_i suit une loi gaussienne.

Exemple 1.5.6. Soit (X_i) i.i.d. de la loi $\mathcal{N}(0, \sigma^2)$, alors (X_1, \dots, X_d) est un vecteur gaussien :

$$\forall \lambda_1, \dots, \lambda_d, \quad \lambda_1 X_1 + \dots + \lambda_d X_d \sim \mathcal{N}(0, \sum \lambda_i^2 \sigma^2).$$

Remarque 1.5.7. On peut avoir pour tout i , X_i suit une loi gaussienne mais X n'est pas un vecteur gaussien.

Exemple 1.5.8 (Contre-exemple). Soit $X \sim \mathcal{N}(0, \sigma^2)$, Y indépendante de X t.q.

$$\mathbb{P}(Y = 1) = \mathbb{P}(Y = -1) = \frac{1}{2}.$$

Alors (X, XY) n'est pas un vecteur gaussien alors que chaque coordonnée suit une loi gaussienne, en effet,

$$X + XY = X(1 + Y) = \begin{cases} 2X & \text{avec proba } \frac{1}{2} \\ 0 & \text{avec proba } \frac{1}{2} \end{cases}$$

Si X est un vecteur gaussien, il admet une matrice de covariant $M : \forall i, j, \mathbb{E}(X_i X_j)$ existe et M est la matrice d'une forme quadratique positive. Pour la réciproque, on a la proposition ci-dessous :

Proposition 1.5.9. *Si M est la matrice d'une forme quadratique positive, alors $\forall m \in \mathbb{R}^d$, il existe un vecteur gaussien de moyenne m et de matrice de covariance M .*

Démonstration. Pour le cas particulier, i.e. $M = id$, on peut l'associer un vecteur où $X_i = m_i + \mathcal{N}(0, 1)$ et les X_i sont indépendantes.

Plus généralement, si M est symétrique positive, alors il existe une matrice A telle que $M = A^2$. Soit (X_n) des variables aléatoires i.i.d. qui suit la loi $\mathcal{N}(0, 1)$, $X = (X_1, \dots, X_d)$, alors le vecteur $Z = m + AX$ convient. Pour le voir, il suffit de vérifier que la matrice de covariance de Z est M . La matrice de covariance de Z est la même que celle de $Z' = Z - m$. Soit $\xi \in \mathbb{R}^d$, $\langle \xi, Z' \rangle$ suit une loi gaussienne de moyenne nulle et de covariance :

$$\mathbb{E}(\langle \xi, Z' \rangle^2) = \mathbb{E}({}^t \xi A X {}^t X A \xi) = {}^t \xi A \mathbb{E}(X {}^t X) A \xi = {}^t \xi A A \xi = {}^t \xi M \xi.$$

Donc M est la matrice de covariance de Z' . □

Quand on a un $i \neq j$, $M_{ij} = 0$ alors Z_i et Z_j sont indépendantes. En fait, pour les vecteurs gaussiens : indépendance \iff covariance nulle. Mais ce n'est pas vrai si les vecteurs ne sont pas gaussiens.

Exemple 1.5.10. Soient X, Y indépendantes avec

$$\mathbb{P}(Y = 1) = \mathbb{P}(Y = -1) = \frac{1}{2}.$$

Soit $\mathbb{E}(X^2) < \infty$ et $(X, XY) = (X, Z)$, alors

$$\mathbb{E}(Z) = \mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) = 0,$$

et $\mathbb{E}(XZ) = 0$ car avec proba $\frac{1}{2}$, $ZX = X^2$ et avec proba $\frac{1}{2}$, $ZX = -X^2$, alors

$$\text{cov}(X, Z) = \mathbb{E}(XZ) - \mathbb{E}(X)\mathbb{E}(Z) = 0,$$

mais X, Z n'est pas indépendante :

$$\mathbb{P}(X > a, Z > a) = \mathbb{P}(X > a, Y = 1) = \frac{1}{2}\mathbb{P}(X > a).$$

$$\begin{aligned} \mathbb{P}(X > a)\mathbb{P}(Z > a) &= \mathbb{P}(X > a) \left(\mathbb{P}(X > a, Y = 1) + \mathbb{P}(X < a, Y = -1) \right) \\ &= \mathbb{P}(X > a) \left(\frac{1}{2}\mathbb{P}(X > a) + \frac{1}{2}\mathbb{P}(X < a) \right) \\ &= \frac{1}{2}\mathbb{P}(X > a) \left(\mathbb{P}(X > a) + \mathbb{P}(X < a) \right). \end{aligned}$$

Donc X et Z n'est pas indépendante en général.

On peut décrire la densité d'un vecteur de moyenne m , matrice de covariance M , au point $x = (x_1, \dots, x_d)$,

$$f(x) = \frac{1}{\sqrt{(2\pi)^d (\det M)}} \exp\left(-\frac{1}{2}(x - m)^t M (x - m)\right).$$

En particulier, si $M = Id$ et $m = 0$, la densité devient $\frac{1}{\sqrt{2\pi}^d} \exp(-\sum x_i^2)$.

Si $X \sim \mathcal{N}(0, Id)$, alors pour toute isométrie donnée par une matrice orthogonale A , AX est un vecteur gaussien $\sim \mathcal{N}(0, I)$. Car si $\xi \in \mathbb{R}^d$, $\langle \xi, AX \rangle$ suit une loi gaussienne de variance

$$\mathbb{E}(\langle \xi, AX \rangle^2) = \mathbb{E}(\xi^t X^t A A X \xi) = \mathbb{E}(\xi^t X X \xi) = \xi^t \xi.$$

Donc la matrice de covariance de AX est Id et AX a la même loi que X . On dit que X est invariante par transformée orthogonale (invariance par isométrie). En particulier, si $d = 2$, X, Y indépendante, $X, Y \sim \mathcal{N}(0, 1)$, alors (X, Y) est invariant par rotation.

On a défini la loi uniforme ω_d sur la sphère unité $S^{d-1} = \{u \in \mathbb{R}^d, \|u\| = 1\}$: soit $B \subset S^{d-1}$, soit

$$\Gamma(B) = \{u \in \mathbb{R}^d, 0 < \|u\| \leq 1, \frac{u}{\|u\|} \in B\} = \text{Cone}(B) \cap B(0, 1) \setminus \{0\},$$

où $\text{Cone}(B) = \{u \in \mathbb{R}^d, \exists r > 0, \exists b \in B, u = rb\}$. On définit

$$\omega_d(B) = \frac{\text{Leb}(\Gamma(B))}{\text{Leb}(B(0, 1))},$$

et on sait que $\omega_d(B)$ est invariant par la transformée orthogonale (conséquence de l'invariance par transformée orthogonale de la mesure de Lebesgue).

Soit $X = (X_1, \dots, X_d) \sim \mathcal{N}(0, Id)$, alors p.s. $\|X\|_2 \neq 0$. Donc $Y = \frac{X}{\|X\|}$ est bien défini p.s. et appartient à S^{d-1} . Comme X est invariant par transformée orthogonale, Y est aussi invariante par transformée orthogonale. On peut montrer qu'il existe une seule mesure de probabilité sur S^{d-1} qui est invariante par transformée orthogonale et Y est uniforme sur la sphère S^{d-1} .

Théorème 1.5.11. Soient $Y_d = (Y_d^{(1)}, \dots, Y_d^{(d)})$ comme ci-dessus, alors

$$\sqrt{d}Y_d^{(1)} \xrightarrow{(l)} \mathcal{N}(0, 1).$$

Démonstration. Comme $X_d \sim \mathcal{N}(0, Id)$ dans \mathbb{R}^d et $Y_d = \frac{X_d}{\|X_d\|}$, on a

$$\sqrt{d}Y_d^{(1)} = \frac{X_d^{(1)}}{\sqrt{\frac{X_1^2 + \dots + X_d^2}{d}}}.$$

Par la loi des grands nombres, $\frac{X_1^2 + \dots + X_d^2}{d}$ converge p.s. et dans L^1 , de même pour $\sqrt{\frac{d}{X_1^2 + \dots + X_d^2}}$.

On a

$$\mathbb{P}(X_d^{(1)} \in [a, b]) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{d^2}{2}\right) dx := \mu([a, b]).$$

D'ailleurs,

$$\forall \epsilon > 0, \exists D, \forall d \geq D, \text{ avec proba } > 1 - \epsilon, \frac{1}{\sqrt{\frac{X_1^2 + \dots + X_d^2}{d}}} \in [1 - \epsilon, 1 + \epsilon].$$

On en déduit que

$$\mathbb{P}\left(\frac{X_d^{(1)}}{\sqrt{\frac{X_1^2 + \dots + X_d^2}{d}}} \in [a(1 - \epsilon), b(1 + \epsilon)]\right) \in (\mu(a, b)(1 - \epsilon), \mu(a, b)(1 + \epsilon)), \quad \text{si } d > D.$$

Donc

$$\mathbb{P}\left(\frac{X_d^{(1)}}{\sqrt{\frac{X_1^2 + \dots + X_d^2}{d}}} \in [a, b]\right) \longrightarrow \mu(a, b).$$

□

Chapitre 2

Espérance conditionnelle

Définition 2.0.1. Si A et B sont 2 éléments d'un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$, si $\mathbb{P}(B) > 0$, alors on définit $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$. On l'appelle « Probabilité de A sachant B » ou bien « A conditionnellement à B ».

Plus généralement, si B est un élément tel que $\mathbb{P}(B) > 0$. On désigne $\mathcal{F} \xrightarrow{\mathbb{P}_B} \mathbb{R}$ qui envoie A à $\mathbb{P}(A|B)$. Alors \mathbb{P}_B est une mesure de probabilité par les trois vérifications suivantes.

- $\mathbb{P}_B(\emptyset) = 0 = \frac{\mathbb{P}(\emptyset \cap B)}{\mathbb{P}(B)} = 0$.
- $\mathbb{P}_B(\Omega) = 1$.
- Si $(A_n)_{n \in \mathbb{N}}$ suite des événements deux à deux disjoints : $\forall m \neq n, A_m \cap A_n = \emptyset$, alors $\mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$.

$$\mathbb{P}_B\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \frac{\mathbb{P}\left(\left(\bigcup_{n \in \mathbb{N}} A_n\right) \cap B\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} (A_n \cap B)\right)}{\mathbb{P}(B)} = \sum_{n \in \mathbb{N}} \mathbb{P}_B(A_n).$$

Définition 2.0.2 (Cas discret). Soit X v.a. à valeur dans $A \subset \mathbb{R}^d$ (exemple $A = \mathbb{N}$, $A = \mathbb{Z}$), A dénombrable, et pour $\mathbb{P}(B) > 0$. On peut définir une nouvelle variable aléatoire X_B par : $a \in A, \mathbb{P}(X_B = a) = \mathbb{P}(X = a|B)$.

Terminologie. On adopte les terminologies suivantes :

- loi conditionnelle de X sachant B : Loi de X_B .
- Espérance conditionnelle : Espérance de la loi conditionnelle.

Remarque 2.0.3. Si $X \in L^1$, alors $X_B \in L^1$.

On sait que $\mathbb{E}(X_B) = \mathbb{E}(X|B) = \sum_{a \in A} \mathbb{P}(X_B = a)a = \sum_{a \in A} \mathbb{P}(X = a|B)a = \sum_{a \in A} \frac{\mathbb{P}(\{X=a\} \cap B)}{\mathbb{P}(B)}a$ est bien défini car $\sum_{a \in A} \frac{\mathbb{P}(\{X=a\} \cap B)}{\mathbb{P}(B)} \|a\| \leq \frac{1}{\mathbb{P}(B)} \sum_{a \in A} \mathbb{P}(X = a) \|a\|$.

Remarque 2.0.4. Espérance conditionnelle sachant B est un réel car $\mathbb{E}(X|B) = \mathbb{E}(X_B) = \frac{\mathbb{E}(X \cdot 1_B)}{\mathbb{P}(B)}$.

2.1 Cas discret

Définition 2.1.1. Soient X, Y v.a à valeurs dans $A \subset \mathbb{R}^d$, et A est un ensemble dénombrable (par exemple $\mathbb{N}^n, \mathbb{Z}^n$). Alors on définit

$$\mathbb{E}(X|Y) = \sum_{b \in A'} 1_{\{Y=b\}} \mathbb{E}_{\{Y=b\}}(X) = \sum_{b \in A'} 1_{\{Y=b\}} \sum_{a \in A} \frac{\mathbb{P}(Y=b, X=a)}{\mathbb{P}(Y=b)} a$$

Ici on pose $A' = \{b \in A, \mathbb{P}(Y=b) > 0\}$.

Remarque 2.1.2. De manière équivalente, si on pose $\varphi(y) = \begin{cases} \mathbb{E}_{Y=y}(X) & \text{pour } y \in A' \\ 0 & \text{pour } y \notin A' \end{cases}$. Alors on a $\mathbb{E}(X|Y) = \sum_{b \in A'} 1_{\{Y=b\}} \varphi(b) = \sum_{b \in A} 1_{\{Y=b\}} \varphi(b)$, dont $1_{\{Y=b\}}$ est une variable aléatoire mesurable par rapport à la tribu $\sigma(Y)$ à valeur dans \mathbb{R}^d .

Exemples 2.1.3. On joue à pile ou face 3 fois. On pose $Y = 1_{\{\text{pile au } 1^{\text{er}} \text{ lance}\}}$ et $X =$ nombre de «pile» .

$$\mathbb{E}(X|Y=1) = \mathbb{E}(\text{nombre de «pile»} | \text{pile au } 1^{\text{er}} \text{ lance}) = 1 + \frac{1}{2} + \frac{1}{2} = 2.$$

$$\mathbb{E}(X|Y=0) = \mathbb{E}(\text{nombre de «pile»} | \text{face au } 1^{\text{er}} \text{ lance}) = 0 + \frac{1}{2} + \frac{1}{2} = 1.$$

$$\mathbb{E}(X|Y) = Y + 1.$$

Rappel 2.1.4. *Un variable aléatoire Z est mesurable par rapport $\sigma(Y)$ si et seulement si elle peut s'exprimer $Z = \varphi(Y)$ où φ est une fonction mesurable. Définition du Z mesurable par rapport à $\sigma(Y)$: $\forall C$ borélien $\{z \in C\} \in \sigma(Y)$.*

Rappel 2.1.5. *Espace de Hilbert : espace vectoriel normé complet où la norme issue d'un produit scalaire : $\exists \phi$ bilinéaire symétrique i.e. $(x, y) \mapsto \phi(x, y) = \phi(y, x)$. De plus ϕ est définie positive : $\phi(x, x) \geq 0$, et $\phi(x, x) = 0$ si et seulement si $x = 0$. Norme de x est définie par : $\sqrt{\phi(x, x)}$.*

Exemples 2.1.6. $E := \{\text{Variables aléatoire dans } L^2(\mathbb{R}), \text{ de moyenne nulle}\}$, $\mathbb{E}(XY) = \phi(X, Y)$ pour ϕ symétrique bilinéaire et positive. ϕ n'est pas définie positive sur E , relation d'équivalence $\sim : X \sim Y$ si $X = Y$ p.s.

(E/\sim) : Ensemble quotient : ensemble des classes d'équivalence de E pour la relation \sim .

Sur (E/\sim) on définit ϕ comme plus haut (il faut vérifier $\phi(X, Y) = \phi(X', Y)$ si $X \sim X'$: $\phi(X, Y) - \phi(X', Y) = \phi(X - X', Y) = 0$). Et ϕ ici est définie positive $\phi(X, X) = 0 \Leftrightarrow X = 0$ p.s. $\Leftrightarrow X = 0$ dans (E/\sim) .

Si E' est un sous-espace vectoriel d'un espace de Hilbert E , la projection orthogonale $\pi_{E'} : E \rightarrow E'$ est une application linéaire continue caractérisée par $x = \pi_{E'}(x) + y$ où y est orthogonale à E' : $\forall z \in E', \phi(z, y) = 0$.

Sur l'espace E des v.a $L^2(\mathbb{R})$, si $Y \in E$, dans l'ensemble E'_Y des variable aléatoire mesurable par rapport à $\sigma(Y)$ est un sous-espace vectoriel : $z, z' \in E'_Y, \lambda z + \lambda' z' \in E'_Y$.

Dans \mathbb{R}^d , E espace des variable aléatoire dans $L^2(\mathbb{R}^d)$, on définit $\phi(X, Y) = E(\langle X, Y \rangle)$. On généralise ce qui a été fait plus haut. On va voir que si $X \in L^2(\mathbb{R}^d)$ à valeur dans un ensemble dénombrable, si on note $m = \mathbb{E}(X)$, alors $\mathbb{E}(X|Y) = m + Z$, où Z est la projection orthogonale de $(X - m)$ sur l'espace des variables aléatoires mesurable par rapport à $\sigma(Y)$.

Proposition 2.1.7. $A \subset \mathbb{R}$ dénombrable, $X \in L^2(\mathbb{R})$ à valeurs dans A , $m = \mathbb{E}(X)$ alors m minimize la fonction $f : x \mapsto \mathbb{E}[(X - x)^2]$

Démonstration.

$$f(x) = \sum_{i \in A} \mathbb{P}(X = i)(x - i)^2 = \sum_{i \in A} \mathbb{P}(X = 1)(x^2 - 2xi + i^2) = x^2 - 2x\mathbb{E}(X) + \mathbb{E}(X^2).$$

$f'(x) = 2(x - \mathbb{E}(x))$ s'annule en $\mathbb{E}(X) = m$. Alors on a un minimum en $x = m$. □

Remarque 2.1.8. Dans \mathbb{R}^n , $X \in L^1(\mathbb{R}^d)$, X à valeurs dans $A \subset \mathbb{R}^d$ dénombrable. $m = \mathbb{E}(X)$, alors m minimise la fonction $\mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \mathbb{E}(\|X - x\|^2)$ (minimum en m_i pour chaque i).

Proposition 2.1.9. Soient X, Y v.a. à valeurs dans A dénombrable contenu dans \mathbb{R} , $X, Y \in L^2(\mathbb{R})$. Alors $\mathbb{E}(X|Y)$ minimise la fonction $Z \mapsto \mathbb{E}((X - Z)^2)$ sur toutes les variable aléatoire Z qui sont $\sigma(Y)$ -mesurable.

Démonstration. Soit $a \in A$, conditionnellement à $\{Y = a\}$, $\mathbb{E}((X - x)^2|Y = a)$ est minimisée par le réel $\mathbb{E}_{\{Y=a\}}(X)$. En effet,

$$\mathbb{E}((X - x)^2|Y = a) = \mathbb{E}_{\{Y=a\}}[(X - x)^2] = \mathbb{E}[(X_{\{Y=a\}} - x)^2]$$

est minimisée par la moyenne de la variable aléatoire $X_{\{Y=a\}}$, cette moyenne est $\mathbb{E}(X_{\{Y=a\}}) = \mathbb{E}_{\{Y=a\}}(X) = \sum_{b \in A'} \frac{\mathbb{P}(X=b, Y=a)}{\mathbb{P}(Y=a)} b$.

C'est vrai pour tout $a \in A'$. On cherche une variable aléatoire Z qui est $\sigma(Y)$ mesurable et qui minimise $\mathbb{E}((X - Z)^2)$. Ce qui est équivalent à chercher une fonction φ qui minimise $\mathbb{E}((X - \varphi(Y))^2) = \sum_{a \in A'} 1_{\{Y=a\}} \mathbb{E}((X - \varphi(a))^2) = \sum_{a \in A'} \mathbb{E}(1_{\{Y=a\}}(X - \varphi(a))^2)$.

Alors φ doit minimiser $\mathbb{E}(1_{\{Y=a\}}(X - \varphi(a))^2)$ pour tout $a \in A'$.

Autrement dit, φ doit valoir $\mathbb{E}_{\{Y=a\}}(X)$ sur l'événement $\{Y = a\}$, $\varphi(a) = \mathbb{E}_{\{Y=a\}}(X)$, donc $\varphi(Y) = \mathbb{E}(X|Y)$.

On a minimisé sur les variables aléatoires $\sigma(Y)$ mesurable, $Z \mapsto \mathbb{E}((X - Z)^2)$ en prenant $Z = \varphi(Y) = \mathbb{E}(X|Y)$.

Si $\mathbb{E}(X) = 0$, Z est bien la projection orthogonale de X sur le sous-espace vectoriel des variables aléatoires de moyenne nulle, $\sigma(Y)$ -mesurable, de carré intégrable.

Si $\mathbb{E}(X) = m$, pour le raisonnement ci-dessus, la variable aléatoire $\sigma(Y)$ -mesurable qui minimise $\mathbb{E}(((X - m) - Z)^2)$ est la projection orthogonale de $(X - m)$ sur le sous-espace vectoriel des variables aléatoires $\sigma(Y)$ -mesurable et de carré intégrable.

$$\mathbb{E}(((X - m) - Z)^2) = \mathbb{E}((X - (m + Z))^2)$$

Donc $\mathbb{E}(X|Y) = m + Z$.

□

Proposition 2.1.10 (cadre discret). *Soient X, Y v.a. à valeurs dans A . Soit A est dénombrable et contenu dans \mathbb{R} . De plus on exige $X, Y \in L^1$.*

1. $\mathbb{E}(|\mathbb{E}(X|Y)|) \leq \mathbb{E}(|X|)$.
2. $\forall Z$ v.a. $\sigma(Y)$ -mesurable, $\mathbb{E}(ZX) = \mathbb{E}(Z\mathbb{E}(X|Y))$.

Démonstration. D'après la définition de l'espérance conditionnelle, on calcule

$$\mathbb{E}(|\mathbb{E}(X|Y)|) = \sum_{a \in A'} \mathbb{P}(Y = a) \frac{|\mathbb{E}(X \cdot 1_{\{Y=a\}})|}{\mathbb{P}(Y = a)} \leq \sum_{a \in A} \mathbb{E}(|X| 1_{\{Y=a\}}) = \mathbb{E}(|X|).$$

Si Z est $\sigma(Y)$ -mesurable, on utilise le fait qu'on peut écrire $Z = \varphi(Y)$, avec φ une fonction mesurable et bornée.

$$\begin{aligned} \mathbb{E}(Z\mathbb{E}(X|Y)) &= \sum_{a \in A'} \varphi(a) \mathbb{E}(X 1_{\{Y=a\}}) = \sum_{a \in A'} \mathbb{E}(\varphi(Y) X 1_{Y=a}) \\ &= \mathbb{E}\left(\sum_{a \in A'} \varphi(Y) X 1_{Y=a}\right) = \mathbb{E}(\varphi(Y) X) = \mathbb{E}(ZX). \end{aligned}$$

□

Remarque 2.1.11. Si on le ramène à des v.a. de moyenne nulle (on peut toujours le faire en remplaçant X par $X - \mathbb{E}(X)$ et si la v.a. sont dans L^2). (2) se réécrit : $\phi(Z, X) = \phi(Z, \mathbb{E}(X|Y))$, $y \in E$, $\phi(y, z) = \phi(y, \pi_E(x))$.

Remarque 2.1.12 (Bilan). Dans le cas discret, si X variable aléatoire dans L^1 , et Y variable aléatoire, alors $\mathbb{E}(X|Y)$ est une variable aléatoire mesurable par rapport à $\sigma(Y)$.

Si de plus $X \in L^2$, on sait que $\mathbb{E}(X|Y)$ est la projection orthogonale de X sur l'espace vectoriel de variables aléatoires dans L^2 qui sont mesurables par rapport à $\sigma(Y)$.

Exemple 2.1.13. Soit n entier plus grand que 2, et U, V des variables aléatoires uniformes sur $\{1, \dots, n\}$, $S = U + V$, $P = UV$, on va déterminer $\mathbb{E}(P|S)$.

$$\{S = k\} = \bigcup_{i=1}^n \{U = i, V = k - i\}$$

Alors

$$\mathbb{P}(S = k) = \sum_{i=1}^n \mathbb{P}(U = i, V = k - i) = \sum_{i=1}^n \mathbb{P}(U = i) \mathbb{P}(V = k - i)$$

Par définition de l'espérance conditionnelle, on a

$$\begin{aligned} \mathbb{E}(P|S = k) &= \mathbb{E}\left(\sum_{i=1}^n 1_{U=i, V=k-i} i(k-i)\right) \\ &= \sum_{i=1}^n \mathbb{E}(1_{U=1, V=k-i} i(k-i)) \\ &= \sum_{i=1}^n i(k-i) \mathbb{P}(U = i, V = k - i) \\ &= \sum_{i=1}^n \underbrace{\frac{i(k-i)}{n^2} 1_{k-i \in [1, n]}}_{=: \varphi(k)} \end{aligned}$$

Alors $\mathbb{E}(P|S) = \varphi(S)$ est une fonction de S , donc mesurable par rapport à la tribu $\sigma(S)$.

2.2 Cas général

On a montré dans le cas discret que pour toute v.a. Z qui est mesurable par rapport à $\sigma(Y)$, on a

$$\mathbb{E}(ZX) = \phi(Z, X) \quad (2.2.1)$$

où ϕ est la forme bilinéaire associée au produit scalaire qui définit l'espace de Hilbert L^2 . Équivalentement, 2.2.1 se réécrit en $\phi(Z, X) = \phi(Z, \pi_E(X))$, où π_E est la projection orthogonale sur l'espace vectoriel des variables aléatoires mesurables par rapport à $\sigma(Y)$.

Théorème 2.2.1. *Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et \mathcal{B} une sous-tribu de \mathcal{A} . Soit $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, alors il existe une unique variable aléatoire dans $L^1(\Omega, \mathcal{B}, \mathbb{P})$, qu'on note $\mathbb{E}(X|\mathcal{B})$, telle que*

$$\forall B \in \mathcal{B}, \mathbb{E}(X1_B) = \mathbb{E}(\mathbb{E}(X|\mathcal{B})1_B) \quad (2.2.2)$$

Plus généralement, pour toute Z variable aléatoire bornée mesurable par rapport à \mathcal{B} , on a

$$\mathbb{E}(XZ) = \mathbb{E}(\mathbb{E}(X|\mathcal{B})Z). \quad (2.2.3)$$

Démonstration. Unicité On le démontre par l'absurde, si Y et Y' vérifient 2.2.2. On pose $B = \{Y > Y'\}$. par l'hypothèse de Y , on a

$$\mathbb{E}(Y1_B - Y'1_B) = \mathbb{E}((Y - Y')1_B) = 0.$$

Comme $Y - Y' > 0$ sur B , alors $\mathbb{P}(B) = 0$. Donc $\mathbb{P}(Y \leq Y') = 1$, de même $\mathbb{P}(Y \geq Y') = 1$. Ce qui est équivalent à $\mathbb{P}(Y = Y') = 1$.

Existence, pour démontrer l'existence, il faut utiliser le théorème de Radon-Nikodym.

Rappel 2.2.2. *Version faible de théorème de Radon-Nikodym : Si ν et μ sont des mesures positives sur (Ω, \mathcal{F}) .*

De plus, on demande $\nu \ll \mu$ est absolument continue [si pour tout $B \in \mathcal{F}$ et $\mu(B) = 0$, on a $\nu(B) = 0$.]

Alors il existe $f : \Omega \rightarrow \mathbb{R}_+$ mesurable par rapport à \mathcal{F} telle que $\forall A \in \mathcal{F}, \nu(A) = \int_A f d\mu$, où f est la dérivée de Radon-Nikodym. (P53 de Le Gall)

Pour $B \in \mathcal{B}$, on pose $Q(B) = \mathbb{E}(X1_B)$. Il faut vérifier que Q est une mesure sur (Ω, \mathcal{B}) .

— Q est σ -additive si $(B_i)_{i \in \mathbb{N}}$ sont deux à deux disjoints :

$$\sum Q(B_i) = \mathbb{E}(X \sum 1_{B_i}) = \mathbb{E}(X 1_{\cup B_i}) = Q(\cup B_i)$$

— $Q \ll \mathbb{P}$: si $\mathbb{P}(B) = 0$, $Q(B) = \mathbb{E}(X 1_B) = 0$.

Appliquons le théorème de Radon-Nikodym à Q, \mathbb{P} , il existe $f : \Omega \rightarrow \mathbb{R}_+$ mesurable par rapport à \mathcal{B} , tel que $\forall B \in \mathcal{B}$, $Q(B) = \int_B f d\mathbb{P}$.

Donc $f : (\Omega, \mathcal{B}) \rightarrow \mathbb{R}$ est une fonction mesurable par rapport à \mathcal{B} . On appelle cette variable aléatoire \hat{X} , donc $\mathbb{E}(\hat{X}) = \int_{\Omega} f d\mathbb{P}$ et $\forall B \in \mathcal{B}$, on a

$$\mathbb{E}(\hat{X} 1_B) = \int_B f d\mathbb{P} = Q(B) = \mathbb{E}(X 1_B). \quad (2.2.4)$$

Pour passer de 2.2.2 à 2.2.3, on approche les variables aléatoires mesurables par rapport à \mathcal{B} par des fonctions étagées. (fonctions de la forme $\sum_{i=1}^n c_i 1_{B_i}$, où $B_i \in \mathcal{B}$). \square

Exemple 2.2.3. Soient $\Omega =]0, 1[$ et \mathcal{A} est tribu borélienne. Fixons $n \geq 2$, et \mathcal{B} tribu engendrée par les intervalles $[\frac{i}{n}, \frac{i+1}{n}]$, $0 \leq i \leq n-1$. (\mathcal{B} est l'ensemble des réunions d'intervalles dont les bornes sont de la forme $\frac{k}{n}$). On adopte \mathbb{P} la mesure uniforme et f une variable aléatoire t.q. $f \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ i.e. $\int_0^1 f(x) dx < \infty$.

$$\mathbb{E}(f|\mathcal{B}) = \sum_{i=0}^{n-1} c_i 1_{[\frac{i}{n}, \frac{i+1}{n}]}, \quad \text{ici } c_i = \int_{i/n}^{(i+1)/n} f(x) dx.$$

Il est évident que $\mathbb{E}(f|\mathcal{B})$ est mesurable par rapport à \mathcal{B} . Les variables aléatoires mesurables par rapport de \mathcal{B} sont les fonctions constantes sur tout intervalle $[\frac{k}{n}, \frac{k+1}{n}]$. Donc ce sont les fonctions de la forme $\sum_{i=1}^{n-1} d_i 1_{[\frac{i}{n}, \frac{i+1}{n}]}$.

Proposition 2.2.4. On a des propriétés de l'espérance conditionnelle suivants :

1. Si X est \mathcal{B} -mesurable, alors $\mathbb{E}(X|B) = X$.
2. L'application $X \mapsto \mathbb{E}(X|\mathcal{B})$ est linéaire.
3. $\mathbb{E}(\mathbb{E}(X|B)) = \mathbb{E}(X)$
4. $\mathbb{E}(X|\mathcal{B}) \leq \mathbb{E}(|X||\mathcal{B}) \implies \mathbb{E}(\mathbb{E}(X|\mathcal{B})) \leq \mathbb{E}(|X|)$
5. Si $X \leq X'$, alors on a $\mathbb{E}(X|\mathcal{B}) \leq \mathbb{E}(X'|\mathcal{B})$

Démonstration. 1. Puisque X vérifie 2.2.2 et comme on a montré l'unicité, on a directement $X = \mathbb{E}(X|\mathcal{B})$.

2. Si X, X' sont des variables aléatoires. Soit $Z = \lambda X + \lambda' X'$ et on pose $Y = \lambda \mathbb{E}(X|\mathcal{B}) + \lambda' \mathbb{E}(X'|\mathcal{B})$. Alors on a

$$\forall B \in \mathcal{B}, \mathbb{E}(Y1_B) = \lambda \mathbb{E}(X1_B) + \lambda' \mathbb{E}(X'1_B) = \mathbb{E}((\lambda X + \lambda' X')1_B) = \mathbb{E}(Z1_B)$$

Donc $Y = \mathbb{E}(Z|\mathcal{B})$.

3. Si on prends $B = \Omega$ et $1_B = 1$, alors le point 1 implique le point 3.
 4. Si $X \geq 0$, alors $\mathbb{E}(X|\mathcal{B}) \geq 0$. Dans le cas général on écrit $X = X1_{X \geq 0} - (-X)1_{X < 0}$.
 Si on note $X_+ = X1_{X \geq 0}$ et $X_- = (-X)1_{X < 0}$, on a

$$\begin{aligned} \mathbb{E}(X|B) &= \mathbb{E}(X_+ - X_-|\mathcal{B}) = \mathbb{E}(X_+|\mathcal{B}) - \mathbb{E}(X_-|\mathcal{B}) \leq \mathbb{E}(X_+|\mathcal{B}) \\ &\leq \mathbb{E}(X_+|\mathcal{B}) + \mathbb{E}(X_-|\mathcal{B}) = \mathbb{E}(|X||\mathcal{B}) \end{aligned}$$

5. Si $X \leq X'$, $\mathbb{E}(X' - X|\mathcal{B}) \geq 0$, ce qui est équivalent à $\mathbb{E}(X'|\mathcal{B}) \geq \mathbb{E}(X|\mathcal{B})$.

□

2.3 Variables aléatoires positives

Notation : $a \wedge b = \inf(a, b)$

Théorème 2.3.1. *Soit X variable aléatoire positive, alors $\mathbb{E}(X|\mathcal{B}) = \lim_{n \rightarrow \infty} \mathbb{E}(X \wedge n|\mathcal{B})$ est une variable aléatoire positive, caractérisée par : $\forall Z$ \mathcal{B} -mesurable positive, on a*

$$\mathbb{E}(XZ) = \mathbb{E}(\mathbb{E}(X|\mathcal{B})Z). \quad (2.3.1)$$

Remarque 2.3.2. Les variables aléatoires positives peuvent être égales à $+\infty$.

Démonstration. D'après 5 précédent, et $\mathbb{E}(X \wedge n|\mathcal{B})$ est croissante. Donc son limite dans 2.3.1 est bien définie dans $\mathbb{R}_+ \cup \{+\infty\}$. Si Z est positive et \mathcal{B} -mesurable, on a

$$\mathbb{E}(\mathbb{E}(X|\mathcal{B})Z) = \mathbb{E}(\lim_{n \rightarrow \infty} \mathbb{E}(X \wedge n|\mathcal{B})Z)$$

Comme $\mathbb{E}(X \wedge n|\mathcal{B})$ est une suite monotone (croissante) de variable aléatoire positives, par

le théorème de convergence monotone :

$$\begin{aligned}\mathbb{E}(\lim_{n \rightarrow \infty} \mathbb{E}(X \wedge n | \mathcal{B})) &= \lim_{n \rightarrow \infty} \mathbb{E}(\mathbb{E}(X \wedge n | \mathcal{B})Z) \\ &= \lim_{n \rightarrow \infty} \mathbb{E}((X \wedge n)Z) = \mathbb{E}(XZ)\end{aligned}$$

L'unicité de la variable aléatoire \mathcal{B} -mesurable vérifiant (3) : si X, X' vérifient (3) et sont \mathcal{B} -mesurables. Alors pour tout $q < q' \in \mathbb{Q}$, si $Z = 1_{X' \leq q < q' \leq X}$, on a

$$q\mathbb{P}(X' \leq q < q' \leq X) \geq q'\mathbb{P}(X' \leq q < q' \leq X) \implies \mathbb{P}(X' \leq q < q' \leq X) = 0$$

Alors

$$\mathbb{P}(X' < X) = \mathbb{P}\left(\bigcup_{q < q' \in \mathbb{Q}} \{X' \leq q < q' \leq X\}\right) = 0$$

Par symétrie, on a $\mathbb{P}(X < X') = \mathbb{P}(X' < X) = 0$. C'est-à-dire $\mathbb{P}(X = X') = 1$. \square

Si on reprend l'exemple $\Omega =]0, 1[$ et \mathcal{A} tribu borélienne. Posons \mathcal{B} sous-tribu engendrée par les intervalles $[\frac{i}{n}, \frac{i+1}{n}]$. Si on prend la fonction positive $f(x) = \frac{1}{x}$, f a ses valeurs dans \mathbb{R}_+ :

$$\mathbb{E}(f | \mathcal{B}) = \sum_{i=0}^{n-1} c_i 1_{[\frac{i}{n}, \frac{i+1}{n}]}, \quad \text{avec } c_i = \int_{i/n}^{(i+1)/n} f(x) dx$$

on a $x_0 = +\infty$ et $c_1, \dots, c_{n-1} \in \mathbb{R}$.

Proposition 2.3.3. *On a des propriétés de l'espérance conditionnelle suivants :*

1. Si X, X' variables aléatoires positives, et $a, b \geq 0$. On a

$$\mathbb{E}(aX + bX' | \mathcal{B}) = a\mathbb{E}(X | \mathcal{B}) + b\mathbb{E}(X' | \mathcal{B})$$

2. Si X est \mathcal{B} -mesurable, alors $\mathbb{E}(X | \mathcal{B}) = X$.
3. Soit X_n une suite croissante de variables aléatoires positives, et $X = \lim_{n \rightarrow \infty} X_n$, alors

$$\mathbb{E}(X | \mathcal{B}) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n | \mathcal{B}).$$

(Vérifier du théorème de convergence monotone pour l'espérance conditionnelle)

4. Si X_n suite de variables aléatoires positives, on a

$$\mathbb{E}(\liminf (X_n) | \mathcal{B}) \leq \liminf \mathbb{E}(X_n | \mathcal{B})$$

5. Soit (X_n) une suite de variables aléatoires dans L^1 qui converge p.s. vers X . S'il existe une variable aléatoire positive Z t.q. $\forall n, |X_n| \leq Z$ p.s. et $\mathbb{E}(Z) < \infty$, alors $\mathbb{E}(X|\mathcal{B}) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n|\mathcal{B})$ p.s. et dans L^1 .
6. Si la fonction f est positive et convexe, si $X \in L^1$, alors

$$\mathbb{E}(f(X)|\mathcal{B}) \geq f(\mathbb{E}(X|\mathcal{B})).$$

(Version de l'inégalité de Janson pour l'espérance conditionnelle)

Démonstration. (3) Si $0 \leq X_1 \leq X_2$, alors $\mathbb{E}(X_1|\mathcal{B}) \leq \mathbb{E}(X_2|\mathcal{B})$. Donc on peut poser $X' = \lim_{n \rightarrow \infty} \mathbb{E}(X_n|\mathcal{B})$ qui est \mathcal{B} -mesurable.

Si Z positive, \mathcal{B} -mesurable, on a

$$\mathbb{E}(ZX') = \mathbb{E}(Z \lim_{n \rightarrow \infty} \mathbb{E}(X_n|\mathcal{B})) = \lim_{n \rightarrow \infty} \mathbb{E}(Z\mathbb{E}(X_n|\mathcal{B})) = \lim_{n \rightarrow \infty} \mathbb{E}(ZX_n) = \mathbb{E}(ZX)$$

- (4) On utilise le fait que pour toute suite de réels (v_n) , $\liminf(v_n) = \lim_{k \rightarrow \infty} (\inf_{n \geq k} v_n)$. Alors

$$\begin{aligned} \mathbb{E}(\liminf X_n|\mathcal{B}) &= \mathbb{E}(\lim_{k \rightarrow \infty} (\inf_{n \geq k} v_n)|\mathcal{B}) = \lim_{k \rightarrow \infty} \mathbb{E}((\inf_{n \geq k} X_n)|\mathcal{B}) \\ &\leq \lim_{k \rightarrow \infty} \inf_{n \geq k} (\mathbb{E}(X_n|\mathcal{B})) = \liminf_{n \rightarrow \infty} \mathbb{E}(X_n|\mathcal{B}) \end{aligned}$$

- (5) Appliquons 4, on a

$$\mathbb{E}(\liminf(Z - X_n)|\mathcal{B}) \leq \mathbb{E}(Z|\mathcal{B}) - \limsup \mathbb{E}(X_n|\mathcal{B})$$

$$\mathbb{E}(\liminf(Z + X_n)|\mathcal{B}) \leq \mathbb{E}(Z|\mathcal{B}) + \liminf \mathbb{E}(X_n|\mathcal{B})$$

Donc

$$\mathbb{E}(X|\mathcal{B}) \leq \liminf \mathbb{E}(X_n|\mathcal{B}) \leq \limsup \mathbb{E}(X_n|\mathcal{B}) \leq \mathbb{E}(X|\mathcal{B})$$

On a bien montré la convergence p.s.

Montrons la convergence dans L^1 . Notons que

$$\mathbb{E}(X_n|\mathcal{B}) \leq \mathbb{E}(|X_n||\mathcal{B}) \leq \mathbb{E}(Z|\mathcal{B}) < \infty,$$

Donc $\mathbb{E}(\mathbb{E}(Z|\mathcal{B})) = \mathbb{E}(Z) < \infty$, ensuite on applique le théorème de convergence dominée.

- (6) Soit $E_f = \{(a, b) \in \mathbb{R}^2, \forall x \in \mathbb{R}, f(x) \geq ax + b\}$ (On considère l'ensemble des droites qui sont en dessous de la courbe de f).

On vérifie que pour tout x ,

$$f(x) = \sup_{(a,b) \in E_f} (ax + b) = \sup_{(a,b) \in E_f \cap \mathbb{Q}^2} (ax + b),$$

Alors, on a immédiatement

$$\mathbb{E}(f(X)|\mathcal{B}) = \mathbb{E}\left(\sup_{(a,b) \in E_f \cap \mathbb{Q}^2} aX + b|\mathcal{B}\right) \geq \sup_{(a,b) \in E_f \cap \mathbb{Q}^2} \mathbb{E}(aX + b|\mathcal{B}) = f(\mathbb{E}(X|\mathcal{B}))$$

□

Théorème 2.3.4. Si $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ et \mathcal{B} sous-tribu de \mathcal{F} , alors $\mathbb{E}(X|\mathcal{B})$ est la projection orthogonale de X sur $L^2(\Omega, \mathcal{B}, \mathbb{P})$ (sous espace vectoriel des variables aléatoires de carré intégrable mesurable sur \mathcal{B}).

Démonstration. Par l'inégalité de Jensen, on a

$$\mathbb{E}[X|\mathcal{B}]^2 \leq \mathbb{E}[X^2|\mathcal{B}] \tag{2.3.2}$$

Ceci implique

$$\mathbb{E}[\mathbb{E}[X|\mathcal{B}]^2] \leq \mathbb{E}[\mathbb{E}[X^2|\mathcal{B}]] = \mathbb{E}[X^2] \tag{2.3.3}$$

Donc l'espérance conditionnelle $\mathbb{E}[X|\mathcal{B}]$ appartient à L^2 .

Si Z une v.a. \mathcal{B} -mesurable bornée,

$$\mathbb{E}[Z(X - \mathbb{E}[X|\mathcal{B}])] = \mathbb{E}[ZX] - \mathbb{E}[Z\mathbb{E}[X|\mathcal{B}]] = 0. \tag{2.3.4}$$

Donc Z est orthogonal à $X - \mathbb{E}[X|\mathcal{B}]$ pour tout Z qui est \mathcal{B} -mesurable bornée par densité. C'est vrai pour tout Z \mathcal{B} -mesurable et de carré intégrable.

C'est-à-dire que $X - \mathbb{E}[X|\mathcal{B}]$ est orthogonal à $L^2(\Omega, \mathcal{B}, \mathbb{P})$.

□

Proposition 2.3.5. Soient X, Y des v.a. réelles, \mathcal{B} est un sous-tribu. Si Y est \mathcal{B} -mesurable,

alors

$$\mathbb{E}[YX|\mathcal{B}] = Y\mathbb{E}[X|\mathcal{B}]. \quad (2.3.5)$$

En supposant que Y et X positives ou que Y et X appartient à L^1 .

Remarque 2.3.6. Si $\lambda \in \mathbb{R}$, alors $\mathbb{E}[\lambda X] = \lambda\mathbb{E}[X]$.

Démonstration. On se place dans le cas où X, Y positives. Soit Z une v.a. \mathcal{B} -mesurable positive, alors on a

$$\mathbb{E}[Z(Y\mathbb{E}[X|\mathcal{B}])] = \mathbb{E}[(ZY)\mathbb{E}[X|\mathcal{B}]] = \mathbb{E}[Z Y X] \quad (2.3.6)$$

par définition de l'espérance conditionnelle, on en déduit que

$$Y\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[XY|\mathcal{B}] \quad (2.3.7)$$

□

Proposition 2.3.7. Si $\mathcal{B}_1, \mathcal{B}_2$ sont deux sous-tribus tels que $\mathcal{B}_1 \subset \mathcal{B}_2$, alors pour toute v.a. X positive (ou L^1), on a

$$\mathbb{E}[\mathbb{E}[X|\mathcal{B}_2]|\mathcal{B}_1] = \mathbb{E}[X|\mathcal{B}_1]. \quad (2.3.8)$$

Remarque 2.3.8. Si on est dans L^2 , $L^2(\Omega, \mathcal{B}_1, \mathbb{P}) \subset L^2(\Omega, \mathcal{B}_2, \mathbb{P})$. Alors la proposition nous dit que si on projette sur $L^2(\Omega, \mathcal{B}_2, \mathbb{P})$ puis sur $L^2(\Omega, \mathcal{B}_1, \mathbb{P})$, c'est la même chose que de projeter sur $L^2(\Omega, \mathcal{B}_1, \mathbb{P})$

Démonstration. Cas positif : Si Z est positive et \mathcal{B}_1 -mesurable, alors Z est aussi \mathcal{B}_2 -mesurable.

$$\mathbb{E}[Z\mathbb{E}[\mathbb{E}[X|\mathcal{B}_2]|\mathcal{B}_1]] = \mathbb{E}[Z\mathbb{E}[X|\mathcal{B}_2]] = \mathbb{E}[ZX]. \quad (2.3.9)$$

□

Théorème 2.3.9. Si $\mathcal{B}_1, \mathcal{B}_2$ sont indépendantes, alors pour toute v.a. X positive \mathcal{B}_2 -mesurable, on a

$$\mathbb{E}[X|\mathcal{B}_1] = \mathbb{E}[X]. \quad (2.3.10)$$

Réciproquement, si $\mathcal{B}_1, \mathcal{B}_2$ sont des sous-tribus et si pour toute X positive et \mathcal{B}_2 -mesurable, on a

$$\mathbb{E}[X|\mathcal{B}_1] = \mathbb{E}[X]. \quad (2.3.11)$$

Alors $\mathcal{B}_1, \mathcal{B}_2$ sont indépendants.

Démonstration. D'abord pour toute Z positive et \mathcal{B}_1 -mesurable, on a

$$\mathbb{E}[ZX] = \mathbb{E}[Z]\mathbb{E}[X] = \mathbb{E}[Z\mathbb{E}[X|\mathcal{B}_1]]. \quad (2.3.12)$$

Donc $\mathbb{E}[X] = \mathbb{E}[X|\mathcal{B}_1]$.

Ensuite, en particulier pour tout $A \in \mathcal{B}_2$, $\mathbb{1}_A$ est positive et \mathcal{B}_2 -mesurable.

$$\mathbb{E}[\mathbb{1}_A|\mathcal{B}_1] = \mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A). \quad (2.3.13)$$

Si $B \in \mathcal{B}_1$, $\mathbb{P}(A \cap B) = \mathbb{E}[\mathbb{1}_{A \cap B}] = \mathbb{E}[\mathbb{1}_A \mathbb{1}_B] = \mathbb{E}[\mathbb{E}[\mathbb{1}_A|\mathcal{B}_1] \mathbb{1}_B] = \mathbb{E}[\mathbb{P}(A) \mathbb{1}_B] = \mathbb{P}(A) \mathbb{E}[\mathbb{1}_B] = \mathbb{P}(A) \mathbb{P}(B)$.

Ceci implique $\mathcal{B}_1, \mathcal{B}_2$ sont indépendants. □

Théorème 2.3.10. Soient X v.a. à valeur dans E et Y v.a. à valeur dans F . On suppose que Y est \mathcal{B} -mesurable et X est indépendant de \mathcal{B} .

Alors pour toute fonction mesurable $g : E \times F \rightarrow \mathbb{R}_+$, on a

$$\mathbb{E}[g(X, Y)|\mathcal{B}] = \int g(z, Y) \mathbb{P}_X(dz) \quad (2.3.14)$$

où \mathbb{P}_X est la loi de X .

Remarque 2.3.11. On dit qu'une variable aléatoire X est déterministe s'il existe $c \in \mathbb{R}$, $X = c$ presque sûrement.

Si Y est déterministe alors la variable aléatoire donnée par 2.3.14 est aussi déterministe. On se ramène au résultat précédent : on peut écrire $\mathbb{E}[g(X, Y)|\mathcal{B}] = \mathbb{E}[h(X)]$ pour une certaine h .

Démonstration. On note $\phi : y \rightarrow \int g(x, y) \mathbb{P}_X(dx)$. Soient $Z \geq 0$ \mathcal{B} -mesurable et Y \mathcal{B} -mesurable. On exige que X est indépendante de \mathcal{B} .

$$\mathbb{E}[g(X, Y)Z] = \int g(x, y)z\mathbb{P}(dx, dy, dz) \quad (2.3.15)$$

$$= \int g(x, y)z\mathbb{P}_X(dx)\mathbb{P}_{(Y,Z)}(dy, dz) \quad (2.3.16)$$

$$= \int_{F \times \mathbb{R}_+} z \left(\int_E g(x, y)\mathbb{P}_X(dx) \right) \mathbb{P}_{(Y,Z)}(dy, dz) \quad (2.3.17)$$

$$= \int_{F \times \mathbb{R}_+} z\phi(y)\mathbb{P}_{(Y,Z)}(dy, dz) \quad (2.3.18)$$

$$(2.3.19)$$

Et $\phi(y)$ est \mathcal{B} -mesurable donc $\phi(Y)$ est l'espérance conditionnelle de $g(X, Y)$ par rapport à \mathcal{B} . \square

2.4 Comment calculer les espérances conditionnelles ?

On a vu le cas discret et le cas des v.a. a densité.

On suppose que X à valeur dans \mathbb{R}^m et Y à valeur dans \mathbb{R}^n . Soit (X, Y) a une densité $p : \mathbb{R}^{m+n} \rightarrow \mathbb{R}$ mesurable, positive et satisfait $\int_{\mathbb{R}^{m+n}} p(x, y)dx dy = 1$.

Alors X et Y ont une densité. La densité de Y et la fonction $q : \mathbb{R}^n \rightarrow \mathbb{R}_+, y \mapsto \int_{\mathbb{R}^m} p(x, y)dx$.

Proposition 2.4.1. *Pour toute fonction $h : \mathbb{R}^m \rightarrow \mathbb{R}_+$ mesurable, on a*

$$\mathbb{E}[h(X)|Y] = \int h(x) \frac{p(x, Y)}{q(Y)} dx \quad (2.4.1)$$

qui est une variable aléatoire mesurable par rapport à $\sigma(Y)$ donc une fonction de Y .

Démonstration. On prend Z v.a. positive et $\sigma(Y)$ -mesurable, alors Z s'écrit comme une fonction de Y par $Z = g(Y)$ pour $g : \mathbb{R}^n \rightarrow \mathbb{R}_+$.

$$\mathbb{E}[Zh(X)] = \mathbb{E}[h(X)g(Y)] = \int_{\mathbb{R}^m \times \mathbb{R}^n} h(x)g(y)p(x, y)dx dy \quad (2.4.2)$$

$$= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} h(x)p(x, y)dx \right) g(y)dy \quad (2.4.3)$$

$$= \int_{\mathbb{R}^n} \left(\frac{\int_{\mathbb{R}^m} h(x)p(x, y)dx}{q(y)} \right) q(y)g(y)1_{q(y)>0}dy \quad (2.4.4)$$

$$= \int_{\mathbb{R}^n} \varphi(y)q(y)g(y)1_{q(y)>0}dy \quad (2.4.5)$$

$$= \mathbb{E}[g(Y)\varphi(Y)] = \mathbb{E}[Z\varphi(Y)] \quad (2.4.6)$$

$$\text{où } \varphi(y) = \begin{cases} \frac{\int_{\mathbb{R}^m} h(x)p(x, y)dx}{q(y)} & \text{si } g(y) > 0 \\ 0 & \text{sinon} \end{cases}.$$

Donc $\varphi(Y) = \mathbb{E}[h(x)|Y]$.

□

Rappel 2.4.2. Dans le cas discret : (X, Y) v.a. discrètes. Pour tout y , on peut définir la loi conditionnelle de X par rapport à $Y = y$ et l'espérance conditionnelle et l'espérance de la loi conditionnelle : $\mathbb{E}[X|Y = y]$.

On va essayer de généraliser la notion de la los conditionnelle dans la cas non discret.

Définition 2.4.3. Soient $(E, \mathcal{E}), (F, \mathcal{F})$ deux espace mesurable. On dit que $\nu : E \times \mathcal{F} \rightarrow [0, 1]$ est une probabilité de transition si elle vérifie :

- Pour tout $x \in E$, la fonction $\mathcal{F} \rightarrow [0, 1], B \mapsto \nu(x, B)$ est mesure de probabilité (probabilité conditionnellement à x).
- Pour tout $B \in \mathcal{F}$, la fonction $E \rightarrow [0, 1], B \mapsto \nu(x, B)$ est \mathcal{E} -mesurable.

Remarque 2.4.4. Cas précédent $p(x, y)$, $p : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}_+$ pour $x \in \mathbb{R}^m$ et $\nu \in \mathcal{B}(\mathbb{R}^n)$, la fonction $\nu(x, B) = \int_B p(x, y)dy$ vérifie ces conditions.

Proposition 2.4.5. Si h mesurable positive sur (F, \mathcal{F}) , alors $\varphi : x \mapsto \int \nu(x, dy)h(y)$ est une focntion mesurable positive sur E .

Si λ est une mesure de probabilité sur (E, \mathcal{E}) , alors $\mu : A \in \mathcal{F} \mapsto \int \lambda(dx)\nu(x, A)$ est une mesure de probabilité sur (F, \mathcal{F}) .

Définition 2.4.6. Soient X v.a. sur (E, \mathcal{E}) et Y v.a. sur (F, \mathcal{F}) . On appelle loi conditionnelle de Y sachant X une probabilité de transitive ν telle que pour tout $h : F \rightarrow \mathbb{R}_+$ \mathcal{F} -mesurable,

on a

$$E[h(Y)|X] = \int \nu(X, dy)h(y). \quad (2.4.7)$$

Remarque 2.4.7. L'espérance conditionnelle est en effet un espérance de la loi conditionnelle.

Unicité : Si ν et ν' sont des lois conditionnelles. Pour tout $A \in \mathcal{F}$, on a

$$\nu(X, A) = \mathbb{P}(Y \in A|X) = \nu'(X, A) \quad p.s. \quad (2.4.8)$$

De même, $\nu(x, \cdot) = \nu'(x, \cdot)$ p.s.

Théorème 2.4.8. Soient $(E, \mathcal{E}), (F, \mathcal{F})$ deux espace métrique complet séparable, \mathcal{E}, \mathcal{F} des tribus boréliennes. Alors il existe une loi conditionnelle de Y sachant X .

2.5 Espérance conditionnelle pour les vecteurs gaussiens

Soit $(X_1, \dots, X_d) \in \mathbb{R}^d$ vecteur aléatoire est gaussien si toute combinaison linéaire des X est gaussienne : $\forall \lambda_1, \dots, \lambda_d, \sum_{i=1}^d \lambda_i X_i$ soit une loi gaussienne.

Si (X_1, \dots, X_d) est gaussien, $(X_i) \in L^2$. Pour tout Y aléatoire dans \mathbb{R}^d et L^2 , $\mathbb{E}[Y|(X_1, \dots, X_d)]$ est la projection orthogonale de Y sur le sous-espace vectoriel des variable aléatoire L^2 et mesurable par rapport à $\sigma(X_1, \dots, X_d)$.

Le sous-espace vectoriel est l'ensemble des variable aléatoire de la forme $h(X_1, \dots, X_d)$ où h est mesurable.

En fait, on va voir que cette projection orthogonale est la projetcion orthogonale sur l'espace vectoriel engendré par (X_1, \dots, X_d) , alors c'est un espace de dimension finie.

Proposition 2.5.1. Soit $(X_1, \dots, X_m, Y_1, \dots, Y_n)$ un vecteur gaussien dans \mathbb{R}^{m+n} .

Alors (X_1, \dots, X_m) et (Y_1, \dots, Y_n) sont des vecteurs gaussiens et ils sont indépendants si et seulement si $\forall i \in [1, m], \forall j \in [1, n]$ on a $\text{cov}(X_i, Y_j) = 0$.

Remarque 2.5.2. C'est toujours vrai que l'indépendance implique le covariance nulle. Et on sait que si deux vecteurs gaussiens sont de covariance nulle, alors ils sont indépendants. Mais on ne peut pas obtenir l'indépendance de la covariance nulle.

Démonstration. Pour Q la matrice de la forme quadratique associée au vecteur gaussien, on a

$$\mathbb{E}[\exp(i \langle \xi, X_1, \dots, X_m, Y_1, \dots, Y_n \rangle)] = \exp\left(\frac{1}{2} {}^t \xi Q \xi\right).$$

Par $Q_{i,j} = \text{cov}(Z_i, Z_j)$, si Z est un vecteur gaussien de covariance nulle, on réécrit

$${}^t \xi Q \xi = \sum_{j,k=1}^m \xi_j \xi_k \text{cov}(X_j, X_k) + \sum_{j,k=1}^n \xi_j \xi_k \text{cov}(Y_j, Y_k).$$

Alors

$$\begin{aligned} \mathbb{E}[\exp(i \langle \xi, X_1, \dots, X_m, Y_1, \dots, Y_n \rangle)] &= \mathbb{E}[\exp(i \sum_{j=1}^m \xi_j X_j)] \mathbb{E}[\exp(i \sum_{j=1}^n \xi_j Y_j)] \\ &= \mathbb{E}[\exp(i \langle \xi^{(1)}, X_1, \dots, X_m \rangle)] \mathbb{E}[\exp(i \langle \xi^{(2)}, Y_1, \dots, Y_n \rangle)] \end{aligned}$$

où $\xi^{(1)} = (\xi_1, \dots, \xi_m)$, $\xi^{(2)} = (\xi_{m+1}, \dots, \xi_{m+n})$.

Par la propriété des transformé de Fourier des v.a., on en déduit que (X_1, \dots, X_m) indépendant de (Y_1, \dots, Y_n) .

□

Théorème 2.5.3. *Soit (X, Y_1, \dots, Y_n) vecteur gaussien centré (i.e. $\mathbb{E}[X] = 0$, $\mathbb{E}[Y_i] = 0$, $\forall i$). Alors $\mathbb{E}[X|Y_1, \dots, Y_n]$ est la projection orthogonale sur l'espace vectoriel engendré par Y_1, \dots, Y_n .*

C'est-à-dire qu'il existe $\lambda_1, \dots, \lambda_n$ t.q. $\mathbb{E}[X|Y_1, \dots, Y_n] = \sum_{i=1}^n \lambda_i Y_i$.

Plus généralement, pour toute fonction mesurable $h : \mathbb{R} \rightarrow \mathbb{R}_+$,

$$\mathbb{E}[h(X)|Y_1, \dots, Y_n] = \int_{\mathbb{R}} h(x) \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx.$$

avec $\sigma^2 = \mathbb{E}[(X - \sum \lambda_i Y_i)^2]$: carré de la distance de X à sa projection orthogonale.

Démonstration. On suppose que X non mesurable par rapport à $\sigma(Y_1, \dots, Y_n)$. Soit $\hat{X} = \sum \lambda_i Y_i$ la projection orthogonale de X sur l'espace vectoriel engendré par les Y_i .

Par $(X - \hat{X}) \perp Y_i$, $\text{cov}(X - \hat{X}, Y_i) = \mathbb{E}[(X - \hat{X})Y_i] = 0$. Donc $(X - \hat{X}, Y_1, \dots, Y_n)$ est un vecteur gaussien et $(X - \hat{X})$ est indépendant de (Y_1, \dots, Y_n) .

Donc

$$\begin{aligned}\mathbb{E}[X|Y_1, \dots, Y_n] &= \mathbb{E}[X - \hat{X}|Y_1, \dots, Y_n] + \mathbb{E}[\hat{X}|Y_1, \dots, Y_n] \\ &= \mathbb{E}[X - \hat{X}|Y_1, \dots, Y_n] + \hat{X} \\ &= \underbrace{E[X - \hat{X}]}_{=0} + \hat{X}\end{aligned}$$

Ensuite, soit $Z = X - \hat{X}$, alors $Z \sim \mathcal{N}(0, \sigma^2)$ et Z indépendante de (Y_1, \dots, Y_n) .

Alors

$$\begin{aligned}\mathbb{E}[h(X)|Y_1, \dots, Y_n] &= \mathbb{E}[h(\hat{X} + Z)|Y_1, \dots, Y_n] \\ &= \int h(\hat{X} + z)\mathbb{P}_Z(dz)\end{aligned}$$

□

Chapitre 3

Martingale

3.1 Définitions et exemples

En français, le mot "martingale" n'a pas le même sens en mathématiques et dans le langage courant.

Dans le langage courant, on parle de "martingale" par une méthode de jeu pour quelqu'un qui joue au casino.

En mathématiques, une martingale est la quantité d'argent que possède le joueur.

Soit X une variable aléatoire, et \mathcal{F} une tribu

Définition 3.1.1. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ espace de probabilité. On appelle filtration une suite croissante de sous tribus de $\mathcal{F} : (\mathcal{F}_n)$ t.q. $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_n \subset \dots \subset \mathcal{F}$.

Exemples 3.1.2. Soit X_0, \dots, X_n variable aléatoire, $\mathcal{F}_n = \sigma(X_i, i \leq n)$. Alors $\mathcal{F}_n \subset \mathcal{F}_{n+1} = \sigma(\mathcal{F}_n, X_{n+1})$.

Soit $\Omega = [0,1]$, \mathbb{P} une mesure uniforme, \mathcal{F} une tribu borélienne est définie par $\sigma([\frac{i}{2^n}, \frac{i+1}{2^n}], 0 \leq i \leq 2^n - 1)$.

Définition 3.1.3. Une processus à temps discret $(X_n)_{n \geq 0}$ est une suite de variable aléatoire.

Une processus à temps continue $(X_r)_{r \in \mathbb{R}_+}$ est une famille de variable aléatoire indexée par \mathbb{R}_+ . (Exemple : mouvement Brownien)

Définition 3.1.4. Un processus (X_n) est adapté à la filtration (\mathcal{F}_n) si $\forall x$ X_n est mesurable par rapport à \mathcal{F}_n .

Définition 3.1.5. Soit $(X_n)_{n \geq 0}$ un processus adapté à une filtration (\mathcal{F}_n) . On suppose que pour tout n , $\mathbb{E}[|X_n|] < \infty$. (X_n) est une martingale par rapport à \mathcal{F}_n si pour tout n , on a $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$.

Remarque 3.1.6. Assez souvent, $\mathcal{F}_n = \sigma(X_i, i \leq n)$, $\mathbb{E}[X_{n+1}|X_1, \dots, X_n] = X_n$.

Définition 3.1.7. Sur-martingale si $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq X_n$.

Sous-martingale si $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n$.

En anglais, martingale, supermartingale, submartingale.

Exemple 3.1.8. Si (Y_i) suite de v.a. i.i.d., $\mathbb{E}[Y_i] = 0$. $X_n = Y_1 + \dots + Y_n$, $\mathcal{F}_n = \sigma(X_1, \dots, X_n) = \sigma(Y_1, \dots, Y_n)$. Alors (X_n) est une martingale par rapport à \mathcal{F}_n .

$$\begin{aligned} \mathbb{E}[X_{n+1}|\mathcal{F}_n] &= \mathbb{E}[X_n + Y_{n+1}|\mathcal{F}_n] = \mathbb{E}[X_n|\mathcal{F}_n] + \mathbb{E}[Y_{n+1}|\mathcal{F}_n] \\ &= X_n + \mathbb{E}[Y_{n+1}|\mathcal{F}_n] = X_n \end{aligned}$$

Y_{n+1} est indépendante de Y_1, \dots, Y_n dans indépendante de \mathcal{F}_n , donc

$$\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = \mathbb{E}[Y_{n+1}] = 0.$$

On a un joueur qui joue au casino. Au temps n , il joue et son gain est Y_n . La qualité $Y_1 + \dots + Y_n = X_n$ est la qualité d'argent du joueur au temps n .

Remarque 3.1.9. Par $\mathbb{E}[X_{n+1}] = \mathbb{E}[\mathbb{E}[X_{n+1}|\mathcal{F}_n]] = \mathbb{E}[X_n] = \mathbb{E}[X_0]$, en moyenne, on ne gagne rien au casino si le jeu est équitable ($\mathbb{E}[Y_i] = 0$).

Exemple 3.1.10 (Processus de Branchement). Idée : on modélise l'évaluation d'une population.

Soit $(Z_n)_{n \geq 0}$ processus. Pour tout n , Z_n est le nombre d'individus à la n-ième génération.

On considère des variables aléatoires $(X_n^{(k)})$ i.i.d, à valeur dans \mathbb{N} et L^1 .

Soit $(X_n^{(k)})$ représente le nombre d'enfants du n-ième individu à la génération k .

Alors

$$\begin{aligned}
Z_0 &= 1 \\
Z_1 &= X_1^{(0)} \\
Z_2 &= X_1^{(1)} + X_1^{(1)} + \cdots + X_{Z_1}^{(1)} \\
&\dots \\
Z_{n+1} &= X_1^{(n)} + X_1^{(n)} + \cdots + X_{Z_n}^{(n)}
\end{aligned}$$

Processus de Galton-Watson : Galton s'intéressait aux noms de famille, transmis par les hommes. Si $Z_1 = 0$ alors $Z_{n+1} = 0$, $Z_n = 0$, $k \geq n$ la population s'éteint (Z_n) processus, $\mathcal{F}_n = \sigma(Z_1, \dots, Z_n)$.

$$\begin{aligned}
\mathbb{E}[Z_{n+1} | \mathcal{F}_n] &= \mathbb{E}[X_1^{(n)} + X_1^{(n)} + \cdots + X_{Z_n}^{(n)} | \mathcal{F}_n] \\
&= \mathbb{E}\left[\sum_{k=1}^{\infty} X_n^{(k)} 1_{k \leq Z_n} | \mathcal{F}_n\right]
\end{aligned}$$

Par le lemme de convergence monotone

$$\begin{aligned}
&= \sum_{k=1}^{\infty} \mathbb{E}[X_n^{(k)} 1_{k \leq Z_n} | \mathcal{F}_n] \\
&= \sum_{k=1}^{\infty} 1_{k \leq Z_n} \mathbb{E}[X_n^{(k)} | \mathcal{F}_n]
\end{aligned}$$

On sait que $\mathcal{F}_n = \sigma(Z_1, \dots, Z_n)$ pour tout i , Z_i est une fonction des $X_k^{(i-1)}$. Dans $\sigma(Z_1, \dots, Z_n) \subset \sigma(X_n^{(i)}, i \leq n-1)$.

Pour tout k , $X_n^{(k)}$ est indépendant des $(X_j^{(i)})$, $i \leq n-1$, de $\sigma(X_j^{(i)}, i \leq n-1)$.

Ceci implique que $X_n^{(k)}$ indépendant de \mathcal{F}_n , alors $\mathbb{E}[X_n^{(k)} | \mathcal{F}_n] = \mathbb{E}[X_n^{(k)}] = m$. Ici m est le nombre moyen d'enfants.

Alors

$$\sum_{k=1}^{\infty} 1_{k \leq Z_n} \mathbb{E}[X_n^{(k)} | \mathcal{F}_n] = m \sum_{k=1}^{\infty} 1_{k \leq Z_n} = m Z_n = \mathbb{E}[Z_{n+1} | \mathcal{F}_n].$$

Pour $m \geq 1$, (Z_n) est une sous-martingale. Pour $m \leq 1$, (Z_n) est une sur-martingale. Pour

$m = 1, (Z_n)$ est une martingale.

Si la loi de $X_1^{(1)}$ n'est pas δ_1 et si $m = 1$ alors p.s il existe n , $Z_n = 0$. (résultat énoncé par Galtron, démonstration fausse. Watson a donnée une démonstration juste.)

Définition 3.1.11. Soit $(H_n)_{n \geq 1}$ processus est prévisible par rapport à une filtration $(\mathcal{F}_n)_{n \geq 0}$ si $\forall n$, H_n borné et \mathcal{F}_{n-1} -mesurable.

Proposition 3.1.12. Soit (X_n) processus adapté, (H_n) prévisible. On pose $(H.X)_0 = 0$ et $n \geq 1$, $(H.X)_n = H_1(X_1 - X_0) + H_2(X_2 - X_1) + \dots + H_n(X_n - X_{n-1})$. Alors

1. Si (X_n) est une martingale, alors $(H.X)_n$ est également une martingale.
2. Si X_n est une sous-martingale et si $H_n \geq 0$ p.s. pour tout $n \geq 1$. Alors $(H.X)_n$ est également une sous-martingale.

Idée : H est la quantité d'argent mise au casino au temps n par un joueur .

On dit H est prévisible : le joueur décide de la quantité d'argent mise au temps n sachant ce qui s'est passé jusqu'à temps $n - 1$, mais sans savoir ce qui se passe un temps n . Le quantité d'argent gagnée au temps n égale le quantité d'argent mise.

Démonstration. Soit H_n bornées, $(X_n) \in L^1$, alors pour tout n , on a $(H.X)_n \in L^1$.

Car (H_n) est prévisible et (X_n) adapté, alors $(H.X)_n$ adapté. Alors on a $\mathbb{E}[(H.X)_{n+1} | \mathcal{F}_n] = \mathbb{E}[(H.X)_n + H_{n+1}(X_{n+1} - X_n) | \mathcal{F}_n] = (H.X)_n + H_{n+1} \mathbb{E}[(X_{n+1} - X_n) | \mathcal{F}_n] = (H.X)_n$.

□

Rappel 3.1.13. Une martingale par rapport à la filtration \mathcal{F}_n est une suite de v.a. M_n dans L^1 telle que pour tout n , $\mathbb{E}[M_{n+1} | \mathcal{F}_n] = M_n$.

On utilise beaucoup de vocabulaire des jeux de hasard. Voir Mme Cleays si vous avez des difficultés avec ce vocabulaire.

Si (H_n) prévisible, alors $(H.X)_n$ est une martingale.

Exemple 3.1.14. Soient $Y_1, \dots, Y_n \dots$ i.i.d. et $\mathbb{E}[Y_i] = 0$. Alors on a $X_n = Y_1 + \dots + Y_n$ est une martingale.

Si on joue à pile ou face, soit $Y_n = \begin{cases} +1 & \text{pile} \\ -1 & \text{face} \end{cases}$, de plus on a $\mathbb{P}(Y_1 = 1) = \mathbb{P}(Y_1 = -1) = \frac{1}{2}$.

Alors on prend $H_n = \begin{cases} 1 & \text{si } X_{n-1} \geq 0 \\ 2 & \text{si } X_{n-1} < 0 \end{cases}$.

Evidemment on sait que H_n est prévisible par bornée et mesurable par rapport à \mathcal{F}_{n-1} .

Par définitions, on a $(H.X)_n = \begin{cases} (H.X)_{n-1} + Y_n & \text{si } X_{n-1} \geq 0 \\ (H.X)_{n-1} + 2Y_n & \text{si } X_{n-1} < 0 \end{cases}$.

Alors on vérifie que $\mathbb{E}[(H.X)_n | \mathcal{F}_n] = (H.X)_{n-1}$.

Exemple 3.1.15. On définit X_n, Y_n comme ci-dessus et on prend $H_n = \begin{cases} 1 & \text{si } Y_n = 1 \\ 0 & \text{si } Y_n = -1 \end{cases}$.

Alors H_n n'est pas prévisible car il est mesurable par rapport à \mathcal{F}_n mais pas à \mathcal{F}_{n-1} . On a

$(H.X)_n = \begin{cases} (H.X)_{n-1} + Y_n & \text{si } Y_n = 1 \\ (H.X)_{n-1} & \text{si } Y_n = 0 \end{cases}$. De plus $\mathbb{E}[(H.X)_n | \mathcal{F}_n] = (H.X)_{n-1} + \frac{1}{2}$. C'est-à-dire

que $(H.X)_n$ n'est pas une martingale.

Proposition 3.1.16. Soient (X_n) une processus adapté, $\varphi : \mathbb{R} \rightarrow \mathbb{R}_+$ convexe, telle que $\mathbb{E}[\varphi(X_n)] < \infty$. Alors on a

1. Si (X_n) est une martingale alors $\varphi(X_n)$ est une sous-martingale.
2. Si (X_n) est une sous-martingale et φ est croissant alors $\varphi(X_n)$ est une sous-martingale.

Démonstration. Par l'inégalité de Jensen pour l'espérance conditionnelle, on a

$$\mathbb{E}[\varphi(X_n) | \mathcal{F}_{n-1}] \geq \varphi(\mathbb{E}[X_n | \mathcal{F}_{n-1}]) = \varphi(X_{n-1}).$$

Par φ est croissant et $X_n \leq \mathbb{E}[X_{n+1} | \mathcal{F}_n]$, on a

$$\mathbb{E}[\varphi(X_{n+1}) | \mathcal{F}_n] \geq \varphi(\mathbb{E}[X_{n+1} | \mathcal{F}_n]) \geq \varphi(X_n)$$

□

3.2 Temps d'arrêt

Idée : Un joueur joue au casino et décide de s'arrêter à un certain moment. (noter de temps d'arrêt)

Définition 3.2.1. Soit (\mathcal{F}_n) une filtration, on dit que une v.a. $T : \Omega \rightarrow \mathbb{N} \cup \{+\infty\} = \bar{\mathbb{N}}$ est un temps d'arrêt de la filtration \mathcal{F}_n si pour tout $n \in \mathbb{N}$ on a

$$\{T = n\} \subset \mathcal{F}_n.$$

L'ensemble $\{T = n\}$ est l'évènement "on s'arrête au temps n ". Cet évènement dépend de la qualité d'information qu'on a au temps n , cette quantité d'informations est donnée par \mathcal{F}_n .

Exemples 3.2.2. 1. $k \in \mathbb{N}$, si pour tout $\omega \in \Omega$, $T(\omega) = k$. T est un temps d'arrêt. Pour tout n , $\{T = n\} = \Omega$ ou \emptyset qui appartient à \mathcal{F}_n par définitions d'une tribu.

2. Temps d'arrêt : (Y_n) processus adapté à valeurs dans \mathbb{R}^d . Si B est un borélien contenu dans \mathbb{R}^d . On définit $T_B = \inf\{n, Y_n \in B\}$ est un temps d'arrêt.

Exemple 3.2.3 (Contre exemple). Dernière temps de suite (Y_n) processus adapté à (\mathcal{F}_n) , à valeurs dans \mathbb{R}^d , B borélien de \mathbb{R}^d . $T_B = \sup\{n, Y_{n-1} \in B, Y_n \notin B\}$.

Alors on a

$$\begin{aligned} \{T_B = n\} &= \{Y_{n-1} \in B, Y_n \notin B, \forall m \geq n, \text{ on n'a pas } Y_m \in B, Y_{m+1} \in B\} \\ &= \{Y_{n-1} \in B, Y_n \in B\} \cap \{\forall m \geq n, \text{ on n'a pas } Y_m \in B, Y_{m+1} \in B\}. \end{aligned}$$

On décide de s'arrêter en croissant le passé mais pas en connaissant le futur.

Proposition 3.2.4. 1. Si S, T sont des temps d'arrêt, alors $S \vee T, S \wedge T$ sont des temps d'arrêt. ($S \vee T = \sup(S, T)$ et $S \wedge T = \inf(S, T)$)

2. Si (T_n) est une suite de temps d'arrêt. Alors $\inf(T_n), \sup(T_n), \limsup(T_n), \liminf(T_n)$ sont des temps d'arrêt.

Démonstration. On voit que

$$\begin{aligned} \{S \wedge T \leq n\} &= \{S \leq n\} \cup \{T \leq n\} \in \mathcal{F}_n \\ \{S \vee T \leq n\} &= \{S \leq n\} \cap \{T \leq n\} \in \mathcal{F}_n \end{aligned}$$

Donc $S \vee T, S \wedge T \in \mathcal{F}_n$.

Par $\{\inf(T_n) \leq n\} = \bigcap_{k \in \mathbb{N}} \{T_n \leq n\}$ et les formules triviales, on a pariellement les autres sont des temps d'arrêt. \square

Définition 3.2.5. Soit T un temps d'arrêt pour (\mathcal{F}_n) . La tribu du passé jusqu'à l'instant T est

$$\mathcal{F}_T = \{A \in \mathcal{F}, \forall n \in \mathbb{N}, A \cap \{T = n\} \in \mathcal{F}_n\}.$$

Proposition 3.2.6. Si S et T deux temps d'arrêt, et si $S \leq T$. Alors $\mathcal{F}_S \subset \mathcal{F}_T$.

Démonstration. Pour $A \in \mathcal{F}_S$, alors $\forall n \in \mathbb{N}$, on a

$$A \cap \{T = n\} = \bigcup_{k \leq n} (A \cap \{S = k\} \cap \{T = n\}) \in \mathcal{F}_n$$

Donc on a $A \in \mathcal{F}_T$. □

Exemple 3.2.7. Soit (Y_n) une processus adapté par rapport à \mathcal{F}_n , à valeurs dans \mathbb{R}^d , B borélien de \mathbb{R}^d . Comme précédent, on défini $T_B = \inf\{n, Y_n \in B\}$ et A un évènement $\{\exists n, Y_n \in B\}$. Alors $A \in \mathcal{F}_{T_B} : A \cap \{T = n\} = \{Y_0 \notin A, \dots, Y_{n-1} \notin A, Y_n \in A\} = \{T = n\} \in \mathcal{F}_n$.

Proposition 3.2.8. Soit (Y_n) adapté par rapport à \mathcal{F}_n et T un temps d'arrêt par rapport à \mathcal{F}_n . Alors la variable aléatoire $1_{\{T < \infty\}} Y_T$ définie par

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

est mesurable par rapport à \mathcal{F}_T .

Démonstration. Soit B borélien, $n \in \mathbb{N}$. Alors pour tout n , on a

$$\{1_{\{T < \infty\}} Y_T \in B\} \cap \{T = n\} = \{Y_n \in B\} \cap \{T = n\} \in \mathcal{F}_n.$$

□

Théorème 3.2.9. Soit (X_n) une martingale adaptée à \mathcal{F}_n et T un temps d'arrêt de la \mathcal{F}_n . Alors $(X_{T \wedge n})_{n \in \mathbb{N}}$ est une martingale. En particulier, si le temps T est borné alors $X_T \in L^1$ et $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

Idée: Si (X_n) est la quantité d'argent d'un joueur au casino et si T est le temps où le joueur décide de s'arrêter. On définit

$$X_{T \wedge n} = \begin{cases} X_n & \text{Si } T \geq n (\text{le joueur joue encore au temps } n) \\ X_T & \text{Si } T \leq n (\text{le joueur a arrêté de jouer au temps } n) \end{cases}.$$

Alors $(X_{T \wedge n})$ est une martingale et $\mathbb{E}[X_T] = \mathbb{E}[X_0]$: en moyenne, le joueur n'a pas gagné d'argent au moment où il s'arrête.

Démonstration. Pour tout n , on a vu que $T \wedge n$ est un temps d'arrêt. Pour $n \geq 1$, on définit $H_n = 1_{\{T \geq n\}} = 1 - 1_{\{T \leq n-1\}}$. C'est un processus prévisible parce que $1_{\{T \leq n-1\}}$ est mesurable par rapport à \mathcal{F}_{n-1} , ce qui implique que H_n mesurable par rapport à \mathcal{F}_{n-1} .

La proposition 3.1.12 nous dit que $(X_0 + (H \cdot X)_n)_{n \geq 0}$ est une martingale. D'ailleurs on trouve que $X_0 + (H \cdot X)_n = X_{T \wedge n}$, ceci conclut la démonstration.

De plus si T est borné, il existe $N \in \mathbb{N}$, $T \leq N$ p.s., alors $\mathbb{E}[X_T] = \mathbb{E}[X_{T \wedge n}] = \mathbb{E}[X_0]$.

□

Application 3.2.10 (Risque de ruine du joueur). Un joueur joue au casino, au début sa fortune est $X_0 = m$. Soit (Y_n) une suite de v.a. i.i.d. et $\mathbb{P}(Y_1 = 1) = \mathbb{P}(Y_1 = -1) = \frac{1}{2}$.

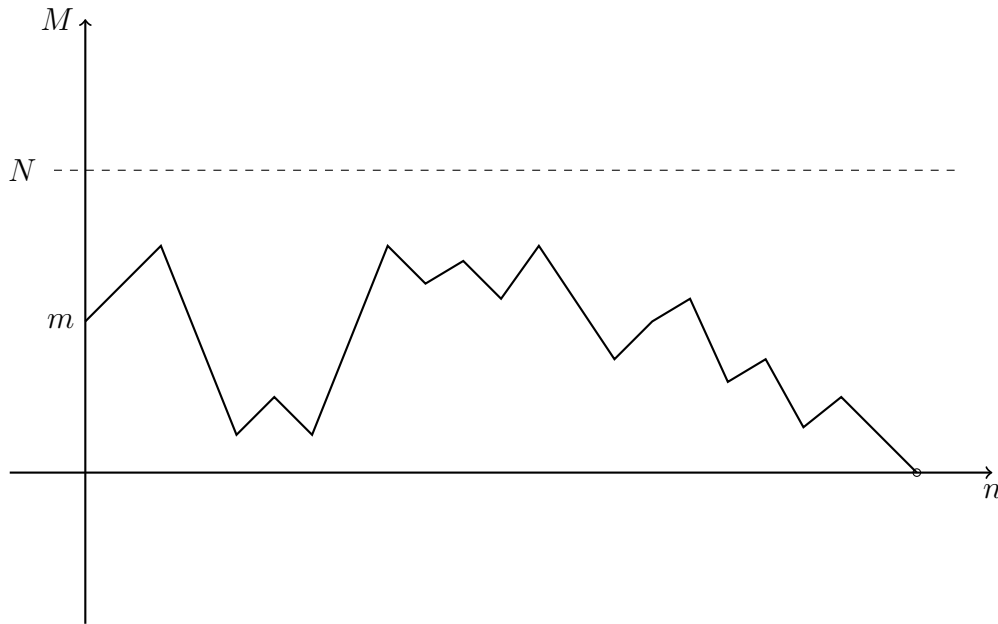
La fortune du joueur au temps n est $X_0 + Y_1 + \dots + Y_n = X_n$. Alors X_n est une martingale.

Le joueur décide de s'arrêter quand sa fortune atteint une valeur N . D'ailleurs, il faut s'arrêter dans le temps où il est ruiné parce qu'il n'a plus d'argent et il ne peut plus jouer.. Donc on pose $B = [N, +\infty)$ et $B' = [-\infty, 0]$. Alors T_B est le temps d'arrêt dans B et $T_{B'}$ est le temps d'arrêt dans B' .

Soit $T = T_B \wedge T_{B'} = T_{B''}$ temps d'arrêt dans $B'' = (-\infty, 0] \cup [N, \infty)$, premier temps où soit le joueur est ruiné, soit il a N dollars.

$$X_T = \begin{cases} N & \text{si } T_B < T_{B'} \Leftrightarrow T_{B''} = T_B \\ 0 & \text{si } T_{B'} < T_B \Leftrightarrow T_{B''} = T_{B'} \end{cases}$$

On peut trouver que la probabilité de ruine est exactement $\mathbb{P}(X_T = 0)$.



Par $X_n = X_0 + Y_1 + Y_2 + \dots + Y_n$, on a vu proposition précédente que $\limsup(Y_1 + \dots + Y_n) = +\infty$ p.s., c'est-à-dire que $T_B < \infty$. De même, $\liminf(Y_1 + \dots + Y_n) = -\infty$ p.s., c'est-à-dire que $T_{B'} < \infty$.

Donc on a $T = T_B \wedge T_{B'} < +\infty$ p.s..

Par le théorème 3.2.9, $(X_{T \wedge n})$ est une martingale. Alors, on pose $M_n = X_{T \wedge n}$. Pour tout $K \in \mathbb{N}$, $\mathbb{E}[M_K] = \mathbb{E}[M_0] = \mathbb{E}[X_0]$.

Pour le temps K , on distingue trois cas :

- Soit $T \leq K$, $T = T_B$, maintenant le joueur a atteint un valeur N : $X_{T \wedge K} = N$.
- Soit $T \leq K$, $T = T_{B'}$, ici il n'a plus d'argent : $X_{T \wedge K} = 0$.
- Soit $T > K$.

Donc on obtient l'égalité suivante

$$\begin{aligned} \mathbb{E}[X_0] &= \mathbb{E}[M_K] = \mathbb{E}[M_K(1_{\{T \leq K, T=T_B\}} + 1_{\{T \leq K, T=T_{B'}\}} + 1_{\{T > K\}})] \\ &= N\mathbb{P}(T \leq K, T = T_B) + 0 \cdot \mathbb{P}(T \leq K, T = T_{B'}) + \mathbb{E}[M_k 1_{T > K}]. \end{aligned}$$

Sur l'évènement $\{T > K\}$, $1 \leq M \leq N - 1$. Donc

$$1 \cdot \mathbb{P}(T \geq K) \leq \mathbb{E}[M_K 1_{\{T > K\}}] \leq (N - 1) \cdot \mathbb{P}(T \geq K) \quad (3.2.1)$$

Comme $T < \infty$ p.s., on a $\mathbb{P}(T \geq K) \rightarrow 0$ quand $K \rightarrow \infty$. D'après 3.2.1, on sait que $\mathbb{E}[M_K 1_{T > K}] \xrightarrow{K \rightarrow \infty} 0$.

Donc

$$\mathbb{E}[X_0] = N\mathbb{P}(T \leq K, T = T_B) + \mathbb{E}[M_k 1_{T > K}] \xrightarrow{K \rightarrow \infty} \mathbb{P}(T = T_B).$$

Ceci conclut que

$$\mathbb{P}(T = T_B) = \frac{m}{N}.$$

On obtient la risque de ruine est $\mathbb{P}(X_T = 0) = \mathbb{P}(T = T_B) = 1 - \frac{m}{N}$.

C'est l'application la plus classique au théorème d'arrêt pour les martingales. Ce n'est pas une application immédiate car T n'est pas borné.

Remarque 3.2.11. Si T_B = temps d'arrêt dans $[N, \infty)$ est un temps d'arrêt, fini p.s. Ensuite on a $(X_{Y_B \wedge n})$ est une martingale.

Finalement on a $X_{T_B} = N \neq \mathbb{E}[X_0] = m$. Mais on ne peut pas appliquer la deuxième partie du théorème d'arrêt ici car T_B n'est pas borné.

On rappelle que si T est un temps d'arrêt, M martingale, alors $(M_{T \wedge n})$ est une martingale. De plus si T est borné, on a $\mathbb{E}(M_T) = \mathbb{E}(M_0)$.

3.3 Convergence presque sûre des martingales

Théorème 3.3.1. *Si (M_n) est une martingale positive (i.e. $\forall n, M_n \geq 0$, p.s.) Alors M_n converge presque sûrement : p.s. $\exists M_\infty, M_n \rightarrow M_\infty$.*

Cas des martingales bornés dans L^2

On va montrer d'abord que pour (M_n) martingales telle que $\forall n, \mathbb{E}(M_n^2) < \infty$ et $\sup \mathbb{E}(M_n^2) < \infty$, le théorème est vrai.

Proposition 3.3.2. *Soit $A_n = M_{n+1} - M_n$. Alors les variables aléatoires A_n sont dans L^2 et sont orthogonales : $\forall m \neq n, \mathbb{E}(A_m A_n) = 0$.*

Remarque 3.3.3. On a $\mathbb{E}(A_{n+1}) = \mathbb{E}(M_{n+1}) - \mathbb{E}(M_n) = 0$.

Démonstration. Notons que

$$\begin{aligned} \mathbb{E}(A_{n+1}^2) &= \mathbb{E}((M_{n+1} - M_n)^2) = \mathbb{E}(\mathbb{E}((M_{n+1} - M_n)^2 | \mathcal{F}_n)) \\ &\leq \mathbb{E}(M_{n+1}^2) + \mathbb{E}(M_n^2) + 2\mathbb{E}(|M_{n+1} M_n|) < \infty. \end{aligned}$$

Donc $A_{n+1} \in L^2$. Ensuite, si $m < n$,

$$\mathbb{E}(A_m A_n) = \mathbb{E}\left((X_{m+1} - X_m)(X_{n+1} - X_n)\right) = \mathbb{E}\left(\mathbb{E}\left((X_{m+1} - X_m)(X_{n+1} - X_n) \mid \mathcal{F}_n\right)\right)$$

En notant que si $m < n$, $X_{m+1} - X_m$ est \mathcal{F}_n -mesurable, donc

$$\mathbb{E}(A_m A_n) = \mathbb{E}\left((X_{m+1} - X_m)\mathbb{E}(X_{n+1} - X_n \mid \mathcal{F}_n)\right) = 0.$$

(Car $\mathbb{E}(X_{n+1} - X_n \mid \mathcal{F}_n) = 0$.) □

Par conséquence, si $m \leq n$,

$$M_n = (M_n - M_{n-1}) + \cdots + (M_{m+1} - M_m) + M_m = A_{n-1} + \cdots + A_m + M_m,$$

$$\mathbb{E}(M_n^2) = \mathbb{E}(A_{n-1}^2) + \cdots + \mathbb{E}(A_m^2) + \mathbb{E}(M_m^2) \geq \mathbb{E}(M_m^2).$$

donc $\mathbb{E}(M_n^2)$ est une suite croissante. L'hypothèse M_n bornée dans L^2 est équivalent à $\mathbb{E}(M_n^2) \rightarrow L$ pour quelque $L < \infty$.

Soient maintenant $p < q \in \mathbb{Q}$, on note $N_{p,q}(n)$ le nombre de montées de l'intervalle $[p, q]$ avant l'instant n . Soient

$$S_1 = \inf\{m, M_m < p\}, \quad T_1 = \inf\{m \geq S_1, M_m > q\},$$

$$S_{k+1} = \inf\{m > T_k, M_m < p\}, \quad T_{k+1} = \inf\{m \geq S_k, M_m > q\}.$$

Alors $N_{p,q}(n) = \max\{k, T_k \leq n\}$.

Idée : à chaque montée, on augmente le carrée de la norme L^2 de (M_n) . Ce carrée de la norme est bornée donc le nombre de montées est borné :

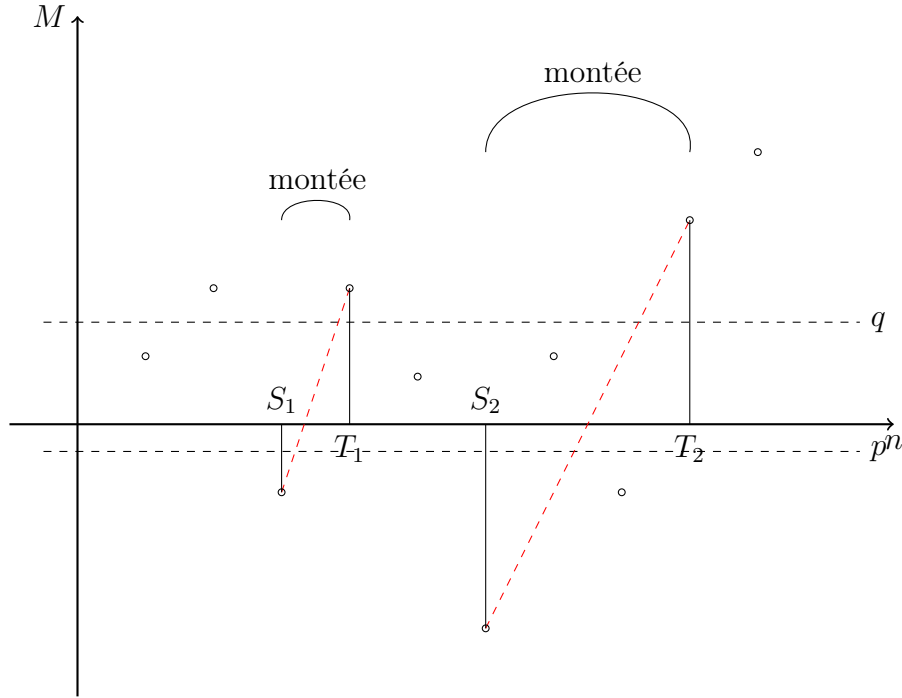
$$\mathbb{E}(N_{p,q}(n)(p - q)^2) \leq \mathbb{E}(M_n^2) = \mathbb{E}((A_1 + \cdots + A_n)^2) = \mathbb{E}(A_1^2 + \cdots + A_n^2).$$

D'autre part,

$$\forall k, \mathbb{E}((M_{T_k} - M_{S_k})^2) \geq (q - p)^2.$$

Donc sur l'événement $\{N_{p,q} = k\}$,

$$\mathbb{E}(M_n^2) \geq \mathbb{E}((M_{T_1} - M_{S_1})^2) + \cdots + \mathbb{E}((M_{T_k} - M_{S_k})^2) \geq k(p - q)^2.$$



Ainsi

$$\mathbb{E}(M_n^2) = \sum_{k=1}^{\infty} \mathbb{E}(M_n^2 \mathbf{1}_{\{N_{p,q}(n)=k\}}) \geq \sum_{k=1}^{\infty} k(q-p)^2 \mathbb{P}(N_{p,q}(n) = k) = (p-q)^2 \mathbb{E}(N_{p,q}(n)).$$

Comme $N_{p,q}(n)$ est une suite croissante et $\forall n$, $\mathbb{E}(N_{p,q}(n)) \leq \frac{L}{(p-q)^2}$. Alors $\mathbb{P}(N_{p,q}(n) \rightarrow \infty) = 0$, cela implique le nombre de montées de $[p, q]$ est fini p.s. C'est à dire M_n ne traverse pas l'intervalle $[p, q]$ pour n assez grand.

Alors le nombre de montées de $[p, q]$ est fini sauf sur Ω_{pq} , avec $\mathbb{P}(\Omega_{pq}) = 0$. Soit $\Omega' = \bigcup_{p,q \in \mathbb{Q}} \Omega_{pq}$, on a $\mathbb{P}(\Omega') = 0$. Le nombre de montées de tout intervalle de la forme $[p, q]$, $p < q$, $p, q \in \mathbb{Q}$, est fini sauf sur Ω' . On peut montrer la même chose pour les descentes. Donc p.s. $\forall p, q \in \mathbb{Q}$, (M_n) ne traverse $[p, q]$ qu'un nombre fini de fois. Donc p.s., (M_n) converge. (Ici on utilise la propriété en analyse : (U_n) converge s.s.i. $\forall p, q \in \mathbb{Q}$, U_n traverse $[p, q]$ un nombre fini de fois.) Cela démontre le théorème.

Cas général

Pour montrer le cas général : on conclut par la même idée de prouver qu'on traverse $[p, q]$ un nombre fini de fois. On peut montrer

$$\mathbb{E}(N_{p,q}(n)(q-p)) \leq \mathbb{E}((M_n - p)^+ - (M_0 - p)^+),$$

avec la notion $x^+ = \sup\{x, 0\}$, puis on utilise une démonstration analogue (P170-172, Le Gall).

3.4 Application : extinction des processus de Galton-Watson critiques

Processus de Galton-Watson : C'est un processus qui modélise l'évolution d'une population où chaque individu a un nombre aléatoire d'enfants suivant une loi μ , indépendamment des autres individus. Soit Z_n le nombre d'individus à la génération n , on a

$$\begin{aligned} Z_0 &= 1 \\ Z_1 &= X_1^{(1)} \sim \mu \\ Z_2 &= X_1^{(2)} + \dots + X_{Z_1}^{(2)} \\ &\dots \\ Z_{n+1} &= X_1^{(n+1)} + \dots + X_{Z_n}^{(n+1)} \end{aligned}$$

Si $Z_k = 0$, alors $Z_n = 0, \forall n \geq k$. Si le nombre moyen d'enfants est $\mathbb{E}(Z_1) = m$, on sait que $(\frac{Z_n}{m^n})$ est une martingale. On a énoncé le fait que quand $m = 1$ (resp. $m > 1$, resp. $m < 1$), c'est un processus critique (resp. surcritique, resp. sous-critique).

Théorème 3.4.1. *Si $m \leq 1$ et si $\mathbb{P}(Z_1 = 1) \neq 1$, alors p.s. le processus s'éteint : $\exists n \geq 1, Z_n = 0$.*

Démonstration. Soit $m = 1$, Z_n est une martingale positive, alors avec probabilité 1, elle converge vers Z_∞ qui est une variable aléatoire. Comme (Z_n) à valeurs dans \mathbb{N} ,

$$Z_n \text{ converge} \iff \exists l \in \mathbb{N}, \exists N, \forall n \geq N, Z_n = l.$$

Soit $l \in \mathbb{N}, l \geq 1$, soit $N \in \mathbb{N}$, on peut vérifier que $\mathbb{P}(\forall n \geq N, Z_n = l) = 0$ en notant que $\mathbb{P}(Z_{k+1} = l | Z_k = l) < 1$. Soit maintenant que $\Omega_{l,N} = \{\forall n \geq N, Z_n = l\}$, alors $\mathbb{P}(\Omega_{l,N}) = 0$.

Posons

$$\Omega' = \bigcup_{l \geq 1} \bigcup_{l \in \mathbb{N} N \in \mathbb{N}} \Omega_{l,N} \implies \mathbb{P}(\Omega') = 0,$$

mais comme la suite est convergente p.s. avec probabilité 1, il existe l, N t.q. $\forall n \geq N, Z_n = l$ donc p.s. $l = 0$, donc le processus s'éteint p.s.. \square

Remarque 3.4.2. La convergence est une convergence p.s., pas dans L^1 . Car $\mathbb{E}(Z_0) = 1$, $Z_\infty = 0$ p.s. $\mathbb{E}(Z_\infty) = 0 \neq \mathbb{E}(Z_0)$.

Retour sur la marche aléatoire simple : $X_n = X_0 + Y_1 + \dots + Y_n$ où les (Y_i) sont i.i.d.

$$\mathbb{P}(Y_1 = 1) = \mathbb{P}(Y_1 = -1) = \frac{1}{2}.$$

Alors (X_n) est une martingale et $T = T_0 = \inf\{n, X_n = 0\}$ est temps d'arrêt. Donc si $X_0 \geq 0$, $M_n = X_{T \wedge n}$ est une martingale positive. Si $M_n = 0$, alors $\forall m \geq n, M_m = 0$.

D'ailleurs on sait que M_n converge p.s. or (M_n) a des valeurs entières et $\mathbb{P}(M_n = M_{n+1} = l) = 0$ si $l \geq 1$, donc M_n ne peut être stationnaire qu'en 0, donc p.s. $M_n \rightarrow 0$ et p.s. $T < \infty$.

Remarque 3.4.3. On avait montré que $T < \infty$ p.s. en utilisant une loi de 0-1 de Kolmogorov. Le théorème de convergence des martingales donne une autre preuve de fait que $T < \infty$ p.s. Ici on n'a pas convergence dans L^1 puisque $M_0 = X_0 > 0$ et $M_\infty = 0$ p.s. donc $\mathbb{E}(M_0) \neq \mathbb{E}(M_\infty) = 0$. Mais dans certains cas, on a bien la convergence dans L^1 .

Théorème 3.4.4. Soit (X_n) martingale, alors les deux assertions sont équivalentes :

1. (X_n) converge p.s. et dans L^1 .
2. il existe une variable aléatoire $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ telle que $\forall n, X_n = \mathbb{E}(Z | \mathcal{F}_n)$.

Dans ce cas, $Z = X_\infty$ p.s. et on dit que (X_n) est une martingale fermée.

Démonstration. (1) \Rightarrow (2) : si $m > n$, $X_n = \mathbb{E}(X_m | \mathcal{F}_n)$. Or l'application $Y \mapsto \mathbb{E}(Y | \mathcal{F}_n)$ est une contraction dans L^1 : $\mathbb{E}|Y| \geq \mathbb{E}|\mathbb{E}(Y | \mathcal{F}_n)|$. Dans l'équation $X_n = \mathbb{E}(X_m | \mathcal{F}_n)$, posons $m \rightarrow \infty$, on a $X_n = \mathbb{E}(X_\infty | \mathcal{F}_n)$.

(2) \Rightarrow (1) : Dans ce cas X_n est bornée dans L^1 . On a montré que les martingales positives convergent p.s. On montre de même que les sous-martingales positives convergent p.s., puis que les sous-martingales bornées dans L^1 convergent p.s. Et puis notons que $X_n^+ = \inf(X_n, 0)$ converge p.s. et $X_n - X_n^+$ converge p.s. cela implique X converge p.s. vers X_∞ .

Si Z est bornée : $\exists k, |Z| < k$ p.s. alors $|X_n| < k$ p.s. et par le théorème de convergence dominées, X_n converge vers X_∞ dans L^1 .

Si Z n'est pas bornée : soit $\epsilon > 0$ et soit M t.q. $\mathbb{E}(Z1_{\{|Z|>M\}}) < \epsilon$, on a

$$\forall n, \mathbb{E}\left(X_n - \mathbb{E}(Z1_{\{|Z|\leq M\}}|\mathcal{F}_n)\right) = \mathbb{E}\left(Z(1 - 1_{\{|Z|\leq n\}})\middle|\mathcal{F}_n\right) \leq \epsilon,$$

donc $(\mathbb{E}(Z1_{\{|Z|<M\}}|\mathcal{F}_n))_{n\geq 0}$ est une martingale bornée, d'où la convergence p.s. et dans L^1 :

$$\exists n_0, \forall m, n \geq n_0, \mathbb{E}\left|\mathbb{E}(Z1_{\{|Z|<M\}}|\mathcal{F}_m) - \mathbb{E}(Z1_{\{|Z|<n\}}|\mathcal{F}_n)\right| \leq \epsilon.$$

On en déduit que $\mathbb{E}(|X_m - X_n|) \leq 3\epsilon$. Donc (X_n) est une suite de Cauchy dans L^1 et comme L^1 est compact, (X_n) converge dans L^1 . \square

Corollaire 3.4.5. Soit $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, $X_n = \mathbb{E}(Z|\mathcal{F}_n)$ est une martingale qui converge p.s. et dans L^1 vers $\mathbb{E}(Z|\mathcal{F}_\infty)$ où \mathcal{F}_∞ est la tribu engendrée par les \mathcal{F}_n : la plus petite tribu contenant toutes les \mathcal{F}_n .

Démonstration. Dans le théorème précédent, on a montré la convergence p.s. et dans L^1 vers X_∞ . Il faut montrer $X_\infty = \mathbb{E}(Z|\mathcal{F}_\infty)$. Pour tout n , X_n est \mathcal{F}_n -mesurable et $X_n \xrightarrow{p.s.} X_\infty$, donc X_∞ est \mathcal{F}_∞ -mesurable. Si $A \in \mathcal{F}_n$, $\mathbb{E}(Z1_A) = \mathbb{E}(X_n1_A) = \mathbb{E}(X_\infty1_A)$. Par le théorème de classe monotone : c'est vrai pour tout A qui est dans la tribu engendré par les \mathcal{F}_n , donc $X_\infty = \mathbb{E}(Z|\mathcal{F}_\infty)$. \square

Exemple 3.4.6. Soit $\Omega = [0, 1]$, \mathcal{F} la tribu borélienne et \mathbb{P} la mesure de Lebesgue. Soient μ une mesure absolument continue par rapport à \mathbb{P} , X la variable aléatoire de loi μ , soient

$$\mathcal{F}_n = \sigma\left(\left[\frac{i}{2^n}, \frac{i+1}{2^n}\right], 0 \leq i \leq 2^n - 1\right), \quad M_n = \mathbb{E}(X|\mathcal{F}_n).$$

Alors M_n converge p.s. et dans L^1 vers M_∞ . i.e. $x \in \Omega = [0, 1]$, $M_n(x)$ converge vers $M_\infty(x)$ p.s. Ici $\mathcal{F}_\infty = \mathcal{F}$, et par la corollaire, $M_\infty = X$ p.s. Donc $\forall A \in \mathcal{F}$:

$$\mu(A) = \int M_\infty(x)1_A(x)dx = \int_A M_\infty(x)dx,$$

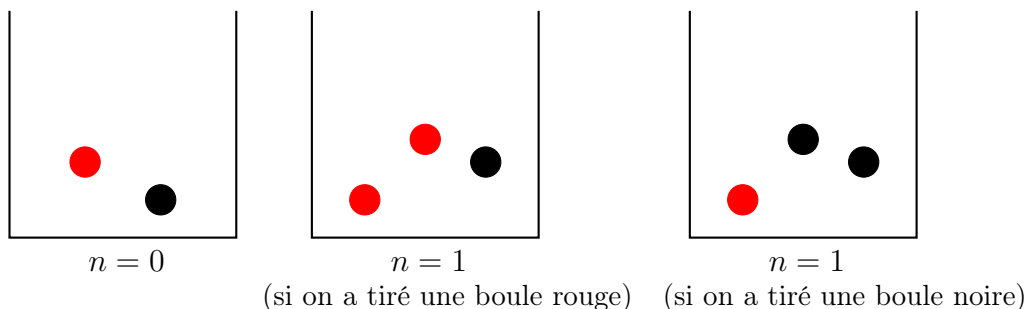
M_∞ est la densité de Radon-Nikodym de μ .

3.5 Urne de Pólya

On a une urne avec des boules noires ou rouges. Au temps 0, il y en a une boule noire et une boule rouge. Au temps 1, on tire au hasard une boule dans l'urne avec la probabilité

$$\mathbb{P}(\text{boule rouge}) = \mathbb{P}(\text{boule noire}) = \frac{1}{2}.$$

Ensuite on remet la boule dans l'urne et on ajoute un boule de la même couleur.



Au temps $n + 1$: on tire une boule au hasard, on la remet dans l'urne et on ajoute une boule de la même couleur. Soit R_n le nombre de boules rouges au temps n . Donc $R_0 = 1$ et au temps n , on a $n + 2$ boules. Au temps $n + 1$, on a

$$\mathbb{P}(\text{tirer une boule rouge}) = \frac{R_n}{n + 2}.$$

Soit M_n la proportion de boules rouges au temps n : $M_n = \frac{R_n}{n+2}$. Et supposons $\mathcal{F}_n = \sigma(M_1, \dots, M_n)$, alors \mathcal{F}_n est une filtration.

Théorème 3.5.1. M_n est une martingale par rapport à \mathcal{F}_n .

Démonstration. Soit A_{n+1} l'événement { on a tiré une boule rouge au temps $n + 1$ }, alors

$$\mathbb{E}(M_{n+1} | \mathcal{F}_n) = \mathbb{E}(M_{n+1}(1_{A_{n+1}} + 1_{A_{n+1}^c}) | \mathcal{F}_n),$$

et

$$\begin{aligned} \mathbb{E}(M_{n+1} 1_{A_{n+1}} | \mathcal{F}_n) &= \mathbb{E}\left(\frac{R_n + 1}{n + 3} 1_{A_{n+1}} | \mathcal{F}_n\right) = \frac{R_n + 1}{n + 3} \mathbb{E}(1_{A_{n+1}} | \mathcal{F}_n) \\ &= \frac{R_n + 1}{n + 3} \mathbb{P}(A_{n+1} | \mathcal{F}_n) = \frac{R_n + 1}{n + 3} \cdot \frac{R_n}{n + 2}. \end{aligned}$$

Comme $\forall k$, $\mathbb{P}(A_{n+1}|R_n = k) = \frac{k}{n+2}$, on calcule

$$\begin{aligned}\mathbb{E}(M_{n+1}|\mathcal{F}_n) &= \frac{R_n}{n+2} \cdot \frac{R_n+1}{n+3} + \left(1 - \frac{R_n}{n+2}\right) \frac{R_n}{n+3} \\ &= \frac{R_n}{(n+2)(n+3)} (R_n+1+n+2-R_n) = \frac{R_n(n+3)}{(n+2)(n+3)} = \frac{R_n}{n+2} = M_n.\end{aligned}$$

□

Par conséquence, comme M_n est positive, elle converge p.s. De plus, $M_n \in [0, 1]$ est bornée. D'après le théorème de convergence dominée, M_n converge dans L^1 . Soit M_∞ la limite de M_n , M_∞ est une variable aléatoire.

Théorème 3.5.2. 1. $\forall n \geq 0$, R_n est uniformément distribuée dans $\{1, \dots, n+1\}$.

2. M_∞ est uniformément distribuée sur $[0, 1]$.

Démonstration. 1. On fait la démonstration par récurrence : c'est vrai pour $n = 0$. Si c'est vrai pour n , on calcule $\mathbb{P}(R_{n+1} = k)$ pour $k \in \{1, \dots, n+2\}$. Si $k = 1$,

$$\mathbb{P}(R_{n+1} = 1) = \mathbb{P}(R_n = 1 \text{ et on a tiré une boule noire}) = \frac{1}{n+1} \left(1 - \frac{1}{n+2}\right) = \frac{1}{n+2}.$$

Si $k \geq 2$,

$$\begin{aligned}\mathbb{P}(R_{n+1} = k) &= \mathbb{P}(R_n = k \text{ et on a tiré une boule noire}) \\ &\quad + \mathbb{P}(R_n = k-1 \text{ et on a tiré une boule rouge}) \\ &= \frac{1}{n+1} \left(1 - \frac{k}{n+2}\right) + \frac{1}{n+1} \cdot \frac{k-1}{n+2} = \frac{1}{n+2}.\end{aligned}$$

2. On sait que $M_n = \frac{R_n}{n+2}$ est uniformément distribué sur $\{\frac{1}{n+2}, \dots, \frac{n+1}{n+2}\}$. Si $x \in [0, 1]$, on a

$$\mathbb{P}(M_n \leq x) = \frac{[(n+2)x]}{n+2} \xrightarrow{n \rightarrow \infty} x.$$

Donc M_n converge en loi vers la loi uniforme sur $[0, 1]$. Donc si M_n converge vers M_∞ p.s., M_∞ suit la loi uniforme sur $[0, 1]$.

□

Remarque 3.5.3. On peut prendre pour $n = 0$ une autre répartition des boules noires et rouges. $\frac{R_n}{n+B}$ est une martingale, avec B le nombre de boules à $n = 0$. Cette martingale est bornée donc elle converge p.s. vers M_∞ , mais M_∞ ne suis pas la loi uniforme sur $[0, 1]$.

Comme (M_n) est bornée, elle est bornée dans L^p pour $p \geq 1$. En particulier, on peut estimer la variation quadratique de (M_n) . D'abord on calcule $\mathbb{E}((M_{n+1} - M_n)^2 | \mathcal{F}_n)$. Soit A_{n+1} l'événement { on a tiré une boule rouge }, alors

$$\begin{aligned} \mathbb{E}\left((M_{n+1} - M_n)^2 1_{A_{n+1}} | \mathcal{F}_n\right) &= \mathbb{E}\left(\frac{(n+2)M_n + 1}{n+3} - M_n\right)^2 1_{A_{n+1}} | \mathcal{F}_n \\ &= \left(\frac{(n+2)M_n + 1}{n+3} - M_n\right)^2 \mathbb{P}(A_{n+1} | \mathcal{F}_n) \\ &= \left(\frac{(n+2)M_n + 1}{n+3} - M_n\right)^2 M_n. \end{aligned}$$

Et

$$\begin{aligned} \mathbb{E}((M_{n+1} - M_n)^2 1_{A_{n+1}^C} | \mathcal{F}_n) &= \mathbb{E}\left(\frac{(n+2)M_n}{n+3} - M_n\right)^2 1_{A_{n+1}^C} | \mathcal{F}_n \\ &= \frac{M_n^2}{(n+3)^2} \mathbb{P}(A_{n+1}^C | \mathcal{F}_n) = \frac{M_n^2}{(n+3)^2} (1 - M_n). \end{aligned}$$

Donc

$$\mathbb{E}((M_{n+1} - M_n)^2 | \mathcal{F}_n) = \frac{M_n}{(n+3)^2} (M_n(1 - M_n) + (M_n - 1)^2) = \frac{M_n(1 - M_n)}{(n+3)^2}.$$

Alors

$$\mathbb{E}((M_{n+1} - M_n)^2) = \mathbb{E}(\mathbb{E}((M_{n+1} - M_n)^2 | \mathcal{F}_n)) \leq \frac{1}{(n-3)^2}.$$

Comme les $(M_{n+1} - M_n)$ sont des variables aléatoires orthogonales, on conclut que

$$\mathbb{E}((M_\infty - M_n)^2) = \sum_{k \geq n} \mathbb{E}(M_{k+1} - M_k)^2 \leq \sum_{k \geq n} \frac{1}{(k+3)^2} < \frac{1}{n+2}.$$

On va ensuite montrer que (M_n) est fermée, i.e. il existe Z variable aléatoire t.q. $M_n = \mathbb{E}(Z | \mathcal{F}_n)$.

Soient maintenant V la variable aléatoire qui suit la loi uniforme sur $[0, 1]$, $(U_n)_{n \geq 1}$ variables aléatoires i.i.d. uniformes sur $[0, 1]$, indépendantes de V . Soit $X_n = 1_{\{U_n \leq V\}}$. On va voir que $(X_n)_{n \geq 1}$ a la même loi que $(1_{A_n})_{n \geq 1}$.

Idée : si on a beaucoup de X_i ($i \leq n$) qui valent 1 : alors on a beaucoup de $U_i \leq V$ ($i \leq n$). On peut en déduire que V est proche de 1, et donc U_{n+1} a une grande probabilité d'être inférieure à V , donc X_{n+1} a une grande probabilité de valoir 1.

On veut calculer

$$\mathbb{P}(X_{n+1} = 1 | X_1 = \epsilon_1, \dots, X_n = \epsilon_n) = \frac{\mathbb{P}(X_{n+1} = 1, X_1 = \epsilon_1, \dots, X_n = \epsilon_n)}{\mathbb{P}(X_1 = \epsilon_1, \dots, X_n = \epsilon_n)}$$

où les ϵ_i valent 0 ou 1. Si $V = x$,

$$\mathbb{P}(X_1 = \epsilon_1) = \begin{cases} \mathbb{P}(X_1 = 1) = \mathbb{P}(U_1 \leq V) = x & \text{si } \epsilon_1 = 1 \\ 1 - x & \text{si } \epsilon_1 = 0 \end{cases}$$

donc $\mathbb{P}(X_1 = \epsilon_1) = x^{\epsilon_1}(1-x)^{1-\epsilon_1}$, et

$$\mathbb{P}(X_1 = \epsilon_1, \dots, X_n = \epsilon_n) = \int_0^1 \prod_{i=1}^n x^{\epsilon_i}(1-x)^{1-\epsilon_i} dx = \int_0^1 x^{S_n}(1-x)^{n-S_n} dx,$$

où $S_n = x_1 + \dots + x_n$.

Proposition 3.5.4. Soient $k, l \in \mathbb{N}$, on a

$$\int_0^1 x^k(1-x)^l dx = \frac{k! l!}{(k+l+1)!}$$

Remarque 3.5.5. On peut faire une démonstration par récurrence. En prenant k, l réels > -1 , on obtient la fonction bêta \mathcal{B} qui a des relations avec la fonction Γ .

Revenons à la calcul, on obtient

$$\begin{aligned} \mathbb{P}(X_{n+1} = 1 | X_1 = \epsilon_1, \dots, X_n = \epsilon_n) &= \frac{\int_0^1 x^{S_n+1}(1-x)^{n-S_n} dx}{\int_0^1 x^{S_n}(1-x)^{n-S_n} dx} \\ &= \frac{(S_n+1)!(n-S_n)!}{(n+2)!} \cdot \frac{(n+1)!}{S_n!(n-S_n)!} = \frac{S_n+1}{n+2}. \end{aligned}$$

Or

$$\mathbb{P}(A_{n+1} | \mathcal{F}_n) = \frac{R_n}{n+2} = \frac{1 + \sum_{i=1}^n A_i}{n+2}$$

par l'urne de Pólya. On en déduit que la suite (X_n) a même loi que la suite 1_{A_n} . Donc $(\frac{S_n+1}{n+2})_{n \geq 1}$ a même loi que (M_n) . Soit $\frac{S_n+1}{n+2}$ converge p.s. vers T , où

$$S_n = 1_{\{U_1 \leq V\}} + \dots + 1_{\{U_n \leq V\}}.$$

Par la loi des grandes nombres : pour V fixé, $\frac{S_n}{n}$ converge p.s. vers V , donc $\frac{S_n+1}{n+2}$ converge

p.s. vers V . Donc pour $y \in [0, 1]$,

$$\begin{aligned} \mathbb{P}(V \leq y | X_1 = \epsilon_1, \dots, X_n = \epsilon_n) &= \frac{\int_0^y x^{S_n} (1-x)^{n-S_n} dx}{\int_0^1 x^{S_n} (1-x)^{n-S_n} dx} \\ &= \left(\int_0^y x^{S_n} (1-x)^{n-S_n} dx \right) \cdot \frac{(n+1)!}{S_n!(n-S_n)!}. \end{aligned}$$

Si on dérive par rapport à y , on trouve que la loi conditionnelle de V sachant (X_1, \dots, X_n) admet une densité qui est

$$y^{S_n} (1-y)^{n-S_n} \frac{(n+1)!}{S_n!(n-S_n)!}$$

(L'espérance conditionnelle est l'espérance pour la loi conditionnelle), donc

$$\begin{aligned} \mathbb{E}(V | \mathcal{F}_n) &= \int_0^1 y \cdot y^{S_n} (1-y)^{n-S_n} \frac{(n+1)!}{S_n!(n-S_n)!} dy \\ &= \frac{(S_n+1)!(n-S_n)!}{(n+2)!} \frac{(n+1)!}{S_n!(n-S_n)!} = \frac{S_n+1}{n+2} = M_n. \end{aligned}$$

Donc $M_n = \mathbb{E}(V | \mathcal{F}_n)$, M_n est une martingale fermée.

Remarque 3.5.6. On peut généraliser l'urne de Pólya : par exemple quand on tire une boule rouge, on ajoute a boules rouges, b boules noires et quand on tire une boule noire, on ajoute c boules rouges et d boules noires. En général on peut trouver une suite (c_n) telle que $c_n R_n$ est une martingale positive. (R_n est le nombre de boules rouges) La martingale converge p.s. vers M_∞ . La loi de M_∞ n'est pas uniforme en général.

3.6 Marches aléatoires et martingales exponentielles

Soient X_1, \dots, X_n, \dots des variables aléatoires i.i.d. à valeurs dans \mathbb{R} et dans L^1 . On suppose que $\mathbb{E}(X_1) = m$, posons $Y_n = X_n - m$ et $S_n = Y_1 + \dots + Y_n$, alors S_n est une martingale.

On suppose qu'il existe $t > 0$ t.q. $\mathbb{E}(e^{tY}) < \infty$, (vrai si Y_1 bornée), et pour $s \in [0, t]$, définissons

$$M_n(s) = \frac{e^{sS_n}}{\mathbb{E}(e^{sY_1})^n}.$$

Alors $M_n(s)$ est une martingale par rapport à filtration $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$, car

$$\begin{aligned} \mathbb{E}(e^{sS_{n+1}} | \mathcal{F}_n) &= \mathbb{E}(e^{s(S_n + Y_{n+1})} | \mathcal{F}_n) = \mathbb{E}(e^{sS_n} e^{sY_{n+1}} | \mathcal{F}_n) \\ &= e^{sS_n} \mathbb{E}(e^{sY_{n+1}} | \mathcal{F}_n) = e^{sS_n} \mathbb{E}(e^{sY_{n+1}}) = e^{sS_n} \mathbb{E}(e^{sY_1}). \end{aligned}$$

Application 3.6.1. Si $S_n = m + Y_1 + \dots + Y_n$ avec $\mathbb{P}(Y_1 = 1) = \mathbb{P}(Y_1 = -1) = \frac{1}{2}$, alors

$$\mathbb{E}(e^{sY_1}) = \frac{1}{2}(e^s + e^{-s}) = \cosh(s),$$

ici m est un entier > 0 . soit $T = T_0 = \inf\{n, S_n = 0\}$, alors $(M(s)_{T \wedge n})$ est une martingale,

$$M(s)_{T \wedge n} = \frac{e^{sS_{T \wedge n}}}{\cosh(s)^{T \wedge n}}.$$

Si $A > m$, soit $T_A = \inf\{n, S_n \in \{0\} \cup [A, \infty[\}$, alors

$$M(s)_{T_A \wedge n} = \frac{e^{sS_{T_A \wedge n}}}{\cosh(s)^{T_A \wedge n}}$$

est une martingale bornée (car $\cosh(s)^{T_A \wedge n} \geq 1$). On prend $s < 0$, alors $M(s)_{T_A \wedge n}$ converge p.s. vers $\frac{e^{sS_{T_A}}}{\cosh(s)^{T_A}}$, et son espérance est

$$\mathbb{E}\left(\frac{e^{sS_{T_A}}}{\cosh(s)^{T_A}}\right) = e^{sA}\mathbb{E}\left(\frac{1_{\{S_{T_A}=A\}}}{\cosh(s)^{T_A}}\right) + \mathbb{E}\left(\frac{1_{\{S_{T_A}=0\}}}{\cosh(s)^{T_A}}\right).$$

Au temps 0, la martingale M_0 vaut e^{sm} , donc on en déduit que

$$e^{sm} = \mathbb{E}(e^{sm}) = e^{sA}\mathbb{E}\left(\frac{1_{\{S_{T_A}=A\}}}{\cosh(s)^{T_A}}\right) + \mathbb{E}\left(\frac{1_{\{S_{T_A}=0\}}}{\cosh(s)^{T_A}}\right).$$

On fait $A \rightarrow \infty$, et on obtient que

$$\forall s < 0, \quad e^{sm} = \mathbb{E}\left(\frac{1}{\cosh(s)^T}\right).$$

Soit $v = \frac{1}{\cosh(s)} < 1$, on trouve que $\mathbb{E}(v^T) = f(v)^m$, où f satisfait $f\left(\frac{1}{\cosh(s)}\right) = e^s$, i.e.

$$f(v) = \frac{1}{v} - \sqrt{\frac{1}{v^2} - 1}.$$

Remarque 3.6.2. Feuille TD4, exercice 6 : ERREUR ! S_n n'est pas une filtration. Le théorème de De Finetti se démontre avec les martingales rétrogrades.

3.7 Martingales bornées dans L^p , $p > 1$

Lemme 3.7.1. *Soit (X_n) une sous-martingale, S, T deux temps d'arrêt bornés tels que $S \leq T$. Alors $\mathbb{E}[X_S] \leq \mathbb{E}[X_T]$.*

Démonstration. On définit (H_n) comme un processus prévisible par

$$H_n = 1_{\{S < n \leq T\}} = 1_{\{S \leq n-1\}} - 1_{\{T \leq n-1\}}.$$

Puisque S, T sont bornés, il existe N tel que on a p.s. $S \leq T \leq N$. Alors $(H.X)$ est une sous-martingale, donc $\mathbb{E}[(H.X)_N] \geq \mathbb{E}[(H.X)_0] = 0$. On en déduit que $\mathbb{E}[(H.X)_N] = \mathbb{E}[X_T - X_S] = \mathbb{E}[X_T] - \mathbb{E}[X_S] \geq 0$. \square

Théorème 3.7.2 (Inégalité maximale). *Soit (X_n) une sous martingale. Si $a > 0$ et $n \in \mathbb{N}$, on a*

$$a\mathbb{P}\left(\sup_{0 \leq k \leq n} X_k \geq a\right) \leq \mathbb{E}(X_n 1_{\{\sup_{0 \leq k \leq n} X_k \geq a\}}) \leq \mathbb{E}(X_n^+).$$

Remarque 3.7.3. On va utiliser l'inégalité de Markov : $a\mathbb{P}(X \geq a) \leq \mathbb{E}(X)$.

Démonstration. On définit $T = \inf\{n \geq 0, X_n \geq a\}$ un temps d'arrêt, et supposons

$$A = \left\{ \sup_{0 \leq k \leq n} X_k \geq a \right\} = \{T \leq n\}.$$

Puisque $T \wedge n \leq n$ sont deux temps d'arrêt, on a $\mathbb{E}[X_{T \wedge n}] \leq \mathbb{E}[X_n]$ par le lemme précédent. On note que $X_{T \wedge n} \geq X_n 1_{A^c} + a 1_A$. Alors

$$\mathbb{E}[X_n 1_{A^c}] + a\mathbb{P}(T \leq n) \leq \mathbb{E}[X_n],$$

et on conclut que

$$a\mathbb{P}\left(\sup_{0 \leq k \leq n} X_k \geq a\right) \leq \mathbb{E}[X_n 1_A] = \mathbb{E}[X_n 1_{\{\sup_{0 \leq k \leq n} X_k \geq a\}}] \leq \mathbb{E}[X_n 1_{X_n \geq 0}] = \mathbb{E}[X_n^+].$$

\square

Proposition 3.7.4. 1. *Soient $p > 1$ et X_n une sous martingale positive. On pose $\tilde{X}_n = \sup_{0 \leq k \leq n} X_k$. Alors $\forall n \geq 0$,*

$$\mathbb{E}[\tilde{X}_n^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[X_n^p].$$

2. Si (Y_n) une martingale, $Y_n^* = \sup_{0 \leq k \leq n} |Y_k|$, $\forall n \geq 0$. Alors on a

$$\mathbb{E}[(Y_n^*)^p] \leq \left(\frac{p}{p-1} \right)^p \mathbb{E}[|Y_n|^p].$$

Démonstration. (1) \Rightarrow (2) : On remarque que $x \mapsto |x|$ est une fonction convexe. Donc $(|Y_n|)$ est une sous martingale. Il suffit de montrer que (1) est vrai :

On peut supposer que $\mathbb{E}[X_n^p] < \infty$. Par le théorème précédent, on a

$$a^{p-1} \mathbb{P}(\tilde{X}_n \geq a) \leq \mathbb{E}[X_n 1_{\tilde{X}_n \geq a}] a^{p-2}.$$

De plus, par Fubini :

$$\int_0^\infty a^{p-1} \mathbb{P}(\tilde{X}_n \geq a) da = \mathbb{E} \left[\int_0^{\tilde{X}_n} a^{p-1} da \right] = \frac{1}{p} \mathbb{E}[\tilde{X}_n^p].$$

Et

$$\int_0^\infty a^{p-2} \mathbb{E}[X_n 1_{\tilde{X}_n \geq a}] da = \mathbb{E} \left[X_n \int_0^{\tilde{X}_n} a^{p-2} da \right] = \mathbb{E} \left[X_n \frac{\tilde{X}_n^{p-1}}{p-1} \right].$$

On utilise l'inégalité de Hölder :

$$\mathbb{E} \left[X_n \frac{\tilde{X}_n^{p-1}}{p-1} \right] \leq \frac{1}{p-1} (\mathbb{E}[X_n^p])^{\frac{1}{p}} (\mathbb{E}[\tilde{X}_n^p])^{\frac{p-1}{p}}.$$

Alors on conclut par

$$\frac{1}{p} \mathbb{E}[\tilde{X}_n^p] \leq \frac{1}{p-1} (\mathbb{E}[X_n^p])^{\frac{1}{p}} (\mathbb{E}[\tilde{X}_n^p])^{\frac{p-1}{p}}.$$

□

Théorème 3.7.5. On note $X_\infty^* = \sup_{n \in \mathbb{N}} |X_n|$. Soit (X_n) une sous martingale. Et on suppose $\exists p > 1$ tel que $\sup \mathbb{E}[|X_n|^p] < \infty$. Alors X_n converge p.s. vers X_∞ variable aléatoire telle que

$$\mathbb{E}[|X_\infty|^p] = \sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|^p]$$

et

$$\mathbb{E}[|X_\infty^*|^p] \leq \left(\frac{p}{p-1} \right)^p \mathbb{E}[|X_\infty|^p].$$

Démonstration. Puisque (X_n) est bornée dans L^p , alors il est bornée dans L^1 . Donc il converge

p.s. Et on note que

$$\mathbb{E}[\tilde{X}_n^p] \leq \left(\frac{p}{p-1}\right)^p \sup_{k \leq n} \mathbb{E}[|X_k|^p].$$

En prenant $n \rightarrow \infty$, on a

$$\mathbb{E}[|X_\infty^*|^p] \leq \left(\frac{p}{p-1}\right)^p \sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|^p] < \infty.$$

Donc $X_\infty^* \in L^p$ et les $|X_n|$ sont dominés par X_∞^* . Alors on a $X_n \xrightarrow{L^p} X_\infty$, et

$$\mathbb{E}[|X_\infty|^p] = \lim_{n \rightarrow \infty} \mathbb{E}[|X_n|^p] = \sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|^p].$$

La dernière assertion vient de l'inégalité de Jensen pour l'espérance conditionnelle : pour la fonction $x \mapsto x^p$, on a pour $k \leq n$:

$$\mathbb{E}[X_k^p] \leq \mathbb{E}[\mathbb{E}[X_n | \mathcal{F}_k]^p] \leq \mathbb{E}[\mathbb{E}[X_n^p | \mathcal{F}_k]] = \mathbb{E}[X_n^p].$$

□

3.8 Uniforme intégrabilité

Définition 3.8.1. Soit $(X_i)_{i \in I}$ une famille de variables aléatoires dans L^1 . On dit que $(X_i)_{i \in I}$ est uniformément intégrable si

$$\lim_{a \rightarrow \infty} \left(\sup_{i \in I} \mathbb{E}[|X_i| 1_{\{|X_i| > a\}}] \right) = 0.$$

Remarque 3.8.2. On notera u.i. pour u.i..

— *u.i.* \Rightarrow borné dans L^1 : $\exists a$ t.q. $\sup_{i \in I} \mathbb{E}(|X_i| 1_{|X_i| > a}) \leq 1$, alors

$$\mathbb{E}(|X_i|) \leq \mathbb{E}(|X_i| 1_{|X_i| \leq a}) + \mathbb{E}(|X_i| 1_{|X_i| > a}) \leq a + 1.$$

— Si I ensembles fini, $(X_i)_{i \in I}$ u.i..

— Si $Z \geq 0$, $Z \in L^1$, si $\forall i, |X_i| < Z$, alors $(X_i)_{i \in I}$ est u.i., car

$$\mathbb{E}[|X_i| 1_{|X_i| > a}] \leq \mathbb{E}[Z 1_{Z > a}].$$

— Soit $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, t.q. $\frac{\varphi(x)}{x} \xrightarrow{x \rightarrow \infty} +\infty$ et soit $C > 0$. Alors $\{X \in L^1, \mathbb{E}|\varphi(X)| \leq C\}$

est u.i.. En effet,

$$\mathbb{E}[|X|1_{|X|>a}] \leq \mathbb{E}[\varphi(|X|)] \sup_{x>a} \left(\frac{x}{\varphi(x)} \right).$$

[En particulier, soit $\varphi(x) = x^p$, $p > 1$, les familles de variable aléatoires bornées dans L^p sont u.i..]

— Contre-exemple : (X_n) v.a. uniforme sur $[n, n+1]$ n'est pas u.i..

Proposition 3.8.3. *Soit $(X_i)_{i \in I}$ une suite de variables aléatoires bornées dans L^1 . On a l'équivalence*

1. $(X_i)_{i \in I}$ u.i..
2. $\forall \epsilon > 0, \exists \delta > 0, \forall A \in \mathcal{F}$ tel que $\mathbb{P}(A) < \delta$, on a $\forall i \in I, \mathbb{E}[X_i 1_A] < \epsilon$.

Démonstration. (1) \Rightarrow (2) : Soit $\epsilon > 0$, il existe a tel que

$$\sup_{i \in I} \mathbb{E}[|X_i|1_{|X_i|>a}] \leq \frac{\epsilon}{2}.$$

Posons $\delta = \frac{\epsilon}{2a}$, alors si $\mathbb{P}(A) < \delta$, on a pour tout $i \in I$

$$\mathbb{E}[|X_i|1_A] = \mathbb{E}[|X_i|1_{A \cap \{|X_i|>a\}}] + \mathbb{E}[|X_i|1_{A \cap \{|X_i| \leq a\}}] \leq \frac{\epsilon}{2} + a\mathbb{P}(A) \leq \epsilon$$

(2) \Rightarrow (1) : Soit $C = \sup_{i \in I} \mathbb{E}|X_i|$. Alors

$$\forall a > 0, \forall i \in I, \mathbb{P}(|X_i| > a) \leq \frac{\mathbb{E}[|X_i|]}{a} \leq \frac{C}{a}.$$

Soit $\epsilon > 0$, et il existe δ t.q. (2) est vrai. Alors si a vérifie $\frac{C}{a} < \delta$, on a bien

$$\forall i \in I, \mathbb{E}[|X_i|1_{|X_i|>a}] < \epsilon.$$

□

Corollaire 3.8.4. *Soit X v.a. $\in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Alors les v.a. $\mathbb{E}[X|\mathcal{G}]$ où \mathcal{G} est une sous-tribu de \mathcal{F} sont uniformément intégrables.*

Démonstration. Comme (X_i) sont u.i.s, on a $\forall \epsilon, \exists \delta > 0$, t.q. $\forall A \in \mathcal{F}, \mathbb{P}(A) < \delta$, on a $\mathbb{E}[|X|1_A] \leq \epsilon$. De plus, par l'inégalité de Markov :

$$\mathbb{P}(\mathbb{E}[X|\mathcal{G}] > a) \leq \frac{1}{a} \mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \frac{\mathbb{E}[|X|]}{a}.$$

Donc si $a > \frac{\mathbb{E}[|X|]}{\delta}$,

$$\mathbb{E}\left[|\mathbb{E}[X|\mathcal{G}]|1_{\mathbb{E}[X|\mathcal{G}]>a}\right] \leq \mathbb{E}\left[\mathbb{E}[|X||\mathcal{G}]1_{\mathbb{E}[X|\mathcal{G}]>a}\right] \leq \mathbb{E}[|X|1_{\mathbb{E}[X|\mathcal{G}]>a}] < \epsilon.$$

□

Théorème 3.8.5. *Soit X_n une suite de v.a L^1 telle que $(X_n) \xrightarrow{(p)} X_\infty$, alors il y a l'équivalence entre*

1. $X_n \xrightarrow{L^1} X_\infty$.
2. (X_n) u.i..

Démonstration. (1) \Rightarrow (2) : soit $\epsilon > 0$, il existe N , t.q. $\forall n > N$, $\mathbb{E}[|X_n - X_N|] \leq \frac{\epsilon}{2}$. Comme $\{X_0, X_1, \dots, X_N\}$ est u.i.. Alors

$$\exists \delta > 0, \forall A \in \mathcal{F}, \mathbb{P}(A) < \delta \quad \Rightarrow \quad \forall n \leq N, \mathbb{E}[|X_n|1_A] < \frac{\epsilon}{2}.$$

Et pour $n > N$, $\mathbb{E}[|X_n|1_A] = \mathbb{E}[|X_N|1_A] + \mathbb{E}[|X_N - X_n|1_A] \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$.

(2) \Rightarrow (1) : Si $(X_n)_{n \in \mathbb{N}}$ est u.i., alors $(X_m - X_n)_{(m,n) \in \mathbb{N}^2}$ est aussi u.i..

Alors $\forall \epsilon > 0$, il existe a , t.q. pour tout $m, n \in \mathbb{N}$, $\mathbb{E}[|X_n - X_m|1_{|X_n - X_m|>a}] < \epsilon$. Donc on a

$$\begin{aligned} \mathbb{E}[|X_n - X_m|] &\leq \mathbb{E}[|X_n - X_m|1_{|X_n - X_m| \leq \epsilon}] + \mathbb{E}[|X_n - X_m|1_{\epsilon < |X_n - X_m| \leq a}] \\ &\quad + \mathbb{E}[|X_n - X_m|1_{|X_n - X_m| > a}] \\ &\leq 2\epsilon + \mathbb{E}[|X_n - X_m|1_{\epsilon < |X_n - X_m| \leq a}] \\ &\leq 2\epsilon + a\mathbb{P}(|X_n - X_m| > \epsilon). \end{aligned}$$

Comme X_n converge vers X_∞ en probabilité, et notons que

$$\mathbb{P}(|X_n - X_m| > \epsilon) \leq \mathbb{P}(|X_n - X_\infty| > \frac{\epsilon}{2}) + \mathbb{P}(|X_n - X_\infty| > \frac{\epsilon}{2}).$$

On a $\mathbb{P}(|X_n - X_m| > \frac{\epsilon}{2}) < \epsilon$ si $n, m > N$. Donc dans ce cas, on a $\mathbb{E}[|X_m - X_n|] < 2\epsilon + 2\epsilon = 4\epsilon$, et (X_n) est une suite de Cauchy dans L^1 . □

Proposition 3.8.6 (Application aux martingales). *Si (X_n) une martingale, on a l'équivalence*

1. X_n converge. p.s. et dans L^1 vers X_∞ .
2. (X_n) est u.i..

3. (X_n) est une martingale fermée : il existe Z v.a. dans L^1 t.q. $X_n = \mathbb{E}[Z|\mathcal{F}_n]$.

On rappelle que une martingale est converge dans L^1 s.i.i. elle est u.i.. Et on a vu que si la martingale est bornée dans L^p ($p > 1$), alors elle est u.i.. Voici quelque exemple qui n'est pas u.i..

Contre-exemple :

1. Soit $S_n = 1 + X_1 + \dots + X_n$ où (X_i) variables aléatoires i.i.d. telle que

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}.$$

Alors (S_n) est une martingale. Soit $T_0 = \inf(k, S_k = 0)$ est un temps d'arrêt, on a $M_n = S_{T_0 \wedge n}$ martingale positive. M_n converge p.s. vers $M_\infty = 0$, mais $\mathbb{E}(M_0) = 1 \neq 0 = \mathbb{E}(M_\infty)$ donc M_n ne converge pas vers M_∞ dans L^1 , donc (M_n) n'est pas u.i.

2. Branchement critique : Soit (Z_n) processus de Galton-Watson :

$$Z_0 = 1, \quad Z_{n+1} = X_1^{(n)} + \dots + X_{Z_n}^{(n)}$$

où les $(X_i^{(n)})_{n,i \in \mathbb{N}}$ sont i.i.d.

$$X_i^{(n)} = \{\text{nombre d'enfants du } i\text{-ème individu à la génération } n\}.$$

Soit $m = \mathbb{E}(Z_1)$, alors $(\frac{Z_n}{m^n})$ est une martingale positive qui converge presque sûrement vers 0. Mais pour tout n , $\mathbb{E}(Z_n) = \mathbb{E}(Z_0) = 1$, donc il n'y a pas de convergence dans L^1 , et elle n'est pas u.i..

Théorème 3.8.7. Soit (X_n) une martingale u.i.. Alors pour tout temps d'arrêt T (qui peut être infini), on a $X_T = \mathbb{E}(X_\infty|\mathcal{F}_T)$. En particulier, $\mathbb{E}(X_T) = \mathbb{E}(X_\infty) = \mathbb{E}(X_n)$ pour tout n . De plus, si S et T sont deux temps d'arrêt avec $S \leq T$ p.s. alors $X_S = \mathbb{E}(X_T|\mathcal{F}_S)$.

Démonstration. On montre d'abord que $X_T \in L^1$:

$$\begin{aligned}
\mathbb{E}(|X_T|) &= \sum_{n=0}^{\infty} \mathbb{E}(|X_n|1_{\{T=n\}}) + \mathbb{E}(|X_{\infty}|1_{\{T=\infty\}}) \\
&= \sum_{n=0}^{\infty} \mathbb{E}\left(\mathbb{E}(|X_n|1_{\{T=n\}}|\mathcal{F}_n)\right) + \mathbb{E}(|X_{\infty}|1_{\{T=\infty\}}) \\
&\leq \sum_{n=0}^{\infty} \mathbb{E}(1_{\{T=n\}}\mathbb{E}(|X_{\infty}||\mathcal{F}_n)) + \mathbb{E}(|X_{\infty}|1_{\{T=\infty\}}) \\
&= \sum_{n=0}^{\infty} \mathbb{E}(1_{T=n}\mathbb{E}(|X_{\infty}|) + \mathbb{E}(|X_{\infty}|1_{T=\infty})) \\
&= \mathbb{E}(|X_{\infty}|)
\end{aligned}$$

Prenons $A \in \mathcal{F}_T$, (notons $\tilde{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$), on a

$$\mathbb{E}(X_T 1_A) = \mathbb{E}\left(\sum_{n \in \tilde{\mathbb{N}}} 1_{\{T=n\}} 1_A X_T\right).$$

Comme $X_T \in L^1$, on peut intervertir l'ordre de l'intégration :

$$\begin{aligned}
\mathbb{E}(X_T 1_A) &= \sum_{n \in \tilde{\mathbb{N}}} \mathbb{E}(1_{\{T=n\}} 1_A X_T) = \sum_{n \in \tilde{\mathbb{N}}} \mathbb{E}(1_A 1_{\{T=n\}} X_n) \\
&= \sum_{n \in \tilde{\mathbb{N}}} \mathbb{E}(1_A 1_{\{T=n\}} \mathbb{E}(X_{\infty}|\mathcal{F}_n)) = \sum_{n \in \tilde{\mathbb{N}}} \mathbb{E}(1_{A \cap \{T=n\}} \mathbb{E}(X_{\infty}|\mathcal{F}_n)) \\
&= \sum_{n \in \tilde{\mathbb{N}}} \mathbb{E}(\mathbb{E}(X_{\infty} 1_{A \cap \{T=n\}}|\mathcal{F}_n)) \\
&= \mathbb{E}(\mathbb{E}(X_{\infty} 1_A|\mathcal{F}_T)) = \mathbb{E}(X_{\infty} 1_A)
\end{aligned}$$

Donc l'égalité est vrai pour tous les $A \in \mathcal{F}_T$, c'est vrai si on remplace 1_A par n'importe quelle variable aléatoire bornée mesurable par rapport à \mathcal{F}_T , donc on a $X_T = \mathbb{E}(X_{\infty}|\mathcal{F}_T)$.

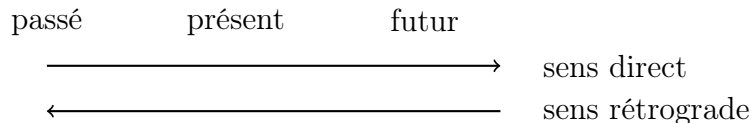
Si $S \leq T$, alors $\mathcal{F}_S \subset \mathcal{F}_T$, donc

$$X_S = \mathbb{E}(X_{\infty}|\mathcal{F}_S) = \mathbb{E}(\mathbb{E}(X_{\infty}|\mathcal{F}_T)|\mathcal{F}_S) = \mathbb{E}(X_T|\mathcal{F}_S)$$

□

3.9 Martingales rétrogrades

On dit aussi martingales inverses (en anglais : “inverse martingale”, “backward martingale”).



Définition 3.9.1. $(\mathcal{F}_n)_{n \in -\mathbb{N}}$ est une filtration rétrograde si pour tout $n \in \mathbb{N}$, \mathcal{F}_n est une tribu et si pour tout $m, n \in -\mathbb{N}$, $n \leq m$ implique $\mathcal{F}_n \subset \mathcal{F}_m$.

Définition 3.9.2. (M_n) est une martingale rétrograde si $\forall n \in -\mathbb{N}$, $\mathbb{E}(|M_n|) < \infty$ et si pour tout $m, n \in -\mathbb{N}$, $n \leq m$ implique $X_n = \mathbb{E}(X_m | \mathcal{F}_n)$. (On peut ainsi définir des sous-martingales et des surmartingales rétrogrades.)

Théorème 3.9.3. Soit $(M_n)_{n \in -\mathbb{N}}$ une martingale rétrograde. Alors (M_n) est u.i. et converge p.s. et dans L^1 vers M_∞ . Et pour tout n , $\mathbb{E}(X_n | \mathcal{F}_{-\infty}) = X_\infty$, où $\mathcal{F}_{-\infty} = \bigcap_{n \in -\mathbb{N}} \mathcal{F}_n$.

Remarque 3.9.4. Pour les martingales positives, on montre la convergence p.s. en montrant que p.s. pour tout $p < q$ rationnels, on traverse $[p, q]$ un nombre fini de fois. (On utilise une inégalité sur les nombres de montrés.)

Dans les notes de cours, on a montré qu'on a un nombre fini de montrés dans le cas où la martingale est bornée dans L^2 . Dans le cours de Le Gall, la démonstration est faite dans le cas des martingales positive (inégalité de Doob). Pour les martingales rétrogrades, on utilise les mêmes arguments, on démontre ainsi la convergence p.s.

Démonstration de l'uniforme intégrabilité. On a vu que si $X \in L^1$, la famille $(\mathbb{E}(X | \mathcal{G}))_{\mathcal{G} \subset \mathcal{F}}$ indexée par les sous-tribus de \mathcal{F} est u.i.. Comme $X_n = \mathbb{E}(X_0 | \mathcal{F}_n)$, on a bien l'uniforme intégrabilité. □

Lemme 3.9.5. Soient Z variable aléatoire dans L^1 , H_1, H_2 sous-tribus, H_2 indépendante de $\sigma(\sigma(Z), H_1)$. Alors

$$\mathbb{E}(Z | \sigma(H_1, H_2)) = \mathbb{E}(Z | \sigma(H_1)).$$

Démonstration. Soit $A \in \sigma(H_1, H_2)$, on regarde $\mathbb{E}(1_A Z)$, on commence par prendre A de la forme $B \cap C$, avec $B \in H_1$, $C \in H_2$, alors

$$\mathbb{E}(1_A Z) = \mathbb{E}(1_B 1_C Z) = \mathbb{E}(1_B 1_C \mathbb{E}(Z | H_1)).$$

Ensuite on utilise le théorème de classe monotone. □

Théorème 3.9.6 (Loi des grands nombres, version L^1). Soient X_1, \dots, X_n, \dots des variables aléatoires i.i.d. dans L^1 , on pose $m = \mathbb{E}(X_1)$. Alors

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[L^1]{p.s.} m.$$

Démonstration. Soit $S_n = X_1 + \dots + X_n$, on a $\mathbb{E}(X_1|S_n) = \mathbb{E}(X_k|S_n)$ pour tout $k \in \{1, 2, \dots, n\}$ car les X_i sont i.i.d. et S_n invariante par permutation :

$$S_n = X_{\sigma(1)} + X_{\sigma(2)} + \dots + X_{\sigma(n)} \quad \text{pour tout } \sigma \in \mathfrak{S}_n.$$

Donc on a

$$n\mathbb{E}(X_1|S_n) = \mathbb{E}(X_1|S_n) + \dots + \mathbb{E}(X_n|S_n) = \mathbb{E}(S_n|S_n) = S_n$$

donc $\mathbb{E}(X_1|S_n) = \frac{S_n}{n}$. Soit $H_1 = \sigma(S_n)$, $H_2 = \sigma(X_{n+1}, \dots)$, alors par le lemme

$$\mathbb{E}(X_1|S_n, X_{n+1}, \dots) = \frac{S_n}{n}.$$

Soit $\mathcal{F}_{-n} = \sigma(S_n, X_{n+1}, \dots)$, $M_{-n} = \frac{S_n}{n}$, alors (M_{-n}) est une martingale rétrograde, donc elle converge p.s. et dans L^1 vers M_∞ . Soit $\mathcal{F}_\infty = \bigcap_{n \in \mathbb{N}} \mathcal{F}_{-n}$, d'après la loi du 0-1 de Kolmogorov, $\mathcal{F}_\infty = \{\emptyset, \Omega\}$ et M_∞ est mesurable par rapport à \mathcal{F}_∞ , donc M_∞ est constante p.s. On a

$$M_n \xrightarrow{L^1} M_\infty, \quad \mathbb{E}(M_{-n}) = \mathbb{E}(M_\infty) = M_\infty = m.$$

□

Autre application : variables aléatoires échangeable (invariance par permutation), Théorème de De Finetti, Loi du 0-1 de Hewitt-Savage.

Proposition 3.9.7 (Hewitt-Savage). Soient $X_1, X_2, \dots, X_n, \dots$ variables aléatoires i.i.d. sur (E, \mathcal{E}) . F fonction symétrique sur $E^{\mathbb{N}^*}$, i.e. pour tout σ permutation de \mathbb{N}^* ayant un support fini (nombre fini d'entiers n t.q. $\sigma(n) \neq n$). Alors $F(X_1, \dots, X_n, \dots)$ est constante p.s.

Démonstration. On suppose F borné, soient $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ et $\mathcal{G}_n = \sigma(X_{n+1}, \dots)$, on pose $Y = F(X_1, X_2, \dots)$ et

$$U_n = \mathbb{E}(Y|\mathcal{F}_n), \quad Z_n = \mathbb{E}(Y|\mathcal{G}_n).$$

D'après des résultats déjà vus,

$$U_n \xrightarrow[L^1]{p.s.} \mathbb{E}(Y|\mathcal{F}_\infty) = Y,$$

$$Z_n \xrightarrow[L^1]{p.s.} \mathbb{E}(Y|\mathcal{G}_\infty) = \mathbb{E}(Y).$$

(Car G_∞ est trivial par la loi du 0-1 de Kolmogorov.)

Donc pour n assez grand, $\mathbb{E}(|U_n - Y|) < \epsilon$, $\mathbb{E}(|Z_n - \mathbb{E}(Y)|) < \epsilon$. Comme U_n est mesurable par rapport à \mathcal{F}_n , il existe $g : E^n \rightarrow \mathbb{R}$, $U_n = g(X_1, \dots, X_n)$, alors

$$\mathbb{E}(|F(X_1, \dots, X_n, \dots) - g(X_1, \dots, X_n)|) < \epsilon.$$

Comme F est invariante par permutation :

$$\mathbb{E}(|F(X_{n+1}, \dots, X_{2n}, X_1, \dots, X_n, X_{2n+1}, \dots) - g(X_1, \dots, X_n)|) < \epsilon.$$

De plus (X_i) est i.i.d. alors

$$\mathbb{E}(|F(X_1, \dots, X_n, X_{n+1}, \dots, X_{2n}, X_{2n+1}, \dots) - g(X_{n+1}, \dots, X_{2n})|) < \epsilon.$$

Donc

$$\mathbb{E}(|Y - g(X_{n+1}, \dots, X_{2n})|) < \epsilon.$$

On en déduit que

$$\mathbb{E}(|\mathbb{E}(Y|\mathcal{G}_n) - \mathbb{E}(g(X_{n+1}, \dots, X_{2n})|\mathcal{G}_n)|) < \epsilon,$$

$$i.e. \quad \mathbb{E}(|Z_n - g(X_{n+1}, \dots, X_{2n})|) < \epsilon.$$

Donc $\mathbb{E}(|Y - \mathbb{E}(Y)|) < 3\epsilon$ est vrai pour tout $\epsilon > 0$, alors p.s. $Y = \mathbb{E}(Y)$. □

Exemple 3.9.8. Soient X_1, \dots, X_n i.i.d. à valeurs dans \mathbb{R}^d , $S_n = X_1 + \dots + X_n$. Pour B borélien, soit

$$F(X_1, \dots, X_n, \dots) = 1_{\{card\{n, X_n \in B\} = \infty\}}.$$

Alors F est symétrique : si σ permutation de \mathbb{N}^* à support fini : il existe N t.q. $\forall n \geq N$, $\sigma(n) = n$ et $S_n = S_{\sigma(n)}$.

Chapitre 4

Chaînes de Markov

4.1 Définition et premières propriétés

Idée : on se déplace sur un espace E : $(X_n)_{n \in \mathbb{N}}$ processus à valeurs dans E . Au temps n , on choisit au hasard le prochain point qu'on va visiter en fonction de l'endroit où on est au temps n mais pas de fonction de X_0, \dots, X_{n-1} . On suppose de plus que l'ensemble des temps est discret, i.e. $n \in \mathbb{N}$. On peut considérer des chaîne de Markov à temps continue : $(X_t)_{t \in \mathbb{R}}$, mais pas dans ce cours.

L'espace d'états E peut être fini, infini dénombrable (par exemple \mathbb{Z}^d), infini non dénombrable (par exemple \mathbb{R}^d). Dans ce cours, E est fini ou infini dénombrable :

$$\forall n \in \mathbb{N}, \forall x \in E, \quad \mathbb{P}(X_n = x) = p_n(x) \geq 0, \quad \sum_{x \in E} p_n(x) = 1.$$

Définition 4.1.1. Soit E un ensemble dénombrable dont la tribu est $\mathcal{P}(E)$. Une matrice stochastique sur E est une famille Q indexée par $E \times E$ à valeurs dans $[0, 1]$ telle que pour tout $x \in E$, $\sum_{y \in E} Q(x, y) = 1$.

Remarque 4.1.2. Dans le cours au l'espérance conditionnelle, on avait défini la notion de probabilité de transition : E, F espaces mesurables, $\nu : E \times \mathcal{F} \rightarrow [0, 1]$ t.q.

1. Pour tout $x \in E$, $\nu(x, \cdot)$ mesure de probabilité sur \mathcal{F} .
2. Pour tout $A \in \mathcal{F}$, ensemble mesurable de F , $x \mapsto \nu(x, A)$ est mesurable.

Dans le cas particulier $E = F$ dénombrable, ces notions sont les mêmes : pour Q donnée, $\nu(x, \cdot)$ est la probabilité donnée par $\nu(x, A) = \sum_{y \in A} Q(x, y)$, et pour ν donnée, $Q(x, y) = \nu(x, \{y\})$.

Définition 4.1.3. Soit Q une matrice stochastique. On dit que $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov sur E de matrice de transition Q si

$$\forall n \in \mathbb{N}, \quad \mathbb{P}(X_{n+1} = z | X_n = y) = Q(y, z)$$

et de plus, $\forall x_0, x_1, \dots, x_{n-1} \in E$,

$$\mathbb{P}(X_{n+1} = z | X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = y) = Q(y, z).$$

Remarque 4.1.4. Si E est fini, $|E| = n$, une matrice stochastique sur E est une matrice $n \times n$, à coefficients dans $[0, 1)$, et telle que $\sum_{y \in Y} Q(x, y) = 1, \forall x \in E$. On a alors $Q^{n+1}(x, z) = \sum_{y \in E} Q^n(x, y)Q(y, z)$. D'après la définition d'une chaîne de Markov, si (X_n) est une chaîne de Markov de matrice Q , $\mathbb{P}(X_n = y | X_0 = x) = Q^n(x, y)$. En effet, on peut faire une récurrence sur n : c'est vrai pour $n = 0, 1$, et

$$\begin{aligned} & \mathbb{P}(X_{n+1} = y | X_0 = x) \\ &= \sum_{y_1, \dots, y_n} \mathbb{P}(X_{n+1} = y | X_0 = x, X_1 = y_1, \dots, X_n = y_n) \mathbb{P}(X_0 = x, X_1 = y_1, \dots, X_n = y_n) \\ &= \sum_{y_1, \dots, y_n} \mathbb{P}(X_{n+1} = y | X_n = y_n) \mathbb{P}(X_0 = x, X_1 = y_1, \dots, X_n = y_n) \\ &= \sum_{y_n} Q(y_n, y) \sum_{y_1, \dots, y_{n-1}} \mathbb{P}(X_n = y_n | X_0 = x) \mathbb{P}(X_1 = y_1, \dots, X_{n-1} = y_{n-1}) \\ &= \sum_{y_n} Q(y_n, y) Q^n(x, y_n) = Q^{n+1}(x, y). \end{aligned}$$

Si E est infini dénombrable, on peut définir par récurrence

$$Q_1 = Q, \quad Q_{n+1}(x, y) = \sum_{y'} Q_n(x, y') Q(y', y)$$

et on a de même :

$$\mathbb{P}(X_{n+1} = y | X_0 = x) = Q_{n+1}(x, y).$$

Exemple 4.1.5 (Marche aléatoire). Si X_1, \dots, X_n, \dots i.i.d. à valeurs dans \mathbb{Z} , alors $S_0 = 0, S_1 = X_1, S_n = X_1 + \dots + X_n$ est une chaîne de Markov à valeurs dans \mathbb{Z} . En effet,

$$\begin{aligned} \mathbb{P}(S_{n+1} = y | S_0 = 0, S_1 = x_1, \dots, S_n = x_n) &= \mathbb{P}(S_{n+1} - S_n = y - x_n | S_0 = 0, \dots, S_n = x_n) \\ &= \mathbb{P}(X_{n+1} = y - x_n | S_0 = 0, \dots, S_n = x_n) \\ &= \mathbb{P}(X_{n+1} = y - x_n | S_n = x_n). \end{aligned}$$

Plus généralement, soit (X_n) i.i.d. à valeurs dans \mathbb{Z}^d , $S_n = X_1 + \dots + X_n$, alors S_n est une chaîne de Markov. Si X_n i.i.d. à valeurs dans un groupe G (pas forcément commutatif), $S_n = X_1 X_2 \dots X_n$, alors S_n est une chaîne de Markov :

$$\mathbb{P}(S_{n+1} = y | S_n = x) = \mathbb{P}(X_{n+1} = yx^{-1} | S_n = x) = \mathbb{P}(X_{n+1} = yx^{-1}).$$

On peut aussi considérer le marche aléatoire sur un graphe $G = (S, A)$: S ensembles des sommets du graphe, A ensemble des arêtes (G dénombrable). On prend la matrice de transition Q définie par

$$Q(x, y) = \begin{cases} 0 & \text{s'il n'y pas d'arête entre } x, y \\ \frac{1}{\text{degré}(x)} & \text{s'il y a une arête entre } x, y \end{cases}$$

Ici le degré de x est le nombre d'arrêtes incidentes à x .

Remarque 4.1.6. Le marche aléatoire dans le graphe \mathbb{Z}^2 est le marche aléatoire avec

$$\mathbb{P}(X_1 = (1, 0)) = \mathbb{P}(X_1 = (0, 1)) = \mathbb{P}(X_1 = (0, -1)) = \mathbb{P}(X_1 = (-1, 0)).$$

Exemple 4.1.7 (Processus de Galton-Watson). Soit $Z_0 = 1$, $Z_{n+1} = X_1^{(n+1)} + \dots + X_{Z_n}^{(n+1)}$. On vérifie que Z_n est une chaîne de Markov :

$$\begin{aligned} \mathbb{P}(Z_{n+1} = y | Z_0 = 1, Z_1 = x_1, \dots, Z_n = x_n) &= \mathbb{P}(Z_{n+1} = y | Z_n = x_n) \\ &= \mathbb{P}(X_1^{(n+1)} + \dots + X_{x_n}^{(n+1)} = y). \end{aligned}$$

Exemple 4.1.8 (Contre-exemple). Soit (X_n) i.i.d. et $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$. Supposons

$$Y_0 = 0, Y_1 = X_1, \quad Y_{n+1} = Y_n + X_n 1_{\{Y_n > Y_{n-1}\}} + 2X_n 1_{\{Y_n < Y_{n-1}\}}.$$

Alors (Y_n) n'est pas une chaîne de Markov. La loi de Y_{n+1} conditionnellement à Y_n a un support de cardinal 4 :

$$Y_{n+1} \in \{Y_n + 1, Y_n + 2, Y_n - 1, Y_n - 2\}.$$

La loi de Y_{n+1} conditionnellement à Y_n, Y_{n-1} a un support de cardinal 2 :

$$\begin{cases} Y_{n+1} \in \{Y_n + 1, Y_n - 1\} & \text{si } Y_n > Y_{n-1} \\ Y_{n+1} \in \{Y_n + 2, Y_n - 2\} & \text{si } Y_n < Y_{n-1}. \end{cases}$$

Remarque 4.1.9. Mais (Y_n) est une martingale par rapport à $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Car

$$\begin{aligned}\mathbb{E}(Y_{n+1}|\mathcal{F}_n) &= \mathbb{E}(X_{n+1}1_{\{Y_n > Y_{n-1}\}} + 2X_{n+1}1_{\{Y_n < Y_{n-1}\}}|\mathcal{F}_n) + Y_n \\ &= Y_n + 1_{\{Y_n > Y_{n-1}\}}\mathbb{E}(X_{n+1}|\mathcal{F}_n) + 1_{\{Y_n < Y_{n-1}\}}2\mathbb{E}(X_{n+1}|\mathcal{F}_n) \\ &= Y_n\end{aligned}$$

Proposition 4.1.10. *Soient E dénombrable, Q une matrice stochastique sur E . Alors il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ tel que $\forall x \in E$, on peut définir un processus $(X_n^x)_{n \in \mathbb{N}}$ qui est une chaîne de Markov de matrice de transition Q avec $X_0 = x$.*

Démonstration. Idée : à partir d'une variable aléatoire Y qui suit la loi uniforme sur $[0, 1]$, on peut construire une famille i.i.d. de variables aléatoires uniformes sur $[0, 1]$, $(Y_n)_{n \in \mathbb{N}}$. A partir de (Y_n) , on peut construire (X_n^x) .

1. Soit Y variable aléatoire uniforme sur $[0, 1]$, à partir du développement dyadique propre $Y = \sum_{i \geq 1} \frac{z_i}{2^i}$, on peut construire une suite de v.a. $z_i \in \{0, 1\}$, z_i i.i.d. et $\mathbb{P}(z_i = 0) = \mathbb{P}(z_i = 1) = \frac{1}{2}$. Comme $\mathbb{N}^* \times \mathbb{N}^*$ est dénombrable, il existe φ bijection $\mathbb{N}^* \rightarrow \mathbb{N}^* \times \mathbb{N}^*$. On pose

$$Y_i = \sum_{n \geq 1} \frac{z_{\varphi^{-1}(n,i)}}{2^i}.$$

Alors pour tout i , Y_i est uniforme sur $[0, 1]$ et (Y_n) indépendantes.

2. On construit $(X_n^x)_{n \in \mathbb{N}}$ à partir de (Y_n) . Comme E est dénombrable, on se ramène au cas $E = \mathbb{N}$. Conditionnellement à $X_n = y$, on définit $X_{n+1} = z$ si

$$\sum_{z' < z} Q(y, z') < Y_{n+1} \leq \sum_{z' \leq z} Q(y, z').$$

Comme

$$\{X_{n+1} = z, X_n = y\} = \left\{ Y_{n+1} \in \left[\sum_{z' < z} Q(y, z'), \sum_{z' \leq z} Q(y, z') \right], X_n = y \right\},$$

alors (on note que les Y_i sont indépendantes)

$$\begin{aligned}\mathbb{P}(X_{n+1} = z | X_n = y) &= \mathbb{P}\left(Y_{n+1} \in \left[\sum_{z' < z} Q(y, z'), \sum_{z' \leq z} Q(y, z') \right] \right) \\ &= \sum_{z' \leq z} Q(y, z') - \sum_{z' < z} Q(y, z') = Q(y, z).\end{aligned}$$

Comme les variables aléatoires (Y_n) sont indépendantes, on a

$$\mathbb{P}(X_{n+1} = z | X_0 = x, X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_{n+1} = z | X_n = x).$$

□

Théorème 4.1.11. 1. Soient E dénombrable et Q matrice stochastique sur E . Alors $\forall x \in E$, il existe une unique probabilité \mathbb{P}_x à valeurs dans $E^{\mathbb{N}}$ t.q. sous \mathbb{P}_x , $(X_n)_{n \in \mathbb{N}}$ soit une chaîne de Markov de matrice de transition Q et t.q. $\mathbb{P}(X_0 = x) = 1$.

2. Soit μ est une loi de probabilité sur E , alors il existe une unique loi de probabilité \mathbb{P}_μ t.q. sous \mathbb{P}_μ , $(X_n)_{n \in \mathbb{N}}$ soit une chaîne de Markov de matrice de transition Q et t.q. $\forall x \in E$, $\mathbb{P}(X_0 = x) = \mu(x)$.

Démonstration. L'existence a été obtenu par la proposition ci-dessus. Montrons l'unicité : la loi de \mathbb{P}_x est déterminé par sa valeurs sur les cylindriques A (de la forme $A = \{X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\}$). Et notons que les événements cylindriques engendrent la tribu par le lemme des classes monotones. Pour (2), on définit $\mathbb{P}_\mu = \sum_{x \in E} \mu(x) \mathbb{P}_x$. □

Opérateur de translation Soient $k \in \mathbb{N}$, on définit

$$\theta_k : E^{\mathbb{N}} \rightarrow E^{\mathbb{N}}, (w_0, w_1, \dots, w_n, \dots) \mapsto (w_k, w_{k+1}, \dots, w_{k+n}, \dots).$$

Théorème 4.1.12 (Propriété de Markov simple). Soient F, G fonctions mesurables positives sur $E^{\mathbb{N}}$. On suppose F mesurable par rapport à \mathcal{F}_n , où \mathcal{F}_n est la tribu engendrée par la fonction $(w_0, \dots, w_n, w_{n+1}, \dots) \mapsto (w_0, \dots, w_n)$. Alors

$$\forall x \in E, \mathbb{E}_x(F \cdot G(\theta_n)) = \mathbb{E}_x(F \mathbb{E}_{X_n}(G)) \quad \text{et} \quad \mathbb{E}_x(G(\theta_n) | \mathcal{F}_n) = \mathbb{E}_{X_n}(G).$$

Ici la notion \mathbb{E}_x désigner l'espérance sous la probabilité \mathbb{P}_x .

Démonstration. On peut prendre $F = 1_{\{X_0=x_0, \dots, X_n=x_n\}}$. Si G est de la même forme $G = 1_{\{X_0=y_0, \dots, X_p=y_p\}}$, alors

$$\forall y \in G, \mathbb{E}_y(G) = 1_{\{Y_0=y\}} Q(y_0, y_1) Q(y_1, y_2) \cdots Q(y_{p-1}, y_p).$$

Par ailleurs,

$$\begin{aligned}\mathbb{E}_x(F \cdot G(\theta_n)) &= \mathbb{P}_x(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n, X_{n+1} = y_0, \dots, X_{n+p} = y_p) \\ &= 1_{\{X_0=x\}} Q(x_0, x_1) \cdots Q(x_{n-1}, x_n) 1_{\{X_n=y_0\}} Q(y_0, y_1) \cdots Q(y_{p-1}, y_p).\end{aligned}$$

On a vérifié pour G de la forme $G = 1_{\{X_0=y_0, \dots, X_p=y_p\}}$, et donc c'est vrai pour toute fonction G (argument de classe monotone). \square

Théorème 4.1.13 (Propriété de Markov forte). *Soient T temps d'arrêt, F, G fonctions positives. On suppose que F est mesurable par rapport à \mathcal{F}_T . Alors*

$$\forall x \in E, \mathbb{E}_x(1_{\{T < \infty\}} F G(\theta_T)) = \mathbb{E}_x(1_{\{T < \infty\}} F \mathbb{E}_{X_T}(G)).$$

De manière équivalente,

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

Démonstration. Pour tout $n \in \mathbb{N}$, on a

$$\mathbb{E}_x(1_{\{T=n\}} F G(\theta_T)) = \mathbb{E}_x(1_{\{T=n\}} F G(\theta_n)) = \mathbb{E}_x(1_{\{T=n\}} F \mathbb{E}_{X_n}(G)).$$

(Propriété de Markov simple appliquée avec la fonction $F' = 1_{\{T=n\}} F$ qui est \mathcal{F}_n -mesurable puisque F est F_T -mesurable.) Il suffit de faire la somme sur $n \in \mathbb{N}$. \square

Rappel 4.1.14. *Pour une chaîne de Markov (X_n) sur E dénombrable, on a vu que la loi de X_{n+1} sachant X_n égale la loi de X_n sachant X_0, \dots, X_{n-1} . Et la propriété de Markov simple nous dit que on a $\mathbb{E}_x[G \circ \theta_n | \mathcal{F}_n] = \mathbb{E}_{X_n}[G]$ pour G mesurable. De plus, la propriété forte de Markov dit que*

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

pour T temps d'arrêt.

Exemple 4.1.15 (Galton-Watson). Soit $Z_0 = 1$, $Z_{n+1} = X_1^{(n+1)} + \dots + X_{Z_n}^{(n+1)}$. On suppose que $\mathbb{P}(Z_1 = k) = \mu_k$ et $\mu_k \geq 0$, $\sum \mu_k = 1$. Soit $E = \mathbb{N}$ et $\Omega = E^{\mathbb{N}}$. On pose $G : \Omega \rightarrow \mathbb{R}$ définie par $(\omega_0, \omega_1, \dots) \mapsto \omega_1 + \omega_2 + \omega_3$ C'est la population totale sur les 3 premières générations.

On utilise la propriété de Markov simple pour $n = 2$ et $x = 1$, on a

$$\begin{aligned}\mathbb{E}_1[G \circ \theta_n | \mathcal{F}_n] &= \mathbb{E}_{X_2}[G] \\ \mathbb{E}_1[G \circ \theta_n | \mathcal{F}_n] &= \mathbb{E}_1[\omega_1 + \omega_2 + \omega_3 | \mathcal{F}_2] = \mathbb{E}_{X_2}[\omega_1 + \omega_2 + \omega_3]\end{aligned}$$

Pour tout $k \in \mathbb{N}$ sur l'évènement $\{X_2 = k\}$, $\mathbb{E}_{X_2}[\omega_1 + \omega_2 + \omega_3] = \mathbb{E}_k[\omega_1 + \omega_2 + \omega_3]$.

Donc

$$\mathbb{E}_{X_2}[\omega_1 + \omega_2 + \omega_3] = \sum_{k \in \mathbb{N}} 1_{\{X_2=k\}} \mathbb{E}_k[\omega_1 + \omega_2 + \omega_3].$$

Exemple 4.1.16 (Marche aléatoire simple symétrique sur \mathbb{Z}). Soit $S_n = X_1 + \dots + X_n$ où (X_i) iid, $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$. On pose temps d'arrêt $T = \inf\{n, X_n = 1\}$ qui est fini p.s et on suppose $x = 0$. On définit

$$G : (\omega_1, \omega_2, \omega_3, \dots) \mapsto \sum_{k=1}^N 1_{\{\omega_k \geq 0\}}.$$

Utilisant la propriété de Markov forte avec T temps d'attente, on a

$$\mathbb{E}_0[G \circ \theta_T | \mathcal{F}_T] = \mathbb{E}_1[G]$$

.

Remarque 4.1.17. Question naturelle : si E est infini si on part de x , revient-on une infinité fois en x ?

Pour un point $x \in E$, on peut définir $H_x = \inf\{n \geq 1, X_n = x\}$ comme "Hitting time" en anglais et $N_x = \sum_{n=0}^{\infty} 1_{\{X_n=x\}}$ le nombre de visites de x .

Proposition 4.1.18. Soit $x \in E$, on a une des deux situations :

- Soit $\mathbb{P}(H_x < \infty) = 1$ et dans ce cas, $N_x = \infty$ \mathbb{P}_x p.s. $\Rightarrow x$ récurrent.
- Soit $\mathbb{P}(H_x < \infty) < 1$ et dans ce cas, $N_x < \infty$ \mathbb{P}_x p.s. $\Rightarrow x$ transitoire.

Démonstration. On utilise la propriété forte de Markov avec $T = H_x$. Si $k \in \mathbb{N}$, on a

$$\begin{aligned} \mathbb{P}_x(N_x \geq k+1) &= \mathbb{E}[1_{\{H_x < \infty\}} 1_{\{N_x \geq k\}} \circ \theta_{H_x}] \\ &= \mathbb{E}[1_{\{H_x < \infty\}} \mathbb{E}_x[1_{\{N_x \geq k\}}]] \\ &= \mathbb{P}_x(H_x < \infty) \mathbb{P}_x(N_x \geq k). \end{aligned}$$

Donc on obtient $\frac{\mathbb{P}(N_x \geq k+1)}{\mathbb{P}(N_x \geq k)} = \mathbb{P}_x(H_x < \infty)$. Cela conclut que soit $\mathbb{P}_x(H_x < \infty) = 1$ on a $\mathbb{P}(N_x \geq k) = 1$ pour tout k que soit $p = \mathbb{P}_x(H_x < \infty) < 1$ on a $\mathbb{P}(N_x \geq k) = p^k(1-p)$ et $\mathbb{E}[N_x] < \infty$. \square

Définition 4.1.19. On définit le noyau potentiel $U : E \times E \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ qui envoie (x, y) à $\mathbb{E}_x[N_y]$.

$$U(x, y) = \mathbb{E}_x[N_y].$$

Proposition 4.1.20. *On a les propriétés suivantes pour le noyau potentiel :*

1. $\forall x, y, U(x, y) = \sum_{n=0}^{\infty} Q_n(x, y)$.
2. x est récurrent si et seulement si $U(x, x) = +\infty$.
3. $\forall x, y, x \neq y, U(x, y) = U(y, y)\mathbb{P}_x(H_y < \infty)$.

Démonstration. 1.

$$U(x, y) = \mathbb{E}_x\left[\sum_{n=0}^{\infty} 1_{\{X_n=y\}}\right] = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y) = \sum_{n=0}^{\infty} Q_n(x, y).$$

2. Directement par proposition précédente.
3. Propriété forte de Markov avec $T = H_y$,

$$\mathbb{E}_x[N_y] = \mathbb{E}_x[1_{\{H_y < \infty\}} N_y \circ \theta_{H_y}] = \mathbb{E}_x[1_{\{H_y < \infty\}} \mathbb{E}_y[N_y]] = \mathbb{P}_x(H_y < \infty)U(y, y).$$

□

Exemple 4.1.21 (Marche aléatoire simple symétrique sur \mathbb{Z}^d). Soit (X_i) iid à valeur dans $E = \mathbb{Z}^d$, on a $2d$ vecteurs de norme L^1 égale à 1 : $(1, 0, 0, \dots), (-1, 0, 0, \dots), (0, 1, 0, \dots) \dots$. Soit U_d l'ensemble de ces vecteurs X_i uniforme dans U_d . Soit $S_n = X_1 + X_2 + \dots + X_n$. On a $Q_n(0, 0) = 0$ si n impair.

Pour le cas $d = 1$, on a

$$Q_{2n}(0, 0) = \frac{1}{2^{2n}} \binom{2n}{n} = \frac{(2n)!}{(n!)^2 2^{2n}} \sim \frac{\left(\frac{2n}{e}\right)^{2n} \sqrt{4\pi n}}{\left(\frac{n}{e}\right)^{2n} 2^{2n} (\sqrt{2\pi n})^2} \sim \frac{1}{\sqrt{\pi n}}.$$

Pour le cas $d \geq 2$, avec des arguments combinatoire, on montre qu'en dimension d ,

$$Q_{2n}^{(d)}(0, 0) \sim \frac{C_d}{n^{\frac{d}{2}}}.$$

De plus $\sum_{n=0}^{\infty} Q_{2n}^{(d)}(0, 0) = U(0, 0) \begin{cases} = +\infty & \text{si } d \leq 2 \\ < \infty & \text{si } d \geq 3 \end{cases}$.

Remarque 4.1.22. Si au lieu de prendre U_d , on prend $U'd = \{\text{vecteur de norme } L^\infty \text{ égale à } 1\}$, donc $|U'd| = 2^d$. Alors la formule devient

$$Q_{2n}(0, 0) = \left(\frac{1}{2^{2n}} \binom{2n}{n}\right)^d.$$

On montre aussi $\begin{cases} \text{récurent en dimension 1 et 2} \\ \text{transitoire en dimension plus grand que 3} \end{cases}$

Soit R l'ensemble des points récurrents.

Lemme 4.1.23. *Soient $x \in R$ et $y \in E$. Si $U(x, y) > 0$, alors $y \in R$ et $\mathbb{P}_y(H_x < \infty) = 1$. De plus on a $U(y, x) > 0$.*

Démonstration. Montrons $\mathbb{P}_y(H_x < \infty) = 1$:

$$\begin{aligned} 0 &= \mathbb{P}_x(N_x < \infty) \geq \mathbb{P}_x(H_y < \infty, H_x \circ \theta_{H_y} = \infty) \\ &= \mathbb{E}[1_{\{H_y < \infty\}} 1_{\{H_x = \infty\}} \circ \theta_{H_y}] \\ &= \mathbb{E}[1_{\{H_y < \infty\}} \mathbb{P}_y(H_x = \infty)] \\ &= \mathbb{P}_x(H_y < \infty) \mathbb{P}_y(H_x = \infty). \end{aligned}$$

Puisque $U(x, y) > 0$ on a $\mathbb{P}_x(H_y < \infty) > 0$ et $\mathbb{P}_y(H_x < \infty) > 0$. Si $n_1, n_2 \geq 1$, on a $Q_{n_1}(x, y) > 0$, $Q_{n_2}(y, x) > 0$. Si $p \geq 0$, on a $Q_{n_1+n_2+p}(y, y) \geq Q_{n_2}(y, x) Q_p(x, x) Q_{n_1}(x, y)$. Alors

$$\sum_m Q_m(y, y) \geq \sum_p Q_{n_1+n_2+p}(y, y) \geq Q_{n_2}(y, x) Q_{n_1}(x, y) \sum_p Q_p(x, x).$$

Donc y est récurrent. □

Théorème 4.1.24. *Il existe une partition de R , $R = \sqcup_{i \in I} R_i$ telle que*

1. *Si $x \in R_i$ et $y \in R_j$, $i \neq j$, alors $\mathbb{P}_x(H_y = \infty) = 1$.*
2. *Si $x \in R_i$ et $y \in R_i$, alors $\mathbb{P}_x(H_y < \infty) = 1$.*
3. *Si $x \notin R$ et $T = \inf\{n, X_n \in R\}$ alors \mathbb{P}_x -p.s.

 - soit $T = \infty$ et $\forall y \in E$, $N_y < \infty$.
 - soit $T < \infty$ et $\exists i \in I$, $\forall n > T, X_n \in R_i$*

Remarque 4.1.25. Dans le 3ème cas, $\forall y, \exists m, \forall n > m, X_n \neq y$. $\forall B$ ensemble fini, $\exists m, \forall n > m, X_n \notin B$.

Démonstration. On définit $x \sim y$ si $U(x, y) > 0$ si $x, y \in R$. Alors \sim est une relation d'équivalence. Les classes d'équivalence sont les R_i .

Pour $x \in R_i$, si $y \in R_j$, $i \neq j$, donc $U(x, y) = 0$. Cela implique $\mathbb{P}_x(H_y = \infty) = 1$. De même si $y \in R_i$, alors $\mathbb{P}_x(H_y < \infty) = 1$.

Pour $x \notin R$, Soit $T = \infty$. $y \in E$. Si $y \in R$, $N_y = 0$; Si $y \notin R$, on utilise la propriété forte de Markov à H_y , $N_y < \infty$ p.s. Soit $T < \infty$, on pose i tel que $X_T \in R_i$, propriété de Markov forte appliquée à l'instant T .

□

Définition 4.1.26. Soit (X_n) chaîne de Markov, (X_n) irréductibles si $\forall x, y \in E, U(x, y) > 0$.

Proposition 4.1.27. Soit (X_n) chaîne de Markov irréductible, il y a 2 possibilités

- tous les états sont récurrents, une seule classe de récurrence, $\forall x, y \mathbb{P}_x(H_y < \infty) = 1$.
(On dit que la chaîne est récurrente)
- tous les états sont transitoires, la chaîne est dite transitoire.

Démonstration. S'il existe un état récurrent x , $\forall y, U(x, y) > 0$. Alors y est récurrent et $\mathbb{P}_x(H_y < \infty) = 1$.

Sinon, tous les états sont transitoires.

□

Exemples 4.1.28. Soient S_n marches aléatoires simple sur \mathbb{Z}^d .

- C'est une chaîne de Markov $\begin{cases} \text{récurrente} & \text{si } d \leq 2 \\ \text{transitoire} & \text{si } d \geq 3 \end{cases}$
- Pour $d = 1$, on pose $T = \inf\{n, S_n \in \{-N, N\}\}$. Alors $X_n := S_{T \wedge n}$ admet deux états récurrents (N ou $-N$). De plus, les autres états sont transitoires. Ceci conclut qu'il n'est pas irréductible.
- Pour le cas $d = 3$, on pose $T = \inf\{n, S_n = (1, 0, 0)\}$. Comme S_n est transitoire et irréductible, on a $0 < \mathbb{P}_0(T < \infty) < 1$. Soit $X_n := S_{T \wedge n}$ est une chaîne de Markov avec $(1, 0, 0)$ récurrent, les autres états sont transitoires. Pour $x \neq (1, 0, 0)$, on a $0 < \mathbb{P}_x(T < \infty) < 1$.