Sorbonne Université

Année 2020–2021

Master de Mathématiques et Applications

Probabilités approfondies

(M1, UE 4M011)

Cours : Thierry Lévy

Travaux dirigés : Shen Lin, Raphaël Roux, Camille Tardif

Ce polycopié a été écrit par Zhan Shi (http://www.proba.jussieu.fr/pageperso/zhan/)

http://www.lpma-paris.fr/pageperso/levy/4M011.html

Table des matières

Chapitre 1

Espérance conditionnelle

Bienvenue à bord !

Ces notes servent comme support de l'UE 4M011 "Probabilités approfondies", niveau Master première année.

Les objets essentiels du cours sont les martingales à temps discret et les chaînes de Markov à espace d'états dénombrable, que l'on étudiera dans les deux chapitres suivants. Le chapitre présent est consacré à un outil fondamental dans ces études à venir, à savoir espérance conditionnelle par rapport à une tribu.

Il est impératif d'avoir une solide connaissance de la théorie de l'intégration abstraite, et d'avoir suivi un cours de probabilités de base.

Rappelons enfin que notre UE est incompatible avec les UE 4M010 "Probabilités de base" et 4M052 "Éléments de probabilités".

1. Avant le décollage

Prêtons quelques moments d'attention à des rappels de certaines notions de la théorie élémentaire des probabilités pour vérifier nos bagages mathématiques.

• Un espace mesurable $(\Omega, \mathscr{F})$ est un ensemble non vide Ω muni d'une tribu $\mathscr{F}$, c'est-à-dire une famille de parties de Ω satisfaisant : (i) $\Omega \in \mathscr{F}$; (ii) $A \in \mathscr{F} \Rightarrow A^c \in \mathscr{F}$; (iii) $A_n \in \mathscr{F}, \forall n \geq 1 \Rightarrow \cup_{n\geq 1} A_n \in \mathscr{F}$.

• Un espace mesuré $(\Omega, \mathscr{F}, \mu)$: μ est une mesure (positive)[1] sur l'espace mesurable $(\Omega, \mathscr{F})$.

[1] C'est-à-dire une application $\mu : \mathscr{F} \to [0, \infty]$ telle que $\mu(\varnothing) = 0$ et que $\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ pour toute suite $(A_n, n \geq 1)$ d'éléments *deux-à-deux disjoints* de $\mathscr{F}$.

• Un espace de probabilité (ou : espace probabilisé) $(\Omega, \mathscr{F}, \mathbb{P})$ = espace mesuré muni d'une mesure de probabilité (ou : probabilité) $\mathbb{P}$, c'est-à-dire une mesure telle que $\mathbb{P}(\Omega) = 1$.

• Une variable aléatoire X définie sur $(\Omega, \mathscr{F}, \mathbb{P})$, à valeurs dans un espace mesurable quelconque $(E, \mathscr{E})$, est une application $X : (\Omega, \mathscr{F}) \to (E, \mathscr{E})$ mesurable, c'est-à-dire que pour tout $B \in \mathscr{E}$, $X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}$ est un élément de $\mathscr{F}$.

• La **loi** de la variable aléatoire X, notée P_X, est la mesure de probabilité sur $(E, \mathscr{E})$ définie par $P_X(B) := \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B)$, $B \in \mathscr{E}$. [Il s'agit de la mesure image de $\mathbb{P}$ par l'application mesurable X.]

• Lorsque $(E, \mathscr{E}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ (espace réel muni de sa tribu borélienne[2]), on dit que X est une variable aléatoire réelle.

• Si X est une variable aléatoire à valeurs dans $[0, \infty]$, on écrit $\mathbb{E}(X) := \int_\Omega X(\omega)\, \mathbb{P}(\mathrm{d}\omega) = \int_{\mathbb{R}_+} x\, P_X(\mathrm{d}x) \in [0, \infty]$, $0 \times \infty := 0$.[3]

• Plus généralement, si X est une variable aléatoire réelle, et **si** $\int_\Omega |X(\omega)|\, \mathbb{P}(\mathrm{d}\omega) = \int_{\mathbb{R}} |x|\, P_X(\mathrm{d}x)$ est finie, on dit alors que X admet un moment d'ordre 1 (ou : X est intégrable), et l'on écrit $\mathbb{E}(X) := \int_\Omega X(\omega)\, \mathbb{P}(\mathrm{d}\omega) = \int_{\mathbb{R}} x\, P_X(\mathrm{d}x) \in \mathbb{R}$.

2. Rappels : Indépendance

Soient $\mathscr{G}_1, \cdots, \mathscr{G}_n$ des sous-tribus de $\mathscr{F}$. On dit qu'elles sont indépendantes si :

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n), \qquad \forall A_i \in \mathscr{G}_i \ (1 \le i \le n).$$

Soient $X_1, \cdots, X_n$ des variables aléatoires à valeurs dans $(E_1, \mathscr{E}_1), \cdots, (E_n, \mathscr{E}_n)$, respectivement. On dit que $X_1, \cdots, X_n$ sont indépendantes si $\sigma(X_1), \cdots, \sigma(X_n)$ sont indépendantes. Ceci équivaut à :

$$\mathbb{P}(X_1 \in B_1, \cdots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \cdots \mathbb{P}(X_n \in B_n), \qquad \forall B_i \in \mathscr{E}_i \ (1 \le i \le n).$$

En effet, il suffit de rappeler que $\sigma(X_i) = \{X_i^{-1}(B) : B \in \mathscr{E}_i\}$.

On dit que les événements $A_1, \cdots, A_n$ sont indépendants si les variables aléatoires $\mathbf{1}_{A_1}, \cdots, \mathbf{1}_{A_n}$ sont indépendantes, ce qui équivaut à $\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})$, $\forall 1 \le k \le n$, $\forall 1 \le i_1 < \cdots < i_k \le n$.[4]

[2] C'est-à-dire la tribu engendrée par les ensembles ouverts de $\mathbb{R}$, qui est également celle engendrée par tous les intervalles ou par toutes les demi-droites de $\mathbb{R}$.

[3] L'identité est une conséquence du fait que P_X est la mesure image de $\mathbb{P}$ par l'application mesurable X.

[4] Notation : $\mathbf{1}_A$ est la fonction indicatrice de A, c'est-à-dire que $\mathbf{1}_A(\omega)$ vaut 1 si $\omega \in A$ et vaut 0 sinon.

Remarque 2.1. Si $\mathscr{G}_1, \cdots, \mathscr{G}_n$ sont des sous-tribus indépendantes, et si pour tout i, X_i est une variable aléatoire $\mathscr{G}_i$-mesurable, alors $X_1, \cdots, X_n$ sont indépendantes.

En particulier, si $X_1, \cdots, X_n$ sont des variables aléatoires indépendantes, à valeurs dans $(E_1, \mathscr{E}_1), \cdots, (E_n, \mathscr{E}_n)$, respectivement, et si $f_i : (E_i, \mathscr{E}_i) \to (\widetilde{E}_i, \widetilde{\mathscr{E}}_i)$ (pour $1 \le i \le n$) sont des applications mesurables, alors $f_1(X_1), \cdots, f_n(X_n)$ sont des variables aléatoires indépendantes. □

Théorème 2.2. *Soient $X_1, \cdots, X_n$ des variables aléatoires à valeurs dans $(E_1, \mathscr{E}_1), \cdots, (E_n, \mathscr{E}_n)$, respectivement. Les conditions suivantes sont équivalentes.*

(i) *$X_1, \cdots, X_n$ sont indépendantes.*

(ii) $P_{(X_1,\cdots,X_n)} = P_{X_1} \otimes \cdots \otimes P_{X_n}$.

(iii) $\mathbb{E}[f_1(X_1)\cdots f_n(X_n)] = \mathbb{E}[f_1(X_1)]\cdots\mathbb{E}[f_n(X_n)]$ *pour toute fonction mesurable positive (ou bornée) f_i sur $(E_i, \mathscr{E}_i)$.*

Preuve. (i) ⇔ (ii) : Soit $B_i \in \mathscr{E}_i$. Par définition,

$$\begin{aligned} P_{(X_1,\cdots,X_n)}(B_1 \times \cdots \times B_n) &= \mathbb{P}(X_1 \in B_1, \cdots, X_n \in B_n), \\ (P_{X_1} \otimes \cdots \otimes P_{X_n})(B_1 \times \cdots \times B_n) &= \mathbb{P}(X_1 \in B_1)\cdots\mathbb{P}(X_n \in B_n). \end{aligned}$$

Donc l'indépendance de $X_1, \cdots, X_n$ équivaut à ce que les deux mesures de probabilité $P_{(X_1,\cdots,X_n)}$ et $P_{X_1} \otimes \cdots \otimes P_{X_n}$ prennent les mêmes valeurs sur tous les pavés, ce qui revient à dire, grâce au lemme de classe monotone,[5] que ces deux mesures de probabilités sont identiques.

(iii) ⇒ (i) : Trivial en prenant $f_i := \mathbf{1}_{B_i}$ avec $B_i \in \mathscr{E}_i$.

(ii) ⇒ (iii) : Si les f_i sont mesurables positives, alors par Fubini–Tonelli,

$$\begin{aligned} \mathbb{E}[f_1(X_1)\cdots f_n(X_n)] &= \int_{E_1\times\cdots\times E_n} f_1(x_1)\cdots f_n(x_n)\, P_{X_1}(\mathrm{d}x_1)\cdots P_{X_n}(\mathrm{d}x_n) \\ &= \Big(\int_{E_1} f_1(x_1)\, P_{X_1}(\mathrm{d}x_1)\Big)\cdots\Big(\int_{E_n} f_n(x_n)\, P_{X_n}(\mathrm{d}x_n)\Big), \end{aligned}$$

qui n'est autre que $\mathbb{E}[f_1(X_1)]\cdots\mathbb{E}[f_n(X_n)]$. Dans le cas où les f_i sont bornées, il suffit de considérer séparément f_i^+ et f_i^-. □

[5]Soient $\mathscr{C}$ et $\mathscr{M}$ deux ensembles de parties de Ω, tels que $\mathscr{C} \subset \mathscr{M}$. Si $\mathscr{C}$ est stable par intersections finies (= "π-système"), et si $\mathscr{M}$ est une classe monotone, alors $\sigma(\mathscr{C}) \subset \mathscr{M}$. Rappelons la définition d'une classe monotone $\mathscr{M}$: (i) $\Omega \in \mathscr{M}$; (ii) $A, B \in \mathscr{M}$, $A \subset B \Rightarrow B\backslash A \in \mathscr{M}$; (iii) $A_n \uparrow A$, $A_n \in \mathscr{M} \Rightarrow A \in \mathscr{M}$.

Remarque 2.3. Si les fonctions f_i sont de signe quelconque, l'égalité

$$\mathbb{E}[f_1(X_1)\cdots f_n(X_n)] = \mathbb{E}[f_1(X_1)]\cdots\mathbb{E}[f_n(X_n)],$$

reste valable à condition que $\mathbb{E}(|f_i(X_i)|) < \infty$, $\forall i$. On a alors aussi

$$\mathbb{E}[\,|f_1(X_1)\cdots f_n(X_n)|\,] = \mathbb{E}[\,|f_1(X_1)|\,]\cdots\mathbb{E}[\,|f_n(X_n)|\,] < \infty\,,$$

ce qui justifie l'écriture $\mathbb{E}[f_1(X_1)\cdots f_n(X_n)]$.

En particulier, si $X_1, \cdots, X_n$ sont des variables aléatoires réelles indépendantes et dans L^1, on a aussi $X_1\cdots X_n \in L^1$, et $\mathbb{E}(X_1\cdots X_n) = \mathbb{E}(X_1)\cdots\mathbb{E}(X_n)$. [Attention : en général, le produit de variables aléatoires dans L^1 n'est pas dans L^1.] $\square$

Corollaire 2.4. *Si X_1 et X_2 sont des variables aléatoires réelles indépendantes et dans L^2, alors* $\mathrm{Cov}(X_1, X_2) = 0$.

Corollaire 2.5. *Soient $X_1, \cdots, X_n$ des variables aléatoires réelles indépendantes, telles que chaque X_i admet une densité notée f_{X_i}. Alors $(X_1, \cdots X_n)$ admet une densité donnée par $f_{(X_1,\cdots X_n)}(x_1, \cdots, x_n) := f_{X_1}(x_1)\cdots f_{X_n}(x_n)$.*

Proposition 2.6. *Soient $\mathscr{C}_1 \subset \mathscr{F}, \cdots, \mathscr{C}_n \subset \mathscr{F}$ des classes stables par intersections finies telles que $\Omega \in \mathscr{C}_i$, $\forall i$. Supposons que*

$$\mathbb{P}(C_1\cap\cdots\cap C_n) = \mathbb{P}(C_1)\cdots\mathbb{P}(C_n), \qquad \forall C_i \in \mathscr{C}_i.$$

Alors les tribus $\sigma(\mathscr{C}_1), \cdots, \sigma(\mathscr{C}_n)$ sont indépendantes.

Preuve. Fixons $C_2 \in \mathscr{C}_2, \cdots, C_n \in \mathscr{C}_n$. Posons

$$\mathscr{M}_1 := \{B_1 \in \sigma(\mathscr{C}_1) : \mathbb{P}(B_1\cap C_2\cap\cdots\cap C_n) = \mathbb{P}(B_1)\,\mathbb{P}(C_2)\cdots\mathbb{P}(C_n)\}.$$

On vérifie très facilement que $\mathscr{M}_1$ est une classe monotone[6], tandis que par hypothèse, $\mathscr{C}_1 \subset \mathscr{M}_1$, et $\mathscr{C}_1$ est stable par intersections finies. Il résulte du lemme de classe monotone que $\mathscr{M}_1 = \sigma(\mathscr{C}_1)$. On a donc montré que

$$\mathbb{P}(B_1\cap C_2\cap\cdots\cap C_n) = \mathbb{P}(B_1)\,\mathbb{P}(C_2)\cdots\mathbb{P}(C_n), \qquad \forall B_1 \in \sigma(\mathscr{C}_1),\ \forall C_2 \in \mathscr{C}_2, \cdots, \forall C_n \in \mathscr{C}_n.$$

Pour continuer, on fixe $B_1 \in \sigma(\mathscr{C}_1)$, $C_3 \in \mathscr{C}_3, \cdots, C_n \in \mathscr{C}_n$, et on pose

$$\mathscr{M}_2 := \{B_2 \in \sigma(\mathscr{C}_2) : \mathbb{P}(B_1\cap B_2\cap C_3\cap\cdots\cap C_n) = \mathbb{P}(B_1)\,\mathbb{P}(B_2)\,\mathbb{P}(C_3)\cdots\mathbb{P}(C_n)\}.$$

[6] On utilise l'hypothèse $\Omega \in \mathscr{C}_1$ pour dire que $\Omega \in \mathscr{M}_1$.

De nouveau, $\mathscr{M}_2$ est une classe monotone, tandis que l'on a déjà montré qu'elle contient $\mathscr{C}_2$ qui par hypothèse est stable par intersections finies : donc $\mathscr{M}_2 = \sigma(\mathscr{C}_2)$. En raisonnant par récurrence, on arrive à :

$$\mathbb{P}(B_1 \cap \cdots \cap B_n) = \mathbb{P}(B_1) \cdots \mathbb{P}(B_n), \qquad \forall B_1 \in \sigma(\mathscr{C}_1), \cdots, \forall B_n \in \sigma(\mathscr{C}_n).$$

Donc les tribus $\sigma(\mathscr{C}_1)$, $\cdots$, $\sigma(\mathscr{C}_n)$ sont indépendantes. $\square$

Théorème 2.7. (Regroupement par paquets). *Soient $\mathscr{G}_1$, $\cdots$, $\mathscr{G}_n$ des sous-tribus indépendantes de $\mathscr{F}$. Soient $n_0 := 0 < n_1 < \cdots < n_p = n$. Alors les tribus*[7]

$$\begin{aligned}
\mathscr{D}_1 &:= \sigma(\mathscr{G}_1, \cdots, \mathscr{G}_{n_1}) := \sigma(\{A : A \in \mathscr{G}_1 \cup \cdots \cup \mathscr{G}_{n_1}\}), \\
\mathscr{D}_2 &:= \sigma(\mathscr{G}_{n_1+1}, \cdots, \mathscr{G}_{n_2}), \\
& \cdots \\
\mathscr{D}_p &:= \sigma(\mathscr{G}_{n_{p-1}+1}, \cdots, \mathscr{G}_{n_p}),
\end{aligned}$$

sont indépendantes.

Preuve. Il suffit d'appliquer la Proposition 2.6 en prenant $\mathscr{C}_j$ la classe des parties de la forme

$$B_{n_{j-1}+1} \cap \cdots \cap B_{n_j}, \qquad B_i \in \mathscr{G}_i,$$

qui est stable par intersections finies et qui contient Ω, avec $\sigma(\mathscr{C}_j) = \mathscr{D}_j$. $\square$

Théorème 2.8. *Si X_1, $\cdots$, X_n sont indépendantes, alors les variables aléatoires*

$$Y_1 := (X_1, \cdots, X_{n_1}), \cdots, Y_p := (X_{n_{p-1}+1}, \cdots, X_{n_p}),$$

sont indépendantes.

Preuve. C'est une conséquence immédiate du Théorème 2.7. $\square$

Exemple 2.9. Si ξ_1, ξ_2, ξ_3, ξ_4, ξ_5, ξ_6, ξ_7, ξ_8 sont des variables aléatoires réelles indépendantes, alors les variables aléatoires $Z_1 := \xi_1\xi_7$, $Z_2 := (\xi_4^2, \xi_3, \cos(\xi_6))$ et $Z_3 := (\xi_2 + 1)^3$ sont indépendantes. $\square$

[7] Donc $\sigma(\mathscr{G}_1, \cdots, \mathscr{G}_{n_1})$ est la plus petite tribu qui contient tous les éléments de toutes les tribus $\mathscr{G}_i$, $1 \le i \le n_1$.

Indépendance d'une famille infinie. Soit $(\mathscr{G}_i,\, i \in I)$ une famille quelconque de sous-tribus de $\mathscr{F}$. On dit que cette famille est indépendante si pour tout sous-ensemble fini $\{i_1, \cdots, i_p\} \subset I$, les tribus $\mathscr{G}_{i_1}, \cdots, \mathscr{G}_{i_p}$ sont indépendantes.

Si $(X_i,\, i \in I)$ est une famille quelconque de variables aléatoires, cette famille est dite indépendante si la famille de tribus $(\sigma(X_i),\, i \in I)$ l'est. La proposition suivante montre que le principe du regroupement par paquets est valable dans un contexte plus général.

Théorème 2.10. (Regroupement par paquets). *Soit $(\mathscr{G}_n,\, n \geq 1)$ une suite de tribus indépendantes. Soient $I_1 \subset \mathbb{N}^*, \cdots, I_p \subset \mathbb{N}^*$ disjoints. Alors les tribus $\sigma(\mathscr{G}_i,\ i \in I_1), \cdots, \sigma(\mathscr{G}_i,\ i \in I_p)$ sont indépendantes.*

Preuve. Il suffit d'appliquer la Proposition 2.6 en prenant

$$\mathscr{C}_1 := \bigcup_{k\geq 1} \sigma(\mathscr{G}_i,\ i \in I_1 \cap [1, k]), \qquad \cdots, \qquad \mathscr{C}_p := \bigcup_{k\geq 1} \sigma(\mathscr{G}_i,\ i \in I_p \cap [1, k]),$$

l'hypothèse étant satisfaite grâce au principe du regroupement par paquets usuel. □

Théorème 2.11. (Loi du tout ou rien de Kolmogorov) *Soit $(\mathscr{G}_n,\, n \geq 1)$ une suite de tribus indépendantes. On pose, pour tout n,*

$$\mathscr{Q}_n := \sigma(\mathscr{G}_i,\ i \geq n).$$

Alors la tribu de queue $\mathscr{Q}_\infty$, définie par $\mathscr{Q}_\infty := \bigcap_{n\geq 1} \mathscr{Q}_n$, est triviale, au sens où $\mathbb{P}(B) = 0$ ou 1 pour tout $B \in \mathscr{Q}_\infty$.

Cas particulier : *Soit $(X_n,\, n \geq 1)$ une suite de variables aléatoires indépendantes. La tribu de queue $\mathscr{Q}_\infty := \cap_{n\geq 1}\sigma(X_i,\ i \geq n)$ est triviale.*

Preuve. Posons $\mathscr{F}_n := \sigma(\mathscr{G}_1, \cdots, \mathscr{G}_n)$. D'après le Théorème 2.10, $\mathscr{F}_n$ est indépendante de $\mathscr{Q}_{n+1}$, donc a fortiori de $\mathscr{Q}_\infty$.

Puisque la classe $\cup_{n\geq 1}\mathscr{F}_n$ est stable par intersections finies et contient Ω, la Proposition 2.6 nous dit que $\mathscr{Q}_\infty$ est indépendante de $\sigma(\cup_{n\geq 1}\mathscr{F}_n) = \sigma(\mathscr{G}_n,\, n \geq 1)$, a fortiori, de $\mathscr{Q}_\infty$ elle-même : pour tout $B \in \mathscr{Q}_\infty$, on a $\mathbb{P}(B) = \mathbb{P}(B \cap B) = [\mathbb{P}(B)]^2$, ce qui n'est possible que si $\mathbb{P}(B) = 0$ ou 1. □

Exemple 2.12. Soit $(\xi_n,\, n \geq 1)$ une suite de variables aléatoires réelles indépendantes. Alors $\mathbb{P}(\sum_n \xi_n \text{ converge}) = 0$ ou 1. Autrement dit, la série $\sum_n \xi_n$ a la propriété suivante : soit elle converge p.s., soit elle diverge p.s. □

3. Un exemple informel

Soit $(\Omega, \mathscr{F}, \mathbb{P})$ un espace de probabilité. Soit $\mathscr{G} \subset \mathscr{F}$ une sous-tribu. On étudie dans ce chapitre espérance conditionnelle d'une variable aléatoire réelle sachant $\mathscr{G}$, lorsque cette espérance conditionnelle a un sens.

D'une façon informelle, l'espérance est une sorte de moyenne (pondérée par la loi de la variable aléatoire), tandis que le conditionnement par rapport à la sous-tribu $\mathscr{G}$ revient à dire que nous avons connaissance du résultat de tous les événements de $\mathscr{G}$ (c'est-à-dire que nous savons si, oui ou non, un événement quelconque de $\mathscr{G}$ est réalisé). La notion d'espérance conditionnelle est un formalisme rigoureux de la question heuristique suivante : comment calculer la moyenne d'une variable aléatoire réelle connaissant le résultat de tous les événements de $\mathscr{G}$?

Commençons par un exemple simple.

On jette, de façon indépendante, deux dés parfaits à six faces chacun. Soit Y le résultat du premier jet, et soit Z celui du second. Plus formellement, Y, $Z : \Omega \to \{1, \cdots, 6\}$ sont des variables aléatoires indépendantes et suivent la même loi uniforme sur $\{1, \cdots, 6\}$, à savoir que $\mathbb{P}(Y = i) = \mathbb{P}(Z = i) = \frac{1}{6}$ pour tout $i \in \{1, \cdots, 6\}$. On voit tout de suite que $\mathbb{E}(Y) = \mathbb{E}(Z) = \frac{7}{2}$. On s'intéresse à la somme $X := Y + Z$. Par linéarité de l'espérance, $\mathbb{E}(X) = \mathbb{E}(Y) + \mathbb{E}(Z) = 7$.

On suppose maintenant que le résultat du second jet est connu (c'est-à-dire que la valeur de Z est déterminée), et l'on veut calculer la moyenne de X sachant Z. La situation étant très simple, on peut facilement donner une réponse informelle : la moyenne de X sachant Z = (la moyenne de Y sachant Z) + (la moyenne de Z sachant Z) ; comme Y ne dépend pas du tout de Z, connaître Z ne change rien à Y, donc la moyenne de Y sachant Z est simplement la moyenne de Y, c'est-à-dire $\frac{7}{2}$; d'autre part, la moyenne de Z sachant Z est évidemment Z puisque sa value est supposée connue. On a donc "prouvé" que la moyenne de X sachant Z est égale à $\frac{7}{2} + Z$.

Employons maintenant un langage probabiliste. Par l'expression "sachant Z", il faut comprendre "sachant l'information apportée par la connaissance de Z", autrement dit, sachant la tribu engendrée par Z (qui sera, plus généralement, remplacée par une sous-tribu quelconque $\mathscr{G}$ dans la suite du chapitre). Quant à la moyenne, il s'agit de l'espérance, tandis que le mot "sachant" signifie "conditionnellement à". La conclusion du paragraphe précédent se lit donc : l'espérance de X conditionnellement à la tribu $\sigma(Z)$ est égale à $\frac{7}{2} + Z$. D'une façon équivalente, on dit également que l'espérance conditionnelle de X sachant $\sigma(Z)$ est égale à $\frac{7}{2} + Z$.

Dans la suite, on notera $\mathbb{E}[X \mid \sigma(Z)]$ pour l'espérance conditionnelle de X sachant $\sigma(Z)$. On observe qu'il s'agit d'une variable aléatoire, qui ne dépend que de Z, ce qui est nullement surprenant puisque l'on a intégré sur tous les paramètres qui sont indépendants de Z et seulement sur ceux-là.

Passons maintenant au formalisme en général.

4. Espérance conditionnelle

Soit $(\Omega, \mathscr{F}, \mathbb{P})$ un espace de probabilité. Soit $\mathscr{G} \subset \mathscr{F}$ une sous-tribu. On définit l'espérance conditionnelle de X sachant $\mathscr{G}$ si X est intégrable ou si X est positive.

4.1. Cas des variables aléatoires intégrables

Soit X une variable aléatoire réelle et intégrable: $\mathbb{E}(|X|) < +\infty$.

Définition 4.1. *On appelle espérance conditionnelle de X sachant $\mathscr{G}$, notée $\mathbb{E}(X \mid \mathscr{G})$ ou $\mathbb{E}^{\mathscr{G}}(X)$, toute variable aléatoire Y intégrable vérifiant les deux conditions suivantes :*

(i) *Y est $\mathscr{G}$-mesurable ;*
(ii) *$\forall A \in \mathscr{G}$, $\mathbb{E}(X\,\mathbf{1}_A) = \mathbb{E}(Y\,\mathbf{1}_A)$ (c'est-à-dire $\int_A X \,\mathrm{d}\mathbb{P} = \int_A Y \,\mathrm{d}\mathbb{P}$).*

On écrira probabilité conditionnelle $\mathbb{P}(A \mid \mathscr{G}) := \mathbb{E}(\mathbf{1}_A | \mathscr{G})$. Attention : il s'agit d'une variable aléatoire !

Montrons l'existence et l'unicité de l'espérance conditionnelle.

Unicité. Soient Y et $\widetilde{Y}$ deux variables aléatoires intégrables vérifiant (i) et (ii). On a, pour tout $A \in \mathscr{G}$, $\mathbb{E}(Y\,\mathbf{1}_A) = \mathbb{E}(\widetilde{Y}\,\mathbf{1}_A)$. Soit $\varepsilon > 0$ et soit $A := \{\omega \in \Omega : \ Y(\omega) - \widetilde{Y}(\omega) \geq \varepsilon\} \in \mathscr{G}$. On a

$$0 = \mathbb{E}(Y\,\mathbf{1}_A) - \mathbb{E}(\widetilde{Y}\,\mathbf{1}_A) = \mathbb{E}((Y - \widetilde{Y})\,\mathbf{1}_A) \geq \varepsilon\,\mathbb{P}(A)\,,$$

ce qui signifie que $\mathbb{P}(A) = 0$. Le choix de $\varepsilon > 0$ étant quelconque, on obtient $Y \leq \widetilde{Y}$ p.s. De même, $\widetilde{Y} \leq Y$ p.s. Donc $Y = \widetilde{Y}$ p.s. Ceci montre l'unicité.[8] □

Existence. La preuve s'appuie sur le théorème de Radon–Nikodým dont on rappelle l'énoncé : si μ et ν sont deux mesures σ-finies sur $(\Omega, \mathscr{G})$ telles que $\mu \ll \nu$ (c'est-à-dire que

[8] Strictement parlant, toute variable aléatoire Y satisfaisant (i) et (ii) est une version de $\mathbb{E}(X \mid \mathscr{G})$ (c'est-à-dire dans la même classe d'équivalence que $\mathbb{E}(X \mid \mathscr{G})$) : on doit, par exemple, écrire $Y = \mathbb{E}(X \mid \mathscr{G})$ p.s., mais l'expression "presque sûr" est souvent omise.

pour tout $A \in \mathscr{G}$, $\nu(A) = 0 \Rightarrow \mu(A) = 0$), alors il existe une fonction mesurable $f \geq 0$ telle que $\int_A f \, \mathrm{d}\nu = \mu(A)$, $\forall A \in \mathscr{G}$. La fonction f est notée par $\frac{\mathrm{d}\mu}{\mathrm{d}\nu}$, et appelée la dérivée (ou : densité) Radon–Nikodým de μ par rapport à ν.

On suppose sans perte de généralité que $X \geq 0$ (sinon, on considèrera séparément X^+ et X^-). Soit μ la mesure sur $(\Omega, \mathscr{G})$ définie par $\mu(A) = \int_A X \, \mathrm{d}\mathbb{P}$, $A \in \mathscr{G}$. L'intégrabilité de X garantit que μ est une mesure finie. On a donc deux mesures finies $\mathbb{P}$ (ou plutôt la restriction de $\mathbb{P}$ sur $\mathscr{G}$) et μ sur $\mathscr{G}$ telles que $\mu \ll \mathbb{P}$. Par le théorème de Radon–Nikodým, $\frac{\mathrm{d}\mu}{\mathrm{d}\mathbb{P}}$ est bien définie, qui est intégrable car μ est une mesure finie.

Or, $\frac{\mathrm{d}\mu}{\mathrm{d}\mathbb{P}}$ est $\mathscr{G}$-mesurable, telle que pour tout $A \in \mathscr{G}$, $\int_A X \, \mathrm{d}\mathbb{P} = \mu(A) = \int_A \frac{\mathrm{d}\mu}{\mathrm{d}\mathbb{P}} \, \mathrm{d}\mathbb{P}$, ce qui prouve que $\frac{\mathrm{d}\mu}{\mathrm{d}\mathbb{P}}$ satisfait les conditions (i) et (ii). □

Exercice 4.2. Montrer que si $\widetilde{X} = X$ p.s., alors $\mathbb{E}(\widetilde{X} \,|\, \mathscr{G}) = \mathbb{E}(X \,|\, \mathscr{G})$ p.s. □

Exercice 4.3. (Exercice très important). Soient X et Y des variables aléatoires réelles intégrables. Montrer que $\mathbb{E}(X \,|\, \mathscr{G}) \leq \mathbb{E}(Y \,|\, \mathscr{G})$ p.s. $\Leftrightarrow$ $\mathbb{E}(X \, \mathbf{1}_A) \leq \mathbb{E}(Y \, \mathbf{1}_A)$, $\forall A \in \mathscr{G}$. □

Exemple 4.4. Soit X une variable aléatoire réelle intégrable et $\mathscr{G}$-mesurable. Sachant la tribu $\mathscr{G}$, X qui est $\mathscr{G}$-mesurable est considérée comme une constante : $\mathbb{E}(X \,|\, \mathscr{G}) = X$ p.s.

En effet, il est clair dans cette situation que X satisfait les conditions (i) et (ii). □

Exemple 4.5. L'exemple précédent concerne la situation particulière où toute information sur X est déjà contenue dans la tribu $\mathscr{G}$. On étudie maintenant l'autre situation extrême: X est indépendante de $\mathscr{G}$, c'est-à-dire que $\sigma(X)$ et $\mathscr{G}$ sont indépendantes.[9] Dans ce cas, le fait de savoir $\mathscr{G}$ n'apporte strictement rien sur X: $\mathbb{E}(X \,|\, \mathscr{G}) = \mathbb{E}(X)$ p.s.

En effet, $\mathbb{E}(X)$ est une variable aléatoire (dégénérée) intégrable et $\mathscr{G}$-mesurable. Il nous suffit de vérifier la condition (ii). Soit $A \in \mathscr{G}$. Comme X et $\mathbf{1}_A$ sont indépendantes, on a $\mathbb{E}(X \, \mathbf{1}_A) = \mathbb{E}(X) \, \mathbb{E}(\mathbf{1}_A) = \mathbb{E}[\,\mathbb{E}(X) \, \mathbf{1}_A\,]$, ce qui prouve (ii). □

Théorème 4.6. *Si X est une variable aléatoire intégrable, alors* $\mathbb{E}[\,\mathbb{E}(X \,|\, \mathscr{G})\,] = \mathbb{E}(X)$.

Preuve. Il suffit de prendre $A = \Omega$ dans la propriété (ii). □

L'espérance conditionnelle jouit de la plupart des propriétés de l'espérance mathématique.

Propriété 4.7. *Soient X et Y des variables aléatoires réelles intégrables.*

- *Pour tous réels a et b, on a* $\mathbb{E}(aX + bY \,|\, \mathscr{G}) = a \, \mathbb{E}(X \,|\, \mathscr{G}) + b \, \mathbb{E}(Y \,|\, \mathscr{G})$ p.s.

[9] C'est le cas, par exemple, si $\mathscr{G} = \{\varnothing, \Omega\}$.

- *Si* $X \geq Y$*, alors* $\mathbb{E}(X \,|\, \mathscr{G}) \geq \mathbb{E}(Y \,|\, \mathscr{G})$ p.s. *En particulier,* $X \geq 0$ *implique* $\mathbb{E}(X \,|\, \mathscr{G}) \geq 0$.
- $|\mathbb{E}(X \,|\, \mathscr{G})| \leq \mathbb{E}(|X| \,|\, \mathscr{G})$ p.s. *En particulier,* $\mathbb{E}\{|\mathbb{E}(X \,|\, \mathscr{G})|\} \leq \mathbb{E}(|X|)$.
- *Si* $\mathscr{H} \subset \mathscr{G} \subset \mathscr{F}$ *sont des sous-tribus, alors* $\mathbb{E}[\,\mathbb{E}(X \,|\, \mathscr{G}) \,|\, \mathscr{H}\,] = \mathbb{E}(X \,|\, \mathscr{H})$ p.s.

Preuve. Les trois premières assertions sont triviales. Montrons la dernière. Il est clair que $\mathbb{E}[\,\mathbb{E}(X \,|\, \mathscr{G}) \,|\, \mathscr{H}]$ est bien définie, car par définition, $\mathbb{E}(X \,|\, \mathscr{G})$ est intégrable. Remarquons que $\mathbb{E}[\,\mathbb{E}(X \,|\, \mathscr{G}) \,|\, \mathscr{H}]$ est $\mathscr{H}$-mesurable, ce qui donne la propriété (i) dans la Définition 4.1. Pour la propriété (ii), soit $A \in \mathscr{H}$, et on a

$$\mathbb{E}\Big[\mathbb{E}[\,\mathbb{E}(X \,|\, \mathscr{G}) \,|\, \mathscr{H}]\ \mathbf{1}_A\Big] = \mathbb{E}\Big[\mathbb{E}(X \,|\, \mathscr{G})\ \mathbf{1}_A\Big] = \mathbb{E}(X\,\mathbf{1}_A),$$

car $A \in \mathscr{G}$. Ceci donne la propriété (ii) dans la Définition 4.1, et confirme donc que $\mathbb{E}[\,\mathbb{E}(X \,|\, \mathscr{G}) \,|\, \mathscr{H}] = \mathbb{E}(X \,|\, \mathscr{H})$ p.s. □

Théorème 4.8. *Soit X une variable aléatoire réelle intégrable. Si Y est une variable aléatoire réelle $\mathscr{G}$-mesurable telle que XY soit intégrable,*[10] *alors* $\mathbb{E}(XY \,|\, \mathscr{G}) = Y\,\mathbb{E}(X \,|\, \mathscr{G})$ p.s. [*Par conséquent,* $\mathbb{E}(XY) = \mathbb{E}[Y\,\mathbb{E}(X \,|\, \mathscr{G})]$.]

Preuve. Sans perte de généralité, on suppose que X et Y sont positives (sinon, on considèrera leurs parties positive et négative, séparément).

IL s'agit de prouver que pour tout $A \in \mathscr{G}$,

$$\mathbb{E}\Big[Y\,\mathbb{E}(X \,|\, \mathscr{G})\,\mathbf{1}_A\Big] = \mathbb{E}\Big[XY\,\mathbf{1}_A\Big].$$

[En particulier, en prenant $A = \Omega$ on déduit que $Y\,\mathbb{E}(X \,|\, \mathscr{G})$ est intégrable.]

Montrons d'abord l'identité dans le cas particulier où Y est une fonction indicatrice : soit $B \in \mathscr{G}$, alors

$$\mathbb{E}[\,\mathbf{1}_B\,\mathbb{E}(X \,|\, \mathscr{G})\,\mathbf{1}_A\,] = \mathbb{E}(X\,\mathbf{1}_{B\cap A}) = \mathbb{E}(\,\mathbf{1}_B\,X\,\mathbf{1}_A)\,,$$

comme prévu.

Par combinaison linéaire, l'identité reste vraie pour toute fonction étagée[11] Y qui est $\mathscr{G}$-mesurable. Le théorème de convergence monotone confirme alors que l'identité est encore valable si Y est positive et $\mathscr{G}$-mesurable, en rappelant que toute fonction mesurable et positive est limite croissante (au sens de convergence simple) de fonctions étagées positives. □

À la fin de la preuve précédente, on a utilisé le théorème de convergence monotone. Rappelons quelques résultats de convergence en théorie des probabilités élémentaire :

[10] En particulier, si Y est $\mathscr{G}$-mesurable et bornée.

[11] C'est-à-dire combinaison linéaire finie de fonctions indicatrices d'ensembles mesurables.

- convergence monotone : $X_n \geq 0$ $(\forall n)$ $X_n \uparrow X \;\Rightarrow\; \mathbb{E}(X_n) \uparrow \mathbb{E}(X)$;
- Fatou : $X_n \geq 0 \;\Rightarrow\; \mathbb{E}(\liminf_{n\to\infty} X_n) \leq \liminf_{n\to\infty} \mathbb{E}(X_n)$;
- convergence dominée : $|X_n| \leq Z$ $(\forall n)$, $\mathbb{E}(Z) < \infty$, $X_n \to X$ p.s. $\Rightarrow\; \mathbb{E}(X_n) \to \mathbb{E}(X)$.

Exemple 4.9. Soit X une variable aléatoire réelle intégrable. Soient $\mathscr{G}$ et $\mathscr{H}$ deux sous-tribus de $\mathscr{F}$. Si $\mathscr{H}$ est indépendante de $\sigma(\sigma(X), \mathscr{G})$, alors $\mathbb{E}[X \,|\, \sigma(\mathscr{G}, \mathscr{H})] = \mathbb{E}(X \,|\, \mathscr{G})$ p.s.

En effet, $\mathbb{E}(X \,|\, \mathscr{G})$ étant une variable aléatoire intégrable et $\sigma(\mathscr{G}, \mathscr{H})$-mesurable (car même $\mathscr{G}$-mesurable), il suffit de vérifier que pour tout $D \in \sigma(\mathscr{G}, \mathscr{H})$, on a

$$\mathbb{E}[\mathbb{E}(X \,|\, \mathscr{G})\, \mathbf{1}_D] = \mathbb{E}(X\, \mathbf{1}_D)\,.$$

On considère

$$\begin{aligned} \mathscr{C} &:= \{A \cap B :\; A \in \mathscr{G},\, B \in \mathscr{H}\}, \\ \mathscr{M} &:= \{D \in \sigma(\mathscr{G}, \mathscr{H}) :\; \mathbb{E}[\mathbb{E}(X \,|\, \mathscr{G})\, \mathbf{1}_D] = \mathbb{E}(X\, \mathbf{1}_D)\}\,. \end{aligned}$$

Il est clair que $\mathscr{C}$ est stable par intersections finies. Si l'on sait prouver (a) $\mathscr{C} \subset \mathscr{M}$, et (b) $\mathscr{M}$ est une classe monotone, alors par le lemme de classe monotone, on aura $\sigma(\mathscr{C}) \subset \mathscr{M}$. Or, $\mathscr{C}$ contient tous les éléments de $\mathscr{G}$ et de $\mathscr{H}$, on a $\sigma(\mathscr{C}) = \sigma(\mathscr{G}, \mathscr{H})$, et donc $\mathscr{M} = \sigma(\mathscr{G}, \mathscr{H})$. On aura donc l'identité cherchée.

Il reste de prouver (a) et (b).

Montrons d'abord (a). On écrit $\mathbb{E}(X \,|\, \mathscr{G})\, \mathbf{1}_{A\cap B} = \mathbb{E}(X \,|\, \mathscr{G})\, \mathbf{1}_A \times \mathbf{1}_B$. Comme $\mathbb{E}(X \,|\, \mathscr{G})\, \mathbf{1}_A$ et $\mathbf{1}_B$ sont, respectivement, $\mathscr{G}$-mesurable et $\mathscr{H}$-mesurable, et toutes deux intégrables, il résulte de l'indépendance entre $\mathscr{G}$ et $\mathscr{H}$ que

$$\begin{aligned} \mathbb{E}[\mathbb{E}(X \,|\, \mathscr{G})\, \mathbf{1}_{A\cap B}] &= \mathbb{E}[\mathbb{E}(X \,|\, \mathscr{G})\, \mathbf{1}_A] \times \mathbb{E}(\mathbf{1}_B) \\ &= \mathbb{E}(X\, \mathbf{1}_A) \times \mathbb{E}(\mathbf{1}_B)\,. \qquad \text{(Définition 4.1, car } A \in \mathscr{G}) \end{aligned}$$

Les variables aléatoires intégrables $X\, \mathbf{1}_A$ et $\mathbf{1}_B$ sont mesurables par rapport à $\sigma(\sigma(X), \mathscr{G})$ et $\mathscr{H}$, respectivement. Par hypothèse, les tribus $\sigma(\sigma(X), \mathscr{G})$ et $\mathscr{H}$ sont indépendantes. Donc $\mathbb{E}(X\, \mathbf{1}_A) \times \mathbb{E}(\mathbf{1}_B) = \mathbb{E}(X\, \mathbf{1}_A \times \mathbf{1}_B)$, ce qui conduit à $\mathbb{E}[\mathbb{E}(X \,|\, \mathscr{G})\, \mathbf{1}_{A\cap B}] = \mathbb{E}(X\, \mathbf{1}_{A\cap B})$.

Montrons ensuite (b). Plus généralement, montrons que $\mathscr{M} := \{D \in \widetilde{\mathscr{G}} :\; \mathbb{E}[Y\, \mathbf{1}_D] = \mathbb{E}(X\, \mathbf{1}_D)\}$ est une classe monotone si $\widetilde{\mathscr{G}} \subset \mathscr{F}$ est une sous-tribu, et Y une variable aléatoire réelle intégrable avec $\mathbb{E}(Y) = \mathbb{E}(X)$. Sans perte de généralité, on peut supposer $X := 0$ (car sinon, on considèrera $Y - X$).

Il est clair que $\Omega \in \mathscr{M}$, car $\mathbb{E}(Y) = 0$. Si $D_1 \subset D_2$ sont deux éléments de $\mathscr{M}$, alors $\mathbb{E}(Y \mathbf{1}_{D_2 \backslash D_1}) = \mathbb{E}(Y \mathbf{1}_{D_2}) - \mathbb{E}(Y \mathbf{1}_{D_1}) = 0 - 0 = 0$, et donc $D_2 \backslash D_1 \in \mathscr{M}$. Enfin, soit $(D_n,\, n \geq 1)$ une suite croissante d'éléments de $\mathscr{M}$. Soit $D := \cup_{n=1}^\infty D_n$. On a $\mathbf{1}_{D_n} \to \mathbf{1}_D$, et donc $Y \mathbf{1}_{D_n} \to Y \mathbf{1}_D$. Comme $|Y \mathbf{1}_{D_n}| \leq |Y|$ qui est intégrable, il résulte du théorème de convergence dominée que $\mathbb{E}(Y \mathbf{1}_{D_n}) \to \mathbb{E}(Y \mathbf{1}_D)$. Comme $\mathbb{E}(Y \mathbf{1}_{D_n}) = 0$ $(\forall n)$ par définition de D_n, on obtient $\mathbb{E}(Y \mathbf{1}_D) = 0$, c'est-à-dire $D \in \mathscr{M}$. Autrement dit, $\mathscr{M}$ est bien une classe monotone. □

Théorème 4.10. *Soient X et Y deux variables aléatoires, à valeurs dans $(E,\, \mathscr{E})$ et $(\widetilde{E},\, \widetilde{\mathscr{E}})$, respectivement. On suppose que X est indépendante de $\mathscr{G}$, et que Y est $\mathscr{G}$-mesurable. Soit $g : (E \times \widetilde{E},\, \mathscr{E} \otimes \widetilde{\mathscr{E}}) \to (\mathbb{R},\, \mathscr{B}(\mathbb{R}))$ une fonction mesurable, telle que $g(X,\, Y)$ soit intégrable. Alors*

$$\mathbb{E}[g(X,\, Y) \,|\, \mathscr{G}] = h(Y) \qquad \text{p.s.},$$

où $h(y) := \mathbb{E}[g(X,\, y)]$, $y \in \widetilde{E}$.

Preuve. Le théorème de Fubini confirme que $h : (\widetilde{E},\, \widetilde{\mathscr{E}}) \to (\mathbb{R},\, \mathscr{B}(\mathbb{R}))$ est une fonction mesurable, et que $h(Y)$ est intégrable. Comme $h(Y)$ est par hypothèse $\mathscr{G}$-mesurable, il reste de vérifier que $\mathbb{E}[h(Y)\, \mathbf{1}_A] = \mathbb{E}[g(X,\, Y)\, \mathbf{1}_A]$, $\forall A \in \mathscr{G}$.

Montrons, plus généralement, que si Z est une variable aléatoire réelle bornée[12] et $\mathscr{G}$-mesurable, alors $\mathbb{E}[h(Y)\, Z] = \mathbb{E}[g(X,\, Y)\, Z]$.

Soit $P_{(X,Y,Z)}$ la loi de la variable aléatoire $(X,\, Y,\, Z)$ (à valeurs dans $E \times \widetilde{E} \times \mathbb{R}$). On a

$$\mathbb{E}[g(X,\, Y)\, Z] = \int_{E \times \widetilde{E} \times \mathbb{R}} g(x,\, y)\, P_{(X,Y,Z)}(\mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z)\,.$$

Par hypothèse, X et $(Y,\, Z)$ sont indépendantes (car $(Y,\, Z)$ est $\mathscr{G}$-mesurable tandis que X est indépendante de $\mathscr{G}$). Donc $P_{(X,Y,Z)} = P_X \otimes P_{(Y,Z)}$, ce qui nous conduit à :

$$\mathbb{E}[g(X,\, Y)\, Z] = \int_{E \times \widetilde{E} \times \mathbb{R}} g(x,\, y)\, z\, P_X(\mathrm{d}x)\, P_{(Y,Z)}(\mathrm{d}y\, \mathrm{d}z)\,.$$

Le théorème de Fubini-Lebesgue nous dit que ceci est

$$= \int_{\widetilde{E} \times \mathbb{R}} \Big(\int_E g(x,\, y)\, z\, P_X(\mathrm{d}x) \Big)\, P_{(Y,Z)}(\mathrm{d}y\, \mathrm{d}z) = \int_{\widetilde{E} \times \mathbb{R}} h(y)\, z\, P_{(Y,Z)}(\mathrm{d}y\, \mathrm{d}z)\,,$$

qui n'est autre que $\mathbb{E}[h(Y)\, Z]$. □

[12] Bornitude : On dit que X est une variable aléatoire réelle bornée s'il existe une constante $C < \infty$ *ne dépendant pas de* ω telle que $|X(\omega)| \leq C$, $\forall \omega \in \Omega$. Attention à ne pas confondre avec la finitude : X est une variable aléatoire réelle finie si $X(\omega) \in \mathbb{R}$, $\forall \omega \in \Omega$ (autrement dit, X est une variable aléatoire réelle tout court !).

Remarque 4.11. Le sens du théorème est intuitivement clair. Si l'on cherche la valeur de $\mathbb{E}[g(X,\, Y)\,|\,\mathscr{G}]$, comme Y est $\mathscr{G}$-mesurable, on peut la considérer comme une constante, disons y. Dans ce cas, $\mathbb{E}[g(X,\, Y)\,|\,\mathscr{G}]$ devient $\mathbb{E}[g(X,\, y)]$ (car $g(X,\, y)$ est indépendante de $\mathscr{G}$), qui n'est autre que $h(y)$. Bien sûr, y est en réalité Y. En remplaçant y par Y dans $h(y)$, on tombe sur $h(Y)$ comme la valeur de $\mathbb{E}[g(X,\, Y)\,|\,\mathscr{G}]$. □

Certaines inégalités générales que l'on a vues en théorie des probabilités élémentaire (Jensen, Cauchy–Schwarz, Hölder) restent valables si l'on remplace espérance par espérance conditionnelle, avec la même preuve qui s'appuie sur deux propriétés satisfaites à la fois par espérance et par espérance conditionnelle : si X et Y sont des variables aléatoires réelles intégrables, et si a et b sont des réels, alors (i) linéarité : $\mathbb{E}(aX + bY\,|\,\mathscr{G}) = a\,\mathbb{E}(X\,|\,\mathscr{G}) + b\,\mathbb{E}(Y\,|\,\mathscr{G})$ p.s. ; (ii) préservation d'ordre : $X \le Y$ p.s. $\Rightarrow \mathbb{E}(X\,|\,\mathscr{G}) \le \mathbb{E}(Y\,|\,\mathscr{G})$ p.s.

Théorème 4.12. (Inégalité de Jensen). *Soit X une variable aléatoire réelle, et soit φ une fonction convexe. Si X et $\varphi(X)$ sont intégrables, alors*

$$\varphi\Big(\mathbb{E}[X\,|\,\mathscr{G}]\Big) \le \mathbb{E}[\,\varphi(X)\,|\,\mathscr{G}\,], \qquad \text{p.s.}$$

En particulier, $\varphi(\mathbb{E}(X)) \le \mathbb{E}[\,\varphi(X)\,]$.

Théorème 4.13. (Inégalité de Hölder). *Soient X et Y des variables aléatoires réelles, et soient $p > 1$ et $q > 1$ des réels tels que $\frac{1}{p} + \frac{1}{q} = 1$. Si $\mathbb{E}(|X|^p) < \infty$ et $\mathbb{E}(|Y|^q) < \infty$, alors XY est intégrable, et*

$$|\mathbb{E}(XY\,|\,\mathscr{G})| \le [\mathbb{E}(|X|^p\,|\,\mathscr{G})]^{1/p}\,[\mathbb{E}(|Y|^q\,|\,\mathscr{G})]^{1/q}\,, \qquad \text{p.s.}$$

Corollaire 4.14. (Inégalité de Cauchy–Schwarz). *Soient X et Y des variables aléatoires réelles telles que $\mathbb{E}(X^2) < \infty$ et que $\mathbb{E}(Y^2) < \infty$, alors XY est intégrable, et*

$$|\mathbb{E}(XY\,|\,\mathscr{G})| \le [\mathbb{E}(X^2\,|\,\mathscr{G})]^{1/2}\,[\mathbb{E}(Y^2\,|\,\mathscr{G})]^{1/2}\,, \qquad \text{p.s.}$$

4.2. Cas des variables aléatoires positives

Soit X une variable aléatoire à valeurs dans $[0,\, \infty]$, et soit $\mathscr{G} \subset \mathscr{F}$ une sous-tribu. On définit[13]

$$\mathbb{E}(X\,|\,\mathscr{G}) := \lim_{n\to\infty} \uparrow \mathbb{E}(X \wedge n\,|\,\mathscr{G}), \qquad \text{p.s.}$$

On voit que $\mathbb{E}(X\,|\,\mathscr{G})$ est une variable aléatoire $\mathscr{G}$-mesurable, à valeurs dans $[0,\, \infty]$.

[13]Notation : $a \wedge b := \min\{a,\, b\}$.

Dans le cas où X est à valeurs dans $[0, \infty[$ et intégrable, le théorème de convergence monotone nous assure que $\mathbb{E}[\mathbb{E}(X \,|\, \mathscr{G})\, \mathbf{1}_A] = \mathbb{E}(X\, \mathbf{1}_A)$, $\forall A \in \mathscr{G}$. [En particulier, $\mathbb{E}[\mathbb{E}(X \,|\, \mathscr{G})] = \mathbb{E}(X) < \infty$.] Autrement dit, $\mathbb{E}(X \,|\, \mathscr{G})$ coïncide avec l'espérance conditionnelle que l'on a définie précédemment.

On voit que $\mathbb{E}(X \,|\, \mathscr{G})$ est l'unique (au sens p.s.) variable aléatoire Y à valeurs dans $[0, \infty]$ qui satisfasse les conditions suivantes :

(i) $\mathbb{E}(X \,|\, \mathscr{G})$ est $\mathscr{G}$-mesurable ;

(ii) $\forall A \in \mathscr{G}$, $\mathbb{E}[\mathbb{E}(X \,|\, \mathscr{G})\, \mathbf{1}_A] = \mathbb{E}(X\, \mathbf{1}_A)$.

En effet, l'existence est une conséquence immédiate du théorème de convergence monotone, tandis que l'unicité se prouve avec exactement le même argument que dans le cas des variables aléatoires intégrables.

La plupart des résultats que l'on a démontrés pour les variables aléatoires intégrables restent valables pour les variables aléatoires positives, avec une preuve similaire.

Propriété 4.15. *Si X et Y sont des variables aléatoires à valeurs dans $[0, \infty]$,*

- $\mathbb{E}[\, \mathbb{E}(X \,|\, \mathscr{G}) \,] = \mathbb{E}(X)$ *;*
- $\mathbb{E}(X \,|\, \mathscr{G}) = X$ p.s., *si X est $\mathscr{G}$-mesurable ;*
- $\mathbb{E}(X \,|\, \mathscr{G}) = \mathbb{E}(X)$ p.s., *si X est indépendante de $\mathscr{G}$;*
- $\mathbb{E}(aX + bY \,|\, \mathscr{G}) = a\, \mathbb{E}(X \,|\, \mathscr{G}) + b\, \mathbb{E}(Y \,|\, \mathscr{G})$ p.s. *si $a \geq 0$ et $b \geq 0$;*
- $\mathbb{E}(XY \,|\, \mathscr{G}) = Y\, \mathbb{E}(X \,|\, \mathscr{G})$ p.s. *si Y est, en plus, $\mathscr{G}$-mesurable.*
- $\mathbb{E}[\, \mathbb{E}(X \,|\, \mathscr{G}) \,|\, \mathscr{H} \,] = \mathbb{E}(X \,|\, \mathscr{H})$ p.s. *si $\mathscr{H} \subset \mathscr{G} \subset \mathscr{F}$ sont des sous-tribus.*

Remarque 4.16. On peut avoir $X < \infty$ p.s., mais $\mathbb{P}[\mathbb{E}(X \,|\, \mathscr{G}) = \infty] > 0$. C'est le cas, par exemple, pour $\mathscr{G} := \{\varnothing, \Omega\}$ et X une variable aléatoire à valeurs dans $\mathbb{R}_+$ avec $\mathbb{E}(X) = \infty$. □

Théorème 4.17. (Théorème de convergence monotone). *Si $(X_n,\, n \geq 1)$ est une suite de variables aléatoires à valeurs dans $[0, \infty]$, alors*

$$\mathbb{E}(X \,|\, \mathscr{G}) = \lim_{n\to\infty} \uparrow \mathbb{E}(X_n \,|\, \mathscr{G}), \qquad \text{p.s.},$$

où $X := \lim_{n\to\infty} \uparrow X_n$.

Preuve. La monotonie p.s. de $n \mapsto \mathbb{E}(X_n \,|\, \mathscr{G})$ garantit l'existence de la limite p.s. de $Y := \lim_{n\to\infty} \uparrow \mathbb{E}(X_n \,|\, \mathscr{G})$. Par définition, Y est une variable aléatoire à valeurs dans $[0, \infty]$ et $\mathscr{G}$-mesurable.

Soit $A \in \mathscr{G}$. Observons que

$$\mathbb{E}[Y\, \mathbf{1}_A] = \lim_{n\to\infty} \uparrow \mathbb{E}[\mathbb{E}(X_n \,|\, \mathscr{G})\, \mathbf{1}_A] = \lim_{n\to\infty} \uparrow \mathbb{E}[X_n\, \mathbf{1}_A] = \mathbb{E}[X\, \mathbf{1}_A]\,,$$

où l'on a appliqué le théorème usuel de convergence monotone pour obtenir les première et troisième identités. Autrement dit, $Y = \mathbb{E}(X \,|\, \mathscr{G})$ p.s. □

Exemple 4.18. Soient $X_1,\ X_2,\ \cdots$ des variables aléatoires à valeurs dans $[0,\ \infty]$. Alors $\mathbb{E}(\sum_{n=1}^\infty X_n \,|\, \mathscr{G}) = \sum_{n=1}^\infty \mathbb{E}(X_n \,|\, \mathscr{G})$ p.s.

Il suffit d'appliquer le Théorème 4.17 à $(\sum_{i=1}^n X_i,\ n \ge 1)$, qui est une suite croissante de variables aléatoires à valeurs dans $[0,\ \infty]$. □

Théorème 4.19. (Lemme de Fatou). *Si* $(X_n,\ n \ge 1)$ *est une suite de variables aléatoires à valeurs dans* $[0,\ \infty]$, *alors*

$$\mathbb{E}\Big(\liminf_{n\to\infty} X_n \,\Big|\, \mathscr{G}\Big) \le \liminf_{n\to\infty} \mathbb{E}(X_n \,|\, \mathscr{G}), \qquad \text{p.s.}$$

Preuve. Par définition, $\liminf_{n\to\infty} X_n = \lim_{k\to\infty} \uparrow \inf_{n\ge k} X_n$. Par le Théorème 4.17,

$$\mathbb{E}\Big(\liminf_{n\to\infty} X_n \,\Big|\, \mathscr{G}\Big) = \mathbb{E}\Big(\lim_{k\to\infty} \uparrow \inf_{n\ge k} X_n \,\Big|\, \mathscr{G}\Big) = \lim_{k\to\infty} \uparrow \mathbb{E}\Big(\inf_{n\ge k} X_n \,\Big|\, \mathscr{G}\Big), \qquad \text{p.s.}$$

Pour tout k, $\inf_{n\ge k} X_n \le X_k$. Donc $\mathbb{E}(\inf_{n\ge k} X_n \,|\, \mathscr{G}) \le \mathbb{E}(X_k \,|\, \mathscr{G})$ p.s. D'où l'inégalité cherchée. □

On termine ce paragraphe avec un résultat qui traite le cas des variables aléatoires intégrables de signe quelconque.

Théorème 4.20. (Théorème de convergence dominée). *Soient* $Z,\ X_1,\ X_2,\ \cdots$ *des variables aleatoires réelles. On suppose que* $X_n \to X_\infty$ p.s., *et que pour tout* n, $|X_n| \le Z$ p.s. *Si* $\mathbb{E}(Z) < \infty$, *alors* X_∞ *est intégrable, et* $\mathbb{E}(|X_n - X_\infty| \,|\, \mathscr{G}) \to 0$ p.s.

En particulier, $\mathbb{E}(X_n \,|\, \mathscr{G}) \to \mathbb{E}(X_\infty \,|\, \mathscr{G})$ p.s.

Preuve. Par hypothèse, $|X_\infty| \le Z$, et donc X_∞ est intégrable. Posons $Y_n := 2Z - |X_\infty - X_n|$, de sorte que $0 \le Y_n \le 2Z$ p.s. et que $Y_n \to 2Z$ p.s. Par le lemme de Fatou conditionnel, $\mathbb{E}(\liminf_{n\to\infty} Y_n \,|\, \mathscr{G}) \le \liminf_{n\to\infty} \mathbb{E}(Y_n \,|\, \mathscr{G})$ p.s., c'est-à-dire

$$\mathbb{E}(2Z \,|\, \mathscr{G}) \le \mathbb{E}(2Z \,|\, \mathscr{G}) - \limsup_{n\to\infty} \mathbb{E}(|X_\infty - X_n| \,|\, \mathscr{G})\,, \qquad \text{p.s.,}$$

ce qui implique que $\limsup_{n\to\infty} \mathbb{E}(|X_\infty - X_n| \,|\, \mathscr{G}) \le 0$ p.s. Donc $\mathbb{E}(|X_n - X_\infty| \,|\, \mathscr{G}) \to 0$ p.s.

Comme $|\mathbb{E}(X_n \,|\, \mathscr{G}) - \mathbb{E}(X_\infty \,|\, \mathscr{G})| \le \mathbb{E}(|X_n - X_\infty| \,|\, \mathscr{G})$, ceci entraîne que $\mathbb{E}(X_n \,|\, \mathscr{G}) \to \mathbb{E}(X_\infty \,|\, \mathscr{G})$ p.s. □

Exemple 4.21. Soient X_1, X_2, $\cdots$ des variables aléatoires réelles telles que $\sum_n \mathbb{E}(|X_n|) < \infty$. Alors $\sum_{n=1}^{\infty} X_n$ est intégrable, et $\mathbb{E}(\sum_{n=1}^{\infty} X_n \,|\, \mathscr{G}) = \sum_{n=1}^{\infty} \mathbb{E}(X_n \,|\, \mathscr{G})$ p.s.

En effet, posons $Z := \sum_{n=1}^{\infty} |X_n|$; le théorème de Fubini–Tonelli nous dit que $\mathbb{E}(Z) = \sum_{n=1}^{\infty} \mathbb{E}(|X_n|) < \infty$. Il suffit alors d'appliquer le Théorème 4.20 à la suite $(\sum_{i=1}^{n} X_i,\, n \geq 1)$ de variables aléatoires réelles. □

5. Conditionnement par rapport à une variable aléatoire

Lorsque la tribu est engendrée par une variable aléatoire (pas nécessairement à valeurs réelles ou dans un espace euclidien), l'espérance conditionnelle a des propriétés particulières.

Sauf mention du contraire, X désignera dans cette section une variable aléatoire réelle intégrable.

Notation. Si $\mathscr{G}$ est engendrée par une variable aléatoire Z à valeurs dans un espace mesurable $(E, \mathscr{E})$ quelconque, alors on écrira $\mathbb{E}(X \,|\, Z)$ au lieu de $\mathbb{E}(X \,|\, \mathscr{G})$. Si $\mathscr{G} = \sigma(Z_i,\, i \in I)$, on écrira aussi $\mathbb{E}(X \,|\, Z_i,\, i \in I)$.

Soit Z une variable aléatoire à valeurs dans $(E, \mathscr{E})$. Soit $A \in \sigma(Z)$. Par définition, il existe $B \in \mathscr{E}$ telle que $A = \{\omega \in \Omega : \, Z(\omega) \in B\}$; autrement dit, $\mathbf{1}_A = \mathbf{1}_B(Z)$. Donc toute fonction étagée qui est $\sigma(Z)$-mesurable s'écrit comme une fonction étagée ($\mathscr{E}$-mesurable) de Z. Par un passage à la limite, si ξ est une variable aléatoire réelle intégrable et $\sigma(Z)$-mesurable, elle s'écrit comme $\xi = h(Z)$, où $h : \, E \to \mathbb{R}$ est une fonction mesurable ($\mathbb{R}$ muni de la tribu borélienne $\mathscr{B}(\mathbb{R})$). Prenons $\xi = \mathbb{E}(X \,|\, Z)$, et on obtient

$$\mathbb{E}(X \,|\, Z) = h(Z), \qquad \text{p.s.}$$

L'unicité de l'espérance conditionnelle nous assure que si $\widetilde{h}$ est une fonction borélienne telle que $\mathbb{E}(X \,|\, Z) = \widetilde{h}(Z)$, alors $\widetilde{h} = h$, P_Z-presque partout.

Exemple 5.1. Soient X et Y deux variables aléatoires réelles indépendantes et identiquement distribuées (i.i.d.), telles que $\mathbb{E}(|X|) < \infty$. Alors $\mathbb{E}(X \,|\, X+Y) = \frac{X+Y}{2}$ p.s.

En fait, si $P_{(X,Z)} = P_{(Y,Z)}$, alors $\mathbb{E}(X \,|\, Z) = \mathbb{E}(Y \,|\, Z)$ p.s. Pour voir cela, on observe d'abord que $\mathbb{E}(X \,|\, Z)$ est une variable aléatoire intégrable et $\sigma(Z)$-mesurable. De plus, pour tout $A \in \sigma(Z)$, $\mathbb{E}(\mathbb{E}(X \,|\, Z)\, \mathbf{1}_A) = \mathbb{E}(X\, \mathbf{1}_A)$. On sait qu'il existe une fonction mesurable $h : \, E \to \mathbb{R}$ telle que $\mathbf{1}_A = h(Z)$, ce qui nous donne $\mathbb{E}(X\, \mathbf{1}_A) = \mathbb{E}(X h(Z)) = \mathbb{E}(Y h(Z)) = \mathbb{E}(Y\, \mathbf{1}_A)$. D'où $\mathbb{E}[\,\mathbb{E}(X \,|\, Z)\, \mathbf{1}_A\,] = \mathbb{E}(Y\, \mathbf{1}_A)$. Par conséquent, $\mathbb{E}(X \,|\, Z) = \mathbb{E}(Y \,|\, Z)$ p.s.

Pour revenir à l'exemple $Z := X+Y$, il suffit de remarquer que $P_{(X,X+Y)} = P_{(Y,X+Y)}$, car $P_{(X,Y)} = P_{(Y,X)}$. Donc $\mathbb{E}(X \,|\, X+Y) = \mathbb{E}(Y \,|\, X+Y)$. Or, $\mathbb{E}(X+Y \,|\, X+Y) = X+Y$, on obtient $\mathbb{E}(X \,|\, X+Y) = \frac{X+Y}{2}$ p.s.

[Plus généralement, avec le même argument, on montre que si $X_1, \cdots, X_n$ sont des variables aléatoires réelles i.i.d. intégrables, alors $\mathbb{E}(X_1 \,|\, \sum_{i=1}^n X_i) = \frac{1}{n}\sum_{i=1}^n X_i$ p.s.] $\square$

Exemple 5.2. Soient N, X_1, X_2, $\cdots$ des variables aléatoires réelles indépendantes, admettant toutes un moment d'ordre 1. On suppose que les X_i suivent la même loi, et que N est à valeurs dans $\mathbb{N} := \{0, 1, 2, \cdots\}$. Définissons

$$Y(\omega) := \sum_{i=1}^{N(\omega)} X_i(\omega)\,,$$

avec la notation $\sum_{i=1}^0 := 0$. On cherche à étudier le moment d'ordre 1 de Y.

Il est clair que Y est bien une variable aléatoire : $Y(\omega) := \sum_{n=1}^\infty \sum_{i=1}^n X_i(\omega)\,\mathbf{1}_{\{N(\omega)=n\}}$, qui est représentée comme somme de variables aléatoires réelles. Il est également clair que Y est intégrable, car par le théorème de Fubini–Tonelli,

$$\begin{aligned}
\mathbb{E}\Big(\sum_{n=1}^\infty \sum_{i=1}^n |X_i|\,\mathbf{1}_{\{N=n\}}\Big) &= \sum_{n=1}^\infty \sum_{i=1}^n \mathbb{E}(|X_i|\,\mathbf{1}_{\{N=n\}}) \\
&= \sum_{n=1}^\infty \sum_{i=1}^n \mathbb{E}(|X_i|)\,\mathbb{P}(N=n) \\
&= \mathbb{E}(|X_1|) \sum_{n=1}^\infty n\,\mathbb{P}(N=n) = \mathbb{E}(|X_1|)\,\mathbb{E}(N) < \infty.
\end{aligned}$$

Le même argument nous donne en fait $\mathbb{E}(Y) = \mathbb{E}(X_1)\mathbb{E}(N)$, en utilisant cette fois-ci le théorème de Fubini–Lebesgue.

Calculons $\mathbb{E}(Y)$. On a $\mathbb{E}(Y \,|\, N) = \sum_{n=1}^\infty \mathbb{E}(\sum_{i=1}^n X_i\,\mathbf{1}_{\{N=n\}} \,|\, N)$ p.s., d'après l'Exemple 4.21. On a $\mathbb{E}(\sum_{i=1}^n X_i\,\mathbf{1}_{\{N=n\}} \,|\, N) = \sum_{i=1}^n \mathbb{E}(X_i\,\mathbf{1}_{\{N=n\}} \,|\, N)$.

Comme $\mathbf{1}_{\{N=n\}}$ est $\sigma(N)$-mesurable, il résulte du Théorème 4.8 que $\mathbb{E}(X_i\,\mathbf{1}_{\{N=n\}} \,|\, N) = \mathbf{1}_{\{N=n\}}\,\mathbb{E}(X_i \,|\, N)$ p.s., et puisque X_i est indépendante de $\sigma(N)$, on a $\mathbb{E}(X_i \,|\, N) = \mathbb{E}(X_i) = \mathbb{E}(X_1)$ p.s. D'où : $\mathbb{E}(X_i\,\mathbf{1}_{\{N=n\}} \,|\, N) = \mathbb{E}(X_1)\,\mathbf{1}_{\{N=n\}}$ p.s., et donc $\mathbb{E}(\sum_{i=1}^n X_i\,\mathbf{1}_{\{N=n\}} \,|\, N) = \sum_{i=1}^n \mathbb{E}(X_1)\,\mathbf{1}_{\{N=n\}} = n\,\mathbb{E}(X_1)\,\mathbf{1}_{\{N=n\}}$. Par conséquent, $\mathbb{E}(Y \,|\, N) = \sum_{n=1}^\infty n\,\mathbb{E}(X_1)\,\mathbf{1}_{\{N=n\}} = N\,\mathbb{E}(X_1)$ p.s.

Par le Théorème 4.6, ceci donne alors $\mathbb{E}(Y) = \mathbb{E}(N\,\mathbb{E}(X_1)) = \mathbb{E}(N)\,\mathbb{E}(X_1)$. $\square$

6. Espérance conditionnelle et projection orthogonale

Lorsque X est une variable aléatoire réelle admettant un moment d'ordre 2, il y a une interprétation géométrique pour représenter $\mathbb{E}(X \,|\, \mathscr{G})$.

Théorème 6.1. *Si $\mathbb{E}(X^2) < \infty$, alors $\mathbb{E}(X \,|\, \mathscr{G})$ est la variable Y, $\mathscr{G}$-mesurable et admettant un moment d'ordre 2, qui minimise "l'erreur quadratique" $\mathbb{E}[(X-Y)^2]$.*

Remarque 6.2. (i) Soit $L^2(\Omega, \mathscr{F}, \mathbb{P}) := \{Y : \mathbb{E}(Y^2) < \infty\}$ qui, muni du produit scalaire $\langle Y_1, Y_2 \rangle := \mathbb{E}(Y_1 Y_2)$, est un espace de Hilbert.[14] Rappelons que $L^2(\Omega, \mathscr{G}, \mathbb{P}) := \{Y : Y \text{ est } \mathscr{G}\text{-mesurable}, \ \mathbb{E}(Y^2) < \infty\}$ est un sous-espace fermé. Sous la condition $\mathbb{E}(X^2) < \infty$, le Théorème 6.1 nous confirme que $\mathbb{E}(X \,|\, \mathscr{G})$ est la projection orthogonale de X sur $L^2(\Omega, \mathscr{G}, \mathbb{P})$, c'est-à-dire l'élément dans $L^2(\Omega, \mathscr{G}, \mathbb{P})$ qui est le plus proche de X.

(ii) Dans la cas particulier où $\mathscr{G} = \sigma(Z)$, le Théorème 6.1 s'énonce comme suit: la fonction mesurable f minimisant la quantité $\mathbb{E}[(X - f(Z))^2]$ est telle que $f(Z) = \mathbb{E}(X \,|\, Z)$. Si Z est une constante, on retombe sur le fait bien connu que la constante c minimisant $\mathbb{E}[(X-c)^2]$ est $c = \mathbb{E}(X)$. $\square$

Preuve du Théorème 6.1. Puisque $\mathbb{E}(X^2) < \infty$, l'inégalité de Jensen (Théorème 4.12) nous dit que $\mathbb{E}(X \,|\, \mathscr{G})$ est un élément de $L^2(\Omega, \mathscr{G}, \mathbb{P})$. Il suffit donc de montrer que $X - \mathbb{E}(X \,|\, \mathscr{G})$ est orthogonale de $L^2(\Omega, \mathscr{G}, \mathbb{P})$. Soit $Z \in L^2(\Omega, \mathscr{G}, \mathbb{P})$. On sait que XZ admet un moment d'ordre 1, et par le Théorème 4.8, $\mathbb{E}(ZX \,|\, \mathscr{G}) = Z\,\mathbb{E}(X \,|\, \mathscr{G})$. Donc

$$\mathbb{E}(ZX) = \mathbb{E}\Big[\mathbb{E}(ZX \,|\, \mathscr{G})\Big] = \mathbb{E}\Big[Z\,\mathbb{E}(X \,|\, \mathscr{G})\Big].$$

Autrement dit, $\mathbb{E}[Z(X - \mathbb{E}(X \,|\, \mathscr{G}))] = 0$. $\square$

7. Exemples de calculs

On traite trois exemples de calcul d'espérance conditionnelle $\mathbb{E}(X \,|\, Z)$, où X est une variable aléatoire réelle intégrable, et Z est une variable aléatoire à valeurs dans un espace mesurable E. Rappelons que $\mathbb{E}(X \,|\, Z) = h(Z)$ p.s., où $h : E \to \mathbb{R}$ est une fonction mesurable.

Dans le premier exemple, Z est une variable aléatoire discrète, à savoir que E est un espace dénombrable[15].

[14] Bien entendu, on prend plutôt les classes d'équivalence dans $L^2(\Omega, \mathscr{F}, \mathbb{P})$, en identifiant les variables aléatoires qui ne diffèrent que sur un ensemble de probabilité nulle.

[15] C'est-à-dire fini, ou infini dénombrable.

Dans les deux derniers exemples, Z est une variable aléatoires à valeurs dans l'espace euclidien $\mathbb{R}^n$ (avec $n \geq 1$), et l'on suppose que (X, Z) est un vecteur aléatoire à densité (dans le deuxième exemple) ou un vecteur aléatoire gaussien (dans le troisième exemple).

7.1. Conditionnement discret

Soit X une variable aléatoire réelle intégrable. Soit Z une variable aléatoire à valeurs dans un espace dénombrable E, muni de la tribu $\mathscr{P}(E)$ de toutes les parties de E (y compris $\varnothing$ et E).

Théorème 7.1. *Si X est intégrable, et si Z est variable aléatoire à valeurs dans un espace dénombrable E, alors*

$$\mathbb{E}(X \mid Z) = h(Z), \qquad \text{p.s.},$$

où[16]

$$h(z) := \frac{\mathbb{E}(X\, \mathbf{1}_{\{Z=z\}})}{\mathbb{P}(Z=z)} \,,$$

pour tout $z \in E$ tel que $\mathbb{P}(Z = z) > 0$.

Preuve. Soit $h : E \to \mathbb{R}$ la fonction (mesurable) définie comme dans l'énoncé du théorème. Comme $\mathbb{E}(|h(Z)|) = \sum_{z\in E} |\mathbb{E}(X\, \mathbf{1}_{\{Z=z\}})| \leq \sum_{z\in E} \mathbb{E}(|X|\, \mathbf{1}_{\{Z=z\}}) = \mathbb{E}(|X|) < \infty$, on voit que la variable aléatoire réelle $h(Z)$ est intégrable.

Pour tout $A \in \sigma(Z)$, il existe $B \in \mathscr{P}(E)$ tel que $A = Z^{-1}(B)$ et donc $\mathbf{1}_A = \mathbf{1}_B(Z)$. On a alors

$$\mathbb{E}[h(Z)\, \mathbf{1}_A] = \mathbb{E}[h(Z)\, \mathbf{1}_B(Z)] = \sum_{z\in E} h(z)\, \mathbf{1}_B(z)\, \mathbb{P}(Z=z) = \sum_{z\in E} \mathbb{E}(X\, \mathbf{1}_{\{Z=z\}})\, \mathbf{1}_B(z) \,,$$

qui n'est autre que $\mathbb{E}(X\, \mathbf{1}_B(Z)] = \mathbb{E}(X\, \mathbf{1}_A)$. Par la Définition 4.1, $h(Z) = \mathbb{E}(X \mid Z)$ p.s. $\square$

Remarque 7.2. Lorsque Z est une variable aléatoire discrète à valeurs dans un espace mesurable dénombrable E, pour tout $z \in E$ tel que $\mathbb{P}(Z = z) > 0$, on écrit souvent

$$\mathbb{E}(X \mid Z = z) := \frac{\mathbb{E}(X\, \mathbf{1}_{\{Z=z\}})}{\mathbb{P}(Z=z)} \,.$$

Le Théorème 7.1 nous dit que p.s.,

$$[\mathbb{E}(X \mid Z)](\omega) = \mathbb{E}(X \mid Z = z), \qquad \text{si } Z(\omega) = z \,.$$

Il y a donc une certaine homogénéité dans les notations. $\square$

[16] La valeur de $h(z)$ pour $z \in Z$ tel que $\mathbb{P}(Z = z) = 0$ n'a aucune importance ; on peut par exemple prendre $h(z) := 0$ dans ce cas.

7.2. Cas des vecteurs aléatoires à densité

Soit $(X, Z_1, \cdots, Z_n)$ un vecteur aléatoire à valeurs dans $\mathbb{R}^{n+1}$. On suppose que X est intégrable, et que la loi de $(X, Z_1, \cdots, Z_n)$ est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}^{n+1}$. Pour simplifier les notations, on écrit (X, Z) à la place de $(X, Z_1, \cdots, Z_n)$, et Z au lieu de $(Z_1, \cdots, Z_n)$.

Soit $f_{(X,Z)}$ la densité. On sait que la loi de Z est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$, et qu'elle possède comme densité

$$f_Z(z) := \int_{\mathbb{R}} f_{(X,Z)}(x, z)\,\mathrm{d}x\,, \qquad z \in \mathbb{R}^n\,.$$

Théorème 7.3. *Soit (X, Z) un vecteur aléatoire à valeurs dans $\mathbb{R}^{n+1}$ qui admet une densité $f_{(X, Z_1, \cdots, Z_n)}$. On suppose que $\mathbb{E}(|X|) < \infty$. On pose*[17]

$$\begin{aligned} f_{X|Z=z}(x) &:= \frac{f_{(X,Z)}(x, z)}{f_Z(z)}, \qquad x \in \mathbb{R}, \\ h(z) &:= \int_{\mathbb{R}} x\, f_{X|Z=z}(x)\,\mathrm{d}x, \qquad z \in \mathbb{R}^n\,. \end{aligned}$$

Alors

$$\mathbb{E}(X \mid Z) = h(Z), \qquad \text{p.s.}$$

Preuve. Il est clair, d'après le théorème de Fubini, que $h(Z)$ est une variable aléatoire intégrable. Soit $A \in \sigma(Z)$. Il s'agit de prouver que $\mathbb{E}[\,h(Z)\,\mathbf{1}_A\,] = \mathbb{E}[\,X\,\mathbf{1}_A\,]$.

Comme $A \in \sigma(Z)$, on peut écrire $A = \{\omega : \; Z(\omega) \in B\}$, où $B \in \mathscr{B}(\mathbb{R}^n)$ est une partie borélienne de $\mathbb{R}^n$. On a alors

$$\begin{aligned} \mathbb{E}[\,X\,\mathbf{1}_A\,] &= \mathbb{E}[\,X\,\mathbf{1}_B(Z)\,] \\ &= \int_{\mathbb{R}} \mathrm{d}x \int_B \mathrm{d}z\; x\, f_{(X,Z)}(x, z) \\ &= \int_B \mathrm{d}z\; h(z) f_Z(z) \\ &= \mathbb{E}[\,h(Z)\,\mathbf{1}_B(Z)\,]. \end{aligned}$$

D'où le résultat désiré. □

[17] La valeur de $f_{X|Z=z}$ pour $z \in \mathbb{R}^n$ tel que $f_Z(z) = 0$ n'a aucune importance ; on peut prendre, par exemple, $f_{X|Z=z}(x) := 0$, $\forall x \in \mathbb{R}$, dans ce cas.

Remarque 7.4. On observe que

$$\begin{aligned} f_{X|Z=z}(x) &= \lim_{\mathrm{d}x\to 0,\,\mathrm{d}z\to 0} \frac{\mathbb{P}(X\in[x,x+\,\mathrm{d}x],\ Z\in[z,z+\,\mathrm{d}z])/(\mathrm{d}x\,\mathrm{d}z)}{\mathbb{P}(Z\in[z,z+\,\mathrm{d}z])/\,\mathrm{d}z} \\ &= \lim_{\mathrm{d}x\to 0,\,\mathrm{d}z\to 0} \mathbb{P}\Big(X\in[x,x+\,\mathrm{d}x]\,|\,Z\in[z,z+\,\mathrm{d}z]\Big)/\,\mathrm{d}x\,. \end{aligned}$$

C'est pour cette raison que $f_{X|Z=z}$ est souvent appelée "densité conditionnelle de X sachant $Z=z$" (il s'agit bien sûr d'une fonction de densité associée à une loi de probabilité sur $\mathbb{R}$). Dans la littérature, on trouve de temps en temps l'écriture[18] $\mathbb{E}(X\,|\,Z=z)$ pour désigner $h(z)$. □

Exemple 7.5. Soient X et Y deux variables aléatoires réelles indépendantes suivant la même loi gaussienne $\mathscr{N}(0,\,1)$. Soit $Z=X+Y$.

Il est facile de calculer la loi de (X,Z). Pour toute fonction borélienne bornée $\varphi:\ \mathbb{R}^2\to\mathbb{R}$, on a

$$\begin{aligned} \mathbb{E}[h(X,Z)] &= \int_{\mathbb{R}}\int_{\mathbb{R}} h(x,x+y)\frac{1}{2\pi}\mathrm{e}^{-(x^2+y^2)/2}\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} h(x,z)\frac{1}{2\pi}\mathrm{e}^{-(x^2+(z-x)^2)/2}\,\mathrm{d}x\,\mathrm{d}z. \end{aligned}$$

Donc (X,Z) admet une densité qui vaut

$$f_{(X,Z)}(x,z)=\frac{1}{2\pi}\mathrm{e}^{-(x^2+(z-x)^2)/2},\qquad (x,z)\in\mathbb{R}^2.$$

D'autre part, $P_Z=\mathscr{N}(0,\,2)$, donc $f_Z(z)=\frac{1}{2\pi^{1/2}}\mathrm{e}^{-z^2/4}$. Posons, pour chaque $z\in\mathbb{R}$,

$$f_{X|Z=z}(x)=\frac{f_{(X,Z)}(x,z)}{f_Z(z)}=\frac{1}{\pi^{1/2}}\mathrm{e}^{-(x^2+(z-x)^2)/2+z^2/4}=\frac{1}{\pi^{1/2}}\mathrm{e}^{-(x-z/2)^2}.$$

La fonction $f_{X|Z=z}$ n'est autre que la densité de la loi gaussienne $\mathscr{N}(\frac{z}{2},\,\frac12)$ (d'où vient une formulation du genre "sachant $Z=z$, X suit la loi gaussienne $\mathscr{N}(\frac{z}{2},\,\frac12)$"). Posons

$$h(z)=\int_{\mathbb{R}} x f_{X|Z=z}(x)\,\mathrm{d}x=\frac{z}{2}\,.$$

D'après le Théorème 7.3, $\mathbb{E}(X\,|\,Z)=h(Z)=\frac{Z}{2}$. On arrive donc à la même conclusion que dans l'Exemple 5.1. □

[18] C'est une écriture informelle pour nous, qui peut être rendue rigoureuse. Cela dépasse cependant le cadre de notre cours.

7.3. Cas des vecteurs aléatoires gaussiens

Soit $(X, Z_1, \cdots, Z_n)$ un vecteur aléatoire à valeurs dans $\mathbb{R}^{n+1}$. On suppose que X est intégrable. Alors la variable aléatoire réelle $\mathbb{E}(X \,|\, Z_1, \cdots, Z_n)$, étant $\sigma(Z_1, \cdots, Z_n)$-mesurable, s'écrit nécessairement comme $h(Z_1, \cdots, Z_n)$, où $h : \mathbb{R}^n \to \mathbb{R}$ est une fonction mesurable.

Il s'avère que dans le cas particulier où $(X, Z_1, \cdots, Z_n)$ est un vecteur aléatoire gaussien, la fonction mesurable h est affine. Autrement dit, $\mathbb{E}(X \,|\, Z_1, \cdots, Z_n)$ coïncide avec la projection orthogonale de X sur l'espace vectoriel (de dimension finie) engendré par $Z_1, \cdots, Z_n$.

Théorème 7.6. *Soit $(X, Z_1, \cdots, Z_n)$ un vecteur aléatoire gaussien centré à valeurs dans $\mathbb{R}^{n+1}$. Alors il existe des réels $a_1, \cdots, a_n$ tels que*

$$\mathbb{E}(X \,|\, Z_1, \cdots, Z_n) = \sum_{j=1}^n a_j Z_j \,, \qquad \text{p.s.}$$

De plus, pour toute fonction mesurable $g : \mathbb{R} \to \mathbb{R}_+$, on a

$$\mathbb{E}[g(X) \,|\, Z_1, \cdots, Z_n] = f\Big(\sum_{j=1}^n a_j Z_j \Big), \qquad \text{p.s.,}$$

où, pour tout $m \in \mathbb{R}$, $f(m) := \mathbb{E}[g(m + \sigma \mathscr{N})]$, $\mathscr{N}$ désignant une variable aléatoire gaussienne réelle standard, et $\sigma^2 := \mathbb{E}[(X - \sum_{j=1}^n a_j Z_j)^2]$.

Preuve. Soit $\widehat{X} := a_0 + \sum_{j=1}^n a_j Z_j$ une droite de régression de X par rapport à $(Z_1, \cdots, Z_n)$; c'est-à-dire $\mathbb{E}(X - \widehat{X}) = 0$ et $\mathbb{E}[(X - \widehat{X}) Z_j] = 0$, $\forall 1 \le j \le n$. Ainsi, $a_0 = 0$, et pour tout $1 \le j \le n$,

$$\mathrm{Cov}(X - \widehat{X}, Z_j) = 0.$$

Remarquons que $(X - \widehat{X}, Z_1, \cdots, Z_n)$ est un vecteur gaussien centré, car toutes combinaison linéaire de ses composantes est une combinaison linéaire de X, $Z_1, \cdots, Z_n$. L'identité précédente assure donc que les variables aléatoires $X - \widehat{X}$ et $(Z_1, \cdots, Z_n)$ sont indépendantes. En particulier, $\mathbb{E}(X - \widehat{X} \,|\, Z_1, \cdots, Z_n) = \mathbb{E}(X - \widehat{X}) = 0$ p.s. Donc

$$\mathbb{E}(X \,|\, Z_1, \cdots, Z_n) = \mathbb{E}(X - \widehat{X} \,|\, Z_1, \cdots, Z_n) + \mathbb{E}(\widehat{X} \,|\, Z_1, \cdots, Z_n) = 0 + \widehat{X} = \widehat{X} \,.$$

[On a utilisé le fait que $\widehat{X}$ est $\sigma(Z_1, \cdots, Z_n)$-mesurable.]

Soit maintenant $g : \mathbb{R} \to \mathbb{R}_+$ une fonction mesurable. On a

$$\mathbb{E}[g(X) \,|\, Z_1, \cdots, Z_n] = \mathbb{E}[g(\sum_{j=1}^{n} a_j Z_j + Y) \,|\, Z_1, \cdots, Z_n]\,,$$

où $Y := X - \widehat{X}$ est indépendante de $\sigma(Z_1, \cdots, Z_n)$. Il résulte alors du Théorème 4.10 que $\mathbb{E}[g(X) \,|\, Z_1, \cdots, Z_n] = f(\sum_{j=1}^{n} a_j Z_j)$ p.s., avec $f(m) := \mathbb{E}[g(m+Y)] = \mathbb{E}[g(m+\sigma\mathscr{N})]$. □

Exemple 7.7. Soit $\begin{pmatrix} X \\ Z \end{pmatrix}$ un vecteur aléatoire gaussien centré, de matrice de covariances $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. On cherche à calculer $\mathbb{E}(X \,|\, Z)$.

Par le Théorème 7.6, il existe un réel $a \in \mathbb{R}$ tel que $\mathbb{E}(X \,|\, Z) = aZ$ p.s. Il s'agit de déterminer la valeur de a. En multipliant par Z dans les deux côtés de l'identité, et de prendre espérance ensuite, on obtient :

$$\mathbb{E}[Z\, \mathbb{E}(X \,|\, Z)] = a\, \mathbb{E}(Z^2)\,.$$

Comme Z est $\sigma(Z)$-mesurable, on a $\mathbb{E}[Z\, \mathbb{E}(X \,|\, Z)] = \mathbb{E}(XZ)$, qui vaut 1, tandis que $\mathbb{E}(Z^2) = 2$, on obtient $a = \frac{1}{2}$. Conclusion : $\mathbb{E}(X \,|\, Z) = \frac{Z}{2}$ p.s.

Il s'agit, bien sûr, du même exemple que l'Exemple 7.5 traité précédemment. □

Chapitre 2

Martingales à temps discret

Ce chapitre est consacré à la théorie des martingales à temps discret. On introduit d'abord la notion de la martingale, et étudie ensuite les divers modes de convergence (p.s., dans L^p pour $1 < p < \infty$, dans L^1), ainsi que le théorème d'arrêt.

1. Filtrations et martingales

Soit $(\Omega, \mathscr{F}, \mathbb{P})$ un espace de probabilité. Une **filtration** $(\mathscr{F}_n)_{n\ge 0}$ sur cet espace est une famille croissante de sous-tribus de $\mathscr{F}$:

$$\mathscr{F}_0 \subset \mathscr{F}_1 \subset \mathscr{F}_2 \subset \cdots \subset \mathscr{F}.$$

On appelle $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathbb{P})$ un espace de probabilité filtré.

Une suite de variables aléatoires $(X_n,\, n \ge 0)$ est appelée un processus aléatoire (ou : processus stochastique, ou : processus). L'indice n est moralement considéré comme un paramètre temporel.

Exemple 1.1. Soit $(X_n,\, n \ge 0)$ un processus aléatoire. Pour tout $n \ge 0$, on pose $\mathscr{F}_n := \sigma(X_i \,:\, 0 \le i \le n)$. Alors $(\mathscr{F}_n)_{n\ge 0}$ est une filtration, que l'on appelle souvent la filtration naturelle (ou : filtration canonique) de $(X_n,\, n \ge 0)$. $\square$

Soit $(\mathscr{F}_n)_{n\ge 0}$ une filtration, et soit $(X_n,\, n \ge 0)$ un processus aléatoire. On dit que $(X_n,\, n \ge 0)$ est **adapté** par rapport à $(\mathscr{F}_n)$ si pour tout n, X_n est $\mathscr{F}_n$-mesurable. [Lorsqu'il n'y a pas de risque d'ambiguïté sur le choix de la filtration, on dit simplement que $(X_n,\, n \ge 0)$ est adapté.] Dans l'Exemple 1.1, le processus est adapté.

Définition 1.2. *On dit que* (X_n) *est une martingale* (*resp. surmartingale ; sous-martingale*) *par rapport à la filtration* $(\mathscr{F}_n)$ *si*

(i) (X_n) *est adapté ;*

(ii) $\forall\, n$, $\mathbb{E}(|X_n|) < \infty$ *;*

(iii) $\forall\, n$, $\mathbb{E}(X_{n+1}\,|\,\mathscr{F}_n) = X_n$, p.s. (*resp.*, $\mathbb{E}(X_{n+1}\,|\,\mathscr{F}_n) \le X_n$ *;* $\mathbb{E}(X_{n+1}\,|\,\mathscr{F}_n) \ge X_n$).

On appellera de temps en temps (iii) l'inégalité caractéristique des sous-martingales (ou surmartingales), ou l'identité caractéristique des martingales.

Remarquons que si (X_n) est une surmartingale, alors $n \mapsto \mathbb{E}(X_n)$ est *dé*croissante.[1]

Exemple 1.3. (Martingale fermée). Soit ξ une variable aléatoire intégrable. Soit $X_n := \mathbb{E}(\xi\,|\,\mathscr{F}_n)$. Alors (X_n) est une martingale (appellée martingale fermée).

En effet, (X_n) est adapté d'après la définition de l'espérance conditionnelle, ce qui donne (i).

Pour tout n, $|X_n| \le \mathbb{E}(\,|\xi|\,|\,\mathscr{F}_n)$ par l'inégalité de Jensen (version conditionnelle). Comme $\mathbb{E}(|\xi|\,|\,\mathscr{F}_n)$ est intégrable (par la définition de l'espérance conditionnelle), on en déduit que X_n est également intégrable, ce qui donne (ii).

Enfin, on a $\mathbb{E}[X_{n+1}\,|\,\mathscr{F}_n] = \mathbb{E}[\mathbb{E}(\xi\,|\,\mathscr{F}_{n+1})\,|\,\mathscr{F}_n]$, qui n'est autre que $\mathbb{E}[\xi\,|\,\mathscr{F}_n]$ (Propriété 4.7 du Chapitre 1), c'est-à-dire X_n : d'où (iii). $\square$

Exemple 1.4. Soient ξ_1, ξ_2, $\cdots$ des variables aléatoires réelles intégrables et indépendantes. Soit $x \in \mathbb{R}$. Posons $\mathscr{F}_0 := \{\varnothing, \Omega\}$ et, pour $n \ge 1$,

$$\begin{aligned} X_n &:= x + \sum_{i=1}^n \xi_i, \\ \mathscr{F}_n &:= \sigma\{X_i,\ 1 \le i \le n\}\,. \end{aligned}$$

Il est clair que (X_n) est adapté, et que pour tout $n \ge 0$, X_n est intégrable. De plus, $\mathbb{E}(X_{n+1}\,|\,\mathscr{F}_n) = X_n + \mathbb{E}(\xi_{n+1})$. Donc (X_n) est une martingale si $\mathbb{E}(\xi_n) = 0$, $\forall\, n \ge 1$; est une sous-martingale si $\mathbb{E}(\xi_n) \ge 0$, $\forall\, n \ge 1$; et est une surmartingale si $\mathbb{E}(\xi_n) \le 0$, $\forall\, n \ge 1$. $\square$

Exemple 1.5. Soient ξ_0, ξ_1, $\cdots$ des variables aléatoires positives et indépendantes, telles que $\mathbb{E}(\xi_i) = 1$, $\forall i \ge 0$. On pose $X_n := \prod_{i=0}^n \xi_i$ et $\mathscr{F}_n := \sigma(\xi_i,\, 0 \le i \le n)$, $n \ge 0$. Alors $(X_n,\, n \ge 0)$ est une martingale.

[1] L'origine de l'expression: une fonction f sur $\mathbb{R}^n$ est surharmonique si et seulement si pour un mouvement brownien B à valeurs dans $\mathbb{R}^n$, $f(B)$ est une surmartingale locale par rapport à la filtration naturelle de B. On en saura un (tout) peu plus à la fin du Chapitre 3.

En effet, X_n est intégrable, et $(X_n,\, n \geq 0)$ adapté. Pour tout n, $\mathbb{E}(X_{n+1} \,|\, \mathscr{F}_n) = \mathbb{E}(\xi_{n+1} X_n \,|\, \mathscr{F}_n) = X_n \, \mathbb{E}(\xi_{n+1} \,|\, \mathscr{F}_n) = X_n$ p.s. (où l'on a utilisé le fait que ξ_{n+1} est indépendante de $\mathscr{F}_n$ pour déduire que $\mathbb{E}(\xi_{n+1} \,|\, \mathscr{F}_n) = \mathbb{E}(\xi_{n+1}) = 1$).

On étudiera cette martingale en profondeur plus tard, en Section 8.2. □

Remarque 1.6. Il est clair que (X_n) est une sous-martingale si et seulement si $(-X_n)$ est une surmartingale, et que (X_n) est une martingale si et seulement si elle est à la fois une sous-martingale et une surmartingale. □

Proposition 1.7. *Si (X_n) est une sous-martingale, alors $\forall\, m \geq n$, $\mathbb{E}(X_m \,|\, \mathscr{F}_n) \geq X_n$,* p.s.[2]

Preuve. Montrons le résultat par un argument par récurrence. L'hypothèse de récurrence est la suivante : pour tout entier $k \geq 0$,

$$\text{Propriété } (k) : \qquad \forall n \geq 0,\ \mathbb{E}(X_{n+k} \,|\, \mathscr{F}_n) \geq X_n\,, \quad \text{p.s.}$$

Cette propriété est trivialement vraie au rang $k = 0$.

On suppose la propriété vraie au rang k. Par la Propriété 4.7 du Chapitre 1, on a, pour tout $n \geq 0$,

$$\mathbb{E}(X_{n+k+1} \,|\, \mathscr{F}_n) = \mathbb{E}[\, \mathbb{E}(X_{n+k+1} \,|\, \mathscr{F}_{n+k}) \,|\, \mathscr{F}_n] \,.$$

L'inégalité caractéristique des sous-martingales nous donne $\mathbb{E}(X_{n+k+1} \,|\, \mathscr{F}_{n+k}) \geq X_{n+k}$ p.s., ce qui implique que

$$\mathbb{E}(X_{n+k+1} \,|\, \mathscr{F}_n) \geq \mathbb{E}[X_{n+k} \,|\, \mathscr{F}_n] \,,$$

qui est, d'après la propriété au rang k, $\geq X_n$ p.s. On obtient donc la propriété au rang $k+1$.

Ceci démontre le résultat cherché par récurrence. □

Proposition 1.8. (i) *Si (X_n) et (Y_n) sont des martingales (resp. sous-martingales ; surmartingales), alors $(X_n + Y_n)$ en est également une.*

(ii) *Si (X_n) et (Y_n) sont des sous-martingales, alors $(X_n \vee Y_n)$ est une sous-martingale.*[3]

Preuve. Élémentaire. □

Proposition 1.9. *Soit (X_n) une martingale, et soit φ une fonction convexe. Si $\mathbb{E}(|\varphi(X_n)|) < \infty$, $\forall\, n$, alors $(\varphi(X_n))$ est une sous-martingale.*

---

[2] Donc, la proposition nous dit également, d'après la Remarque 1.6, que si (X_n) est une surmartingale, alors $\forall\, m \geq n$, $\mathbb{E}(X_m \,|\, \mathscr{F}_n) \leq X_n$ p.s., et que si (X_n) est une martingale, alors $\forall\, m \geq n$, $\mathbb{E}(X_m \,|\, \mathscr{F}_n) = X_n$ p.s.

[3] Rappel de notations : $x \vee y := \max\{x,\, y\}$ et $x \wedge y := \min\{x,\, y\}$.

Preuve. Par l'inégalité de Jensen,

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

D'où la conclusion. □

Exemple 1.10. Soit $p \geq 1$, et soit (X_n) une martingale telle que $\mathbb{E}(|X_n|^p) < \infty$. Alors $(|X_n|^p)$ est une sous-martingale.

[En particulier, si (X_n) est une martingale, alors $(|X_n|)$ est une sous-martingale.]

Il suffit d'appliquer la Proposition 1.9 à la fonction convexe $\varphi(x) = |x|^p$. □

Proposition 1.11. *Soit (X_n) une sous-martingale, et soit φ une fonction convexe et croissante telle que $\mathbb{E}(|\varphi(X_n)|) < \infty$, $\forall\, n$. Alors $(\varphi(X_n))$ est une sous-martingale.*

Preuve. On a

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

comme prévu. □

Exemple 1.12. Soit (X_n) est une sous-martingale.

(i) Posons[4] $Y_n := X_n^+$. Alors (Y_n) est également une sous-martingale, car $x \mapsto x^+$ est convexe et croissante sur $\mathbb{R}$.

(ii) Plus généralement, pour tout réel $a \in \mathbb{R}$, soit $Z_n := (X_n - a)^+$; alors (Z_n) est une sous-martingale. □

2. Temps d'arrêt

Soit $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathbb{P})$ un espace de probabilité filtré, et l'on note

$$\mathscr{F}_\infty := \sigma(\mathscr{F}_n,\, n \geq 0)\,,$$

c'est-à-dire la tribu engendrée par les éléments de l'ensemble des tribus $\mathscr{F}_n$.

Une application $T : \Omega \to \{0, 1, 2, \cdots\} \cup \{\infty\}$ est appelée un **temps d'arrêt** si pour tout $n \leq \infty$, $\{T \leq n\} \in \mathscr{F}_n$. Il est clair que T est un temps d'arrêt si et seulement si $\{T = n\} \in \mathscr{F}_n$, $\forall\, n \leq \infty$.

[4]Notations : $x^+ := \max\{x, 0\}$, $x^- := \max\{-x, 0\}$ pour tout réel x.

Exemple 2.1. (*Temps d'arrêt déterministes*) Fixons un entier positif $k \geq 0$. Posons $T(\omega) := k$, $\omega \in \Omega$. Alors T est un temps d'arrêt.

Un autre exemple de temps d'arrêt déterministe est $T(\omega) := \infty$, $\omega \in \Omega$. $\square$

Exemple 2.2. (*Temps d'atteinte*) Soit (X_n) un processus adapté par rapport à la filtration $(\mathscr{F}_n)$, à valeurs dans un espace mesurable $(E, \mathscr{E})$. Soit $B \in \mathscr{E}$. Posons

$$T_B := \inf\{n \geq 0 : X_n \in B\},$$

avec la convention $\inf \varnothing := \infty$ (c'est-à-dire, si $X_n(\omega) \notin B$ pour tout $n \geq 0$, alors $T_B(\omega) = \infty$).

On vérifie très facilement par définition que T_B est un temps d'arrêt. En effet, T_B est à valeurs dans $\{0, 1, 2, \cdots\} \cup \{\infty\}$; de plus, pour tout entier $n \geq 0$, $\{T_B \leq n\} = \cup_{k=0}^{n}\{X_k \in B\}$ qui est un élément de $\mathscr{F}_n$ car $\{X_k \in B\} \in \mathscr{F}_k \subset \mathscr{F}_n$ pour $k \leq n$. $\square$

Exemple 2.3. (*Temps de passages successifs*) Soit (X_n) un processus adapté par rapport à la filtration $(\mathscr{F}_n)$, à valeurs dans un espace mesurable $(E, \mathscr{E})$. Soit $(B_k, k \geq 0)$ une suite d'éléments (pas nécessairement distincts) de $\mathscr{E}$. On définit, par récurrence, la suite $(T_k, k \geq 0)$ par $T_0 := 0$ et pour $k \geq 1$,

$$T_k := \inf\{n > T_{k-1} : X_n \in B_k\},$$

avec la convention $\inf \varnothing := \infty$.

Montrons, par un argument par récurrence, que $(T_k, k \geq 0)$ est une suite de temps d'arrêt, l'hypothèse de récurrence au rang $k \in \mathbb{N}$ étant : "T_k est un temps d'arrêt".

Au rang $k = 0$, $T_0 := 0$ est un temps d'arrêt déterministe.

On suppose maintenant que T_k est un temps d'arrêt. Pour tout $n \in \mathbb{N}$, on a

$$\{T_{k+1} > n\} = \{T_k \geq n\} \cup \bigcup_{j=0}^{n-1} \Big(\{T_k = j\} \cap \{X_{j+1} \notin B_{k+1}\} \cap \cdots \cap \{X_n \notin B_{k+1}\}\Big).$$

Comme T_k est supposé un temps d'arrêt, on a $\{T_k \geq n\} = \{T_k \leq n-1\}^c \in \mathscr{F}_{n-1} \subset \mathscr{F}_n$, et pour $j \leq n-1$, $\{T_k = j\} = \{T_k \leq j\} \backslash \{T_k \leq j-1\} \in \mathscr{F}_j \subset \mathscr{F}_n$, ainsi que $\{X_\ell \notin B_{k+1}\} \in \mathscr{F}_\ell \subset \mathscr{F}_n$ pour tout $\ell \leq n$. Donc $\{T_{k+1} > n\} \in \mathscr{F}_n$. On déduit alors que $\{T_{k+1} \leq n\} = \{T_{k+1} > n\}^c \in \mathscr{F}_n$. Autrement dit, T_{k+1} est un temps d'arrêt.

Par un raisonnement par récurrence, on voit que $(T_k, k \geq 0)$ est une suite de temps d'arrêt. $\square$

Soit T un temps d'arrêt. La tribu $\mathscr{F}_T$ des événements antérieurs à T, que l'on appellera également "tribu engendrée par le temps d'arrêt T", est définie par

$$\begin{aligned}\mathscr{F}_T &= \{A \in \mathscr{F}_\infty : \forall\, n, A \cap \{T = n\} \in \mathscr{F}_n\} \\ &= \{A \in \mathscr{F}_\infty : \forall\, n, A \cap \{T \leq n\} \in \mathscr{F}_n\}.\end{aligned}$$

Propriété 2.4. *Soient S et T des temps d'arrêt.*

(i) *Alors $S \vee T$ et $S \wedge T$ sont aussi des temps d'arrêt, et $\mathscr{F}_{S\wedge T} = \mathscr{F}_S \cap \mathscr{F}_T$.*

(ii) *Si $S \leq T$, alors $\mathscr{F}_S \subset \mathscr{F}_T$.*

(iii) *Si (X_n) est un processus à valeurs dans $\mathbb{R}^d$ et est adapté par rapport à $(\mathscr{F}_n)$, alors $X_T\, \mathbf{1}_{\{T<\infty\}}$ est $\mathscr{F}_T$-mesurable.*[5]

Preuve. (ii) Si $A \in \mathscr{F}_S$, alors $A \cap \{T \leq n\} = (A \cap \{S \leq n\}) \cap \{T \leq n\} \in \mathscr{F}_n$.

(i) On a $\{S \vee T \leq n\} = \{S \leq n\} \cap \{T \leq n\}$, et $\{S \wedge T \leq n\} = \{S \leq n\} \cup \{T \leq n\}$.

D'après (ii), $\mathscr{F}_{S\wedge T} \subset \mathscr{F}_S$ et $\mathscr{F}_{S\wedge T} \subset \mathscr{F}_T$; donc $\mathscr{F}_{S\wedge T} \subset \mathscr{F}_S \cap \mathscr{F}_T$.

Inversement, soit $A \in \mathscr{F}_S \cap \mathscr{F}_T$. Alors pour tout $n \geq 0$, $A \cap \{S \leq n\} \in \mathscr{F}_n$ et $A \cap \{T \leq n\} \in \mathscr{F}_n$; d'où $A \cap \{(S\wedge T) \leq n\} = A \cap (\{S \leq n\} \cup \{T \leq n\}) = (A \cap \{S \leq n\}) \cup (A \cap \{T \leq n\})$ qui est un élément de $\mathscr{F}_n$, étant réunion de deux éléments de $\mathscr{F}_n$.

(iii) Il suffit de remarquer que $X_T\, \mathbf{1}_{\{T<\infty\}} = \sum_{n=0}^\infty X_n\, \mathbf{1}_{\{T=n\}}$ est $\mathscr{F}_\infty$-mesurable, et que pour tout $A \subset \mathbb{R}^d$ borélien et tout entier $n \geq 0$, $\{X_T\, \mathbf{1}_{\{T<\infty\}} \in A\} \cap \{T = n\} = \{X_n \in A\} \cap \{T = n\} \in \mathscr{F}_n$. □

3. Inégalité maximale et inégalité de Doob

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

Théorème 3.1. *Si T est un temps d'arrêt, et si (X_n) est une sous-martingale, alors $(X_{T\wedge n})$ est une sous-martingale.*

Preuve. Comme $|X_{T\wedge n}| \leq |X_0| + |X_1| + \cdots + |X_n|$, $X_{T\wedge n}$ est intégrable. Il est clair que $(X_{T\wedge n})$ est adapté, car $X_{T\wedge n}$ est $\mathscr{F}_{T\wedge n}$-mesurable, et cette dernière est une sous-tribu de $\mathscr{F}_n$ (Propriété 2.4). Enfin, puisque $\{T \geq n+1\} = \{T \leq n\}^c \in \mathscr{F}_n$,

$$\begin{aligned}\mathbb{E}\Big[(X_{T\wedge(n+1)} - X_{T\wedge n}) \,|\, \mathscr{F}_n\Big] &= \mathbb{E}\Big[\, (X_{n+1} - X_n)\, \mathbf{1}_{\{T\geq n+1\}} \,|\, \mathscr{F}_n\Big] \\ &= \mathbf{1}_{\{T\geq n+1\}} \mathbb{E}[(X_{n+1} - X_n) \,|\, \mathscr{F}_n] \geq 0\,,\end{aligned}$$

[5]Notation : Pour $\omega \in \{T < \infty\}$, $X_T(\omega) := X_{T(\omega)}(\omega)$.

comme prévu. □

Théorème 3.2. (Théorème d'arrêt). *Soient $S \le T$ deux temps d'arrêt* **bornés**. *Soit (X_n) une sous-martingale. Alors*

$$\mathbb{E}(X_T \,|\, \mathscr{F}_S) \ge X_S\,, \qquad \text{p.s.}$$

Preuve. Soit k un entier tel que $\mathbb{P}(S \le T \le k) = 1$. Alors $|X_T| \le |X_0| + |X_1| + \cdots + |X_k|$ qui est intégrable. Soit $A \in \mathscr{F}_S$. On a

$$\mathbb{E}[(X_T - X_S)\,\mathbf{1}_A] = \sum_{n=0}^{k} \mathbb{E}\Big[(X_{T\wedge k} - X_n)\,\mathbf{1}_{A\cap\{S=n\}}\Big].$$

Comme $A \cap \{S = n\} \in \mathscr{F}_n$, et $(X_{T\wedge n})$ est une sous-martingale, on a

$$\mathbb{E}\Big[(X_{T\wedge k} - X_n)\,\mathbf{1}_{A\cap\{S=n\}}\Big] \ge \mathbb{E}\Big[(X_{T\wedge n} - X_n)\,\mathbf{1}_{A\cap\{S=n\}}\Big] = 0,$$

la dernière identité provient du fait que $X_{T\wedge n} = X_n$ sur $\{S = n\}$. On a donc montré que $\mathbb{E}[(X_T - X_S)\,\mathbf{1}_A] \ge 0$. □

Le théorème d'arrêt est l'un des résultats les plus importants de la théorie des martingales. On utilisera souvent les conséquences suivantes :

• Conséquence 1 : $S \le T$ temps d'arrêt **bornés**, (X_n) sous-martingale $\Rightarrow \mathbb{E}(X_0) \le \mathbb{E}(X_S) \le \mathbb{E}(X_T)$.

• Conséquence 2 : T temps d'arrêt **borné**, (X_n) martingale $\Rightarrow \mathbb{E}(X_T) = \mathbb{E}(X_0)$.

Remarque 3.3. (Importante). En pratique, si l'on a une martingale (X_n) quelconque et un temps d'arrêt T quelconque, on peut appliquer la Conséquence 2 pour voir que pour tout entier $n \ge 0$, $\mathbb{E}(X_{T\wedge n}) = \mathbb{E}(X_0)$. On fait ensuite $n \to \infty$, et tâche à justifier le passage à la limite (ce qui n'est nullement automatique, en général) en essayant d'appliquer Fatou, convergence monotone ou convergence dominée — tout en se croisant bien les doigts. □

Théorème 3.4. (Inégalité maximale).[6] *Si (X_n) est une sous-martingale, alors pour tout réel $\lambda > 0$ et tout entier $k \ge 0$,*

$$\lambda\,\mathbb{P}\Big(\max_{0\le n\le k} |X_n| \ge \lambda\Big) \le \mathbb{E}(|X_0|) + 2\,\mathbb{E}(|X_k|).$$

[6]Qui porte souvent le nom de l'inégalité maximale de Doob.

Preuve. Montrons d'abord les inégalités suivantes :

$$\lambda\,\mathbb{P}(A) \;\le\; \mathbb{E}(X_k\,\mathbf{1}_A), \tag{3.1}$$

$$\lambda\,\mathbb{P}(B) \;\le\; -\mathbb{E}(X_0)+\mathbb{E}(X_k\,\mathbf{1}_{B^c}), \tag{3.2}$$

où $A := \{\max_{0\le n\le k} X_n \ge \lambda\}$ et $B := \{\min_{0\le n\le k} X_n \le -\lambda\}$.

Soit $T := (\inf\{n : 0\le n\le k,\ X_n \ge \lambda\}) \wedge k$ (avec $\inf\varnothing := \infty$), qui est un temps d'arrêt borné. D'après la Conséquence 1 du théorème d'arrêt (Théorème 3.2), $\mathbb{E}(X_T)\le \mathbb{E}(X_k)$. Or $X_T = X_k$ sur A^c, ce qui implique que $\mathbb{E}(X_T\,\mathbf{1}_A)\le \mathbb{E}(X_k\,\mathbf{1}_A)$.

Sur l'ensemble A, $X_T \ge \lambda$, d'où $\mathbb{E}(X_T\,\mathbf{1}_A)\ge \mathbb{E}(\lambda\,\mathbf{1}_A) = \lambda\,\mathbb{P}(A)$. Par conséquent, on a (3.1).

La preuve de (3.2) est similaire. Soit $S := (\inf\{n\ge 0 : X_n \le -\lambda\})\wedge k$ qui est un temps d'arrêt borné. Il s'en suit que $\mathbb{E}(X_S)\ge \mathbb{E}(X_0)$. D'autre part $X_S\,\mathbf{1}_B \le -\lambda\,\mathbf{1}_B$, et $X_S\,\mathbf{1}_{B^c} = X_k\,\mathbf{1}_{B^c}$. Donc $\mathbb{E}(X_0)\le \mathbb{E}(X_S) = \mathbb{E}(X_S\,\mathbf{1}_B)+\mathbb{E}(X_S\,\mathbf{1}_{B^c}) \le -\lambda\,\mathbb{P}(B)+\mathbb{E}(X_k\,\mathbf{1}_{B^c})$.

Maintenant que (3.1) et (3.2) sont prouvées, on obtient que

$$\begin{aligned}
\lambda\,\mathbb{P}\Big(\max_{0\le n\le k}|X_n|\ge\lambda\Big) &\le \lambda\,\mathbb{P}(A\cup B)\\
&\le \mathbb{E}(X_k\,\mathbf{1}_A)-\mathbb{E}(X_0)+\mathbb{E}(X_k\,\mathbf{1}_{B^c})\\
&\le \mathbb{E}(|X_k|)+\mathbb{E}(|X_0|)+\mathbb{E}(|X_k|)\,.
\end{aligned}$$

D'où l'inégalité maximale. □

Exemple 3.5. (Inégalité de Kolmogorov). Soit (X_n) une martingale telle que $\mathbb{E}(X_n^2)<\infty$. Alors pour tout $k\ge 0$ et tout $\lambda>0$,

$$\mathbb{P}\Big(\max_{0\le n\le k}|X_n|\ge\lambda\Big)\le \frac{\mathbb{E}(X_k^2)}{\lambda^2}.$$

En effet, on sait que (X_n^2) est une sous-martingale (Exemple 1.10). D'après (3.1),

$$\mathbb{P}\Big(\max_{0\le n\le k}|X_n|\ge\lambda\Big)\le \lambda^{-2}\,\mathbb{E}(X_k^2\,\mathbf{1}_{\{\max_{0\le n\le k}X_n^2\ge\lambda^2\}})\le \lambda^{-2}\,\mathbb{E}(X_k^2)\,,$$

d'où la conclusion désirée. □

Théorème 3.6. (Inégalité de Doob)[7]. *Si (X_n) est une sous-martingale, et si $p>1$ et $q>1$ sont telles que $\frac{1}{p}+\frac{1}{q}=1$, alors pour tout $k\ge 0$,*[8]

$$\Big\|\max_{0\le n\le k}X_n^+\Big\|_p \le q\,\|X_k^+\|_p.$$

[7] Qui porte également le nom de l'inégalité L^p de Doob.

[8] Notation : $\|\xi\|_p := \{\mathbb{E}(|\xi|^p)\}^{1/p}$.

En particulier, si (X_n) *est une martingale, alors*

$$\Big\| \max_{0\le n\le k} |X_n| \Big\|_p \le q\, \|X_k\|_p.$$

Preuve. La seconde assertion est une conséquence immédiate de la première : il suffit de remplacer (X_n) par $(|X_n|)$, qui est une sous-martingale si (X_n) est une martingale (Exemple 1.10).

Soit (X_n) une sous-martingale. Alors (X_n^+) en est aussi une (Proposition 1.11). Soit $\lambda > 0$. Par (3.1),

$$\lambda\, \mathbb{P}\Big(\max_{0\le n\le k} X_n^+ \ge \lambda \Big) \le \mathbb{E}\Big[X_k^+\, \mathbf{1}_{\{\max_{0\le n\le k} X_n^+ \ge \lambda\}} \Big].$$

Écrivons $Y_k := \max_{0\le n\le k} X_n^+$ pour simplifier la notation. Alors pour tout $M > 0$,

$$\begin{aligned}
\mathbb{E}[(Y_k\wedge M)^p] &= \int_0^\infty p\lambda^{p-1}\, \mathbb{P}[(Y_k\wedge M)\ge\lambda]\, \mathrm{d}\lambda \\
&= \int_{[0,\, M]} p\lambda^{p-1}\, \mathbb{P}[Y_k\ge\lambda]\, \mathrm{d}\lambda \\
&\le \int_{[0,\, M]} p\lambda^{p-1} \Big(\lambda^{-1}\, \mathbb{E}[X_k^+\, \mathbf{1}_{\{Y_k\ge\lambda\}}] \Big)\, \mathrm{d}\lambda\,.
\end{aligned}$$

L'intégrand étant positif, on peut appliquer le théorème de Fubini–Tonelli pour voir que l'intégrale vaut

$$\begin{aligned}
&= \mathbb{E}\Big\{ \int_{[0,M]} p\lambda^{p-2} X_k^+\, \mathbf{1}_{\{Y_k\ge\lambda\}}\, \mathrm{d}\lambda \Big\} \\
&= \mathbb{E}\Big\{ \int_0^{Y_k\wedge M} p\lambda^{p-2} X_k^+\, \mathrm{d}\lambda \Big\} \\
&= \frac{p}{p-1} \mathbb{E}[\, (Y_k\wedge M)^{p-1} X_k^+ \,]\,.
\end{aligned}$$

À l'aide de l'inégalité de Hölder (notons que $\frac{p}{p-1} = q$), ceci nous emmène à:

$$\mathbb{E}[\, (Y_k\wedge M)^p \,] \le \frac{p}{p-1} \Big\{ \mathbb{E}[\, (Y_k\wedge M)^p \,] \Big\}^{1/q} \Big\{ \mathbb{E}[\, (X_k^+)^p \,] \Big\}^{1/p}.$$

Autrement dit,

$$\|Y_k\wedge M\|_p \le q\, \|X_k^+\|_p.$$

En faisant $M\to\infty$ et à l'aide du théorème de convergence monotone, on obtient l'inégalité de Doob. $\square$

Corollaire 3.7. (Inégalité de Doob). *Si* (X_n) *est une martingale, et si* $\frac{1}{p} + \frac{1}{q} = 1$*, alors*

$$\Big\| \sup_{n\geq 0} |X_n| \Big\|_p \leq q \sup_{n\geq 0} \|X_n\|_p \,.$$

Preuve. Le Théorème 3.6 nous dit que pour tout k,

$$\Big\| \max_{0\leq n\leq k} |X_n| \Big\|_p \leq q \sup_{n\geq 0} \|X_n\|_p \,.$$

Il suffit de faire $k \to \infty$ et d'appliquer le théorème de convergence monotone. □

Remarque 3.8. Dans l'inégalité de Doob (Théorème 3.6, Corollaire 3.7), p est *strictement* supérieur à 1. Il est possible de l'étendre au cas $p = 1$, à condition de remplacer, dans la borne supérieure, la norme L^1 par la norme de $L \log L$: si (X_n) est une sous-martingale,

$$\Big\| \max_{0\leq n\leq k} X_n^+ \Big\|_1 \leq \frac{\mathrm{e}}{\mathrm{e}-1} \|1 + X_k^+ \log^+(X_k^+)\|_1,$$

où $\log^+(x) := \log(x \vee 1)$. La preuve de cette inégalité est semblable à celle du Théorème 3.6 (pour les détails, voir TD). □

4. Convergences

On présente des résultats de convergence en divers modes (p.s. ; dans L^p pour $1 < p < \infty$; dans L^1) des martingales.

4.1. Convergence presque sûre

Le but de cette partie est de prouver le résultat suivant concernant convergence p.s. des martingales.

Théorème 4.1. *Si* (X_n) *est une sous-martingale telle que*

$$\sup_n \mathbb{E}(X_n^+) < \infty,$$

alors $X_\infty := \lim_{n\to\infty} X_n$ *existe* p.s.*, et* $\mathbb{E}(|X_\infty|) < \infty$*.*

Exercice 4.2. Trouver un exemple de martingale (X_n) pour laquelle la convergence p.s. n'a pas lieu. □

La preuve du théorème nécessite quelques préparations. Soit (X_n) une sous-martingale. Soit $a < b$ deux réels. On définit $(T_n,\, n \ge 0)$ par récurrence: $T_0 := 0$, et pour $k \ge 1$,

$$\begin{aligned} T_{2k-1} &:= \inf\{m > T_{2k-2} :\, X_m \le a\}, \\ T_{2k} &:= \inf\{m > T_{2k-1} :\, X_m \ge b\}. \end{aligned}$$

Il est clair que (T_n) est une suite croissante de temps d'arrêt. Comme $X_{T_{2k-1}} \le a$ et $X_{T_{2k}} \ge b$, entre les "instants" T_{2k-1} et T_{2k}, (X_n) effectue une montée le long de $[a,\, b]$.[9] Soit $M_n := \sup\{k :\, T_{2k} \le n\}$, qui est le nombre de montées avant n. Montrons que

$$(b-a)\,\mathbb{E}(M_n) \le \mathbb{E}[(X_n - a)^+]. \tag{4.1}$$

Par le théorème d'arrêt (Théorème 3.2),

$$\begin{aligned} 0 &\le \mathbb{E}[\, X_{T_{2k+1}\wedge n} - X_{T_{2k}\wedge n}\,] \\ &= \mathbb{E}[\,(X_{T_{2k+1}\wedge n} - X_{T_{2k}\wedge n})\,\mathbf{1}_{\{T_{2k}\le n\}}\,] \\ &= \mathbb{E}[\,(X_{T_{2k+1}\wedge n} - X_{T_{2k}\wedge n})\,(\,\mathbf{1}_{\{T_{2k}\le n< T_{2k+1}\}} + \,\mathbf{1}_{\{T_{2k+1}\le n\}})\,] \\ &\le \mathbb{E}[\,(X_n - b)\,\mathbf{1}_{\{T_{2k}\le n< T_{2k+1}\}}\,] + (a-b)\,\mathbb{P}(T_{2k+1}\le n) \\ &= \mathbb{E}[\,(X_n - a)\,\mathbf{1}_{\{T_{2k}\le n< T_{2k+1}\}}\,] + (a-b)\,\mathbb{P}(T_{2k}\le n)\,. \end{aligned}$$

On remarque que $\{T_{2k} \le n < T_{2k+1}\} \subset \{M_n = k\}$, et donc $(X_n - a)\,\mathbf{1}_{\{T_{2k}\le n< T_{2k+1}\}} \le (X_n - a)^+\,\mathbf{1}_{\{T_{2k}\le n< T_{2k+1}\}} \le (X_n - a)^+\,\mathbf{1}_{\{M_n=k\}}$.[10] D'autre part, $\{T_{2k} \le n\} = \{M_n \ge k\}$. Donc

$$(b-a)\,\mathbb{P}(M_n \ge k) \le \mathbb{E}[\,(X_n - a)^+\,\mathbf{1}_{\{M_n=k\}}\,].$$

En sommant sur k, on obtient (4.1).

Preuve du Théorème 4.1. Puisque $(X_n - a)^+ \le X_n^+ + |a|$, on a, d'après (4.1),

$$\mathbb{E}(M_n) \le \frac{\mathbb{E}(X_n^+) + |a|}{b-a} \le \frac{\sup_m \mathbb{E}(X_m^+) + |a|}{b-a}.$$

Quand $n \uparrow \infty$, $M_n \uparrow M$ qui est le nombre total de montées le long de $[a,\, b]$ par la suite (X_n). Par convergence monotone,

$$\mathbb{E}(M) \le \frac{\sup_m \mathbb{E}(X_m^+) + |a|}{b-a}\,,$$

[9] En termes de finance, on achète lorsque $X_n \le a$ et on vend si $X_n \ge b$; ainsi, chaque montée nous fournit un profit $\ge (b-a)$.

[10] En fait, la première inégalité est une identité car sur $\{T_{2k} \le n < T_{2k+1}\}$, on a $X_n \ge a$.

et en particulier, $M < \infty$, p.s. Ceci étant vrai pour tout couple de rationnels $a < b$, on obtient:

$$\mathbb{P}\Big(\bigcup_{a<b,\, a,\, b\in\mathbb{Q}} \{\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n\}\Big) = 0.$$

Par conséquent, $\limsup_n X_n = \liminf X_n$, p.s. ; autrement dit, $\lim_{n\to\infty} X_n$ existe p.s., et soit X_∞ la limite. Par le lemme de Fatou, $\mathbb{E}(X_\infty^+) \le \liminf_{n\to\infty} \mathbb{E}(X_n^+) < \infty$.

D'autre part, puisque (X_n) est une sous-martingale,

$$\mathbb{E}(X_n^-) = \mathbb{E}(X_n^+) - \mathbb{E}(X_n) \le \mathbb{E}(X_n^+) - \mathbb{E}(X_0) \le \sup_m \mathbb{E}(X_m^+) - \mathbb{E}(X_0),$$

ce qui, de nouveau à l'aide d'une application du lemme de Fatou, nous confirme que

$$\mathbb{E}(X_\infty^-) \le \liminf_{n\to\infty} \mathbb{E}(X_n^-) \le \sup_m \mathbb{E}(X_m^+) - \mathbb{E}(X_0) < \infty\,.$$

Par conséquent, $\mathbb{E}(|X_\infty|) < \infty$. □

Remarque 4.3. Pour la preuve du Théorème 9.1 ci-dessous, on observe que l'on a prouvé que si pour tous réels $a < b$, le nombre total de montées le long de $[a,\, b]$ par (X_n) est p.s. fini, alors la limite de X_n existe p.s. □

Remarque 4.4. Une observation sur la condition du Théorème 9.1: si $(X_n,\, n \ge 0)$ est une sous-martingale, alors $\sup_{n\ge 0} \mathbb{E}(X_n^+) < \infty$ si et seulement si $\sup_{n\ge 0} \mathbb{E}(|X_n|) < \infty$.[11]

En effet, $\sup_{n\ge 0} \mathbb{E}(|X_n|) < \infty$ implique trivialement $\sup_{n\ge 0} \mathbb{E}(X_n^+) < \infty$ car $\mathbb{E}(X_n^+) \le \mathbb{E}(|X_n|)$.

Inversement, supposons que $\sup_{n\ge 0} \mathbb{E}(X_n^+) < \infty$. Puisque $|X_n| = 2X_n^+ - X_n$ et $\mathbb{E}(X_n) \ge \mathbb{E}(X_0)$, on a $\sup_{n\ge 0} \mathbb{E}(|X_n|) \le 2 \sup_{n\ge 0} \mathbb{E}(X_n^+) - \mathbb{E}(X_0) < \infty$. □

Un cas spécial (et important) du Théorème 4.1 est le corollaire suivant.

Corollaire 4.5. *Si (X_n) est une surmartingale positive,*[12] *alors $X_\infty := \lim_{n\to\infty} X_n$ existe* p.s.*, et $\mathbb{E}(X_\infty) \le \mathbb{E}(X_0)$.*

Preuve. Comme $(-X_n)$ est une sous-martingale telle que $(-X_n)^+ = 0$, l'existence p.s. de la limite $X_\infty := \lim_{n\to\infty} X_n$ résulte immédiatement du Théorème 4.1. Puisque $\mathbb{E}(X_n) \le \mathbb{E}(X_0)$, le lemme de Fatou nous dit que $\mathbb{E}(X_\infty) \le \liminf_n \mathbb{E}(X_n) \le \mathbb{E}(X_0)$. □

[11] Dans la littérature, si un processus $(X_n,\, n \ge 0)$ satisfait $\sup_{n\ge 0} \mathbb{E}(|X_n|) < \infty$, on dit qu'il est borné dans L^1. Plus généralement, si $\sup_{n\ge 0} \mathbb{E}(|X_n|^p) < \infty$, on dit que le processus est borné dans L^p.

[12] Positivité : pour tout n, $X_n \ge 0$ p.s.

Exemple 4.6. (Jeu de pile ou face). Soit $X_0 := k \in \mathbb{N}$, et soit $(\xi_i,\, i \geq 1)$ une suite de variables aléatoires i.i.d. avec $\mathbb{P}(\xi_1 = 1) = \frac{1}{2} = \mathbb{P}(\xi_1 = -1)$. On pose

$$X_n := k + \sum_{i=1}^{n} \xi_i\,, \qquad n \geq 1,$$

ainsi que $\mathscr{F}_0 := \{\Omega, \varnothing\}$ et $\mathscr{F}_n := \sigma(X_1, \cdots, X_n) = \sigma(\xi_1, \cdots, \xi_n)$. Comme $\mathbb{E}(\xi_1) = 0$, on sait que $(X_n,\, n \geq 0)$ est une martingale (voir l'Exemple 1.4). Posons

$$T := \inf\{n \geq 0 : X_n = 0\}, \qquad \inf \varnothing := \infty.$$

Alors T est un temps d'arrêt (Exemple 2.2). Du Théorème 3.1, on déduit que $(Y_n := X_{T\wedge n},\, n \geq 0)$ est une martingale positive, à laquelle on peut appliquer le Corollaire 4.5 : Y_n converge p.s. vers $Y_\infty < \infty$.

Puisque pour $\omega \in \{T = \infty\}$, on a $|Y_{n+1}(\omega) - Y_n(\omega)| = 1$, $\forall n \geq 0$, ce qui empêche $Y_n(\omega)$ de converger quand $n \to \infty$. Donc $T < \infty$ p.s.

Ainsi, $Y_\infty = 0$ p.s., et donc $\mathbb{E}(Y_\infty \,|\, \mathscr{F}_n) \leq Y_n$ p.s., qui n'est pas une identité, bien que $(Y_n,\, n \geq 0)$ soit une martingale. Cet exemple, simple mais utile, montre également que la convergence dans le Corollaire 4.5 (ou dans le Théorème 4.1 en général) n'a pas nécessairement lieu dans L^1 : on a dans cet exemple $\mathbb{E}(Y_n) = k$, $\forall n$, tandis que $\mathbb{E}(Y_\infty) = 0$.
□

4.2. Convergence dans L^p

On étudie à présent convergence dans L^p, $1 < p < \infty$, pour les martingales.

Théorème 4.7. *Soit* $1 < p < \infty$. *Si* (X_n) *est une martingale telle que*

$$\sup_n \mathbb{E}(|X_n|^p) < \infty,$$

alors il existe une variable aléatoire X_∞ *telle que* $X_n \to X_\infty$ p.s. *et dans* L^p.

Preuve. On a $[\mathbb{E}(X_n^+)]^p \leq [\mathbb{E}(|X_n|)]^p \leq \mathbb{E}(|X_n|^p)$, et donc $\sup_n \mathbb{E}(X_n^+) < \infty$ par hypothèse. Le Théorème 4.1 nous dit alors que $X_\infty := \lim_n X_n$ existe p.s.

Pour avoir la convergence dans L^p, on remarque que d'après l'inégalité de Doob (Corollaire 3.7), $\sup_n |X_n| \in L^p$. Puisque $|X_n - X_\infty|^p \leq (2 \sup_n |X_n|)^p$, le théorème de convergence dominée implique que $\mathbb{E}(|X_n - X_\infty|^p) \to 0$, $n \to \infty$. □

4.3. Intégrabilité uniforme

Cette section est consacrée à la notion d'intégrabilité uniforme pour une famille de variables aléatoires réelles. Cette notion jouera ensuite un rôle crucial dans l'étude de convergence dans L^1 pour les martingales.

Soit $(X_i,\ i \in I)$ une famille de variables aléatoires réelles, indexée par un ensemble I non vide, définie sur $(\Omega,\ \mathscr{F},\ \mathbb{P})$.

Définition 4.8. *On dit que* $(X_i,\ i \in I)$ *est uniformément intégrable* (*ou : équi-intégrable*) *si*

$$\lim_{a\to\infty} \sup_{i\in I} \mathbb{E}[\,|X_i|\,\mathbf{1}_{\{|X_i|>a\}}\,] = 0.$$

Exemple 4.9. Soit Y une variable aléatoire réelle intégrable. Alors Y est uniformément intégrable.

En effet, $\mathbb{E}[\,|Y|\,\mathbf{1}_{\{|Y|>a\}}\,] \to 0$ (quand $a \to \infty$) par convergence dominée. □

Exemple 4.10. Soient $(X_i,\ i \in I)$ et $(Y_i,\ i \in I)$ des familles de variables aléatoires réelles telle que $|X_i| \le Y_i$, $\forall i \in I$. Si, $(Y_i,\ i \in I)$ est uniformément intégrable, alors $(X_i,\ i \in I)$ l'est également.

En particulier, si $|X_i| \le Y$, $\forall i \in I$, où Y est intégrable. Alors $(X_i,\ i \in I)$ est uniformément intégrable. Plus généralement, si $\sup_{i\in I} |X_i|$ est mesurable et si $\mathbb{E}(\sup_{i\in I} |X_i|) < \infty$, alors $(X_i,\ i \in I)$ est uniformément intégrable. □

Exemple 4.11. Soit $(X_i,\ i \in I)$ une famille de variables aléatoires réelles uniformément intégrable, alors $(X_i^+,\ i \in I)$ l'est également.

En effet, $|X_i^+| \le |X_i|$, $\forall i \in I$. □

Exemple 4.12. Soit $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ une fonction mesurable telle que $\lim_{x\to\infty} \frac{\Phi(x)}{x} = \infty$. Soit $(X_i,\ i \in I)$ une famille de variables aléatoires réelles telle que $\sup_{i\in I} \mathbb{E}[\Phi(|X_i|)] < \infty$. Alors $(X_i,\ i \in I)$ est uniformément intégrable.

Il suffit, en effet, de remarquer que $\mathbb{E}[\,|X_i|\,\mathbf{1}_{\{|X_i|>a\}}] \le (\sup_{x>a} \frac{x}{\Phi(x)})\, \mathbb{E}[\Phi(|X_i|)]$. □

Exemple 4.13. Soit $1 < p < \infty$. Soit $(X_i,\ i \in I)$ une famille de variables aléatoires réelles telle que $\sup_{i\in I} \mathbb{E}[\,|X_i|^p\,] < \infty$. Alors $(X_i,\ i \in I)$ est uniformément intégrable.

C'est un cas particulier (avec $\Phi(x) := x^p$, $x \in \mathbb{R}_+$) de l'exemple précédent. □

Théorème 4.14. *Une famille de variables aléatoires réelles* $(X_i,\ i \in I)$ *est uniformément intégrable si et seulement si les deux conditions suivantes sont satisfaites :*

(i) $\sup_{i\in I} \mathbb{E}[\,|X_i|\,] < \infty$ *;*

(ii) $\forall\, \varepsilon > 0$, $\exists\, \delta > 0$ *tel que* $A \in \mathscr{F}$, $\mathbb{P}(A) \le \delta \;\Rightarrow\; \sup_{i\in I} \mathbb{E}[\,|X_i|\,\mathbf{1}_A] \le \varepsilon$.

Preuve. "$\Rightarrow$" On peut choisir $a > 0$ suffisamment grand tel que $\sup_{i\in I} \mathbb{E}[\,|X_i|\,\mathbf{1}_{\{|X_i|>a\}}] \le 1$, et donc $\sup_{i\in I} \mathbb{E}[\,|X_i|\,] \le 1 + a < \infty$.

Pour vérifier la condition (ii), on constate que pour tout $\varepsilon > 0$, il existe $a = a(\varepsilon) > 0$ tel que $\sup_{i\in I} \mathbb{E}[\,|X_i|\,\mathbf{1}_{\{|X_i|>a\}}] \le \frac{\varepsilon}{2}$. Donc si $\mathbb{P}(A) \le \frac{\varepsilon}{2a}$, on a, pour tout $i \in I$,

$$\begin{aligned}
\mathbb{E}(|X_i|\,\mathbf{1}_A) &= \mathbb{E}(|X_i|\,\mathbf{1}_{A\cap\{|X_i|>a\}}) + \mathbb{E}(|X_i|\,\mathbf{1}_{A\cap\{|X_i|\le a\}}) \\
&\le \mathbb{E}(|X_i|\,\mathbf{1}_{\{|X_i|>a\}}) + a\,\mathbb{P}(A) \\
&\le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

D'où (ii).

"$\Leftarrow$" Soit $\varepsilon > 0$, et soit $\delta = \delta(\varepsilon) > 0$ qui satisfait (ii). Soit $M := \sup_{i\in I} \mathbb{E}[\,|X_i|\,] < \infty$. Lorsque $a > \frac{M}{\delta}$, on a, d'après l'inégalité de Markov, $\mathbb{P}(|X_i| > a) \le \frac{M}{a} < \delta$, et donc par (ii), $\sup_{i\in I} \mathbb{E}[\,|X_i|\,\mathbf{1}_{\{|X_i|>a\}}\,] \le \varepsilon$. Par conséquent, $(X_i,\ i \in I)$ est uniformément intégrable. □

Exemple 4.15. Si $(X_i,\ i \in I)$ et $(Y_i,\ i \in I)$ sont deux familles de variables aléatoires uniformément intégrables. Alors $(X_i + Y_i,\ i \in I)$ est aussi uniformément intégrable.

Il suffit d'appliquer le Théorème 4.14. [On peut aussi directement prouver le résultat en utilisant seulement la définition.] □

Théorème 4.16. *Soient X, X_1, X_2, $\cdots$ des variables aléatoires réelles. Alors*

$$X_n \xrightarrow{L^1} X \iff (X_n) \text{ uniformément intégrable, et } X_n \xrightarrow{P.} X.$$

Preuve. "$\Rightarrow$" Il suffit de vérifier l'intégrabilité uniforme. Comme $\mathbb{E}[\,|X_n|\,] \to \mathbb{E}[\,|X|\,]$, il est clair que $\sup_n \mathbb{E}[\,|X_n|\,] < \infty$. D'autre part, pour tout $\varepsilon > 0$, il existe $N = N(\varepsilon) < \infty$ tel que $\mathbb{E}[\,|X_n - X|\,] \le \varepsilon$ pour tout $n > N$. Soit $\delta > 0$ tel que $\mathbb{P}(A) \le \delta \Rightarrow \max_{1\le i\le N} \mathbb{E}[\,|X_i|\,\mathbf{1}_A] \le \varepsilon$ (ce qui est possible pour un nombre fini de variables aléatoires) et $\mathbb{E}[\,|X|\,\mathbf{1}_A] \le \varepsilon$. Alors $\sup_{n>N} \mathbb{E}[\,|X_n|\,\mathbf{1}_A] \le \mathbb{E}[\,|X|\,\mathbf{1}_A] + \sup_{n>N} \mathbb{E}[\,|X_n - X|\,] \le 2\varepsilon$ et donc $\sup_{n\ge1} \mathbb{E}[\,|X_n|\,\mathbf{1}_A] \le 2\varepsilon$ pour tout A avec $\mathbb{P}(A) < \delta$. D'après le Théorème 4.14, (X_n) est uniformément intégrable.

"$\Leftarrow$" Comme $\sup_n \mathbb{E}(|X_n|) < \infty$, le lemme de Fatou (version convergence en probabilité) nous dit que X est intégrable. Pour tout $a > 0$, soit $h_a : \to \mathbb{R}_+$ la fonction continue et à support compact définie par:

$$h_a(x) := \begin{cases} a & \text{if } x > a, \\ x & \text{if } x \in [-a,\, a], \\ -a & \text{if } x < -a\,. \end{cases}$$

Il est clair que $|h_a(x) - x| = (|x| - a)^+ \leq |x| \, \mathbf{1}_{\{|x|>a\}}$, on a

$$\begin{aligned} |X_n - X| &\leq |X_n - h_a(X_n)| + |h_a(X_n) - h_a(X)| + |X - h_a(X)| \\ &\leq |X_n| \, \mathbf{1}_{\{|X_n|>a\}} + |X| \, \mathbf{1}_{\{|X|>a\}} + |h_a(X_n) - h_a(X)| \,. \end{aligned}$$

Soit $\varepsilon > 0$. Comme (X_n) est uniformément intégrable, et X est intégrable, on peut *fixer* $a > 0$ tel que $\mathbb{E}[\, |X_n| \, \mathbf{1}_{\{|X_n|>a\}} \,] + \mathbb{E}[\, |X| \, \mathbf{1}_{\{|X|>a\}} \,] \leq \varepsilon$, $\forall n$. D'autre part, par convergence dominée (version convergence en probabilité), on a $\mathbb{E}[\, |h_a(X_n) - h_a(X)| \,] \to 0$, $n \to \infty$. Il existe donc $N < \infty$ tel que $|\mathbb{E}[h_a(X_n)] - \mathbb{E}[h_a(X)]| \leq \varepsilon$ pour tout $n > N$. On a alors prouvé que pour tout $n > N$, $\mathbb{E}(|X_n - X|) \leq 2\varepsilon$. D'où convergence dans L^1. $\square$

Exemple 4.17. ("Théorème de convergence dominée généralisé"). Si $X_n \to X$ en probabilité, et si (X_n) est uniformément intégrable, alors $\mathbb{E}(X_n) \to \mathbb{E}(X)$, $n \to \infty$.

Il suffit de rappeler que convergence dans L^1 implique convergence des espérances. $\square$

Théorème 4.18. *Si Y est une variable aléatoire réelle intégrable, alors*

$$\Big(\mathbb{E}(Y \,|\, \mathscr{G});\ \mathscr{G} \subset \mathscr{F} \text{ sous-tribu}\Big) \text{ est uniformément intégrable.}$$

Preuve. Sans perte de généralité, on peut supposer que $Y \geq 0$, p.s. (sinon, on remplace Y par $|Y|$). Soit $\varepsilon > 0$. Il existe $\delta > 0$ tel que $\mathbb{P}(A) \leq \delta \Longrightarrow \mathbb{E}(Y \, \mathbf{1}_A) \leq \varepsilon$.

Soit a suffisamment grand tel que $\mathbb{E}(\frac{Y}{a}) \leq \delta$. Soit $A := \{\mathbb{E}(Y \,|\, \mathscr{G}) \geq a\}$. Par l'inégalité de Markov, $\mathbb{P}(A) \leq \frac{\mathbb{E}(Y)}{a} \leq \delta$. D'autre part, puisque $A \in \mathscr{G}$, on a

$$\mathbb{E}\Big[\, \mathbb{E}(Y \,|\, \mathscr{G}) \, \mathbf{1}_A \Big] = \mathbb{E}[\, Y \, \mathbf{1}_A \,] \leq \varepsilon,$$

vu que $\mathbb{P}(A) \leq \delta$. On a donc prouvé que pour tout $\varepsilon > 0$, il existe a tel que $\forall \, n$,

$$\mathbb{E}\Big[\mathbb{E}(Y \,|\, \mathscr{G}) \, \mathbf{1}_{\{\mathbb{E}(Y \,|\, \mathscr{G}) \geq a\}}\Big] \leq \varepsilon.$$

Par définition, $(\mathbb{E}(Y \,|\, \mathscr{G});\ \mathscr{G} \subset \mathscr{F}$ sous-tribu) est uniformément intégrable. $\square$

4.4. Convergence dans L^1

On étudie maintenant convergence dans L^1 pour les martingales.

Théorème 4.19. *Soit (X_n) une sous-martingale uniformément intégrable. Alors*

(i) $X_n \to X_\infty$ *dans* L^1 *;*

(ii) $X_n \to X_\infty$ p.s. *;*

(iii) $\forall n,\ X_n \leq \mathbb{E}(X_\infty \,|\, \mathscr{F}_n)$, p.s.

Preuve. L'intégrabilité uniforme implique que $\sup_n \mathbb{E}(|X_n|) < \infty$. Par le Théorème 4.1, $X_n \to X_\infty$ p.s., qui est intégrable. La convergence dans L^1, quant à elle, est une conséquence de converge p.s. et de l'intégrabilité uniforme (et n'a rien à voir avec la propriété de sous-martingale).

Par la Proposition 1.7, si $m > n$ et $A \in \mathscr{F}_n$, alors $\mathbb{E}(X_m \mathbf{1}_A) \ge \mathbb{E}(X_n \mathbf{1}_A)$. En faisant $m \to \infty$, et vu la convergence dans L^1, on a $\mathbb{E}(X_m \mathbf{1}_A) \to \mathbb{E}(X_\infty \mathbf{1}_A)$. Donc $\mathbb{E}(X_n \mathbf{1}_A) \le \mathbb{E}(X_\infty \mathbf{1}_A)$ pour tout $A \in \mathscr{F}_n$. Autrement dit, $X_n \le \mathbb{E}(X_\infty \,|\, \mathscr{F}_n)$, p.s. □

Exemple 4.20. (Ruine du joueur). Considérons de nouveau le jeu de pile ou face (voir l'Exemple 4.6), avec $X_0 = k \ge 1$. Soit $m \ge k$ un entier. On pose cette fois-ci

$$T := \inf\{n \ge 0 : \; X_n = 0 \text{ ou } X_n = m\}, \qquad \inf \varnothing := \infty . \tag{4.2}$$

On a vu dans l'Exemple 4.6 que $T < \infty$ p.s. La martingale $(Y_n := X_{T\wedge n},\, n \ge 0)$ est uniformément intégrable car bornée. Le Théorème 4.19 nous dit donc que $\mathbb{E}(Y_\infty) = \mathbb{E}(Y_0) = k$, soit

$$m\,\mathbb{P}(X_T = m) = k .$$

Il résulte alors que

$$\mathbb{P}(X_T = m) = \frac{k}{m}, \qquad \mathbb{P}(X_T = 0) = \frac{m-k}{m} . \tag{4.3}$$

On peut généraliser le calcul pour le *jeu de pile ou face biaisé* : On suppose que $X_0 := k$ et $X_n := k + \sum_{i=1}^n \xi_i$, $n \ge 1$, où les variables aléatoires ξ_i, $i \ge 1$, sont i.i.d. avec

$$\mathbb{P}(\xi_1 = 1) = p, \qquad \mathbb{P}(\xi_1 = 1) = q := 1 - p ,$$

où $p \in]0,\, 1[\setminus \{\frac12\}$ est un paramètre fixé. On vérifie assez facilement que $Y_n := (\frac{q}{p})^{X_n}$ est une martingale. Si T est le temps d'arrêt défini par (4.2), le fait que la martingale positive $(Y_{T\wedge n},\, n \ge 0)$ converge p.s. entraîne que $T < \infty$ p.s.

Appliquons le Théorème 4.19 à la martingale bornée $(Y_{T\wedge n},\, n \ge 0)$ qui est uniformément intégrable (puisque bornée), on obtient

$$(\frac{q}{p})^k = \mathbb{E}[(\frac{q}{p})^{X_T}] = (\frac{q}{p})^m\, \mathbb{P}(X_T = m) + \mathbb{P}(X_T = 0) ,$$

ce qui donne

$$\mathbb{P}(X_T = m) = \frac{(\frac{q}{p})^k - 1}{(\frac{q}{p})^m - 1}, \qquad \mathbb{P}(X_T = 0) = \frac{(\frac{q}{p})^m - (\frac{q}{p})^k}{(\frac{q}{p})^m - 1} .$$

Ce résultat est, informellement, en accord avec ce que l'on a trouvé dans (4.3) pour le jeu sans biaisé en faisant $p \to \frac12$. □

D'après le Théorème 4.19, une martingale qui converge dans L^1, donc uniformément intégrable, est fermée. La réciproque est également vraie.

Théorème 4.21. (Lévy). *Soit Y une variable aléatoire réelle intégrable. Alors*

$$\mathbb{E}(Y \,|\, \mathscr{F}_n) \to \mathbb{E}(Y \,|\, \mathscr{F}_\infty), \qquad \text{p.s. } \textit{et dans } L^1.$$

Preuve. Soit $X_n := \mathbb{E}(Y \,|\, \mathscr{F}_n)$, qui est une martingale fermée (voir l'Exemple 1.3). D'après le Théorème 4.18, (X_n) est uniformément intégrable. Donc, par le Théorème 4.19, $X_n \to X_\infty$ p.s. et dans L^1, et $X_n = \mathbb{E}(X_\infty \,|\, \mathscr{F}_n)$ p.s. Par conséquent,

$$\mathbb{E}(X_\infty \mathbf{1}_A) = \mathbb{E}(Y \mathbf{1}_A), \qquad \forall\, A \in \mathscr{F}_n.$$

Soient $\mathscr{M} := \{A \in \mathscr{F}_\infty : \mathbb{E}(X_\infty \mathbf{1}_A) = \mathbb{E}(Y \mathbf{1}_A)\}$ et $\mathscr{C} := \cup_{n=1}^\infty \mathscr{F}_n$. On voit que $\mathscr{C} \subset \mathscr{M}$, et que $\mathscr{C}$ est de toute évidence stable par intersections finies. Vérifions que $\mathscr{M}$ est une classe monotone : (i) $\Omega \in \mathscr{M}$ est évident vu l'identité précédente ; (ii) si A, $B \in \mathscr{M}$ avec $A \subset B$, on voit immédiatement que $\mathbb{E}(X_\infty \mathbf{1}_{B\backslash A}) = \mathbb{E}(Y \mathbf{1}_{B\backslash A})$ en faisant la soustraction entre deux équations, et donc $(B\backslash A) \in \mathscr{M}$; (iii) enfin, si $A_n \uparrow$ et si $A_n \in \mathscr{M}$, $\forall n$, alors $\mathbb{E}(X_\infty \mathbf{1}_{A_n}) \to \mathbb{E}(X_\infty \mathbf{1}_A)$ avec $A := \cup_{n=1}^\infty A_n$ (convergence dominée), et $\mathbb{E}(Y \mathbf{1}_{A_n}) \to \mathbb{E}(Y \mathbf{1}_A)$, et donc $\mathbb{E}(X_\infty \mathbf{1}_A) = \mathbb{E}(Y \mathbf{1}_A)$; autrement dit, $A \in \mathscr{M}$.

Par classe monotone, $\mathscr{M} \supset \sigma(\mathscr{C}) = \mathscr{F}_\infty$. Donc $\mathbb{E}(X_\infty \mathbf{1}_A) = \mathbb{E}(Y \mathbf{1}_A)$ pour tout $A \in \mathscr{F}_\infty$. Comme X_∞ est $\mathscr{F}_\infty$-mesurable, on obtient $X_\infty = \mathbb{E}(Y \,|\, \mathscr{F}_\infty)$. □

Exemple 4.22. Soit $(\mathscr{F}_n)_{n\geq 0}$ une filtration, et soit $A \in \mathscr{F}_\infty$. Alors

$$\mathbb{P}(A \,|\, \mathscr{F}_n) \to \mathbf{1}_A\,, \qquad \text{p.s. et dans } L^1.$$

Il suffit d'appliquer le théorème de Lévy (Théorème 4.21) à la variable aléatoire intégrable $Y := \mathbf{1}_A$. □

On regroupe certains résultats obtenus dans cette section sous la forme d'un joli petit corollaire :

Corollaire 4.23. *Soit $(X_n,\, n \geq 0)$ une martingale. Alors les assertions suivantes sont équivalentes :*

(i) $X_n \to X_\infty$ *dans* L^1 *;*

(ii) $(X_n,\, n \geq 0)$ *une martingale fermée ;*

(iii) $(X_n,\, n \geq 0)$ *est uniformément intégrable.*

Preuve. "(i) ⇒ (iii)" Voir le Théorème 4.16.

"(iii) ⇒ (ii)" Voir le Théorème 4.19.

"(ii) ⇒ (i)" Voir le Théorème 4.21. □

5. Exemple : Processus de branchement

Soit μ une mesure de probabilité sur $\mathbb{N}$ telle que[13]

$$m := \sum_{i=0}^{\infty} i\,\mu(i) < \infty\,.$$

On exclut les cas triviaux où μ est la mesure de Dirac en 1 ou la mesure de Dirac en 0.

Soit $(\xi_{n,\,i},\, n \ge 0,\, i \ge 1)$ une famille de variables aléatoires i.i.d. de loi μ. On fixe un entier $k \ge 1$, et l'on définit par récurrence une suite $(X_n,\, n \ge 0)$ de variables aléatoires à valeurs dans $\mathbb{N}$ en posant $X_0 := k$ et

$$X_{n+1} := \sum_{i=1}^{X_n} \xi_{n,\,i}\,, \qquad n \ge 0\,,$$

avec la convention $\sum_{i=1}^{0} := 0$.

La quantité X_n s'interprète comme le nombre d'individus dans une population à la génération n, sachant que le nombre d'enfants de chaque individu suit la loi μ (que l'on appelle loi de reproduction), en supposant que les nombres d'enfants des différents individus sont des variables aléatoires indépendantes.

Posons $\mathscr{F}_0 := \{\varnothing, \Omega\}$ et pour $n \ge 1$,

$$\mathscr{F}_n := \sigma(\xi_{\ell,\,i},\, 0 \le \ell < n,\, i \ge 1)\,.$$

On voit aisément que le processus (X_n) est adapté par rapport à la filtration $(\mathscr{F}_n,\, n \ge 0)$. Pour tout $n \ge 0$, comme $X_{n+1} = \sum_{i=1}^{\infty} \xi_{n,\,i}\, \mathbf{1}_{\{i \le X_n\}}$, on a, d'après l'Exemple 4.21 du Chapitre 1,

$$\begin{aligned}
\mathbb{E}(X_{n+1} \,|\, \mathscr{F}_n) &= \sum_{i=1}^{\infty} \mathbb{E}(\xi_{n,\,i}\, \mathbf{1}_{\{i \le X_n\}} \,|\, \mathscr{F}_n) \\
&= \sum_{i=1}^{\infty} \mathbf{1}_{\{i \le X_n\}}\, \mathbb{E}(\xi_{n,\,i} \,|\, \mathscr{F}_n) \\
&= \sum_{i=1}^{\infty} \mathbf{1}_{\{i \le X_n\}}\, \mathbb{E}(\xi_{n,\,i}) = m\, X_n\,, \qquad \text{p.s.,}
\end{aligned}$$

[13] On écrit $\mu(i)$ à la place de $\mu(\{i\})$.

où l'on a utilisé les faits que X_n est $\mathscr{F}_n$-mesurable, et que $\xi_{n,i}$ est indépendante de $\mathscr{F}_n$, avec $\mathbb{E}(\xi_{n,i}) = m$.[14] Par conséquent,

$$\mathbb{E}\Big(\frac{X_{n+1}}{m^{n+1}} \,\Big|\, \mathscr{F}_n\Big) = \frac{X_n}{m^n}, \qquad \text{p.s.}$$

Ceci montre d'abord que X_n est intégrable, et ensuite que $(\frac{X_n}{m^n}, \, n \ge 0)$ est une martingale positive. [Une récurrence immédiate montre que $\mathbb{E}(X_n) = k\, m^n$ pour tout $n \ge 0$.] D'après le Théorème 4.1 (ou le Corollaire 4.5), $\frac{X_n}{m^n}$ converge p.s. vers une limite finie.

On distingue trois situations possibles :

• $m < 1$ (cas sous-critique). Puisque X_n est à valeurs entières, la convergence de $\frac{X_n}{m^n}$ vers une quantité finie n'est possible que si $X_n = 0$ pour tout n suffisamment grand (extinction presque sûre de la population).

• $m = 1$ (cas critique). Dans ce cas, (X_n) est une martingale positive, et on aura la même conclusion (extinction presque sûre) une fois que l'on saura prouver que

$$\mathbb{P}(\exists N \ge 1, \ j \ge 1 : \ X_n = j, \, \forall n \ge N) = 0. \tag{5.1}$$

Pour voir pourquoi l'identité (5.1) est vraie, on constate que

$$\mathbb{P}(\exists N \ge 1, \ j \ge 1 : \ X_n = j, \, \forall n \ge N) \le \sum_{N=1}^{\infty} \sum_{j=1}^{\infty} \mathbb{P}(X_n = j, \, \forall n \ge N) \, .$$

Pour tous $N \ge 1$, $j \ge 1$ et $k \ge 1$,

$$\begin{aligned} & \mathbb{P}(X_n = j, \, \forall n \ge N) \\ \le \ & \mathbb{P}(X_n = j, \, \forall N \le n \le N+k) \\ = \ & \mathbb{P}(X_{N+k} = j \,|\, X_n = j, \, \forall N \le n \le N+k-1) \times \mathbb{P}(X_n = j, \, \forall N \le n \le N+k-1) \, . \end{aligned}$$

Comme $X_{N+k} = \sum_{i=1}^{X_{N+k-1}} \xi_{N+K-1,\, i}$, et $(\xi_{N+K-1,\, i}, \, i \ge 1)$ est indépendante de $(X_n, \, N \le n \le N+k-1)$, on a $\mathbb{P}(X_{N+k} = j \,|\, X_n = j, \, \forall N \le n \le N+k-1) = \mathbb{P}(\sum_{i=1}^{j} \xi_{0,\, i} = j) =: p < 1$ (car $\mu \ne \delta_1$). Donc

$$\mathbb{P}(X_n = j, \, \forall N \le n \le N+k) = p \times \mathbb{P}(X_n = j, \, \forall N \le n \le N+k-1) \, .$$

En itérant l'argument, on trouve :

$$\mathbb{P}(X_n = j, \, \forall N \le n \le N+k) = p^k \times \mathbb{P}(X_N = j) \le p^k \, .$$

[14] En fait, on vient de refaire le même calcul que celui dans l'Exemple 5.2 du Chapitre 1.

Ainsi, $\mathbb{P}(X_n = j, \forall n \geq N) \leq p^k$. Comme $p < 1$ et k peut être arbitrairement grand, on a $\mathbb{P}(X_n = j, \forall n \geq N) = 0$: d'où (5.1).

• $m > 1$ (cas surcritique). On a, lorsque $n \to \infty$,

$$\frac{X_n}{m^n} \to Z, \qquad \text{p.s.}, \tag{5.2}$$

et sur l'ensemble $\{Z > 0\}$, on voit que X_n est de l'ordre de m^n quand n est grand. On voudrait alors vérifier que $\mathbb{P}(Z > 0) > 0$. Remarquons que si la convergence (5.2) a lieu dans L^1, on a $\mathbb{P}(Z > 0) > 0$ car dans ce cas, $\mathbb{E}(Z) = \lim_{n\to\infty} \mathbb{E}(\frac{X_n}{m^n}) = k$.

On suppose que la loi de reproduction μ satisfait

$$\sum_{i=1}^{\infty} i^2 \, \mu(i) < \infty \,.$$

En écrivant $X_{n+1}^2 = \sum_{i=1}^\infty \sum_{j=1}^\infty \xi_{n,i}\xi_{n,j} \, \mathbf{1}_{\{i\leq X_n\}} \, \mathbf{1}_{\{j\leq X_n\}}$, on obtient, grâce à l'Exemple 4.21 du Chapitre 1,

$$\begin{aligned} \mathbb{E}(X_{n+1}^2 \,|\, \mathscr{F}_n) &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \mathbf{1}_{\{i\leq X_n\}} \, \mathbf{1}_{\{j\leq X_n\}} \, \mathbb{E}(\xi_{n,i}\,\xi_{n,j} \,|\, \mathscr{F}_n) \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \mathbf{1}_{\{i\leq X_n\}} \, \mathbf{1}_{\{j\leq X_n\}} \, \mathbb{E}(\xi_{n,i}\,\xi_{n,j}) \,, \qquad \text{p.s.} \end{aligned}$$

Pour tout couple i, $j \geq 1$, $\mathbb{E}(\xi_{n,i}\,\xi_{n,j})$ vaut m^2 si $i \neq j$ (avec $m := \sum_{i=1}^\infty i\,\mu(i) \in]1, \infty[$ comme avant), et vaut $m^2 + \sigma^2$ si $i = j$, où $\sigma^2 := \sum_{i=1}^\infty i^2\mu(i) - m^2$ représente la variance de $\xi_{0,1}$. Donc

$$\mathbb{E}(X_{n+1}^2 \,|\, \mathscr{F}_n) = m^2 X_n^2 + \sigma^2 X_n \,, \qquad \text{p.s.}$$

En posant $a_n := m^{-2n}\, \mathbb{E}(X_n^2)$, on trouve

$$a_{n+1} = a_n + \frac{k\sigma^2}{m^{n+2}} \,.$$

Puisque $m > 1$, la suite (a_n) est convergeante, a fortiori bornée : $\sup_{n\geq 0} \mathbb{E}[(\frac{X_n}{m^n})^2] < \infty$. D'après le Théorème 4.7, $\frac{X_n}{m^n} \to Z$ p.s. et dans L^2. Ceci entraîne que $\mathbb{P}(Z > 0) > 0$.

[Dans la littérature, un théorème célèbre de Kesten–Stigum dit que la convergence (5.2) a lieu dans L^1 si et seulement si $\sum_{i=1}^\infty i\,\mu(i) \log i < \infty$, et qu'alors $Z > 0$ p.s. sur l'ensemble $\{\liminf_{n\to\infty} X_n > 0\}$ de non-extinction. Pour ce dernier point ainsi que d'autres propriétés intéressantes, voir les discussions en TD à la fin du semestre.]

6. Théorème d'arrêt

On a vu en Section 3 que si (X_n) est une sous-martingale, et si $S \le T$ sont deux temps d'arrêt *bornés*, alors $\mathbb{E}(X_T \,|\, \mathscr{F}_S) \ge X_S$, p.s. Dans cette section, on étudie le cas où S et T ne sont pas bornés.

Théorème 6.1. *Soit (X_n) une sous-martingale uniformément intégrable. Alors pour tout temps d'arrêt T, $(X_{T\wedge n})$ est une sous-martingale uniformément intégrable.*

Preuve. On a déjà vu dans le Théorème 3.1 que $(X_{T\wedge n})$ est une sous-martingale (sans l'hypothèse d'intégrabilité uniforme). Il reste donc de prouver qu'elle est uniformément intégrable.

Comme (X_n^+) est une sous-martingale (Exemple 1.12), la Conséquence 1 du Théorème 3.2 nous dit que $\mathbb{E}(X_{T\wedge n}^+) \le \mathbb{E}(X_n^+)$. Et puisque (X_n^+) est uniformément intégrable, on obtient

$$\sup_n \mathbb{E}(X_{T\wedge n}^+) \le \sup_n \mathbb{E}(X_n^+) < \infty.$$

Le Théorème 4.1 nous dit alors que $X_{T\wedge n} \to X_T$, p.s. et que $\mathbb{E}(|X_T|) < \infty$. Pour $a > 0$,

$$\begin{aligned} \mathbb{E}[\,|X_{T\wedge n}|\,\mathbf{1}_{\{|X_{T\wedge n}|\ge a\}}\,] &= \mathbb{E}[\,|X_T|\,\mathbf{1}_{\{|X_T|\ge a\}\cap\{T\le n\}}\,] + \mathbb{E}[\,|X_n|\,\mathbf{1}_{\{|X_n|\ge a\}\cap\{T>n\}}\,] \\ &\le \mathbb{E}[\,|X_T|\,\mathbf{1}_{\{|X_T|\ge a\}}\,] + \mathbb{E}[\,|X_n|\,\mathbf{1}_{\{|X_n|\ge a\}}\,]\,. \end{aligned}$$

Comme (X_n) est uniformément intégrable, l'inégalité ci-dessus implique que $(X_{T\wedge n})$ est aussi uniformément intégrable. □

Exemple 6.2. Soit (X_n) une sous-martingale uniformément intégrable. Si T est un temps d'arrêt, alors

$$\mathbb{E}(X_0) \le \mathbb{E}(X_T) \le \mathbb{E}(X_\infty),$$

où X_∞ est la limite p.s. et dans L^1 de X_n quand $n \to \infty$.

En effet, par la Conséquence 1 du Théorème 3.2, $\mathbb{E}(X_0) \le \mathbb{E}(X_{T\wedge n}) \le \mathbb{E}(X_n)$. On fait $n \to \infty$. Par les Théorèmes 6.1 et 4.19, $X_n \to X_\infty$ dans L^1 et $X_{T\wedge n} \to X_T$ dans L^1. D'où la conclusion cherchée. □

Théorème 6.3. (Théorème d'arrêt). *Soit (X_n) une sous-martingale uniformément intégrable. Alors pour tous temps d'arrêt $S \le T$, on a*

$$\mathbb{E}(X_T \,|\, \mathscr{F}_S) \ge X_S, \qquad \text{p.s.}$$

Preuve. D'après le Théorème 6.1 et l'Exemple 6.2, $\mathbb{E}(X_S) \le \mathbb{E}(X_T)$.

Soit $A \in \mathscr{F}_S$, et soit

$$R = S\,\mathbf{1}_A + T\,\mathbf{1}_{A^c}\,,$$

qui est un temps d'arrêt, tel que $R \le T$. D'après ce que l'on vient de démontrer, $\mathbb{E}(X_R) \le \mathbb{E}(X_T)$. Puisque $R = T$ sur A^c, on a $\mathbb{E}(X_S\,\mathbf{1}_A) = \mathbb{E}(X_R\,\mathbf{1}_A) \le \mathbb{E}(X_T\,\mathbf{1}_A)$. Ceci étant vrai pour tout $A \in \mathscr{F}_S$, on obtient $\mathbb{E}(X_T \,|\, \mathscr{F}_S) \ge X_S$, p.s. □

7. Martingales de carré intégrable

On se met dans un espace de probabilité filtré $(\Omega,\, \mathscr{F},\, (\mathscr{F}_n),\, \mathbb{P})$.

Un processus aléatoire $(Y_n,\ n \ge 1)$ est dit prévisible si pour tout $n \ge 1$, Y_n est $\mathscr{F}_{n-1}$-mesurable.

Théorème 7.1. (Décomposition de Doob). *Toute sous-martingale (X_n) peut s'écrire comme $X_n = M_n + A_n$, où (M_n) est une martingale, et (A_n) est prévisible telle que $A_0 = 0$.*

Cette décomposition est unique. De plus, le processus (A_n) est croissant.[15]

Preuve. (Unicité). Si (M, A) vérifie les conditions du théorème, alors

$$\mathbb{E}[X_n \,|\, \mathscr{F}_{n-1}] = \mathbb{E}[M_n \,|\, \mathscr{F}_{n-1}] + \mathbb{E}[A_n \,|\, \mathscr{F}_{n-1}] = M_{n-1} + A_n = X_{n-1} - A_{n-1} + A_n\,.$$

Donc on a nécessairement $A_n - A_{n-1} = \mathbb{E}[X_n \,|\, \mathscr{F}_{n-1}] - X_{n-1}$. La condition $A_0 = 0$ implique que le choix de (A_n) (dont aussi celui de (M_n)) est unique.

(Existence) Soit le processus (A_n) tel que $A_n - A_{n-1} = \mathbb{E}[X_n \,|\, \mathscr{F}_{n-1}] - X_{n-1}$, et que $A_0 = 0$. Comme (X_n) est une sous-martingale, on a $A_n - A_{n-1} \ge 0$, et (A_n) est donc croissant. De plus, $A_n = \sum_{i=1}^n \{\mathbb{E}[X_i \,|\, \mathscr{F}_{i-1}] - X_{i-1}\}$ est $\mathscr{F}_{n-1}$-mesurable.

Pour montrer que $(M_n := X_n - A_n)$ est une martingale, on constate que M_n est intégrable et adapté, tel que

$$\mathbb{E}[M_n \,|\, \mathscr{F}_{n-1}] = \mathbb{E}[X_n - A_n \,|\, \mathscr{F}_{n-1}] = \mathbb{E}[X_n \,|\, \mathscr{F}_{n-1}] - A_n = X_{n-1} - A_{n-1} = M_{n-1}\,.$$

Donc (M_n) est un emartingale. □

[15] C'est-à-dire que pour tout $n \ge 0$, $A_n \le A_{n+1}$ p.s.

Exemple 7.2. Une martingale (X_n) est dite de carré intégrable si $\mathbb{E}(X_n^2) < \infty$ pour tout n. Dans ce cas, (X_n^2) est une sous-martingale (Exemple 1.10). Soit $X_n^2 = M_n + A_n$ sa décomposition de Doob. La preuve du Théorème 7.1 nous dit que

$$A_n = \sum_{i=1}^n \Big\{ \mathbb{E}[X_i^2 \,|\, \mathscr{F}_{i-1}] - X_{i-1}^2 \Big\} = \sum_{i=1}^n \mathbb{E}[\, (X_i - X_{i-1})^2 \,|\, \mathscr{F}_{i-1}\,].$$

Dans la littérature, (A_n) est appelé processus croissant associé à la martingale (X_n). □

8. Deux exemples

On étudie deux familles de martingales un peu particulières. Dans la première, les accroissements des martingales sont bornées. On montre alors une dichotomie pour le comportement asymptotique. Comme application, on établit le lemme de Borel–Cantelli de Paul Lévy. La seconde famille consiste des martingales de type multiplicatif, pour lesquelles on montre la dichotomie de Kakutani.

8.1. Martingales à accroissements bornés

On considère dans ce paragraphe les martingales dont les accroissements sont bornés. On montre une dichotomie pour une telle martingale : p.s., soit elle converge vers une limite finie, soit elle oscille entre ∞ et $-\infty$.

Théorème 8.1. *Soit* $(X_n,\, n \ge 0)$ *une martingale. On suppose qu'il existe une constante* $C > 0$ *telle que* $\sup_{n\ge 0} |X_{n+1} - X_n| \le C$. *Posons*

$$E_{\rm cvg} := \{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) \text{ existe (et est finie)}\}\,, \tag{8.1}$$

$$E_{\rm osc} := \{\omega \in \Omega : \limsup_{n\to\infty} X_n(\omega) = \infty,\ \liminf_{n\to\infty} X_n(\omega) = -\infty\}\,. \tag{8.2}$$

Alors $\mathbb{P}(E_{\rm cvg} \cup E_{\rm osc}) = 1$.

Preuve. On peut supposer, sans perte de généralité, que $X_0 = 0$, car sinon on considèrera la martingale $(X_n - X_0,\, n \ge 0)$.

Soit $k \ge 1$ un entier. Définissons le temps d'arrêt

$$T_k := \inf\{n \ge 0 :\ X_n \le -k\}, \qquad \inf \varnothing := \infty.$$

On sait, d'après le Théorème 3.1, que $(X_{n\wedge T_k},\, n \geq 0)$ est une martingale à valeurs dans $[-k-C,\, \infty[$.[16] Donc $(X_{n\wedge T_k}+k+C,\, n\geq 0)$ est une martingale positive, qui converge donc p.s., d'après le Corollaire 4.5, vers une limite finie. Ceci implique que p.s. sur $\{T_k=\infty\}$, $\lim_{n\to\infty} X_n$ existe et est finie.

On a donc prouvé que p.s. sur $\cup_{k\geq 1}\{T_k=\infty\}$, $\lim_{n\to\infty} X_n$ existe et est finie. Observons que l'ensemble $\cup_{k\geq 1}\{T_k=\infty\}$ coïncide avec $\{\liminf_{n\to\infty} X_n > -\infty\}$.

En considérant la martingale $(-X_n)$ à la place de (X_n), on déduit également que p.s. sur $\{\limsup_{n\to\infty} X_n < \infty\}$, $\lim_{n\to\infty} X_n$ existe et est finie. Par conséquent,

$$\text{p.s. sur } \{\limsup_{n\to\infty} X_n < \infty\} \cup \{\liminf_{n\to\infty} X_n > -\infty\}, \qquad \lim_{n\to\infty} X_n \text{ existe et est finie,}$$

ce qui donne le résultat cherché. □

On utilise maintenant le Théorème 8.1 pour prouver la version Paul Lévy du lemme de Borel–Cantelli. Rappelons que si $(A_n,\, n\geq 1)$ est une suite d'événements, $\limsup_{n\to\infty} A_n$ est défini comme l'événement $\cap_{k=1}^\infty \cup_{n=k}^\infty A_n$, qui représente l'ensemble des $\omega\in\Omega$ qui appartiennent à une infinité de A_n. Autrement dit, $\limsup_{n\to\infty} A_n = \{\omega\in\Omega : \sum_{n=1}^\infty \mathbf{1}_{A_n}(\omega)=\infty\}$.

Théorème 8.2. (Lemme de Borel–Cantelli de Lévy). *Soit $(\mathscr{F}_n)_{n\geq 0}$ une filtration. Soit $(A_n,\, n\geq 1)$ une suite telle que pour tout $n\geq 1$, $A_n\in\mathscr{F}_n$. Alors*

$$\limsup_{n\to\infty} A_n = \Big\{\omega\in\Omega : \sum_{n=1}^\infty \mathbb{P}(A_n \,|\, \mathscr{F}_{n-1})(\omega)=\infty\Big\}, \qquad \text{p.s.}^{17}$$

Preuve. Posons $X_0 := 0$ et pour $n\geq 1$,

$$X_n := \sum_{i=1}^n \Big[\mathbf{1}_{A_i} - \mathbb{P}(A_i \,|\, \mathscr{F}_{i-1})\Big].$$

Pour tout n, X_n est intégrable (car bornée) et $\mathscr{F}_n$-mesurable, avec $\mathbb{E}(X_{n+1}-X_n \,|\, \mathscr{F}_n) = \mathbb{P}(A_{n+1}\,|\,\mathscr{F}_n) - \mathbb{P}(A_{n+1}\,|\,\mathscr{F}_n) = 0$. Donc $(X_n,\, n\geq 0)$ est une martingale, avec $|X_{n+1}-X_n|\leq 1$ pour tout $n\geq 0$.

Soient $E_{\rm cvg}$ et $E_{\rm osc}$ comme dans (8.1) et (8.2), respectivement. Par définition, sur $E_{\rm cvg}$, les ensembles $\limsup_{n\to\infty} A_n$ et $\{\sum_{n=1}^\infty \mathbb{P}(A_n\,|\,\mathscr{F}_{n-1})=\infty\}$ sont identiques, tandis que sur $E_{\rm osc}$, on a $\sum_{n=1}^\infty \mathbf{1}_{A_n}=\infty$ et $\sum_{n=1}^\infty \mathbb{P}(A_n\,|\,\mathscr{F}_{n-1})=\infty$.

Comme $E_{\rm cvg}\cup E_{\rm osc}=\Omega$ p.s. (Théorème 8.1), on obtient la conclusion désirée. □

---

[16] Il est possible que $X_{n\wedge T_k}$ descende en-dessous de $-k$ lorsque $T_k\leq n$; l'hypothèse de bornitude des accroissements nous permet cependant de dire que $X_{n\wedge T}\geq -k-C$, $\forall n\geq 0$.

[17] On dit que deux événements A et B sont p.s. identiques si $\mathbf{1}_A=\mathbf{1}_B$ p.s., ce qui équivaut à dire que $\mathbb{P}[(A\backslash B)\cup(B\backslash A)]=0$.

8.2. Martingales de type multiplicatif

Soient ξ_0, ξ_1, $\cdots$ des variables aléatoires positives et indépendantes, telles que $\mathbb{E}(\xi_i) = 1$, $\forall i \geq 0$. On pose, pour $n \geq 0$,

$$X_n := \prod_{i=0}^{n} \xi_i \, ,$$

et $\mathscr{F}_n := \sigma(\xi_i,\, 0 \leq i \leq n)$. On a vu, dans l'Exemple 1.5, que $(X_n,\, n \geq 0)$ est une martingale. On étudie maintenant des questions de convergence pour cette martingale.

La martingale $(X_n,\, n \geq 0)$ étant positive, il résulte du Théorème 4.1 que $X_n \to X_\infty$ p.s.

Posons $a_n := \mathbb{E}(\xi_n^{1/2})$, $n \geq 0$. Par l'inégalité de Jensen, $a_n \leq [\mathbb{E}(\xi_n)]^{1/2} = 1$. D'autre part, comme $\mathbb{E}(\xi_n) = 1$, on a $\mathbb{P}(\xi_n > 0) > 0$ (et $\xi_n \geq 0$ p.s.) et donc $a_n > 0$. Ainsi, $0 < a_n \leq 1$, $\forall n \geq 0$. On peut définir

$$Y_n := \prod_{i=0}^{n} \frac{\xi_i^{1/2}}{a_i} = \frac{X_n^{1/2}}{\prod_{i=0}^{n} a_i} \, , \qquad n \geq 0 \, .$$

Il s'agit de nouveau d'une martingale de type multiplicatif. En particulier, $Y_n \to Y_\infty$ p.s. Par définition, $\mathbb{E}(Y_n^2) = \mathbb{E}(\prod_{i=0}^n \frac{\xi_i}{a_i^2}) = [\prod_{i=0}^n a_i]^{-2} < \infty$, $\forall n \geq 0$.

Supposons d'abord que $\prod_{n=0}^\infty a_n > 0$. Dans ce cas, on a $\sup_{n\geq 0} \mathbb{E}(Y_n^2) \leq [\prod_{i=0}^\infty a_i]^{-2} < \infty$. D'après l'inégalité de Doob (Théorème 3.6), on trouve $\mathbb{E}(\sup_{n\geq 0} Y_n^2) \leq 4 \sup_{n\geq 0} \mathbb{E}(Y_n^2) < \infty$. Comme

$$|X_n| = Y_n^2 \prod_{i=0}^{n} a_i^2 \leq Y_n^2 \, ,$$

(en rappelant que $a_i \leq 1$, $\forall i$), on obtient $\mathbb{E}(\sup_{n\geq 0} |X_n|) < \infty$. En particulier, la martingale $(X_n,\, n \geq 0)$ est uniformément intégrable (voir l'Exemple 4.9). D'après le Théorème 4.19, $X_n \to X_\infty$ p.s. et dans L^1, et $\mathbb{E}(X_\infty) = \mathbb{E}(X_0) = 1$.

Supposons ensuite que $\prod_{n=0}^\infty a_n = 0$. On savait que $\frac{X_n^{1/2}}{\prod_{i=0}^n a_i} = Y_n$ converge p.s. vers une limite finie, et que $X_n \to X_\infty$ p.s. L'hypothèse $\prod_{n=0}^\infty a_n = 0$ conduit alors à : $X_\infty = 0$ p.s. Autrement dit, la martingale $(X_n,\, n \geq 0)$ n'est pas uniformément intégrable.

Enfin, on observe que $\prod_{n=0}^\infty a_n > 0$ si et seulement si $\sum_n (1 - a_n) < \infty$.

Résumons toutes ces discussions dans le résultat suivant :

Théorème 8.3. (Kakutani). *Les quatre assertions sont équivalentes :*

(i) $\sum_n (1 - a_n) < \infty$ *;*

(ii) $\mathbb{E}(X_\infty) = 1$ *;*

(iii) $X_n \to X_\infty$ *dans* L^1 *;*

(iv) $(X_n,\, n \geq 0)$ *est uniformément intégrable.*

9. Martingales indexées par les entiers négatifs*

**Le matériel de cette section ne fait pas partie du programme de l'examen.*

Soit $(\mathscr{F}_n,\, n \le 0)$ une suite de sous-tribus de $\mathscr{F}$ telle que $\mathscr{F}_n \subset \mathscr{F}_{n+1}$, et soit $(X_n,\, n \le 0)$ tel que $\forall\, n$, X_n soit $\mathscr{F}_n$-mesurable et intégrable. On suppose que $(X_n,\, n \le 0)$ est une sous-martingale au sens où $\mathbb{E}(X_{n+1} \,|\, \mathscr{F}_n) \ge X_n$, $\forall\, n = -1, -2, \cdots$[18]

Théorème 9.1. *Soit* $(X_n,\, n \le 0)$ *une sous-martingale telle que* $\inf_{n\le 0} \mathbb{E}(X_n) > -\infty$. *Alors* $X_n \to X_{-\infty}$ p.s. *et dans* L^1 *(quand* $n \to -\infty$*). De plus,* $X_{-\infty} \le \mathbb{E}(X_0 \,|\, \mathscr{F}_{-\infty})$ p.s.*, où* $\mathscr{F}_{-\infty} := \bigcap_{n\le 0} \mathscr{F}_n$.

Preuve. Soient a et b des réels, et soit M_n le nombre de montées le long de $[a, b]$ par X_n, $\cdots$, X_{-1}, X_0. Par (4.1), on a $\mathbb{E}(M_n) \le \frac{\mathbb{E}[(X_0-a)^+]}{b-a}$. En faisant $n \to -\infty$, et par convergence monotone, $M_n \to M_{-\infty}$ et $\mathbb{E}(M_{-\infty}) \le \frac{\mathbb{E}[(X_0-a)^+]}{b-a}$, donc a fortiori $M_{-\infty} < \infty$, p.s. D'après la Remarque 4.3, $X_{-\infty} := \lim_{n\to-\infty} X_n$ existe p.s. On constate que la convergence p.s. est automatique, sans l'hypothèse que $\inf_{n\le 0} \mathbb{E}(X_n) > -\infty$.

Montrons maintenant convergence dans L^1. Il suffit de vérifier l'intégrabilité uniforme. Puisque $(\mathbb{E}[X_0 \,|\, \mathscr{F}_n],\ n \le 0)$ est uniformément intégrable (Théorème 4.18), il suffit de vérifier que la sous-martingale $(X_n - \mathbb{E}[X_0 \,|\, \mathscr{F}_n],\ n \le 0)$ est uniformément intégrable. On peut donc supposer sans perte de généralité que $X_n \le 0$.

Quand $n \to -\infty$, $\mathbb{E}(X_n) \to A = \inf_{n\le 0} \mathbb{E}(X_n) \in]-\infty, 0]$. Soit $\varepsilon > 0$. Il existe N tel que $\mathbb{E}(X_{-N}) - A \le \varepsilon$, et a fortiori $\mathbb{E}(X_{-N}) - \mathbb{E}(X_n) \le \varepsilon$, $\forall\, n \le 0$. Soit $a > 0$. On a, pour $n \le -N$,

$$\begin{aligned}
\mathbb{E}[\,|X_n|\,\mathbf{1}_{\{|X_n|>a\}}\,] &= -\mathbb{E}[\,X_n\,\mathbf{1}_{\{X_n<-a\}}\,] \\
&= -\mathbb{E}(X_n) + \mathbb{E}[\,X_n\,\mathbf{1}_{\{X_n\ge -a\}}\,] \\
&\le -\mathbb{E}(X_n) + \mathbb{E}[\,X_{-N}\,\mathbf{1}_{\{X_n\ge -a\}}\,] \\
&= -\mathbb{E}(X_n) + \mathbb{E}(X_{-N}) - \mathbb{E}[\,X_{-N}\,\mathbf{1}_{\{X_n<-a\}}\,] \\
&\le \varepsilon + \mathbb{E}[\,|X_{-N}|\,\mathbf{1}_{\{|X_n|>a\}}\,].
\end{aligned}$$

Par l'inégalité de Markov, $\mathbb{P}(|X_n| > a) \le \frac{-\mathbb{E}(X_n)}{a} \le \frac{-A}{a} = \frac{|A|}{a}$. On peut donc choisir a suffisamment grand tel que $\mathbb{E}[\,|X_{-N}|\,\mathbf{1}_{\{|X_n|>a\}}\,] \le \varepsilon$. Alors

$$\sup_{n\le -N} \mathbb{E}[\,|X_n|\,\mathbf{1}_{\{|X_n|>a\}}\,] \le 2\varepsilon.$$

[18] Les martingales indexées par les entiers négatifs sont parfois appelées “martingales rétrogrades” ou “martingales inverses”.

D'autre part, il est clair que l'on peut choisir a grand tel que $\mathbb{E}[\,|X_n|\,\mathbf{1}_{\{|X_n|>a\}}\,] \le \varepsilon$ pour $n = 0, -1, \cdots, -N$. Par conséquent, X_n est uniformément intégrable.

Vérifions maintenant que $X_{-\infty} \le \mathbb{E}(X_0 \,|\, \mathscr{F}_{-\infty})$. Puisque $X_n \le \mathbb{E}(X_0 \,|\, \mathscr{F}_n)$, on a, pour tout $A \in \mathscr{F}_{-\infty}$ (A est donc un élément de $\mathscr{F}_n$),

$$\mathbb{E}[X_n \,\mathbf{1}_A] \le \mathbb{E}[X_0 \,\mathbf{1}_A].$$

Comme $X_n \to X_{-\infty}$ dans L^1, en faisant $n \to -\infty$, on obtient : $\mathbb{E}[X_{-\infty} \,\mathbf{1}_A] \le \mathbb{E}[X_0 \,\mathbf{1}_A]$. Puisque $X_{-\infty}$ est $\mathscr{F}_n$-mesurable (pour tout n) donc $(\mathscr{F}_{-\infty})$-mesurable, ceci implique que $X_{-\infty} \le \mathbb{E}(X_0 \,|\, \mathscr{F}_{-\infty})$, p.s. □

Le résultat suivant est l'analogue du Théorème 4.21 pour les tribus indexées par $\mathbb{N}_-$.

Théorème 9.2. (Lévy). *Soit Y une variable aléatoire intégrable. Alors quand $n \to -\infty$,*

$$\mathbb{E}(Y \,|\, \mathscr{F}_n) \to \mathbb{E}(Y \,|\, \mathscr{F}_{-\infty}), \qquad \text{p.s. } \textit{et dans } L^1.$$

Preuve. Soit $X_n := \mathbb{E}(Y \,|\, \mathscr{F}_n)$, qui est une martingale indexée par $\mathbb{N}_-$. D'après le Théorème 9.1, $X_n \to X_{-\infty}$ p.s. et dans L^1, où

$$X_{-\infty} = \mathbb{E}[X_0 \,|\, \mathscr{F}_{-\infty}] = \mathbb{E}[\mathbb{E}(Y \,|\, \mathscr{F}_0) \,|\, \mathscr{F}_{-\infty}] = \mathbb{E}[Y \,|\, \mathscr{F}_{-\infty}], \qquad \text{p.s.}$$

C'est ce qu'il fallait démontrer. □

Exemple 9.3. (Application à la loi des grands nombres forte). Soit $(\xi_i,\, i \ge 1)$ une suite de variables aléatoires réelles i.i.d. intégrables. On pose $S_0 := 0$ et pour $n \ge 1$,

$$S_n := \sum_{i=1}^n \xi_i\,.$$

On a vu, dans l'Exemple 5.1 du Chapitre 1, que pour tout $n \ge 1$, $\mathbb{E}(\xi_1 \,|\, S_n) = \frac{S_n}{n}$ p.s. Comme $(S_n,\, \xi_1)$ est indépendante de $\sigma(\xi_i,\, i > n)$, on a, d'après l'Exemple 4.9 du Chapitre 1, $\mathbb{E}(\xi_1 \,|\, S_n,\, \xi_i,\, i > n) = \frac{S_n}{n}$ p.s.

Pour tout entier $n \le -1$, soit $\mathscr{F}_n := \sigma(S_{-n},\, \xi_i,\, i > -n)$. Alors $\mathscr{F}_n \subset \mathscr{F}_{n+1}$, $\forall n \le -2$. Le Théorème 9.2 nous dit que $\mathbb{E}(\xi_1 \,|\, S_n,\, \xi_i,\, i > n)$ converge p.s. et dans L^1 lorsque $n \to \infty$, c'est-à-dire que $\frac{S_n}{n}$ converge p.s. et dans L^1 lorsque $n \to \infty$, dont la limite est une constante d'après la loi 0–1 de Kolmogorov. Comme convergence dans L^1 implique convergence des espérances, la limite constante en question vaut $\lim_{n\to\infty} \mathbb{E}(\frac{S_n}{n})$, qui n'est autre que $\mathbb{E}(\xi_1)$.

Conclusion : $\frac{1}{n} \sum_{i=1}^n \xi_i \to \mathbb{E}(\xi_1)$ p.s. et dans L^1. □

Chapitre 3

Chaînes de Markov

Dans ce chapitre, on étudie les chaînes de Markov homogènes à valeurs dans un espace d'état dénombrable.

1. Définition et propriétés élémentaires

Soit E un espace dénombrable[1] non vide, muni de la tribu $\mathscr{P}(E)$ de toutes les parties de E (y compris $\varnothing$ et E).

Définition 1.1. *Une fonction $Q : E \times E \to \mathbb{R}$, que l'on écrira souvent sous la forme $Q := (p(x, y),\ x, y \in E)$, est une* **matrice de transition** *(= noyau de transition = probabilité de transition) si*

(i) $0 \le p(x, y) \le 1$, $\forall x, y \in E$;

(ii) $\sum_{y\in E} p(x, y) = 1$, $\forall x \in E$.

Pour tout couple de matrices de transitions Q et $\widetilde{Q}$ sur E, on définit la nouvelle matrice de transition $Q\widetilde{Q}$ par

$$(Q\widetilde{Q})(x, y) := \sum_{z\in E} p(x, z)\, \widetilde{Q}(z, y).$$

Soit I la matrice de transition identité, définie par $I(x, y) = 1$ si $x = y$ et $I(x, y) = 0$ sinon. On pose $Q^0 := I$, $Q^1 := Q$, et par récurrence, $Q^{n+1} := Q^n\, Q$.

Définition 1.2. *Soit Q une matrice de transition sur E. Soit $(X_n, n \ge 0)$ une suite de variables aléatoires à valeurs dans E. On dit que $(X_n, n \ge 0)$ est une chaîne de Markov*

[1] C'est-à-dire soit fini, soit infini dénombrable.

avec matrice de transition Q si, pour tout entier $n \geq 0$ et tout $y \in E$,

$$\mathbb{P}(X_{n+1} = y \,|\, X_0,\, X_1,\, \cdots,\, X_n) = p(X_n,\, y), \qquad \text{p.s.}^2$$

En d'autres mots,[3]

$$\mathbb{P}(X_{n+1} = y \,|\, X_0 = x_0,\, X_1 = x_1,\, \cdots,\, X_n = x_n) = p(x_n,\, y),$$

pour tous $x_0,\, x_1,\, \cdots,\, x_n,\, y \in E$ tels que $\mathbb{P}(X_0 = x_0,\, X_1 = x_1,\, \cdots,\, X_n = x_n) > 0$.

Remarque 1.3. (i) En général, $\mathbb{P}(X_{n+1} = y \,|\, X_0,\, X_1,\, \cdots,\, X_n)$ est une fonction mesurable de $X_0,\, X_1,\, \cdots,\, X_n$, et non seulement de X_n. Le fait qu'elle ne dépend que de X_n porte le nom de **propriété de Markov** : pour prédire le "futur" X_{n+1}, la connaissance des "passé et présent" $(X_0,\, X_1,\, \cdots,\, X_n)$ ne donne pas plus d'information que celle du "présent" X_n. On verra d'autres formulations de la propriété de Markov, avec la même idée intuitive.

(ii) La probabilité conditionnelle $\mathbb{P}(X_{n+1} = y \,|\, X_0,\, X_1,\, \cdots,\, X_n)$ ne dépend pas de n : on dit que la chaîne de Markov est homogène. On peut également s'intéresser aux chaînes de Markov inhomogènes, dont la transition de n à $n+1$ dépend de n. □

Proposition 1.4. *Un processus aléatoire $(X_n,\, n \geq 0)$ à valeurs dans E est une chaîne de Markov avec matrice de transition Q si et seulement si pour tout entier $n \geq 0$ et tous x_0, $x_1,\, \cdots,\, x_n \in E$,*

$$\begin{aligned} & \mathbb{P}(X_0 = x_0,\, X_1 = x_1,\, \cdots, X_n = x_n) \\ = \ & \mathbb{P}(X_0 = x_0)\, p(x_0,\, x_1)\, p(x_1,\, x_2) \cdots p(x_{n-1},\, x_n). \end{aligned} \tag{1.1}$$

En particulier, si $\mathbb{P}(X_0 = x_0) > 0$, alors

$$\mathbb{P}(X_n = x_n \,|\, X_0 = x_0) = [Q^n](x_0,\, x_n).$$

Preuve. "$\Rightarrow$" Immédiate, vu la relation de récurrence :

$$\begin{aligned} & \mathbb{P}(X_0 = x_0,\, X_1 = x_1,\, \cdots, X_n = x_n,\, X_{n+1} = x_{n+1}) \\ = \ & \mathbb{P}(X_0 = x_0,\, \cdots, X_n = x_n)\, \mathbb{P}(X_{n+1} = x_{n+1} \,|\, X_0 = x_0,\, \cdots, X_n = x_n). \end{aligned}$$

[2] Ainsi, $\mathbb{P}(X_{n+1} = y \,|\, X_n) = p(X_n,\, y)$, p.s.

[3] Si Y est une variable aléatoire réelle intégrable, et Z une variable aléatoire à valeurs dans un espace dénombrable E, alors dire que $\mathbb{E}(Y \,|\, Z) = h(Z)$ p.s. équivaut à dire que $\mathbb{E}(Y \,|\, Z = z) = h(z)$ pour tout $z \in E$ tel que $\mathbb{P}(Z = z) > 0$. Voir la Remarque 7.2 du Chapitre 1.

"$\Leftarrow$" On a, pour $y \in E$,

$$\begin{aligned}
& \mathbb{P}(X_{n+1} = y \,|\, X_0 = x_0,\, X_1 = x_1,\, \cdots, X_n = x_n) \\
= \; & \frac{\mathbb{P}(X_0 = x_0,\, X_1 = x_1,\, \cdots, X_n = x_n,\, X_{n+1} = y)}{\mathbb{P}(X_0 = x_0,\, X_1 = x_1,\, \cdots, X_n = x_n)} \\
= \; & \frac{\mathbb{P}(X_0 = x_0)\, p(x_0,\, x_1)\, p(x_1,\, x_2)\, \cdots\, p(x_{n-1},\, x_n)\, p(x_n,\, y)}{\mathbb{P}(X_0 = x_0)\, p(x_0,\, x_1)\, p(x_1,\, x_2)\, \cdots\, p(x_{n-1},\, x_n)} \qquad \text{(par hypothèse)} \\
= \; & p(x_n,\, y),
\end{aligned}$$

ce qui donne la propriété de Markov cherchée.

La dernière assertion est claire, car

$$\begin{aligned}
[Q^n](x_0,\, x_n) &= \sum_{x_1,\, x_2,\, \cdots,\, x_{n-1} \in E} p(x_0,\, x_1)\, p(x_1,\, x_2)\, \cdots\, p(x_{n-1},\, x_n) \qquad \text{(définition de } Q^n\text{)} \\
&= \sum_{x_1,\, x_2,\, \cdots,\, x_{n-1} \in E} \frac{\mathbb{P}(X_0 = x_0,\, X_1 = x_1,\, \cdots, X_n = x_n)}{\mathbb{P}(X_0 = x_0)} \qquad \text{(par (1.1))} \\
&= \frac{\mathbb{P}(X_0 = x_0,\, X_n = x_n)}{\mathbb{P}(X_0 = x_0)} \\
&= \mathbb{P}(X_0 = x_0 \,|\, X_n = x_n),
\end{aligned}$$

comme promis. □

Remarque 1.5. La formule (1.1) montre que pour une chaîne de Markov $(X_n,\, n \geq 0)$, la loi de $(X_0,\, X_1,\, \cdots, X_n)$ est déterminée par la *loi initiale* (i.e., la loi de X_0) et la matrice de transition Q. □

Proposition 1.6. *Soit $(X_n,\, n \geq 0)$ une chaîne de Markov de matrice de transition Q. Soient $n \geq 0$ et $m \geq 1$ des entiers. Pour tous $y_1,\, \cdots,\, y_m \in E$,*

$$\mathbb{P}(X_{n+1} = y_1,\, \cdots,\, X_{n+m} = y_m \,|\, X_0,\, \cdots,\, X_n) = p(X_n,\, y_1)\, p(y_1,\, y_2)\, \cdots\, p(y_{m-1},\, y_m),$$

et donc

$$\mathbb{P}(X_{n+m} = y_m \,|\, X_n) = [Q^m](X_n,\, y_m).$$

Si l'on écrit $Y_i := X_{n+i}$, alors $(Y_i,\, i \geq 0)$ est une chaîne de Markov de matrice de transition Q.

Preuve. D'après (1.1), on a

$$\begin{aligned}
& \mathbb{P}(X_{n+1} = y_1,\, \cdots\, X_{n+m} = y_m \,|\, X_0 = x_0,\, \cdots,\, X_n = x_n) \\
= \; & p(x_n,\, y_1)\, p(y_1,\, y_2) \cdots p(y_{m-1},\, y_m),
\end{aligned}$$

ce qui donne la première identité. Le résultat cherché pour $\mathbb{P}(X_{n+m} = y_m \,|\, X_n)$ s'obtient en sommant sur tous les choix possibles de $y_1, \cdots, y_{m-1}$. Enfin, pour vérifier que $(Y_i, i \geq 0)$ est une chaîne de Markov de matrice de transition Q, on observe que

$$\mathbb{P}(Y_0 = y_0, Y_1 = y_1, \cdots, Y_m = y_m) = \mathbb{P}(X_n = y_0)\, p(y_0, y_1) \cdots p(y_{m-1}, y_m),$$

et conclut à l'aide de la caractérisation des chaînes de Markov obtenue en Proposition 1.4.□

2. Exemples de chaînes de Markov

On se donne quelques exemples de chaînes de Markov.

Exemple 2.1. (Variables aléatoires indépendantes). Soient $X_0, X_1, X_2, \cdots$ des variables aléatoires indépendantes à valeurs dans E, ayant la même loi μ. On vérifie assez facilement que $(X_n, n \geq 0)$ est une chaîne de Markov de matrice de transition

$$p(x, y) := \mu(\{y\}), \qquad \forall x, y \in E.$$

[On écrira très souvent $\mu(y)$ au lieu de $\mu(\{y\})$.] Ce n'est évidemment pas l'exemple de chaîne de Markov le plus intéressant ! □

Exemple 2.2. (Marches aléatoires sur $\mathbb{Z}^d$). Soit $E := \mathbb{Z}^d$, et soient $X_0, \xi_1, \xi_2, \cdots$ des variables aléatoires indépendantes à valeurs dans $\mathbb{Z}^d$, telles que ξ_n, pour tout $n \geq 1$, ait la loi μ. Posons

$$X_n := X_0 + \sum_{i=1}^n \xi_i\,, \qquad n \geq 1.$$

Alors $(X_n, n \geq 0)$ est une chaîne de Markov de matrice de transition

$$p(x, y) := \mu(y - x), \qquad \forall x, y \in \mathbb{R}^d.$$

En effet, puisque ξ_{n+1} est indépendante de $(X_0, X_1, \cdots X_n)$, on a

$$\begin{aligned}
& \mathbb{P}(X_{n+1} = y \,|\, X_0 = x_0, X_1 = x_1, \cdots, X_n = x_n) \\
= \; & \mathbb{P}(\xi_{n+1} = y - x_n \,|\, X_0 = x_0, X_1 = x_1, \cdots, X_n = x_n) \\
= \; & \mathbb{P}(\xi_{n+1} = y - x_n) \\
= \; & \mu(y - x_n),
\end{aligned}$$

comme prévu.

Cas spécial. Soient $e_1, \cdots, e_d$ des vecteurs unité de $\mathbb{R}^d$. Dans le cas $\mu(e_i) = \mu(-e_i) = \frac{1}{2d}$, pour $1 \leq i \leq d$, la chaîne de Markov est appelée **marche aléatoire simple** sur $\mathbb{Z}^d$. □

Exemple 2.3. (Marche aléatoire simple sur un graphe). Soit A une famille de parties de E à deux éléments. Pour tout $x \in E$, on définit

$$A_x := \{y \in E : \{x, y\} \in A\}.$$

On suppose que $0 < \#A_x < \infty$ pour tout $x \in E$.

On définit maintenant une matrice de transition Q sur E par

$$p(x, y) := \begin{cases} \frac{1}{\#A_x} & \text{si } \{x, y\} \in A, \\ 0 & \text{sinon.} \end{cases}$$

Une chaîne de Markov de matrice de transition Q est appelée une **marche aléatoire simple sur le graphe** (E, A). □

Exemple 2.4. (Processus de branchement). Dans cet exemple, $E = \mathbb{N} := \{0, 1, 2, \cdots\}$ est l'ensemble des nombres naturels. Soit μ une mesure de probabilité sur $\mathbb{N}$. On définit par récurrence une suite de variables aléatoires $(X_n, n \geq 0)$ à valeurs dans $\mathbb{N}$ de la façon suivante :

$$X_{n+1} := \sum_{i=1}^{X_n} \xi_{n,i}, \quad n \in \mathbb{N}, \qquad (\text{notation : } \textstyle\sum_{i=1}^{0} := 0)$$

où $\xi_{n,i}$, $n \geq 0$, $i \geq 1$, sont des variables aléatoires i.i.d. ayant la même loi μ, et qui sont indépendantes de X_0. Alors $(X_n, n \geq 0)$ est une chaîne de Markov sur $\mathbb{N}$ de matrice de transition

$$p(x, y) := \mu^{*x}(y), \qquad \forall x, y \in \mathbb{N},$$

où μ^{*x} désigne[4] la convolution d'ordre x de μ, ou en langage probabiliste, la loi de la somme de x variables aléatoires i.i.d. de loi μ.

En effet, $(\xi_{n,i}, i \geq 1)$ étant indépendante de $\sigma(X_0, X_1, \cdots, X_n)$, on a

$$\begin{aligned} & \mathbb{P}(X_{n+1} = y \,|\, X_0 = x_0, X_1 = x_1, \cdots, X_n = x_n) \\ = \;& \mathbb{P}\Big(\sum_{i=1}^{x_n} \xi_{n,i} = y \,\Big|\, X_0 = x_0, X_1 = x_1, \cdots, X_n = x_n \Big) \\ = \;& \mathbb{P}\Big(\sum_{i=1}^{x_n} \xi_{n,i} = y \Big), \end{aligned}$$

qui n'est autre que $\mu^{*x_n}(y)$. □

[4] Pour $x = 0$, μ^{*0} est la mesure de Dirac au point 0.

Exemple 2.5. (Chaîne d'Ehrenfest). Dans cet exemple, $E := \{0, 1, \cdots, r\}$, où $r \ge 1$ est une entier fixé, et

$$p(i, j) := \begin{cases} \frac{r-i}{r} & \text{si } j = i+1 \\ \frac{i}{r} & \text{si } j = i-1 \\ 0 & \text{sinon.} \end{cases}$$

En mots, il y a au total r boules dans deux urnes ; on tire une boule par hasard parmi les r, et remet la boule dans l'autre urne. $\square$

3. Chaîne de Markov canonique

On démarre cette section avec un résultat sur l'existence des chaînes de Markov avec une matrice de transition donnée.

Proposition 3.1. *Soit Q une matrice de transition sur E. Il existe un espace de probabilité $(\widetilde{\Omega}, \widetilde{\mathscr{F}}, \widetilde{\mathbb{P}})$, sur lequel on peut définir, pour tout $x \in E$, un processus $(X_n^x, n \ge 0)$ qui est une chaîne de Markov de matrice de transition Q, issue de $X_0^x = x$.*

Preuve. Soit $\widetilde{\Omega} := [0, 1[$, muni de la tribu borélienne et de (la restriction sur $[0, 1[$ de) la mesure de Lebesgue. Pour tout $\omega \in \widetilde{\Omega} = [0, 1[$, on écrit le développement dyadique (propre)

$$\omega = \sum_{n=0}^{\infty} \frac{\varepsilon_n(\omega)}{2^{n+1}}, \qquad \varepsilon_n(\omega) \in \{0, 1\}.$$

On vérifie assez facilement que $(\varepsilon_n, n \ge 0)$ est une suite i.i.d. de variables aléatoires telle que $\widetilde{\mathbb{P}}(\varepsilon_n = 0) = \frac{1}{2} = \widetilde{\mathbb{P}}(\varepsilon_n = 1)$.

Soit $\varphi : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ une injection. Alors $(\eta_{i,j} := \varepsilon_{\varphi(i,j)}, i, j \in \mathbb{N})$ est de nouveau une famille i.i.d. de variables aléatoires. Posons

$$U_i := \sum_{j=0}^{\infty} \frac{\eta_{i,j}}{2^{j+1}}, \qquad i \ge 0.$$

Il est clair[5] que U_i, $i \ge 0$, sont des variables aléatoires i.i.d. de loi uniforme sur $[0, 1[$.

On écrit $E = \{y_k, k \ge 1\}$. Fixons $x \in E$. On définit $X_0^x := x$ et

$$X_1^x := y_k, \qquad \text{si } \sum_{j=1}^{k-1} p(x, y_j) \le U_1 < \sum_{j=1}^{k} p(x, y_j),$$

[5]Pour voir que U_i est uniformément distribuée, il suffit de remarquer que pour tout entier $m \ge 1$, $\sum_{j=0}^{m} \frac{\eta_{i,j}}{2^{j+1}}$ a la même loi que $\sum_{n=0}^{m} \frac{\varepsilon_n}{2^{n+1}}$, et de laisser ensuite m tendre vers l'infini.

ce qui donne $\widetilde{\mathbb{P}}(X_1^x = y) = p(x, y)$ pour tout $y \in E$. On procède ensuite par récurrence :

$$X_{n+1}^x := y_k \qquad \text{si } \sum_{j=1}^{k-1} p(X_n^x, y_j) \le U_{n+1} < \sum_{j=1}^{k} p(X_n^x, y_j).$$

Puisque U_i, $i \ge 1$, sont indépendantes, on a, pour tout $k \ge 1$,

$$\begin{aligned}
& \widetilde{\mathbb{P}}(X_{n+1}^x = y_k \,|\, X_0^x = x_0, X_1^x = x_1, \cdots, X_n^x = x_n) \\
= \ & \widetilde{\mathbb{P}}\Big(\sum_{j=1}^{k-1} p(x_n, y_j) \le U_{n+1} < \sum_{j=1}^{k} p(x_n, y_j) \,|\, X_0^x = x_0, X_1^x = x_1, \cdots, X_n^x = x_n \Big) \\
& \qquad \text{(définition de } X_{n+1}^x\text{)} \\
= \ & \widetilde{\mathbb{P}}\Big(\sum_{j=1}^{k-1} p(x_n, y_j) \le U_{n+1} < \sum_{j=1}^{k} p(x_n, y_j) \Big), \qquad (U_{n+1} \text{ indépendante de } (X_0^x, X_1^x, \cdots, X_n^x))
\end{aligned}$$

qui vaut $p(x_n, y_k)$ car U_{n+1} est uniformément distribuée. Ainsi, $(X_n^x, n \ge 0)$ est une chaîne de Markov de matrice de transition Q. $\square$

Dans le reste du chapitre, il conviendra de choisir un espace de probabilité canonique sur lequel la chaîne de Markov est définie. On choisira

$$\Omega := E^{\mathbb{N}}.$$

Un élément $\omega \in \Omega$ est alors une suite $\omega := (\omega_0, \omega_1, \omega_2, \cdots)$ d'éléments de E. Introduisons les projections X_n, $n \in \mathbb{N}$, par

$$X_n : \Omega \to E, \qquad \text{avec } X_n(\omega) := \omega_n, \qquad \forall \omega \in \Omega := E^{\mathbb{N}}.$$

On munit Ω de la plus petite tribu, notée $\mathscr{F}$, qui rende ces projections mesurables. [Autrement dit, $\mathscr{F}$ est la tribu engendrée par les projections.] Alors $\mathscr{F}$ est la tribu engendrée par les cylindres, c'est-à-dire par les ensembles de la forme $\{\omega \in \Omega : \omega_0 = x_0, \omega_1 = x_1, \cdots, \omega_n = x_n\}$, pour $n \in \mathbb{N}$ et $x_0, x_1, \cdots, x_n \in E$.

Lemme 3.2. *Soit $(\widetilde{E}, \widetilde{\mathscr{E}})$ un espace mesurable. Soit $\Phi : \widetilde{E} \to \Omega$ une application. Alors Φ est mesurable si et seulement si pour tout $n \in \mathbb{N}$, $X_n \circ \Phi : \widetilde{E} \to E$ est mesurable.*

Preuve. "$\Rightarrow$" La mesurabilité de $X_n \circ \Phi$, qui est composition de deux applications mesurables, est immédiate.

"$\Leftarrow$" Considérons $\mathscr{G} := \{A \in \mathscr{F} : \Phi^{-1}(A) \in \widetilde{\mathscr{E}}\}$, qui est une tribu contenant, par hypothèse, tous les ensembles de la forme $X_n^{-1}(\{x\})$ (pour $n \in \mathbb{N}$ et $x \in E$) : elle rend les projections X_n mesurables, et inclut donc $\mathscr{F}$. Autrement dit, $\mathscr{G} = \mathscr{F}$. $\square$

Théorème 3.3. (Chaîne de Markov canonique). *Soit Q une matrice de transition sur E. Pour tout $x \in E$, il existe une unique mesure de probabilité sur $\Omega := E^{\mathbb{N}}$, notée $\mathbb{P}_x$, telle que le processus de projections $(X_n,\, n \in \mathbb{N})$ soit une chaîne de Markov de matrice de transition Q et que $\mathbb{P}_x(X_0 = x) = 1$.*

Preuve. (Unicité) Supposons que $\widehat{\mathbb{P}}_x$ est une mesure de probabilité sur $\Omega := E^{\mathbb{N}}$ qui satisfait les propriétés du théorème. La famille $\mathscr{M} := \{A \in \mathscr{F} : \mathbb{P}_x(A) = \widehat{\mathbb{P}}_x(A)\}$ est de toute évidence une classe monotone. Soit $\mathscr{C}$ la famille des cylindres, qui est stable par intersections finies. Par hypothèse, $\mathscr{C} \subset \mathscr{M}$, et donc par classe monotone, $\sigma(\mathscr{C}) \subset \mathscr{M}$. Comme $\sigma(\mathscr{C}) = \mathscr{F}$, on a $\mathbb{P}_x = \widehat{\mathbb{P}}_x$.

(Existence) Fixons $x \in E$. D'après la Proposition 3.1, il existe un espace de probabilité $(\widetilde{\Omega}, \widetilde{\mathscr{F}}, \widetilde{\mathbb{P}})$ sur lequel on peut trouver une chaîne de Markov $(X_n^x,\, n \geq 0)$ de matrice de transition Q, issue de $X_0^x = x$. Soit $\mathbb{P}_x$ la mesure image de $\widetilde{\mathbb{P}}$ par l'application mesurable

$$\begin{array}{rcl} \widetilde{\Omega} & \to & \Omega := E^{\mathbb{N}} \\ \widetilde{\omega} & \mapsto & (X_n^x(\widetilde{\omega}),\, n \in \mathbb{N}) . \end{array}$$

Le mesurabilité de l'application est une conséquence du Lemme 3.2. On a $\mathbb{P}_x(X_0 = x) = \widetilde{\mathbb{P}}(X_0^x = x) = 1$, et pour tout $n \in \mathbb{N}$ et tous $x_0, x_1, \cdots, x_n \in E$,

$$\begin{aligned} \mathbb{P}_x(X_0 = x_0,\, X_1 = x_1,\, \cdots,\, X_n = x_n) &= \widetilde{\mathbb{P}}(X_0^x = x_0,\, X_1^x = x_1,\, \cdots,\, X_n^x = x_n) \\ &= \widetilde{\mathbb{P}}(X_0^x = x_0)\, p(x_0, x_1) \cdots p(x_{n-1}, x_n) \\ &= \mathbb{P}_x(X_0 = x_0)\, p(x_0, x_1) \cdots p(x_{n-1}, x_n). \end{aligned}$$

Par la Proposition 1.4, ceci signifie que sous $\mathbb{P}_x$, $(X_n,\, n \in \mathbb{N})$ est une chaîne de Markov de matrice de transition Q. □

Remarque 3.4. (a) La dernière assertion de la Proposition 1.4 nous dit que pour tous $n \geq 0$ et $x, y \in E$,

$$\mathbb{P}_x(X_n = y) = [Q^n](x, y).$$

(b) Si μ est une mesure de probabilité sur E, on pose[6]

$$\mathbb{P}_\mu(A) := \int_E \mathbb{P}_x(A)\, \mu(\mathrm{d}x) , \qquad \forall A \in \mathscr{F},$$

[6]Puisque l'espace E est dénombrable, l'intégale est, en fat, une somme.

qui est une mesure de probabilité sur Ω. En écrivant explicitement $\mathbb{P}_\mu(X_0 = x_0, \cdots, X_n = x_n)$, on se rend compte que sous $\mathbb{P}_\mu$, $(X_n, n \geq 0)$ est une chaîne de Markov de matrice de transition Q, et de loi initiale (c'est-à-dire, loi de X_0) μ. On observe également que $\mathbb{P}_x = \mathbb{P}_{\delta_x}$, où δ_x est la mesure de Dirac au point x.

(c) Si $(\widehat{X}_n, n \geq 0)$ est une chaîne de Markov définie sur l'espace de probabilité $(\widehat{\Omega}, \widehat{\mathscr{F}}, \widehat{\mathbb{P}})$, de matrice de transition Q et de loi initiale μ, alors pour toute partie $A \subset \Omega := E^{\mathbb{N}}$ mesurable,

$$\widehat{\mathbb{P}}\{(\widehat{X}_n)_{n\geq 0} \in A\} = \mathbb{P}_\mu(A).$$

En effet, l'identité est triviale si A est un cylindre, et reste valable pour toute partie mesurable A par classe monotone (à l'aide d'un argument qui est similaire à celui que l'on a employé dans la preuve du Théorème 3.3).

Cette identité montre que les résultats que l'on établira dans la suite pour la chaîne de Markov canonique pourront être reformulés pour toute chaîne de Markov ayant la même matrice de transition. □

Un avantage important de travailler avec la chaîne de Markov canonique est l'utilisation des opérateurs de translation, que l'on appelle en bon français **opérateurs shift**. Pour tout $k \in \mathbb{N}$, on définit l'opérateur $\theta_k : \Omega \to \Omega$ par

$$\theta_k((\omega_n)_{n\in\mathbb{N}}) = (\omega_{k+n})_{n\in\mathbb{N}}.$$

Le Lemme 3.2 nous dit que tous ces opérateurs sont des applications mesurables.

Posons $\mathscr{F}_n := \sigma(X_0, X_1, \cdots, X_n)$; on appelle $(\mathscr{F}_n)_{n\in\mathbb{N}}$ la filtration canonique. On note également $\mathbb{E}_x$ pour espérance associée à la probabilité $\mathbb{P}_x$.

Théorème 3.5. (Propriété de Markov) *Soit $Y : \Omega \to \mathbb{R}_+$ une application mesurable. Soit $n \geq 0$ un entier positif. Pour tout $x \in E$, on a*

$$\mathbb{E}_x[Y \circ \theta_n \,|\, \mathscr{F}_n] = \mathbb{E}_{X_n}(Y), \qquad \mathbb{P}_x\text{-p.s.}$$

Remarque 3.6. Le théorème reste valable si l'on remplace $\mathbb{E}_x$ par $\mathbb{E}_\mu$, pour toute loi initiale μ. La même remarque s'appliquera, plus tard, à la propriété de Markov forte.

Preuve. Il s'agit de prouver $\mathbb{E}_x[(Y \circ \theta_n)\, \mathbf{1}_B] = \mathbb{E}_x[\mathbf{1}_B\, \mathbb{E}_{X_n}(Y)]$ pour tout $B \in \mathscr{F}_n$.

Par un argument usuel (des fonctions indicatrices aux fontions étagées, puis aux fonctions mesurables positives grâce au théorème de convergence monotone), il suffit de prouver l'identité pour $Y = \mathbf{1}_A$, avec $A \in \mathscr{F}$.

Par classe monotone, il suffit de prouver l'identité pour les ensembles A de la forme (pour $m \in \mathbb{N}$ et $x_0, x_1, \cdots, x_m \in E$)

$$\begin{aligned} A &= \{x_0\} \times \{x_1\} \times \cdots \times \{x_m\} \times E \times E \times \cdots \\ &= \{\omega \in \Omega : X_0(\omega) = x_0, X_1(\omega) = x_1, \cdots, X_m(\omega) = x_m\}, \end{aligned}$$

car la famille des ensembles A de cette forme est stable par intersections finies et engendre $\mathscr{F}$, tandis que la famille de tous les ensembles A tels que $Y := \mathbf{1}_A$ satisfasse l'identité désirée est une classe monotone.

Encore par classe monotone, il suffit de prouver l'identité pour le cas spécial

$$B := \{\omega \in \Omega : X_0(\omega) = y_0, X_1(\omega) = y_1, \cdots, X_n(\omega) = y_n\},$$

avec $y_0, y_1, \cdots, y_n \in E$.

Soit $x \in E$. Pour notre choix spécial de A et B, on a

$$\begin{aligned} &\mathbb{E}_x[(Y \circ \theta_n)\, \mathbf{1}_B] \\ = \;& \mathbb{P}_x\{X_0 = y_0, X_1 = y_1, \cdots, X_n = y_n, X_n = x_0, X_{n+1} = x_1, \cdots, X_{n+m} = x_m\} \\ = \;& \mathbf{1}_{\{y_0 = x\}}\, p(y_0, y_1) \cdots p(y_{n-1}, y_n)\, \mathbf{1}_{\{x_0 = y_n\}}\, p(x_0, x_1) \cdots p(x_{m-1}, x_m). \end{aligned}$$

D'autre part, pour tout $y \in E$,

$$\mathbb{E}_y(Y) = \mathbb{P}_y\{X_0 = x_0, X_1 = x_1, \cdots, X_m = x_m\} = \mathbf{1}_{\{y = x_0\}}\, p(x_0, x_1) \cdots p(x_{m-1}, x_m),$$

ce qui implique que

$$\begin{aligned} &\mathbb{E}_x[\mathbf{1}_B\, \mathbb{E}_{X_n}(Y)] \\ = \;& \mathbb{P}_x\{X_0 = y_0, X_1 = y_1, \cdots, X_n = y_n;\ X_n = x_0\}\, p(x_0, x_1) \cdots p(x_{m-1}, x_m) \\ = \;& \mathbf{1}_{\{y_0 = x\}}\, p(y_0, y_1) \cdots p(y_{n-1}, y_n)\, \mathbf{1}_{\{x_0 = y_n\}}\, p(x_0, x_1) \cdots p(x_{m-1}, x_m). \end{aligned}$$

Par conséquent, $\mathbb{E}_x[(Y \circ \theta_n)\, \mathbf{1}_B] = \mathbb{E}_x[\mathbf{1}_B\, \mathbb{E}_{X_n}(Y)]$. □

La propriété de Markov énoncée dans le Theorem 3.5 donne un énoncé rigoureux de ce que l'on disait auparavant : pour prédire le "futur" θ_n, la connaissance des "passé et présent" $(X_0, X_1, \cdots, X_n)$ ne donne pas plus d'information que celle du "présent" X_n. Il sera très important de pouvoir étendre cette propriété à des cas où n est remplacé par un temps aléatoire T.

Pour illustrer l'intérêt d'une telle extension, on considère le problème de savoir si, issue d'un point x, la chaîne de Markov revient à x infiniment souvent. Posons

$$N_x := \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n = x\}},$$

qui représente le nombre total des visites au site x par la chaîne. Il suffira alors de vérifier si la chaîne de Markov revient au site x au moins une fois : soit

$$T_x := \inf\{n \geq 1 : \, X_n = x\}, \qquad (\inf \varnothing := \infty)$$

on a

$$\mathbb{P}_x\{N_x = \infty\} = 1 \;\Leftrightarrow\; \mathbb{P}_x\{T_x < \infty\} = 1.$$

En effet, le passage "$\Rightarrow$" est trivial. Quant au passage inverse "$\Leftarrow$", supposons que $\mathbb{P}_x\{T_x < \infty\} = 1$. Si la propriété de Markov est étendue dans le sens évoqué ci-dessus, on verra dans le Corollaire 3.9 que cela impliquera que la loi de θ_{T_x} est $\mathbb{P}_x$. Puisque

$$N_x(\omega) = 1 + N_x(\theta_{T_x(\omega)}(\omega)),$$

on voit que sous $\mathbb{P}_x$, N_x a la même loi que $1 + N_x$, ce qui n'est possible que si $N_x = \infty$, $\mathbb{P}_x$-p.s.

Le théorème suivant nous permettra de rendre cet argument rigoureux ; et le résultat sera discuté en détail dans la section prochaine. Rappelons qu'un temps d'arrêt par rapport à la filtration $(\mathscr{F}_n)_{n\geq 0}$ est une application $T : \Omega \to \mathbb{N} \cup \{\infty\}$ telle que $\{\omega : \, T(\omega) \leq n\} \in \mathscr{F}_n$, $\forall n \in \mathbb{N}$. Ceci équivaut à $\{\omega : \, T(\omega) = n\} \in \mathscr{F}_n$, $\forall n \in \mathbb{N}$. Si T est un temps d'arrêt, $\mathscr{F}_T := \{A \subset \Omega : \, A \cap \{T \leq n\} \in \mathscr{F}_n, \ \forall n \in \mathbb{N}\}$ est la tribu engendrée par T ; on vérifie facilement que $\mathscr{F}_T = \{A \subset \Omega : \, A \cap \{T = n\} \in \mathscr{F}_n, \ \forall n \in \mathbb{N}\}$.

On s'intéresse à la variable aléatoire $\theta_T \, \mathbf{1}_{\{T<\infty\}}$ qui est à valeurs dans $E^{\mathbb{N}}$. [Pour voir la mesurabilité, il suffit d'écrire $\theta_T \, \mathbf{1}_{\{T<\infty\}} = \sum_{n\in\mathbb{N}} \theta_n \, \mathbf{1}_{\{T=n\}}$.]

Théorème 3.7. (Propriété de Markov forte). *Soit T un temps d'arrêt. Soit $Y : \Omega \to \mathbb{R}_+$ une application mesurable. Pour tout $x \in E$,*

$$\mathbb{E}_x[(Y \circ \theta_T) \, \mathbf{1}_{\{T<\infty\}} \,|\, \mathscr{F}_T] = \mathbf{1}_{\{T<\infty\}} \mathbb{E}_{X_T}(Y), \qquad \mathbb{P}_x\text{-p.s.}$$

Remarque 3.8. On voit assez facilement que la variable aléatoire X_T, définie sur l'ensemble $\{T < \infty\} \in \mathscr{F}_T$, est $\mathscr{F}_T$-mesurable. La mesurabilité de $\mathbb{E}_{X_T}(Y)$, également définie sur $\{T < \infty\}$, est claire car $\mathbb{E}_{X_T}(Y)$ est composition des applications mesurables[7] $\omega \to X_{T(\omega)}(\omega)$ et $x \mapsto \mathbb{E}_x(Y)$. □

[7]La mesurabilité de $x \mapsto \mathbb{E}_x(Y)$ est triviale, car on travaille dans un espace dénombrable.

Preuve. Il s'agit de prouver que $\mathbb{E}_x[(Y \circ \theta_T)\, \mathbf{1}_{\{T<\infty\}}\, \mathbf{1}_B] = \mathbb{E}_x[\mathbf{1}_B\, \mathbf{1}_{\{T<\infty\}}\, \mathbb{E}_{X_T}(Y)]$ pour tout $B \in \mathscr{F}_T$.

Soit $n \in \mathbb{N}$. On a

$$\mathbb{E}_x[(Y \circ \theta_T)\, \mathbf{1}_{\{T=n\}}\, \mathbf{1}_B] = \mathbb{E}_x[(Y \circ \theta_n)\, \mathbf{1}_{\{T=n\}\cap B}] = \mathbb{E}_x[\mathbf{1}_{\{T=n\}}\, \mathbf{1}_B\, \mathbb{E}_{X_n}(Y)],$$

la seconde identité étant une conséquence de la propriété de Markov (en remarquant que $\{T = n\} \cap B \in \mathscr{F}_n$). En sommant sur $n \in \mathbb{N}$, on obtient l'identité cherchée. □

Corollaire 3.9. *Soit $x \in E$. Soit $i \in E$. Soit T un temps d'arrêt tel que $T < \infty$ et $X_T = i$, $\mathbb{P}_x$-p.s. Alors sous $\mathbb{P}_x$, θ_T est indépendante de $\mathscr{F}_T$, et a pour loi $\mathbb{P}_i$.*

Remarque 3.10. La loi sous $\mathbb{P}_x$ de la variable aléatoire θ_T (à valeurs dans Ω) est une mesure de probabilité sur Ω. La seconde partie du corollaire dit que cette loi n'est autre que $\mathbb{P}_i$. Puisque $\theta_T = (X_T,\, X_{T+1},\, \cdots)$, ceci peut se reformuler de la façon suivante : $(X_T,\, X_{T+1},\, \cdots)$ sous $\mathbb{P}_x$ a la même loi que $(X_0,\, X_1,\, \cdots)$ sous $\mathbb{P}_i$.

Preuve du Corollaire 3.9. Il s'agit de prouver que $\mathbb{E}_x[(Y \circ \theta_T)\, \mathbf{1}_B] = \mathbb{E}_i(Y)\, \mathbb{P}_x(B)$ pour tout $B \in \mathscr{F}_T$ et toute application mesurable $Y : \Omega \to \mathbb{R}_+$. Cette identité, cependant, est triviale car d'après la propriété de Markov forte, $\mathbb{E}_x[(Y \circ \theta_T)\, \mathbf{1}_B] = \mathbb{E}_x[\mathbf{1}_B \mathbb{E}_{X_T}(Y)] = \mathbb{E}_x[\mathbf{1}_B \mathbb{E}_i(Y)]$, qui n'est autre que $\mathbb{P}_x(B)\, \mathbb{E}_i(Y)$. □

4. États récurrents et transients

Sauf mention du contraire (notamment dans certains exemples), on travaille désormais avec la chaîne de Markov canonique. Rappelons nos notations :[8] pour $x \in E$,

$$\begin{aligned} T_x &:= \inf\{n \geq 1 : X_n = x\}, \qquad (\inf \varnothing := \infty) \\ N(x) &:= \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n = x\}}. \end{aligned}$$

Proposition 4.1. (et définition). *Soit $x \in E$. Il y a deux situations possibles :*

- *soit $\mathbb{P}_x(T_x < \infty) = 1$; dans ce cas,*

$$N(x) = \infty, \qquad \mathbb{P}_x\text{-p.s.},$$

[8] On écrit $N(x)$ à la place de N_x.

et on dit que x est un état **récurrent***;*

- *soit $\mathbb{P}_x(T_x < \infty) < 1$; dans ce cas,*

$$N(x) < \infty, \qquad \mathbb{P}_x\text{-p.s.},$$

et de plus, $\mathbb{E}_x[N(x)] = \frac{1}{\mathbb{P}_x\{T_x=\infty\}} < \infty$; on dit que x est un état **transient** *(ou : transitoire).*

Preuve. Soit $k \geq 1$. On a

$$\begin{aligned}
\mathbb{P}_x\{N(x) \geq k+1\} &= \mathbb{E}_x\{\mathbf{1}_{\{T_x<\infty\}}\,(\mathbf{1}_{\{N(x)\geq k\}} \circ \theta_{T_x})\} \\
&= \mathbb{E}_x\{\mathbf{1}_{\{T_x<\infty\}}\,\mathbb{E}_{X_{T_x}}(\mathbf{1}_{\{N(x)\geq k\}})\} \qquad \text{(propriété de Markov forte)} \\
&= \mathbb{E}_x\{\mathbf{1}_{\{T_x<\infty\}}\,\mathbb{P}_x(N(x) \geq k)\} \qquad (X_{T_x} = x \text{ sur } \{T_x < \infty\}) \\
&= \mathbb{P}_x\{T_x < \infty\}\,\mathbb{P}_x\{N(x) \geq k\}.
\end{aligned}$$

Puisque $\mathbb{P}_x\{N(x) \geq 1\} = 1$, on obtient immédiatement par récurrence que $\mathbb{P}_x\{N(x) \geq k\} = \mathbb{P}_x\{T_x < \infty\}^{k-1}$.

Si $\mathbb{P}_x\{T_x < \infty\} = 1$, alors $\mathbb{P}_x\{N(x) \geq k\} = 1$, $\forall k \geq 1$, ce qui signifie $\mathbb{P}_x\{N(x) = \infty\} = 1$.

Si $\mathbb{P}_x\{T_x < \infty\} < 1$, alors

$$\mathbb{E}_x[N(x)] = \sum_{k=1}^{\infty} \mathbb{P}_x\{N(x) \geq k\} = \frac{1}{\mathbb{P}_x\{T_x = \infty\}},$$

ce qui est fini par hypothèse. □

Définition 4.2. *La* **fonction de Green** *(= noyau de Green = noyau potentiel) de la chaîne de Markov est la fonction $G:\ E \times E \to [0, \infty]$ définie par*

$$G(x, y) := \mathbb{E}_x[N(y)].$$

Proposition 4.3. (i) *Pour tous $x, y \in E$,*

$$G(x, y) = \sum_{n=0}^{\infty} [Q^n](x, y).$$

(ii) *$G(x, x) = \infty$ si et seulement si x est un état récurrent.*

(iii) *Si $x \neq y$, alors*[9]

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

En particulier, $G(x, y) \leq G(y, y)$.

[9] Notation : $\infty \times 0 := 0$.

Preuve. (i) Par définition,

$$G(x,\,y) = \mathbb{E}_x\Big(\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=y\}}\Big) = \sum_{n=0}^{\infty} \mathbb{P}_x\{X_n = y\} = \sum_{n=0}^{\infty}[Q^n](x,\,y).$$

(ii) C'est une conséquence de la Proposition 4.1 et de la définition de la fonction de Green.

(iii) Par la propriété de Markov forte,

$$G(x,\,y) = \mathbb{E}_x[N(y)] = \mathbb{E}_x[\mathbf{1}_{\{T_y<\infty\}}\,(N(y)\circ\theta_{T_y})] = \mathbb{E}_x[\mathbf{1}_{\{T_y<\infty\}}\,\mathbb{E}_y(N(y))],$$

qui vaut $\mathbb{P}_x\{T_y < \infty\}\,G(y,\,y)$. □

Exemple 4.4. Considérons une chaîne de Markov de matrice de transition

$$p(x,\,y) = \frac{1}{2^d}\prod_{i=1}^{d}\mathbf{1}_{\{|y_i-x_i|=1\}}, \qquad x := (x_1,\,\cdots,\,x_d),\; y := (y_1,\,\cdots,\,y_d) \in \mathbb{Z}^d.$$

Il s'agit d'une marche aléatoire sur $\mathbb{Z}^d$ un peu spéciale, qui n'est pas la marche aléatoire simple sur $\mathbb{Z}^d$. En effet, sous $\mathbb{P}_0$, elle a la loi de $(Y_n^1,\,\cdots,\,Y_n^d)_{n\in\mathbb{N}}$, où les processus aléatoires $(Y_n^1)_{n\in\mathbb{N}}$, $\cdots$, $(Y_n^d)_{n\in\mathbb{N}}$ sont des copies indépendantes de la marche aléatoire simple (et symétrique) sur $\mathbb{Z}$ issue de $0 \in \mathbb{Z}$. Par conséquent,

$$[Q^n](0,\,0) = \mathbb{P}(Y_n^1 = 0,\,\cdots,\,Y_n^d = 0) = \mathbb{P}(Y_n^1 = 0)^d.$$

On calcule facilement $\mathbb{P}(Y_n^1 = 0)$: cette probabilité est 0 si n est impair, et si $n = 2k$, alors

$$\mathbb{P}(Y_{2k}^1 = 0) = \frac{C_{2k}^k}{2^{2k}}.$$

Ainsi,

$$G(0,\,0) = \sum_{k=0}^{\infty}[Q^{2k}](0,\,0) = \sum_{k=0}^{\infty}\Big(\frac{C_{2k}^k}{2^{2k}}\Big)^d.$$

Par la formule de Stirling, lorsque $k \to \infty$,

$$\frac{C_{2k}^k}{2^{2k}} = \frac{(2k)!}{2^{2k}\,(k!)^2} \sim \frac{(\frac{2k}{\mathrm{e}})^{2k}(4\pi k)^{1/2}}{2^{2k}\,[(\frac{k}{\mathrm{e}})^k(2\pi k)^{1/2}]^2} = \frac{1}{(\pi k)^{1/2}}.$$

Par conséquent, $G(0,\,0) = \infty$ si $d = 1$ ou 2, et $G(0,\,0) < \infty$ si $d \geq 3$. À l'aide de la Proposition 4.3, on peut conclure que $0 \in \mathbb{Z}^d$ est un état récurrent pour la chaîne si $d = 1$ ou 2, et est un état transient si $d \geq 3$. □

On procède maintenant à un résultat de décomposition pour la chaîne de Markov. Commençons par un résultat préliminaire.

Lemme 4.5. *Soit $x \in E$ un état récurrent. Soit y tel que $G(x, y) > 0$. Alors y est également un état récurrent, et $\mathbb{P}_y(T_x < \infty) = 1$. En particulier, $G(y, x) > 0$, et $\mathbb{P}_x(T_y < \infty) = 1$.*

Preuve. Il n'y a rien à prouver si $y = x$. On suppose donc $y \neq x$.

Vérifions d'abord $\mathbb{P}_y(T_x < \infty) = 1$. Puisque x est un état récurrent,

$$\begin{aligned} 0 = \mathbb{P}_x\{N(x) < \infty\} &\geq \mathbb{P}_x\{T_y < \infty,\ T_x \circ \theta_{T_y} = \infty\} \\ &= \mathbb{E}_x[\mathbf{1}_{\{T_y<\infty\}}\, \mathbf{1}_{\{T_x=\infty\}} \circ \theta_{T_y}] \\ &= \mathbb{E}_x[\mathbf{1}_{\{T_y<\infty\}}\, \mathbb{P}_y\{T_x = \infty\}] \qquad \text{(propriété de Markov forte)} \\ &= \mathbb{P}_x\{T_y < \infty\}\, \mathbb{P}_y\{T_x = \infty\}. \end{aligned}$$

Donc $\mathbb{P}_x\{T_y < \infty\}\, \mathbb{P}_y\{T_x = \infty\} = 0$. Par hypothèse, $\mathbb{P}_x\{T_y < \infty\} > 0$ (ce qui équivaut à $G(x, y) > 0$). Ainsi, $\mathbb{P}_y\{T_x = \infty\} = 0$ comme désiré.

Il reste de prouver que y est un état récurrent. On sait déjà que $G(x, y) > 0$ et $G(y, x) > 0$. Il existe donc des entiers $n \geq 1$ et $m \geq 1$ tels que $[Q^n](x, y) > 0$ et que $[Q^m](y, x) > 0$. Pour tout entier $i \geq 0$,

$$[Q^{m+i+n}](y, y) \geq [Q^m](y, x)\, [Q^i](x, x)\, [Q^n](x, y).$$

Ainsi

$$\begin{aligned} G(y, y) &\geq \sum_{i=0}^{\infty} [Q^{m+i+n}](y, y) \\ &\geq [Q^m](y, x)\, [Q^n](x, y) \sum_{i=0}^{\infty} [Q^i](x, x) \\ &= [Q^m](y, x)\, [Q^n](x, y)\, G(x, x). \end{aligned}$$

Par hypothèse, $G(x, x) = \infty$ (car x est un état récurrent), tandis que $[Q^m](y, x)\, [Q^n](x, y) > 0$, donc $G(y, y) = \infty$. En d'autres mots, y est un état récurrent. □

Une conséquence du Lemme 4.5 est que si x est un état récurrent et si y est un état transient, alors $G(x, y) = 0$: il est impossible de connecter d'un état récurrent à un état transient. Cette propriété joue un rôle important dans le théorème suivant. Posons

$$R := \{x \in E :\ G(x, x) = \infty\} = \{x \in E :\ x \text{ est un état récurrent}\}.$$

Théorème 4.6. (Classification des états). *L'ensemble des états récurrents peut être décomposé de la façon suivante :*

$$R = \bigcup_{i \in I} R_i,$$

de sorte que

- *si x est un état récurrent et si $x \in R_i$, alors $\mathbb{P}_x$-p.s.,*
  — $N(y) = \infty$, $\forall y \in R_i$;
  — $N(y) = 0$, $\forall y \in E \backslash R_i$;
- *si x est un état transient et $T := \inf\{n \geq 1 : X_n \in R\}$, alors $\mathbb{P}_x$-p.s.,*
  — *soit $T = \infty$ et $N(y) < \infty$, $\forall y \in E$;*
  — *soit $T < \infty$ et il existe $j \in I$ tel que $X_n \in R_j$, $\forall n \geq T$.*

Preuve. Pour $x, y \in R$, on écrit $x \sim y$ si $G(x, y) > 0$. La réflexivité $x \sim x$ est déjà acquise (Proposition 4.3), ainsi que la symétrie (voir le Lemme 4.5) : si $x \sim y$ alors $y \sim x$. Pour la transitivité, supposons que $x \sim y$ et $y \sim z$, il existe alors des entiers $m \geq 1$ et $n \geq 1$ tels que $[Q^m](x, y) > 0$ et $[Q^n](y, z) > 0$, de sorte que $[Q^{m+n}](x, z) \geq [Q^m](x, y)\,[Q^n](y, z) > 0$, ce qui implique que $G(x, z) > 0$, ainsi $x \sim z$. En mots, $\sim$ définit une relation d'équivalence sur R. La partition énoncée dans le théorème correspond aux classes d'équivalence pour cette relation d'équivalence, appelées dans la littérature **classes de récurrence** pour la chaîne de Markov.

Soit x un état récurrent, disons $x \in R_i$ pour un certain $i \in I$. Pour $y \in E \backslash R_i$, on a $G(x, y) = 0$ (si $y \in E \backslash R$ d'après le Lemme 4.5, tandis que si $y \in R \backslash R_i$, ceci est une conséquence de la définition des classes d'équivalence), ce qui signifie que $N(y) = 0$, $\mathbb{P}_x$-p.s. Pour $y \in R_i$, on a $\mathbb{P}_x(T_y < \infty) = 1$ d'après le Lemme 4.5, et donc par la propriété de Markov forte,

$$\mathbb{P}_x\{N(y) = \infty\} = \mathbb{E}_x\{\mathbf{1}_{\{T_y < \infty\}}\,[\mathbf{1}_{\{N(y)=\infty\}} \circ \theta_{T_y}]\} = \mathbb{P}_x\{T_y < \infty\}\,\mathbb{P}_y\{N(y) = \infty\} = 1.$$

Regardons maintenant le cas où x est un état transient. Notons que d'après la propriété de Markov forte, $N(y) < \infty$ si y est un état transient. Si $T = \infty$, alors $N(y) < \infty$ pour tout $y \in E$. Si $T < \infty$, soit $j \in I$ l'indice (aléatoire) tel que $X_T \in R_j$, il résulte alors de la propriété de Markov forte et de la première partie du théorème que l'on a déjà prouvée, $X_n \in R_j$ pour tout $n \geq T$. $\square$

Définition 4.7. *Une chaîne de Markov est* **irréductible** *si $G(x, y) > 0$ pour tous $x, y \in E$.*

Corollaire 4.8. *Supposons que la chaîne de Markov est irréductible. Il y a alors deux situations possibles :*

- *tous les états sont récurrents, et il n'y a qu'une classe de récurrence, et pour tout $x \in E$,*

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

- *tous les états sont transients, et pour tout $x \in E$,*

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

Lorsque E est un espace fini, seule la première situation se réalise.

Preuve. La dernière assertion est évidente car lorsque $\#E < \infty$, la seconde situation ne peut se produire.

Supposons qu'il existe un état récurrent. Dans ce cas, le Lemme 4.5 nous dit que tous les états sont récurrents (car par l'irréductibilité, $G(x, y) > 0$ pour tout couple d'états x et y) et qu'il n'y a qu'une seule classe de récurrence (pour la même raison). Le théorème de classification des états (Théorème 4.6) implique alors que $\mathbb{P}_x\{N(y) = \infty,\ \forall y \in E\} = 1$, $\forall x \in E$.

Supposons maintenant que tous les états sont transients. On est alors dans la seconde situation décrite dans le théorème de classification des états, et $T = \infty$ (car $R = \varnothing$). D'après ce théorème, $\mathbb{P}_x\{N(y) < \infty,\ \forall y \in E\} = 1$, $\forall x \in E$. □

Lorsque les états d'une chaîne de Markov irréductible sont récurrents, on dit que la chaîne est **irréductible et récurrente** ; de temps en temps, on dit également que la matrice de transition Q est irréductible et récurrente.

Exemples. On reprend les exemples présentés en Section 2, et discute sur la classification des états pour chacune des chaînes dans ses exemples. On insiste sur le fait que les résultats obtenus peuvent être formulés pour une chaîne de Markov $(Y_n)_{n\in\mathbb{N}}$ (définie sur un certain espace, sous la probabilité $\mathbf{P}$) avec la même matrice de transition, et vice versa. Par exemple, si $Y_0 = y$ et en écrivant $N_x^Y := \sum_{n=0}^{\infty} \mathbf{1}_{\{Y_n = x\}}$, on voit que pour $k \in \mathbb{N} \cup \{\infty\}$,

$$\mathbf{P}\{N_x^Y = k\} = \mathbb{P}_y\{N(x) = k\},$$

car $\mathbf{P}\{N_x^Y = k\} = \mathbf{P}\{(Y_n)_{n\in\mathbb{N}} \in A\}$, avec $A := \{\omega \in E^{\mathbb{N}} :\ N_x(\omega) = k\}$, et il suffit d'appliquer la Remarque 3.4 (c).

Premier exemple : variables aléatoires i.i.d. de loi μ. On a vu que $p(x, y) = \mu(y)$. Il est clair que y est un état récurrent si et seulement si $\mu(y) > 0$, et qu'il n'y a qu'une seule classe de récurrence. La chaîne est irréductible si et seulement si $\mu(y) > 0$ pour tout $y \in E$.

Deuxième exemple : marches aléatoires sur $\mathbb{Z}$. Dans cet exemple, $E = \mathbb{Z}$, et la matrice de transition est $p(x, y) = \mu(y - x)$. On vérifie aisément que $G(x, y)$ dépend seulement de $y - x$, ainsi $G(x, x) = G(y, y)$; en particulier, tous les états sont du même type : ils sont soit tous récurrents, soit tous transients. Soit ξ une variable aléatoire générique de loi μ.

Théorème 4.9. *Supposons que* $\mathbb{E}(|\xi|) < \infty$.

(i) *Si* $\mathbb{E}(\xi) \neq 0$, *tous les états sont transients.*

(ii) *Si* $\mathbb{E}(\xi) = 0$, *tous les états sont récurrents. La chaîne est irréductible si et seulement si le sous-groupe*[10] *engendré par* $\{y \in \mathbb{Z} : \mu(y) > 0\}$ *est* $\mathbb{Z}$.

Preuve. (i) Soit $x \in \mathbb{Z}$. La loi des grands nombres implique que $|Y_n| \to \infty$ $\mathbb{P}_x$-p.s. Donc x est un état transient.

(ii) Montrons d'abord, par un raisonnement par l'absurde, que 0 est un état récurrent.

Supposons que 0 soit un état transient, c'est-à-dire $G(0, 0) < \infty$. Pour tout $x \in \mathbb{Z}$, on a, par la Proposition 4.3 (iii),

$$G(0, x) \leq G(x, x) = G(0, 0).$$

Donc, pour tout $n \geq 1$,

$$\sum_{|x| \leq n} G(0, x) \leq (2n + 1)G(0, 0) \leq cn,$$

où $c := 3\, G(0, 0) < \infty$.

D'autre part, la loi faible des grands nombres nous dit que, $\frac{Y_n}{n} \to 0$ en probabilité sous $\mathbb{P}_0$. Donc pour tout $\varepsilon > 0$, et tout n suffisamment grand (disons, $n \geq n_0$),

$$\mathbb{P}\{|Y_n| \leq \varepsilon n\} > \frac{1}{2},$$

ce qui équivaut à :

$$\sum_{|x| \leq \varepsilon n} [Q^n](0, x) > \frac{1}{2}.$$

[10]Rappelons que $E = \mathbb{Z}$.

Pour $n \geq i \geq n_0$, on a

$$\sum_{|x| \leq \varepsilon n} [Q^i](0, x) \geq \sum_{|x| \leq \varepsilon i} [Q^i](0, x) > \frac{1}{2},$$

de sorte que pour $n \geq n_0$,

$$\sum_{|x| \leq \varepsilon n} G(0, x) \geq \sum_{i=n_0}^{n} \sum_{|x| \leq \varepsilon i} [Q^i](0, x) > \frac{n - n_0 + 1}{2},$$

ce qui contredit l'inégalité

$$\sum_{|x| \leq \varepsilon n} G(0, x) \leq c\varepsilon n,$$

si l'on choisit ε suffisamment petit tel que $c\varepsilon < \frac{1}{2}$. Par conséquent, 0 est un état récurrent.

Puisque tous les états sont d'un même type, on déduit qu'ils sont tous récurrents.

Il reste de caractériser l'irréductibilité. Soit $\mathbb{G}$ le sous-groupe engendré par $\{x \in \mathbb{Z} : \mu(x) > 0\}$. On a

$$\mathbb{P}_0\{Y_n \in \mathbb{G},\ \forall n \in \mathbb{N}\} = 1.$$

Si $\mathbb{G} \neq \mathbb{Z}$, la chaîne n'est pas irréductible.

Supposons maintenant que $\mathbb{G} = \mathbb{Z}$. Soit

$$\mathbb{G}_1 := \{x \in \mathbb{Z} : G(0, x) > 0\}.$$

Alors $\mathbb{G}_1$ est un sous-groupe, car

- si $x, y \in \mathbb{G}_1$, alors

$$[Q^{n+m}](0, x + y) \geq [Q^n](0, x)\, [Q^m](x, x + y) = [Q^n](0, x)\, [Q^m](0, y),$$

et donc $x + y \in \mathbb{G}_1$;

- associativité est triviale ;
- élément neutre : 0 est évidemment un élément de $\mathbb{G}_1$;
- si $x \in \mathbb{G}_1$, et puisque 0 est un état récurrent, $G(0, x) > 0$ implique que $G(x, 0) > 0$ (Lemme 4.5), et comme $G(x, 0) = G(0, -x)$, on a $-x \in \mathbb{G}_1$.

Enfin, puisque $\mathbb{G}_1 \supset \{x \in \mathbb{Z} : \mu(x) > 0\}$, on voit que $\mathbb{G}_1$ inclut le sous-groupe $\mathbb{G}$ engendré par $\{x \in \mathbb{Z} : \mu(x) > 0\}$ qui par hypothèse est $\mathbb{Z}$. Par conséquent, $\mathbb{G}_1 = \mathbb{Z}$: la chaîne est bien irréductible. □

On termine les discussions sur l'exemple des marches aléatoires sur $\mathbb{Z}$ en remarquant que lorsque $\mu := \frac{1}{2}\,\delta_2 + \frac{1}{2}\,\delta_{-2}$, tous les états sont récurrents, mais il y a deux classes de récurrence (celle des entiers pairs, et celle des entiers impairs).

Troisième exemple : marche aléatoire simple sur un graphe. On considère seulement la situation où E est fini. On suppose que pour tout $x \in E$, $A_x := \{y \in E : \{x, y\} \in A\}$ est non vide. On dit que le graphe est **connecté** si pour tout couple $x, y \in E$, il existe $n \geq 0$ et $x_0 := x, x_1, \cdots, x_n := y \in E$ tels que $\{x_{i-1}, x_i\} \in A$, $\forall i \in \{1, \cdots, n\}$.

Proposition 4.10. *La marche aléatoire simple sur un graphe fini et connecté est irréductible et récurrente.*

Preuve. L'irréductibilité provient de la connectivité du graphe, tandis que la récurrence est une conséquence du Corollaire 4.8. □

Quatrième exemple : processus de branchement. Dans cet exemple, $E = \mathbb{N}$ et $p(x, y) = \mu^{*x}(y)$ (avec $\mu^{0*} := \delta_0$). On observe que 0 est un état **absorbant**, c'est-à-dire

$$\mathbb{P}_0\{X_n = 0, \ \forall n \in \mathbb{N}\} = 1.$$

A fortiori, 0 est un état récurrent.

Dans la proposition suivante, on exclut le cas trivial $\mu = \delta_1$, où tous les états sont absorbants.

Proposition 4.11. *Pour le processus de branchement avec $\mu \neq \delta_1$, 0 est le seul état récurrent. Par conséquent,* p.s.,

- *soit il existe $T < \infty$ tel que $X_n = 0$, $\forall n \geq T$;*
- *soit $X_n \to \infty$, $n \to \infty$.*

Preuve. Il suffit de prouver que 0 est le seul état récurrent, le reste de l'énoncé étant une conséquence du théorème de classification des états (Théorème 4.6).

Soit $x \geq 1$. On veut prouver que x est un état transient. On distingue deux situations possibles. D'après le Lemme 4.5, il suffit de trouver $y \in \mathbb{N}$ tel que $G(x, y) > 0$ et que $G(y, x) = 0$.

Première situation : $\mu(0) > 0$. Alors $G(x, 0) \geq \mathbb{P}_x\{X_1 = 0\} = \mu(0)^x > 0$, tandis que $G(0, x) = 0$.

Seconde (et dernière) situation : $\mu(0) = 0$. Puisque $\mu \neq \delta_1$, il existe $k \geq 2$ tel que $\mu(k) > 0$. Comme $\mathbb{P}_x\{X_1 > x\} \geq \mu(k)^x > 0$, il existe $y > x$ tel que $p(x, y) > 0$, et a fortiori, $G(x, y) > 0$. D'autre part, on a $G(y, x) = 0$ car $\mu(0) = 0$. □

Remarque 4.12. On a vu, en Section 5 du Chapitre 2, que la première situation décrite dans la Proposition 4.11 est assurée p.s. si $m := \sum_{k\in\mathbb{N}} k\,\mu(k) \le 1$, tandis que la seconde situation se produit avec probabilité strictement positive si $1 < m < \infty$ (au moins sous une condition d'intégrabilité supplémentaire, qui est en fait superficielle) et que dans ce dernier cas, X_n est exponentiellement grand en n.[11]

5. Mesures invariantes

Sauf mention du contraire, on note μ pour une mesure générique sur E telle que $\mu(x) < \infty$, $\forall x \in E$, et on suppose qu'elle n'est pas identiquement nulle. On dit que μ est une mesure **stationaire** ou **invariante** pour la matrice de transition Q (ou simplement, invariante, lorsque qu'il n'y a pas de risque de confusion), si

$$\mu(y) = \sum_{x\in E} \mu(x)\, p(x,\, y), \qquad \forall y \in E.$$

Il convient souvent d'écrire sous la forme matricielle : $\mu Q = \mu$. Si μ est une mesure invariante, on a $\mu Q^n = \mu$ pour tout $n \in \mathbb{N}$.

Interprétation. Supposons que $\mu(E) < \infty$ (ce qui est automatiquement le cas si E est un espace fini). Sans perte de généralité, on peut supposer que μ est une mesure de probabilité (car sinon, on remplace μ par $\frac{\mu}{\mu(E)}$), et on écrit dans ce cas π à la place de μ. Pour toute application $f :\, E \to \mathbb{R}_+$,

$$\mathbb{E}_\pi[f(X_1)] = \sum_{x\in E} \pi(x) \sum_{y\in E} f(y)\, p(x,\, y) = \sum_{y\in E} f(y) \sum_{x\in E} \pi(x)\, p(x,\, y) = \sum_{y\in E} f(y)\, \pi(y),$$

ce qui implique que sous $\mathbb{P}_\pi$, X_1 a la même loi π que X_0. Le même argument montre, grâce à la relation $\pi Q^n = \pi$, que pour tout $n \in \mathbb{N}$, la loi de X_n sous $\mathbb{P}_\pi$ est π.

On peut en dire plus. Soit $Y :\, \Omega := E^{\mathbb{N}} \to \mathbb{R}_+$ une application mesurable. On a

$$\mathbb{E}_\pi[Y \circ \theta_1] = \mathbb{E}_\pi[\mathbb{E}_{X_1}(Y)] = \sum_{x\in E} \pi(x)\, \mathbb{E}_x(Y) = \mathbb{E}_\pi(Y).$$

Ceci montre que sous $\mathbb{P}_\pi$, $(X_{1+n},\, n \in \mathbb{N})$ a la même loi que $(X_n,\, n \in \mathbb{N})$. Plus généralement, pour tout $k \in \mathbb{N}$, sous $\mathbb{P}_\pi$, $(X_{k+n},\, n \in \mathbb{N})$ et $(X_n,\, n \in \mathbb{N})$ ont la même loi. □

[11] Il est connu dans la littérature que la seconde situation se produit avec probabilité strictement positive si $1 < m \le \infty$. Voir les discussions en TD à la fin du semestre.

Exemple 5.1. Considérons une marche aléatoire sur $\mathbb{Z}^d$. Puisque $p(x, y)$ dépend seulement de $y - x$, on voit immédiatement que la mesure de comptage sur $\mathbb{Z}^d$ est invariante. □

Définition 5.2. *Soit $\mu \neq 0$ une mesure sur E telle que $\mu(x) < \infty$, $\forall x \in E$. On dit que μ est* **réversible** *pour la matrice de transition Q si*

$$\mu(x)\, p(x, y) = \mu(y)\, p(y, x), \qquad \forall x, y \in E.$$

Proposition 5.3. *Une mesure réversible est invariante.*

Preuve. Si μ est réversible, alors

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

ce qui signifie que μ est invariante. □

Remarque 5.4. Attention : Il existe des mesures invariantes qui ne sont pas réversibles. Par exemple, considérons une marche aléatoire sur $\mathbb{Z}^d$: la mesure de comptage sur $\mathbb{Z}^d$ est invariante, mais elle est réversible si et seulement si $p(x, y) = p(y, x)$, $\forall x, y \in E$ (i.e., la loi de saut est symétrique). □

Exemples. (a) **Jeu de pile ou face biaisé.** Considérons une marche aléatoire sur $\mathbb{Z}$ de matrice de transition

$$p(i, i+1) = p, \qquad p(i, i-1) = q = 1 - p,$$

où $p \in]0, 1[$ est un paramètre fixé. On vérifie aisément que la mesure

$$\mu(i) := (\frac{p}{q})^i, \qquad i \in \mathbb{Z},$$

est réversible, donc invariante. Remarquons que μ est differente de la mesure de comptage (qui est également invariante), sauf dans le cas non biaisé $p = \frac{1}{2}$.

(b) **Marche aléatoire simple sur le graphe.** La mesure

$$\mu(x) := \#A_x, \qquad x \in E,$$

est réversible. En effet, si $\{x, y\} \in A$, alors

$$\mu(x)\, p(x, y) = \#A_x \, \frac{1}{\#A_x} = 1 = \mu(y)\, p(y, x),$$

et $\mu(x)\, p(x,\, y) = 0 = \mu(y)\, p(y,\, x)$ if $\{x,\, y\} \notin A$.

(c) **Urnes d'Ehrenfest.** Dans cet exemple, $E := \{0,\, 1,\, \cdots,\, r\}$, où r est un entier fixé, et

$$p(i,\, j) := \begin{cases} \frac{r-i}{r} & \text{si } j = i+1 \text{ (et } i \le r-1) \\ \frac{i}{r} & \text{si } j = i-1 \text{ (et } i \ge 1) \\ 0 & \text{sinon.} \end{cases}$$

En mots, il y a au total r boules dans deux urnes ; on tire par hasard une boule parmi les r, et la remet dans l'autre urne.

Une mesure μ est réversible si et seulement si

$$\mu(i)\, \frac{r-i}{r} = \mu(i+1)\, \frac{i+1}{r},$$

pour $0 \le i \le r-1$. It est immédiat que

$$\mu(i) := C_r^i, \qquad 0 \le i \le r\,,$$

convenient. □

Le théorème suivant construit explicitement une mesure invariante.

Théorème 5.5. *Soit $x \in E$ un état récurrent. La formule*

$$\mu(y) := \mathbb{E}_x\Big(\sum_{i=0}^{T_x - 1} \mathbf{1}_{\{X_i = y\}} \Big), \qquad y \in E,$$

donne une mesure invariante. De plus, $\mu(y) > 0$ si et seulement si y est dans la même classe de récurrence que x.

Remarque 5.6. Une partie de la conclusion du théorème est que $\mu(y) < \infty$, $\forall y \in E$. □

Preuve du Théorème 5.5. Remarquons tout d'abord que si y n'est pas dans la même classe de récurrence que x (ce qui est le cas si y est un état transient), alors $\mathbb{E}_x[N(y)] = G(x,\, y) = 0$; d'où $\mu(y) = 0$.

Calculons maintenant μ. Pour tout $y \in E$, on a

$$\mu(y) = \mathbb{E}_x\Big(\sum_{i=1}^{T_x} \mathbf{1}_{\{X_i = y\}} \Big).$$

Ceci reste valable dans les cas $y = x$ et $y \neq x$. Donc

$$\begin{aligned}
\mu(y) &= \sum_{z\in E} \mathbb{E}_x\Big(\sum_{i=1}^{T_x} \mathbf{1}_{\{X_{i-1}=z,\, X_i=y\}}\Big) && \text{(Fubini–Tonelli)}\\
&= \sum_{z\in E}\sum_{i=1}^{\infty} \mathbb{E}_x\Big(\mathbf{1}_{\{T_x\ge i,\, X_{i-1}=z\}}\mathbf{1}_{\{X_i=y\}}\Big) && \text{(Fubini–Tonelli)}\\
&= \sum_{z\in E}\sum_{i=1}^{\infty} \mathbb{E}_x\Big(\mathbf{1}_{\{T_x\ge i,\, X_{i-1}=z\}}\Big) p(z,\, y), && \text{(propriété de Markov)}
\end{aligned}$$

l'application de la propriété de Markov étant justifiée par le fait que $\{T_x \geq i\} = \{T_x \leq i-1\}^c \in \mathscr{F}_{i-1}$. Donc, de nouveau par le théorème de Fubini,

$$\mu(y) = \sum_{z\in E} \mathbb{E}_x\Big(\sum_{i=1}^{T_x} \mathbf{1}_{\{X_{i-1}=z\}}\Big) p(z,\, y) = \sum_{z\in E} \mu(z)\, p(z,\, y).$$

Attention : nous n'avons pas encore prouvé, à ce stade, que μ est une mesure invariante, car nous devons vérifier que $\mu(y) < \infty$, $\forall y \in E$. Néanmoins, en itérant l'argument précédent, on voit que $\mu Q^n = \mu$, $\forall n \in \mathbb{N}$. En particulier,

$$\sum_{z\in E} \mu(z)\, [Q^n](z,\, x) = \mu(x) = 1, \qquad \forall n \in \mathbb{N}.$$

[Par définition, $\mu(x) = 1$.]

Soit $y \neq x$ un état dans la même classe de récurrence que x. Il reste de démontrer que $0 < \mu(y) < \infty$.

Par définition des classes de récurrence, il existe $n \in \mathbb{N}$ tel que $[Q^n](y,\, x) > 0$. De l'identité $\sum_{z\in E} \mu(z)\, [Q^n](z,\, x) = 1$, on tire que $\mu(y) < \infty$.

Puisque y est dans la même classe de récurrence que x, il existe $m \in \mathbb{N}$ tel que $[Q^m](x,\, y) > 0$, ce qui implique que

$$\mu(y) = \sum_{z\in E} \mu(z)\, [Q^m](z,\, y) \geq \mu(x)\, [Q^m](x,\, y) = [Q^m](x,\, y) > 0.$$

La preuve du théorème est complète. □

Remarque 5.7. Lorsque qu'il y a plusieurs classes de récurrence R_i, $i \in I$, on peut choisir, pour chaque $i \in I$, un état $x_i \in R_i$, et définir

$$\mu_i(y) := \mathbb{E}_{x_i}\Big(\sum_{k=0}^{T_{x_i}-1} \mathbf{1}_{\{X_k=y\}}\Big), \qquad y \in E.$$

Ainsi, on obtient des mesures invariantes mutuellement singulières, c'est-à-dire portées par des ensembles disjoints. □

Théorème 5.8. *Supposons que la chaîne est irréductible et récurrente. Il y a unicité, à une constante mulpliticative près, pour la mesure invariante.*

Preuve. Soit μ une mesure invariante.

Montrons d'abord, avec un raisonnement par récurrence sur n, que pour tout $n \in \mathbb{N}$ et tous x, $y \in E$,[12]

$$\mu(y) \ge \mu(x)\, \mathbb{E}_x\Big(\sum_{i=0}^{n\wedge(T_x-1)} \mathbf{1}_{\{X_i=y\}}\Big). \tag{5.1}$$

Si $y = x$, (5.1) est trivialement vraie (on a même une identité). Supposons donc que $y \neq x$. Si $n = 0$, (5.1) est triviale, car l'expression $\mathbb{E}_x(\cdots)$ à droite s'annule.

Supposons que l'inégalité est prouvée au rang n. On a

$$\begin{aligned}
\mu(y) &= \sum_{z\in E} \mu(z)\, p(z,\, y) \\
&\ge \mu(x) \sum_{z\in E} \mathbb{E}_x\Big(\sum_{i=0}^{n\wedge(T_x-1)} \mathbf{1}_{\{X_i=z\}}\Big) p(z,\, y) && \text{(hypothèse de récurrence)} \\
&= \mu(x) \sum_{z\in E} \sum_{i=0}^{n} \mathbb{E}_x\Big(\mathbf{1}_{\{X_i=z,\, i\le T_x-1\}}\Big) p(z,\, y) && \text{(Fubini–Tonelli)} \\
&= \mu(x) \sum_{z\in E} \sum_{i=0}^{n} \mathbb{E}_x\Big(\mathbf{1}_{\{X_i=z,\, i\le T_x-1\}}\, \mathbf{1}_{\{X_{i+1}=y\}}\Big) && \text{(propriété de Markov)} \\
&= \mu(x) \sum_{i=0}^{n} \mathbb{E}_x\Big(\mathbf{1}_{\{i\le T_x-1\}}\, \mathbf{1}_{\{X_{i+1}=y\}}\Big). && \text{(Fubini–Tonelli)}
\end{aligned}$$

[On a utilisé le fait que $\{i \le T_x - 1\} \in \mathscr{F}_i$ au moment d'appliquer la propriété de Markov, exactement comme dans la preuve du Théorème 5.5.] Par le théorème de Fubini–Tonelli,

$$\begin{aligned}
\mu(y) &\ge \mu(x)\, \mathbb{E}_x\Big(\sum_{i=0}^{n\wedge(T_x-1)} \mathbf{1}_{\{X_{i+1}=y\}}\Big) \\
&= \mu(x)\, \mathbb{E}_x\Big(\sum_{i=1}^{(n+1)\wedge T_x} \mathbf{1}_{\{X_i=y\}}\Big) \\
&= \mu(x)\, \mathbb{E}_x\Big(\sum_{i=0}^{(n+1)\wedge(T_x-1)} \mathbf{1}_{\{X_i=y\}}\Big), && (y \neq x)
\end{aligned}$$

[12] Notation: $a \wedge b := \min\{a,\, b\}$.

ce qui donne l'inégalité au rang $n+1$.

En faisant $n \to \infty$ dans (5.1), on obtient, par convergence monotone,

$$\mu(y) \geq \mu(x)\, \mathbb{E}_x\Big(\sum_{i=0}^{T_x - 1} \mathbf{1}_{\{X_i = y\}} \Big).$$

Fixons $x \in E$. La mesure ν_x définie par

$$\nu_x(y) := \mathbb{E}_x\Big(\sum_{i=0}^{T_x - 1} \mathbf{1}_{\{X_i = y\}} \Big), \qquad y \in E,$$

est invariante (voir le Théorème 5.5), et on a $\mu(y) \geq \mu(x)\, \nu_x(y)$, $\forall y \in E$. Donc, pour tout $n \in \mathbb{N}$,

$$\mu(x) = \sum_{z \in E} \mu(z)\, [Q^n](z,\, x) \geq \sum_{z \in E} \mu(x)\, \nu_x(z)\, [Q^n](z,\, x) = \mu(x)\, \nu_x(x) = \mu(x),$$

ce qui signifie que l'identité $\mu(z) = \mu(x)\, \nu_x(z)$ est valable pour tout z tel que $[Q^n](z,\, x) > 0$. L'irréductibilité nous dit que pour tout $z \in E$, il existe $n \in \mathbb{N}$ tel que $[Q^n](z,\, x) > 0$, ce qui permet de déduire que $\mu = c\,\nu_x$, où $c := \mu(x) < \infty$. Puisque $c \neq 0$ (sinon μ s'annulerait identiquement), cela complète la preuve du théorème. □

Corollaire 5.9. *Supposons que la chaîne est irréductible et récurrente. Alors*

- *soit la chaîne est* **récurrente positive** *: il existe une mesure de probabilité invariante π ; dans ce cas*

$$\mathbb{E}_x(T_x) = \frac{1}{\pi(x)} < \infty\ , \qquad \forall x \in E;$$

- *soit la chaîne est* **récurrente nulle** *: la masse totale de toute mesure invariante est infinie ; dans ce cas,*

$$\mathbb{E}_x(T_x) = \infty, \qquad \forall x \in E.$$

Lorsque E est un espace fini, seule la première situation se produit.

Preuve. Puisque la chaîne est irréductible et récurrente, il résulte du Théorème 5.8 que toutes les mesures invariantes sont identiques à une constante multiplicative près : soit elles ont toutes une masse finie, dans lequel cas on peut faire une normalisation appropriée pour obtenir une mesure de probabilité invariante (cas récurrent positif), soit toutes les mesures invariantes ont une masse totale infinie (cas récurrent nul).

Soit $x \in E$. Soit ν_x la mesure invariante introduite dans le Théorème 5.5, i.e.,

$$\nu_x(y) := \mathbb{E}_x\Big(\sum_{i=0}^{T_x-1} \mathbf{1}_{\{X_i=y\}}\Big), \qquad y \in E.$$

Par définition,

$$\nu_x(E) = \sum_{y\in E} \mathbb{E}_x\Big(\sum_{i=0}^{T_x-1} \mathbf{1}_{\{X_i=y\}}\Big) = \mathbb{E}_x\Big(\sum_{i=0}^{T_x-1}\sum_{y\in E} \mathbf{1}_{\{X_i=y\}}\Big) = \mathbb{E}_x\Big(\sum_{i=0}^{T_x-1} 1\Big) = \mathbb{E}_x(T_x).$$

Dans le cas récurrent positif, soit π l'unique probabilité invariante. Il existe alors une constante $c > 0$ telle que $\pi = c\,\nu_x$. Comme $1 = \pi(E) = c\,\nu_x(E)$, on obtient $c = \frac{1}{\nu_x(E)}$, ce qui donne

$$\pi(x) = \frac{\nu_x(x)}{\nu_x(E)} = \frac{1}{\nu_x(E)},$$

qui vaut $\frac{1}{\mathbb{E}_x(T_x)}$. Bien sûr, $\mathbb{E}_x(T_x) = \nu_x(E) < \infty$ dans ce cas.

Dans le cas récurrent nul, ν_x a une masse totale infinie, donc $\mathbb{E}_x(T_x) = \nu_x(E) = \infty$. □

Proposition 5.10. *Supposons que la chaîne est irréductible. S'il existe une probabilité invariante, alors la chaîne est récurrente (donc récurrente positive).*

Preuve. Soit π la probabilité invariante. Soit $y \in E$ tel que $\pi(y) > 0$. Pour tout $x \in E$,

$$\sum_{n=0}^{\infty}[Q^n](x,\, y) = G(x,\, y) \le G(y,\, y),$$

l'inégalité étant une conséquence de la Proposition 4.3. On multiplie par $\pi(x)$ dans les deux côtés de l'inégalité, et somme sur $x \in E$, pour voir que

$$\sum_{n=0}^{\infty}(\pi Q^n)(y) \le \pi(E)\, G(y,\, y) = G(y,\, y).$$

Comme π est invariante, on a $\pi Q^n = \pi$, $\forall n \in \mathbb{N}$, de sorte que

$$\sum_{n=0}^{\infty} \pi(y) \le G(y,\, y).$$

Ceci implique que $G(y,\, y) = \infty$ car $\pi(y) > 0$ par hypothèse. Par conséquent, y est un état récurrent. Puisque la chaîne est irréductible (par hypothèse), elle est récurrente. □

Remarque 5.11. Supposons que la chaîne est irréductible. La Proposition 5.10 nous dit que l'existence d'une probabilité invariante (qui est nécessairement unique) équivaut à la récurrence positive. L'existence d'une mesure invariante de masse totale infinie ne permet pas de conclure. Par exemple, on a vu, dans l'exemple du jeu de pile ou face biaisé, que la chaîne possède une mesure invariante (de masse totale infinie) définie par $\mu(i) := (\frac{p}{q})^i$, $i \in \mathbb{Z}$, mais que la chaîne est récurrente si et seulement si $p = \frac{1}{2}$. □

Exemple 5.12. (Jeu de pile ou face biaisé avec réflexion à l'origine). Fixons un paramètre $0 < p < 1$, et considérons une chaîne de Markov sur $\mathbb{N}$, de matrice de transition

$$\begin{aligned} &p(i,\, i+1) = p, \qquad p(i,\, i-1) = q = 1-p, \qquad \text{si } i \geq 1,\\ &p(0,\, 1) = 1. \end{aligned}$$

La chaîne est évidemment irréductible car tous les états sont connectés avec l'origine. La mesure définie par

$$\mu(i) := (\frac{p}{q})^i, \qquad i \geq 1, \qquad \mu(0) := p,$$

est réversible, donc invariante.

Si $p < \frac{1}{2}$, μ est une mesure finie, donc la chaîne est récurrente positive d'après la Proposition 5.10.

Si $p = \frac{1}{2}$, μ a une masse totale infinie. La chaîne étant irréductible, si l'on peut prouver qu'elle est récurrente, elle sera alors récurrente nulle. Montrons que 0 est un état récurrent, c'est-à-dire $\mathbb{P}_0\{T_0 < \infty\} = 1$, ce qui revient à vérifier que $\mathbb{P}_1\{T_0 < \infty\} = 1$. La valeur de $\mathbb{P}_1\{T_0 < \infty\}$ ne change pas si l'on remplace le jeu de pile ou face biaisé avec réflexion à l'origine par le jeu de pile ou face biaisé sans réflexion à l'origine (la chaîne sera alors sur $\mathbb{Z}$). Pour la chaîne associée au jeu sans réflexion à l'origine, on a vu dans le Théorème 4.9 qu'elle est récurrente (car la loi des sauts est centrée). Donc $\mathbb{P}_1\{T_0 < \infty\} = 1$, et la chaîne est récurrente nulle.

Enfin, si $p > \frac{1}{2}$, de nouveau en considérant la chaîne sans réflexion à l'origine (la chaîne est alors transiente d'après le Théorème 4.9), on sait que $\mathbb{P}_1\{T_0 < \infty\} < 1$: la chaîne est transiente. □

6. Comportement asymptotique

Le but de cette section est d'étudier le comportement asymptotique de $[Q^n](x,\, y)$ lorsque $n \to \infty$ pour tout couple d'états x et y.

Si y est un état transient, on a $G(x, y) \le G(y, y) < \infty$, et donc $[Q^n](x, y) \to 0$, $n \to \infty$.

Qu'est-ce qui se passe avec les états récurrents ? On commence avec un résultat général. Soit ν une mesure de probabilité quelconque sur E.

Théorème 6.1. *Supposons que la chaîne est irréductible et récurrente. Soit μ une mesure invariante. Soient $f : E \to \mathbb{R}_+$ et $g : E \to \mathbb{R}_+$ telles que $0 < \int_E g \, \mathrm{d}\mu < \infty$. Alors*

$$\frac{\sum_{i=0}^n f(X_i)}{\sum_{i=0}^n g(X_i)} \to \frac{\int_E f \, \mathrm{d}\mu}{\int_E g \, \mathrm{d}\mu}, \qquad \mathbb{P}_\nu\text{-p.s.}$$

Corollaire 6.2. *Supposons que la chaîne est irréductible et récurrente positive. Soit π l'unique probabilité invariante. Soit $f : E \to \mathbb{R}_+$. On a*

$$\frac{1}{n} \sum_{i=0}^n f(X_i) \to \int_E f \, \mathrm{d}\pi, \qquad \mathbb{P}_\nu\text{-p.s.}$$

Preuve. Le corollaire étant une conséquence immédiate du théorème (en prenant $g := 1$), seul le théorème nécessite une preuve.

Soit $x \in E$. On définit par récurrence une suite de temps d'arrêt : $T^{(0)} := 0$ et[13]

$$T^{(n)} := \inf\{i > T^{(n-1)} : X_i = x\}, \qquad n = 1, 2, \cdots$$

[Donc $T^{(1)} = T_x$, $\mathbb{P}_x$-p.s.] Comme x est un état récurrent, la suite est bien définie $\mathbb{P}_x$-p.s. Posons

$$\xi_n := \sum_{i=T^{(n)}}^{T^{(n+1)}-1} f(X_i), \qquad n \in \mathbb{N}.$$

À l'aide de la propriété de Markov forte et d'un raisonnement par récurrence, on voit que ξ_n, $n \in \mathbb{N}$, est une suite de variables aléatoires i.i.d.

Soit ν_x la mesure définie dans le Théorème 5.5 :

$$\nu_x(y) := \mathbb{E}_x\Big(\sum_{i=0}^{T_x-1} \mathbf{1}_{\{X_i=y\}} \Big), \qquad y \in E,$$

Puisque la chaîne est irréductible et récurrente, on a $\mu = c\,\nu_x$ (Théorème 5.8), et $c = \mu(x)$ (car $\nu_x(x) = 1$ par définition). Donc

$$\mathbb{E}_x(\xi_0) = \mathbb{E}_x\Big(\sum_{i=0}^{T_x-1} \sum_{y\in E} f(y)\, \mathbf{1}_{\{X_i=y\}} \Big) = \sum_{y\in E} f(y)\, \nu_x(y) = \frac{\int_E f \, \mathrm{d}\mu}{\mu(x)}.$$

[13] D'après l'Exemple 2.3 du Chapitre 2, $(T^{(n)}, n \ge 0)$ est effectivement une suite croissante de temps d'arrêt.

En utilisant la loi des grands nombres forte, on obtient

$$\frac{1}{n}\sum_{i=0}^{n-1}\xi_i \to \frac{\int_E f\,\mathrm{d}\mu}{\mu(x)}, \qquad \mathbb{P}_x\text{-p.s.}$$

Pour tout entier $n \in \mathbb{N}$, soit $N_n(x) := \sum_{i=0}^n \mathbf{1}_{\{X_i=x\}}$, qui représente le nombre de visites à l'état x par la chaîne avant l'instant n ; on a $T^{(N_n(x))} \le n < T^{(N_n(x)+1)}$. Donc

$$\sum_{i=0}^{n} f(X_i) \le \sum_{i=0}^{T^{(N_n(x)+1)}-1} f(X_i) = \sum_{j=0}^{N_n(x)} \xi_j,$$

ce qui entraîne que

$$\limsup_{n\to\infty} \frac{\sum_{i=0}^n f(X_i)}{N_n(x)} \le \mathbb{E}(\xi_0) = \frac{\int_E f\,\mathrm{d}\mu}{\mu(x)}, \qquad \mathbb{P}_x\text{-p.s.}$$

De même, puisque

$$\sum_{i=0}^{n} f(X_i) \ge \sum_{i=0}^{T^{(N_n(x))}} f(X_i) \ge \sum_{i=0}^{T^{(N_n(x))}-1} f(X_i) = \sum_{j=0}^{N_n(x)-1} \xi_j,$$

on a aussi

$$\liminf_{n\to\infty} \frac{\sum_{i=0}^n f(X_i)}{N_n(x)} \ge \mathbb{E}(\xi_0) = \frac{\int_E f\,\mathrm{d}\mu}{\mu(x)}, \qquad \mathbb{P}_x\text{-p.s.}$$

D'où

$$\lim_{n\to\infty} \frac{\sum_{i=0}^n f(X_i)}{N_n(x)} = \frac{\int_E f\,\mathrm{d}\mu}{\mu(x)}, \qquad \mathbb{P}_x\text{-p.s.}$$

L'identité reste valable avec g à la place de f, et l'on prend le rapport pour obtenir convergence presque sûre (sous $\mathbb{P}_x$) pour le rapport. Cette convergence p.s. reste également valable sous $\mathbb{P}_\nu$ par définition de $\mathbb{P}_\nu$. □

Corollaire 6.3. *Supposons que la chaîne est irréductible et récurrente. Alors pour tout $y \in E$,*

- *dans le cas récurrent positif,*

$$\frac{1}{n}\sum_{i=0}^{n-1}\mathbf{1}_{\{X_i=y\}} \to \pi(y), \qquad \mathbb{P}_\nu\text{-p.s.},$$

où π est l'unique probabilité invariante ;

- *dans le cas récurrent nul,*

$$\frac{1}{n}\sum_{i=0}^{n-1}\mathbf{1}_{\{X_i=y\}} \to 0, \qquad \mathbb{P}_\nu\text{-p.s.}$$

Fixons $x \in E$. Puisque $\frac{1}{n}\sum_{i=0}^{n-1} \mathbf{1}_{\{X_i=y\}} \leq 1$, le théorème de convergence dominée nous dit que pour une chaîne irréductible et récurrente, $\frac{1}{n}\sum_{i=0}^{n-1}[Q^i](x, y)$ converge p.s., soit vers $\pi(y)$ (qui est strictement positif) dans le cas récurrent positif, soit vers 0 dans le cas récurrent nul. Le corollaire donne également une explication au sens des expressions telles récurrence positive ou récurrence nulle : la proportion asymptotique du temps passé à un état est positive dans le premier cas, et est nulle dans le second cas.

La convergence de la somme de Cesàro $\frac{1}{n}\sum_{i=0}^{n-1}[Q^i](x, y)$, n'implique cependant pas nécessairement la convergence de $[Q^n](x, y)$. En voici un exemple simple à deux états, avec $E := \{1, 2\}$ et $p(1, 2) = p(2, 1) = 1$ (donc $p(1, 1) = p(2, 2) = 0$). Il convient de formuler sous la forme matricielle :

$$Q := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad Q^2 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

La chaîne est irréductible et récurrente, mais Q^n est la matrice identité si n est pair, et est Q si n est impair : $[Q^n](x, y)$ ne converge pas.

Dans le Théorème 6.7 ci-dessous, on verra que ce phénomène "cyclique" est la seule possibilité pour empêcher $[Q^n](x, y)$ de converger. On introduit d'abord la notion de périodicité.

Définition 6.4. *Soit $x \in E$ un état récurrent. Considérons l'ensemble*

$$I_x := \{n \geq 0 : [Q^n](x, x) > 0\}.$$

Le plus grand commun diviseur de I_x, noté par d_x, est appelé la **période** *de x.*

Remarque 6.5. Il est clair que I_x est stable par addition : si $m, n \in I_x$, alors $m + n \in I_x$ (car $[Q^{m+n}](x, x) \geq [Q^m](x, x)\,[Q^n](x, x) > 0$). Donc I_x est un semi-groupe (l'associativité étant évidente). Le sous-groupe engendré par I_x est $I_x - I_x = d_x\,\mathbb{Z}$. □

Proposition 6.6. *Supposons que la chaîne est irréductible et récurrente.*

(i) *Tous les états ont la même période, appelée la période de la chaîne et notée d.*

(ii) *Si $d = 1$ (la chaîne est alors dite* **apériodique***), alors pour tous $x, y \in E$, il existe $n_0 < \infty$ tel que $[Q^n](x, y) > 0$, $\forall n \geq n_0$.*

Preuve. (i) Soient $x \neq y \in E$. Par l'irréductibilité, il existe des entiers $m \geq 1$ et $\ell \geq 1$ tels que $[Q^m](x, y) > 0$ et $[Q^\ell](y, x) > 0$.

Si $n \in I_x$, alors $[Q^{\ell+n+m}](y, y) \geq [Q^\ell](y, x)\,[Q^n](x, x)\,[Q^m](x, y) > 0$; donc $\ell+n+m \in I_y$. Ceci implique que $I_x - I_x \subset I_y - I_y$, de sorte que d_x est un multiple de d_y (i.e., d_y est diviseur de d_x). En interchangeant les rôles de x et de y, on obtient $d_x = d_y$.

(ii) Puisque la chaîne est irréductible, il existe m tel que $[Q^m](x, y) > 0$; donc tout ce qui reste à traiter, c'est le cas $x = y$.

Montrons d'abord l'existence de $i \in \mathbb{N}$ tel que $[Q^i](y, y) > 0$ et que $[Q^{i+1}](y, y) > 0$.

Soient ℓ et $\ell + k$ deux éléments de I_y. Si $k = 1$, il n'y a rien à prouver. Sinon (i.e., $k \geq 2$), puisque $d_y = 1$, ce qui signifie que le plus grand commun diviseur de I_y est 1, il existe $\ell_1 \in I_y$ qui n'est pas un multiple de k. On écrit $\ell_1 = mk + r$, avec $m \in \mathbb{N}$ et $0 < r < k$. Comme I_y est stable par addition, on a $(m+1)(\ell + k) \in I_y$ (car $\ell + k \in I_y$) et $(m+1)\ell + \ell_1 \in I_y$ (car $\ell \in I_y$ et $\ell_1 \in I_y$). La différence de ces deux éléments de I_y est

$$[(m+1)(\ell + k)] - [(m+1)\ell + \ell_1] = (m+1)k - \ell_1 = k - r < k.$$

En itérant l'argument au plus $k-1$ fois, on arrivera à trouver un couple d'entiers consécutifs i et $i+1$, tous deux éléments de I_y.

Si $i = 0$, la conclusion cherchée est triviale, avec $n_0 = 0$. Si $i \geq 1$, pour tout $k \in \{0, 1, \cdots, i-1\}$, on a

$$[Q^{i^2+k}](y, y) = [Q^{k(i+1)+(i-k)i}](y, y) > 0.$$

Ainsi, en prenant $n_0 := i^2$, on obtient $[Q^n](y, y) > 0$, $\forall n \geq n_0$. □

Théorème 6.7. *Supposons que la chaîne est irréductible, récurrente positive, et apériodique. Pour tout $x \in E$,*

$$\sum_{y \in E} |[Q^n](x, y) - \pi(y)| \to 0, \qquad n \to \infty,$$

où π est l'unique probabilité invariante.

Preuve. Définissons la matrice de transition $\overline{Q}$ sur E^2 par

$$\overline{p}(\mathbf{x}, \mathbf{y}) := p(x_1, y_1)\, p(x_2, y_2), \qquad \mathbf{x} := (x_1, x_2) \in E^2, \ \mathbf{y} := (y_1, y_2) \in E^2,$$

Soit $((X_n^1, X_n^2)_{n \in \mathbb{N}}, (\overline{\mathbb{P}}_{\mathbf{x}})_{\mathbf{x} \in E^2})$ la chaîne de Markov canonique de matrice de transition $\overline{Q}$.

Observons que $\overline{Q}$ est irréductible : si $\mathbf{x}$ et $\mathbf{y}$ sont des états quelconques de $E \times E$, il résulte de la Proposition 6.6 qu'il existe des entiers n_1 et n_2 tels que $[Q^n](x_1, y_1) > 0$, $\forall n \geq n_1$ et que $[Q^n](x_2, y_2) > 0$, $\forall n \geq n_2$; donc pour tout $n \geq \max\{n_1, n_2\}$, $[\overline{Q}^n](\mathbf{x}, \mathbf{y}) > 0$.

Observons également que $\pi \otimes \pi$ est invariante pour $\overline{Q}$:

$$\begin{aligned}
\sum_{\mathbf{x} \in E^2} (\pi \otimes \pi)(\mathbf{x})\, \overline{p}(\mathbf{x}, \mathbf{y}) &= \sum_{\mathbf{x} \in E^2} \pi(x_1)\pi(x_2)\, p(x_1, y_1) p(x_2, y_2) \\
&= \sum_{x_1 \in E} \pi(x_1)\, p(x_1, y_1) \sum_{x_2 \in E} \pi(x_2)\, p(x_2, y_2) \\
&= \pi(y_1)\pi(y_2) = (\pi \otimes \pi)(\mathbf{y}).
\end{aligned}$$

La chaîne étant irréductible et possédant une probabilité invariante, il résulte de la Proposition 5.10 qu'elle est récurrente positive.

La raison pour laquelle on a introduit la nouvelle chaîne est que

$$[Q^n](x,\, y) - \pi(y) = \overline{\mathbb{P}}_{\pi\otimes\delta_x}(X_n^2 = y) - \overline{\mathbb{P}}_{\pi\otimes\delta_x}(X_n^1 = y) = \overline{\mathbb{E}}_{\pi\otimes\delta_x}[\mathbf{1}_{\{X_n^2=y\}} - \mathbf{1}_{\{X_n^1=y\}}].$$

Considérons le temps d'arrêt

$$T := \inf\{n \in \mathbb{N} :\; X_n^1 = X_n^2\}, \qquad \inf\varnothing := \infty.$$

L'identité précédente nous dit que

$$\begin{aligned} [Q^n](x,\, y) - \pi(y) \;&=\; \overline{\mathbb{E}}_{\pi\otimes\delta_x}[\mathbf{1}_{\{T>n\}}(\mathbf{1}_{\{X_n^2=y\}} - \mathbf{1}_{\{X_n^1=y\}})] \\ &\quad + \sum_{k=0}^{n}\sum_{z\in E} \overline{\mathbb{E}}_{\pi\otimes\delta_x}[\mathbf{1}_{\{T=k,\, X_k^1=X_k^2=z\}}(\mathbf{1}_{\{X_n^2=y\}} - \mathbf{1}_{\{X_n^1=y\}})]. \end{aligned} \tag{6.1}$$

Pour $k \in \{0,\, 1,\, \cdots,\, n\}$ et $z \in E$, on a, par la propriété de Markov,

$$\begin{aligned} \overline{\mathbb{E}}_{\pi\otimes\delta_x}[\mathbf{1}_{\{T=k,\, X_k^1=X_k^2=z\}}\mathbf{1}_{\{X_n^2=y\}}] \;&=\; \overline{\mathbb{E}}_{\pi\otimes\delta_x}[\mathbf{1}_{\{T=k,\, X_k^1=X_k^2=z\}}]\,[Q^{n-k}](z,\, y) \\ &=\; \overline{\mathbb{E}}_{\pi\otimes\delta_x}[\mathbf{1}_{\{T=k,\, X_k^1=X_k^2=z\}}\mathbf{1}_{\{X_n^1=y\}}], \end{aligned}$$

ce qui signifie que la double somme à droite de (6.1) s'annule. Donc

$$\begin{aligned} \sum_{y\in E} |[Q^n](x,\, y) - \pi(y)| \;&=\; \sum_{y\in E} |\overline{\mathbb{E}}_{\pi\otimes\delta_x}[\mathbf{1}_{\{T>n\}}(\mathbf{1}_{\{X_n^2=y\}} - \mathbf{1}_{\{X_n^1=y\}})]| \\ &\le\; \sum_{y\in E} \overline{\mathbb{E}}_{\pi\otimes\delta_x}[\mathbf{1}_{\{T>n\}}(\mathbf{1}_{\{X_n^2=y\}} + \mathbf{1}_{\{X_n^1=y\}})] \\ &=\; 2\,\overline{\mathbb{P}}_{\pi\otimes\delta_x}(T > n), \end{aligned}$$

qui tend vers 0 lorsque $n \to \infty$, car la nouvelle chaîne est récurrente. □

7. Martingales et chaînes de Markov

On continue à travailler avec la chaîne canonique. Pour toute fonction $f :\, E \to \mathbb{R}_+$, on note

$$[Qf](x) := \sum_{y\in E} p(x,\, y) f(y) = \mathbb{E}_x[f(X_1)], \qquad x \in E.$$

Définition 7.1. *Une fonction* $f : E \to \mathbb{R}_+$ *est harmonique* (*resp. surharmonique*) *si pour tout* $x \in E$,

$$f(x) = [Qf](x) \qquad (\textit{resp. } f(x) \ge [Qf](x)).$$

Plus généralement, si $A \subset E$ *est non vide, on dit que* f *est harmonique* (*resp. surharmonique*) *sur* A *si, pour tout* $x \in A$,

$$f(x) = [Qf](x) \qquad (\textit{resp. } f(x) \ge [Qf](x)).$$

Remarque 7.2. Plus généralement, on peut considérer les fonctions harmoniques ou surharmoniques de signe quelconque. □

Proposition 7.3. *Soit* $f : E \to \mathbb{R}_+$.

(i) *Une condition nécessaire et suffisante pour que* f *soit harmonique* (*resp. surharmonique*) *est : pour tout* $x \in E$, $(f(X_n),\, n \ge 0)$ *est une martingale* (*resp. surmartingale*) *sous* $\mathbb{P}_x$ *par rapport à la filtration canonique* $(\mathscr{F}_n)$.

(ii) *Soit* $A \subset E$ *non vide, et soit*

$$\tau_{A^c} := \inf\{n \in \mathbb{N} :\, X_n \notin A\}, \qquad \inf \varnothing := \infty.$$

Si f *est harmonique* (*resp. surharmonique*) *sur* A, *alors pour tout* $x \in A$, $(f(X_{n\wedge \tau_{A^c}}),\, n \ge 0)$ *est une martingale* (*resp. surmartingale*) *sous* $\mathbb{P}_x$.

Preuve. On écrit la preuve seulement pour les fonctions harmoniques, la preuve pour les fonctions surharmoniques étant tout à fait similaire.

(i) "$\Rightarrow$" Soit $x \in E$. On a $\mathbb{E}_x[f(X_1)] = f(x) < \infty$, et par récurrence en n, $\mathbb{E}_x[f(X_n)] < \infty$, $\forall n \in \mathbb{N}$. De plus, $\mathbb{E}_x[f(X_{n+1}) \,|\, \mathscr{F}_n] = \sum_{y\in E} f(y)\, p(X_n,\, y) = f(X_n)$, donc $(f(X_n),\, n \ge 0)$ est une $(\mathscr{F}_n)$-martingale sous $\mathbb{P}_x$.

"$\Leftarrow$" Soit $x \in E$. La propriété de martingale de $(f(X_n),\, n \ge 0)$ implique que $f(x) = \mathbb{E}_x[f(X_1)] = \sum_{y\in E} f(y)\, p(x,\, y)$, i.e., f est harmonique.

(ii) Soit $x \in A$. On écrit

$$f(X_{(n+1)\wedge \tau_{A^c}}) = f(X_{n+1})\, \mathbf{1}_{\{\tau_{A^c} > n\}} + f(X_{\tau_{A^c}})\, \mathbf{1}_{\{\tau_{A^c} \le n\}},$$

et calcule espérance conditionnelle (sachant $\mathscr{F}_n$) pour chacun des deux termes à droite.

Pour le premier terme : comme $\{\tau_{A^c} > n\} \in \mathscr{F}_n$, on a $\mathbb{E}_x[f(X_{n+1})\, \mathbf{1}_{\{\tau_{A^c} > n\}} \,|\, \mathscr{F}_n] = \mathbf{1}_{\{\tau_{A^c} > n\}}\, \mathbb{E}_x[f(X_{n+1}) \,|\, \mathscr{F}_n] = \mathbf{1}_{\{\tau_{A^c} > n\}} \sum_{y\in E} f(y) p(X_n,\, y)$, ce qui est $\mathbf{1}_{\{\tau_{A^c} > n\}}\, f(X_n)$ car f est harmonique sur A.

Pour le second terme : on a $f(X_{\tau_{A^c}})\,\mathbf{1}_{\{\tau_{A^c}\le n\}} = f(X_{\tau_{A^c}\wedge n})\,\mathbf{1}_{\{\tau_{A^c}\le n\}}$ qui est $\mathscr{F}_n$-mesurable ; ainsi $\mathbb{E}_x[f(X_{\tau_{A^c}})\,\mathbf{1}_{\{\tau_{A^c}\le n\}} \,|\, \mathscr{F}_n] = f(X_{\tau_{A^c}})\,\mathbf{1}_{\{\tau_{A^c}\le n\}}$.

Par conséquent,

$$\mathbb{E}_x[f(X_{(n+1)\wedge\tau_{A^c}}) \,|\, \mathscr{F}_n] = \mathbf{1}_{\{\tau_{A^c}>n\}}\, f(X_n) + f(X_{\tau_{A^c}})\,\mathbf{1}_{\{\tau_{A^c}\le n\}} = f(X_{n\wedge\tau_{A^c}}).$$

En particulier, on voit par récurrence que $\mathbb{E}_x[f(X_{n\wedge\tau_{A^c}})] = f(x) < \infty$. Conclusion : $(f(X_{n\wedge\tau_{A^c}}),\, n\ge 0)$ est une martingale. □

Théorème 7.4. *Soit $A \subsetneqq E$ non vide. Soit $g:\, A^c \to \mathbb{R}_+$ une fonction bornée. Posons*

$$h(x) := \mathbb{E}_x[g(X_{\tau_{A^c}})\,\mathbf{1}_{\{\tau_{A^c}<\infty\}}], \qquad x\in E.$$

(i) *La fonction h est harmonique sur A.*

(ii) *Supposons que $\mathbb{P}_x\{\tau_{A^c}<\infty\}=1$, $\forall x\in A$. Alors h est l'unique fonction bornée sur E qui soit harmonique sur A et qui coïncide avec g sur A^c.*

Preuve. (i) Soit $x\in A$. Par définition, $\mathbb{P}_x\{\tau_{A^c} = 1 + (\tau_{A^c}\circ\theta_1)\} = 1$, et donc[14]

$$g(X_{\tau_{A^c}})\,\mathbf{1}_{\{\tau_{A^c}<\infty\}} = [g(X_{\tau_{A^c}})\,\mathbf{1}_{\{\tau_{A^c}<\infty\}}]\circ\theta_1, \qquad \mathbb{P}_x\text{-p.s.}$$

Ainsi,

$$\begin{aligned} h(x) &= \mathbb{E}_x\{[g(X_{\tau_{A^c}})\,\mathbf{1}_{\{\tau_{A^c}<\infty\}}]\circ\theta_1\} \\ &= \mathbb{E}_x\{\mathbb{E}_{X_1}[g(X_{\tau_{A^c}})\,\mathbf{1}_{\{\tau_{A^c}<\infty\}}]\} \qquad \text{(propriété de Markov)} \\ &= \mathbb{E}_x\{h(X_1)\} \qquad \text{(définition de } h) \\ &= \sum_{y\in E} h(y)\,p(x,\,y). \end{aligned}$$

Ceci montre que h est harmonique sur A.

(ii) (Existence) On sait déjà que h est harmonique sur A. Si $x\in A^c$, alors $\tau_{A^c}=0$ $\mathbb{P}_x$-p.s., donc $h(x)=g(x)$. La bornitude de h provient de celle de g.

(Unicité) Soit $\widetilde{h}$ une autre fonction satisfaisant les conditions du théorème. Pour $x\in A^c$, on a $\widetilde{h}(x)=g(x)=h(x)$. Pour $x\in A$, comme $Y_n := \widetilde{h}(X_{n\wedge\tau_{A^c}})$ est une martingale sous $\mathbb{P}_x$ (d'après la Proposition 7.3), qui est bornée, elle converge $\mathbb{P}_x$-p.s. et dans L^1, vers $Y_\infty := \widetilde{h}(X_{\tau_{A^c}}) = g(X_{\tau_{A^c}})$.[15] Par conséquent,

$$\widetilde{h}(x) = \mathbb{E}_x(Y_0) = \mathbb{E}_x(Y_\infty) = \mathbb{E}_x[g(X_{\tau_{A^c}})],$$

[14] Plus généralement, si S et T sont des temps d'arrêt, alors $U := [T + (S\circ\theta_T)]\,\mathbf{1}_{\{T<\infty\}} + \infty\times\mathbf{1}_{\{T=\infty\}}$ est un temps d'arrêt, et $X_U = X_S\circ\theta_T$ sur $\{U<\infty\}$.

[15] Rappelons que par hypothèse, $\tau_{A^c}<\infty$, $\mathbb{P}_x$-p.s.

ce qui, par définition, vaut $h(x)$. □

Exemple 7.5. (Problème de Dirichlet discret). Soit $A \subset \mathbb{Z}^d$ un sous-ensemble fini non vide. La frontière de A est définie par[16]

$$\partial A := \{y \in \mathbb{Z}^d \backslash A : \ |y - x| = 1 \text{ pour un } x \in A\}.$$

On écrit $\overline{A} := A \cup \partial A$.

Une fonction h définie $\overline{A}$ est dite **harmonique** sur A (au sens discret) si pour tout $x \in A$, $h(x)$ est la moyenne des valeurs de h sur les $(2d)$ sites voisins de x. Cette définition coïncide avec celle de la Définition 7.1 si l'on prend le noyau de transition de la marche aléatoire simple sur $\mathbb{Z}^d$: $p(x, x + e_i) = p(x, x - e_i) = \frac{1}{2d}$, pour $i = 1, \cdots, d$, où $e_1, \cdots, e_d$ sont les vecteurs unité de $\mathbb{Z}^d$.

D'après le Théorème 7.4, toute fonction positive[17]g définie sur ∂A peut être étendue, d'une façon unique, à une fonction positive sur $\overline{A}$ qui soit harmonique sur A : cette fonction est donnée par

$$h(x) := \mathbb{E}_x[g(X_{\tau_{\partial A}})], \qquad x \in A,$$

où

$$\tau_{\partial A} := \inf\{n \in \mathbb{N} : \ X_n \in \partial A\}.$$

En appliquant le Théorème 7.4, on a, a priori, besoin de définir g sur $\mathbb{Z}^d \backslash A$ et non seulement sur ∂A ; cependant, les valeurs de g sur $\mathbb{Z}^d \backslash \overline{A}$ n'ont aucune influence sur les valeurs de h sur A.

Application en physique : On met le voltage 1 à ∂A et le voltage 0 à un point $x_0 \in A$ fixé : quel est le voltage à un point $x \in A$ générique ?

Réponse : le voltage $h(x)$ au point $x \in A$ est donné par

$$h(x) := \mathbb{P}_x\{\tau_{\partial A} < \tau_{x_0}\}.$$

Ceci correspond à la fonction harmonique sur A associée à $g = 1$ sur ∂A. □

[16]Notation : "$|\cdot|$" désigne la norme L^1 sur $\mathbb{Z}^d$.

[17]Comme A est un ensemble fini, la bornitude de g est automatique.

Appendice A

Rappels de théorie de l'intégration

On rappelle *quelques* notions et résultats de base de la théorie de l'intégration. Les notations étant celles usuelles de la théorie de la mesure, elles se diffèrent légèrement des notations probabilistes ; par exemple, l'espace mesurable générique est noté $(E, \mathscr{A})$ ici, plutôt que $(\Omega, \mathscr{F})$ en théorie des probabilités.

1. Espaces mesurés

Soit E un ensemble non vide. Une **tribu** (= σ-algèbre = σ-algèbre de Boole) $\mathcal{A}$ sur E est une famille de parties de E telle que :

(i) $E \in \mathscr{A}$;

(ii) $A \in \mathscr{A} \Rightarrow A^c \in \mathscr{A}$;

(iii) $A_n \in \mathscr{A}$, $\forall n \geq 1 \Rightarrow \cup_{n\geq 1} A_n \in \mathscr{A}$.

On appelle $(E, \mathscr{A})$ un **espace mesurable**. Les éléments de $\mathscr{A}$ sont appelés parties mesurables.

Définition 1.1. *Soit $\mathscr{C}$ une famille de parties de E non vide. Il existe une plus petite tribu sur E qui contienne $\mathscr{C}$, notée $\sigma(\mathscr{C})$. On a alors*

$$\sigma(\mathscr{C}) = \bigcap_{\mathscr{A} \text{ tribu}, \, \mathscr{C} \subset \mathscr{A}} \mathscr{A}.$$

On appelle $\sigma(\mathscr{C})$ la tribu engendrée par $\mathscr{C}$.

Une **mesure** (positive) sur l'espace mesurable $(E, \mathscr{A})$ est une application $\mu : \mathscr{A} \to [0, \infty]$ telle que :

(i) $\mu(\varnothing) = 0$;

(ii) **(σ-additivité)** pour toute suite $(A_n)_{n\geq 1}$ de parties mesurables *deux-à-deux disjointes*, on a $\mu(\cup_{n\geq 1} A_n) = \sum_{n\geq 1} \mu(A_n)$.
On appelle $(E, \mathscr{A}, \mu)$ un **espace mesuré**.

Terminologie. On dit que μ est une mesure finie si $\mu(E) < \infty$ (dans ce cas, $\mu(E)$ est appelée la masse totale de μ), et qu'elle est σ-finie s'il existe une suite (A_n) de parties mesurables telle que $E = \cup_{n\geq 1} A_n$, et que $\mu(A_n) < \infty$, $\forall n$. □

On verifie assez facilement que pour des suites de parties mesurables $(A_n,\, n \geq 1)$ et $(B_n,\, n \geq 1)$:[1]

- $(A_n) \uparrow \;\Rightarrow\; \mu(\cup_{n\geq 1} A_n) = \lim_{n\to\infty} \mu(A_n)$.
- $(B_n) \downarrow$, $\exists n_0\ \mu(B_{n_0}) < \infty \;\Rightarrow\; \mu(\cap_{n\geq 1} B_n) = \lim_{n\to\infty} \mu(B_n)$.

Exemple 1.2. (Tribu borélienne réelle). Lorsque $E = \mathbb{R}$ (plus généralement, lorsque E est un espace topologique quelconque), on lui associe souvent la tribu borélienne $\mathscr{B}(E)$ qui est celle engendrée par la classe des ouverts de E. Autrement dit, c'est la plus petite tribu associée à E qui contienne tous les ouverts de E.

On vérifie assez facilement que $\mathscr{B}(\mathbb{R})$ est également engendrée par les intervalles $]a, b[$ (avec a, $b \in \mathbb{R}$, et $a < b$), ou par les intervalles $]-\infty, a[$ (avec $a \in \mathbb{R}$), ou encore par les intervalles $]-\infty, a[$ (avec $a \in \mathbb{Q}$), les intervalles ouverts pouvant être remplacés par les intervalles fermés.

On vérifie aisément que $\mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R}) = \mathscr{B}(\mathbb{R}^2)$.

À chaque fois que l'on parle de l'espace $\mathbb{R}$, sauf mention du contraire, on lui associe la tribu borélienne $\mathscr{B}(\mathbb{R})$. □

Soient $(E, \mathscr{A})$, $(\widetilde{E}, \widetilde{\mathscr{E}})$, $(\widetilde{E}_1, \widetilde{\mathscr{E}_1})$ et $(\widetilde{E}_2, \widetilde{\mathscr{E}_2})$ des espaces mesurables.

Une **application** $f : E \mapsto \widetilde{E}$ est **mesurable** si

$$\forall B \in \widetilde{\mathscr{E}}, \quad f^{-1}(B) \in \mathscr{A}.$$

Propriétés :

- La composition de deux applications mesurables est mesurable.
- Soit $f : (E, \mathscr{A}) \mapsto (\widetilde{E}, \widetilde{\mathscr{E}})$. Pour que f soit mesurable, il suffit qu'il existe une sous-classe $\mathscr{C} \subset \widetilde{\mathscr{E}}$ avec $\sigma(\mathscr{C}) = \widetilde{\mathscr{E}}$, telle que $f^{-1}(B) \in \mathscr{A}$, $\forall B \in \mathscr{C}$.

[1]Notations : $(A_n) \uparrow$ signifie que $A_n \subset A_{n+1}$, $\forall n$; de même, $(B_n) \downarrow$ signifie que $B_n \supset B_{n+1}$, $\forall n$. Observons que la condition $\exists n_0\ \mu(B_{n_0}) < \infty$ est automatiquement satisfaite lorsque μ est une mesure finie (en particulier, lorsque μ est une mesure de probabilité).

• Soient $g_1 : (E, \mathscr{A}) \to (\widetilde{E}_1, \widetilde{\mathscr{E}}_1)$ et $g_2 : (E, \mathscr{A}) \to (\widetilde{E}_2, \widetilde{\mathscr{E}}_2)$ deux applications. L'application $g : (E, \mathscr{A}) \to (\widetilde{E}_1 \times \widetilde{E}_2, \widetilde{\mathscr{E}}_1 \otimes \widetilde{\mathscr{E}}_2)$ définie par $g(x) := (g_1(x), g_2(x))$ est mesurable si et seulement si g_1 et g_2 sont mesurables.[2]

Ces proporiétés élémentaires ont des conséquences intéressantes :

• Soit $f : (E, \mathscr{A}) \mapsto (\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Pour montrer la mesurabilité d'une application f, il suffit que les ensembles $f^{-1}(\,]a, b[\,)$, ou même $f^{-1}(\,]-\infty, a[\,)$, soient mesurables.

• Si E et $\widetilde{E}$ sont des espaces topologiques munis de leurs tribus boréliennes, toute application continue est mesurable.

• Si $f, g : (E, \mathscr{A}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ sont des fonctions mesurables, alors les fonctions $f+g$, fg, $\max\{f, g\}$, $\min\{f, g\}$, $f^+ := \max\{f, 0\}$, $f^- = \max\{-f, 0\}$ sont également mesurables.

• Si (f_n) est une suite de fonctions mesurables de E dans $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$, alors $\sup_{n\geq 1} f_n$, $\inf_{n\geq 1} f_n$, $\limsup_{n\to\infty} f_n$, $\liminf_{n\to\infty} f_n$ sont également mesurables. [En particulier, si la suite (f_n) converge simplement, sa limite $\lim_{n\to\infty} f_n$ est mesurable.][3]

Définition 1.3. *Soit $f : (E, \mathscr{A}) \to (\widetilde{E}, \widetilde{\mathscr{E}})$ une application mesurable, et soit μ une mesure sur $(E, \mathscr{A})$. La* **mesure image** *de μ par f, est la mesure sur $(\widetilde{E}, \widetilde{\mathscr{E}})$ qui donne valeur $\mu(f^{-1}(B))$ à tout $B \in \widetilde{\mathscr{E}}$.*

2. Intégration par rapport à une mesure

2.1. Intégration des fonctions positives

Soit $(E, \mathscr{A}, \mu)$ un espace mesuré.

Une fonction mesurable $f : E \to \mathbb{R}$ est dite **fonction étagée** si elle s'écrit

$$f(x) = \sum_{i=1}^{n} \alpha_i \, \mathbf{1}_{A_i}(x), \qquad x \in E\,,$$

où $n \geq 1$, $\alpha_1 < \alpha_2 < \cdots < \alpha_n$, et pour tout i, $A_i := f^{-1}(\{\alpha_i\}) \in \mathscr{A}$.

[2] Pour la mesure produit $\widetilde{\mathscr{E}}_1 \otimes \widetilde{\mathscr{E}}_2$, voir les rappels en Section 3.

[3] Rappelons que si (a_n) est une suite d'éléments de $\overline{\mathbb{R}}$, on définit $\limsup_{n\to\infty} a_n := \lim_{n\to\infty} \downarrow (\sup_{k\geq n} a_k)$ et $\liminf_{n\to\infty} a_n := \lim_{n\to\infty} \uparrow (\inf_{k\geq n} a_k)$, les limites existant dans $\overline{\mathbb{R}}$. [Autrement dit, $\limsup_{n\to\infty} a_n$ et $\liminf_{n\to\infty} a_n$ représentent, respectivement, la plus grande et la plus petite valeur d'adhérence de la suite (a_n).]

Définition 2.1. *Soit f une fonction étagée à valeurs dans $\mathbb{R}_+$. L'intégrale de f par rapport à μ est définie par*

$$\int f \,\mathrm{d}\mu = \int_E f \,\mathrm{d}\mu := \sum_{i=1}^n \alpha_i \,\mu(A_i) \in [0,\, \infty],$$

avec la convention $0 \times \infty := 0$ si $\alpha_i = 0$ et $\mu(A_i) = \infty$.

Cas particulier : si $f := \mathbf{1}_A$ avec $A \in \mathscr{A}$, alors $\int \mathbf{1}_A \,\mathrm{d}\mu = \mu(A)$.

Propriété 2.2. *Soient f et g des fonctions étagées positives.*

(i) $\int (af + bg) \,\mathrm{d}\mu = a \int f \,\mathrm{d}\mu + b \int g \,\mathrm{d}\mu$, $\forall a,\, b \geq 0$.

(ii) $f \leq g \;\Rightarrow\; \int f \,\mathrm{d}\mu \leq \int g \,\mathrm{d}\mu$.

Définition 2.3. *Soit $f :\, E \to [0,\, \infty]$ une fonction mesurable. On pose*

$$\int f \,\mathrm{d}\mu := \sup \Big\{ \int g \,\mathrm{d}\mu :\; g \leq f, \quad g \text{ étagée positive} \Big\}.$$

Si f est une fonction étagée positive, cette définition de $\int f \,\mathrm{d}\mu$ coïncide avec celle de la Définition 2.1 (par l'assertion (ii) de la Propiété 2.2).

On notera indifféremment $\int f \,\mathrm{d}\mu$, $\int f(x)\, \mu(\mathrm{d}x)$ ou $\int f(x) \,\mathrm{d}\mu(x)$.

Attention : désormais, "fonction mesurable positive" signifie fonction mesurable à valeurs dans $[0,\, \infty]$ (tandis que fonction étagée positive signifie, par définition, fonction étagée à valeurs dans $\mathbb{R}_+$).

Propriété 2.4. *Soient f et g des fonctions mesurables positives. Si $f \leq g$, alors $\int f \,\mathrm{d}\mu \leq \int g \,\mathrm{d}\mu$.*

Théorème 2.5. (Théorème de convergence monotone de Beppo Levi) *Soient (f_n) une suite de fonctions mesurables positives, et soit $f := \lim_{n\to\infty} \uparrow f_n$. Alors*

$$\int f \,\mathrm{d}\mu = \lim_{n\to\infty} \uparrow \int f_n \,\mathrm{d}\mu.$$

Preuve. Comme $f_n \leq f$, $\forall n$, on a $\int f \,\mathrm{d}\mu \geq \lim_{n\to\infty} \uparrow \int f_n \,\mathrm{d}\mu$. Pour l'inégalité dans l'autre sens, soit $g \leq f$ une fonction étagée positive : $g := \sum_{i=1}^m \alpha_i \,\mathbf{1}_{A_i}$. Soit $0 \leq a < 1$, et soit

$$B_n := \{x \in E :\, ag(x) \leq f_n(x)\}.$$

Il est clair que (B_n) est une suite croissante de parties mesurables, et $\cup_{n\geq 1} B_n = E$ (car, si x est tel que $g(x) > 0$, alors $f(x) > 0$, et comme $a < 1$ tandis que $f_n(x) \uparrow f(x)$, on a $ag(x) \leq f_n(x)$ pour tout n suffisamment grand ; si $g(x) = 0$, on a $ag(x) \leq f_n(x)$, $\forall n$).

Pour tout n, on a $ag\,\mathbf{1}_{B_n} \le f_n$, et donc

$$\int f_n \,\mathrm{d}\mu \ge \int ag\,\mathbf{1}_{B_n} \,\mathrm{d}\mu = a \sum_{i=1}^{m} \alpha_i \,\mu(A_i \cap B_n).$$

On fait maintenant $n \to \infty$. À droite, comme $B_n \uparrow$ avec $\cup_{n=1}^{\infty} B_n = E$, on a $\mu(A_i \cap B_n) \uparrow \mu(A_i)$, $\forall i$, et donc $\sum_{i=1}^{m} \alpha_i \mu(A_i \cap B_n) \uparrow \sum_{i=1}^{m} \alpha_i \mu(A_i) = \int g \,\mathrm{d}\mu$. Par conséquent, $\lim_{n\to\infty} \uparrow \int f_n \,\mathrm{d}\mu \ge a \int g \,\mathrm{d}\mu$.

Comme a peut être aussi proche de 1 que possible, on obtient $\lim_{n\to\infty} \uparrow \int f_n \,\mathrm{d}\mu \ge \int g \,\mathrm{d}\mu$. La fonction $g \le f$ étant étagée positive quelconque, on a donc $\lim_{n\to\infty} \uparrow \int f_n \,\mathrm{d}\mu \ge \int f \,\mathrm{d}\mu$.□

Théorème 2.6. *Si f_n, $n \ge 1$, sont des fonctions mesurables positives, on a $\int \sum_{n\ge1} f_n \,\mathrm{d}\mu = \sum_{n\ge1} \int f_n \,\mathrm{d}\mu$.*

Preuve. It suffit d'appliquer le théorème de Beppo Levi à la suite $(\sum_{i=1}^{n} f_i,\ n \ge 1)$. □

Théorème 2.7. *Soit f une fonction mesurable positive. Il existe une suite croissante (f_n) de fonctions étagées positives telle que $f_n \uparrow f$. [En conséquence, $\int f_n \,\mathrm{d}\mu \uparrow \int f \,\mathrm{d}\mu$, ce qui est en accord avec la Définition 2.3.]*

Preuve. Pour tout $n \ge 1$, on définit la fonction étagée $f_n := \min\{\frac{\lfloor 2^n f \rfloor}{2^n}, n\}$. Alors $f_n \uparrow f$. □

Terminologie. On dit qu'une propriété est vraie μ-presque partout, ou μ-p.p., voire simplement p.p. s'il n'y a pas d'ambiguïté, si elle est valable sauf sur un semble de mesure nulle. Par exemple, pour un couple de fonctions mesurables f et g, $f = g$ μ-p.p. signifie $\mu(\{x \in E : f(x) \neq g(x)\}) = 0$.

Propriété 2.8. *Soit f une fonction mesurable positive.*

(i) $\int f \,\mathrm{d}\mu < \infty \;\Rightarrow\; f < \infty$ p.p.

(ii) $\int f \,\mathrm{d}\mu = 0 \;\Leftrightarrow\; f = 0$ p.p.

Preuve. Montrons d'abord l'inégalité de Markov : $\mu(\{x \in E : f(x) \ge a\}) \le \frac{1}{a} \int f \,\mathrm{d}\mu$, $\forall a > 0$.

Soit $A := \{x \in E : f(x) \ge a\}$. On a $f \ge a\,\mathbf{1}_A$. Donc $\int f \,\mathrm{d}\mu \ge \int a\,\mathbf{1}_A \,\mathrm{d}\mu = a\,\mu(A)$.

[Même si l'inégalité est toujours valable, elle n'est intéressante que si $\int f \,\mathrm{d}\mu < \infty$.]

(i) Soit $A_n := \{x \in E : f(x) \ge n\}$. Alors (A_n) est une suite décroissante de parties mesurables, et $\cap_{n\ge1} A_n = \{x \in E : f(x) = \infty\}$. Comme $\int f \,\mathrm{d}\mu < \infty$, on a, d'après (i),

$\mu(A_1) < \infty$. Donc $\mu(\{x \in E : f(x) = \infty\}) = \lim_{n\to\infty} \downarrow \mu(A_n)$. Or, d'après l'inégalité de Markov, $\mu(A_n) \le \frac{1}{n} \int f \,\mathrm{d}\mu \to 0$; d'où $\mu(\{x \in E : f(x) = \infty\}) = 0$.

[On peut également prouver cette propriété en utilisant un argument par l'absurde.]

(ii) "$\Leftarrow$" Trivial par définition.

"$\Rightarrow$" Soit $B_n := \{x \in E : f(x) \ge \frac{1}{n}\}$. Alors d'après (i), $\mu(B_n) \le n \int f \,\mathrm{d}\mu = 0$; donc $\mu(B_n) = 0$, $\forall n$. Or, (B_n) étant une suite croissante d'ensembles mesurables, on a $\mu(\cup_{n\ge 1} B_n) = \lim \uparrow \mu(B_n) = 0$. Il suffit alors de remarquer que $\cup_{n\ge 1} B_n = \{x \in E : f(x) > 0\}$. $\square$

Théorème 2.9. (Lemme de Fatou). *Soit (f_n) une suite de fonctions mesurables positives. Alors*

$$\int (\liminf_{n\to\infty} f_n) \,\mathrm{d}\mu \le \liminf_{n\to\infty} \int f_n \,\mathrm{d}\mu.$$

Preuve. [L'intégrale $\int (\liminf_{n\to\infty} f_n) \,\mathrm{d}\mu$ est bien définie, car $\liminf_{n\to\infty} f_n$ est une fonction mesurable, et évidemment positive.]

Soit $g_n := \inf_{m\ge n} f_m$. Par définition, $\liminf_{n\to\infty} f_n = \lim_{n\to\infty} \uparrow g_n$, et donc par le théorème de convergence monotone,

$$\int (\liminf_{n\to\infty} f_n) \,\mathrm{d}\mu = \lim_{n\to\infty} \uparrow \int g_n \,\mathrm{d}\mu.$$

Or, pour tout $m \ge n$, on a $g_n \le f_m$ et donc $\int g_n \,\mathrm{d}\mu \le \int f_m \,\mathrm{d}\mu$, avec $m \ge n$ quelconque ; d'où $\int g_n \,\mathrm{d}\mu \le \inf_{m\ge n} \int f_m \,\mathrm{d}\mu$. On fait $n \to \infty$ dans les deux côtés. À gauche, on a vu que cela vaut $\int (\liminf_{n\to\infty} f_n) \,\mathrm{d}\mu$. À droite, cela vaut $\liminf_{n\to\infty} \int f_n \,\mathrm{d}\mu$. $\square$

2.2. Fonctions intégrables

Soit $(E, \mathscr{A}, \mu)$ un espace mesuré.

Définition 2.10. *Soit $f : E \to [-\infty, \infty]$ une fonction mesurable. On dit que f est intégrable par rapport à μ (ou : μ-intégrable) si*

$$\int |f| \,\mathrm{d}\mu < \infty.$$

Dans ce cas, on pose

$$\int f \,\mathrm{d}\mu := \int f^+ \,\mathrm{d}\mu - \int f^- \,\mathrm{d}\mu \in \mathbb{R}.$$

On voit que f est intégrable si et seulement si $\int f^+ \,\mathrm{d}\mu < \infty$ et $\int f^- \,\mathrm{d}\mu < \infty$. Dans le cas où f est positive, la définition de $\int f \,\mathrm{d}\mu$ coïncide avec celle précédente.

On note $\mathscr{L}^1(E, \mathscr{A}, \mu)$ l'espace des fonctions μ-intégrables, voire $\mathscr{L}^1$ s'il n'y a pas de risque d'ambiguïté.

Propriété 2.11. *Soient $f, g \in \mathscr{L}^1(E, \mathscr{A}, \mu)$.*

(i) $|\int f \,\mathrm{d}\mu| \le \int |f| \,\mathrm{d}\mu$.

(ii) $\int (af + bg) \,\mathrm{d}\mu = a \int f \,\mathrm{d}\mu + b \int g \,\mathrm{d}\mu$, $\forall a, b \in \mathbb{R}$.

(iii) $f \le g \Rightarrow \int f \,\mathrm{d}\mu \le \int g \,\mathrm{d}\mu$.

(iv) *h fonction mesurable telle que $f = h$ p.p. $\Rightarrow h \in \mathscr{L}^1(E, \mathscr{A}, \mu)$ et $\int h \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu$.*

Théorème 2.12. *Soit $\Phi : (E, \mathscr{A}) \to (F, \mathscr{B})$ une application mesurable, et soit ν la mesure image de μ par Φ. Alors pour toute fonction mesurable $f : F \to [-\infty, \infty]$,*

$$\int_E (f \circ \Phi) \,\mathrm{d}\mu = \int_F f \,\mathrm{d}\nu,$$

si au moins l'une des deux intégrales est bien définie.

Preuve. Si $f = \mathbf{1}_B$ avec $B \in \mathscr{B}$, l'identité est trivial, car

$$\int f \,\mathrm{d}\nu = \nu(B) = \mu(\Phi^{-1}(A)) = \int_E (\mathbf{1}_B \circ \Phi) \,\mathrm{d}\mu.$$

Par linéarité de l'intégrale, l'identité reste valable pour toute fonction étagée f, et donc encore valable si f est une fonction mesurable positive (il suffit d'utiliser le fait que f est la limite croissante d'une suite de fonctions étagées positives et appliquer le théorème de convergence monotone). Enfin, si f est de signe quelconque, on peut considérer séparément f^+ et f^-. $\square$

Théorème 2.13. (Continuité de l'intégrale par rapport à la mesure). *Soit $f \in \mathscr{L}^1(E, \mathscr{A}, \mu)$. Pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que pour tout $A \in \mathscr{A}$ avec $\mu(A) < \delta$, on a $\int_A |f| \,\mathrm{d}\mu < \varepsilon$.*

Preuve. Montrons d'abord qu'il existe $B \in \mathscr{A}$ avec $\mu(B) < \infty$, tel que $|f| \,\mathbf{1}_B$ est bornée, et que $\int_{B^c} |f| \,\mathrm{d}\mu < \varepsilon$.

Soient $B_n := \{\frac{1}{n} \le |f| \le n\}$, $n \ge 1$. Comme $|f| \,\mathbf{1}_{B_n} \uparrow |f| \,\mathbf{1}_{\{|f|<\infty\}}$ (quand $n \to \infty$), il résulte du théorème de convergence monotone que $\lim_{n\to\infty} \uparrow \int_{B_n} |f| \,\mathrm{d}\mu = \int_{\{|f|<\infty\}} |f| \,\mathrm{d}\mu = \int |f| \,\mathrm{d}\mu < \infty$ (la dernière égalité étant due au fait que $|f| < \infty$, p.p.). Donc $\int_{B_n^c} |f| \,\mathrm{d}\mu \to 0$, alors que pour chaque n, $\mu(B_n) \le \mu(\{|f| \ge \frac{1}{n}\}) \le n \int |f| \, \mathrm{d}\mu < \infty$, et $|f| \, \mathbf{1}_{B_n}$ est bornée (par n).

On a donc trouvé $B \in \mathscr{A}$ avec $\mu(B) < \infty$, tel que $|f| \, \mathbf{1}_B$ est bornée (disons par C), et que $\int_{B^c} |f| \, \mathrm{d}\mu < \varepsilon$. Donc si $\mu(A) < \delta := \frac{\varepsilon}{C}$, on aura

$$\int_A |f| \, \mathrm{d}\mu = \int_{A \cap B} |f| \, \mathrm{d}\mu + \int_{A \cap B^c} |f| \, \mathrm{d}\mu \le C \, \mu(A) + \int_{B^c} |f| \, \mathrm{d}\mu \le 2\varepsilon \, ,$$

ce qu'il fallait démontrer. □

Théorème 2.14. (Théorème de convergence dominée de Lebesgue). *Soit (f_n) une suite de fonctions dans $\mathscr{L}^1(E, \mathscr{A}, \mu)$. On suppose :*

(1) Il existe une fonction mesurable f telle que $f_n \to f$, μ-p.p. ;

(2) Il existe g intégrable telle que pour tout n, $|f_n| \le g$, μ-p.p.

Alors $f \in \mathscr{L}^1$, et $\lim_{n\to\infty} \int |f_n - f| \, \mathrm{d}\mu = 0$. En particulier,

$$\lim_{n\to\infty} \int f_n \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu.$$

Preuve. On suppose d'abord que les hypothèses suivantes plus fortes sont satisfaites :

$(1')$ $f_n(x) \to f(x)$, $\forall x \in E$.

$(2')$ Il existe une fonction mesurable $g : E \to \mathbb{R}_+$ telle que $\int g \, \mathrm{d}\mu < \infty$, et que $|f_n(x)| \le g(x)$, $\forall n$, $\forall x \in E$.

La propriété $f \in \mathscr{L}^1(E, \mathscr{A}, \mu)$ est alors claire puisque $|f| \le g$ et $\int g \, \mathrm{d}\mu < \infty$. Ensuite, puisque $|f_n - f| \le 2g$ et $|f_n - f| \to 0$, on peut appliquer le lemme de Fatou à la suite de fonctions positives $(2g - |f_n - f|)$, pour voir que

$$\liminf_{n\to\infty} \int (2g - |f_n - f|) \, \mathrm{d}\mu \ge \int \liminf_{n\to\infty} (2g - |f_n - f|) \, \mathrm{d}\mu = 2 \int g \, \mathrm{d}\mu.$$

Par linéarité de l'intégrale et le fait que $\int g \, \mathrm{d}\mu < \infty$, on obtient $\limsup_{n\to\infty} \int |f_n - f| \, \mathrm{d}\mu \le 0$. Donc $\int |f_n - f| \, \mathrm{d}\mu \to 0$.

Dans le cas général où l'on suppose seulement (1) et (2), on pose

$$A := \{x \in E : \, f_n(x) \to f(x)\} \cap \{x \in E : \, g(x) < \infty\} \cap \bigcap_{n \ge 1} \{x \in E : \, |f_n(x)| \le g(x)\} \in \mathscr{A}.$$

Par hypothèse, $\mu(A^c) = 0$.

Soient $\widetilde{f}_n := f_n \, \mathbf{1}_A$ et $\widetilde{f} := f \, \mathbf{1}_A$. On peut alors l'argument ci-dessus à $(\widetilde{f}_n)$ et $\widetilde{f}$, pour voir que $\int |\widetilde{f}_n - \widetilde{f}| \, \mathrm{d}\mu \to 0$. Or, $\widetilde{f}_n = f_n$ p.p., et $\widetilde{f} = f$ p.p., donc $\int |\widetilde{f}_n - \widetilde{f}| \, \mathrm{d}\mu = \int |f_n - f| \, \mathrm{d}\mu$.□

3. Espaces produit et théorème de Fubini

Soient $(E, \mathscr{A})$ et $(F, \mathscr{B})$ deux espaces mesurables. On peut alors munir le produit cartésien $E \times F$ de la tribu produit

$$\mathscr{A} \otimes \mathscr{B} := \sigma(A \times B : \, A \in \mathscr{A}, \; B \in \mathscr{B}).$$

Les ensembles de la forme $A \times B$ (avec $A \in \mathscr{A}$ et $B \in \mathscr{B}$) sont appelés pavés (ou : pavés mesurables), qui sont donc mesurables par rapport à la tribu produit.

Si $C \subset E \times F$, on pose, pour $x \in E$,

$$C_x := \{y \in F : \, (x, y) \in C\},$$

et pour $y \in F$,

$$C^y := \{x \in E : \, (x, y) \in C\}.$$

Si f est une application définie sur $E \times F$, on note pour $x \in E$, $f_x(y) := f(x, y)$, et pour $y \in F$, $f^y(x) := f(x, y)$.

Théorème 3.1. (i) *Soit $C \in \mathscr{A} \otimes \mathscr{B}$. Alors pour tout $x \in E$, on a $C_x \in \mathscr{B}$, et pour tout $y \in F$, on a $C^y \in \mathscr{A}$.*

(ii) *Soit $f : \, E \times F \to G$ une application mesurable pour la tribu produit $\mathscr{A} \otimes \mathscr{B}$. Alors pour tout $x \in E$, f_x est $\mathscr{B}$-mesurable, et pour tout $y \in F$, f^y est $\mathscr{A}$-mesurable.*

Preuve. (i) Fixons $x \in E$, et posons

$$\mathscr{C} := \{C \in \mathscr{A} \otimes \mathscr{B} : \, C_x \in \mathscr{B}\}.$$

Il est clair que $\mathscr{C}$ contient tous les pavés car si $C = A \times B$ avec $A \in \mathscr{A}$ et $B \in \mathscr{B}$, alors C_x est B ou $\varnothing$ selon que $x \in A$ ou $x \notin A$.

Par ailleurs, on vérifie facilement par définition que $\mathscr{C}$ est une tribu (car contenant $E \times F$, stable par passage au complémentaire, et stable par réunion dénombrable). Donc $\mathscr{C} = \mathscr{A} \otimes \mathscr{B}$.

(ii) Pour toute partie mesurable D de G,

$$f_x^{-1}(D) = \{y \in F : \, f(x, y) \in D\} = \{y \in F : \, (x, y) \in f^{-1}(D)\} = (f^{-1}(D))_x,$$

ce qui est un élément de $\mathscr{B}$ d'après (i). $\square$

Remarque 3.2. Il ne faut pas confondre le Théorème 3.1 avec le fait élémentaire que (f_1, f_2) est mesurable si et seulement si f_1 et f_2 sont mesurables. D'autre part, la "réciproque" du Théorème 3.1 est fausse. $\square$

Théorème 3.3. *Soient μ et ν des mesures σ-finies. Alors il existe une unique mesure m sur $(E \times F, \mathscr{A} \otimes \mathscr{B})$ telle que*[4]

$$m(A \times B) = \mu(A)\,\nu(B), \qquad \forall A \in \mathscr{A},\ \forall B \in \mathscr{B}.$$

Preuve. (Unicité) Supposons que m et $\widetilde{m}$ sont deux mesures sur $\mathscr{A} \otimes \mathscr{B}$ vérifiant la propriété énoncée du théorème.

Comme μ et ν sont σ-finies, il existe des suites croissantes $(A_n) \subset \mathscr{A}$ et $(B_n) \subset \mathscr{B}$, telles que $\mu(A_n) + \nu(B_n) < \infty$, $\forall n$, et que $E = \cup_{n\geq 1} A_n$, $F = \cup_{n\geq 1} B_n$.

Posons $C_n := A_n \times B_n$. Alors (C_n) est croissante, et $\cup_{n\geq 1} C_n = E \times F$.

• m et $\widetilde{m}$ coïncident sur la classe des pavés, qui est stable par intersections finies et engendre la tribu produit $\mathscr{A} \otimes \mathscr{B}$;

• $m(C_n) = \mu(A_n)\,\nu(B_n) = \widetilde{m}(C_n) < \infty$, $\forall n$.

Par classe monotone, les deux mesures m et $\widetilde{m}$ sont identiques.

(Existence) On pose, pour tout $C \in \mathscr{A} \otimes \mathscr{B}$,

$$m(C) := \int_E \nu(C_x)\,\mu(\mathrm{d}x). \tag{3.1}$$

Remarquons que $\nu(C_x)$ est bien définie pour tout $x \in E$ d'après le théorème précédent. Pour que la formule (3.1) ait un sens, il faut encore vérifier que la l'application $x \mapsto \nu(C_x)$ est mesurable.

Supposons d'abord ν finie. Posons

$$\mathscr{G} := \{C \in \mathscr{A} \otimes \mathscr{B} : x \mapsto \nu(C_x) \text{ est } \mathscr{A}\text{-mesurable}\}.$$

Alors,

• $\mathscr{G}$ contient tous les pavés, car si $C = A \times B \in \mathscr{A} \otimes \mathscr{B}$, alors $\nu(C_x) = \mathbf{1}_A(x)\,\nu(B)$ qui est $\mathscr{A}$-mesurable en x ;

• $\mathscr{G}$ est une classe monotone : si $C \subset C'$ sont mesurables, on a $\nu((C' \backslash C)_x) = \nu(C'_x) - \nu(C_x)$ (car ν est finie !) et si (C_n) est une suite croissante, $\nu((\cup_n C_n)_x) = \lim_{n\to\infty} \nu((C_n)_x)$.

La classe des pavés est stable par intersections finies. Par classe monotone, $\mathscr{G} = \mathscr{A} \otimes \mathscr{B}$, ce qui donne la mesurabilité recherchée pour l'application $x \mapsto \nu(C_x)$.

[4] Convention usuelle : $0 \times \infty := 0$.

Dans le cas général où ν n'est pas nécessairement finie mais seulement σ-finie, on choisit une suite croissante (B_n) d'éléments de $\mathscr{B}$ telle que $F = \cup_{n\ge 1} B_n$ et que $\nu(B_n) < \infty$, $\forall n$. En considérant la mesure finie $\nu_n(B) := \nu(B \cap B_n)$, $B \in \mathscr{B}$, on voit que $x \mapsto \nu_n(C_x)$ est $\mathscr{A}$-mesurable pour tout $C \in \mathscr{A} \otimes \mathscr{B}$. Il suffit alors de remarquer que $\nu(C_x) = \lim_{n\to\infty} \uparrow \nu_n(C_x)$. D'où la mesurabilité recherchée pour l'application $x \mapsto \nu(C_x)$ dans le cas général.

Montrons ensuite que m est une mesure sur $(E \times F,\, \mathscr{A} \otimes \mathscr{B})$: on a $m(\varnothing) = 0$, et si (C_n) est une suite de parties disjointes de $\mathscr{A} \otimes \mathscr{B}$, les $(C_n)_x$ sont aussi disjoints pour tout $x \in E$, et donc

$$m\Big(\bigcup_{n\ge 1} C_n\Big) = \int_E \nu\Big((\bigcup_{n\ge 1} C_n)_x\Big)\, \mu(\mathrm{d}x) = \int_E \nu\Big(\bigcup_{n\ge 1} (C_n)_x\Big)\, \mu(\mathrm{d}x) = \int_E \sum_{n\ge 1} \nu((C_n)_x)\, \mu(\mathrm{d}x).$$

D'après le Théorème 2.6, on peut inverser somme et intégrale, pour voir que tout cela vaut

$$= \sum_{n\ge 1} \int_E \nu((C_n)_x)\, \mu(\mathrm{d}x) = \sum_{n\ge 1} m(C_n),$$

ce qui prouve la σ-additivité de m : m est bien une mesure. De plus, par définition, on a clairement $m(A \times B) = \mu(A)\, \nu(B)$ pour tous les pavés. □

Notation et remarque. La mesure m sera notée $m = \mu \otimes \nu$, que l'on appellera **mesure produit**, est une mesure σ-finie sur l'espace produit $(E \times F,\, \mathscr{A} \otimes \mathscr{B})$. La preuve du théorème montre que pour tout $C \in \mathscr{A} \otimes \mathscr{B}$,[5]

$$(\mu \otimes \nu)(C) = \int_E \nu(C_x)\, \mu(\mathrm{d}x) = \int_F \mu(C^y)\, \nu(\mathrm{d}y).$$

On a déjà prouvé la première identité. Pour la seconde, on définit $\widetilde{m}$ par

$$\widetilde{m}(C) := \int_F \mu(C^y)\, \nu(\mathrm{d}y), \qquad \forall C \in \mathscr{A} \otimes \mathscr{B}.$$

Exactement comme pour m, on voit que $\widetilde{m}$ est une mesure sur $(E \times F,\, \mathscr{A} \otimes \mathscr{B})$, et qui attribue aux pavés les mêmes valeurs que m. Par classe monotone, $m = \widetilde{m}$. □

Théorème 3.4. (Théorème de Fubini–Tonelli) *Supposons que μ et ν sont σ-finies. Soit $f : E \times F \to [0, \infty]$ une fonction mesurable.*

(i) *Les fonctions $x \mapsto \int_F f(x, y)\, \nu(\mathrm{d}y)$ et $y \mapsto \int_E f(x, y)\, \mu(\mathrm{d}x)$ sont respectivement $\mathscr{A}$-mesurable et $\mathscr{B}$-mesurable.*

(ii) *On a*

$$\int_{E\times F} f\, \mathrm{d}(\mu \otimes \nu) = \int_E \Big(\int_F f(x, y)\, \nu(\mathrm{d}y)\Big) \mu(\mathrm{d}x) = \int_F \Big(\int_E f(x, y)\, \mu(\mathrm{d}x)\Big) \nu(\mathrm{d}y).$$

[5] Faisant partie de l'énoncé de la remarque : $x \mapsto \nu(C_x)$ et $y \mapsto \mu(C^y)$ sont mesurables.

Preuve. (i) Si $f = \mathbf{1}_C$ avec $C \in \mathscr{A} \otimes \mathscr{B}$, on a déjà vu que la fonction $x \mapsto \int_F f(x, y)\, \nu(\mathrm{d}y) = \nu(C_x)$ est $\mathscr{A}$-mesurable, et de même $y \mapsto \int_E f(x, y)\, \mu(\mathrm{d}x) = \mu(C^y)$ est $\mathscr{B}$-mesurable. Par linéarité, on déduit que le résultat de (i) est vrai pour toute fonction étagée (positive). Enfin, si f est quelconque (mais positive), on peut écrire $f = \lim_{n\to\infty} \uparrow f_n$, où (f_n) est une suite croissante de fonctions étagées positives, et on utilise le fait qu'alors,

$$\int_F f(x, y)\, \nu(\mathrm{d}y) = \lim_{n\to\infty} \uparrow \int_F f_n(x, y)\, \nu(\mathrm{d}y),$$

et de même pour $\int_E f(x, y)\, \mu(\mathrm{d}x)$.

(ii) Pour $f = \mathbf{1}_C$, avec $C \in \mathscr{A} \otimes \mathscr{B}$, le résultat de (ii) annoncé est

$$(\mu \otimes \nu)(C) = \int_E \nu(C_x)\, \mu(\mathrm{d}x) = \int_F \mu(C^y)\, \nu(\mathrm{d}y),$$

ce qui a déjà été prouvé précédemment. On en déduit, par linéarité, le résultat voulu quand f est étagée (positive), puis par limite croissante pour f quelconque (mais positive) : on remarque, par exemple, si $f = \lim_{n\to\infty} \uparrow f_n$, avec $f_n \geq 0$, $\forall n$, on a

$$\int_E \Big(\int_F f(x, y)\, \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x) = \lim_{n\to\infty} \uparrow \int_E \Big(\int_F f_n(x, y)\, \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x),$$

par une double application du théorème de convergence monotone. □

Théorème 3.5. (Théorème de Fubini–Lebesgue) *Supposons que μ et ν sont σ-finies. Soit $f \in \mathscr{L}^1(E \times F, \mathscr{A} \otimes \mathscr{B}, \mu \otimes \nu)$. Alors*

(i) $\mu(\mathrm{d}x)$-p.p. *la fonction $y \mapsto f(x, y)$ est dans $\mathscr{L}^1(F, \mathscr{B}, \nu)$, et* $\nu(\mathrm{d}y)$-p.p. *la fonction $x \mapsto f(x, y)$ est dans $\mathscr{L}^1(E, \mathscr{A}, \mu)$.*

(ii) *Les fonctions $x \mapsto \int_F f(x, y)\, \nu(\mathrm{d}y)$ et $y \mapsto \int_E f(x, y)\, \mu(\mathrm{d}x)$, bien définies sauf sur un ensemble mesurable de mesure nulle, sont respectivement dans $\mathscr{L}^1(E, \mathscr{A}, \mu)$ et dans $\mathscr{L}^1(F, \mathscr{B}, \nu)$.*

(iii) *On a*

$$\int_{E\times F} f\, \mathrm{d}(\mu \otimes \nu) = \int_E \Big(\int_F f(x, y)\, \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x) = \int_F \Big(\int_E f(x, y)\, \mu(\mathrm{d}x) \Big) \nu(\mathrm{d}y).$$

Preuve. (i) [On reconnaît que les fonctions $y \mapsto f(x, y)$ et $x \mapsto f(x, y)$ ont été précedemment notées par f_x et par f^y, respectivement.]

En appliquant le théorème de Fubini–Tonelli à $|f|$, on a

$$\int_E \Big(\int_F |f(x, y)|\, \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x) = \int_{E\times F} |f|\, \mathrm{d}(\mu \otimes \nu),$$

qui est finie par hypothèse. Donc $\mu(\mathrm{d}x)$-p.p., $\int_F |f(x,\, y)|\, \nu(\mathrm{d}y) < \infty$, et la fonction $y \mapsto f(x,\, y)$, dont on sait déjà la mesurabilité grâce au Théorème 3.1, est dans $\mathscr{L}^1(F,\, \mathscr{B},\, \nu)$.

(ii) En écrivant $f = f^+ - f^-$ et en utilisant le théorème de Fubini–Tonelli, on voit que

$$x \mapsto \int_F f(x,\, y)\, \nu(\mathrm{d}y) = \int_F f^+(x,\, y)\, \nu(\mathrm{d}y) - \int_F f^-(x,\, y)\, \nu(\mathrm{d}y)$$

est $\mathscr{A}$-mesurable.[6] De plus,

$$\int_E \Big| \int_F f(x,\, y)\, \nu(\mathrm{d}y) \Big|\, \mu(\mathrm{d}x) \le \int_E \Big(\int_F |f(x,\, y)|\, \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x),$$

qui, par le théorème de Fubini–Tonelli, n'est autre que $\int_{E\times F} |f|\, \mathrm{d}(\mu \otimes \nu)$, qui est donc finie par hypothèse.

(iii) On applique le théorème de Fubini–Tonelli à f^+ et à f^- :

$$\begin{aligned} \int_{E\times F} f^+\, \mathrm{d}(\mu\otimes\nu) &= \int_E \Big(\int_F f^+(x,\, y)\, \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x), \\ \int_{E\times F} f^-\, \mathrm{d}(\mu\otimes\nu) &= \int_E \Big(\int_F f^-(x,\, y)\, \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x). \end{aligned}$$

Il suffit de faire la différence des deux identités. □

Remarque 3.6. Faisons attentions aux hypothèses de base dans cette section : on suppose que μ et ν sont σ-finies. Sans cette hypothèse, les résultats ne sont, en général, plus valables. □

Remarque 3.7. Plus généralement, on peut considérer $E_1 \times \cdots \times E_n$ (pour un entier $n \ge 2$ quelconque), et prouver les deux versions du théorème de Fubini sur cet espace produit. □

[6] Strictement parlant, il faut donner une valeur, arbitraire (par exemple, 0), à l'intégrale $\int f(x,\, y)\, \nu(\mathrm{d}y)$ pour les x tels que $\int |f(x,\, y)|\, \nu(\mathrm{d}y) = \infty$, qui forment un ensemble de ν-mesure nulle.

Appendice B

Rappels de théorie des probabilités

Dans tout le chapitre, les variables aléatoires sont définies sur l'espace de probabilité $(\Omega, \mathscr{F}, \mathbb{P})$.

1. Espérance mathématique

Si X est une variable aléatoire à valeurs dans $[0, \infty]$, on note

$$\mathbb{E}(X) := \int_\Omega X(\omega)\, \mathbb{P}(\mathrm{d}\omega) \in [0, \infty],$$

qui est l'espérance mathématique (ou : espérance) de X. Si X est une variable aléatoire réelle *telle que* $\mathbb{E}(|X|) < \infty$, on définit son espérance mathématique

$$\mathbb{E}(X) := \int_\Omega X(\omega)\, \mathbb{P}(\mathrm{d}\omega) \in \mathbb{R}.$$

On étend cette définition au cas où $X := (X_1, \cdots, X_d)$ est une variable aléatoire à valeurs dans $\mathbb{R}^d$, en prenant $\mathbb{E}(X) := (\mathbb{E}(X_1), \cdots, \mathbb{E}(X_d))$, pourvu que chacune des espérances $\mathbb{E}(X_i)$ soit bien définie.

Théorème 1.1. *Soit X une variable aléatoire à valeurs dans $(E, \mathscr{E})$. Alors P_X, la loi de X, est l'unique mesure sur $(E, \mathscr{E})$ telle que*

$$\mathbb{E}[f(X)] = \int_E f(x)\, P_X(\mathrm{d}x),$$

pour toute fonction mesurable[1] $f : E \to \mathbb{R}_+$.

[1] La preuve montre que l'on peut remplacer $f : E \to \mathbb{R}_+$ par $f : E \to [0, \infty]$.

Preuve. L'unicité est triviale en prenant $f := \mathbf{1}_B$ pour $B \in \mathscr{E}$.

Pour prouver l'identité, on constate qu'elle est valable pour les fonctions f qui sont des fonctions indicatrices (c'est la partie unicité), et donc pour les fonctions étagées, et enfin pour les fonctions mesurables positives par convergence monotone (en utilisant le fait que toute fonction mesurable positive est la limite croissante simple d'une suite de fonctions étagées positives). □

Remarque 1.2. Le théorème reste évidemment valable si f est de signe quelconque, à condition que $\mathbb{E}[|f(X)|] < \infty$, ou, d'une façon équivalente, $\int_E |f| \, \mathrm{d}P_X < \infty$. Cas particulier : X est une variable aléatoire réelle, si $\mathbb{E}(X)$ est bien définie,[2] alors

$$\mathbb{E}(X) = \int_\Omega X(\omega) \, \mathbb{P}(\mathrm{d}\omega) = \int_\mathbb{R} x \, P_X(\mathrm{d}x).$$

Cette dernière écriture nous permet de dire que, lorsque X est une variable aléatoire réelle pour laquelle $\mathbb{E}(X)$ est bien définie, $\mathbb{E}(X)$ est une moyenne des valeurs de X, pondérée par la loi P_X. □

Remarque 1.3. Le théorème est souvent utilisé pour calculer la loi de X. Si l'on arrive à écrire

$$\mathbb{E}[f(X)] = \int_E f \, \mathrm{d}\nu,$$

pour une certaine mesure de probabilité ν sur E et pour toute fonction f "suffisamment générale" (par exemple, continue à support compact), alors ν n'est autre que P_X. □

2. Tribu engendrée par une variable aléatoire

Soit X une variable aléatoire à valeurs dans un espace mesurable $(E, \mathscr{E})$. La tribu engendrée par X, notée $\sigma(X)$, est définie comme la plus petite tribu sur Ω qui rende X mesurable :

$$\sigma(X) := \{A := X^{-1}(B), \ B \in \mathscr{E}\}.$$

Plus généralement, si $(X_i, \, i \in I)$ est une famille quelconque de variables aléatoires, X_i à valeurs dans $(E_i, \mathscr{E}_i)$, alors

$$\sigma(X_i, \, i \in I) := \sigma(X_i^{-1}(B_i) : \ B_i \in \mathscr{E}_i, \ i \in I).$$

[2] C'est-à-dire, si X est positive, ou si $\mathbb{E}(|X|) < \infty$.

Proposition 2.1. *Soit X une variable aléatoire à valeurs dans $(E, \mathscr{E})$. Soit Y une variable aléatoire réelle. Alors Y est $\sigma(X)$-mesurable si et seulement si $Y = f(X)$ pour une fonction mesurable $f : E \to \mathbb{R}$.*

Preuve. "$\Leftarrow$" Évident.

"$\Rightarrow$" Supposons que Y est $\sigma(X)$-mesurable. Si $Y = \mathbf{1}_A$ avec $A \in \mathscr{F}$, alors $A \in \sigma(X)$, et donc par définition, il existe $B \in \mathscr{E}$ tel que $A = X^{-1}(B)$. Donc $Y = \mathbf{1}_A = \mathbf{1}_{X^{-1}(B)} = \mathbf{1}_B(X)$: la proposition est valable avec $f := \mathbf{1}_B$.

Si Y est une fonction étagée $Y = \sum_{i=1}^p \alpha_i \mathbf{1}_{A_i}$ qui est $\sigma(X)$-mesurable, alors $A_i = Y^{-1}(\{\alpha_i\}) \in \sigma(X)$, et donc par le paragraphe précédent, $Y = f(X)$ avec f une fonction étagée sur E, mesurable.

Si Y est $\sigma(X)$-mesurable et positive, on sait que Y est la limite simple d'une suite croissante (Y_n) de fonctions $\sigma(X)$-mesurables étagées positives. Soit $Y_n = f_n(X)$, avec f_n fonction mesurable sur E. Alors $Y = f(X)$, où[3] $f(x) := \mathbf{1}_A(x) \lim_{n\to\infty} f_n(x)$, avec $A := \{x \in E : f_n(x) \text{ converge dans } \mathbb{R}\}$. La fonction f est mesurable sur E.

Enfin, si Y est $\sigma(X)$-mesurable de signe quelconque, il suffit de considérer Y^+ et Y^- séparément. □

3. Vecteurs aléatoires gaussiens

Soit X une variable aléatoire à valeurs dans $\mathbb{R}^d$. On dit que X est un vecteur gaussien si toute combinaison linéaire de ses composantes est gaussienne.[4]

Supposons que X est un vecteur aléatoire à valeurs dans $\mathbb{R}^d$:

• sa fonction caractéristique vaut $\exp[t' \,\mathbb{E}(X) - \frac{1}{2} t' K_X t]$, $\forall t \in \mathbb{R}^d$, où K_X est la matrice de covariances de X ;[5]

• X admet une densité si et seulement si K_X est inversible ; dans ce cas, la densité vaut

$$\frac{1}{(2\pi)^{d/2} [\det(K_X)]^{1/2}} \exp\Big(- \frac{1}{2} (x - \mathbb{E}(X))' K_X^{-1} (x - \mathbb{E}(X))\Big), \qquad x \in \mathbb{R}^d \,;$$

• soit $U = (X_1, \cdots, X_m, Y_1, \cdots, Y_n, Z_1, \cdots, Z_\ell)$ un vecteur gaussien. Alors les variables aléatoires $X := (X_1, \cdots, X_m)$ et $Y := (Y_1, \cdots, Y_n)$ sont indépendantes si et seulement si $\mathrm{Cov}(X_i, Y_j) = 0$, $\forall 1 \le i \le m$, $\forall 1 \le j \le n$.

[3] La valeur de f sur A^c n'a aucune importance, car c'est la composition $f \circ X$ qui nous intéresse, et aucune valeur de A^c ne peut être atteinte par X, vu que $Y_n = f_n(X)$ converge simplement.

[4] Les constantes sont considérées comme des gaussiennes, au sens dégénéré.

[5] Notation : t' désigne le transposé de t.

Conséquence importante : Si $(X_1, \cdots, X_n)$ est un vecteur aléatoire gaussien à valeurs dans $\mathbb{R}^n$, et si sa matrice de covariances est diagonale, alors $X_1, \cdots, X_n$ sont des variables aléatoires réelles indépendantes. En effet, la dernière propriété implique d'abord que X_n est indépendante de $(X_1, \cdots, X_{n-1})$, puis que X_{n-1} est indépendante de $(X_1, \cdots, X_{n-2})$, etc, ce qui permet de conclure.

4. Lemme de Borel–Cantelli

Soit $(A_n)_{n\geq 1}$ une suite d'éléments de $\mathscr{F}$. On note

$$\begin{aligned}\limsup_{n\to\infty} A_n &:= \bigcap_{n\geq 1}\bigcup_{k\geq n} A_k \in \mathscr{F},\\ \liminf_{n\to\infty} A_n &:= \bigcup_{n\geq 1}\bigcap_{k\geq n} A_k \in \mathscr{F}.\end{aligned}$$

Il est clair que $\liminf_{n\to\infty} A_n \subset \limsup_{n\to\infty} A_n$, et que $(\limsup_{n\to\infty} A_n)^c = \liminf_{n\to\infty} A_n^c$. On a également

$$\begin{aligned}\mathbf{1}_{\limsup_{n\to\infty} A_n} &= \limsup_{n\to\infty} \mathbf{1}_{A_n},\\ \mathbf{1}_{\liminf_{n\to\infty} A_n} &= \liminf_{n\to\infty} \mathbf{1}_{A_n}.\end{aligned}$$

Enfin, on a l'interprétation suivante :

$$\begin{aligned}\limsup_{n\to\infty} A_n &= \{\text{les événements } A_n \text{ sont réalisés infiniment souvent}\}\\ &= \Big\{\sum_n \mathbf{1}_{A_n} = \infty\Big\},\\ \liminf_{n\to\infty} A_n &= \{\text{les événements } A_n \text{ sont toujours réalisés à partir d'un certain rang}\}.\end{aligned}$$

Théorème 4.1. (Lemme de Borel–Cantelli) *Soit $(A_n)_{n\geq 1}$ une suite d'événements.*

(i) *Si $\sum_n \mathbb{P}(A_n) < \infty$, alors $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$, c'est-à-dire $\sum_n \mathbf{1}_{A_n} < \infty$,* p.s.

(ii) *Si $\sum_n \mathbb{P}(A_n) = \infty$, et si A_n sont indépendants, alors $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$, c'est-à-dire $\sum_n \mathbf{1}_{A_n} = \infty$,* p.s.

Preuve. (i) Par Fubini–Tonelli, $\mathbb{E}(\sum_n \mathbf{1}_{A_n}) = \sum_n \mathbb{P}(A_n) < \infty$, et donc $\sum_n \mathbf{1}_{A_n} < \infty$, p.s.

(ii) Pour $j \leq n$, on a

$$\mathbb{P}\Big(\bigcap_{k=j}^{n} A_k^c\Big) = \prod_{k=j}^{n} \mathbb{P}(A_k^c) = \prod_{k=j}^{n}[1 - \mathbb{P}(A_k)].$$

La divergence de la série $\sum_n \mathbb{P}(A_n)$ implique que $\prod_{k=j}^n [1-\mathbb{P}(A_k)] \to 0$, $n\to\infty$, et donc

$$\mathbb{P}\Big(\bigcap_{k=j}^{\infty} A_k^c\Big) = 0, \qquad \forall j.$$

Par conséquent,

$$\mathbb{P}\Big(\bigcup_{j=1}^{\infty}\bigcap_{k=j}^{\infty} A_k^c\Big) = 0,$$

ce qui équivaut à $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$. $\square$

Remarque 4.2. Dans l'assertion (ii) du lemme de Borel–Cantelli, la condition $\sum_n \mathbb{P}(A_n) = \infty$ seule n'implique pas $\sum_n \mathbf{1}_{A_n} = \infty$ p.s. [Un contre-exemple simple : $A_n := A$, $n \ge 1$, où $A \in \mathscr{F}$ est tel que $0 < \mathbb{P}(A) < 1$.] Il nous faut une condition supplémentaire. On a prouvé que l'indépendance de A_n, $n\ge 1$, suffit. Cependant, la preuve présentée ci-dessus pourrait donner la fausse impression qu'il nous faut une relation globale entre les événements A_n. En réalité, il suffit que les *couples* A_i et A_j, pour $i \neq j$, ne soient pas trop "positivement corrélés" (notre contre-exemple est un ca extrême de corrélation positive). Par exemple, on démontre assez facilement que si $\mathbb{P}(A_i\cap A_j) \le \mathbb{P}(A_i)\,\mathbb{P}(A_j)$, $\forall i\neq j$, alors la condition $\sum_n \mathbb{P}(A_n) = \infty$ implique que $\frac{\sum_{i=1}^n \mathbf{1}_{A_i}}{\sum_{i=1}^n \mathbb{P}(A_i)} \to 1$ p.s. ; en particulier, $\sum_n \mathbf{1}_{A_n} = \infty$ p.s. $\square$

5. Convergence de variables variables

Soient X, X_1, X_2, $\cdots$ des variables aléatoires à valeurs dans $\mathbb{R}^d$.

- $X_n \to X$ p.s., s'il existe $A\in\mathscr{A}$ avec $\mathbb{P}(A)=1$ tel que $X_n(\omega)\to X(\omega)$, $\forall \omega\in A$;
- $X_n \to X$ dans L^p (pour $1\le p<\infty$), si $X_n \in L^p$, $\forall n$, et $\mathbb{E}[\,|X_n - X|^p] \to 0$.
- $X_n \to X$ en probabilité si $\lim_{n\to\infty}\mathbb{P}(|X_n - X| > \varepsilon) = 0$, $\forall \varepsilon > 0$.

Théorème 5.1. (i) *Si $X_n \to X$* p.s. *ou dans L^p, alors $X_n\to X$ en probabilité.*

(ii) *Si $X_n \to X$ en probabilité, il existe une sous-suite $X_{n_k}\to X$* p.s.

(iii) *$X_n \to X$ en probabilité si et seulement si de toute sous-suite (X_{n_k}), on peut en extraire une sous-sous-suite $X_{n_{k_\ell}} \to X$* p.s.

Conséquences importantes. Le théorème précédent permet souvent d'alléger les conditions dans des résultats connus :

Lemme de Fatou (version convergence en probabilité) : si $X_n\to X$ en probabilité et si $X_n \ge 0$ p.s., alors $\mathbb{E}(X) \le \liminf_{n\to\infty}\mathbb{E}(X_n)$.

Théorème de convergence dominée (version convergence en probabilité) : si $X_n \to X$ en probabilité et si $|X_n| \le Y$, $\forall n$, et $\mathbb{E}(|Y|) < \infty$, alors $X_n \to X$ dans L^1 et $\mathbb{E}(X_n) \to \mathbb{E}(X)$.□

Exercice 5.2. (i) Soit (a_n) une suite de nombre réels, et soit $a \in \mathbb{R}$ un réel. Montrer que $a_n \to a$ (quand $n \to \infty$) si et seulement si de toute sous-suite (a_{n_k}), on peut en extraire une sous-sous-suite $a_{n_{k_\ell}} \to a$, $\ell \to \infty$.

(ii) Camille dit : en appliquant le résultat de (i) à la suite $a_n := \mathbb{P}(|X_n - X| > \varepsilon)$ (pour $\varepsilon > 0$ fixé), on voit que $X_n \to X$ en probabilité si et seulement si de toute sous-suite (X_{n_k}), on peut en extraire une sous-sous-suite $X_{n_{k_\ell}} \to X$ en probabilités. Qu'en pensez-vous ?

(iii) Dominique dit : en appliquant le résultat de (i) à la suite $a_n := X_n(\omega)$ (pour $\omega \in \Omega$ fixé), on voit que $X_n \to X$ p.s. si et seulement si de toute sous-suite (X_{n_k}), on peut en extraire une sous-sous-suite $X_{n_{k_\ell}} \to X$ p.s. Qu'en pensez-vous ? □

Théorème 5.3. (Loi forte des grands nombres) *Soit $(X_n)_{n\ge 1}$ une suite de variables aléatoires réelles i.i.d.,*[6] *telle que $\mathbb{E}(|X_1|) < \infty$, alors*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}(X_1), \qquad \text{p.s.}$$

6. Convergence en loi

Soient X, X_1, X_2, $\cdots$, des variables aléatoires à valeurs dans $\mathbb{R}^d$. On dit que (X_n) converge en loi vers X si $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ pour toute fonction $f : \mathbb{R}^d \to \mathbb{R}$ continue bornée. [On peut vérifier que ceci équivaut également à $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ pour toute fonction f continue à support compact.]

Il y a un abus de langage à dire que la suite de variables aléatoires (X_n) converge en loi vers X, car la variable aléatoire limite X n'est pas définie de manière unique : seule sa loi l'est. La convergence en loi est, en réalité, une notion de convergence pour suites de mesures, et non pas pour suites de variables aléatoires.[7] Ainsi, les variables aléatoires ne sont, a priori, pas nécessairement définies sur un même espace de probabilité. On supposera cependant, sauf mention du contraire, que toutes les variables aléatoires sont définies sur un même espace de probabilité générique, noté $(\Omega, \mathscr{F}, \mathbb{P})$.

Propriété 6.1. *Soient X, X_1, X_2, $\cdots$ des variables aléatoires à valeurs dans $\mathbb{R}^d$. Soit $a \in \mathbb{R}^d$.*

[6] C'est-à-dire, indépendantes et identiquement distribuées.

[7] C'est pour cette raison que l'on dit de temps en temps que X_n converge en loi vers P_X.

- $X_n \to X$ *en probabilité* $\Rightarrow$ $X_n \to X$ *en loi.*
- $X_n \to a$ *en loi* $\Leftrightarrow$ $X_n \to a$ *en probabilité.*

Théorème 6.2. (Théorème de portmanteau) *Les conditions suivantes sont équivalentes :*

(i) $X_n \to X$ *en loi ;*

(ii) $\limsup_{n\to\infty} \mathbb{P}(X_n \in F) \le \mathbb{P}(X \in F)$, $\forall F \subset \mathbb{R}^d$ *fermé ;*

(iii) $\liminf_{n\to\infty} \mathbb{P}(X_n \in G) \ge \mathbb{P}(X \in G)$, $\forall G \subset \mathbb{R}^d$ *ouvert ;*

(iv) $\lim_{n\to\infty} \mathbb{P}(X_n \in A) = \mathbb{P}(X \in A)$, $\forall A \in \mathscr{B}(\mathbb{R}^d)$ *tel que* $\mathbb{P}(X \in A) = 0$.

Preuve. [Le théorème de portmanteau (avec un “p” en minuscule) ne fait pas nécessairement partie du programme de la théorie élémentaire des probabilités, mais le résultat est bien utile, et la preuve peu technique.]

“(i) $\Rightarrow$ (iii)” Pour tout $k \ge 1$, soit $f_k(x) := [k\,\mathrm{dist}(x, G^c)] \wedge 1$, $x \in \mathbb{R}^d$. On voit que $f_k \in C_b(\mathbb{R}^d)$ (f_k est même lipschitzienne, à valeurs dans $[0, 1]$), et que $f_k \uparrow \mathbf{1}_G$. Donc

$$\liminf_{n\to\infty} \mathbb{P}(X_n \in G) \ge \sup_{k\ge 1} \liminf_{n\to\infty} \mathbb{E}[f_k(X_n)] = \sup_{k\ge 1} \mathbb{E}[f_k(X)],$$

qui n'est autre que $\mathbb{P}(X \in G)$ (par convergence monotone).

“(ii) $\Leftrightarrow$ (iii)” Trivial.

“(ii)+(iii) $\Rightarrow$ (iv)” Pour $\forall A \in \mathscr{B}(\mathbb{R}^d)$, on a[8]

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{P}(X_n \in A) &\le \limsup_{n\to\infty} \mathbb{P}(X_n \in \overline{A}) \le \mathbb{P}(X \in \overline{A}),\\ \liminf_{n\to\infty} \mathbb{P}(X_n \in A) &\ge \liminf_{n\to\infty} \mathbb{P}(X_n \in \mathring{A}) \ge \mathbb{P}(X \in \mathring{A}).\end{aligned}$$

Si $\mu(\partial A) = 0$, on a alors $\mathbb{P}(X \in \overline{A}) = \mathbb{P}(X \in \mathring{A}) = \mathbb{P}(X \in A)$. D'où (iv).

“(iv) $\Rightarrow$ (i)” Soit f une fonction continue et bornée sur $\mathbb{R}^d$. On veut montrer que $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$.

Quitte à décomposer $f = f^+ - f^-$, on peut supposer que $f \ge 0$. Soit $K := \|f\|_\infty < \infty$. Par le théorème de Fubini–Tonelli,

$$\mathbb{E}[f(X)] = \mathbb{E}\Big(\int_0^K \mathbf{1}_{\{t<f(X)\}}\,\mathrm{d}t\Big) = \int_0^K \mathbb{P}(X \in A_t)\,\mathrm{d}t,$$

où $A_t := \{x \in \mathbb{R}^d :\ f(x) > t\}$. De même, pour tout n, $\mathbb{E}[f(X_n)] = \int_0^K \mathbb{P}(X_n \in A_t)\,\mathrm{d}t$.

[8]Notations : $\overline{A}$ est l'adhérence de A, et $\mathring{A}$ la partie intérieure de A.

Observons que $\partial A_t \subset B_t := \{x \in \mathbb{R}^d : f(x) = t\}$ (il est possible d'avoir une inclusion stricte !), et que l'ensemble

$$\Big\{t : \mathbb{P}(X \in B_t) > 0\Big\}$$

est au plus infini dénombrable[9], et donc de mesure de Lebesgue nulle. Par conséquent, d'après (iv), on a $\mathbb{P}(X_n \in A_t) \to \mathbb{P}(X \in A_t)$, pour presque tout t. Par convergence dominée,

$$\int_0^K \mathbb{P}(X_n \in A_t)\,\mathrm{d}t \to \int_0^K \mathbb{P}(X \in A_t)\,\mathrm{d}t, \qquad n \to \infty,$$

c'est-à-dire, $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$, $n \to \infty$. □

Théorème 6.3. *Soient X, X_1, X_2, $\cdots$, des variables aléatoires réelles. Alors $X_n \to X$ en loi si et seulement si $F_{X_n}(x) \to F_X(x)$, $\forall x \in \mathbb{R}$ tel que F_X soit continue au point x.*[10]

Preuve. "$\Rightarrow$" Il suffit d'appliquer le théorème de portmanteau au borélien $A :=]-\infty, x]$ pour lequel $\mathbb{P}(X \in A) = 0$ si (et seulement si) F_X est continue au point x.

"$\Leftarrow$" L'ensemble des points x auxquels F_X n'est pas continue est dénombrable (car F_X est monotone). Soient $a < b$ des réels.

Il existe $(a_k) \downarrow a$ tel que $\forall k$, F_X soit continue au point a_k. Par hypothèse, $F_{X_n}(a_k) \to F_X(a_k)$, $\forall k$. Donc

$$\limsup_{n\to\infty} F_{X_n}(a) \le \limsup_{n\to\infty} F_{X_n}(a_k) = F_X(a_k),$$

et il suffit de faire $k \to \infty$ et utiliser la continuité à droite de la fonction de répartition pour voir que

$$\limsup_{n\to\infty} F_{X_n}(a) \le F_X(a).$$

De même, en considérant une suite strictement croissante de points $(b_k) \uparrow\uparrow b$ auxquels F_X est continue, on a $\liminf_{n\to\infty} F_{X_n}(b-) \ge \liminf_{n\to\infty} F_{X_n}(b_k) = F_X(b_k)$, $\forall b$, et donc

$$\liminf_{n\to\infty} F_{X_n}(b-) \ge F_X(b-).$$

Il en découle que la condition (iii) du théorème de portmanteau est satisfaite pour $G :=]a, b[\subset \mathbb{R}$ un intervalle ouvert. Il suffit ensuite d'écrire un ouvert $G \subset \mathbb{R}$ quelconque comme réunion dénombrable disjointe d'intervalles ouverts pour voir que la condition (iii) est satisfaite pour tout ouvert $G \subset \mathbb{R}$. Par conséquent, $X_n \to X$ en loi. □

[9] En effet, pour tout k, il existe au plus k valeurs distinctes de t telles que $\mathbb{P}(X \in B_t) \ge \frac{1}{k}$.

[10] Notation : Pour toute variable aléatoire réelle X, on note sa fonction de répartition $F_X(x) := \mathbb{P}(X \le x)$, $x \in \mathbb{R}$.

Théorème 6.4. (Slutsky). *Soient $\{X_n\}_{n\ge 1}$ et $\{Y_n\}_{n\ge 1}$ des suites de variables aléatoires réelles. Si $X_n \to X$ en loi, et si $Y_n \to a \in \mathbb{R}$ en loi, alors*

(i) $X_n + Y_n \to X + a$ *en loi;*

(ii) $X_n Y_n \to a\, X$ *en loi;*

(iii) $\frac{X_n}{Y_n}\mathbf{1}_{\{Y_n \neq 0\}} \to \frac{X}{a}$ *en loi, si $a \neq 0$.*

Preuve. On vérifie seulement (i), la preuve de (ii) et (iii) étant tout à fait dans le même esprit.

Soient $x \in \mathbb{R}$ et $r > 0$. On a, pour tout $0 < \varepsilon < r$,

$$\begin{aligned}\mathbb{P}(X_n + Y_n \in B(x, r)) &\ge \mathbb{P}(X_n \in B(x-a, r-\varepsilon), \ Y_n \in B(a, \varepsilon)) \\ &\ge \mathbb{P}(X_n \in B(x-a, r-\varepsilon)) - \mathbb{P}(Y_n \notin B(a, \varepsilon)).\end{aligned}$$

Comme $\mathbb{P}(Y_n \notin B(a, \varepsilon)) \to 0$, $n \to \infty$, on a, par le théorème de portmanteau,

$$\begin{aligned}\liminf_{n\to\infty} \mathbb{P}(X_n + Y_n \in B(x, r)) &\ge \liminf_{n\to\infty} \mathbb{P}(X_n \in B(x-a, r-\varepsilon)) - 0 \\ &\ge \mathbb{P}(X \in B(x-a, r-\varepsilon)) \\ &= \mathbb{P}(X + a \in B(x, r-\varepsilon)).\end{aligned}$$

On prend $\varepsilon = \frac{1}{k}$. Comme $k \mapsto B(x-a, r-\frac{1}{k}) \uparrow$, et $\cup_k B(x-a, r-\frac{1}{k}) = B(x-a, r)$, ceci donne

$$\liminf_{n\to\infty} \mathbb{P}(X_n + Y_n \in B(x, r)) \ge \mathbb{P}(X + a \in B(x, r)).$$

La condition (iii) du théorème de portmanteau est donc satisfaite pour $\mu_n := P_{X_n+Y_n}$, $\mu := P_{X+a}$, et $G \subset \mathbb{R}$ un intervalle ouvert. Il suffit ensuite d'écrire un ouvert $G \subset \mathbb{R}$ quelconque comme réunion dénombrable disjointe d'intervalles ouverts pour voir que la condition (iii) est satisfaite pour tout ouvert $G \subset \mathbb{R}$. Par conséquent, $X_n + Y_n \to X + a$ en loi. □

Théorème 6.5. *Soient X, X_1, X_2, $\cdots$, des variables aléatoires à valeurs dans $\mathbb{R}^d$. Alors $X_n \to X$ en loi si et seulement si $\Phi_{X_n}(t) \to \Phi_X(t)$, $\forall t \in \mathbb{R}^d$, où Φ désigne fonction caractéristique.*

Remarque 6.6. En général, si Φ_{X_n} converge simplement, il est possible que la fonction limite (notée Φ) ne soit pas une fonction caractéristique associée à une loi de probabilité sur $\mathbb{R}^d$. On n'a, dans ce cas, pas de convergence en loi.[11]

Il est donc important de savoir si la fonction limite Φ est une fonction caractéristique. Un résultat célèbre (théorème de Lévy), et admis ici, tout comme le Théorème 6.5, de la

[11] Le même phénomène peut se produire également avec la fonction de répartition ou la fonction de densité.

théorie des probabilités dit que si la fonction limite Φ est continue au point 0, alors elle est nécessairement une fonction caractéristique. Par conséquent, X_n converge en loi.

De même, si (X_n) est une suite de variables à densité aléatoires à valeurs $\mathbb{R}^d$ telle que la suite des densités f_{X_n} converge presque partout (par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$), il n'y a aucune raison pour que $f(x) := \lim_{n\to\infty} f_{X_n}(x)$ ($x \in \mathbb{R}^d$) soit une densité de loi de probabilité. Si, de plus, pour tout n, $f_{X_n}(x) \le g(x)$ presque partout et $\int_{\mathbb{R}^d} g(x)\,\mathrm{d}x < \infty$,[12] alors il résulte du théorème de convergence dominée que f est une densité de loi de probabilité, et (X_n) converge en loi. □

Exemple 6.7. Soient $Y, Y_1, Y_2, \cdots$ des variables aléatoires à valeurs dans $\mathbb{R}^d$. Alors $Y_n \to Y$ en loi si et seulement si $Y_n \cdot t \to Y \cdot t$ en loi, $\forall t \in \mathbb{R}^d$, où "$\cdot$" est le produit scalaire de $\mathbb{R}^d$.

En effet, si $Y_n \to Y$ en loi, et si f est une fonction continue et bornée sur $\mathbb{R}$, alors pour tout $t \in \mathbb{R}^d$, la fonction $g : \mathbb{R}^d \to \mathbb{R}$ définie par $g(x) := f(x \cdot t)$ est continue et bornée sur $\mathbb{R}^d$, et donc $\mathbb{E}[g(Y_n)] \to \mathbb{E}[g(Y)]$, c'est-à-dire, $\mathbb{E}[f(Y_n \cdot t)] \to \mathbb{E}[f(Y \cdot t)]$. Donc $Y_n \cdot t \to Y \cdot t$ en loi.

Inversement, si $Y_n \cdot t \to Y \cdot t$ en loi, $\forall t \in \mathbb{R}^d$, alors par le Théorème 6.5 (en étudiant la fonction caractéristique de $Y_n \cdot t$ au point $1 \in \mathbb{R}$), $\mathbb{E}(\mathrm{e}^{i\,Y_n \cdot t}) \to \mathbb{E}(\mathrm{e}^{i\,Y \cdot t})$, c'est-à-dire, $\Phi_{Y_n}(t) \to \Phi_Y(t)$. Donc, de nouveau d'après le Théorème 6.5, $Y_n \to Y$ en loi.

Avec le même argument, on peut étendre le théorème de Slutsky en dimension quelconque.[13] □

Théorème 6.8. (Théorème central limite) *Soit $(X_n)_{n\ge 1}$ une suite de variables aléatoires réelles i.i.d. admettant un moment d'ordre 2, telle que $\sigma^2 := \mathrm{Var}(X_1) \in]0, \infty[$. Alors*

$$\frac{1}{n^{1/2}} \sum_{i=1}^{n} (X_i - \mathbb{E}(X_i)) \to \mathscr{N}(0, \sigma^2), \qquad \text{en loi,}$$

où $\mathscr{N}(0, \sigma^2)$ désigne la loi gaussienne centrée de variance σ^2.

Preuve. Quitte à remplacer X_n par $X_n - \mathbb{E}(X_n)$, on peut supposer que $\mathbb{E}(X_1) = 0$. Posons alors

$$Z_n := \frac{\sum_{i=1}^n X_i}{n^{1/2}}.$$

La fonction caractéristique de Z_n est

$$\Phi_{Z_n}(t) = [\Phi_{X_1}(\frac{t}{n^{1/2}})]^n.$$

[12] Notation : $\int_{\mathbb{R}^d} g(x)\,\mathrm{d}x$ désigne l'intégrale *au sens de Lebesgue*.

[13] Pour le produit, il y a plusieurs versions : $X_n \to X$ en loi dans $\mathbb{R}^d$ et $Y_n \to a \in \mathbb{R}$ en loi ; ou $X_n \to X$ en loi dans $\mathbb{R}$ et $Y_n \to a \in \mathbb{R}^d$ en loi.

Or, X_1 admettant un moment d'ordre 2, on sait, par convergence dominée, que Φ_{X_1} est de classe C^2 :

$$\Phi_{X_1}(t) = 1 + i\,t\,\mathbb{E}(X_1) - \frac{1}{2}\,t^2\,\mathbb{E}(X_1^2) + o(t^2) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2), \qquad t \to 0.$$

Pour $t \in \mathbb{R}$ fixé, on a donc $\Phi_{X_1}(\frac{t}{n^{1/2}}) = 1 - \frac{\sigma^2 t^2}{2n} + o(\frac{1}{n})$, $n \to \infty$. Par conséquent, pour tout $t \in \mathbb{R}$,

$$\Phi_{Z_n}(t) \to \exp\Big(- \frac{\sigma^2 t^2}{2}\Big), \qquad n \to \infty,$$

qui est la fonction caractéristique de $\mathscr{N}(0,\, \sigma^2)$. D'après le Théorème 6.5, Z_n converge en loi vers $\mathscr{N}(0,\, \sigma^2)$. $\square$

Théorème 6.9. (Théorème central limite vectoriel) *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires i.i.d. à valeurs dans $\mathbb{R}^d$. On suppose que toute composante de X_1 admet un moment d'ordre 2. Alors, quand $n \to \infty$,*

$$\frac{1}{n^{1/2}} \sum_{j=1}^n (X_j - \mathbb{E}(X_j)) \to \mathscr{N}(0,\, K_{X_1}), \qquad \textit{en loi,}$$

où K_{X_1} est la matrice de covariances de X_1.

Preuve. C'est la même preuve que dans le cas réel. On peut encore supposer que $\mathbb{E}(X_1) = 0$. Ensuite, pour tout $t \in \mathbb{R}^d$,

$$\mathbb{E}\Big[\exp\Big(it \cdot \frac{1}{n^{1/2}} \sum_{j=1}^n X_j \Big) \Big] = \Big[\Phi_{X_1}(\frac{t}{n^{1/2}}) \Big]^n.$$

Or, puisque les composantes de X_1 admettent un moment d'ordre 2, on a $\Phi_{X_1} \in C^2$ et

$$\Phi_{X_1}(\frac{t}{n^{1/2}}) = 1 - \frac{1}{2n}\, t'\, K_{X_1} t + o(\frac{1}{n}), \qquad n \to \infty,$$

ce qui implique que

$$\mathbb{E}\Big[\exp\Big(it \cdot \frac{1}{n^{1/2}} \sum_{j=1}^n X_j \Big) \Big] \to \exp\Big(- \frac{1}{2}\, t'\, K_{X_1} t \Big), \qquad n \to \infty.$$

On obtient alors la convergence en loi grâce au Théorème 6.5. $\square$

Appendice C

Références bibliographiques

Baldi, P., Mazliak, L. et Priouret, P. : *Martingales et Chaînes de Markov : Théorie élémentaire et exercices corrigés.* Hermann, 2001.

Barbe, P. et Ledoux, M. : *Probabilité.* EDP Sciences, 2007.

Benaïm, M. et El Karoui, N. : *Promenade Aléatoire : Chaînes de Markov et simulations ; martingales et stratégies.* Ecole Polytechnique, 2005.

Billingsley, P. : *Probability and Measure.* Wiley, 1995.

Briane, M. et Pagès, G. : *Théorie de l'Intégration.* Vuibert, 2006.

Durrett, R.T. : *Probability: Theory and Examples.* Cambridge University Press, 2010.

Gramain, A. : *Intégration.* Hermann, 1988.

Neveu, J. : *Martingales à Temps Discret.* Masson, 1972.

Norris, J.R. : *Markov Chains.* Cambridge University Press, 1998.

Rudin, W. : *Real and Complex Analysis.* McGraw-Hill, 1987.

Williams, D. : *Probability with Martingales.* Cambridge University Press, 1991.