Martingales discrètes

Parmi les objets probabilistes qu'un expert aime rencontrer, les martingales ont une place de choix. Pour le comprendre, comme pas mal d'exposés probabilistes, un point de départ est la loi des grands nombres.

Théorème .1 : Loi des grands nombres

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Considérons $(X_n)_n$ une suite de variables aléatoires réelles indépendantes et de même loi qu'une variable aléatoire $X \in L^1(\mathbb{P})$. Alors

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{\mathbb{P}-\text{p.s.}, L^1(\mathbb{P})} \mathbb{E}[X].$$

Dans ce Théorème, l'objet d'étude principal est alors $S_n = \sum_{k=1}^{n} X_k$. Alors $S_{n+1} = S_n + X_{n+1}$ est la somme de deux variables aléatoires indépendantes. Et si le deuxième terme était dépendant du premier ? Cette question est une raison de considérer les martingales. Il existe d'ailleurs une loi des grands nombres pour les martingales L^2, mais cela ne sera hélas pas abordé du tout ici...

Nous allons introduire une notion d'espérance conditionnelle qui permet de préciser la question de non-indépendance. À partir de celle-ci, la route sera dégagée pour définir ce qu'est une martingale, et de montrer que ces objets convergent sous des hypothèses raisonnables. On conclura sur une application des théorèmes de convergence des martingales sur un théorème d'analyse : le Théorème de Rademacher pour les fonctions lipschitziennes.

Commençons d'abord par motiver ces notions sur un exemple non probabiliste en apparence, avant d'entrer dans les détails.

I Un calcul préliminaire

Considérons l'espace mesuré $([0,1), \mathcal{B}([0,1)), \mathrm{Leb})$. Considérons une découpe dyadique de $[0,1)$: pour $n \geqslant 1$ et $0 \leqslant k \leqslant 2^n - 1$:

$$I_{n,k} \overset{\text{déf.}}{=} \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right).$$

Soit pour tout $n \geqslant 1$

$$\mathcal{B}_n \overset{\text{déf.}}{=} \left\{\bigsqcup_{k\in K} I_{n,k}, K \subset [\![0, 2^n-1]\!]\right\}.$$

Alors $\mathcal{B}_n$ est une tribu, puisque $\varnothing \in \mathcal{B}_n$, puisque stable par complémentaire et par intersection dénombrable. De plus, $\mathcal{B}_n \subset \mathcal{B}_{n+1}$: on découpe de plus en plus précisément $[0,1)$, car en effet $I_{n,k} = I_{n+1,2k} \sqcup I_{n+1,2k+1}$.

Soit $g : [0,1) \longrightarrow \mathbb{R}$. L'application g est $\mathcal{B}_n$-mesurable si pour tout $B \in \mathcal{B}(\mathbb{R})$, $\{g \in B\} \in \mathcal{B}_n$, donc si pour tout $B \in \mathcal{B}(\mathbb{R})$, il existe $K \subset [\![0, 2^n-1]\!]$ (éventuellement vide) tel que $\{g \in B\} = \bigsqcup_{k\in K} I_{n,k}$. Dans le cas particulier où $B = \{x\}$ pour $x \in \mathbb{R}$, on en déduit l'existence de $K(x)$ fini tel que $\{g = x\} = \bigsqcup_{k\in K(x)} I_{n,k}$. Si $x \neq x'$ alors $K(x) \cap K(x') = \varnothing$. Ainsi, il n'existe qu'un nombre fini de x tels que $K(x) \neq \varnothing$. Ainsi, il n'existe qu'un nombre fini de valeurs possibles pour g. On vient de montrer que si g est $\mathcal{B}_n$-mesurable alors g est constante par morceaux, g est constante sur chaque intervalle $I_{n,k}$.

La question qu'on se pose est la suivante : étant donné $f : [0,1) \longrightarrow \mathbb{R}$, quelle est une bonne approximation de f comme fonction constante par morceaux sur les $I_{n,k}$? Est-ce que cette approximation converge bien vers f quand on découpe l'intervalle en tous petits morceaux ? Les réponses sont oui quand $f \in L^1(0,1)$, c'est la notion d'espérance conditionnelle, et oui, c'est le théorème de convergence des martingales. (bien sûr il faut préciser les sens de ces approximations et de ces convergences)

Supposons $f \in L^2(0,1)$. Notons E_n le sous-espace vectoriel des fonctions de $L^2(0,1)$ qui sont $\mathcal{B}_n$-mesurables, donc constantes sur chaque intervalle $I_{n,k}$. On cherche $f_n \in E_n$ telle que

$$\|f - f_n\|_{L^2(0,1)} = \inf_{g\in E_n} \|f - g\|.$$

Il est bien connu de la théorie hilbertienne qu'un tel problème admet une unique solution, il s'agit de la projection orthogonale de f sur E_n. On peut le faire à la main ici. Si $g \in E_n$, $g = \sum_{k=0}^{2^n-1} \alpha_k \mathbf{1}_{I_{n,k}}$. On cherche les coefficients α_k qui minimisent $\|f - g\|$. On a en fait

$$\|f-g\|^2 = \sum_{k=0}^{2^n-1} \int_{I_{n,k}} (f(t) - \alpha_k)^2 \ \mathrm{d}t.$$

En minimisant terme par terme, on est en fait réduit au problème suivant : si I est un intervalle fini de $\mathbb{R}$ et $f \in L^2(I)$, on cherche $c \in \mathbb{R}$ tel que $\int_I (f(t) - c)^2 \ \mathrm{d}t = \inf_{\gamma \in \mathbb{R}} \int_I (f(t) - \gamma)^2 \ \mathrm{d}t$. En développant à gauche, on obtient un trinôme en c qu'on minimise en un unique $c = \frac{1}{|I|} \int_I f(t) \ \mathrm{d}t$. Cette valeur de c n'est rien d'autre que l'espérance de f, où plus simplement, sa valeur moyenne. Pour cette valeur de c, on a alors

$$\inf_{\gamma \in \mathbb{R}} \int_I (f(t) - \gamma)^2 \ \mathrm{d}t = |I| \left(\frac{1}{|I|} \int_I f(t)^2 \ \mathrm{d}t - \left(\frac{1}{|I|} \int_I f(t) \ \mathrm{d}t \right)^2 \right).$$

On retrouve en fait $|I|$ fois la variance de f.

Pour revenir à notre problème, on obtient alors aussi une unique minimisation f_n donnée par

$$f_n \overset{(\natural)}{=} \sum_{k=0}^{2^n-1} 2^n \left(\int_{I_{n,k}} f(t) \ \mathrm{d}t \right) \mathbf{1}_{I_{n,k}}.$$

Dorénavant, on notera $f_n = \mathbb{E}[f|\mathcal{B}_n]$. Il s'agit de l'unique application $\mathcal{B}_n$-mesurable qui minimise la distance entre f et E_n, l'espace des fonctions $\mathcal{B}_n$-mesurable. De plus,

$$\|f - \mathbb{E}[f|\mathcal{B}_n]\|^2 = \frac{1}{2^n} \left(\int_0^1 f(t)^2 \ \mathrm{d}t - \left(\int_0^1 f(t) \ \mathrm{d}t \right)^2 \right) \xrightarrow[n \to \infty]{} 0.$$

Ainsi, on dispose de la convergence L^2 de $\mathbb{E}[f|\mathcal{B}_n]$ vers f, donc de la convergence L^1 d'après Cauchy–Schwarz. Remarquons en fait que $\mathbb{E}[f|\mathcal{B}_n]$ tel que donné par ($\natural$) est en fait très bien défini pour $f \in L^1(0,1)$. On peut aussi prouver la convergence presque partout pour la mesure de Lebesgue. Si $x \in [0,1)$, il existe une unique suite d'entiers $(k_n(x))_{n \geqslant 1}$ telle que $\bigcap_{n \geqslant 1}^{\downarrow} I_{n,k_n(x)} = \{x\}$. Dans ce cas, on a en fait $\frac{k_n(x)}{2^n} \xrightarrow[n \to \infty]{} x$. Ainsi, on a en fait pour tout $n \geqslant 1$:

$$\mathbb{E}[f|\mathcal{B}_n](x) = \frac{1}{2^n} \int_{I_{n,k_n(x)}} f(t) \ \mathrm{d}t.$$

On conclut par dérivabilité en x de $y \longmapsto \int_0^y f(t) \ \mathrm{d}t$ que $\mathbb{E}[f|\mathcal{B}_n](x)$ converge vers $f(x)$ quand n tend vers $+\infty$ pour presque tout $x \in [0,1)$, car la valeur de la dérivée de $y \longmapsto \int_0^y f(t) \ \mathrm{d}t$ en x ne vaut $f(x)$ que presque partout. On peut aussi écrire

$$\mathbb{E}[f|\mathcal{B}_n](x) = \int_0^1 f(t) \left(2^n \sum_{k=0}^{2^n-1} \mathbf{1}_{I_{n,k}}(t) \ \mathbf{1}_{I_{n,k}}(x) \right) \mathrm{d}t = \int_0^1 f(t) \ \nu_n \left(\mathrm{d}t, \left(\mathbf{1}_{I_{n,k}} \right)_{0 \leqslant k \leqslant 2^n-1} (x) \right),$$

où ν_n est appelé *noyau de probabilité* et est définie comme une application $\mathcal{B}([0,1)) \times \{0,1\}^{2^n} \longrightarrow \mathbb{R}$ telle que pour tout $A \in \mathcal{B}([0,1))$ et $\mathbf{x} \in \{0,1\}^{2^n}$:

$$\nu_n(A, \mathbf{x}) \overset{\text{déf.}}{=} \int_A \frac{2^n \sum_{k=0}^{2^n-1} \mathbf{1}_{I_{n,k}}(t) \ x_k}{\sum_{k=0}^{2^n-1} x_k} \ \mathrm{d}t,$$

lorsque $\sum_{k=0}^{2^n-1} x_k \neq 0$ et $\nu_n(A, \mathbf{x}) = 0$ sinon. Notez que à $\mathbf{x}$ fixé, $\nu_n(\cdot, \mathbf{x})$ définit une probabilité sur $[0,1)$. Ce noyau de probabilité ν_n permet en fait d'interpréter l'espérance conditionnelle comme une espérance par rapport à une mesure dépendante de la tribu connue $\mathcal{B}_n$.

II Espérance conditionnelle

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. On notera $\mathbb{E}$ son espérance associée : pour tout $X \in L^1(\mathbb{P})$,

$$\mathbb{E}[X] \overset{\text{déf.}}{=} \int_\Omega X(\omega)\ \mathrm{d}\mathbb{P}(\omega) = \int_\mathbb{R} x\ \mathrm{d}\mathbb{P}_X(x),$$

où $\mathbb{P}_X \overset{\text{déf.}}{=} \mathbb{P} \circ X^{-1} : \mathcal{B}(\mathbb{R}) \longrightarrow [0,1]$ est la *loi de* X. On définit l'espérance conditionnelle comme dans l'introduction : à savoir à une sous-tribu $\mathcal{G}$ donnée, on cherche la variable aléatoire qui approche le mieux une variable X au sens L^2. Ainsi, cette approximation dans le cadre L^2 est connu dans le cadre hilbertien, il s'agit de la projection orthogonale sur l'espace des variables $\mathcal{G}$-mesurables, et elle est unique.

Définition II.1

Soient $X \in L^2(\mathbb{P})$ et $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. On appelle *espérance conditionnelle de X sachant $\mathcal{G}$* l'unique variable aléatoire $Y \in L^2(\mathcal{G})$ (c'est-à-dire que Y est $\mathcal{G}$-mesurable et vérifie $\mathbb{E}[Y^2] < \infty$) telle que pour tout $Z \in L^2(\mathcal{G})$

$$\mathbb{E}[XZ] = \mathbb{E}[YZ].$$

On note alors $Y \overset{\text{déf.}}{=} \mathbb{E}[X|\mathcal{G}]$.

Rappelons qu'être $\mathcal{G}$-mesurable signifie que pour tout borélien $B \in \mathcal{B}(\mathbb{R})$, $\{Y \in B\}(= \{\omega \in \Omega, Y(\omega) \in B\}) \in \mathcal{G}$. Autrement dit, la loi de Y est entièrement connue dès lors qu'on connaît tous les événements qui composent $\mathcal{G}$. Ainsi, rechercher une espérance conditionnelle signifie qu'on cherche une variable dont la loi en connaissant uniquement des événements d'une sous-tribus, donc *a priori*, lorsqu'on dispose de moins d'informations, est la plus proche possible de la variable initiale.

On peut étendre la définition d'espérance conditionnelle au variables aléatoires L^1.

Proposition II.2 : Espérance conditionnelle dans L^1

Soient $X \in L^2(\mathbb{P})$ et $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. Alors $Y = \mathbb{E}[X|\mathcal{G}]$ est l'unique variable aléatoire $L^2(\mathbb{P})$ telle que pour tout $A \in \mathcal{G}$

$$\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[Y\mathbf{1}_A].$$

On peut alors étendre $X \longmapsto \mathbb{E}[X|\mathcal{G}]$ sur L^1 *via* cette caractérisation, et dans ce cas $\mathbb{E}[X|\mathcal{G}] \in L^1(\mathcal{G})$.

Calculons explicitement quelques espérances conditionnelles en utilisant cette caractérisation.

1. Si $\mathcal{G} = \{\varnothing, \Omega\}$ est la tribu grossière alors $\mathbb{E}[X|\{\varnothing, \Omega\}] = \mathbb{E}[X]$. En effet, les seules fonctions mesurables par rapport à la tribu grossière sont les fonctions constantes.

2. Pour $\mathcal{G}$ sous-tribu, si X est $\mathcal{G}$-mesurable, alors $\mathbb{E}[X|\mathcal{G}] = X$.

3. Considérons un événement $A \in \mathcal{F}$ avec $\mathbb{P}(A) \notin \{0,1\}$. Considérons $\mathcal{G} = \sigma(A) = \{\varnothing, A, A^c, \Omega\}$, et calculons $\mathbb{E}[X|\sigma(A)]$. Pour cela, observons que des fonctions $\sigma(A)$-mesurables sont les indicatrices $\mathbf{1}_A$ et $\mathbf{1}_{A^c}$. En fait, puisque $\mathbf{1}_A \times \mathbf{1}_{A^c} = 0$, on peut écrire

$$\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}\left[\left(\frac{\mathbb{E}[X\mathbf{1}_A]}{\mathbb{P}(A)}\mathbf{1}_A + \frac{\mathbb{E}[X\mathbf{1}_{A^\times}]}{\mathbb{P}(A^\times)}\mathbf{1}_{A^c}\right)\mathbf{1}_A\right].$$

Idem pour $\mathbb{E}[X\mathbf{1}_{A^c}]$. On en déduit que

$$\mathbb{E}[X|\sigma(A)] = \frac{\mathbb{E}[X\mathbf{1}_A]}{\mathbb{P}(A)}\mathbf{1}_A + \frac{\mathbb{E}[X\mathbf{1}_{A^c}]}{\mathbb{P}(A^c)}\mathbf{1}_{A^c}.$$

En particulier, si $X = \mathbf{1}_E$ alors, avec $\mathbb{E}[\mathbf{1}_E|\sigma(A)] \overset{\text{déf.}}{=} \mathbb{P}(E|\sigma(A))$, on en déduit que

$$\mathbb{P}(E|\sigma(A)) = \frac{\mathbb{P}(E \cap A)}{\mathbb{P}(A)}\mathbf{1}_A + \frac{\mathbb{P}(E \cap A^c)}{\mathbb{P}(A^c)}\mathbf{1}_{A^c}.$$

4. Plus généralement, si $\Omega = N \sqcup \bigsqcup_{i\in I} E_i$, avec $\mathbb{P}(N) = 0$, $\mathbb{P}(E_i) > 0$ et I dénombrable alors

$$\mathbb{E}[X|\sigma(E_i, i \in I)] = \sum_{i\in I} \frac{\mathbb{E}[X\mathbf{1}_{E_i}]}{\mathbb{P}(E_i)} \mathbf{1}_{E_i}.$$

5. Si X est indépendant de $\mathcal{G}$ alors $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$, car si $E \in \mathcal{G}$ alors $\mathbb{E}[X\mathbf{1}_E] = \mathbb{E}[X]\mathbb{E}[\mathbf{1}_E] = \mathbb{E}[\mathbb{E}[X]\mathbf{1}_E]$.

6. Un exemple de calcul de *régression gaussienne*. Soit (X, Y) un couple gaussien (*id est* pour tous réels $t, s \in \mathbb{R}$, $tX + sY$ est une variable aléatoire gaussienne réelle) centré (d'espérance nulle) non dégénéré (X et Y ne sont pas presque sûrement constantes ou encore n'ont pas de variance nulle). On cherche $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$, où $\sigma(Y) = \{\{Y \in B\}, B \in \mathcal{B}(\mathbb{R})\}$. Pour cela, on cherche $\lambda \in \mathbb{R}$ tel que $X - \lambda Y$ soit indépendant de Y. Puisque le couple est gaussien, cela arrive si et seulement si $\mathbb{E}[(X - \lambda Y)Y] = \mathbb{E}[X - \lambda Y]\mathbb{E}[Y] = 0$. Donc cela arrive pour $\lambda = \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}$. Ainsi,

$$\mathbb{E}[X|Y] = \mathbb{E}[X - \lambda Y|Y] + \mathbb{E}[\lambda Y|Y] = 0 + \lambda Y = \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]} Y.$$

Concluons cette sous-section par une Proposition qui permet de simplifier les calculs d'espérances conditionnelles.

Proposition II.3 : Produit par une variable $\mathcal{G}$-mesurable

Soient $X \in L^1(\mathbb{P})$, $\mathcal{G}$ une sous-tribu de $\mathcal{F}$ et $Y \in L^1(\mathcal{G})$ tel que $XY \in L^1(\mathbb{P})$. Alors

$$\mathbb{E}[XY|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}].$$

Démonstration : Il s'agit encore et toujours d'utiliser la caractérisation L^1. Si $A \in \mathcal{G}$ alors par définition de $\mathbb{E}[X|\mathcal{G}]$, puisque $Y\mathbf{1}_A$ est $\mathcal{G}$-mesurable, on a

$$\mathbb{E}[XY\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]Y\mathbf{1}_A].$$

Or $\mathbb{E}[X|\mathcal{G}]Y$ est $\mathcal{G}$-mesurable, si bien qu'on conclut la démonstration par unicité de l'espérance conditionnelle. □

Concluons cette section par une remarque sur la terminologie.

Remarque II.4. Le nom d'espérance conditionnelle n'est pas usurpé. On peut en effet constater qu'on dispose de tous les théorèmes d'intégration en version conditionnelle : linéarité, positivité, lemme de Fatou, théorème de convergence dominée, inégalité de Markov, etc... On peut même lui associer une probabilité, à ω fixé, et avoir une égalité du type

$$\mathbb{E}[X|\mathcal{G}](\omega) = \int_\Omega X(\omega')\ \mathbb{P}(\mathrm{d}\omega'|\mathcal{G})(\omega).$$

De cette dernière égalité, on peut encore se poser beaucoup de questions, mais qui se retrouvent éloignées de cet exposé...

♣

III Convergence de martingales

III.1 Martingales

Définition III.1

1. Soit $(\mathcal{F}_n)_{n\geqslant 1}$ une suite de sous-tribus de $\mathcal{F}$. On dit que $(\mathcal{F}_n)_n$ est une *filtration* si pour tout $n \geqslant 1$, $\mathcal{F}_n \subset \mathcal{F}_{n+1}$. Dans ce cas, on note

$$\mathcal{F}_\infty \overset{\text{déf.}}{=} \sigma\left(\bigcup_{n\geqslant 1}^{\uparrow} \mathcal{F}_n\right)$$

la tribu dite *asymptotique*.

2. Soit $(\mathcal{F}_n)_n$ une filtration. Une suite de variables aléatoires réelles $(X_n)_{n\geqslant 1}$ est une *martingale* par rapport à la filtration $(\mathcal{F}_n)_{n\geqslant 1}$ si

(a) pour tout $n \geqslant 1$, X_n est $\mathcal{F}_n$-mesurable (pour tout borélien $B \in \mathcal{B}(\mathbb{R})$, $\{X_n \in B\} \in \mathcal{F}_n$) ;

(b) pour tout $n \geqslant 1$, $X_n \in L^1(\mathbb{P})$;

(c) pour tout $n \geqslant 1$,
$$X_n = \mathbb{E}[X_{n+1}|\mathcal{F}_n].$$

Remarque III.2. Si on dispose d'une suite de variables aléatoires, on dispose en fait d'une filtration dite *canonique* : il s'agit de $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$. ♣

Remarque III.3. On note plus rapidement les conditions **(i)** et **(ii)** en $X_n \in L^1(\mathcal{F}_n)$. ♣

Remarque III.4. Puisque l'espérance est la même chose que l'espérance de l'espérance conditionnelle, on conclut que pour tout $n \geqslant 1$ $\mathbb{E}[X_n] = \mathbb{E}[X_1]$, qui ne dépend donc pas de n. ♣

Quelques exemples.

1. Soient $X \in L^1(\mathbb{P})$ et $(\mathcal{F}_n)_{n\geqslant 1}$ une filtration. On définit pour tout $n \geqslant 1$, $X_n \overset{\text{déf.}}{=} \mathbb{E}[X|\mathcal{F}_n]$. Alors $(X_n)_n$ est une $(\mathcal{F}_n)_n$-martingale. En effet, si on fixe $n \geqslant 1$, par définition de l'espérance conditionnelle, X_n est bien $\mathcal{F}_n$-mesurable, d'où **(a)**. De plus, $\mathbb{E}[|\mathbb{E}[X|\mathcal{F}_n]|] \leqslant \mathbb{E}[|X|] < \infty$, donc $X_n \in L^1(\mathbb{P})$. Enfin,
$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_{n+1}]|\mathcal{F}_n].$$
Or, si $A \in \mathcal{F}_n$, on a en fait puisque $\mathcal{F}_n \subset \mathcal{F}_{n+1}$:
$$\mathbb{E}[\mathbb{E}[X|\mathcal{F}_{n+1}]\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_n]\mathbf{1}_A],$$
si bien que $\mathbb{E}[\mathbb{E}[X|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbb{E}[X|\mathcal{F}_n]$, d'où $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$, d'où **(iii)**. On a bien prouvé que $(X_n)_n$ est une martingale. On dit qu'elle est *fermée par* X. On verra que ces martingales convergent dans L^1 et $\mathbb{P}$-presque sûrement vers $\mathbb{E}[X|\mathcal{F}_\infty]$.

2. Soient $(\xi_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes centrées L^1. On note $X_n = \sum_{k=1}^n \xi_k$. Alors X_n est une martingale par rapport à la filtration canonique associée aux ξ_n. En effet, les points **(a)** et **(b)** sont vérifiés sans problèmes. Pour **(c)** :
$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[X_n|\mathcal{F}_n] + \mathbb{E}[\xi_{n+1}|\mathcal{F}_n] = X_n + \mathbb{E}[\xi_{n+1}] = X_n,$$
où on a utilisé le fait que X_n est $\mathcal{F}_n$-mesurable donc $\mathbb{E}[X_n|\mathcal{F}_n] = X_n$ et que ξ_{n+1} est indépendant des $\xi_1, \cdots, \xi_n$, donc de $\mathcal{F}_n$, donc $\mathbb{E}[\xi_{n+1}|\mathcal{F}_n] = \mathbb{E}[\xi_{n+1}] = 0$.

3. (Ruine du joueur version loto inversé). On considère un joueur qui joue à pile ou face. Il mise initialement 1 Franc (le jeu se passe en 1254, si on souhaite évoquer l'univers enrichi de cet exposé). Si la pièce tombe sur pile, il gagne sa mise et 1 Franc supplémentaire. Si la pièce tombe sur face, il rejoue en doublant sa mise. Au temps n, si la pièce n'est pas déjà tombée sur pile, et qu'elle tombe sur pile, alors la partie s'arrête et le joueur gagne la mise totale et 1 Franc. Si la pièce tombe sur face, alors le joueur perd sa mise et rejoue en la doublant.

Notons ξ_n le résultat de la pièce pour le lancer n. Les ξ_n sont indépendants, et de même loi $\frac{\delta_{\text{pile}}+\delta_{\text{face}}}{2}$. Notons X_n le gain du joueur au lancer n. Enfin, notons S_n le gain total du joueur au temps n : $S_n = \sum_{k=1}^n X_k$. Alors $(S_n)_n$ est une martingale par rapport à la filtration canonique associée au ξ_n. En effet, les points **(a)** et **(b)** sont bien vérifiés (X_k ne dépend que des lancers précédents $\xi_1, \cdots, \xi_k$ donc est $\mathcal{F}_k$-mesurable, et S_n ne prend qu'un nombre fini de valeurs). Reste à montrer **(c)**. Notons que puisque la mise au temps k est 2^{k-1} :
$$\begin{aligned} X_{n+1} &= \left(\left(1 + \sum_{k=1}^n 2^{k-1}\right)\mathbf{1}_{\{\xi_{n+1}=\text{pile}\}} - 2^n\mathbf{1}_{\{\xi_{n+1}=\text{face}\}}\right)\prod_{k=1}^n \mathbf{1}_{\{\xi_k=\text{face}\}} \\ &= \left(2^n\mathbf{1}_{\{\xi_{n+1}=\text{pile}\}} - 2^n\mathbf{1}_{\{\xi_{n+1}=\text{face}\}}\right)\prod_{k=1}^n \mathbf{1}_{\{\xi_k=\text{face}\}}. \end{aligned}$$
Alors, avec $\mathcal{F}_n = \sigma(\xi_1, \cdots, \xi_n)$, en utilisant l'indépendance des lancers :
$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = (2^n\mathbb{P}(\xi_{n+1} = \text{pile}) - 2^n\mathbb{P}(\xi_{n+1} = \text{face}))\prod_{k=1}^n \mathbf{1}_{\{\xi_k=\text{face}\}} = 0,$$

si bien que $\mathbb{E}[S_{n+1}|\mathcal{F}_n] = \mathbb{E}[S_n|\mathcal{F}_n] + \mathbb{E}[X_{n+1}|\mathcal{F}_n] = S_n$. On verra que $(S_n)_n$ converge $\mathbb{P}$-presque sûrement vers 1.

4. (Intégration stochastique discrète) Supposons qu'on dispose d'une $(\mathcal{F}_n)_{n\geqslant 1}$-martingale $(X_n)_n$ et d'une suite de variables aléatoires $(H_n)_n$ intégrables telle que pour tout $n \geqslant 1$, H_n est $\mathcal{F}_{n-1}$-mesurable (on dit que $(H_n)_n$ est *prévisible*). Notons $I_1 = 0$ et pour $n \geqslant 2$

$$I_n \overset{\text{déf.}}{=} \sum_{k=2}^{n} H_k(X_k - X_{k-1}).$$

On peut en fait interpréter I_n comme une intégrale stochastique discrète (qu'on note $(H \cdot X)_n$ ou de manière plus étrange $\int_{1\leqslant k\leqslant n} H_k \, \mathrm{d}X_k$). C'est le point de départ du calcul stochastique, avec par exemple l'intégration stochastique d'Itô, par rapport à un mouvement brownien, ou plus généralement par rapport à une martingale L^2, ou encore plus généralement par rapport à une semi-martingale. On déborde bien sûr du cadre discret que cet exposé respecte.

La suite $(I_n)_n$ est une martingale par rapport à la même filtration que $(X_n)_n$. L'intégrabilité est vérifiée sans soucis. Pour la mesurabilité, pas de soucis non plus puisque la somme s'arrête au terme $H_n(X_n - X_{n-1})$. Montrons l'égalité de martingales.

$$\mathbb{E}[I_{n+1}|\mathcal{F}_n] = I_n + \mathbb{E}\left[H_{n+1}(X_{n+1} - X_n)|\mathcal{F}_n\right].$$

Puisque H est prévisible, il reste

$$\mathbb{E}[I_{n+1}|\mathcal{F}_n] = I_n + H_{n+1}\left(X_{n+1} - \mathbb{E}\left[X_n|\mathcal{F}_n\right]\right) = I_n.$$

Pour redéborder encore un peu, notez qu'on peut forcer la prévisibilité de H en définissant

$$S_n \overset{\text{déf.}}{=} \sum_{k=2}^{n} \mathbb{E}[H_k|\mathcal{F}_{k-1}](X_k - X_{k-1}).$$

Il s'agit alors d'une nouvelle intégrale stochastique, nommée *intégrale de Skorohod*, et qui est quasi omniprésente dans le calcul de Malliavin.

III.2 Convergence presque sûre de martingales

Le théorème principal est le suivant.

Théorème III.5 : Convergence presque sûre de martingales

Soit $(X_n)_{n\geqslant 1}$ une $(\mathcal{F}_n)_{n\geqslant 1}$-martingale telle que

$$\sup_{n\geqslant 1} \mathbb{E}\left[X_n^+\right] < \infty \text{ ou } \sup_{n\geqslant 1} \mathbb{E}\left[X_n^-\right] < \infty.$$

Alors $(X_n)_n$ converge presque sûrement vers une limite $X \in L^1(\mathbb{P})$.

Remarque III.6. Rappelons qu'on note $x^+ \overset{\text{déf.}}{=} \max\{x, 0\}$ et $x^- \overset{\text{déf.}}{=} \max\{-x, 0\} \geqslant 0$. Une martingale bornée dans L^1 (donc vérifiant $\sup_{n\geqslant 1} \mathbb{E}[|X_n|] < \infty$) vérifie alors ces conditions. ♣

Remarque III.7. Il s'agit bien d'un « ou ». Ce que je cache ici est la notion de sous-martingale et de sur-martingale, qui des analogues aléatoires de suite croissante et décroissante. Dans le cadre de sous-martingales, il suffit d'avoir $\sup_{n\geqslant 1} \mathbb{E}[X_n^+] < \infty$ pour avoir convergence presque sûre, c'est l'analogue de "une suite croissante majorée converge". Idem pour les sur-martingales positives "une suite décroissante minorée (par zéro ici) converge". Une martingale étant une sur et une sous-martingale, c'est de là que provient ce « ou ». ♣

L'idée de démonstration de ce résultat repose sur une caractérisation déterministe de la convergence dans $\bar{\mathbb{R}}$ en termes de nombre de montées dans un intervalle. Plaçons-nous dans ce cadre déterministe. Soit $(x_n)_n$ une suite réelle. On fixe $a < b$ et on définit les temps d'atteinte suivants : on note $N_0 = 0$. Pour $n = 1$, on définit

$$N_1^{a-} \overset{\text{déf.}}{=} \inf\left\{n \geqslant 1, x_n \leqslant a\right\}$$

et

$$N_1^{b+} \overset{\text{déf.}}{=} \inf\left\{n \geqslant N_1^{a-}, x_n \geqslant b\right\}.$$

Il s'agit des premiers temps où la suite se retrouve en dessous de l'intervalle $[a,b]$ puis se retrouve au dessus de cet intervalle. Pour $n \geqslant 1$, on note

$$N_n^{a-} \overset{\text{déf.}}{=} \inf\left\{n \geqslant N_{n-1}^{b+}, x_n \leqslant a\right\}$$

et

$$N_n^{b+} \overset{\text{déf.}}{=} \inf\left\{n \geqslant N_n^{a-}, x_n \geqslant b\right\}.$$

On autorise à ce que ces quantités soient infinies. On définit alors pour tout $n \geqslant 1$

$$M_n^x([a,b]) \overset{\text{déf.}}{=} \sup\left\{n \in \mathbb{N}, N_n^{b+} < \infty\right\}.$$

$M_n^x([a,b])$ compte le nombre de montées de l'intervalle $[a,b]$ jusqu'au temps n. Il s'agit d'une suite (d'entiers) croissante, donc admet une limite, finie ou non. On note $M_\infty^x([a,b])$ sa limite.

Lemme III.8 : Convergence et nombre de montées

La suite réelle $(x_n)_n$ converge dans $\overline{\mathbb{R}}$ si et seulement si

$$\forall a < b \in \mathbb{Q}, M_\infty^x([a,b]) < \infty.$$

Démonstration du lemme : La suite converge dans $\bar{\mathbb{R}}$ si et seulement si limite supérieure et limite inférieure coïncident (et sont éventuellement infinies). Or, si la suite converge, on a bien un nombre fini de montées de $[a,b]$. Et si la suite ne converge pas, cela signifie qu'il existe $a < b$ rationnels tels que

$$\liminf_{n\to\infty} x_n < a < b < \limsup_{n\to\infty} x_n.$$

Or, puisque les limites inférieures et supérieures sont des valeurs d'adhérence, $(x_n)_n$ séjournera une infinité de fois près de ces deux valeurs, donc montera une infinité de fois l'intervalle $[a,b]$. D'où l'équivalence. □

L'objectif est alors de montrer que pour la martingale X, on a pour tout $a < b$ rationnels $M_\infty^X([a,b]) < \infty$ $\mathbb{P}$-presque sûrement. Pour cela, on utilise le lemme, encore une fois déterministe, suivant.

Lemme III.9 : Une inégalité pour le nombre de montées

On considère la suite $(y_j)_{j\geqslant 1} \subset \{0,1\}$ définie de la manière suivante : pour tout $j \geqslant 1$

$$y_j \overset{\text{déf.}}{=} \sum_{k=0}^{+\infty} \mathbf{1}_{\left\{N_k^{b+} < j \leqslant N_{k+1}^{a-}\right\}}.$$

Alors pour $n \geqslant 2$, on a

$$\sum_{k=2}^{n} y_k(x_k - x_{k-1}) \leqslant (x_n - a)^+ - (b-a)M_n^x([a,b]).$$

Pour démontrer ce lemme, il suffit de distinguer différents cas selon le fait qu'on soit dans une phase de descente (c'est-à-dire qu'il existe k tel que $N_k^{b+} < n \leqslant N_{k+1}^{a-}$), de montée, ou une phase initiale ($0 = N_0 < n \leqslant N_1^{a-}$).

À partir de là, on peut démontrer le théorème.

Démonstration : Supposons, quitte à considérer $-X$ que l'hypothèse vérifiée est $\sup_{n\geqslant 1} \mathbb{E}\left[X_n^+\right] < \infty$. Notons Y_j les variables aléatoires données par les y_j, et associées à la martingale X : pour $j \geqslant 1$

$$Y_j = \sum_{k=0}^{+\infty} \mathbf{1}_{\left\{N_k^{b+} < j \leqslant N_{k+1}^{a-}\right\}},$$

où les N_k sont définies en remplaçant les "x" par des "X". Voici le plan de démonstration.

1. Pour tout $j \geqslant 1$, Y_j est $\mathcal{F}_{j-1}$-mesurable.
2. Par l'exemple **4.** de la sous-section III.1, $\sum_{k=2}^n Y_k(X_k - X_{k-1})$ définit une martingale. En prenant l'espérance dans le lemme III.8, on en déduit que $\mathbb{E}[M_\infty^X([a,b])] < \infty$ donc $M_\infty^X([a,b]) < \infty$ $\mathbb{P}$-presque sûrement, donc $(X_n)_n$ converge presque sûrement dans $\overline{\mathbb{R}}$.
3. Si X est la limite presque sûre de $(X_n)_n$ alors $X \in L^1$ *via* des lemmes de Fatou, donc est presque sûrement finie.

1. Montrons que Y_j est $\mathcal{F}_{j-1}$-mesurable. Pour cela, observons que $Y_j = 1$ signifie qu'on se situe dans une phase de descente, donc qu'il existe k tel que $N_k^{b+} < X_j \leqslant N_{k+1}^{a-}$. Ainsi, selon le temps précédent, on peut distinguer deux cas et écrire :

$$\{Y_j = 1\} = \{Y_{j-1} = 1\} \cap \{X_{j-1} > a\} \sqcup \{Y_{j-1} = 0\} \cap \{X_{j-1} \geqslant b\}.$$

En français cela signifie que si $Y_j = 1$ alors :

* Soit $Y_{j-1} = 1$, ce qui signifie que X_{j-1} est aussi en phase de descente, donc n'est pas descendu plus bas que a ;
* Soit $Y_{j-1} = 0$, ce qui signifie que X_{j-1} est en phase de montée, et vient de l'achever pour avoir $Y_j = 1$, donc X_{j-1} est monté plus haut que b.

En traitant le cas $j = 2$ (laissé en exercice au brave lecteur qui lira ce papier jusque là), on conclut par récurrence sur la $\mathcal{F}_{j-1}$-mesurabilité de Y_j.

2. Ainsi, on définit une martingale en définissant $I_n \overset{\text{déf.}}{=} \sum_{k=2}^n Y_k(X_k - X_{k-1})$, par l'exemple **4.**. Il suit que son espérance est la même que pour $n = 1$, c'est-à-dire $\mathbb{E}[I_n] = 0$. Il suit alors, d'après le lemme précédent sur le nombre de montées, en prenant l'espérance de cette inégalité que pour tout $n \geqslant 2$

$$\mathbb{E}[M_n^X(a,b)] \leqslant \frac{\mathbb{E}\left[X_n^+\right] - a}{b-a}.$$

Soit encore

$$\mathbb{E}[M_n^X(a,b)] \leqslant \frac{\sup_{n\geqslant 1}\mathbb{E}\left[X_n^+\right] + |a|}{b-a}.$$

Par convergence monotone, on en déduit que

$$\mathbb{E}[M_\infty^X(a,b)] \leqslant \frac{\sup_{n\geqslant 1}\mathbb{E}\left[X_n^+\right] + |a|}{b-a} < \infty.$$

Il suit que $(X_n)_n$ converge presque sûrement.

3. Soit X la limite presque sûre de $(X_n)_n$. D'une part

$$\mathbb{E}[X^+] = \mathbb{E}\left[\liminf_{n\to\infty} X_n^+\right] \leqslant \liminf_{n\to\infty} \mathbb{E}\left[X_n^+\right] \leqslant \sup_{n\geqslant 1}\mathbb{E}[X_n^+] < \infty.$$

D'autre part, puisque $(X_n)_n$ est une martingale, pour tout $n \geqslant 1$, $\mathbb{E}[X_n] = \mathbb{E}[X_1]$. Ainsi, encore *via* un lemme de Fatou

$$\mathbb{E}[X^-] \leqslant \mathbb{E}[X_0] + \liminf_{n\to\infty}\mathbb{E}[X_n^+] \leqslant \mathbb{E}[X_0] + \sup_{n\geqslant 1}\mathbb{E}[X_n^+] < \infty.$$

Ainsi, $\mathbb{E}[|X|] < \infty$ et $X \in L^1$, donc $|X| < \infty$ presque sûrement. □

III.3 Convergence et martingales fermées

On vient de voir un critère de convergence presque sûr des martingales. On va exposer un dernier théorème de convergence de martingales, pour cet exposé, qui évoque la convergence presque sûre et L^1. Voici son énoncé. On le démontrera en fin de sous-section.

Théorème III.10 : Convergence L^1 et presque sûre des martingales

Soit $(X_n)_{n\geqslant 1}$ une $(\mathcal{F}_n)_{n\geqslant 1}$-martingale. Les assertions suivantes sont équivalentes.

(i) $(X_n)_n$ converge presque sûrement et dans L^1.

(ii) $(X_n)_n$ est uniformément intégrable.

(iii) $(X_n)_n$ est fermée : il existe $X \in L^1$ telle que pour tout $n \geqslant 1$, $X_n = \mathbb{E}[X|\mathcal{F}_n]$.

Dans ce cas, la limite est donnée par $\mathbb{E}[X|\mathcal{F}_\infty]$.

Expliquons ce qu'est être uniformément intégrable, et voyons pourquoi ce théorème est vrai.

Définition III.11

Soit $(X_i)_{i\in I}$ une familles de variables aléatoires. On dit que $(X_i)_{i\in I}$ est *uniformément intégrable* si :

$$\sup_{i\in I}\mathbb{E}[|X_i|\mathbf{1}_{|X_i|>c}] \xrightarrow[c\to\infty]{} 0.$$

Il existe une manière relativement simple de montrer l'uniforme intégrabilité, qui est moins ingrate que sa définition, il s'agit du critère $L^{1+\delta}$.

Proposition III.12 : Critère $L^{1+\delta}$ de l'uniforme intégrabilité

Soit $(X_i)_{i\in I}$ une famille de variables aléatoires. S'il existe $\delta > 0$ telle que $(X_i)_{i\in I}$ soit bornée dans $L^{1+\delta}$ (c'est-à-dire si $\sup_{i\in I}\mathbb{E}\left[|X_i|^{1+\delta}\right] < \infty$) alors $(X_i)_{i\in I}$ est uniformément intégrable.

Démonstration : Il s'agit simplement d'écrire pour $c > 0$

$$\mathbb{E}\left[|X_i|\mathbf{1}_{|X_n|>c}\right] = \mathbb{E}\left[\frac{|X_i|^{1+\delta}}{|X_i|^{\delta}}\mathbf{1}_{|X_n|>c}\right] \leqslant \frac{\sup_{i\in I}\mathbb{E}[|X_i|^{1+\delta}]}{c^{\delta}}.$$

qui tend bien vers 0 quand c tend vers $+\infty$. Observons d'ailleurs qu'on dispose d'une vitesse de convergence, et qu'elle est d'autant meilleure que δ est grand. □

L'intérêt de la notion d'uniforme intégrabilité provient du théorème de Vitali, qui ressemble au théorème de convergence dominé, sauf que ce théorème nous donne une condition nécessaire et suffisante de convergence dans L^1. Rappelons d'abord que *converger en probabilité* signifie que pour tout $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon)$ tend vers 0 quand n tend vers $+\infty$. Il s'agit de la convergence de variables aléatoires la plus faible à manipuler, dans le sens où la convergence presque sûre implique cette convergence, tout comme la convergence L^1 l'implique aussi.

Théorème III.13 : Théorème de Vitali

Soit $(X_n)_n$ une suite de variables aléatoires. Les assertions suivantes sont équivalentes.

(a) $(X_n)_n$ converge en probabilité et est uniformément intégrable ;

(b) $(X_n)_n$ converge dans L^1.

Nous admettons ce théorème, qui se prouve à coup d'epsilon coupés en 36 et qui n'éclaircit pas le propos de cet exposé. Concluons sur la démonstration du Théorème de convergence L^1 des martingales.

Démonstration : On va d'abord montrer que **(i)** et **(ii)** sont équivalent, avant de montrer que **(iii)** implique **(i)** et **(ii)** implique **(iii)**. Nous concluerons sur la limite après.

[(i) $\Longrightarrow$ (ii)] Il s'agit du théorème de Vitali.

[(ii) $\Longrightarrow$ (i)] Supposons que $(X_n)_n$ soit uniformément intégrable. Alors en particulier, $(X_n)_n$ est bornée dans L^1 : $\sup_{n\geqslant 1}\mathbb{E}[|X_n|] < \infty$. Par le théorème de convergence presque sûre des martingales, on en déduit que $(X_n)_n$ converge presque sûrement, donc converge en probabilité. Par le théorème de Vitali, il suit que $(X_n)_n$ converge aussi dans L^1.

[(iii) $\Longrightarrow$ (ii)] Supposons que $(X_n)_n$ soit fermée par X. Alors pour tout $n \geqslant 1$:

$$\mathbb{E}\left[|X_n|\mathbf{1}_{\{|X_n|\geqslant c\}}\right] = \mathbb{E}\left[|\mathbb{E}[X|\mathcal{F}_n]|\mathbf{1}_{\{|X_n|\geqslant c\}}\right].$$

Par inégalité triangulaire conditionnelle

$$\mathbb{E}\left[|X_n|\mathbf{1}_{\{|X_n|\geqslant c\}}\right] \leqslant \mathbb{E}\left[\mathbb{E}[|X||\mathcal{F}_n]\mathbf{1}_{\{|X_n|\geqslant c\}}\right].$$

Or, puisque $\{X_n > c\} \in \mathcal{F}_n$, on a par définition de l'espérance conditionnelle

$$\mathbb{E}\left[|X_n|\mathbf{1}_{\{|X_n|\geqslant c\}}\right] \leqslant \mathbb{E}\left[|X|\mathbf{1}_{\{|X_n|\geqslant c\}}\right].$$

Pour conclure, on fait un petit tour de passe passe en faisant

$$1 = \mathbf{1}_{\{|X|\leqslant A\}} + \mathbf{1}_{\{|X|>A\}},$$

où $A > 0$ va être fixé juste après. Avec ce choix de $A > 0$, on a alors

$$\mathbb{E}\left[|X_n|\mathbf{1}_{\{|X_n|\geqslant c\}}\right] \leqslant \mathbb{E}\left[|X|\mathbf{1}_{\{|X|\geqslant A\}}\right] + A\mathbb{P}(|X_n| > c).$$

Si $\varepsilon > 0$, le théorème de convergence dominée nous permet de choisir $A = A(\varepsilon)$ tel que $\mathbb{E}\left[|X|\mathbf{1}_{\{|X|\geqslant A\}}\right] < \frac{\varepsilon}{2}$. De plus, par inégalité de Markov, on conclut que

$$\mathbb{E}\left[|X_n|\mathbf{1}_{\{|X_n|\geqslant c\}}\right] \leqslant \frac{\varepsilon}{2} + \frac{A\mathbb{E}[|X_n|]}{c}.$$

Or, $\mathbb{E}[|X_n|] \leqslant \mathbb{E}[|X|]$, donc on a réussi à montrer que

$$\sup_{n\geqslant 1}\mathbb{E}\left[|X_n|\mathbf{1}_{\{|X_n|\geqslant c\}}\right] \leqslant \frac{\varepsilon}{2} + \frac{A\mathbb{E}[|X|]}{c}.$$

On conclut en choisissant $M = M(\varepsilon) = \frac{2}{A\mathbb{E}[|X|]}$ que pour tout $\varepsilon > 0$, il existe $M > 0$ tel que pour tout $c > M$,

$$\sup_{n\geqslant 1}\mathbb{E}\left[|X_n|\mathbf{1}_{\{|X_n|\geqslant c\}}\right] \leqslant \varepsilon,$$

d'où l'uniforme intégrabilité.

[(i) $\Longrightarrow$ (iii)] Soit X la limite de $(X_n)_n$, presque sûre et L^1. Montrons que $X_n = \mathbb{E}[X|\mathcal{F}_n]$. Il suffit de montrer que pour tout $A \in \mathcal{F}_n$

$$\mathbb{E}[X_n\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A].$$

Pour cela, il suffit d'observer que la propriété de martingale nous affirme, en utilisant la définition de l'espérance conditionnelle que pour tout $p \geqslant 1$:

$$\mathbb{E}[X_n\mathbf{1}_A] = \mathbb{E}[X_{n+p}\mathbf{1}_A].$$

Or,

$$|\mathbb{E}[X_{n+p}\mathbf{1}_A] - \mathbb{E}[X\mathbf{1}_A]| \leqslant \mathbb{E}[|X_{n+p} - X|] \xrightarrow[p\to\infty]{} 0,$$

par la convergence L^1. D'où $\mathbb{E}[X_n\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A]$, donc $X_n = \mathbb{E}[X|\mathcal{F}_n]$ définit bien une martingale fermée.

- Montrons que la limite de la martingale fermée $X_n = \mathbb{E}[X|\mathcal{F}_n]$ est $\mathbb{E}[X|\mathcal{F}_\infty]$. Soit Z la limite L^1 et presque sûre de $(X_n)_n$. Il suffit de montrer pour tout $A \in \mathcal{F}_\infty$ que

$$\mathbb{E}[Z\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A].$$

Puisque $\mathcal{F}_\infty$ est engendré par $\bigcup_{n\geqslant 1}\mathcal{F}_n$, il suffit de le montrer pour A dans cette union et conclure par lemme de classe monotone. Soient $n_0 \in \mathbb{N}$ et $A \in \mathcal{F}_{n_0} \subset \bigcap_{p\geqslant 0}\mathcal{F}_{n_0+p}$. Alors pour tout $n \geqslant n_0$,

$$\mathbb{E}[X_n\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A].$$

On conclut en passant à la limite (ce qui est possible car la convergence est L^1). □

Revenons à notre exemple de loto inversé **3.**. Le gain total S_n au temps n est soit négatif, dans le cas où tous les résultats $\xi_1, \cdots, \xi_n$ sont face, soit est positif et vaut 1 lorsque pile est apparu avant ou exactement au temps n. Ainsi, $(S_n)_n$ est une martingale déterministiquement minorée par 1. $(S_n)_n$ converge alors presque sûrement vers une variable S quand $[n \to \infty]$. Montrons que $S = 1$ presque sûrement. Au temps n, S_n vaut soit 1 soit -2^n, avec

$$\mathbb{P}(S_n = -2^n) = \frac{1}{2^n}.$$

Ainsi, si $\varepsilon > 0$, on a pour n assez grand

$$\mathbb{P}(|1 - S_n| > \varepsilon) = \mathbb{P}(S_n < 1 - \varepsilon) + \mathbb{P}(S_n > 1 + \varepsilon) = \mathbb{P}(S_n = -2^n) = \frac{1}{2^n}.$$

Pour tout $\varepsilon > 0$, il suit que $\mathbb{P}(|S_n - 1| > \varepsilon)$ tend vers 0. Ainsi, par définition, $(S_n)_n$ converge en probabilité vers 1. La convergence en probabilité est une convergence sur les variables aléatoires, qui admet la propriété d'unicité presque sûre de la limite, et elle est impliquée par la convergence presque sûre. Conclusion : $S = 1$ presque sûrement.

IV Application au théorème de Rademacher pour les fonctions lipschitziennes

On va démontrer avec le théorème de convergence de martingales le théorème d'analyse suivant.

Théorème IV.1 : Théorème de Rademacher

Soit $f : [0, 1] \longrightarrow \mathbb{R}$ une fonction lipschitzienne. Alors il existe $g \in L^\infty(0, 1)$ telle que

$$\forall x \in [0, 1], f(x) = f(0) + \int_0^x g(u)\, du.$$

Remarque IV.2. Il existe une notion générale sous-jacente liée à cette question d'existence de dérivée presque partout. Il s'agit de l'absolue continuité, ou de 1-variation. On dit qu'une fonction $f : [0, 1] \longrightarrow \mathbb{R}$ est *à variation bornée*, ou est *absolument continue* si

$$\sup_{\substack{\pi=\{0=t_0<t_1<\cdots<t_n=1\} \\ \text{subdivision de } [0,1]}} \left\{ \sum_{k=0}^{n-1} |f(t_{k+1}) - f(t_k)| \right\} < \infty.$$

Dans ce cas, f est absolument continue si et seulement s'il existe $g \in L^1(0, 1)$ telle que pour tout $x \in [0, 1]$,

$$f(x) = f(0) + \int_0^x g(u)\, du.$$

On note souvent $g = f'$. ♣

Rappelons le Théorème de Doob–Dynkin : si X est une variable aléatoire et Z est une autre variable aléatoire $\sigma(X)$-mesurable alors il existe une application déterministe $g : \mathbb{R} \longrightarrow \mathbb{R}$ telle que $Z = g(X)$.

Démontrons le théorème de Rademacher.

Démonstration : L'idée est de subdiviser $[0, 1]$ en sous-intervalles dyadiques, comme dans l'introduction. Plus précisément, on considère un espace probabilisé auxiliaire $(\Omega, \mathcal{F}, \mathbb{P})$ et X une variable aléatoire suivant la loi uniforme sur $[0, 1]$: $\mathbb{P}(X \in A) = \mathrm{Leb}(A)$. L'approximation se passe alors de la sorte : pour tout $n \geqslant 1$, considérons l'approximation dyadique à l'ordre n de X :

$$X_n \overset{\text{déf.}}{=} \frac{\lfloor 2^n X \rfloor}{2^n}.$$

Alors, pour construire g, on construit la suite des taux d'accroissement de f de taille $\frac{1}{2^n}$:

$$G_n \overset{\text{déf.}}{=} \frac{f\left(X_n + \frac{1}{2^n}\right) - f(X_n)}{\frac{1}{2^n}}.$$

Le plan de démonstration est le suivant :

1. Si $\mathcal{F}_n$ est la filtration canonique associée à $(X_n)_{n\geqslant 1}$ alors $(G_n)_n$ est une martingale bornée par L donc converge dans L^1 vers $G \in L^1(\mathbb{P})$;
2. Or, par le théorème de Doob-Dynkin, il existe $g : [0, 1] \longrightarrow \mathbb{R}$ mesurable telle que $G = g(X)$. De plus, $g \in L^\infty(0, 1)$.
3. On conclut alors en télescopant la somme $f(x) - f(0)$ sur les dyadiques, modulo un reste. Chaque terme de la somme s'écrit comme une intégrale de g et le reste tend vers 0. On conclut par relation de Chasles.

1. Montrons que $(G_n)_n$ est une $(\mathcal{F}_n)_{n\geqslant 1}$-martingale. La mesurabilité est vérifiée car G_n est une fonction continue donc mesurable de

X_n. L'intégrabilité aussi, car X_n ne prend qu'un nombre fini de valeurs, donc G_n aussi. Il ne reste qu'à prouver l'égalité de martingales. Pour cela, observons d'abord que $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$ se réduit en fait qu'à $\sigma(X_n)$, car en effet, pour tout $n \geqslant 1$, on a pour $0 \leqslant k \leqslant 2^n$:

$$\begin{aligned}\left\{X_n = \frac{k}{2^n}\right\} = & \left\{X \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)\right\} \\ = & \left\{X_{n+1} = \frac{2k}{2^{n+1}}\right\} \sqcup \left\{X_{n+1} = \frac{2k+1}{2^{n+1}}\right\}.\end{aligned}$$

Donc pour tout $n \geqslant 1$, $\sigma(X_n) = \sigma(X_n, X_{n+1})$, d'où par récurrence finie $\mathcal{F}_n = \sigma(X_n)$. Montrons alors que pour tout $n \geqslant 1$:

$$\mathbb{E}[G_{n+1}|X_n] = G_n.$$

Par définition de l'espérance conditionnelle par rapport à un événement, on a la décomposition suivante :

$$\mathbb{E}[G_{n+1}|X_n] = \sum_{k=0}^{2^n} \mathbb{E}\left[G_{n+1} \,\middle|\, X_n = \frac{k}{2^n}\right] \mathbf{1}_{\{X_n = \frac{k}{2^n}\}}.$$

Rappelons que $\mathbb{E}\left[G_{n+1} \,\middle|\, X_n = \frac{k}{2^n}\right]$ n'est plus aléatoire mais est bien un réel. L'événement $\{X_n = 1\}$ est négligeable, mais on peut tout de même définir cette espérance conditionnelle *via* la notion de noyau de probabilité que je n'ai pas pu intégrer ici. En tous cas, puisque cet événement est négligeable, cela signifie que $\mathbf{1}_{\{X_n=1\}} = 0$ presque sûrement, donc on a en fait l'égalité suivante presque sûrement

$$\mathbb{E}[G_{n+1}|X_n] = \sum_{k=0}^{2^n-1} \mathbb{E}\left[G_{n+1} \,\middle|\, X_n = \frac{k}{2^n}\right] \mathbf{1}_{\{X_n = \frac{k}{2^n}\}},$$

de sorte qu'on ait un conditionnement classique par un événement non négligeable. Or, pour $0 \leqslant k \leqslant 2^n - 1$,

$$\mathbb{E}\left[G_{n+1} \,\middle|\, X_n = \frac{k}{2^n}\right] = \frac{\mathbb{E}\left[G_{n+1}\mathbf{1}_{\{X_n = \frac{k}{2^n}\}}\right]}{\mathbb{P}\left(X_n = \frac{k}{2^n}\right)}.$$

Notons provisoirement $h(x) = 2^{n+1}\left(f\left(x + \frac{1}{2^{n+1}}\right) - f(x)\right)$, de sorte que $G_{n+1} = h(X_{n+1})$. Alors en écrivant que presque sûrement, $1 = \sum_{j=0}^{2^{n+1}-1} \mathbf{1}_{\left\{X_{n+1} = \frac{j}{2^{n+1}}\right\}}$, on a alors

$$\begin{aligned}&\mathbb{E}\left[G_{n+1} \,\middle|\, X_n = \frac{k}{2^n}\right] \\ = \quad & 2^n \sum_{j=0}^{2^{n+1}-1} \mathbb{E}\left[h(X_{n+1})\mathbf{1}_{\left\{X_{n+1} = \frac{j}{2^{n+1}}\right\}}\mathbf{1}_{\{X_n = \frac{k}{2^n}\}}\right].\end{aligned}$$

Soit encore

$$\begin{aligned}&\mathbb{E}\left[G_{n+1} \,\middle|\, X_n = \frac{k}{2^n}\right] \\ = \quad & 2^n \sum_{j=0}^{2^{n+1}-1} h\left(\frac{j}{2^{n+1}}\right) \mathbb{P}\left(X_n = \frac{k}{2^n}, X_{n+1} = \frac{j}{2^{n+1}}\right).\end{aligned}$$

Ce qui s'écrit en termes de X en

$$\begin{aligned}&\mathbb{E}\left[G_{n+1} \,\middle|\, X_n = \frac{k}{2^n}\right] \\ = \quad & 2^n \sum_{j=0}^{2^{n+1}-1} h\left(\frac{j}{2^{n+1}}\right) \mathbb{P}\left(X \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right), X \in \left[\frac{j}{2^{n+1}}, \frac{j+1}{2^{n+1}}\right)\right).\end{aligned}$$

Les termes valent tous zéro sauf si $j \in \{2k, 2k+1\}$, où chaque probabilité vaut alors $\frac{1}{2^{n+1}}$. Ainsi,

$$\mathbb{E}\left[G_{n+1} \,\middle|\, X_n = \frac{k}{2^n}\right] = \frac{h\left(\frac{k}{2^n}\right) + h\left(\frac{k}{2^n} + \frac{1}{2^{n+1}}\right)}{2},$$

d'où finalement

$$\mathbb{E}\left[G_{n+1}\,|X_n\right] = \frac{h\left(X_n\right) + h\left(X_n + \frac{1}{2^{n+1}}\right)}{2}$$

En rappelant que $h(x) = 2^{n+1}\left(f\left(x + \frac{1}{2^{n+1}}\right) - f(x)\right)$, on en déduit que

$$\mathbb{E}\left[G_{n+1}\,|X_n\right] = 2^n\left(f\left(X_n + \frac{1}{2^n}\right) - f(X_n)\right) = G_n.$$

On a bien réussi à montrer que $(G_n)_n$ était une $(\mathcal{F}_n)_{n\geqslant 1}$-martingale. Or, puisque f est L-lipschitzienne, pour tout $n \geqslant 1$, $|G_n| \leqslant L$. Ainsi, $(G_n)_n$ est une martingale bornée, donc converge presque sûrement et dans L^1. Notons $G \in L^1(\mathbb{P})$ sa limite.

2. Pour tout $n \geqslant 1$, G_n est $\sigma(X_n)$-mesurable. Puisque $\sigma(X_n) \subset \sigma(X)$ (toujours *via* l'égalité $\left\{X_n = \frac{k}{2^n}\right\} = \left\{X \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)\right\}$), G_n est alors aussi $\sigma(X)$-mesurable. Puisque G est limite presque sûre de $(G_n)_{n\geqslant 1}$, G est elle aussi $\sigma(X)$-mesurable. Par le théorème de Doob-Dynkin, il suit qu'il existe $g : [0,1] \longrightarrow \mathbb{R}$ mesurable telle que $G = g(X)$ presque sûrement. Or, $|G| \leqslant L$ donc $|g(X)| \leqslant L$ presque sûrement. Ainsi,

$$\mathbb{P}(|g(X)| \leqslant L) = \mathrm{Leb}(g \leqslant L) = 1.$$

Autrement dit, $\|g\|_{L^\infty} \leqslant L$ donc $g \in L^\infty(0,1)$.

3. La variable limite $G = g(X)$ ferme la martingale $(G_n)_{n\geqslant 1}$: $G_n = \mathbb{E}[g(X)|\mathcal{F}_n]$. Pour cela, il suffit d'observer que pour tout $p \geqslant 1$:

$$\mathbb{E}\left[|G_n - \mathbb{E}[g(X)|\mathcal{F}_n]|\right] \leqslant \mathbb{E}[|F_{n+p} - g(X)|],$$

qui tend bien vers 0 quand $[p \to \infty]$ par convergence L^1. On peut alors calculer explicitement cette espérance conditionnelle.

$$G_n = \sum_{k=0}^{2^n-1} \mathbb{E}\left[g(X) \,\middle|\, X_n = \frac{k}{2^n}\right] \mathbf{1}_{\{X_n = \frac{k}{2^n}\}}.$$

Pour $0 \leqslant k \leqslant 2^n - 1$, le calcul par conditionnement classique donne

$$\begin{aligned}&\mathbb{E}\left[g(X) \,\middle|\, X_n = \frac{k}{2^n}\right] \\ = \quad & \frac{1}{\mathbb{P}\left(X_n = \frac{k}{2^n}\right)}\mathbb{E}\left[g(X)\mathbf{1}_{\{X_n = \frac{k}{2^n}\}}\right] \\ = \quad & 2^n\mathbb{E}\left[g(X)\mathbf{1}_{\{X \in [\frac{k}{2^n}, \frac{k+1}{2^n})\}}\right].\end{aligned}$$

Ainsi, puisque X suit une loi uniforme sur $[0,1]$, on en déduit que

$$\mathbb{E}\left[g(X) \,\middle|\, X_n = \frac{k}{2^n}\right] = 2^n \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} g(u)\,\mathrm{d}u.$$

Or, rappelons qu'on a

$$\mathbb{E}\left[g(X)\,|X_n\right] = G_n = 2^n\left(f\left(X_n + \frac{1}{2^n}\right) - f(X_n)\right).$$

On a alors montré que pour tout $0 \leqslant k \leqslant 2^n - 1$

$$f\left(\frac{k+1}{2^n}\right) - f\left(\frac{k}{2^n}\right) = \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} g(u)\,\mathrm{d}u.$$

On est alors enfin en mesure de conclure. Soit $x \in [0,1]$. Pour tout $n \geqslant 1$, il existe $k_n(x) = \lfloor 2^n x \rfloor$ tel que $\frac{k_n(x)}{2^n} \leqslant x < \frac{k_n(x)+1}{2^n}$. On a alors $\frac{k_n(x)}{2^n} \xrightarrow[n\to\infty]{} x$. Ainsi,

$$\begin{aligned}&f(x) - f(0) \\ = \quad & \sum_{k=0}^{k_n(x)-1}\left(f\left(\frac{k+1}{2^n}\right) - f\left(\frac{k}{2^n}\right)\right) + \left(f(x) - f\left(\frac{k_n(x)}{2^n}\right)\right).\end{aligned}$$

Avec ce qu'on vient de prouver avec la martingale $(G_n)_n$, on en déduit que

$$\begin{aligned} & f(x) - f(0) \\ = \quad & \sum_{k=0}^{k_n(x)-1} \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} g(u)\mathrm{d}u + \left(f(x) - f\left(\frac{k_n(x)}{2^n}\right)\right) \\ = \quad & \int_0^{\frac{k_n(x)}{2^n}} g(u)\ \mathrm{d}u + \left(f(x) - f\left(\frac{k_n(x)}{2^n}\right)\right). \end{aligned}$$

En faisant tendre n vers l'infini, on en déduit le théorème de Rademacher. □