COURS DE PROBABILITES (M1)

Master mathématiques, $1^{\text{ère}}$ année

Semestre 1

N. CHENAVIER

Table des matières

Chapitre 1

Quelques rappels de probabilités

Sommaire

1.1 Formalisme de Komogorov

Définition 1.1.1. Soit Ω un ensemble non vide. On appelle **tribu** sur Ω toute partie $\mathcal{A}$ de $\mathcal{P}(\Omega)$ satisfaisant les trois propriétés suivantes :

(i) $\Omega \in \mathcal{A}$ et $\emptyset \in \mathcal{A}$;

(ii) $\mathcal{A}$ est stable par passage au complémentaire : pour tout $A \in \mathcal{A}$, on a $A^c \in \mathcal{A}$;

(iii) $\mathcal{A}$ est stable par réunion dénombrable : pour toute suite (A_n) d'éléments de $\mathcal{A}$, on a

$$\bigcup_{n\geq 0} A_n \in \mathcal{A};$$

(iv) $\mathcal{A}$ est stable par intersection dénombrable : pour toute suite (A_n) d'éléments de $\mathcal{A}$, on a

$$\bigcap_{n\geq 0} A_n \in \mathcal{A}.$$

Exemple 1.1.2.

1. Soit Ω un ensemble. Les familles $\mathcal{A} = \mathcal{P}(\Omega)$ et $\mathcal{A}' = \{\emptyset, \Omega\}$ sont des tribus sur Ω.
2. Prenons $\Omega = \mathbf{R}$ et $\mathcal{A} = \{\emptyset, [0,1], \mathbf{R} \setminus [0,1], \mathbf{R}\}$. Alors $\mathcal{A}$ est une tribu.
3. Prenons $\Omega = \{1,2\}$ et $\mathcal{A} = \{\emptyset, \{1\}, \{1,2\}\}$. Alors $\mathcal{A}$ n'est pas une tribu.

Définition 1.1.3. Soit Ω un ensemble non vide et soit $\mathcal{C} \subset \mathcal{P}(\Omega)$. On appelle **tribu engendrée** par $\mathcal{C}$ la plus petite tribu contenant $\mathcal{C}$. On la note $\sigma(\mathcal{C})$.

Exemple 1.1.4. Prenons $\Omega = \mathbf{R}$ et $\mathcal{C} = \{[0,1]\}$. Alors $\sigma(\mathcal{C}) = \{\emptyset, [0,1], \mathbf{R} \setminus [0,1], \mathbf{R}\}$.

Définition 1.1.5. Soient Ω un ensemble et $\mathcal{A}$ une tribu sur Ω :
- l'ensemble Ω s'appelle l'**univers** ;
- le couple $(\Omega, \mathcal{A})$ s'appelle un **espace probabilisable** ;
- les éléments ω de Ω s'appellent des **éventualités** ;
- les éléments A de $\mathcal{A}$ s'appellent des **événements** ;
- on dit que deux événements A et B sont **incompatibles** si $A \cap B = \emptyset$.

Définition 1.1.6. Soit $(\Omega, \mathcal{A})$ un espace probabilisable. On appelle (mesure de) **probabilité** sur $(\Omega, \mathcal{A})$ toute application $\mathbb{P} : \mathcal{A} \to [0, 1]$ telle que :

(i) $\mathbb{P}(\Omega) = 1$;

(ii) pour toute suite d'événements (A_n) deux à deux disjoints (c'est-à-dire tels que $A_n \cap A_m = \emptyset$ pour tous $n \neq m$),

$$\mathbb{P}\left(\bigsqcup_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} \mathbb{P}(A_n).$$

Définition 1.1.7. Soit $(\Omega, \mathcal{A})$ un espace probabilisable et $\mathbb{P}$ une probabilité sur $(\Omega, \mathcal{A})$. Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ s'appelle un **espace probabilisé**.

Exemple 1.1.8.
1. Soit Ω un ensemble fini. Prenons $\mathcal{A} = \mathcal{P}(\Omega)$. L'application $\mathbb{P}$ définie, pour tout $A \in \mathcal{P}(\Omega)$, par $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ est une probabilité et le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace probabilisé.
2. Prenons $\Omega = \mathbf{N}$, $\mathcal{A} = \mathcal{P}(\mathbf{N})$. L'application $\mathbb{P} = \delta_5$ définie par

$$\mathbb{P}(A) = \begin{cases} 1 & \text{si } 5 \in A \\ 0 & \text{sinon} \end{cases}$$

   est une probabilité et le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace probabilisé.
3. Prenons $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}(\mathbf{R})$ et $\mathbb{P} = \lambda_{[0,1]}$. Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace probabilisé.

Définition 1.1.9. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. On dit qu'un événement $A \in \mathcal{A}$ est :
- **négligeable** si $\mathbb{P}(A) = 0$;
- **presque sûr** si $\mathbb{P}(A) = 1$.

1.2 Variables aléatoires

Loi de probabilité

Définition 1.2.1. Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(E, \mathcal{E})$ un espace probabilisable. On appelle **variable aléatoire** toute application $X : \Omega \to E$ mesurable, c'est-à-dire telle que, pour tout $B \in \mathcal{E}$, on a $X^{-1}(B) \in \mathcal{A}$.

Définition 1.2.2. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire à valeurs dans un ensemble E. La mesure de probabilité $\mathbb{P}_X : \mathcal{E} \to [0, 1]$ définie par

$$\mathbb{P}_X(B) = \mathbb{P}\left(X^{-1}(A)\right), \text{ pour tout } B \in \mathcal{E}$$

s'appelle la **loi** de X.

Par définition de $\mathbb{P}_X$, on a donc $\mathbb{P}_X(A) = \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \in A\}\right)$. Il est d'usage, en probabilités, d'alléger les écritures en prenant la notation suivante :

$$\{X \in A\} = \{\omega \in \Omega : X(\omega) \in A\} = X^{-1}(A).$$

Par exemple, si X est à valeurs réelles, l'événement $\{X \leq x\}$ est égal à $\{\omega \in \Omega : X(\omega) \leq x\}$. On écrit également $\mathbb{P}(X \in A)$ au lieu de $\mathbb{P}(\{X \in A\})$ en omettant les parenthèses. En d'autres termes, on a

$$\begin{aligned} \mathbb{P}_X(A) &= \mathbb{P}\left(X^{-1}(A)\right) \\ &= \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \in A\}\right) \\ &= \mathbb{P}\left(\{X \in A\}\right) \\ &= \mathbb{P}\left(X \in A\right). \end{aligned}$$

Remarque 1.2.3. En pratique, les lois de variables aléatoires que nous considérons sont de deux types.

- Ou bien elles sont *discrètes*. Dans ce cas, la loi de X est de la forme $\mathbb{P}_X = \sum_{i \geq 0} p_i \delta_{x_i}$, où (x_i) est une suite d'éléments de E et (p_i) une suite de nombres réels dans $[0,1]$ tels que $\sum_{i=1}^{\infty} p_i = 1$. En particulier, $\mathbb{P}(X = x_i) = p_i$ et plus généralement

$$\mathbb{P}(X \in B) = \sum_{i=1}^{\infty} \mathbf{1}_B(x_i) p_i$$

  pour tout $B \in \mathcal{E}$.
- Ou bien elles *possèdent une densité* f_X par rapport à la mesure de Lebesgue de $\mathbf{R}^d$. Dans ce cas, la loi de X est de la forme $\mathrm{d}\mathbb{P}_X(x) = f_X(x)\mathrm{d}x$. En particulier,

$$\mathbb{P}(X \in B) = \int_E \mathbf{1}_B(x) f_X(x) \mathrm{d}x.$$

Définition 1.2.4. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire réelle.[1] On appelle **fonction de répartition** de X la fonction $F_X : \mathbf{R} \to [0,1]$ définie par

$$F_X(x) = \mathbb{P}(X \leq x), \ x \in \mathbf{R}.$$

Remarque 1.2.5. Si X est une variable aléatoire réelle de densité f_X, alors la fonction de répartition F_X est de la forme

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t, \ x \in \mathbf{R}.$$

Proposition 1.2.6. *La fonction de répartition caractérise la loi : si X et Y sont deux variables aléatoires, alors $X \overset{loi}{=} Y$ si et seulement si $F_X = F_Y$.*

Espérance

Dans cette sous-section, on se limite au cas des variables aléatoires réelles.

1. En d'autres termes, X est une variable aléatoire à valeurs dans $\mathbf{R}$.

Notation. Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $p > 0$. On désigne par $L^p(\Omega, \mathcal{A}, \mathbb{P})$ l'ensemble des variables aléatoires $X : \Omega \to \mathbf{R}$ telles que X^p est intégrable c'est-à-dire telles que

$$\int_\Omega |X(\omega)|^p \mathrm{d}\mathbb{P}(\omega) < \infty\}.$$

Lorsque cela ne prête pas à confusion, on écrira simplement L^p au lieu de $L^p(\Omega, \mathcal{A}, \mathbb{P})$. En particulier, L^1 est l'espace des variables aléatoires intégrables et L^2 l'espace des variables aléatoires de carré intégrable.

Définition 1.2.7. Soit X une variable aléatoire réelle.

(i) Si $X \geq 0$ p.s., on appelle **espérance** de X la quantité (positive)

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

(ii) Si $X \in L^1$, on appelle **espérance** de X le nombre réel

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

Exemple 1.2.8. Soit A un événement. Alors $\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A)$.

On peut étendre la notion d'espérance à des variables aléatoires X à valeurs dans $\mathbf{C}$ en posant à nouveau $\mathbb{E}[X] = \int_\Omega X(\omega)\mathrm{d}\mathbb{P}(\omega)$ dès lors que $\int_\Omega |X(\omega)| < \infty$.

Théorème 1.2.9. *(théorème de transfert) Soit $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ une variable aléatoire et $\varphi : (E, \mathcal{E}) \to (\mathbf{R}, \mathcal{B}(\mathbf{R}))$ une fonction mesurable. Si $\varphi(X) \in L^1$, alors*

$$\mathbb{E}[\varphi(X)] = \int_E \varphi(x)\mathrm{d}\mathbb{P}_X(x).$$

En particulier, si X est une variable aléatoire réelle, alors $\mathbb{E}[X] = \int_{\mathbf{R}} x\mathrm{d}\mathbb{P}_X(x)$.

Remarque 1.2.10. Reprenons les deux cas de figures considérées dans la remarque 1.2.3.

- Si X est une variable aléatoire discrète, à valeurs dans un ensemble dénombrable $E = \{x_i, i \geq 1\} \subset \mathbf{R}$, alors

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i).$$

- Si la loi de X possède une densité f_X par rapport à la mesure de Lebesgue sur $\mathbf{R}$, alors

$$\mathbb{E}[X] = \int_{\mathbf{R}} x f_X(x)\mathrm{d}x.$$

Définition 1.2.11. Soit X une variable aléatoire réelle dans L^p, $p \in \mathbf{N}^*$. On appelle **moment d'ordre** p le nombre réel $\mathbb{E}[X^p]$.

Définition 1.2.12. Soit X une variable aléatoire réelle dans L^2. On appelle :

- **variance** de X le nombre réel positif

$$\mathbb{V}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right];$$

- **écart-type** de X le nombre réel positif

$$\sigma(X) = \sqrt{\mathbb{V}[X]}.$$

Définition 1.2.13. Soient X et Y deux variables aléatoires réelles dans L^2. On appelle **covariance** de X et de Y le nombre réel

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y].$$

Définition 1.2.14. Soit $X : \Omega \to \mathbf{R}^d$ un vecteur aléatoire. On appelle **fonction caractéristique** de X l'application $\varphi_X : \mathbf{R}^d \to \mathbf{C}$ définie, pour tout $t \in \mathbf{R}^d$, par

$$\varphi_X(t) = \mathbb{E}\left[e^{i\langle t,X\rangle}\right] = \int_{\mathbf{R}^d} e^{i\langle t,x\rangle} \mathrm{d}\mathbb{P}_X(x).$$

Proposition 1.2.15. *La fonction caractéristique caractérise la loi : si X et Y sont deux variables aléatoires, alors $X \overset{loi}{=} Y$ si et seulement si $\varphi_X = \varphi_Y$.*

1.3 Conditionnement et indépendance

Probabilités conditionnelles

Définition 1.3.1. Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et A et B deux événements tels que $\mathbb{P}(A) \neq 0$. On appelle probabilité de B **conditionnellement** à A (ou probabilité de B sachant A) la quantité définie par :

$$\mathbb{P}(B|A) = \mathbb{P}_A(B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

Proposition 1.3.2. *(Formule des probabilités totales) Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et (A_n) une partition de Ω telle que $\mathbb{P}(A_n) \neq 0$ pour tout n. Alors, pour tout $B \in \mathcal{A}$,*

$$\mathbb{P}(B) = \sum_{n=1}^{\infty} \mathbb{P}(B \cap A_n) = \sum_{n=1}^{\infty} \mathbb{P}(B|A_n)\,\mathbb{P}(A_n).$$

Proposition 1.3.3. *(Formule de Bayes) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $B \in \mathcal{A}$ un événement tel que $\mathbb{P}(B) \neq 0$.*

(i) Pour tout événement $A \in \mathcal{A}$ tel que $\mathbb{P}(A \neq 0)$, on a

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\,\mathbb{P}(A)}{\mathbb{P}(B)}.$$

(ii) Soit (A_n) une partition de Ω telle que $\mathbb{P}(A_n) \neq 0$ pour tout n. Alors, pour tout $j \geq 1$,

$$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(B|A_j)\,\mathbb{P}(A_j)}{\sum_{n\geq 1} \mathbb{P}(B|A_n)\,\mathbb{P}(A_n)}.$$

Indépendance d'événements, de tribus et de variables aléatoires

Définition 1.3.4. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. On dit que deux événements $A, B \in \mathcal{A}$ sont **indépendants** si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$.

Définition 1.3.5. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(A_i)_{i \in I}$ une famille d'événements. On dit que les événements $(A_i)_{i \in I}$ sont :
- **deux à deux indépendants** si A_i et A_j sont indépendants pour tous $i, j \in I$, $i \neq j$;
- **mutuellement indépendants** si pour toute famille finie $J \subset I$,

$$\mathbb{P}\left(\bigcap_{i \in J} A_j\right) = \prod_{i \in J} \mathbb{P}(A_i).$$

L'indépendance mutuelle entraîne l'indépendance deux à deux. L'exemple suivant montre que la réciproque n'est pas vraie.

Exemple 1.3.6. On lance deux pièces au hasard et on note P (resp. F) lorsqu'on on obtient pile (resp. face). L'univers est $\Omega = \{(P, P), (P, F), (F, P), (F, F)\}$. On considère les événements suivants :

$$A = \{\text{le premier lancer est pile}\} = \{(P, P), (P, F)\};$$

$$B = \{\text{le second lancer est face}\} = \{(P, F), (F, F)\};$$

$$C = \{\text{le résultat est le même aux deux lancers}\} = \{(P, P), (F, F)\}.$$

On a $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$ et $\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{4}$. Les événements A, B, C sont donc deux à deux indépendants. En revanche, ils ne sont pas mutuellement indépendants puisque $\mathbb{P}(A \cap B \cap C) = 0 \neq \frac{1}{8} = \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C)$.

Définition 1.3.7. Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $\mathcal{B}_1, \ldots, \mathcal{B}_n$ des sous-tribus de $\mathcal{A}$. On dit que les tribus $\mathcal{B}_1, \ldots, \mathcal{B}_n$ sont **indépendantes** si, pour tous $B_1 \in \mathcal{B}_1, \ldots, B_n \in \mathcal{B}_n$, on a

$$\mathbb{P}(B_1 \cap \cdots \cap B_n) = \mathbb{P}(B_1) \times \cdots \times \mathbb{P}(B_n).$$

Définition 1.3.8. Soit $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ une variable aléatoire. On appelle **tribu engendrée** par X la tribu

$$\sigma(X) = \{X^{-1}(B) : B \in \mathcal{E}\}.$$

Définition 1.3.9. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et soient $X_1 \ldots, X_n$ des variables aléatoires définies sur $(\Omega, \mathcal{A})$ et à valeurs dans $(E_1, \mathcal{E}_1), \ldots, (E_n, \mathcal{E}_n)$ respectivement. On dit que les variables aléatoires $X_1, \ldots, X_n$ sont **indépendantes** si les tribus $\sigma(X_1), \ldots, \sigma(X_n)$ sont indépendantes, c'est-à-dire si

$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \times \cdots \times \mathbb{P}(X_n \in B_n)$$

pour tous $B_1 \in \mathcal{E}_1, \ldots, B_n \in \mathcal{E}_n$.

Proposition 1.3.10. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et soient $X_1 \ldots, X_n$ des variables aléatoires définies sur $(\Omega, \mathcal{A})$ et à valeurs dans $(E_1, \mathcal{E}_1), \ldots, (E_n, \mathcal{E}_n)$ respectivement. Les propriétés suivantes sont équivalentes.*

(i) $X_1, \ldots, X_n$ sont indépendantes.

(ii) *Pour tous* $B_1 \in \mathcal{E}_1, \ldots, B_n \in \mathcal{E}_n$*, on a*

$$\mathbb{P}\left(X_1 \in B_1, \ldots, X_n \in B_n\right) = \mathbb{P}\left(X_1 \in B_1\right) \times \cdots \times \mathbb{P}\left(X_n \in B_n\right).$$

(iii) *Pour toutes fonctions mesurables* $f_1, \ldots, f_n$ *de* E_i *dans* $\mathbf{R}$*,* $i \leq n$*, on a*

$$\mathbb{E}\left[\prod_{i=1}^{n} |f_i(X_i)|\right] = \prod_{i=1}^{n} \mathbb{E}\left[|f_i(X_i)|\right].$$

(iv) $\mathbb{P}_{(X_1,\ldots,X_n)} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$.

1.4 Quelques théorèmes généraux

Théorèmes d'interversion limite-intégrale

Théorème 1.4.1. *(théorème de convergence monotone) Soit* (X_n) *une suite de variables aléatoires réelles positives. Supposons que* X_n *tend vers* X *p.s. et que, pour presque tout* $\omega \in \Omega$*, la suite* $(X_n(\omega))$ *est croissante. Alors*

$$\lim_{n\to\infty} \mathbb{E}\left[X_n\right] = \mathbb{E}\left[X\right].$$

Théorème 1.4.2. *(théorème de convergence dominée) Soit* (X_n) *une suite de variables aléatoires réelles. Supposons que* X_n *tend vers* X *p.s. et qu'il existe une variable aléatoire* $Z \in L^1$ *telle que* $|X_n| \leq Z$ *p.s. pour tout entier* n*. Alors*

$$\lim_{n\to\infty} \mathbb{E}\left[X_n\right] = \mathbb{E}\left[X\right].$$

Inégalités

Théorème 1.4.3. *(inégalité de Jensen) Soit* X *une variable aléatoire réelle et* $\varphi : \mathbf{R} \to \mathbf{R}$ *une fonction convexe. Si* $X \in L^1$ *et* $\varphi(X) \in L^1$*, alors*

$$\mathbb{E}\left[\varphi(X)\right] \geq \varphi(\mathbb{E}\left[X\right]).$$

Proposition 1.4.4. *(Inégalité de Markov) Soit* X *une variable aléatoire réelle positive. Alors, pour tout* $a > 0$*,*

$$\mathbb{P}\left(X \geq a\right) \leq \frac{\mathbb{E}\left[X\right]}{a}.$$

Proposition 1.4.5. *(Inégalité de Tchebychev) Soit* X *une variable aléatoire réelle dans* L^2*. Alors, pour tout* $a > 0$*,*

$$\mathbb{P}\left(|X - \mathbb{E}\left[X\right]| \geq a\right) \leq \frac{\mathbb{V}\left[X\right]}{a^2}.$$

Lemmes de Borel-Cantelli

Définition 1.4.6. Soit (A_n) une suite d'événements.

(i) On appelle **limite supérieure** des A_n l'événement

$$\limsup A_n = \bigcap_{n\geq 1} \bigcup_{k\geq n} A_k.$$

(ii) On appelle **limite inférieure** des A_n l'événement

$$\liminf A_n = \bigcup_{n\geq 1} \bigcap_{k\geq n} A_k.$$

D'un point de vue probabilise, la limite supérieure s'interprète comme étant l'ensemble des éventualités ω pour lesquels les événements A_k se réalisent infiniment souvent. La limite inférieure s'interprète comme étant l'ensemble des éventualités ω pour lesquels, à partir d'un certain rang, tous les événements A_k se réalisent.

Proposition 1.4.7. *Soit (A_n) une suite d'événements. Alors*

$$\limsup \mathbf{1}_{A_n} = \mathbf{1}_{\limsup A_n} \quad \text{et} \quad \liminf \mathbf{1}_{A_n} = \mathbf{1}_{\liminf A_n}.$$

Théorème 1.4.8. *(Premier lemme de Borel-Cantelli) Soit (A_n) une suite d'événements telle que $\sum_{n\geq 1} \mathbb{P}(A_n) < \infty$. Alors $\mathbb{P}(\limsup A_n) = 0$.*

Théorème 1.4.9. *(Second lemme de Borel-Cantelli) Soit (A_n) une suite d'événements indépendants telle que $\sum_{n\geq 1} \mathbb{P}(A_n) = \infty$. Alors $\mathbb{P}(\limsup A_n) = 1$.*

Chapitre 2

Convergence de suites de variables aléatoires

Sommaire

2.1 Modes de convergence

On distingue deux types de convergences : celles de type "spatial" où l'univers Ω joue un rôle essentiel, et la convergence "en loi" où seules les lois des variables aléatoires importent.

Dans ce qui suit, on considère un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Les variables aléatoires sont définies sur Ω et à valeurs dans un espace d'état $E \subset \mathbf{R}^d$, $d \geq 1$. L'espace $\mathbf{R}^d$ est muni de sa norme euclidienne, notée $|\cdot|$, et de sa tribu borélienne $\mathcal{B}(\mathbf{R}^d)$.

Convergences spatiales

Définition 2.1.1. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit (X_n) une suite de variables aléatoires définies sur Ω et à valeurs dans $\mathbf{R}^d$, $d \geq 1$. On dit que :

- la suite de variables aléatoires (X_n) **converge presque sûrement** vers la variable aléatoire X si l'événement $\{X_n$ converge vers $X\}$ est presque sûr, c'est-à-dire si

$$\mathbb{P}\left(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\right) = 1,$$

  on note alors $X_n \overset{p.s.}{\to} X$;
- la suite de variables aléatoires (X_n) **converge en probabilité** vers la variable aléatoire X si, pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(|X_n - X| > \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 0,$$

  on note alors $X_n \overset{\mathbb{P}}{\to} X$;

- la suite de variables aléatoires (X_n) **converge dans** L^p, $p > 0$, vers la variable aléatoire X si, pour tout entier n, on a $X_n \in L^p$, $X \in L^p$ et

$$\mathbb{E}\left[|X_n - X|^p\right] \underset{n\to\infty}{\longrightarrow} 0,$$

on note alors $X_n \overset{L^p}{\to} X$.

Exemple 2.1.2. Soit (X_n) une suite de variables aléatoires indépendantes telles que $\mathbb{P}(X_n = 1) = \frac{1}{n}$ et $\mathbb{P}(X_n = 0) = 1 - \frac{1}{n}$. La suite (X_n) converge en probabilité vers 0, converge dans L^p, $p > 0$, vers 0 mais ne converge pas presque sûrement vers 0.

Remarque 2.1.3. L'ensemble $\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$ est bien un événement puisque

$$\begin{aligned}\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} &= \{\omega \in \Omega : \forall \varepsilon > 0, \exists N > 0, \forall n \geq N, |X_n(\omega) - X(\omega)| \leq \varepsilon\} \\ &= \bigcap_{\varepsilon \in \mathbf{Q}_+^*} \bigcup_{N \geq 0} \bigcap_{n \geq N} \{\omega \in \Omega : |X_n(\omega) - X(\omega)| \leq \varepsilon\},\end{aligned}$$

qui est mesurable en tant qu'intersections et réunions dénombrables d'ensembles mesurables.

Proposition 2.1.4. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit (X_n) une suite de variables aléatoires définies sur Ω et à valeurs dans $\mathbf{R}^d$, $d \geq 1$.*

(i) Si $X_n \overset{L^p}{\to} X$ alors $X_n \overset{\mathbb{P}}{\to} X$.

(ii) Si $X_n \overset{p.s.}{\to} X$ alors $X_n \overset{\mathbb{P}}{\to} X$.

(iii) Si $X_n \overset{\mathbb{P}}{\to} X$ alors il existe une sous-suite $(X_{n(k)})$ telle que $X_{n(k)} \overset{p.s.}{\to} X$.

Les réciproques du résultat ci-dessus ne sont en général pas vraies (voir l'exemple 2.1.2).

Convergence en loi

Définition 2.1.5.
- Soit (μ_n) une suite de mesures de probabilités sur $\mathbf{R}^d$. On dit que (μ_n) **converge étroitement** vers μ si, pour toute fonction $f : \mathbf{R}^d \to \mathbf{R}$ continue bornée, on a

$$\int_{\mathbf{R}^d} f(x)\mathrm{d}\mu_n(x) \underset{n\to\infty}{\longrightarrow} \int_{\mathbf{R}^d} f(x)\mathrm{d}\mu(x).$$

On note alors $\mu_n \Rightarrow \mu$.
- Soit (X_n) une suite de variables aléatoires à valeurs dans $\mathbf{R}^d$. On dit que (X_n) **converge en loi** vers X si $(\mathbb{P}_{X_n})$ converge étroitement vers $\mathbb{P}_X$ c'est-à-dire si, pour toute fonction continue bornée $f : \mathbf{R}^d \to \mathbf{R}$, on a

$$\mathbb{E}\left[f(X_n)\right] \underset{n\to\infty}{\longrightarrow} \mathbb{E}\left[f(X)\right].$$

Exemple 2.1.6. La loi uniforme sur $\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\}$ converge étroitement vers la loi uniforme sur $[0, 1]$. En effet, pour toute fonction $f : [0, 1] \to \mathbf{R}$ continue bornée, on a

$$\mathbb{E}\left[f(X_n)\right] = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \mathbb{P}(X_n = k) = \frac{1}{n+1} \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \underset{n\to\infty}{\longrightarrow} \int_0^1 f(x)\mathrm{d}x = \mathbb{E}\left[f(X)\right].$$

Remarque 2.1.7.
1. La convergence en loi est fondamentalement différente des trois types de convergence spatiale. En effet, pour la convergence en loi, seules les lois comptent, tandis que pour les convergences de type spatial, il faut prendre en compte les variables aléatoires et l'espace Ω joue un rôle prépondérant.

2. Si (μ_n) est une suite de mesures de probabilités qui converge étroitement vers μ, alors μ est une mesure de probabilité.

Le fait que (X_n) converge en loi vers X n'entraîne pas nécessairement que $\mathbb{P}(X_n \in B) \underset{n\to\infty}{\longrightarrow} \mathbb{P}(X \in B)$ pour tout $B \in \mathcal{B}(\mathbf{R}^d)$ comme en témoigne le contre-exemple ci-dessous.

Contre-exemple 2.1.8. Soit (a_n) une suite (déterministe) strictement croissante de nombres réels convergeant vers un nombre réel a. Posons $X_n = a_n$ et $X = a$. Alors (X_n) converge en loi vers X. En revanche, si l'on prend $B =]-\infty, a[$, alors $\mathbb{P}(X_n \in B) = \mathbb{P}(\{\omega \in \Omega : a_n \in]-\infty, a[\}) = 1$ mais $\mathbb{P}(X \in B) = \mathbb{P}(\{\omega \in \Omega : a \in]-\infty, a[\}) = 0$.

Proposition 2.1.9. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit (X_n) une suite de variables aléatoires définies sur Ω et à valeurs dans $\mathbf{R}^d$, $d \geq 1$. Si $X_n \xrightarrow{\mathbb{P}} X$ alors $X_n \Rightarrow X$.*

La réciproque du résultat ci-dessus n'est en général pas vraie comme en témoigne le contre-exemple ci-dessous.

Contre-exemple 2.1.10. Prenons $(\Omega, \mathcal{A}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda_{[0,1]})$. Considérons [1] la suite de variables aléatoires (X_n) définies, pour tout entier n et pour tout $\omega \in [0,1]$, par :

$$\begin{cases} X_n(\omega) = 1 & \text{si } \omega \in [0, 1/2]; \\ X_n(\omega) = 0 & \text{si } \omega \in]1/2, 1]. \end{cases}$$

Posons X la variable aléatoire définie, pour tout pour tout $\omega \in [0,1]$, par :

$$\begin{cases} X(\omega) = 0 & \text{si } \omega \in [0, 1/2]; \\ X_n(\omega) = 1 & \text{si } \omega \in]1/2, 1]. \end{cases}$$

Alors (X_n) converge en loi vers X mais ne converge pas en probabilité vers X.

Le résultat ci-dessous donne une réciproque partielle à la proposition 2.1.9.

Proposition 2.1.11. *Si (X_n) est une suite de variables aléatoires qui converge en loi vers une constante $c \in \mathbf{R}^d$, alors (X_n) converge en probabilité vers X.*

Caractérisations équivalentes de la convergence en loi

On termine cette section avec quelques caractérisations de la convergence en loi. Le premier concerne les variables aléatoires réelles.

Proposition 2.1.12. *Soit (X_n) (resp. X) une suite de variables aléatoires réelles (resp. une variable aléatoire réelle). Pour tout entier n, notons $F_{X_n}(x) = \mathbb{P}(X_n \leq x)$ et $F_X(x) = \mathbb{P}(X \leq x)$, $x \in \mathbf{R}$, les fonctions de répartition. Les assertions suivantes sont équivalentes :*

(i) (X_n) converge en loi vers X ;

(ii) pour tout point $x \in \mathbf{R}$ tel que $\mathbb{P}_X(\{x\}) = 0$, on a $F_{X_n}(x) \underset{n\to\infty}{\longrightarrow} F_X(x)$;

(iii) pour tout point $x \in \mathbf{R}$ en lequel F_X est continue, on a $F_{X_n}(x) \underset{n\to\infty}{\longrightarrow} F_X(x)$.

Le résultat ci-dessous donne une caractérisation de la convergence en loi pour des variables aléatoires discrètes.

1. De façon plus condensée, on a $X_n = \mathbf{1}_{[0,1/2]}$

Proposition 2.1.13. *Soit* (X_n) *une suite de variables aléatoires à valeurs dans* $\mathbf{Z}$*. Alors* (X_n) *converge en loi vers* X *si et seulement si* $\mathbb{P}(X_n = k) \underset{n\to\infty}{\longrightarrow} \mathbb{P}(X = k)$ *pour tout entier* $k \in \mathbf{Z}$*.*

Le résultat ci-dessous donne une caractérisation de la convergence en loi, en termes de fonctions caractéristiques, pour des variables aléatoires à valeurs dans $\mathbf{R}^d$.

Théorème 2.1.14. *(théorème de Lévy) Soit* (X_n) *une suite de variables aléatoires à valeurs dans* $\mathbf{R}^d$*. Pour tout entier* n*, notons* $\varphi_{X_n}(t) = \mathbb{E}\left[e^{iX_n t}\right]$ *la fonction caractéristique de* X_n*.*

(i) Si (X_n) *converge en loi vers une variable aléatoire* X *alors* $\varphi_{X_n}(t) \underset{n\to\infty}{\longrightarrow} \varphi_X(t)$ *pour tout* $t \in \mathbf{R}^d$*.*

(ii) Supposons qu'il existe une fonction $\varphi : \mathbf{R}^d \to \mathbf{C}$*, continue en 0, telle que* $\varphi_{X_n}(t) \underset{n\to\infty}{\longrightarrow} \varphi(t)$ *pour tout* $t \in \mathbf{R}^d$*. Alors il existe une variable aléatoire* X *telle que* X_n *converge en loi vers* X *et* $\varphi_X = \varphi$*.*

Corollaire 2.1.15. *Une suite de variables aléatoires* (X_n)*, à valeurs dans* $\mathbf{R}^d$*, converge vers une variable aléatoire* X *si et seulement si* φ_{X_n} *converge simplement vers* φ_X*.*

Exemple 2.1.16. Soit (X_n) une suite de variables aléatoires telle que X_n suive la loi binomiale $\mathcal{B}(n, p_n)$, avec $np_n \underset{n\to\infty}{\longrightarrow} \lambda$, $\lambda > 0$. La fonction caractéristique de X_n est

$$\varphi_{X_n}(t) = \left((1 - p_n) + p_n e^{it}\right)^n, \ t \in \mathbf{R}.$$

Pour tout $t \in \mathbf{R}$, on a

$$\lim_{n\to\infty} \varphi_{X_n}(t) = e^{\lambda(e^{it}-1)}.$$

Le terme de droite est la fonction caractéristique (évaluée en t) de la loi de Poisson de paramètre λ. En d'autres termes, si X désigne une variable aléatoire suivant la loi $\mathcal{P}(\lambda)$, on a $\varphi_X(t) = e^{\lambda(e^{it}-1)}$. Par le théorème 2.1.14, on en déduit que X_n converge en loi vers X, et par conséquent que la loi binomiale $\mathcal{B}(n, p_n)$ converge vers $\mathcal{P}(\lambda)$, avec $\lambda = \lim_{n\to\infty} np_n$.

L'hypothèse de continuité en 0 dans l'assertion *(ii)* du théorème 2.1.14 est importante. Voici un contre-exemple.

Contre-exemple 2.1.17. Soit (X_n) une suite de variables aléatoires réelles telles que $X_n \sim \mathcal{N}(0, \sigma_n^2)$ avec $\sigma_n \underset{n\to\infty}{\longrightarrow} \infty$. Pour tout $t \in \mathbf{R}$, on a

$$\varphi_{X_n}(t) = e^{-\frac{\sigma_n^2 t^2}{2}}.$$

En particulier, φ_{X_n} converge simplement vers la fonction φ définie par $\varphi(t) = 0$ si $t \neq 0$ et $\varphi(0) = 1$. Cependant, la suite de variables alétoires (X_n) ne converge pas en loi car φ n'est pas continue en 0.

2.2 Théorèmes limites pour des sommes de variables aléatoires indépendantes

Loi des grands nombres

Dans cette sous-section, on se limite aux variables aléatoires réelles. Etant donnée une suite (X_n) de variables aléatoires réelles, on pose

$$\overline{X_n} = \frac{X_1 + \cdots + X_n}{n}.$$

La quantité $\overline{X_n}$ s'appelle la **moyenne empirique**.

Théorème 2.2.1. *(loi faible des grands nombres) Soit (X_n) une suite de variables aléatoires réelles dans L^2. Supposons que, pour tout entier i, on ait $\mathbb{E}[X_i] = m$, qu'il existe $c > 0$ tel que $\mathbb{V}[X_i] \leq c$ et que $cov(X_i, X_j) = 0$ pour tout $i \neq j$. Alors*

$$\overline{X_n} \xrightarrow{L^2} m.$$

Remarque 2.2.2.
1. Sous les mêmes hypothèses que le théorème 2.2.1, on a $\overline{X_n} \xrightarrow{\mathbb{P}} m$.
2. En particulier, si (X_n) est une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) de carré intégrable, alors la loi faible des grands nombres est valable.

Théorème 2.2.3. *(loi forte des grands nombres) Soit (X_n) une suite de variables aléatoires réelles i.i.d. dans L^1. Notons $m = \mathbb{E}[X_1]$. Alors*

$$\overline{X_n} \xrightarrow{p.s.} m.$$

Exemple 2.2.4. Soit (X_n) une suite de variables aléatoires i.i.d. suivant la loi de Bernoulli de paramètre p. Alors $\overline{X_n} \xrightarrow{p.s.} p$.

L'hypothèse d'intégrabilité des X_i est nécessaire. Voici un contre-exemple où la moyenne empirique $\overline{X_n} = \frac{X_1+\cdots+X_n}{n}$ ne converge pas vers une quantité déterministe.

Contre-exemple 2.2.5. Soit (X_n) une suite de variables aléatoires i.i.d. de loi de Cauchy.[2] La fonction caractéristique de X_i, $i \geq 1$, est égale à $\varphi_{X_i}(t) = e^{-|t|}$ pour tout $t \in \mathbf{R}$. La fonction caractéristique de $\overline{X_n}$ est égale, pour tout $t \in \mathbf{R}$, à

$$\varphi_{\overline{X_n}}(t) = \mathbb{E}\left[e^{i \cdot \frac{X_1+\cdots+X_n}{n} t}\right] = \left(\mathbb{E}\left[e^{\frac{iX_1 t}{n}}\right]\right)^n = e^{-|t|}.$$

En particulier, $\overline{X_n}$ suit la loi de Cauchy et ne converge pas vers une quantité déterministe.

La loi des grands nombres affirme que la moyenne empirique converge vers la moyenne théorique mais sans en garantir la vitesse. Cette dernière est traitée dans le théorème central limite.

Théorème central limite

Théorème 2.2.6. *(théorème central limite univarié) Soit (X_n) une suite de variables aléatoires réelles i.i.d.. Supposons que $X_1 \in L^2$ et notons $m = \mathbb{E}[X_1]$ et $\sigma^2 = \mathbb{V}[X_1]$. Alors*

$$\left(\overline{X_n} - m\right) \times \sqrt{\frac{n}{\sigma^2}} \Rightarrow \mathcal{N}(0,1).$$

Exemple 2.2.7. Soit (X_n) une suite de variables aléatoires i.i.d. suivant la loi de Bernoulli de paramètre p, avec $p \in]0,1[$. Alors $\left(\overline{X_n} - p\right) \times \sqrt{\frac{n}{p(1-p)}} \Rightarrow \mathcal{N}(0,1)$.

2. On rappelle qu'une variable aléatoire X suit la loi de Cauchy si sa loi est de densité $f_X(x) = \frac{1}{\pi(1+x^2)}$.

Chapitre 3

Vecteurs gaussiens

Sommaire

3.1 Définition et propriétés

Proposition 3.1.1. *Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire de $\mathbf{R}^d$. Les assertions suivantes sont équivalentes.*

(i) Pour tout $(\lambda_1, \ldots, \lambda_d) \in \mathbf{R}^d$, la variable aléatoire $\lambda_1 X_1 + \cdots \lambda_d X_d$ suit une loi normale.

(ii) La fonction caractéristique de X, notée $\phi_X : \mathbf{R}^d \to \mathbf{C}$, est donnée par

$$\phi_X(t) = e^{i\langle t,m\rangle} \cdot e^{-\frac{1}{2}\langle t,\Lambda t\rangle},$$

où $m = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])$ et $\Lambda = (cov(X_i, X_j))_{1\leq i,j\leq d}$.

(iii) Il existe un vecteur $m \in \mathbf{R}^d$, une matrice $c \in \mathcal{M}_d(\mathbf{R})$ et un vecteur aléatoire $Y = (Y_1, \ldots, Y_d)$, où les Y_i sont i.i.d. et suivent la loi $\mathcal{N}(0,1)$, tels que $X \overset{loi}{=} m + cY$.

Dans l'assertion *(iii)*, l'égalité est à comprendre d'un point de vue matriciel (écrire les vecteurs X et Y en colonnes).

Définition 3.1.2. Dans le cas où l'une des assertions de la proposition ci-dessus est satisfaite, on dit que $X = (X_1, \ldots, X_d)$ est un vecteur gaussien de moyenne m et de matrice de covariance Λ. On note alors $X \sim \mathcal{N}(m, \Lambda)$.

Exemple 3.1.3. 1. Soient Y_1, Y_2, Y_3 des variables aléatoires indépendantes suivant la loi $\mathcal{N}(0,1)$. Posons $X = (Y_1, Y_2, Y_3)$. Alors X est un vecteur gaussien de moyenne $m = (0,0,0)$ et de matrice de covariance $\Lambda = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. D'un point de vue matriciel, on a

$$X = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}.$$

2. Soient Y_1, Y_2, Y_3 des variables aléatoires indépendantes suivant la loi $\mathcal{N}(0,1)$. Posons $X = (1 - Y_1 + Y_2, Y_2)$. Alors X est un vecteur gaussien de moyenne $m = (1,0)$ et de matrice de covariance $\Lambda = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. D'un point de vue matriciel, on a

$$X = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}.$$

Remarque 3.1.4. 1. Deux vecteurs gaussiens sont identiques en loi si et seulement si ils ont même moyenne et même matrice de covariance.

2. La matrice de covariance Λ est symétrique positive.
3. Toute constante sur $\mathbf{R}^d$ est considérée comme un vecteur gaussien de variance infinie.
4. Les matrices c et Λ sont reliées par l'identité $\Lambda = cc^T$. Plus précisément, si $X = (X_1, \dots, X_d) = m + cY$, où $Y = (Y_1, \dots, Y_d)$ est un d-uplet de variables aléatoires i.i.d. suivant la loi $\mathcal{N}(0,1)$, alors $\mathbb{E}[X] := (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d]) = m$ et $\Lambda := (\mathrm{cov}(X_i, X_j))_{1 \le i,j \le d} = cc^T$.
5. Plus généralement, si $X \sim \mathcal{N}(m, \Lambda)$, alors $AX + B \sim \mathcal{N}(Am + B, A\Lambda A^T)$.
6. La matrice c dans l'expression $X \overset{loi}{=} m + cY$ n'est pas unique : il peut exister des matrices c_1 et c_2 telles que $X \overset{loi}{=} m + c_1 Y \overset{loi}{=} m + c_2 Y$. Par exemple, si $X = Y$ est un d-uplet de variables aléatoires i.i.d. suivant la loi $\mathcal{N}(0,1)$ et si c_2 est une matrice orthogonale (c'est-à-dire telle que $cc^T = I_d$), alors $X \overset{loi}{=} Y \overset{loi}{=} c_2 Y$.

Lorsque $X = (X_1, \dots, X_d)$ est un vecteur gaussien, ses marginales suivent une loi normale. A contrario, il existe des vecteurs aléatoires dont les marginales suivent des lois normales mais qui ne sont pas gaussiens. Voici un contre-exemple.

Contre-exemple 3.1.5. Soit $Y \sim \mathcal{N}(0,1)$ et $a > 0$. Posons $Y' = Y\mathbf{1}_{|Y| \le a} - Y\mathbf{1}_{|Y| > a}$. La variable aléatoire Y' suit la loi $\mathcal{N}(0,1)$. Cependant, le vecteur (Y, Y') n'est pas gaussien puisque $Y + Y' = 2Y\mathbf{1}_{|Y| \le a}$ est presque sûrement compris entre $-a$ et a.

Proposition 3.1.6. *Soit X un vecteur gaussien suivant la loi $\mathcal{N}(m, \Lambda)$, où Λ est symétrique positive. Si Λ est de plus définie, alors X est à densité sur $\mathbf{R}^d$ de densité*

$$\begin{aligned} f_X : \mathbf{R}^d &\to \mathbf{R} \\ x &\mapsto \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Lambda)}} e^{-\frac{1}{2}\langle (x-m), \Lambda^{-1}(x-m) \rangle}. \end{aligned}$$

Remarque 3.1.7. L'hypothèse *définie* est nécessaire. Si on prend, par exemple, X une variable aléatoire normale centrée réduite, alors le vecteur $(X', Y') = (X, 0)$ est un vecteur gaussien de moyenne $m = (0,0)$ et de matrice de covariance $\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ mais ne possède pas de densité.

Théorème 3.1.8. *(théorème d'orthogonalité des vecteurs gaussiens) Soit $X = (X_1, \dots, X_d)$ un vecteur gaussien de matrice de covariance Λ et soient $\lambda, \lambda' \in \mathbf{R}^d$. Notons $Y = \langle \lambda, X \rangle = \sum_{i=1}^d \lambda_i X_i$ et $Y' = \langle \lambda', X \rangle = \sum_{i=1}^d \lambda'_i X_i$. Les assertions suivantes sont équivalentes.*

(i) Y *et* Y' *sont indépendantes.*

(ii) $cov(Y, Y') = 0$.

(iii) $\lambda^T \Lambda \lambda' = 0$.

(iv) λ *et* λ' *sont orthogonaux pour* $\langle \cdot, \cdot \rangle_\Lambda$.

Remarque 3.1.9. 1. L'implication "Y indépendant de $Y' \Rightarrow \mathrm{cov}(Y, Y') = 0$" est toujours vraie (pour des variables aléatoires pas nécessairement gaussiennes).

2. Le sens intéressant est "$\mathrm{cov}(Y, Y') = 0 \Rightarrow Y$ indépendant de Y'". Cette implication est vraie dans le cas gaussien mais fausse dans le cas général.
3. Si Λ est définie alors $\langle \cdot, \cdot \rangle_\Lambda$ est un produit scalaire.

Plus généralement, on a le théorème suivant.

Théorème 3.1.10. *Soit $X = (X_1, \ldots, X_d)$ un vecteur gaussien de matrice de covariance Λ. Soient $\lambda_1, \ldots, \lambda_k, \lambda'_1, \ldots, \lambda'_{k'} \in \mathbf{R}^d$. Pour tout $i \leq k$ (resp. $i' \leq k'$), posons $Y_i = \langle \lambda_i, X \rangle$ (resp. $Y'_i = \langle \lambda'_i, X \rangle$). Les propriétés suivantes sont équivalentes.*

(i) Les vecteurs $(Y_1, \ldots, Y_k)$ et $(Y'_1, \ldots, Y'_k)$ sont indépendants.

(ii) Pour tout $i \leq k$, $j \leq k'$, on a $\langle \lambda_i, \lambda'_j \rangle_\Lambda = 0$.

(iii) Les espaces vectoriels vect$(\lambda_1, \ldots, \lambda_k)$ et vect$(\lambda'_1, \ldots, \lambda'_{k'})$ sont orthogonaux pour $\langle \cdot, \cdot \rangle_\Lambda$.

3.2 Théorèmes limites multidimensionnels

Théorème 3.2.1. *(loi des grands nombres multidimensionnelle) Soit (X_n) une suite de vecteurs aléatoires i.i.d., où $X_i = (X_i^{(1)}, \ldots, X_i^{(d)})$. Si X_1 est intégrable (c'est-à-dire si $\mathbb{E}\left[|X_1^{(j)}|\right] < \infty$ pour tout $1 \leq j \leq d$), alors*

$$\frac{S_n}{n} \overset{p.s.}{\to} \mathbb{E}\left[X_1\right],$$

où $S_n = X_1 + \cdots + X_n$ et $\mathbb{E}\left[X_1\right] = \left(\mathbb{E}\left[X_1^{(1)}\right], \ldots, \mathbb{E}\left[X_1^{(d)}\right]\right)$.

Théorème 3.2.2. *(théorème central limite multidimensionnel) Soit (X_n) une suite de vecteurs aléatoires i.i.d. de carré intégrable (c'est-à-dire tel que $\mathbb{E}\left[|X_1^{(j)}|^2\right] < \infty$ pour tout $1 \leq j \leq d$). Posons $m = \mathbb{E}\left[X_1\right]$ et $\Lambda = \left(cov(X_1^{(i)}, X_1^{(j)})\right)_{1 \leq i,j \leq d}$. Alors*

$$\left(\frac{S_n}{n} - m\right) \cdot \sqrt{n} \Rightarrow \mathcal{N}(0, \Lambda).$$

Chapitre 4

Espérance conditionnelle

Sommaire

4.1 Préambule

Dans ce qui suit, les variables aléatoires sont toutes définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

Rappelons que, si A désigne un événement de probabilité non nulle et B un événement, on appelle **probabilité de** B **sachant** A le nombre réel

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

La quantité

$$\begin{aligned} \mathbb{P}_B : \mathcal{A} &\to [0,1] \\ A &\mapsto \mathbb{P}(B|A) \end{aligned}$$

est une mesure de probabilité sur $\mathcal{A}$, appelée **probabilité conditionnelle**.

Dans ce chapitre, on souhaite définir une notion d'espérance de variable aléatoire conditionnellement à une autre variable aléatoire. Pour cela, on commence par définir une notion d'espérance conditionnellement à un événement.

Définition 4.1.1. Soit X une variable aléatoire réelle dans L^1 et A un événement de probabilité non nulle. On appelle **espérance de** X **sachant** A le nombre réel suivant :

$$\mathbb{E}[X|A] = \int_\Omega X(\omega) \mathrm{d}\mathbb{P}_A(\omega) = \frac{\mathbb{E}[X \mathbf{1}_A]}{\mathbb{P}(A)}.$$

En particulier, si Y est une variable aléatoire à valeurs dans un ensemble dénombrable E (quitte à restreindre E, on peut supposer que $\mathbb{P}(Y = y) \neq 0$ pour tout $y \in E$) et si $X \in L^1$, on a, pour tout $y \in E$,

$$\mathbb{E}[X|Y = y] = \frac{\mathbb{E}[X\mathbf{1}_{Y=y}]}{\mathbb{P}(Y = y)}.$$

On définira ci-dessous une notion d'espérance conditionnelle de X sachant Y comme étant une variable aléatoire $\sigma(Y)$-mesurable.

4.2 Cas d'une variable aléatoire discrète

Dans cette section, on se limite au cas où Y est une variable aléatoire à valeurs dans un espace dénombrable E.

Définition 4.2.1. Soit X une variable aléatoire réelle dans L^1 et Y une variable aléatoire à valeurs dans un espace dénombrable E. On appelle **espérance conditionnelle de X sachant Y** la variable aléatoire notée $\mathbb{E}[X|Y] : \Omega \to \mathbf{R}$ définie, pour tout $\omega \in \Omega$, par :

$$\mathbb{E}[X|Y](\omega) = \begin{cases} \mathbb{E}[X|Y = Y(\omega)] & \text{si } \mathbb{P}(Y = Y(\omega)) \neq 0 \\ 0 & \text{sinon.} \end{cases}$$

Remarque 4.2.2.

1. La valeur 0 est arbitraire car le “sinon” n'arrive presque sûrement jamais. En d'autres termes, $\mathbb{E}[X|Y]$ est définie à un ensemble de mesure nulle près.
2. La variable aléatoire $\mathbb{E}[X|Y]$ est une fonction de Y dans le sens où $\mathbb{E}[X|Y](\omega) = \varphi(Y(\omega))$, avec

$$\varphi(y) = \begin{cases} \mathbb{E}[X|Y = y] & \text{si } \mathbb{P}(Y = y \neq 0) \\ 0 & \text{sinon.} \end{cases}$$

   En particulier, la variable aléatoire $\mathbb{E}[X|Y]$ est $\sigma(Y)$-mesurable.

Exemple 4.2.3. Soit $X : \Omega \to \{1, \dots, 6\}$ une variable aléatoire suivant la loi uniforme sur $\{1, \dots, 6\}$. Soit Y la variable aléatoire définie par :

$$\begin{aligned} Y : \Omega &\to \{0, 1\} \\ \omega &\mapsto 1 \text{ si } X(\omega) \text{ est pair} \\ \omega &\mapsto 0 \text{ si } X(\omega) \text{ est impair.} \end{aligned}$$

L'espérance conditionnelle de X sachant Y est donnée par :

$$\mathbb{E}[X|Y](\omega) = \begin{cases} 4 & \text{si } Y(\omega) = 1 \\ 3 & \text{si } Y(\omega) = 0. \end{cases}$$

Proposition 4.2.4. *Soient $X \in L^1$ et Y une variable aléatoire à valeurs dans un ensemble dénombrable E. Alors*

(i) $\mathbb{E}[|\mathbb{E}[X|Y]|] \leq \mathbb{E}[|X|]$.

(ii) Pour toute variable aléatoire réelle Z bornée, $\sigma(Y)$-mesurable, on a

$$\mathbb{E}[Z\mathbb{E}[X|Y]] = \mathbb{E}[ZX].$$

L'assertion *(i)* garantit que $\mathbb{E}[X|Y]$ est intégrable (et $\sigma(Y)$-mesurable).

Remarque 4.2.5. La variable aléatoire $\mathbb{E}[X|Y]$ ne dépend que de la tribu $\sigma(Y)$. En d'autres termes, si Y et Y' sont des variables aléatoires telles que $\sigma(Y) = \sigma(Y')$, alors $\mathbb{E}[X|Y] = \mathbb{E}[X|Y']$ p.s..

4.3 Cas général

Définition et premières propriétés

Théorème 4.3.1. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $\mathcal{B} \subset \mathcal{A}$ une sous-tribu de $\mathcal{A}$ et soit $X \in L^1$ une variable aléatoire réelle. Alors il existe une variable aléatoire réelle notée $\mathbb{E}[X|\mathcal{B}] \in L^1$, unique (à un ensemble de mesure nulle près), satisfaisant la propriété suivante : pour toute variable aléatoire Z, bornée et $\mathcal{B}$-mesurable, on a*

$$\mathbb{E}[Z\mathbb{E}[X|\mathcal{B}]] = \mathbb{E}[ZX].$$

En particulier, si l'on prend $Z = \mathbf{1}_B$ avec $B \in \mathcal{B}$, on a $\mathbb{E}[\mathbf{1}_B\mathbb{E}[X|\mathcal{B}]] = \mathbb{E}[\mathbf{1}_B X]$. A fortiori, $\mathbb{E}[\mathbb{E}[X|\mathcal{B}]] = \mathbb{E}[X]$.

Exemple 4.3.2. Soit X une variable aléatoire dans L^1.

1. Si $\mathcal{B}$ est la tribu grossière, c'est-à-dire $\mathcal{B} = \{\emptyset, \Omega\}$, alors $\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X]$ p.s..
2. Si $\mathcal{B}$ est la tribu totale, c'est-à-dire $\mathcal{B} = \mathcal{A}$, alors $\mathbb{E}[X|\mathcal{B}] = X$ p.s..

Définition 4.3.3. (i) On appelle la variable aléatoire $\mathbb{E}[X|\mathcal{B}]$ **l'espérance conditionnelle de X par rapport à la tribu $\mathcal{B}$.**

(ii) Etant donnée une variable aléatoire Y, on appelle **espérance conditionnelle de X sachant Y** la variable aléatoire définie par $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$.

Proposition 4.3.4. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $\mathcal{B} \subset \mathcal{A}$ une sous-tribu de $\mathcal{A}$ et soit $X \in L^1$ une variable aléatoire réelle.*

(i) L'espérance conditionnelle est linéaire.

(ii) L'espérance conditionnelle est positive : si $X \geq 0$ p.s., alors $\mathbb{E}[X|\mathcal{B}] \geq 0$ p.s..

(iii) Si X est $\mathcal{B}$-mesurable, alors $\mathbb{E}[X|\mathcal{B}] = X$ p.s..

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

(v) $|\mathbb{E}[X|\mathcal{B}]| \leq \mathbb{E}[|X||\mathcal{B}]$.

Remarque 4.3.5. Si X est une variable aléatoire réelle *positive*, on peut définir l'espérance conditionnelle de la façon suivante :

$$\mathbb{E}[X|\mathcal{B}] = \lim_{n\to\infty} \mathbb{E}[\max(X, n)|\mathcal{B}].$$

Il s'agit, dans ce cas, d'une variable aléatoire à valeurs dans $\overline{\mathbf{R}_+} = [0, \infty]$. De façon équivalente, on peut définir l'espérance conditionnelle comme étant l'unique variable aléatoire dans $\overline{\mathbf{R}_+}$ qui est $\mathcal{B}$-mesurable et telle que, pour toute variable aléatoire Z positive et $\mathcal{B}$-mesurable, on a $\mathbb{E}[Z\mathbb{E}[X|\mathcal{B}]] = \mathbb{E}[ZX]$.

Tous les grands théorèmes de convergence de la théorie de la mesure restent vrais pour l'espérance conditionnelle.

Théorèmes généraux

Théorème 4.3.6. *(théorème de convergence monotone) Soit* (X_n) *une suite de variables aléatoires réelles et positives p.s.. Si* X_n *converge vers* X *p.s. et si, pour tout* $\omega \in \Omega$, *la suite* $(X_n(\omega))$ *est croissante, alors*

$$\mathbb{E}\left[X_n|\mathcal{B}\right] \overset{p.s.}{\to} \mathbb{E}\left[X|\mathcal{B}\right].$$

Théorème 4.3.7. *(lemme de Fatou) Soit* (X_n) *une suite de variables aléatoires réelles et positives p.s.. Alors*

$$\liminf_{n\to\infty} \mathbb{E}\left[X_n|\mathcal{B}\right] \geq \mathbb{E}\left[\liminf_{n\to\infty} X_n|\mathcal{B}\right] \quad p.s..$$

Théorème 4.3.8. *(théorème de convergence dominée) Soit* (X_n) *une suite de variables aléatoires réelles. Supposons qu'il existe une variable aléatoire* $Z \in L^1$ *telle que* $|X_n| \leq Z$ *p.s., pour tout entier* n. *Si* X_n *converge vers* X *p.s., alors*

$$\mathbb{E}\left[X_n|\mathcal{B}\right] \overset{p.s.}{\to} \mathbb{E}\left[X|\mathcal{B}\right].$$

Théorème 4.3.9. *(Iinégalité de Jensen) Soit* $\varphi : \mathbf{R} \to \mathbf{R}_+$ *une fonction convexe. Soit* $X \in L^1$ *une variable aléatoire telle que* $\varphi(X) \in L^1$. *Alors*

$$\mathbb{E}\left[\varphi(X)|\mathcal{B}\right] \geq \varphi(\mathbb{E}\left[X|\mathcal{B}\right]) \quad p.s..$$

L'inégalité de Jensen est encore vraie si φ est positive et si X est seulement dans L^1.

4.4 Cas du conditionnement dans L^2

On considère toujours un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et une sous-tribu $\mathcal{B}$ de $\mathcal{A}$. Dans cette section, on s'intéresse au cas particulier où la variable aléatoire X est dans $L^2 = L^2(\Omega, \mathcal{A}, \mathbb{P})$. Dans ce cas-là, on peut donner une interprétation géométrique de l'espérance conditionnelle en termes de projection orthogonale ; l'idée étant de remarquer que l'espace $L^2(\Omega, \mathcal{B}, \mathbb{P})$ est un sous-espace vectoriel fermé de l'espace de Hilbert $L^2(\Omega, \mathcal{A}, \mathbb{P})$.

Théorème 4.4.1. *Soit* $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ *et soit* $\mathcal{B}$ *une sous-tribu de* $\mathcal{A}$. *Alors* $\mathbb{E}\left[X|\mathcal{B}\right]$ *est le projeté orthogonal de* X *sur* $L^2(\Omega, \mathcal{B}, \mathbb{P})$. *En particulier,*

$$||\mathbb{E}\left[X|\mathcal{B}\right] - X||_{L^2} = \inf_{Z \in L^2(\Omega, \mathcal{B}, \mathbb{P})} ||Z - X||_{L^2}.$$

Informellement, il faut donc voir l'espérance conditionnelle $\mathbb{E}\left[X|\mathcal{B}\right]$ comme étant la meilleure approximation de X (pour la norme L^2) par une variable aléatoire $\mathcal{B}$-mesurable.

4.5 Propriétés importantes de l'espérance conditionnelle

Proposition 4.5.1. *Soit* $\mathcal{B}$ *une sous-tribu de* $\mathcal{A}$. *Soient* X *et* Y *deux variables aléatoires réelles telles que* $X, Y \in L^1$ *ou telles que* $X, Y \geq 0$ *p.s.. Si* Y *est* $\mathcal{B}$*-mesurable, alors*

$$\mathbb{E}\left[XY|\mathcal{B}\right] = Y\mathbb{E}\left[X|\mathcal{B}\right].$$

Proposition 4.5.2. *Soient* $\mathcal{B}_1$ *et* $\mathcal{B}_2$ *deux sous-tribus de* $\mathcal{A}$ *telles que* $\mathcal{B}_1 \subset \mathcal{B}_2$. *Alors, pour toute variable aléatoire* $X \in L^1$, *on a :*

(i) $\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{B}_1\right]|\mathcal{B}_2\right] = \mathbb{E}\left[X|\mathcal{B}_1\right]$.

(ii) $\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{B}_2\right]|\mathcal{B}_1\right] = \mathbb{E}\left[X|\mathcal{B}_1\right]$.

Si l'on applique *(ii)* au cas particulier où $\mathcal{B}_1 = \{\emptyset, \Omega\}$, on retrouve bien le fait que $\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{B}_2\right]\right] = \mathbb{E}\left[X\right]$.

Proposition 4.5.3. *Soient $\mathcal{B}_1$ et $\mathcal{B}_2$ deux sous-tribus de $\mathcal{A}$. Les propriétés suivantes sont équivalentes.*

(i) Les tribus $\mathcal{B}_1$ et $\mathcal{B}_2$ sont indépendantes.

(ii) Pour toute variable aléatoire $\mathcal{B}_1$-mesurable X, dans L^1 ou positive p.s., on a $\mathbb{E}\left[X|\mathcal{B}_2\right] = \mathbb{E}\left[X\right]$ p.s..

Remarque 4.5.4. Deux variables aléatoires X et Y sont indépendantes si et seulement si, pour toute fonction h mesurable bornée, on a $\mathbb{E}\left[h(X)|Y\right] = \mathbb{E}\left[h(X)\right]$.

Chapitre 5

Martingales

Sommaire

Comme dans les chapitres précédents, on se place sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

5.1 Martingales : définition et premières propriétés

Définition 5.1.1. (i) On appelle **filtration** toute suite croissante $(\mathcal{F}_n)$ de sous-tribus de $\mathcal{A}$.

(ii) On appelle **processus adapté** à la filtration $(\mathcal{F}_n)$ toute suite de variables aléatoires (X_n), avec $X_n : \Omega \to \mathbf{R}$, telle que X_n soit $\mathcal{F}_n$-mesurable pour tout entier n.

Il faut voir n comme un temps (discret) et $\mathcal{F}_n$ comme l'information disponible en n.

Définition 5.1.2. Soit (X_n) une suite de variables aléatoires réelles. On appelle **filtration naturelle** de (X_n) la filtration $(\sigma(X_1, \ldots, X_n))$.

En particulier, une suite de variables aléatoires (X_n) est adaptée à sa filtration naturelle.

Définition 5.1.3. Soit $(\mathcal{F}_n)$ une filtration.

(i) On appelle **$\mathcal{F}_n$-martingale** tout processus (M_n), adapté à la filtration $(\mathcal{F}_n)$, tel que $M_n \in L^1$ et $\mathbb{E}\left[M_{n+1}|\mathcal{F}_n\right] = M_n$ p.s. pour tout entier n.

(ii) On appelle **sous-martingale** (resp. **sur-martingale**) tout processus (M_n), adapté à la filtration $(\mathcal{F}_n)$, tel que $M_n \in L^1$ et $\mathbb{E}\left[M_{n+1}|\mathcal{F}_n\right] \geq M_n$ (resp. $\mathbb{E}\left[M_{n+1}|\mathcal{F}_n\right] \leq M_n$) p.s. pour tout entier n.

Proposition 5.1.4. *Soit (M_n) une $\mathcal{F}_n$-martingale.*

(i) Pour tous $m \geq n$, on a $\mathbb{E}\left[M_m|\mathcal{F}_n\right] = M_n$.

(ii) Pour toute fonction convexe f telle que $f(M_n) \in L^1$, le processus $(f(M_n))$ est une $(\mathcal{F}_n)$-sous-martingale.

(iii) Pour tout $n \geq 0$, on a $\mathbb{E}[M_n] = \mathbb{E}[M_0]$.

Une conséquence immédiate de *(ii)* est que, si (M_n) est une martingale telle que $M_n \in L^p$, $p \geq 1$, alors le processus $(|M_n|^p)$ est une sous-martingale.

Exemple 5.1.5.
1. *Marche aléatoire simple.* Soit (X_n) une suite de variables aléatoires i.i.d. telle que $\mathbb{E}[X_1] = 0$. Posons $M_0 = 0$, $M_n = X_1 + \cdots + X_n$ et $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Le processus (M_n) est une $\mathcal{F}_n$-martingale.
2. *Martingale exponentielle.* Soit (X_n) une suite i.i.d. telle qu'il existe $\alpha \in \mathbf{R}$ tel que $\mathbb{E}\left[e^{\alpha X_1}\right] < \infty$. Posons $\Phi(\alpha) = \log(\mathbb{E}\left[e^{\alpha X_1}\right])$, $M_0 = 1$, $M_n = e^{\alpha(X_1+\cdots+X_n)-n\Phi(\alpha)}$ et $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Alors (M_n) est une $\mathcal{F}_n$-martingale.
3. Soit $X \in L^1$ et $(\mathcal{F}_n)$ une filtration. Posons $M_n = \mathbb{E}[X|\mathcal{F}_n]$. Le processus (M_n) est une $\mathcal{F}_n$-martingale.
4. Soit (M_n) une $\mathcal{F}_n$-martingale. Soit (H_n) un processus [1] tel que :
   - H_n est borné p.s. ;
   - H_n est $\mathcal{F}_{n-1}$ mesurable pour tout n.

   Posons $Z_n = \sum_{i=1}^n H_i(M_i - M_{i-1})$. Alors (Z_n) est une $\mathcal{F}_n$-martingale.

5.2 Temps d'arrêt et théorème d'arrêt

Définition 5.2.1. Soit $(\mathcal{F}_n)$ une filtration. On appelle **temps d'arrêt** toute variable aléatoire $T : \Omega \to \mathbf{N} \cup \{\infty\}$ telle que, pour tout entier n, l'événement $\{T \leq n\}$ est $\mathcal{F}_n$-mesurable.

De façon équivalente, une variable aléatoire T est un temps d'arrêt si l'événement $\{T = n\}$ est $\mathcal{F}_n$-mesurable.

Exemple 5.2.2. Soit (X_n) un processus (à valeurs dans $\mathbf{R}$) adapté à une filtration $(\mathcal{F}_n)$. Soit $a \in \mathbf{R}$ et $A \in \mathcal{B}(\mathbf{R})$. Les variables aléatoires T et T' définies par

$$T(\omega) = \inf\{n \geq 1 : X_n(\omega) = a\} \quad \text{et} \quad T'(\omega) = \inf\{n \geq 1 : X_n(\omega) \in A\}$$

sont des temps d'arrêt.

Considérons le problème suivant. Un joueur possède une certaine somme, disons 50 euros et lance une pièce. Il gagne un euro s'il tombe sur pile et perd un euro s'il tombe sur face. Le joueur lance la pièce jusqu'à ce qu'il ait 0 euro ou 100 euros. Ici, X_n désigne la fortune du joueur au temps n. La durée du jeu est alors un temps d'arrêt : il s'agit de la variable aléatoire T' considérée dans l'exemple ci-dessus, avec $A = \{0, 100\}$.

Théorème 5.2.3. *(théorème d'arrêt) Soit $(\mathcal{F}_n)$ une filtration. Soit (M_n) une $\mathcal{F}_n$-martingale et T un temps d'arrêt pour $\mathcal{F}_n$. Posons $Z_n = M_{n\wedge T}$ pour tout $n \geq 0$ (c'est-à-dire $Z_n(\omega) = M_{n\wedge T(\omega)}(\omega)$). Alors (Z_n) est une $\mathcal{F}_n$-martingale.*

Ci-dessus, nous avons utilisé la notation $a \wedge b = \min\{a, b\}$.

Corollaire 5.2.4. *Soit $(\mathcal{F}_n)$ une filtration. Soit (M_n) une $\mathcal{F}_n$-martingale et T un temps d'arrêt.*

(i) Si T est borné, c'est-à-dire s'il existe $k \geq 0$ tel que $T \leq k$ p.s., alors $\mathbb{E}[M_T] = \mathbb{E}[M_0]$.

(ii) Si $M_{n\wedge T}$ est dominé, c'est-à-dire s'il existe une variable aléatoire $Z \in L^1$ tel que $|M_{n\wedge T}| \leq Z$ p.s.), alors $\mathbb{E}[M_T] = \mathbb{E}[M_0]$.

1. On dit du processus (H_n) qu'il est prévisible (voir Section 5.3

Exemple 5.2.5. On s'intéresse au problème de la ruine du joueur. Soient $k \in]0, N[$ et (X_n) une suite de variables aléatoires i.i.d. telle que $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$. Posons $S_n = k+X_1+\cdots+X_n$ et $T = \inf\{n \geq 0 : S_n \in \{0, N\}\}$. On souhaite calculer la probabilité que le joueur perde c'est-à-dire $\mathbb{P}(S_T = 0)$. Nous avons vu que (S_n) est une martingale adaptée à la filtration $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ et que T est un temps d'arrêt. On peut facilement montrer que $T < \infty$ p.s.. Par le théorème 5.2.3, le processus $(S_{n\wedge T})$ est une martingale. De plus, $0 \leq S_{n\wedge T} \leq N$ p.s.. En particulier, $|S_{n\wedge T}|$ est bornée donc dominée. Par le corollaire 5.2.4, on a $\mathbb{E}[S_T] = \mathbb{E}[S_0] = k$. Comme $\mathbb{E}[S_0] = N\mathbb{P}(S_T = N)$, on en déduit que $\mathbb{P}(S_T = N) = \frac{k}{N}$ et que $\mathbb{P}(S_T = 0) = 1 - \frac{k}{N}$.

5.3 Décomposition de Doob

Définition 5.3.1. Soit $(\mathcal{F}_n)$ une filtration.

(i) Un processus (X_n) est dit **prévisible** si X_0 est $\mathcal{F}_0$-mesurable et si, pour tout entier $n \geq 1$, la variable aléatoire X_n est $\mathcal{F}_{n-1}$-mesurable.

(ii) Un processus $A = (A_n)$ est dit **croissant prévisible** s'il est prévisible, si $A_0 = 0$ et s'il vérifie, pour tout entier $n \geq 1$,

$$0 \leq A_n \leq A_{n+1} < \infty \text{ p.s..}$$

Lorsque (A_n) est un processus croissant prévisible, sa limite existe p.s. dans $\overline{R_+}$. On la note A_∞.

Théorème 5.3.2. *(décomposition de Doob) Soit $X = (X_n)$ une sous-martingale.*

1. *Il existe une martingale intégrable $M = (M_n)$ et un processus croissant prévisible $A = (A_n)$ uniques tels que $X = M + A$.*
2. *On a l'équivalence*

$$\sup_{n\geq 0} \mathbb{E}\left[X_n^+\right] < \infty \Longleftrightarrow \sup_{n\geq 0} \mathbb{E}\left[|M_n|\right] < \infty \text{ et } A_\infty \in L^1.$$

Exemple 5.3.3. Soit (X_n) une suite i.i.d. de variables aléatoires intégrables de moyenne $m > 0$. Posons $S_0 = m$, $S_n = X_1 + \cdots + X_n$ et $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. La suite (S_n) est une $\mathcal{F}_n$ martingale puisque $\mathbb{E}[S_{n+1}|\mathcal{F}_n] = S_n + m$. La décomposition de Doob de (S_n) est donnée par $A_n = nm$ et $M_n = S_n - nm$.

5.4 Théorèmes de convergence

Théorème 5.4.1. *(théorème de convergence p.s.) Soit (M_n) une $\mathcal{F}_n$-martingale. Si (M_n) est bornée dans L^1, c'est-à-dire si $\sup_{n\geq 0} \mathbb{E}[|M_n|] < \infty$, alors il existe une variable aléatoire $M_\infty \in L^1$ telle que $M_n \xrightarrow{p.s.} M_\infty$.*

Remarque 5.4.2. L'hypothèse "bornée" est importante. Si l'on prend, par exemple, $S_n = X_1 + \cdots + X_n$ où les X_i sont i.i.d. et tels que $\mathbb{P}(X_n = -1) = \mathbb{P}(X_n = 1) = \frac{1}{2}$, alors S_n ne converge pas.

Exemple 5.4.3. Soit (X_n) une suite de variables aléatoires i.i.d. telle que $\mathbb{P}(X_i = -1) = \mathbb{P}(X_i = 1) = \frac{1}{2}$.

1. Posons $M_n = \sum_{i=1}^n \frac{X_i}{i^2}$. D'après le théorème 5.4.1, la martingale (M_n) converge p.s. vers une variable aléatoire M_∞.

2. Posons $S_n = k + X_1 + \cdots + X_n$, où $k \in \mathbf{N}$, et $T = \inf\{n \geq 0 : S_n = 0\}$. Pour tout $n \geq 0$, notons $M_n = S_{n \wedge T}$. La martingale (M_n) est p.s. positive. De plus, $\mathbb{E}[|M_n|] = \mathbb{E}[M_n] = \mathbb{E}[M_0] = k$. Donc (M_n) est bornée dans L^1. D'après le théorème 5.4.1, la martingale (M_n) converge p.s. vers une variable aléatoire M_∞. Comme M_n est à valeurs dans $\mathbf{Z}$, la suite (M_n) est nécessairement stationnaire. Une conséquence est que, p.s., à partir d'un certain rang, on a $M_n = 0$. En particulier, $T < \infty$ p.s..

Théorème 5.4.4. *(théorème de convergence dans L^p) Soit $p > 1$. Soit (M_n) une $\mathcal{F}_n$-martingale. Si (M_n) est p.s. bornée dans $L^p(\Omega, \mathcal{A}, \mathbb{P})$, c'est-à-dire si $\sup_{n \geq 0} \mathbb{E}[|M_n|^p] < \infty$, alors il existe une variable aléatoire $M_\infty \in L^p$ telle que $M_n \xrightarrow{p.s.} M_\infty$ et $M_n \xrightarrow{L^p} M_\infty$. De plus, $M_n = \mathbb{E}[M_\infty | \mathcal{F}_n]$ p.s..*

Théorème 5.4.5. *Soit (X_n) une $\mathcal{F}_n$-sous-martingale telle que $\sup_{n \geq 0} \mathbb{E}[X_n^+] < \infty$. Alors (X_n) converge p.s..*

Corollaire 5.4.6. *Soit (X_n) une $\mathcal{F}_n$-sur-martingale positive. Alors (X_n) converge p.s. vers une variable aléatoire X_∞ à valeurs dans $\overline{\mathbf{R}_+}$. De plus, pour tout $n \geq 0$, on a $X_n \geq \mathbb{E}[X_\infty | \mathcal{F}_n]$ p.s..*