Pré-rentrée du MVA
ENS Cachan

Rappels de probabilités

Raphaël Deswarte
deswarte@cmap.polytechnique.fr

Raphaël Forien
raphael.forien@cmap.polytechnique.fr

22-23 septembre 2016

Chapitre 1

Rappels de Probabilités

1.1 Définitions, exemples

Espace de probabilité

On modélise une expérience aléatoire par un triplet $(\Omega, \mathcal{A}, \mathbb{P})$, où Ω est l'ensemble des "issues" possibles, $\mathcal{A} \subset \mathcal{P}(\Omega)$ est l'ensemble des événements et $\mathbb{P} : \mathcal{A} \to [0,1]$ donne la probabilité de chaque événement. Pour que les opération mathématiques correspondant au calcul des probabilités soient bien définies, on demande que Ω, $\mathcal{A}$ et $\mathbb{P}$ vérifient certaines hypothèses.

Définition 1.1.1. *On dit que $\mathcal{A} \subset \mathcal{P}(\Omega)$ est une* ***tribu*** *(σ-field en anglais) si*
- $\Omega \in \mathcal{A}$,
- *pour tout $A \in \mathcal{A}$, $A^c \in \mathcal{A}$ (stabilité par passage au complémentaire),*
- *pour toute suite dénombrable $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{A}$, $\cup_n A_n \in \mathcal{A}$.*

La tribu correspond à l'information que l'on a sur l'expérience. Par exemple si $\Omega = \{1,2,3,4,5,6\}$ (comme c'est le cas pour un lancer de dé), on peut vérifier que $\mathcal{A} = \{\emptyset, \Omega, \{1,3,5\}, \{2,4,6\}\}$ est une tribu. Dans ce cas, la seule information que l'on "retient" est la parité du résultat.

EXERCICE 1 - Montrer qu'une tribu est stable par intersection dénombrable.

Définition 1.1.2. *Soit $(\Omega, \mathcal{A})$ un espace muni d'une tribu. On dit que $\mathbb{P} : \mathcal{A} \to \mathbb{R}_+ \cup \{+\infty\}$ est une* ***mesure*** *si*
- $\mathbb{P}(\emptyset) = 0$,
- *pour toute suite dénombrable $(A_n)_{n\in\mathbb{N}}$ d'éléments deux à deux disjoints, $\mathbb{P}(\cup_n A_n) = \sum_n \mathbb{P}(A_n)$.*

Si de plus $\mathbb{P}(\Omega) = 1$, on dit que $\mathbb{P}$ est une ***probabilité****.*

EXERCICE 2 - Soit $\mathbb{P}$ une mesure de probabilité sur $(\Omega, \mathcal{A})$. Montrer les propriétés suivantes.

1. Si $A \subset B$, $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A) \geq 0$.
2. Si $A_n \in \mathcal{A}$ pour tout $n \in \mathbb{N}$, $\mathbb{P}(\cup_n A_n) \leq \sum_n \mathbb{P}(A_n)$.
3. Si $(A_n)_{n\in\mathbb{N}}$ est une suite croissante d'éléments de $\mathcal{A}$, $\mathbb{P}(\cup_n A_n) = \lim \uparrow \mathbb{P}(A_n)$. Montrer la propriété analogue pour une suite décroissante.

Définition 1.1.3. *Un* ***espace de probabilité*** *est un triplet $(\Omega, \mathcal{A}, \mathbb{P})$, où $\mathcal{A}$ est une tribu sur Ω et $\mathbb{P}$ est une probabilité sur $(\Omega, \mathcal{A})$.*

Exemple.
- *Si Ω est fini ou dénombrable, on peut toujours prendre $\mathcal{A} = \mathcal{P}(\Omega)$, et, pour tout $A \in \mathcal{A}$,*
$$\mathbb{P}(A) = \sum_{\omega \in A} p_\omega,$$
où $0 \leq p_\omega \leq 1$ et $\sum_{\omega\in\Omega} p_\omega = 1$.
- *$\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ (Définition 1.1.4 ci-dessous) et*
$$\mathbb{P}(]a,b[) = \int_a^b p(x)\,\mathrm{d}\,x,$$
où $0 \leq p(x) \leq 1$ et $\int_{-\infty}^{+\infty} p(x)\,\mathrm{d}\,x = 1$.

On vérifie aisément (exercice !) qu'une instersection quelconque (*i.e.* même infinie non dénombrable) de tribus est encore une tribu. Cela justifie la définition suivante.

Définition 1.1.4 (Tribu des boréliens)**.** *Notons $\mathcal{O}$ l'ensemble des ouverts de $\mathbb{R}$. La* ***tribu des boréliens*** *est la plus petite tribu sur $\mathbb{R}$ qui contient les ouverts,*
$$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{O}) := \bigcap_{\mathcal{O}\subset\mathcal{A} \text{ tribu}} \mathcal{A}.$$

On dit aussi que la tribu des boréliens est la tribu engendrée par les ouverts.

Remarque. *Pour définir une probabilité sur la tribu des boréliens, il suffit de la définir sur la classe des ouverts de $\mathbb{R}$. C'est la conséquence d'un résultat de théorie de la mesure que l'on ne détaille pas ici. Ainsi la définition de $\mathbb{P}$ dans le deuxième exemple ci-dessus est suffisante pour caractériser $\mathbb{P}$.*

Théorème 3. *Il existe une unique mesure λ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que, pour tout $a \leq b \in \mathbb{R}$,*

$$\lambda(]a, b[) = b - a. \tag{1.1}$$

*On appelle λ la **mesure de Lebesgue**.*

Définition 1.1.5 (Tribu produit). *Soient $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$ et $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$ deux espaces de probabilité, alors $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$ est un espace de probabilité, où*
- $\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(\{A \times B, A \in \mathcal{A}_1, B \in \mathcal{A}_2\})$,
- *pour* $A \in \mathcal{A}_1$, $B \in \mathcal{A}_2$, $\mathbb{P}_1 \otimes \mathbb{P}_2 (A \times B) = \mathbb{P}_1(A)\mathbb{P}_2(B)$.

La tribu $\mathcal{A}_1 \otimes \mathcal{A}_2$ s'appelle la tribu produit, et $\mathbb{P}_1 \otimes \mathbb{P}_2$ est la mesure produit sur $\Omega_1 \times \Omega_2$.

On note qu'ici également, la donnée de $\mathbb{P}_1 \otimes \mathbb{P}_2$ sur la classe des rectangles $A \times B$ suffit à définir la mesure produit.

Lemme 1.1.6. $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$.

Variables aléatoires

Dans la pratique, on ne s'occupe pas vraiment de l'espace $(\Omega, \mathcal{A}, \mathbb{P})$. On lui demande simplement d'être assez "riche" pour pouvoir y définir les variables aléatoires qui nous intéressent.

Définition 1.1.7. *Une application $f : (E, \mathcal{A}) \to (F, \mathcal{B})$ est dite **mesurable** (par rapport à la tribu $\mathcal{A}$) si*

$$\forall B \in \mathcal{B}, \quad f^{-1}(B) := \{x \in E : f(x) \in B\} \in \mathcal{A}.$$

Lemme 1.1.8. *Pour qu'une application $f : (E, \mathcal{A}) \to (F, \mathcal{B})$ soit mesurable, il suffit que $f^{-1}(B) \in \mathcal{A}$ pour $B \in C \subset \mathcal{B}$ si $\sigma(C) = \mathcal{B}$. (Indication : montrer que $\{B \subset F : f^{-1}(B) \in \mathcal{A}\}$ est une tribu.)*

Proposition 1.1.9. *Si E et F sont des espaces métriques munis de leurs tribus boréliennes respectives, alors toute fonction continue de E vers F est mesurable (dans ce cas on utilise aussi l'adjectif borélienne).*

On vérifie que le fait d'être mesurable est stable par addition, multiplication, composition, passage à la limite, et (pour un ensemble dénombrable de fonctions) passage à la borne sup/inf.

Définition 1.1.10. *Une **variable aléatoire** (v.a. en abrégé) sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $(E, \mathcal{E})$ est une application mesurable $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$.*

Exemple. *Somme de deux dés, somme des gains d'un joueur, etc.*

Définition 1.1.11. *La **loi** de la v.a. X est la mesure image P_X de $\mathbb{P}$ par X, i.e.*

$$\forall B \in \mathcal{E}, \quad P_X(B) = \mathbb{P}\left(X^{-1}(B)\right) = \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \in B\}\right) =: \mathbb{P}\left(X \in B\right). \tag{1.2}$$

Exemple.
- *Variable aléatoire discrète :* $P_X = \sum_{x \in E} \mathbb{P}(X = x)\, \delta_x$, *autrement dit,*
$$P_X(B) = \sum_{x \in B} \mathbb{P}(X = x).$$
- *Variable aléatoire réelle à densité :* $P_X(B) = \int_B p(x)\, \mathrm{d}x$.

EXERCICE 4 - Vérifier que (1.2) définit une probabilité sur $(E, \mathcal{E})$.

Définition 1.1.12 (Espérance). *Soit X une v.a. à valeurs dans $(E, \mathcal{E})$ et $f : E \to \mathbb{R}$ une application mesurable. On suppose que f est soit positive, soit intégrable par rapport à la loi de X (i.e. $\int |f|\, \mathrm{d}P_X < \infty$). On pose*

$$\mathbb{E}\left[f(X)\right] = \int_E f(x) P_X(\mathrm{d}x).$$

Exemple.
- *Cas discret :* $\mathbb{E}\left[f(x)\right] = \sum_{x \in E} f(x)\mathbb{P}(X = x)$.
- *Cas réel à densité :* $\mathbb{E}\left[f(X)\right] = \int_{-\infty}^{+\infty} f(x)p(x)\, \mathrm{d}x$.

On a les propriétés usuelles de l'intégrale qui restent vraies pour l'espérance : linéarité, positivité, inégalité triangulaire, etc.

Identification des lois

Le résultat suivant est très utile dans la pratique. Il permet de calculer la loi d'une v.a. X en connaissant l'espérance de $h(X)$ pour un nombre suffisant de fonctions h.

Théorème 5 (Théorème d'identification). *i) Si μ est une probabilité sur $(E, \mathcal{E})$ telle que, pour toute fonction $h : E \to \mathbb{R}$ continue bornée,*

$$\mathbb{E}\left[h(X)\right] = \int_E h(x)\mu(\mathrm{d}\,x),$$

alors $P_X = \mu$.

ii) Si $f : \mathbb{R} \to \mathbb{R}_+$ est telle que, pour toute fonction $h : \mathbb{R} \to \mathbb{R}$ continue bornée,

$$\mathbb{E}\left[h(X)\right] = \int_{\mathbb{R}} h(x)f(x)\,\mathrm{d}\,x,$$

alors X admet la fonction f pour densité.

Dans la pratique, on connaît la loi de X et on cherche celle de $f(X)$ pour une fonction $f : E \to F$ mesurable. On écrit alors

$$\mathbb{E}\left[h(f(X))\right] = \int_E h(f(x))P_X(\mathrm{d}\,x)$$

et on change de variable d'intégration (attention f n'est pas forcément un difféomorphisme sur E).

1.2 Indépendance

Définition

Définition 1.2.1. — *Deux événements $A, B \in \mathcal{A}$ sont dits **indépendants** si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.*

— *Les événements $A_1, \ldots, A_n$ sont indépendants si*

$$\forall 1 \leq i_1 < \ldots < i_k \leq n, \quad \mathbb{P}\left(A_{i_1} \cap \ldots \cap A_{i_k}\right) = \mathbb{P}\left(A_{i_1}\right) \ldots \mathbb{P}\left(A_{i_k}\right).$$

— *Les v.a. $X_1, \ldots, X_n$ sont indépendantes si*

$$\forall B_1 \in \mathcal{E}_1, \ldots, B_n \in \mathcal{E}_n, \quad \mathbb{P}\left(X_1 \in B_1, \ldots, X_n \in B_n\right) = \mathbb{P}\left(X_1 \in B_1\right) \ldots \mathbb{P}\left(X_n \in B_n\right).$$

— *La suite de v.a. $(X_n)_{n\in\mathbb{N}}$ est indépendante si, pour tout $n \in \mathbb{N}$, $X_1, \ldots, X_n$ sont indépendantes.*

Théorème 6. *Les variables $X_1, \ldots, X_n$ sont indépendantes si et seulement si*

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

Le Théorème 8 permet de caractériser les v.a. indépendantes de différentes façons.

Corollaire 1.2.2. — *Si $X_1, \ldots, X_n$ sont des v.a. réelles, elles sont indépendantes si et seulement si, pour toutes fonctions $f_1, \ldots, f_n$ continues à support compact de $\mathbb{R}$ dans $\mathbb{R}$,*

$$\mathbb{E}\left[\Pi_i f_i(X_i)\right] = \Pi_i \mathbb{E}\left[f(X_i)\right].$$

Si de plus les X_i admettent des densités p_i, alors $(X_1, \ldots, X_n)$ admet $\Pi_i p_i(x_i)$ pour densité.

— *Réciproquement, si $(X_1, \ldots, X_n)$ a une densité de la forme*

$$p(x_1, \ldots, x_n) = \Pi_{i=1}^n q_i(x_i),$$

où les q_i sont boréliennes positives, alors $X_1, \ldots, X_n$ sont indépendantes et X_i admet une densité de la forme $p_i = C_i q_i$, $C_i > 0$.

— *Si $X_1, \ldots, X_n$ sont des v.a. discrètes, alors elles sont indépendantes si et seulement si, pour tout $x_1 \in E_1, \ldots, x_n \in E_n$,*

$$\mathbb{P}\left(X_1 = x_1, \ldots, X_n = x_n\right) = \mathbb{P}\left(X_1 = x_1\right) \ldots \mathbb{P}\left(X_n = x_n\right).$$

Démonstration. Utiliser le Théorème 7 et le Théorème de Fubini. □

Remarque. ***Attention !*** *Pour que $X_1, \ldots, X_n$ soient indépendantes, il ne **suffit pas** que X_i et X_j soient indépendantes pour tout $i \neq j$. En effet, supposons que Y_1, Y_2, Y_3 soient trois v.a. de Bernoulli indépendantes et de paramètre $1/2$. Alors les événements $\{Y_1 = Y_2\}$, $\{Y_2 = Y_3\}$ et $\{Y_1 = Y_3\}$ sont indépendants deux à deux, mais pas indépendants.*

Proposition 1.2.3. *Soient X et Y deux v.a. réelles indépendantes, de lois respectives P_X et P_Y. Alors la loi de $X + Y$ est la convolution des lois P_X et P_Y, i.e.*

$$P_X * P_Y(B) = \int P_Y(B - x)P_X(\mathrm{d}\,x).$$

Si de plus X et Y admettent chacun une densité, notées p_X et p_Y, alors $X + Y$ admet $p_X * p_Y$ pour densité.

Lemme de Borel-Cantelli

Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements, on note

$$\limsup_n A_n = \bigcap_{n=0}^{\infty} \left(\bigcup_{k=n}^{\infty} A_k \right),$$

et

$$\liminf_n A_n = \bigcup_{n=0}^{\infty} \left(\bigcap_{k=n}^{\infty} A_k \right).$$

Lemme 1.2.4 (Borel-Cantelli)**.** *Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements.*

i) Si $\sum_{n \geq 0} \mathbb{P}(A_n) < \infty$, alors

$$\mathbb{P}\left(\limsup_n A_n \right) = 0,$$

ou, de manière équivalente, avec probabilité 1

$$\{n \in \mathbb{N} : \omega \in A_n\} \text{ est fini.}$$

ii) Si $\sum_{n \geq 0} \mathbb{P}(A_n) = \infty$ et si les événements $(A_n)_n$ sont indépendants, alors

$$\mathbb{P}\left(\limsup_n A_n \right) = 1,$$

ou, de manière équivalente, avec probabilité 1

$$\{n \in \mathbb{N} : \omega \in A_n\} \text{ est infini.}$$

Ce résultat est très utile pour montrer que quelque chose arrive "presque sûrement" (*i.e.* avec probabilité 1), comme par exemple la convergence d'une variable aléatoire. L'écriture équivalente donnée dans l'énoncé du Lemme 1.2.3 s'obtient en notant que

$$\limsup_n A_n = \{\omega \in \Omega : \forall n \in \mathbb{N}, \exists k \geq n : \omega \in A_k\}.$$

1.3 Convergence de variables aléatoires

Loi faible des grands nombres

Théorème 7 (Loi faible des grands nombres)**.** *Soit $(X_n)_n$ une suite de v.a. i.i.d. telles que $\mathbb{E}\left[X_1^2\right] < \infty$, alors*

$$\frac{1}{n}(X_1 + \ldots + X_n) \xrightarrow[n\to\infty]{L^2} \mathbb{E}[X_1],$$

c'est-à-dire $\mathbb{E}\left[\left(\frac{1}{n}S_n - \mu\right)^2\right] \underset{n\to\infty}{\longrightarrow} 0$, où $S_n = X_1 + \ldots + X_n$ et $\mu = \mathbb{E}[X_1]$.

Démonstration. On écrit

$$\mathbb{E}\left[\left(\frac{1}{n}(X_1 - \mu + \ldots + X_n - \mu)\right)^2\right] = \frac{1}{n^2}\mathbb{V}[X_1 + \ldots + X_n]$$
$$= \frac{1}{n}\mathbb{V}[X_1] \underset{n\to\infty}{\longrightarrow} 0.$$

(On a utilisé l'indépendance des X_i pour passer à la deuxième ligne.) □

Définition 1.3.1. *Une suite de v.a. $(X_n)_n$ converge en* ***probabilité*** *vers X si*

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

Proposition 1.3.2. *La convergence L^p (i.e. $\mathbb{E}[|X_n - X|^p] \to 0$) pour $p \geq 1$ implique la convergence en probabilité.*

Démonstration. Par l'inégalité de Chebyshev (A.1),

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \varepsilon^{-p}\mathbb{E}[|X_n - X|^p] \underset{n\to\infty}{\longrightarrow} 0.$$

□

Loi forte des grands nombres

On a en réalité un résultat beaucoup plus fort que celui du Théorème 10. La convergence de la moyenne empirique des X_i a en fait lieu pour presque toute issue $\omega \in \Omega$, dans un sens que l'on précise à l'aide de la définition suivante.

Définition 1.3.3. *Une suite de v.a. $(X_n)_n$ converge* ***presque sûrement*** *vers X s'il existe un événement $N \in \mathcal{A}$ de probabilité nulle tel que*

$$\forall \omega \notin N, \quad X_n(\omega) \underset{n\to\infty}{\longrightarrow} X(\omega).$$

Théorème 8 (Loi forte des grands nombres)**.** *Soit $(X_n)_n$ une suite de v.a. i.i.d. telle que $\mathbb{E}[|X_1|] < \infty$, alors*

$$\frac{1}{n}(X_1 + \ldots + X_n) \xrightarrow[n\to\infty]{ps} \mathbb{E}[X_1].$$

La preuve utilise le Lemme de Borel-Cantelli.

Le résultat suivant montre que la convergence presque sûre est plus forte que la convergence en probabilité.

Proposition 1.3.4. *Si la suite $(X_n)_n$ converge p.s., alors elle converge en probabilité (vers la même v.a.). Si $(X_n)_n$ converge en probabilité vers X, alors il existe une sous-suite $(X_{n_k})_k$ qui converge p.s. vers X.*

La première assertion se prouve par convergence dominée, la seconde par le Lemme de Borel-Cantelli.

Voici un exemple de v.a. convergeant en probabilité mais pas presque sûrement.

Exemple. *On munit l'espace $\Omega = [0,1]$ de la tribu borélienne et de la mesure de Lebesgue (cf (1.1)). Posons $X_n(\omega) = \mathbb{1}_{A_n}(\omega)$, $n \geq 1$ où $\lambda(A_n) = \frac{1}{n}$ et les A_n sont mis bout à bout le long de l'intervalle $[0,1]$, en repartant de zéro dès que le bord est atteint. Alors $X_n \xrightarrow[n\to\infty]{L^1} 0$ mais, comme la série des $\frac{1}{n}$ diverge, $X_n \overset{ps}{\nrightarrow} 0$.*

EXERCICE 9 - Exhiber une sous-suite qui converge ps vers zéro.

Convergence en loi

Présentons maintenant un mode de convergence pour les variables aléatoires beaucoup plus faible que ceux vus précédemment.

Définition 1.3.5. *Une suite de v.a. $(X_n)_n$ à valeurs dans E converge **en loi** vers X si pour toute fonction $f : E \to \mathbb{R}$ continue bornée,*

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

On note alors indifféremment $X_n \Rightarrow X$, $X_n \xrightarrow[n\to\infty]{\mathcal{L}} X$ et $X_n \xrightarrow[n\to\infty]{d} X$ en anglais.

Proposition 1.3.6.

- *Si X_n et X sont à valeurs dans un espace dénombrable E, alors $X_n \Rightarrow X$ si et seulement si,*

$$\forall x \in E, \quad \mathbb{P}\left(X_n = x\right) \xrightarrow[n\to\infty]{} \mathbb{P}\left(X = x\right).$$

- *Si X_n admet p_n pour densité sur $\mathbb{R}$, alors si $p_n \to p$ λ-presque partout et s'il existe $q \in L^1(\mathbb{R})$ positive telle que, pour tout $n \in \mathbb{N}$, $p_n \leq q$ λ-p.p., alors p est une densité et $X_n \Rightarrow X$, où X admet p pour densité.*

Proposition 1.3.7. *Si $(X_n)_n$ converge en probabilité vers X, alors $(X_n)_n$ converge en loi vers X.*

Le résultat suivant est utile pour caractériser la convergence en loi.

Théorème 10. *Soient $(X_n)_n$ et X des v.a. à valeurs dans $\mathbb{R}^d$. Les propositions suivantes sont équivalentes.*

i) $X_n \Rightarrow X$,

ii) *pour tout ouvert* $G \subset \mathbb{R}^d$, $\liminf_{n\to\infty} \mathbb{P}\left(X_n \in G\right) \geq \mathbb{P}\left(X \in G\right)$,

iii) *pour tout fermé* $F \subset \mathbb{R}^d$, $\limsup_{n\to\infty} \mathbb{P}\left(X_n \in F\right) \leq \mathbb{P}\left(X \in F\right)$,

iv) *pour tout borélien* $A \subset \mathbb{R}^d$ *tel que* $\mathbb{P}\left(X \in \partial A\right) = 0$, $\lim_{n\to\infty} \mathbb{P}\left(X_n \in A\right) = \mathbb{P}\left(X \in A\right)$.

(Pour se rappeler du sens des inégalités dans (ii), il suffit de considérer le cas où $X_n = x_n \in G$ presque sûrement et $x_n \xrightarrow[n\to\infty]{} x \in \partial G$. Pour (iii), il suffit de considérer $F = G^c$.)

Fonction de répartition   On rappelle que la fonction de répartition d'une v.a. réelle X est donnée par $F_X(x) = \mathbb{P}\left(X \leq x\right)$. Le résultat suivant découle du Théorème 13.

Proposition 1.3.8. *Soient $(X_n)_n$ et X des v.a. réelles. X_n converge en loi vers X si et seulement si $F_{X_n}(x) \xrightarrow[n\to\infty]{} F_X(x)$ pour tout x où F_X est continue.*

Fonction caractéristique

Définition 1.3.9. *Soit X une v.a. à valeurs dans $\mathbb{R}^d$. Sa **fonction caractéristique** $\Phi_X : \mathbb{R}^d \to \mathbb{C}$ est définie par*

$$\Phi_X(\xi) = \mathbb{E}\left[e^{i\xi\cdot X}\right].$$

Remarque. *Cette définition coïncide avec celle de la transformée de Fourier de la loi de X. Le résultat suivant est donc une conséquence de la formule d'inversion de Fourier.*

Proposition 1.3.10. *La fonction caractéristique caractérise la loi de X.*

Exemple. *Si $X \sim \mathcal{N}(0,\sigma^2)$, $\Phi_X(\xi) = \exp\left(-\frac{\sigma^2\xi^2}{2}\right)$.*

Comme corollaire de la proposition précédente, on a le résultat suivant.

Proposition 1.3.11. *Les v.a. $X_1, \ldots, X_n$ sont indépendantes si et seulement si $\Phi_{(X_1,\ldots,X_n)} = \Phi_{X_1} \ldots \Phi_{X_n}$.*

La fonction caractéristique nous donne un moyen très utile de caractériser la convergence en loi. Il suffit en effet de vérifier que la suite des fonctions caractéristiques converge simplement vers la fonction caractéristique d'une v.a.

Théorème 11 (Lévy)**.** *Soient $(X_n)_n$ une suite de v.a. et X une v.a. à valeurs dans $\mathbb{R}^d$. Alors $X_n \Rightarrow X$ si et seulement si, pour tout $\xi \in \mathbb{R}^d$, $\Phi_{X_n}(\xi) \underset{n\to\infty}{\longrightarrow} \Phi_X(\xi)$.*

Théorème de la limite centrale

Théorème 12 (Théorème de la limite centrale)**.** *Soit $(X_n)_n$ une suite de v.a. réelles i.i.d. dans L^2. On note $\mu = \mathbb{E}[X_1]$ et $\sigma^2 = \mathbb{V}[X_1]$. Alors*

$$\frac{1}{\sqrt{n}}(X_1 + \ldots + X_n - n\mu) \underset{n\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}(0, \sigma^2).$$

Démonstration. Le résultat s'obtient en faisant un développement limité au deuxième ordre de la fonction caractéristique du membre de gauche. □

Une application importante de ce théorème est la définition d'intervalles de confiance (cf. cours de statistiques).

1.4 Vecteurs Gaussiens

Définition 1.4.1. *Une v.a. $X = (X_1, \ldots, X_d)$ à valeurs dans $\mathbb{R}^d$ est un* ***vecteur gaussien*** *si, pour tout $a \in \mathbb{R}^d$, $\langle a, X\rangle$ suit une loi normale (en considérant que $\delta_x = \mathcal{N}(x, 0)$).*

Exemple. *Si les X_i sont des v.a. gaussiennes indépendantes, alors $(X_1, \ldots, X_d)$ est un vecteur gaussien.*

Théorème 13. *Un vecteur aléatoire X est gaussien si et seulement s'il existe $m \in \mathbb{R}^d$, $C \in \mathcal{S}_d^+(\mathbb{R})$ tels que*

$$\Phi_X(\xi) = \exp\left(i\langle \xi, m\rangle - \frac{1}{2}\langle \xi, C\xi\rangle\right).$$

On note $m = \mathbb{E}[X]$ et C est la matrice de covariance de X.

Théorème 14. *Soit X un vecteur aléatoire gaussien, d'espérance m et de matrice de convariance C.*
Si le rang de C est d (autrement dit, si $Det(C) > 0$), alors X admet une densité f_X par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ avec :

$$f_X(x) = \frac{1}{(2\pi)^{d/2}\sqrt{Det(C)}} \exp\left(-\frac{1}{2}(t-m)^T C^{-1}(t-m)\right) \qquad \forall\ x \in \mathbb{R}^d$$

Si le rang de C est $k < d$ (et donc $Det(C) = 0$, alors la loi de X n'admet pas de densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ et concentre toute sa masse sur un sous-espace affine de $\mathbb{R}^d$ de dimension k.

Un corollaire direct du Théorème 16 est le résultat suivant.

Proposition 1.4.2. *Si X est un vecteur gaussien, ses composantes sont indépendantes si et seulement si sa matrice de covariance est diagonale.*

Remarque. ***Attention !*** *Ce n'est pas vrai en général, même pour des variables gaussiennes. Avant d'appliquer ce résultat, il faut bien vérifier que toute <u>combinaison linéaire</u> des X_i est gaussienne, comme le montre l'exercice suivant.*

EXERCICE 15 - Soit X une v.a. réelle de loi $\mathcal{N}(0,1)$ et ε une v.a. indépendante de X de loi $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. On pose $Y = \varepsilon X$. Montrer que Y suit une loi normale $\mathcal{N}(0,1)$ et calculer $\mathrm{Cov}(X, Y)$. Les variables X et Y sont-elles indépendantes ? Le vecteur (X, Y) est-il gaussien ?

Le TCL se généralise sans peine au cas vectoriel.

Théorème 16 (TCL vectoriel)**.** *Soit $(X_n)_n$ une suite de v.a. i.i.d. à valeurs dans $\mathbb{R}^d$, de covariance $K \in \mathcal{S}_d^+(\mathbb{R})$, de carré intégrable. Alors*

$$\frac{1}{\sqrt{n}}(X_1 + \ldots + X_n - n\mathbb{E}[X_1]) \underset{n\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}(0, K).$$

Chapitre 2

Conditionnement

Dans tout ce chapitre, $(\Omega, \mathcal{A}, \mathbb{P})$ désigne un espace de probabilité.

2.1 Conditionnement discret

Définition 2.1.1. *Soit $B \in \mathcal{A}$ tel que $\mathbb{P}(B) > 0$. On définit une probabilité $\mathbb{P}(\cdot \mid B)$ sur $(\Omega, \mathcal{A})$ par*

$$\forall A \in \mathcal{A}, \quad \mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

EXERCICE 17 - Vérifier que $\mathbb{P}(\cdot \mid B)$ vérifie les axiomes de la Définition 1.1.2 d'une probabilité.

Si $X : \Omega \to E$ est un v.a. dans L^1, on peut calculer l'espérance de X par rapport à cette nouvelle probabilité :

$$\mathbb{E}[X \mid B] = \frac{\mathbb{E}[X \mathbb{1}_B]}{\mathbb{P}(B)}.$$

Remarque. *C'est un nombre. Il s'interprète comme l'espérance de la variable X sachant que l'événement B est réalisé.*

On voudrait maintenant donner un sens à $\mathbb{E}[X \mid Y]$, où Y est une variable aléatoire, qui s'interpréterait comme l'espérance de la variable X *sachant la valeur prise par la variable* Y. Dans toute la suite, Y est une v.a. à valeurs dans un espace E au plus dénombrable. Soit

$$E' = \{y \in E : \mathbb{P}(Y = y) > 0\}.$$

Pour $y \in E'$, on peut prendre $B = \{Y = y\}$ dans la définition ci-dessus et, pour $X \in L^1$, on a

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

Définition 2.1.2. *Soit $X \in L^1$, **l'espérance conditionnelle de X sachant Y** est la <u>variable aléatoire</u> réelle définie par*

$$\mathbb{E}[X \mid Y] = \phi(Y),$$

où $\phi : E \to \mathbb{R}$ est donnée par

$$\phi(y) = \begin{cases} \mathbb{E}[X \mid Y = y] & \text{si } y \in E', \\ 0 & \text{sinon.} \end{cases}$$

Exemple. *Si X est le résultat d'un lancer de dé, $X = k$ avec probabilité $\frac{1}{6}$ pour $k \in \{1, \ldots, 6\}$. Posons*

$$Y = \begin{cases} 1 & \text{si } X \text{ est pair,} \\ 0 & \text{sinon.} \end{cases}$$

Alors,

$$\mathbb{E}[X \mid Y] = \begin{cases} 4 & \text{si } Y = 1, \\ 3 & \text{si } Y = 0, \end{cases}$$
$$= 3 + Y.$$

EXERCICE 18 - Vérifier que $\mathbb{E}[Y \mid Y] = Y$ et que, si X et Y sont indépendants, $\mathbb{E}[X \mid Y] = \mathbb{E}[X]$.

Proposition 2.1.3. *Si $X \in L^1$, alors $\mathbb{E}[X \mid Y] \in L^1$ et*

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$$
$$\mathbb{E}[|\mathbb{E}[X \mid Y]|] \leq \mathbb{E}[|X|].$$

De plus, pour toute fonction $f : E \to \mathbb{R}$ mesurable bornée,

$$\mathbb{E}[f(Y)X] = \mathbb{E}[f(Y)\mathbb{E}[X \mid Y]] \tag{2.1}$$

Démonstration. Exercice ! □

Remarque. *Le cas particulier $f = 1$ permet de retrouver la formule bien connue des probabilités totales,*

$$\mathbb{E}[X] = \sum_{y \in E'} \mathbb{E}[X \mid Y = y] \mathbb{P}(Y = y).$$

2.2 Conditionnement par une tribu

Commençons par l'exercice suivant.

EXERCICE 19 - Soient Y_1 et Y_2 deux variables aléatoires discrètes et soit X une variable aléatoire dans L^1. On suppose qu'il existe une application injective f telle que $Y_2 = f(Y_1)$. Montrer que $\mathbb{E}[X \mid Y_1] = \mathbb{E}[X \mid Y_2]$.

Dans l'exercice ci-dessus, les variables aléatoires Y_1 et Y_2 apportent la même information sur Ω, et donc les deux espérances conditionnelles sont égales. Pour éclaircir ce point, on définit la notion de tribu engendrée par une variable aléatoire.

Définition 2.2.1. *Soit X une variable aléatoire à valeurs dans un espace mesurable $(E, \mathcal{E})$.* ***La tribu engendrée par*** *X, notée $\sigma(X)$, est la plus petite tribu qui rende X mesurable,*

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

Cette tribu correspond à l'information apportée par la connaissance de X. (On peut notamment montrer qu'une variable aléatoire X est $\mathcal{F}$-mesurable si et seulement si $\sigma(X) \subset \mathcal{F}$.) On va voir par la suite que la « bonne » notion pour l'espérance conditionnelle est le conditionnement par une tribu. On verra alors que l'espérance conditionnelle sachant X est l'espérance conditionnellement à la tribu engendrée par X. La propriété (2.1) sert de définition de l'espérance conditionnelle dans le cas général.

Proposition 2.2.2. *Soit $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et $\mathcal{F}$ une sous-tribu de $\mathcal{A}$. Il existe une unique variable aléatoire (à égalité p.s. près), notée $\mathbb{E}[X \mid \mathcal{F}]$, telle que*

i) $\mathbb{E}[X \mid \mathcal{F}]$ est $\mathcal{F}$-mesurable,

ii) pour tout $A \in \mathcal{F}$, $\mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{F}] \mathbb{1}_A]$.

On insiste sur le fait que $\mathbb{E}[X \mid \mathcal{F}]$ est une variable aléatoire. En fait, si pour chaque $A \in \mathcal{F}$ on sait si A est réalisé ou non, $\mathbb{E}[X \mid \mathcal{F}]$ est la meilleure estimation de X que l'on peut donner avec cette information. L'exercice suivant justifie cette interprétation.

EXERCICE 20 -

— Si X est mesurable par rapport à $\mathcal{F}$, alors $\mathbb{E}[X \mid \mathcal{F}] = X$.

— Si Z est mesurable par rapport à $\mathcal{F}$ et bornée, alors $\mathbb{E}[ZX \mid \mathcal{F}] = Z\mathbb{E}[X \mid \mathcal{F}]$.

— Si X est indépendant de $\mathcal{F}$, alors $\mathbb{E}[X \mid \mathcal{F}] = \mathbb{E}[X]$.

— Si $\mathcal{F}_2$ est une sous-tribu de $\mathcal{F}_1$, alors $\mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_1] \mid \mathcal{F}_2] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_2] \mid \mathcal{F}_1] = \mathbb{E}[X \mid \mathcal{F}_2]$.

L'espérance conditionnelle hérite de toutes les propriétes de base de l'espérance : linéarité, positivité, inégalité triangulaire. Elle conserve également l'inégalité de Jensen.

Exemple. *Soit $\mathcal{A}_1, \mathcal{A}_2, \ldots \in \mathcal{A}$ une partition (finie ou infinie) de Ω telle que $\mathbb{P}(A_i) > 0$ pour tout $i \geq 1$. Supposons que $\mathcal{F} = \sigma(A_1, A_2, \ldots)$. Alors, pour $X \in L^1$,*

$$\mathbb{E}[X \mid \mathcal{F}] = \sum_{i \geq 1} \frac{\mathbb{E}[X \mathbb{1}_{A_i}]}{\mathbb{P}(A_i)} \mathbb{1}_{A_i}.$$

Avec cette définition, on vérifie que pour toute variable aléatoire discrète Y, $\mathbb{E}[X \mid Y] = \mathbb{E}[X \mid \sigma(Y)]$. Cela permet d'étendre la définition de l'espérance conditionnelle sachant une variable aléatoire à valeurs dans un espace général (pas forcément dénombrable).

Définition 2.2.3. *Soit $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et Y une variable aléatoire sur Ω. On définit :*

$$\mathbb{E}[X \mid Y] := \mathbb{E}[X \mid \sigma(Y)].$$

2.3 Conditionnement dans L^2 et vecteurs Gaussiens

Lorsque X est de carré intégrable, on peut utiliser le produit scalaire dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ pour donner une autre interprétation de l'espérance conditionnelle.

Soit $\mathcal{F}$ une sous-tribu de $\mathcal{A}$ et soit X une variable aléatoire dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$. D'après l'inégalité de Jensen, on a $\mathbb{E}\left[X \mid \mathcal{F}\right]^2 \leq \mathbb{E}\left[X^2 \mid \mathcal{F}\right]$, et donc

$$\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{F}\right]^2\right] \leq \mathbb{E}\left[\mathbb{E}\left[X^2 \mid \mathcal{F}\right]\right] = \mathbb{E}\left[X^2\right] < \infty.$$

Donc $\mathbb{E}\left[X \mid \mathcal{F}\right]$ est dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$, vu comme sous-espace fermé de $L^2(\Omega, \mathcal{A}, \mathbb{P})$. De plus, pour toute variable aléatoire Z bornée et $\mathcal{F}$-mesurable,

$$\mathbb{E}\left[Z(X - \mathbb{E}\left[X \mid \mathcal{F}\right])\right] = 0.$$

Donc $X - \mathbb{E}\left[X \mid \mathcal{F}\right]$ est orthogonal à toutes les variables aléatoires bornées $\mathcal{F}$-mesurables. Par densité, on en déduit que $X - \mathbb{E}\left[X \mid \mathcal{F}\right]$ est orthogonal à toutes les variables aléatoires de $L^2(\Omega, \mathcal{F}, \mathbb{P})$. L'espérance conditionnelle $\mathbb{E}\left[X \mid \mathcal{F}\right]$ est donc le projeté orthognal de X sur $L^2(\Omega, \mathcal{F}, \mathbb{P})$. En d'autres termes, $\mathbb{E}\left[X \mid Y\right]$ est la meilleure approximation de X, au sens de la norme L^2, par une variable aléatoire de la forme $f(Y)$. Cette observation a une conséquence remarquable dans le cadre des vecteurs Gaussiens.

Théorème 21. *Soit $(Y_1, \ldots, Y_n, X)$ un vecteur Gaussien centré. Alors $\mathbb{E}\left[X \mid Y_1, \ldots, Y_n\right]$ coïncide avec la projection orthogonale de X (dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$) sur l'espace vectoriel engendré par $Y_1, \ldots, Y_n$. Il existe donc des réels $\lambda_1, \ldots, \lambda_n$ tels que*

$$\mathbb{E}\left[X \mid Y_1, \ldots, Y_n\right] = \sum_{k=1}^{n} \lambda_k Y_k.$$

De plus, la loi conditionnelle de X sachant $Y_1, \ldots, Y_n$ est une loi Gaussienne de moyenne $m = \sum_{k=1}^{n} \lambda_k Y_k$ et de variance

$$\sigma^2 = \mathbb{E}\left[\left(X - \sum_{k=1}^{n} \lambda_k Y_k\right)^2\right].$$

C'est-à-dire que pour toute fonction borélienne $h : \mathbb{R} \to \mathbb{R}_+$,

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

En d'autres termes, si $(Y_1, \ldots, Y_n, X)$ est un vecteur Gaussien, la meilleurs manière d'approximer X (dans L^2) par une fonction mesurable de $Y_1, \ldots, Y_n$ est de le faire par une combinaison linéaire des Y_i.

Chapitre 3

Martingales à temps discret

Les martingales sont un outils fondamental dans l'étude des processus aléatoires. Une martingale correspond à l'évolution de la somme des gains d'un joueur dans un jeu équitable. Dans ce chapitre nous rappelons un certain nombre de propriétés fondamentales liées aux martingales.

3.1 Définitions

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

Définition 3.1.1. *Une filtration $(\mathcal{F}_n)_n$ de $(\Omega, \mathcal{A}, \mathbb{P})$ est une suite croissante de sous-tribus de $\mathcal{A}$, c'est-à-dire telle que*

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots \subset \mathcal{A}.$$

Un processus aléatoire est une suite $(X_n)_n$ de variable aléatoire (ici à valeurs réelles) définies sur $(\Omega, \mathcal{A}, \mathbb{P})$.

Exemple. *La filtration canonique associée à $(X_n)_n$ est définie par*

$$\mathcal{F}_n = \sigma(X_0, \ldots, X_n).$$

La tribu $\mathcal{F}_n$ correspond dans ce cas à l'information acquise par l'observation du processus jusqu'au temps n. De manière générale, une filtration représente l'évolution de l'information acquise au cours du temps.

Définition 3.1.2. *Soit $(X_n)_{n\in\mathbb{N}}$ un processus aléatoire. On dit que le processus $(X_n)_{n\in\mathbb{N}}$ est une **martingale** si*

i) $(X_n)_n$ est adapté, c'est-à-dire que pour tout $n \in \mathbb{N}$, X_n est mesurable par rapport à $\mathcal{F}_n$,

ii) $\mathbb{E}[|X_n|] < \infty$ pour tout $n \in \mathbb{N}$,

iii) pour tout $n \in \mathbb{N}$, $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$.

En d'autres termes, une martingale est un processus dont les accroissements sont équilibrés ($X_{n+1} - X_n$ est une variable centrée étant donnée toute l'information disponible jusqu'au temps n). C'est par exemple le cas de la somme des gains d'un joueur dans un jeu équitable. On note que la filtration canonique est la plus petite filtration qui rende le processus $(X_n)_n$ adapté. Si l'on remplace $=$ dans (iii) par $\leq$ ou bien $\geq$, $(X_n)_n$ est appelé respectivement une sur-martingale ou une sous-martingale.

Exemple.

1. *Marche aléatoire sur $\mathbb{R}$. Soit $(\xi_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées à valeurs réelles telles que $\mathbb{E}[\xi_n] = 0$ et $\mathbb{E}[|\xi_n|] < \infty$. On pose $X_n = \xi_0 + \ldots + \xi_n$ et $\mathcal{F}_n = \sigma(\xi_0, \ldots, \xi_n)$ pour $n \geq 0$. Alors $(X_n)_{n\in\mathbb{N}}$ est une martingale par rapport à $(\mathcal{F}_n)_n$.*
2. *Soit $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et $(\mathcal{F}_n)_n$ une filtration. On pose $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$. Alors $(X_n)_{n\in\mathbb{N}}$ est une martingale. Une telle martingale est dite fermée.*

EXERCICE 22 - Démontrer les deux assertions ci-dessus.

3.2 Temps d'arrêt

Le temps d'arrêt est une notion naturelle d'un instant aléatoire lié à une filtration. Elle intervient dans le Théorème d'arrêt (voir le Théorème 26 ci-dessous) qui dit qu'on ne peut pas espérer gagner d'argent à un jeu équilibré en un temps borné.

Définition 3.2.1. *Soit $(\mathcal{F}_n)_n$ une filtration. Une variable aléatoire T à valeurs dans $\mathbb{N} \cup \{\infty\}$ est un* ***temps d'arrêt*** *si pour tout $n \in \mathbb{N}$, $\{T = n\} \in \mathcal{F}_n$.*

Si on imagine que T est l'instant auquel notre joueur s'arrête de jouer, la définition ci-dessus dit que la décision de s'arrêter doit pouvoir être prise à partir de l'information disponible à ce moment. Par exemple, le jour de l'année ou une action donnée est à son maximum n'est a priori pas un temps d'arrêt, car il n'est pas possible de savoir le jour même si l'action ne redépassera pas son cours actuel d'ici la fin de l'année. En particulier, toute variable aléatoire déterministe à valeurs dans $\mathbb{N}$ est un temps d'arrêt.

Exemple. *Un exemple important de temps d'arrêt est le temps d'atteinte d'un ensemble. Si $(X_n)_n$ est un processus adapté à valeurs dans un espace mesurable $(E, \mathcal{E})$,*

$$T_A := \inf \{n \in \mathbb{N} : X_n \in A\}$$

est un temps d'arrêt (avec la convention $\inf \emptyset = +\infty$). En revanche, le temps de sortie de A,

$$L_A := \sup \{n \in \mathbb{N} : X_n \in A\}$$

n'est en général pas un temps d'arrêt.

On rappelle la notation $a \wedge b = min(a, b)$.

Théorème 23 (Théorème d'arrêt)**.** *Soit $(X_n)_n$ une martingale (resp. une surmartingale) et soit T un temps d'arrêt. Alors $(X_{n\wedge T})_n$ est aussi une martingale (resp. une surmartingale). En particulier, si le temps d'arrêt T est borné, on a $X_T \in L^1$ et*

$$\mathbb{E}[X_T] = \mathbb{E}[X_0], \qquad \text{(resp. } \mathbb{E}[X_T] \leq \mathbb{E}[X_0] \text{).}$$

Exemple. *Soit $(\xi_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées telles que $\mathbb{P}(\xi_1 = 1) = \mathbb{P}(\xi_1 = -1) = 1/2$. On pose $X_0 = x_0$ et $X_n = x_0 + \xi_1 + \ldots + \xi_n$. Le processus $(X_n)_{n\in\mathbb{N}}$ est alors une martingale par rapport à sa filtration canonique. Pour $N \geq 1$ fixé, soit T la variable aléatoire définie par*

$$T := \inf \{n \geq 1 : X_n = 0 \text{ ou } X_n = N\}.$$

C'est un temps d'arrêt. On souhaite utiliser le théorème ci-dessus pour déterminer la probabilité que X_n atteigne le niveau N avant de passer par 0. On suppose au passage que $0 \leq x_0 \leq N$. D'après le Théorème 26, pour tout $n \geq 1$,

$$\mathbb{E}[X_{n\wedge T}] = \mathbb{E}[X_0] = x_0.$$

Le temps d'arrêt T n'est pas borné, mais $T < \infty$ p.s. donc, pour $n \to \infty$, $X_{n\wedge T} \to X_T$ presque sûrement. De plus, $|X_{n\wedge T}| \leq N$ pour tout $n \geq 1$. Par convergence dominée, on peut donc conclure que

$$\mathbb{E}[X_T] = x_0.$$

D'autre part,

$$\mathbb{E}[X_T] = N \times \mathbb{P}(X_T = N) + 0 \times \mathbb{P}(X_T = 0).$$

On en déduit donc que $\mathbb{P}(X_T = N) = \frac{x_0}{N}$.

3.3 Convergence presque sûre des martingales

En plus du Théorème d'arrêt, on dispose de résultats très puissants sur la convergence des martingales. Ce résultat peut être vu comme un analogue stochastique de la convergence des suites monotones bornées.

Théorème 24. *Soit $(X_n)_{n\in\mathbb{N}}$ une martingale telle que*

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|] < \infty,$$

alors il existe une variable aléatoire X_∞ telle que $\mathbb{E}[|X_\infty|] < \infty$ et X_n converge p.s. vers X_∞ quand $n \to \infty$.

De manière équivalente, si $(X_n)_n$ est une sous-martingale et que $\sup_{n\in\mathbb{N}} \mathbb{E}[(X_n)^+] < \infty$, alors la conclusion du théorème reste vraie (et inversement si $(X_n)_n$ est une surmartingale en remplacant la partie positive par une partie négative). En particulier, toute surmartingale positive converge p.s. vers une variable aléatoire dans L^1.

Exemple. *On reprend la martingale $(X_n)_{n\in\mathbb{N}}$ de l'exemple précédent, avec $x_0 = 1$. On pose cette fois*

$$T = \inf \{n \geq 1 : X_n = 0\}.$$

D'après le Théorème 26, $Y_n = X_{n\wedge T}$ est une martingale, qui de plus est positive. Elle converge donc p.s. vers une variable aléatoire $X_\infty \in L^1$. Cela n'est possible que sur $\{T < \infty\}$ car $|X_n - X_{n+1}| = 1$ (on a donc montré au passage que $T < \infty$ p.s.). On a donc $X_\infty = 0$ p.s. On remarque qu'il ne peut y avoir convergence dans L^1 puisque $\mathbb{E}[X_n] = 1 > \mathbb{E}[X_\infty] = 0$.

3.4 Convergence L^p des martingales

Comme le montre le dernier exemple, le fait pour une martingale d'être bornée dans L^1 n'assure pas que cette dernière converge dans L^1. Nous allons voir que le fait d'être bornée dans L^p pour $p > 1$ permet en revanche de conclure qu'une martingale converge dans L^p. Au passage nous présentons une séries d'inégalités très utiles dans l'étude des martingales.

Théorème 25 (Inégalité maximale de Doob)**.** *Soit $(X_n)_n$ une martingale (ou une sous-martingale). Pour tout $a > 0$ et pour tout $n \in \mathbb{N}$,*

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

Si $(X_n)_n$ est un processus aléatoire, on note

$$X_n^* = \sup_{0\leq k\leq n} |X_k|, \qquad X_\infty^* = \sup_{n\in\mathbb{N}} |X_n|.$$

La proposition suivante est alors une conséquence de l'inégalité maximale de Doob.

Proposition 3.4.1. *Soit $p > 1$ et soit $(X_n)_n$ une martingale. Alors, pour tout $n \in \mathbb{N}$,*

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

Enfin, le résultat suivant donne la convergence dans L^p des martingales.

Théorème 26. *Soit $(X_n)_n$ une martingale. Supposons qu'il existe $p > 1$ tel que*

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left[|X_n|^p\right] < \infty.$$

Alors X_n converge p.s. et dans L^p vers une variable aléatoire $X_\infty \in L^p$ telle que

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

et

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

3.5 Convergence L^1 des martingales

Comme on l'a vu, être bornée dans L^1 n'assure pas la convergence dans L^1 d'une martingale. Il existe cependant une condition nécessaire et suffisante pour la convergence L^1 des martingales, l'uniforme intégrabilité.

Définition 3.5.1. *Une famille $(X_i)_{i\in I}$ de variables aléatoires dans L^1 est dite uniformément intégrable si*

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

Exemple.
- *Soit Z une variable aléatoire dans L^1. Alors si $|X_i| \leq Z$ pour tout $i \in I$, $(X_i)_{i\in I}$ est uniformément intégrable.*
- *Soit $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ une fonction telle que $x^{-1}\Phi(x) \underset{x\to\infty}{\longrightarrow} 0$. Alors, pour tout $C > 0$,*
$$\left\{X \in L^1 : \mathbb{E}\left[\Phi(|X|)\right] \leq C\right\}$$
est uniformément intégrable. En effet, il suffit d'écrire
$$\mathbb{E}\left[|X| \mathbb{1}_{|X|>a}\right] \leq \left(\sup_{x>a} \frac{x}{\Phi(x)}\right) \mathbb{E}\left[\Phi(|X|)\right].$$
- *Tout sous ensemble borné dans L^p est uniformément intégrable.*

Le nom "uniformément intégrable " est justifié par la proposition suivante.

Proposition 3.5.2. *Soit $(X_i)_{i\in I}$ une famille de variable aléatoire bornée dans L^1. Les propositions suivantes sont équivalentes.*

i) La famille $(X_i)_{i\in I}$ est uniformément intégrable.

ii) Pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que, pour tout $A \in \mathcal{A}$ tel que $\mathbb{P}(A) < \delta$, on a
$$\forall i \in I, \qquad \mathbb{E}\left[|X_i| \mathbb{1}_A\right] \leq \varepsilon.$$

On remarque que le premier exemple ci-dessus correspond à l'hypothèse principale du théorème de convergence dominée. En fait, l'uniforme intégrabilité est une condition nécessaire et suffisante pour pouvoir appliquer ce résultat (voir le Théorème 12.5.3 dans [10]). Le théorème suivant applique ce résultat aux martingales.

Théorème 27. *Soit $(X_n)_{n\in\mathbb{N}}$ une martingale par rapport à une filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$. Les trois propositions suivantes sont équivalentes.*

i) X_n converge p.s. et dans L^1 vers une variable aléatoire X_∞.

ii) *La suite* $(X_n)_n$ *est uniformément intégrable.*

iii) *Il existe une variable aléatoire* $Z \in L^1$ *telle que* $X_n = \mathbb{E}[Z \mid \mathcal{F}_n]$ *pour tout* $n \in \mathbb{N}$. *On dit dans ce cas que* $(X_n)_{n\in\mathbb{N}}$ *est une martingale fermée.*

Dans ce cas, $X_\infty = \mathbb{E}[Z \mid \mathcal{F}_\infty]$, *où* $\mathcal{F}_\infty = \sigma(\cup_{n\in\mathbb{N}}\mathcal{F}_n)$.

Chapitre 4

Chaînes de Markov

4.1 Définitions

Dans toute la suite, E est un espace fini ou dénombrable muni de la tribu $\mathcal{P}(E)$.

Définition 4.1.1. *$Q = \{Q(x,y), x,y \in E\}$ est une matrice de transition sur E (également matrice stochastique ou markovienne) si*

i) pour tout $x, y \in E$, $0 \leq Q(x,y) \leq 1$,

ii) pour tout $x \in E$, $\sum_{y \in E} Q(x,y) = 1$.

On remarque que Q définit une famille de lois sur E, indexées par E, via la formule

$$Q(x,A) = \sum_{y \in A} Q(x,y), \quad A \in \mathcal{P}(E).$$

Une chaîne de Markov est alors un processus dont l'état au temps $n+1$ est tiré suivant la loi $Q(x,\cdot)$, où x est l'état du processus au temps n.

Définition 4.1.2 (Chaîne de Markov)**.** *Soit Q une matrice de transition sur E et soit $(X_n)_n$ une suite de v.a. à valeurs dans E. $(X_n)_n$ est une chaîne de Markov de matrice de transition Q si, pour tout $n \in \mathbb{N}$,*

$$\mathbb{P}\left(X_{n+1} = y \mid X_0 = x_0, \ldots, X_n = x_n\right) = Q(x_n, y),$$

pour tout $x_0, \ldots, x_n, y \in E$ tels que $\mathbb{P}\left(X_0 = x_0, \ldots, X_n = x_n\right) > 0$.

La propriété essentielle ici est que le futur du processus ne dépend du passé qu'à travers l'état présent.

Comme exemple, une suite de v.a. i.i.d. est une chaîne de Markov, dans ce cas $Q(x,y)$ ne dépend pas de x. Les processus de branchement sont un autre exemple de chaîne de Markov.

Exemple. *Soit $(\xi_n)_n$ une suite de v.a. i.i.d. uniformes sur $\{-1,1\}$. On pose $S_0 = 0$, $S_n = \xi_1 + \ldots + \xi_n$, et $X_n = \max\{S_m : 0 \leq m \leq n\}$. Alors $(X_n)_n$ n'est **pas** une chaîne de Markov.*

EXERCICE 28 - Soit $(\xi_i)_{i \in \mathbb{N}}$ une suite de v.a. iid à valeurs dans un espace F et $f : E \times F \to E$ une application mesurable. On suppose que la suite $(X_n)_n$ vérifie

$$X_{n+1} = f(X_n, \xi_{n+1}).$$

Montrer que c'est une chaîne de Markov.

Un exercice du TD montrera que toute chaîne de Markov peut se construire de cette façon. Un autre exemple important de chaîne de Markov est la marche aléatoire sur un graphe. En fait, n'importe quelle chaîne de Markov peut se définir via un graphe orienté, dont les sommets sont les éléments de E et où l'arrête $x \to y$ est présente si $Q(x,y) > 0$ (il est également d'usage d'indiquer la valeur de $Q(x,y)$ à côté de l'arrête).

Notation Si Q est une matrice de transition, on pose, comme pour un produit matriciel,

$$Q^2(x,y) = \sum_{z \in E} Q(x,z) Q(z,y),$$

qui est encore une matrice de transition, et on définit Q^n par récurrence.

Proposition 4.1.3. *$(X_n)_n$ est une chaîne de Markov de matrice de transition Q si et seulement si, pour tout $x_0, \ldots, x_n \in E$,*

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

Démonstration. Immédiate par récurrence à partir de la définition. ☐

La proposition permet d'obtenir les résultats suivants.

Proposition 4.1.4. *Soit $(X_n)_n$ une chaîne de Markov de matrice de transition Q.*

i) Si $\mathbb{P}(X_0 = x_0) > 0$, $\mathbb{P}(X_n = x_n \mid X_0 = x_0) = Q^n(x_0, x_n)$.

ii) Si $f : E \to \mathbb{R}$ est mesurable et intégrable,

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

iii) Pour $p \geq 1$,

$$\mathbb{E}[f(X_{n+p}) \mid X_n] = Q^p f(X_n).$$

Démonstration. Exercice ! □

Dans la suite, on notera souvent $\mathbb{P}_x(\cdot) = \mathbb{P}(\cdot \mid X_0 = x)$, et $\mathbb{P}_\mu$ la loi de $(X_n)_n$ lorsque $X_0 \sim \mu$. Ainsi,

$$\mathbb{E}_\mu[f(X_n)] = \mu Q^n f = \sum_{x \in E} \mu(x) Q^n f(x).$$

4.2 Propriété de Markov

Propriété de Markov simple

Définition 4.2.1. *Si $X = (X_1, X_2, \ldots) \in E^{\mathbb{N}}$, on définit un* ***opérateur de translation*** *$\theta_n : E^{\mathbb{N}} \to E^{\mathbb{N}}$ par*

$$\theta_n X = (X_{n+1}, X_{n+2}, \ldots) = (X_{n+k})_{k \geq 1}.$$

Théorème 29 (Propriété de Markov simple)**.** *Soit $f : E^{\mathbb{N}} \to \mathbb{R}_+$ une fonction mesurable positive, alors*

$$\mathbb{E}[f(\theta_n X) \mid X_0, \ldots, X_n] = \mathbb{E}_{X_n}[f(X)].$$

Autrement dit, la loi de la chaîne de Markov à partir de l'instant $n+1$ sachant le passé est la loi de la chaîne de valeur initiale X_n.

Propriété de Markov forte Ce résultat s'étend également au cas où l'instant n est aléatoire, sous certaines hypothèses. Notons $\mathcal{F}_n = \sigma(X_0, \ldots, X_n) = \sigma(\{\{X_0 = x_0, \ldots, X_n = x_n\}, x_0, \ldots, x_n \in E\})$ la filtration canonique associée à $(X_n)_n$.

Définition 4.2.2. *Soit T un temps d'arrêt, on définit la tribu $\mathcal{F}_T$ par*

$$\mathcal{F}_T = \{A \in \sigma(\cup_n \mathcal{F}_n) : A \cap \{T = n\} \in \mathcal{F}_n\}.$$

La tribu $\mathcal{F}_T$ contient l'information disponible sur le processus $(X_n)_n$ au temps T. On a alors la généralisation suivante de la propriété de Markov simple.

Théorème 30 (Propriété de Markov forte)**.** *Soit $X = (X_n)_n$ une chaîne de Markov de matrice de transition Q et soit T un temps d'arrêt pour $(\mathcal{F}_n)_n$. Soient f et g deux fonctions mesurables positives sur $E^{\mathbb{N}}$. Supposons que la variable aléatoire $f \circ X$ est mesurable par rapport à la tribu $\mathcal{F}_T$. Alors,*

$$\mathbb{E}[\mathbb{1}_{T<\infty} f(X) g(\theta_T X)] = \mathbb{E}[\mathbb{1}_{T<\infty} f(X) \mathbb{E}_{X_T}[g(X)]],$$

où la v.a. $\mathbb{E}_{X_T}[g(X)]$, définie sur l'événement $\{T < \infty\}$, est la composée des applications $\omega \mapsto X_{T(\omega)}(\omega)$ et $x \mapsto \mathbb{E}_x[g(X)]$.

Remarque. *En comparant cet énoncé à la Proposition 2.2.1, on peut réécrire le Théorème ci-dessus de la manière suivante,*

$$\mathbb{E}[\mathbb{1}_{T<\infty} g(\theta_T X) \mid \mathcal{F}_T] = \mathbb{1}_{T<\infty} \mathbb{E}_{X_T}[g(X)].$$

4.3 Irréductibilité, mesure invariante

Définition 4.3.1. *Soit Q une matrice de transition sur E. Une mesure $\mu = (\mu(x))_{x \in E}$ sur E est dite invariante si*

$$\forall y \in E, \quad \sum_{x \in E} \mu(x) Q(x, y) = \mu(y).$$

Ce qui s'écrit également $\mu Q = \mu$.

Exemple. *Si $(X_n)_n$ est une marche aléatoire symétrique sur $\mathbb{Z}^d$, alors la mesure de comptage $\mu(x) = 1$ est invariante.*

Remarque. *Si μ est une probabilité invariante (i.e. $\sum_x \mu(x) = 1$) et si $X_0 \sim \mu$, alors $X_n \sim \mu$ pour tout $n \in \mathbb{N}$.*

On dit qu'une mesure μ sur E est **réversible** si, pour tous $x, y \in E$,

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

On vérifie aisément que toute mesure réversible est invariante.

La première question qu'il est naturel de se poser est celle de l'existence de mesures invariantes.

Définition 4.3.2. *Soit $x \in E$. On note T_x le premier instant de retour en x de la chaîne de Markov :*

$$T_x = \inf\{n \geq 1 : X_n = x\}.$$

*L'état x est dit **récurrent** si $\mathbb{P}_x(T_x < \infty) = 1$, il est dit **transitoire** (ou transient) si $\mathbb{P}_x(T_x < \infty) < 1$.*

On montre (cf. TD) que si x est récurrent, alors, $\mathbb{P}_x$-presque sûrement, X_n passe une infinité de fois par l'état x, et que, si x est transitoire, la chaîne ne passe qu'un nombre fini de fois en x.

L'existence d'un état récurrent permet de construire une mesure invariante en sommant sur toutes les excursions possibles entre deux passages en x.

Théorème 31. *Soit $x \in E$ un état récurrent, alors*

$$\mu_x(y) = \mathbb{E}_x\left[\sum_{n=0}^{T_x - 1} \mathbb{1}_{X_n = y}\right]$$

définit une mesure invariante.

La deuxième question naturelle est celle de l'unicité de la mesure invariante. Le premier cas à évacuer est celui où plusieurs états récurrents ne "communiquent" pas.

Définition 4.3.3. *Une matrice de transition Q est dite **irréductible** si, pour tous $x, y \in E$, il existe $n \geq 1$ tel que $Q^n(x,y) > 0$.*

Exemple. *Une marche aléatoire simple sur un graphe connexe est irréductible.*

Proposition 4.3.4. *Soit $(X_n)_n$ une chaîne de Markov irréductible, alors soit tous les états sont récurrents, soit ils sont tous transitoires. Si E est fini, seul le premier cas peut se produire.*

L'idée de la preuve consiste à dire que si X_n visite l'état x une infinité de fois, et que $\mathbb{P}(X_{n+p} = y \mid X_n = x) > 0$, alors, en utilisant la propriété de Markov forte, on montre que X_n visite l'état y une infinité de fois.

Le résultat suivant montre que l'unicité est automatique dans le cas irréductible récurrent.

Théorème 32. *Soit $(X_n)_n$ une chaîne de Markov irréductible récurrente, alors la mesure invariante est unique à une constante multiplicative près.*

Une preuve dans un cas particulier est donnée dans le TD.

La dernière question à laquelle on souhaite répondre est celle de l'existence d'une probabilité invariante, *i.e.* d'une mesure invariante de masse totale finie (si c'est le cas on peut toujours renormaliser pour obtenir une probabilité).

Proposition 4.3.5. *Soit $(X_n)_n$ une chaîne de Markov irréductible récurrente.*

i) Soit il existe une probabilité invariante π, et alors, pour tout $x \in E$,

$$\mathbb{E}_x[T_x] = \frac{1}{\pi(x)}.$$

Ce cas est appelé cas récurrent positif, c'est toujours le cas qui se produit si E est fini.

ii) Soit toute mesure invariante a une masse totale infinie, et, pour tout $x \in E$,

$$\mathbb{E}_x[T_x] = \infty.$$

Ce cas est appelé cas récurrent nul.

Si E est fini, seul le premier cas peut se produire. Comme corollaire de ce résultat, on montre que, pour une chaîne de Markov irréductible, l'existence d'une probabilité invariante suffit à assurer que tous les états sont récurrents.

Proposition 4.3.6. *Soit $(X_n)_n$ une chaîne de Markov irréductible. Les trois propositions suivantes sont équivalentes.*

i) Il existe $x \in E$ tel que $\mathbb{E}_x[T_x] < \infty$.

ii) Il existe une probabilité invariante π.

iii) la chaîne $(X_n)_n$ est récurrente positive.

4.4 Comportement asymptotique

On conclut avec un résultat d'ergodicité dans le cas récurrent positif.

Théorème 33. *Soit $(X_n)_n$ une chaîne de Markov irréductible récurrente positive, de probabilité invariante π. Alors,*

$$\frac{1}{n}\sum_{k=0}^{n} f(X_k) \underset{n\to\infty}{\longrightarrow} \langle \pi, f\rangle, \qquad \mathbb{P}_x - p.s.$$

où $\langle \pi, f\rangle = \pi f = \sum_{x\in E} \pi(x) f(x)$.

Ainsi, dans ce cas, la chaîne de Markov "oublie" son état initial et "converge" vers sa loi stationnaire.

Proposition 4.4.1. *Soit* $(X_n)_n$ *une chaîne de Markov irréductible récurrente, alors, pour tout* $x \in E$,

i) dans le cas récurrent positif,

$$\frac{1}{n}\sum_{k=0}^{n} \mathbb{1}_{X_n=x} \xrightarrow[n\to\infty]{p.s.} \pi(x) > 0,$$

ii) dans le cas récurrent nul,

$$\frac{1}{n}\sum_{k=0}^{n} \mathbb{1}_{X_n=x} \xrightarrow[n\to\infty]{p.s.} 0.$$

Ce résultat est à l'origine de la terminologie "récurrent positif/nul".

Annexe A

Inégalités usuelles

Bienaymé-Chebyshev Pour tout $p \geq 1$ et tout $a > 0$,

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|^p]}{a^p}. \tag{A.1}$$

Cauchy-Schwarz Pour $X, Y \in L^2$,

$$|\mathbb{E}[XY]| \leq \mathbb{E}\left[X^2\right]^{1/2} \mathbb{E}\left[Y^2\right]^{1/2}.$$

Jensen Soit $\phi : \mathbb{R} \to \mathbb{R}$ mesurable convexe, alors

$$\mathbb{E}[\phi(X)] \geq \phi(\mathbb{E}[X]),$$

si le membre de gauche est bien défini.

Hölder Si $\frac{1}{p} + \frac{1}{q} = 1$, $p, q \in [1, \infty]$, pour $X \in L^p$, $Y \in L^q$,

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{1/p} \mathbb{E}[|Y|^q]^{1/q}.$$

Minkowski Si $p \in [1, \infty]$, pour $X, Y \in L^p$,

$$\mathbb{E}[|X+Y|^p]^{1/p} \leq \mathbb{E}[|X|^p]^{1/p} + \mathbb{E}[|Y|^p]^{1/p}.$$

Bibliographie

[1] Paolo Baldi, Laurent Mazliak et Pierre Priouret. *Martingales et chaînes de Markov : théorie élémentaire et exercices corrigés*. Hermann, 2000.

[2] Philippe Barbe et Michel Ledoux. *Probabilités*. Belin, 1998.

[3] Michel Benaïm et Nicole El Karoui. *Promenade aléatoire : chaînes de Markov et simulations : martingales et stratégies*. Editions Ecole Polytechnique, 2005.

[4] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.

[5] Kai Lai Chung. *Markov chains*. Springer, 1967.

[6] Rick Durrett. *Probability : theory and examples*. Cambridge university press, 2010. url : https://www.math.duke.edu/~rtd/PTE/PTE4_1.pdf.

[7] Willliam Feller. *An introduction to probability theory and its applications*. John Wiley & Sons, 2008.

[8] Carl Graham. *Chaînes de Markov : cours, exercices et corrigés détaillés*. Dunod, 2008.

[9] Olav Kallenberg. *Foundations of modern probability*. springer, 2002.

[10] Jean-François Le Gall. “Intégration, probabilités et processus aléatoires”. In : *Ecole Normale Supérieure de Paris* (2006). url : http://www.math.u-psud.fr/~stafav/IMG/pdf/Legall.pdf.

[11] Sylvie Méléard. *Aléatoire : Introduction à la théorie et au calcul des probabilités*. Editions Ecole Polytechnique, 2010.

[12] James R. Norris. *Markov chains*. Cambridge university press, 1998.

[13] Daniel Revuz. *Probabilités*. Hermann, 1997.

[14] Daniel W. Stroock. *An introduction to Markov processes*. T. 230. Springer Science & Business Media, 2005.