

MASTER MVA
ENS PARIS-SACLAY

Rappels de probabilités

Thibaut Mastrolia

thibaut.mastrolia@polytechnique.edu

Basé sur le polycopié des années précédentes rédigé par Raphaël Deswarte et Raphaël Forien.

21-22 septembre 2017

Chapitre 1

Rappels de Probabilités

1.1 Définitions, exemples

1.1.1 Espace de probabilité

Considérons une expérience aléatoire. L'ensemble des réalisations possibles de cette expérience est noté classiquement Ω , appelé *l'univers*. On note alors $\mathcal{P}(\Omega)$ la collection des sous-ensembles de Ω .

Exemple 1.1.1. *Supposons une expérience aléatoire ayant trois réalisations possibles a , b ou c . On a $\Omega = \{a, b, c\}$. Ainsi $\mathcal{P}(\Omega)$ est composé de huit éléments : l'ensemble vide noté \emptyset , l'univers Ω , les singletons $\{a\}$, $\{b\}$ et $\{c\}$ et les ensembles $\{a, b\}$, $\{b, c\}$, $\{a, c\}$.*

Considérons maintenant un sous ensemble \mathcal{A} de $\mathcal{P}(\Omega)$ vu comme l'ensemble des réalisations aléatoires possibles pour lesquelles nous pouvons associer à chacune une certaine pondération ou probabilité. Ce sous ensemble \mathcal{A} ne doit pas être arbitraire, par exemple en reprenant l'exemple ci-dessus si $\{a\}$ et $\{b\}$ sont possibles, alors $\{a, b\}$ doit l'être également. On introduit donc la définition suivante

Définition 1.1.1. *On dit que $\mathcal{A} \subset \mathcal{P}(\Omega)$ est une **tribu ou σ -algèbre** (ou encore σ -field en anglais) si*

- Ω et \emptyset sont inclus dans \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.

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

EXERCICE 2 - Soit $\Omega = \{1, 2, 3, 4, 5, 6\}$ (comme c'est le cas pour un lancer de dé), 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.

Pour la suite, on supposera implicitement que \mathcal{A} est une tribu de Ω . Introduisons maintenant une notion de pondération (et plus généralement de mesure) sur chacun des évènements de \mathcal{A} .

Définition 1.1.2. Soit (Ω, \mathcal{A}) où \mathcal{A} est une tribu de Ω . On dit que $\mathbb{P} : \mathcal{A} \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ est une *mesure* si

- $\mathbb{P}(\emptyset) = 0$,
- pour toute suite dénombrable $(A_n)_{n \in \mathbb{N}}$ d'éléments deux à deux disjoints, $\mathbb{P}(\cup_n A_n) = \sum_n \mathbb{P}(A_n)$.

On dit alors que $(\Omega, \mathcal{A}, \mathbb{P})$ est une espace *mesurable*.

Si de plus $\mathbb{P}(\Omega) = 1$, on dit que \mathbb{P} est une *mesure de probabilité* (ou *probabilité pour simplifier*) et $(\Omega, \mathcal{A}, \mathbb{P})$ est un *espace probabilisé* ou *espace de probabilité*.

EXERCICE 3 - Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. 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} alors

$$\mathbb{P}(\cup_n A_n) = \lim_{n \uparrow +\infty} \mathbb{P}(A_n).$$

Montrer la propriété analogue pour une suite décroissante.

4. (**Formule des probabilités totales**) Soit $A \subset \mathcal{A}$ et une suite $(B_n)_n$ d'évènements disjoints tels que $\cup_n B_n = \Omega$. Alors

$$\mathbb{P}(A) = \sum_n \mathbb{P}(A \cap B_n).$$

Exemples 1.1.1. Quelques exemples de mesures usuelles.

- 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 $\sum_{\omega \in \Omega} p_\omega = 1$.

- $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ (Théorème 1.1 ci-dessous) et

$$\mathbb{P}(]a, b[) = \int_a^b p(x) dx,$$

où $0 \leq p(x) \leq 1$ et $\int_{-\infty}^{+\infty} p(x) dx = 1$.

- La *mesure uniforme* notée μ sur un espace fini Ω est définie par

$$\mu(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}, \quad A \subset \Omega.$$

- La *mesure de Dirac en un point* a notée δ_a est définie par

$$\delta_a(A) = \mathbf{1}_{a \in A}.$$

1.1.2 Tribu borélienne

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

Définition 1.1.3 (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. Un élément de $\mathcal{B}(\mathbb{R})$ est appelé **un borélien**.

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} .*

Plus en détails, la tribu $\mathcal{B}(\mathbb{R})$ contient bien évidemment les ouverts de \mathbb{R} mais aussi par exemple les intervalles fermés puisque $[a, b] = (] - \infty, a[\cup]b, +\infty[)^c$.

Équipons maintenant l'espace $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ d'une mesure fondamentale.

Théorème 1.1.1. *Il existe une **unique** mesure λ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Cette mesure vérifie pour tout réels $a \leq b$,*

$$\lambda(]a, b]) = b - a. \quad (1.1)$$

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

Introduisons maintenant une notion de tribu sur un ensemble d'espaces, afin de lier notamment la tribu borélienne sur \mathbb{R} avec la tribu borélienne sur \mathbb{R}^2 (et plus généralement sur \mathbb{R}^n).

Définition 1.1.4 (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 1.1.5. $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$.

1.1.3 Variables aléatoires

En pratique, on omet d'expliciter l'espace $(\Omega, \mathcal{A}, \mathbb{P})$. On suppose qu'il est *assez "riche"* pour pouvoir y définir un ensemble de fonctions sur cet espace. Par la suite, nous dirons qu'un évènement $A \in \Omega$ a lieu \mathbb{P} -presque sûrement (ou presque sûrement s'il n'y a pas d'ambiguïté sur \mathbb{P}) si $\mathbb{P}(A) = 1$.

Définition 1.1.6 (Mesurabilité de fonctions). *Soient E et F deux espace mesurables muni de leurs tribus respectives \mathcal{E} et \mathcal{F} . Une application $f : (E, \mathcal{E}) \rightarrow (F, \mathcal{F})$ est dite **mesurable** (par rapport à la tribu \mathcal{E}) si*

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

On a en particulier le lemme suivant

Lemme 1.1.7. *Pour qu'une application $f : (E, \mathcal{E}) \rightarrow (F, \mathcal{F})$ soit mesurable, il suffit que $f^{-1}(B) \in \mathcal{E}$ pour $B \in C \subset \mathcal{F}$ si $\sigma(C) = \mathcal{F}$.*

(Indication pour la preuve : montrer que $\{B \subset F : f^{-1}(B) \in \mathcal{E}\}$ est une tribu).

Il en découle alors (preuve en exercice) le résultat fondamental suivant

Proposition 1.1.8. *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 1.1.9. *Une **variable aléatoire** (v.a. en abrégé) sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans (E, \mathcal{E}) est une application mesurable $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$.*

Si de plus, une variable aléatoire est à valeur dans un intervalle de \mathbb{R} , on dit que cette variable aléatoire est continue.

Deux variables aléatoires X, Y à valeurs dans un même espace sont égales presque sûrement si $\mathbb{P}(X = Y) = 1$.

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

À une variable aléatoire sur une espace probabilisé, on peut alors lui associer l'ensemble des pondérations correspondant à chaque évènements, permettant ainsi de caractériser pleinement cette v.a. Plus formellement, on pose la définition suivante

Définition 1.1.10. *La **loi** de la v.a. X est la mesure image P_X de \mathbb{P} par X , i.e.*

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

Exemple. Pour une variable aléatoire discrète : $P_X = \sum_{x \in E} \mathbb{P}(X = x) \delta_x$, autrement dit,

$$P_X(B) = \sum_{x \in B} \mathbb{P}(X = x).$$

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

Pour des variables aléatoires réelles, une notion importante associées à la loi de celle-ci est la notion de densité définie comme suit

Définition 1.1.11 (Densité de probabilité). Soit X une variable aléatoire réelle sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ de loi P_X . On dit que P_X possède une densité s'il existe une fonction $p : \mathbb{R} \mapsto \mathbb{R}^+$ intégrable telle que

$$P_X(B) = \mathbb{P}(X \in B) = \int_B p(x) dx, \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

On dit que p est la **densité de probabilité** associée à X .

En particulier, on a toujours pour une variable aléatoire réelle $\int_{\mathbb{R}} p(x) dx = 1$. Définissons maintenant la notion de répartition d'une v.a. sur son domaine

Définition 1.1.12. On appelle **fonction de répartition** d'une v.a. réelle X , l'application $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par

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

Définition 1.1.13 (Espérance). Soit X une v.a. à valeurs dans (E, \mathcal{E}) et $f : E \rightarrow \mathbb{R}$ une application mesurable. On suppose que f est soit positive, soit intégrable par rapport à la loi de X (i.e. $\int |f| dP_X < \infty$). On pose

$$\mathbb{E}[f(X)] = \int_E f(x) P_X(dx).$$

Exemple. — Cas discret : $\mathbb{E}[f(X)] = \sum_{x \in E} f(x) \mathbb{P}(X = x)$.
— Cas réel à densité : $\mathbb{E}[f(X)] = \int_{-\infty}^{+\infty} f(x) p(x) dx$.

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

1.1.4 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 $h(X)$ où h est dans une certaine classe de fonctions.

Théorème 1.1.2 (Théorème d'identification). *i) Si μ est une probabilité sur (E, \mathcal{E}) telle que, pour toute fonction $h : E \rightarrow \mathbb{R}$ continue bornée,*

$$\mathbb{E}[h(X)] = \int_E h(x)\mu(dx),$$

alors $P_X = \mu$.

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

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x)f(x) dx,$$

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 $f : E \rightarrow F$ mesurable. On écrit alors

$$\mathbb{E}[h(f(X))] = \int_E h(f(x))P_X(dx)$$

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

1.2 Indépendance

1.2.1 Définition générale et résultats fondamentaux

Définition 1.2.1. *On dit que deux événements $A, B \in \mathcal{A}$ sont **indépendants** si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Plus généralement, les événements A_1, \dots, A_n sont indépendants si*

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

On dit que les v.a. X_1, \dots, X_n sont indépendantes si

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

On dit que la suite de v.a. $(X_n)_{n \in \mathbb{N}}$ est indépendante si, pour tout $n \in \mathbb{N}$, X_1, \dots, X_n sont indépendantes.

Théorème 1.2.1. *Les variables X_1, \dots, X_n sont indépendantes si et seulement si*

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

Le Théorème 1.2.1 permet de caractériser les v.a. indépendantes de différentes façons.

Corollaire 1.2.2. Soit X_1, \dots, X_n des v.a. réelles.

1. X_1, \dots, X_n sont indépendantes si et seulement si, pour toutes fonctions f_1, \dots, f_n continues à support compact de \mathbb{R} dans \mathbb{R} ,

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

2. les composantes du vecteur aléatoire (X_1, \dots, X_n) sont indépendantes si et seulement si

$$F_{(X_1, \dots, X_n)} = \prod_{i=1}^n F_{X_i},$$

où F_X (resp. F_{X_i} , $1 \leq i \leq n$) est la fonction de répartition de $X := (X_1, \dots, X_n)$ (resp. X_i , $1 \leq i \leq n$).

3. Soit (X_1, \dots, X_n) un vecteur aléatoire de densité donnée par p_X . Les composantes de densités p_i sont indépendantes si et seulement si

$$p_X(x_1, \dots, x_n) = \prod_{i=1}^n p_i(x_i).$$

Démonstration. Utiliser le Théorème 1.1.2 et le Théorème de Fubini. \square

Remarque. En pratique, on a également le résultat suivant qui est une généralisation de la réciproque du point 3. précédent. Si (X_1, \dots, X_n) a une densité de la forme

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

où les q_i sont boréliennes positives, alors X_1, \dots, X_n sont indépendantes et X_i admet une densité de la forme $p_i = C_i q_i$, $C_i > 0$ (où C_i est choisi pour s'assurer que p_i est bien de masse 1).

Remarque. Attention ! Pour que X_1, \dots, 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.

$$P_X * P_Y(B) = \int P_Y(B - x) P_X(dx).$$

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.

1.2.2 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,

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

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 1.2.4 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

1.3.1 Loi faible des grands nombres

Théorème 1.3.1 (Loi faible des grands nombres). *Soit $(X_n)_n$ une suite de v.a. i.i.d. telles que $\mathbb{E}[X_1^2] < \infty$, alors*

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

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

Démonstration. On écrit

$$\begin{aligned}\mathbb{E} \left[\left(\frac{1}{n} (X_1 - \mu + \dots + X_n - \mu) \right)^2 \right] &= \frac{1}{n^2} \mathbb{V} [X_1 + \dots + X_n] \\ &= \frac{1}{n} \mathbb{V} [X_1] \xrightarrow{n \rightarrow \infty} 0.\end{aligned}$$

(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

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

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

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

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

□

1.3.2 Loi forte des grands nombres

On a en réalité un résultat beaucoup plus fort que celui du Théorème 1.3.1. 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

$$\forall \omega \notin N, \quad X_n(\omega) \xrightarrow{n \rightarrow \infty} X(\omega).$$

Théorème 1.3.2 (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

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

Une preuve de ce résultat utilise par exemple 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 (voir section suivante), la seconde par le Lemme de Borel-Cantelli.

On donne ci-dessous 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 $X_n(\omega) = \mathbf{1}_{\{A_n\}}(\omega)$, $n \geq 1$ 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 $X_n \xrightarrow[n \rightarrow \infty]{L^1} 0$ mais, comme la série des $\frac{1}{n}$ diverge, $X_n \not\xrightarrow{ps} 0$.

EXERCICE 5 - Exhiber une sous-suite qui converge ps vers zéro.

1.3.3 Limite et espérance

Un résultat essentiel est celui de passage à la limite sous l'espérance lorsque X_n converge vers X en probabilité ou presque sûrement. On a les deux résultats fondamentaux suivant

Proposition 1.3.5 (Lemme de Fatou). *Soit une suite $(X_n)_n$ de v.a. positives presque sûrement. Si X_n converge en probabilité vers une variable aléatoire X , on a*

$$\liminf_{n \rightarrow +\infty} \mathbb{E}[X_n] \geq \mathbb{E}[X].$$

Théorème 1.3.3 (Convergence dominée). *Soit une suite $(X_n)_n$ de v.a. convergeant presque sûrement vers une variable aléatoire X . Supposons qu'il existe une v.a. Y intégrable, i.e. $\mathbb{E}[|Y|] < +\infty$, telle que*

$$|X_n| \leq |Y|, \mathbb{P} - p.s. \quad (1.3)$$

alors X est intégrable et $\lim_{n \rightarrow +\infty} \mathbb{E}[|X_n - X|] = 0$.

Remarque. Le théorème de convergence dominée donné ici est en réalité une version plus forte du théorème de convergence dominée classique. En effet, il suffit d'avoir l'uniforme intégrabilité de la famille (X_n) au lieu du critère de domination (1.3) pour conclure. Par ailleurs, de manière générale, on peut montrer que si une suite $(X_n)_n$ de v.a. converge presque sûrement vers une variable aléatoire X telle que (1.3) est satisfaite, X_n converge dans L^p vers X pour tout p .

1.3.4 Convergence en loi

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

Définition 1.3.6. *Une suite de v.a. $(X_n)_n$ à valeurs dans E converge **en loi** vers X si pour toute fonction $f : E \rightarrow \mathbb{R}$ continue bornée,*

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

On note alors $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$ ou $X_n \xrightarrow[n \rightarrow \infty]{d} X$.

Proposition 1.3.7. — Si X_n et X sont à valeurs dans un espace dénombrable E , alors $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$ si et seulement si,

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

— Si X_n admet p_n pour densité sur \mathbb{R} , alors si $p_n \rightarrow p$ 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 $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$, où X admet p pour densité.

Proposition 1.3.8. 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 1.3.4. Soient $(X_n)_n$ et X des v.a. à valeurs dans \mathbb{R}^d . Les propositions suivantes sont équivalentes.

- i) $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$,
- ii) pour tout ouvert $G \subset \mathbb{R}^d$, $\liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G)$,
- iii) pour tout fermé $F \subset \mathbb{R}^d$, $\limsup_{n \rightarrow \infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$,
- iv) pour tout borélien $A \subset \mathbb{R}^d$ tel que $\mathbb{P}(X \in \partial A) = 0$, $\lim_{n \rightarrow \infty} \mathbb{P}(X_n \in A) = \mathbb{P}(X \in A)$.

Le résultat suivant découle du Théorème 1.3.4 et donne une caractérisation de la convergence en loi en terme de fonction de répartition.

Proposition 1.3.9. Soient $(X_n)_n$ et X des v.a. réelles. X_n converge en loi vers X si et seulement si $F_{X_n}(x) \xrightarrow[n \rightarrow \infty]{} F_X(x)$ pour tout x où F_X est continue.

Enfin, de façon générale, la convergence en loi n'implique pas la convergence en probabilité (contre exemple en TD), sauf si la convergence a lieu vers une constante, on a alors

Corollaire 1.3.10. Si $(X_n)_n$ converge en loi vers une v.a. X constante égale à c , alors $(X_n)_n$ converge en probabilité vers c .

Démonstration. On note F_n la fonction de répartition de X_n et F_X la fonction de répartition de X donnée par $F(x) = \mathbf{1}_{c \leq x}$ pour $x \in \mathbb{R}$. Alors :

$$\begin{aligned} 0 &\leq \mathbb{P}(|X_n - c| > \varepsilon) \\ &= \mathbb{P}(X_n > \varepsilon + c \cup X_n < c - \varepsilon) \\ &\leq 1 - F_n(c + \varepsilon) + F_n(c - \varepsilon). \end{aligned}$$

On en déduit la convergence en probabilité de la convergence en loi d'après la Proposition 1.3.9. □

1.3.5 Fonction caractéristique

Définition 1.3.11. Soit X une v.a. à valeurs dans \mathbb{R}^d . Sa **fonction caractéristique** $\Phi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ est définie par

$$\Phi_X(x) = \mathbb{E} [e^{ix \cdot X}], \quad x \in \mathbb{R}^d.$$

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 1.3.12. La fonction caractéristique caractérise la loi de X .

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

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

Proposition 1.3.13. Les v.a. X_1, \dots, X_n sont indépendantes si et seulement si $\Phi_{(X_1, \dots, X_n)} = \Phi_{X_1} \dots \Phi_{X_n}$.

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 1.3.5 (Lévy). Soient $(X_n)_n$ une suite de v.a. et X une v.a. à valeurs dans \mathbb{R}^d . Alors X_n converge en loi vers X si et seulement si, pour tout $x \in \mathbb{R}^d$, $\Phi_{X_n}(x) \xrightarrow[n \rightarrow \infty]{} \Phi_X(x)$.

1.3.6 Fonction génératrice des moments

Définition 1.3.14. Soit X une v.a. à valeurs dans \mathbb{R}^d . Sa **fonction génératrice des moments** $M_X : \mathbb{R}^d \rightarrow \mathbb{R}^+$ est définie par

$$M_X(x) = \mathbb{E} [e^{x \cdot X}], \quad x \in \mathbb{R}^d.$$

Remarque. $x \mapsto M_X(-x)$ coïncide avec celle de la transformée de Laplace de la loi de X .

Proposition 1.3.15. La fonction génératrice des moments caractérise la loi de X .

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

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

Proposition 1.3.16. Les v.a. X_1, \dots, X_n sont indépendantes si et seulement si $M_{(X_1, \dots, X_n)} = M_{X_1} \dots M_{X_n}$.

La fonction génératrice des moments nous donne un moyen très utile de caractériser les moments (comme son nom l'indique) de la v.a. considérée. Il suffit en effet de vérifier qu'elle est bien définie au voisinage de $x = 0$ et de dériver en x puis prendre la limite en 0 (sous réserve d'existence).

1.3.7 Théorème de la limite centrale/centrée

Théorème 1.3.6 (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

$$\frac{1}{\sqrt{n}}(X_1 + \dots + X_n - n\mu) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

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

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. $X = (X_1, \dots, X_d)$ 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, \dots, X_d) est un vecteur gaussien.

Théorème 1.4.1. 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

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

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

Théorème 1.4.2. Soit X un vecteur aléatoire gaussien, d'espérance m et de matrice de covariance C .

Si le rang de C est d (autrement dit, si $\text{Det}(C) > 0$), alors X admet une densité f_X par rapport à la mesure de Lebesgue sur \mathbb{R}^d avec :

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

Si le rang de C est $k < d$ (et donc $\text{Det}(C) = 0$), 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 1.4.1 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 combinaison linéaire des X_i est gaussienne, cf TD.

La proposition suivante montre qu'une transformation linéaire d'un vecteur gaussien donne un vecteur gaussien. ce résultat sera très utile en particulier pour transformer un vecteur gaussien en un vecteur à composante indépendante (cf TD).

Proposition 1.4.3. *Soit X un vecteur gaussien de \mathbb{R}^d de moyenne μ et de matrice de variance/covariance C et soit M une matrice réelle de taille $p \times d$. Alors MX est un vecteur gaussien de moyenne $M\mu$ et de covariance MCM^\top .*

Le théorème de la limite centrée se généralise sans peine au cas vectoriel.

Théorème 1.4.3 (Théorème de la limite centrée vectoriel). *Soit $(X_n)_n$ une suite de v.a. i.i.d. à valeurs dans \mathbb{R}^d , de covariance $C \in \mathcal{S}_d^+(\mathbb{R})$, de carré intégrable. Alors*

$$\frac{1}{\sqrt{n}}(X_1 + \dots + X_n - n\mathbb{E}[X_1]) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, C).$$

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 $\mathbb{P}(\cdot | B)$ sur (Ω, \mathcal{A}) par

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

EXERCICE 6 - Vérifier que $\mathbb{P}(\cdot | B)$ vérifie les axiomes de la Définition 1.1.2 d'une probabilité.

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

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

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 $\mathbb{E}[X | Y]$, 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

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

Supposons tout d'abord que E' est non vide, et prenons $y \in E'$. On choisit $B = \{Y = y\}$ dans la définition ci-dessus et, pour $X \in L^1$, on a

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

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

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

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

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

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

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

Alors,

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

EXERCICE 7 - Vérifier que $\mathbb{E}[Y | Y] = Y$ et que, si X et Y sont indépendants, $\mathbb{E}[X | Y] = \mathbb{E}[X]$.

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

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X | Y]] &= \mathbb{E}[X] \\ \mathbb{E}[|\mathbb{E}[X | Y]|] &\leq \mathbb{E}[|X|]. \end{aligned}$$

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

$$\mathbb{E}[f(Y)X] = \mathbb{E}[f(Y)\mathbb{E}[X | Y]] \quad (2.1)$$

Démonstration. Exercice! □

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

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

Proposition 2.1.4. Soient X et Y deux variables aléatoires discrètes, on pose

$$p(x, y) = \mathbb{P}(X = x, Y = y), \quad p_Y(y) = \mathbb{P}(Y = y).$$

Alors,

$$\mathbb{E}[X|Y] = \frac{\sum_x xp(x, Y)}{p_Y(Y)}.$$

2.2 Conditionnement par une tribu

Commençons par l'exercice suivant.

EXERCICE 8 - 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 $\mathbb{E}[X | Y_1] = \mathbb{E}[X | Y_2]$.

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,

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

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 $\mathbb{E}[X | \mathcal{F}]$, telle que

- i) $\mathbb{E}[X | \mathcal{F}]$ est \mathcal{F} -mesurable,
- ii) pour tout $A \in \mathcal{F}$, $\mathbb{E}[X \mathbf{1}_{\{A\}}] = \mathbb{E}[\mathbb{E}[X | \mathcal{F}] \mathbf{1}_{\{A\}}]$.

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

Corollaire 2.2.3. Soit $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et \mathcal{F} une sous-tribu de \mathcal{A} .

1. Si X est mesurable par rapport à \mathcal{F} , alors $\mathbb{E}[X | \mathcal{F}] = X$.
2. Si X est indépendant de \mathcal{F} , alors $\mathbb{E}[X | \mathcal{F}] = \mathbb{E}[X]$.
3. (Itération des espérances conditionnelles) Si \mathcal{F}_2 est une sous-tribu de \mathcal{F}_1 , alors $\mathbb{E}[\mathbb{E}[X | \mathcal{F}_1] | \mathcal{F}_2] = \mathbb{E}[\mathbb{E}[X | \mathcal{F}_2] | \mathcal{F}_1] = \mathbb{E}[X | \mathcal{F}_2]$.
4. Si Z est mesurable par rapport à \mathcal{F} et bornée alors $\mathbb{E}[ZX | \mathcal{F}] = Z\mathbb{E}[X | \mathcal{F}]$.

L'espérance conditionnelle hérite de toutes les propriétés de base de l'espérance : linéarité, positivité, inégalité triangulaire. Elle conserve également l'inégalité de Jensen. Avec cette définition, on vérifie que pour toute variable aléatoire discrète Y , $\mathbb{E}[X | Y] = \mathbb{E}[X | \sigma(Y)]$. Cela permet d'étendre la définition de l'espérance conditionnelle sachant une variable aléatoire à valeurs dans un espace général (pas forcément dénombrable).

Définition 2.2.4. Soit $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et Y une variable aléatoire sur Ω . On définit :

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

Pour des v.a. à densités, on peut calculer explicitement la v.a. $\mathbb{E}[X | Y]$ (cf **exercice 31** du TD très important en pratique).

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 $|\mathbb{E}[X | \mathcal{F}]|^2 \leq \mathbb{E}[X^2 | \mathcal{F}]$, et donc

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

Donc $\mathbb{E}[X | \mathcal{F}]$ 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, par la propriété 4. du corollaire 2.2.3.

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

Donc $X - \mathbb{E}[X | \mathcal{F}]$ est orthogonale à toutes les variables aléatoires bornées \mathcal{F} -mesurables. Par densité, on en déduit que $X - \mathbb{E}[X | \mathcal{F}]$ est orthogonale à toutes les variables aléatoires de $L^2(\Omega, \mathcal{F}, \mathbb{P})$. L'espérance conditionnelle $\mathbb{E}[X | \mathcal{F}]$ est donc le projeté orthogonal de X sur $L^2(\Omega, \mathcal{F}, \mathbb{P})$. En d'autres termes, $\mathbb{E}[X | \mathcal{F}]$ 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 2.3.1. Soit (Y_1, \dots, Y_n, X) un vecteur gaussien centré. Alors $\mathbb{E}[X | Y_1, \dots, Y_n]$ coïncide avec la projection orthogonale de X (dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$) sur l'espace vectoriel engendré par Y_1, \dots, Y_n . Il existe donc des réels $\lambda_1, \dots, \lambda_n$ tels que

$$\mathbb{E}[X | Y_1, \dots, Y_n] = \sum_{k=1}^n \lambda_k Y_k.$$

De plus, la loi conditionnelle de X sachant Y_1, \dots, 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 $h : \mathbb{R} \rightarrow \mathbb{R}_+$,

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

En d'autres termes, si (Y_1, \dots, Y_n, X) est un vecteur Gaussien, la meilleure manière d'approximer X (dans L^2) par une fonction mesurable de Y_1, \dots, 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 et joue un rôle essentielle dans de nombreux champs des mathématiques appliquées, notamment en mathématiques financières. 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 (Filtration). *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*

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

Définition 3.1.2 (Processus aléatoire). *Un processus aléatoire est une suite $(X_n)_n$ de variable aléatoire (ici à valeurs réelles) définies sur $(\Omega, \mathcal{A}, \mathbb{P})$. On dit qu'il est \mathbb{F} -adapté si pour tout n , chaque v.a. X_n est \mathcal{F}_n -mesurable.*

On appelle processus aléatoire croissant prévisible un processus aléatoire $(A_n)_n$ tel que $A_0 = 0$, A_n est \mathcal{F}_n -mesurable et $A_{n+1} \geq A_n$.

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

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

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 3.1.3. *Soit $(X_n)_{n \in \mathbb{N}}$ un processus aléatoire à valeurs réelles sur un espace $(\Omega, \mathcal{A}, \mathbb{P})$ filtré par \mathbb{F} .*

- $(X_n)_{n \in \mathbb{N}}$ est une **martingale** si $(X_n)_n$ est \mathbb{F} -adapté, c'est-à-dire que pour tout $n \in \mathbb{N}$, X_n est mesurable par rapport à \mathcal{F}_n tel que $\mathbb{E}[|X_n|] < \infty$ pour tout $n \in \mathbb{N}$, et vérifiant

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n, \quad n \in \mathbb{N}.$$

- $(X_n)_{n \in \mathbb{N}}$ est une **sous-martingale** si $(X_n)_n$ est \mathbb{F} -adapté, c'est-à-dire que pour tout $n \in \mathbb{N}$, X_n est mesurable par rapport à \mathcal{F}_n tel que $\mathbb{E}[|X_n|] < \infty$ pour tout $n \in \mathbb{N}$, et vérifiant

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] \geq X_n, \quad n \in \mathbb{N}.$$

- $(X_n)_{n \in \mathbb{N}}$ est une **sur-martingale** si $(X_n)_n$ est \mathbb{F} -adapté, c'est-à-dire que pour tout $n \in \mathbb{N}$, X_n est mesurable par rapport à \mathcal{F}_n tel que $\mathbb{E}[|X_n|] < \infty$ pour tout $n \in \mathbb{N}$, et vérifiant

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] \leq X_n, \quad n \in \mathbb{N}.$$

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.

- 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 $X_n = \xi_0 + \dots + \xi_n$ et $\mathcal{F}_n = \sigma(\xi_0, \dots, \xi_n)$ 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 $X_n = \mathbb{E}[X | \mathcal{F}_n]$. Alors $(X_n)_{n \in \mathbb{N}}$ est une martingale. Une telle martingale est dite fermée.

EXERCICE 9 - Démontrer les deux assertions ci-dessus.

On a le résultat fondamental suivant en terme de décomposition de sur/sous-martingale

Théorème 3.1.1 (Décomposition de Doob). Soit $(X_n)_n$ une sous-martingale. Il existe de façon unique une martingale $(M_n)_n$ et un processus croissant prévisible et intégrable $(A_n)_n$ (appelé le compensateur) tel que

$$X_n = M_n + A_n.$$

On a bien entendu le même résultat pour des sur-martingales avec un compensateur donné par un processus décroissant.

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 3.2.1 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 τ à valeurs dans $\mathbb{N} \cup \{\infty\}$ est un **temps d'arrêt** si pour tout $n \in \mathbb{N}$, $\{\tau = n\} \in \mathcal{F}_n$.

Si on imagine que τ 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}) ,*

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

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 3.2.1 (Théorème d'arrêt). *Soit $(X_n)_n$ une martingale (resp. une sous/surmartingale) et soit τ un temps d'arrêt. Alors $(X_{n \wedge \tau})_n$ est aussi une martingale (resp. une sous/surmartingale). En particulier, si le temps d'arrêt τ est borné, on a $X_\tau \in L^1$ et*

$$\mathbb{E}[X_\tau] = \mathbb{E}[X_0], \quad (\text{resp. } \mathbb{E}[X_\tau] \geq / \leq \mathbb{E}[X_0]).$$

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 $X_n = x_0 + \xi_1 + \dots + \xi_n$. Le processus $(X_n)_{n \in \mathbb{N}}$ est alors une martingale par rapport à sa filtration canonique. Pour $N \geq 1$ fixé, soit τ la variable aléatoire définie par*

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

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 3.2.1, pour tout $n \geq 1$,

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

Le temps d'arrêt τ n'est pas borné, mais $\tau < \infty$ p.s. donc, pour $n \rightarrow \infty$, $X_{n \wedge \tau} \rightarrow X_\tau$ presque sûrement. De plus, $|X_{n \wedge \tau}| \leq N$ pour tout $n \geq 1$. Par convergence dominée, on peut donc conclure que

$$\mathbb{E}[X_\tau] = x_0.$$

D'autre part,

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

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

3.3 Inégalités maximales

Théorème 3.3.1 (Inégalité maximale de Doob). *Soit $(X_n)_n$ une sous-martingale (ou une martingale) alors pour tout $a > 0$ et pour tout $n \in \mathbb{N}$,*

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

Le résultat suivant découle de ce théorème en remarquant que si $(X_n)_n$ est une martingale, $(|X_n|)_n$ est une sous-martingale

Corollaire 3.3.1. *Soit $(X_n)_n$ une martingale, alors pour tout $a > 0$ et pour tout $n \in \mathbb{N}$,*

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

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

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

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

Proposition 3.3.2. *Soit $p > 1$ et soit $(X_n)_n$ une martingale. Alors, pour tout $n \in \mathbb{N}$,*

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

3.4 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 3.4.1. *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 $n \rightarrow \infty$.

On a alors le corollaire suivant pour les sous/sur-martingale

Corollaire 3.4.1. *Soit $(X_n)_n$ une sous martingale (resp. surmartingale) majorée par une constante M (resp. minorée par une constante m). Alors, X_n converge p.s. vers X_∞ avec $\mathbb{E}[X_\infty | \mathcal{F}_n] \geq X_n$ (resp. $\mathbb{E}[X_\infty | \mathcal{F}_n] \leq X_n$).*

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

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

D'après le Théorème 3.2.1, $Y_n = X_{n \wedge \tau}$ 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 $\{\tau < \infty\}$ car $|X_n - X_{n+1}| = 1$ (on a donc montré au passage que $\tau < \infty$ p.s.). On a donc $X_\infty = 0$ p.s. On remarque qu'il ne peut y avoir convergence dans L^1 puisque $\mathbb{E}[X_n] = 1 > \mathbb{E}[X_\infty] = 0$.

3.5 Convergence L^p des martingales

Comme le montre le dernier exemple, une martingale peut converger presque sûrement sans converger dans L^1 et a fortiori L^p pour $p > 1$. Nous allons voir dans une première partie que le fait d'être bornée dans L^p pour $p > 1$ permet en revanche de conclure qu'une martingale converge dans L^p avec $p > 1$, la démonstration (omise ici) reposant principalement sur l'inégalité de Doob. Le cas $p = 1$ est en revanche plus technique, puisque la Proposition 3.3.2 ne peut plus être appliquée, et sera abordé après.

3.5.1 Le cas $p > 1$

Théorème 3.5.1. Soit $(X_n)_n$ une martingale. Supposons qu'il existe $p > 1$ tel que

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

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

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

et

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

3.5.2 Le cas $p = 1$

Comme le montre ce théorème, être bornée dans L^p avec $p > 1$ assure la convergence dans L^p pour $p > 1$. Cependant, pour le cas $p = 1$, être borné dans L^1 n'est pas suffisant pour avoir la convergence dans L^1 d'une martingale (voir contre exemple en TD). 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

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

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 $\Phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ une fonction telle que $x^{-1}\Phi(x) \xrightarrow{x \rightarrow \infty} 0$. Alors, pour tout $C > 0$,

$$\{X \in L^1 : \mathbb{E}[\Phi(|X|)] \leq C\}$$

est uniformément intégrable. En effet, il suffit d'écrire

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

— 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

$$\forall i \in I, \quad \mathbb{E}[|X_i| \mathbf{1}_{\{A\}}] \leq \varepsilon.$$

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 [le_gall_integration_2006]). Le théorème suivant applique ce résultat aux martingales.

Théorème 3.5.2. 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 $X_n = \mathbb{E}[Z | \mathcal{F}_n]$ 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, $X_\infty = \mathbb{E}[Z | \mathcal{F}_\infty]$, où $\mathcal{F}_\infty = \sigma(\cup_{n \in \mathbb{N}} \mathcal{F}_n)$.

Chapitre 4

Chaînes de Markov

4.1 Définitions

L'idée générale de ce chapitre est d'étudier l'évolution aléatoire d'un système par étapes successives sur un espace E fini ou dénombrable muni de la tribu $\mathcal{P}(E)$. Les évolutions possible de ce système d'un instant n à un instant $n + 1$ sera transcrit à travers une matrice de transition représentant les pondérations associées à chaque changement d'états. L'hypothèse fondamentale sera de considérer des systèmes markoviens, "*dont la valeur au temps $n + 1$ ne dépend du passé qu'à travers l'instant n* ".

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 $x, y \in E$, $0 \leq Q(x, y)$,
- ii) pour tout $x \in E$, $\sum_{y \in E} Q(x, y) = 1$.

Bien entendu (i) et (ii) implique que $0 \leq Q(x, y) \leq 1$.
On définit une famille de lois sur E , indexées par E , en posant

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

Une chaîne de Markov sur E de matrice Q 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 $n \in \mathbb{N}$,

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

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

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 $S_0 = 0$, $S_n = \xi_1 + \dots + \xi_n$, et $X_n = \max\{S_m : 0 \leq m \leq n\}$. Alors $(X_n)_n$ n'est **pas** une chaîne de Markov.

EXERCICE 10 - Soit $(\xi_i)_{i \in \mathbb{N}}$ une suite de v.a. iid à valeurs dans un espace F et $f : E \times F \rightarrow E$ 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 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 marqué dont les noeuds sont les éléments de E et où les arcs dirigés marqués sont les couples (x, y) tels que $Q(x, y) > 0$ (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 $x_0, \dots, x_n \in E$,

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

Démonstration. Immédiate par récurrence à partir de la définition 4.1.1. □

Cette 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 $f : E \rightarrow \mathbb{R}$ est mesurable et intégrable,

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

- iii) Pour $p \geq 1$,

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

iv) Les lois instantanées données par $\mu_n(y) := \mathbb{P}(X_n = y)$ vérifient

$$\mu_n(y) = \sum_{x \in E} \mu_{n-1}(x)Q(x, y),$$

soit en notation matricielle en identifiant μ_n avec un vecteur ligne

$$\mu_n = \mu_{n-1}Q = \cdots = \mu_0Q^n.$$

Démonstration. Exercice! □

Dans la suite, on notera souvent $\mathbb{P}_x(\cdot) = \mathbb{P}(\cdot | X_0 = x)$, et \mathbb{P}_μ la loi de $(X_n)_n$ lorsque $X_0 \sim \mu$. De même, on note \mathbb{E}_x l'espérance conditionnelle sachant que $X_0 = x$ et \mathbb{E}_μ l'espérance conditionnelle sachant que la loi de X_0 est μ . Ainsi,

$$\begin{aligned} \mathbb{E}_\mu[f(X_n)] &= \sum_{x \in E} \mu(x)\mathbb{E}_x[f(X_n)] \\ &= \sum_{x \in E} \mu(x)\mathbb{E}[f(X_n)|X_0 = x] \\ &= \sum_{x \in E} \mu(x)Q^n f(x) =: \mu Q^n f, \end{aligned}$$

où la dernière ligne est une application de la propriété *iii*) de la proposition précédente.

4.2 Propriété de Markov

Propriété de Markov simple

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

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

Théorème 4.2.1 (Propriété de Markov simple). Soit $X := (X_k)_{k \geq 0}$ une chaîne de Markov. Soit $f : E^{\mathbb{N}} \rightarrow \mathbb{R}_+$ une fonction mesurable positive, alors

$$\begin{aligned} \mathbb{E}[f(\theta_n X) | X_0, \dots, X_n] &= \mathbb{E}[f(\theta_n X) | X_n] \\ &= h(X_n) := \mathbb{E}_{X_n}[f(X)] \end{aligned}$$

où $h(x) = \mathbb{E}_x[f(X)]$.

Autrement dit, la chaîne de Markov (X_n, \dots) sachant l'évènement passé $X_n = x$ a la même loi que la chaîne (X_0, \dots) partant de x .

Propriété de Markov forte Ce résultat s'étend également au cas où l'instant n est aléatoire, sous certaines hypothèses. On pose

$$\mathcal{F}_n := \sigma(X_0, \dots, X_n) := \sigma(\{X_0 = x_0, \dots, X_n = x_n\}, x_0, \dots, x_n \in E)$$

la filtration canonique associée à $(X_n)_{n \geq 0}$.

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 $(X_n)_{n \geq 0}$ au temps T . On a alors la généralisation suivante de la propriété de Markov simple.

Théorème 4.2.2 (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,

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

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,

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

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 $x, y \in E$,

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

Remarque. *On vérifie aisément que toute mesure réversible est invariante.*

On cherche maintenant à obtenir un résultat nous assurant qu'une chaîne de Markov admet une mesure invariante. On introduit alors les notions de récurrence et de transience d'états associés à cette chaîne de Markov.

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 montrera en 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 . Ceci entraînera en particulier le résultat suivant

Proposition 4.3.3. *x est récurrent si et seulement si $\sum_{n \geq 0} P^n(x, x) = +\infty$. x est transitoire si et seulement si $\sum_{n \geq 0} P^n(x, x) < +\infty$.*

On a également le résultat fondamental suivant que l'on démontrera en TD

Proposition 4.3.4. *Si E est finie alors toute chaîne de Markov sur E admet au moins un état récurrent.*

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 4.3.1. *Soit X une chaîne de Markov et $x \in E$ 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 telle que $\mu_x(E) = \mathbb{E}_x[T_x]$ et $\mu_x(x) = 1$.

Démonstration. Soit $y \in E$. On a alors par définition

$$\mu_x Q(y) = \sum_{z \in E} \mu_x(z) Q(z, y).$$

Ainsi en utilisant la définition de m_x et le théorème de Fubini-Tonelli, on a

$$\begin{aligned} \mu_x Q(y) &= \sum_{z \in E} \left(\mathbb{E}_x \left[\sum_{n=0}^{T_x-1} \mathbb{1}_{\{X_n=z\}} \right] \right) Q(z, y) \\ &= \mathbb{E}_x \left[\sum_{n=0}^{+\infty} \mathbb{1}_{\{n < T_x\}} \sum_{z \in E} \mathbb{1}_{\{X_n=z\}} Q(z, y) \right] \end{aligned}$$

Pour tout $n \in \mathbb{N}$ on a alors en utilisant la propriété de Markov simple

$$\begin{aligned} \sum_{z \in E} \mathbb{1}_{\{X_n=z\}} Q(z, y) &= \mathbb{P}_{X_n}(X_1 = y) \\ &= \mathbb{E}_{X_n}[\mathbf{1}_{X_1=y}] \\ &= \mathbb{E}[\mathbf{1}_{X_{n+1}=y} | X_n]. \end{aligned}$$

Ainsi,

$$\begin{aligned} \mu_x Q(y) &= \mathbb{E}_x \left[\sum_{n=0}^{+\infty} \mathbb{1}_{\{n < T_x\}} \mathbf{1}_{X_{n+1}=y} \right] \\ &= \mathbb{E}_x \left[\sum_{n=0}^{T_x-1} \mathbf{1}_{X_{n+1}=y} \right] \\ &= \mathbb{E}_x \left[\sum_{n=1}^{T_x} \mathbf{1}_{X_n=y} \right] \\ &= \mathbb{E}_x \left[\sum_{n=0}^{T_x-1} \mathbf{1}_{X_n=y} \right] \\ &= \mu_x(y). \end{aligned}$$

en remarquant que $X_{T_x} = x$. De plus

$$\begin{aligned} \mu_x(E) &= \sum_{y \in E} \mu_x(y) \\ &= \mathbb{E}_x \left[\sum_{n=0}^{T_x-1} \sum_{y \in E} \mathbf{1}_{X_n=y} \right] \\ &= \mathbb{E}_x [T_x]. \end{aligned}$$

□

Bien entendu, bien que la récurrence entraîne l'existence d'une mesure invariante, elle est insuffisante pour avoir l'unicité. Par exemple, si on considère une chaîne de Markov sur $\{1, 2\}$ telle que $Q(1, 1) = Q(2, 2) = 1$ on a une infinité de mesure invariante bien que les états soient récurrents. Cet exemple nous motive à introduire les notions de communication entre les états et d'irréductibilité d'une chaîne de Markov pour étudier l'unicité sous hypothèse d'existence d'une mesure invariante.

Définition 4.3.5. On dit que y est accessible à partir de x s'il existe $n \geq 0$ tel que $Q^n(x, y) > 0$. On utilise alors la notation $x \rightarrow y$.

On dit que les états x et y communiquent si $x \rightarrow y$ et $y \rightarrow x$. On note alors $x \leftrightarrow y$.

Une matrice de transition Q est dite **irréductible** si, $x \leftrightarrow y$ pour tout $x, y \in E$, i.e. il existe $n \geq 1$ tel que $Q^n(x, y) > 0$. La chaîne de Markov associée est alors dite irréductible.

En d'autres termes, une chaîne de Markov est irréductible si tous ses états communiquent les uns avec les autres en un nombre fini d'étapes avec probabilité strictement positive. On vérifie facilement que \rightarrow est une relation réflexive et transitive et que \leftrightarrow est une relation d'équivalence. Une chaîne de Markov est donc irréductible lorsqu'il n'y a qu'une seule classe d'équivalence pour \leftrightarrow .

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

Proposition 4.3.6. 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 peut consister 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. Elle est en fait une conséquence directe de la conservation de la propriété de récurrence au sein d'une classe d'équivalence de \leftrightarrow prouvée en TD. Le cas de E fini est une conséquence directe de la proposition 4.3.4.

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

Théorème 4.3.2. 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).

Définition 4.3.7. Un état $x \in E$ est dit **récurrent positif** si $\mathbb{E}_x[T_x] < +\infty$. Si un état est récurrent mais que $\mathbb{E}_x[T_x] = +\infty$ il est **récurrent nul**.

Cette propriété est conservée au sein d'une classe de \leftrightarrow .

Lemme 4.3.8. Si x est récurrent positif et $x \leftrightarrow y$ alors y est récurrent positif.

On voit en particulier que si un état x est récurrent positif alors la mesure μ_x définie par le théorème 4.3.1 est de masse finie. Ainsi en divisant μ_x par sa masse totale $\mu_x(E) = \mathbb{E}_x[T_x]$, on obtient une probabilité invariante π unique dès lors que la mesure invariante sous jacente l'est, telle que $\pi(x) = \frac{1}{\mathbb{E}_x[T_x]}$. On dit qu'une chaîne de Markov irréductible est récurrente positive si un état est récurrent positif (donc tous ses états sont récurrents positifs). On définit de même la notion de matrice récurrente nulle. On en déduit alors la proposition suivante

Proposition 4.3.9. Soit $(X_n)_n$ une chaîne de Markov irréductible récurrente.

- i) $(X_n)_n$ est récurrente positive si et seulement si il existe une probabilité invariante π . Dans ce cas, pour tout $x \in E$,

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

ii) $(X_n)_n$ est récurrente nulle si et seulement si toute mesure invariante a une masse totale infinie, et, pour tout $x \in E$,

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

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 4.3.10. *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

Le dernier résultat fondamental de ce chapitre est un résultat d'ergodicité dans le cas récurrent positif.

Théorème 4.4.1. *Soit $(X_n)_n$ une chaîne de Markov irréductible récurrente positive, de probabilité invariante π . Alors,*

$$\frac{1}{n} \sum_{k=0}^n f(X_k) \xrightarrow[n \rightarrow \infty]{} \langle \pi, f \rangle, \quad \mathbb{P}_x - p.s.$$

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,

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

ii) dans le cas récurrent nul,

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

En d'autres termes, si la chaîne de Markov est récurrente positive, alors elle passe une proportion strictement positive de son temps en chaque états (donnée par la loi invariante). Si elle est récurrente nulle, cette proportion tend vers 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 $a > 0$,

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

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

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

Jensen Soit $\phi : \mathbb{R} \rightarrow \mathbb{R}$ 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 $X, Y \in L^p$,

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