Université Paris Dauphine

Département MIDO

# Cours de Mathématiques

# Intégrale de Lebesgue et Probabilités

H. DOSS

# Table des matières

# Chapitre 1

# Espaces de probabilité et Intégration

## 1.1 Présentation

La théorie des probabilités propose un modèle mathématique qui rend compte de la notion intuitive d'expérience aléatoire (i.e. dont le résultat est soumis au « hasard »).

La première démarche consiste à introduire l'ensemble $\Omega$ dont les élements constituent tous les résultats possibles de l'expérience puis à distinguer une classe $\mathcal{A}$ de sous-ensembles de $\Omega$ qu'on appelle « événements », vérifiant certaines propriétés naturelles et enfin à affecter un poids $\mathcal{P}(A) \in [0,1]$ à tout événement $A \in \mathcal{A}$ qui sera la probabilité de $A$. Plus précisément, la classe $\mathcal{A}$ doit être une *tribu* sur $\Omega$ et la correspondance $A \mapsto \mathcal{P}(A)$ une *mesure positive de masse totale égale à 1.*

### 1.1.1 Définition 1

1. Soit $\Omega$ un ensemble. On appelle *tribu* sur $\Omega$ un sous ensemble $\mathcal{A}$ de l'ensemble $\mathcal{P}(\Omega)$ des parties de $\Omega$ tel que :
   - $\phi$ et $\Omega$ appartiennent à $\mathcal{A}$.
   - si $A \in \mathcal{A}$ alors $A^c \in \mathcal{A}$.
   - si $(A_n)$ est une suite d'éléments de $\mathcal{A}$ alors

   $$\bigcup_n A_n \in \mathcal{A} \quad \text{et} \quad \bigcap_n A_n \in \mathcal{A}$$

   Le couple $(\Omega, \mathcal{A})$ s'appelle un *espace mesurable*
2. Un espace mesurable $(\Omega, \mathcal{A})$ étant donné, une *mesure positive de masse totale égale à 1* ou *probabilité* sur $\mathcal{A}$ est une application $P$ de $\mathcal{A}$ dans $[0,1]$ vérifiant :
   - si $(A_n)$ est une suite d'éléments de $\mathcal{A}$ deux à deux disjoints, alors

   $$P\left(\bigcup_n A_n\right) = \sum_n P(A_n)$$

– $P(\Omega) = 1$

Le triplet $(\Omega, \mathcal{A}, P)$ s'appelle un *espace de probabilité*. L'espace entier $\Omega$ représente l'événement certain, $\phi$ l'événement impossible ($P(\phi) = 0$ car $\phi \cup \phi = \phi$ et $\phi \cap \phi = \phi$ donc $P(\phi) = P(\phi \cup \phi) = P(\phi) + P(\phi)$ ), le complémentaire $A^c$ représente l'événement contraire de $A$, l'intersection $A \cap B$ des événements $A$ et $B$ représente l'événement « $A$ et $B$ ont lieu », la réunion $A \cup B$ représente l'événement « $A$ ou $B$ a lieu » (attention : ou n'est pas exclusif !), les intersections ou réunions dénombrables d'événements sont introduits pour l'étude des phénomènes asymptotiques, enfin l'inclusion correspond à l'implication.

### 1.1.2 Exemples associés à la définition 1

1. Un joueur effectue quatre parties de pile ou face :
$$\Omega = \big\{(x_1, x_2, x_3, x_4),\ x_i \in \{0, 1\}\big\}$$
2. Un joueur lance un dé trois fois de suite :
$$\Omega = \big\{(x_1, x_2, x_3),\ x_i \in \{1, 2, 3, 4, 5, 6\}\big\}$$
3. Observation de la durée de vie d'un individu :
$$\Omega = \mathbb{R}_+$$
4. Observation du nombre d'appels passant par un central téléphonique tous les jours d'une semaine :
$$\Omega = \mathbb{N}^7$$
5. Observation, pendant un intervalle de temps $[t_1, t_2]$, du mouvement de diffusion d'une particule dans l'espace :
$$\Omega = \mathcal{C}\left([t_1, t_2], \mathbb{R}^3\right)$$
6. Un jeu de pile ou face de durée finie ou infinie :
$$\Omega = \{0, 1\}^n \qquad \text{avec} \qquad (n \in \mathbb{N}) \qquad \text{ou} \qquad \Omega = \{0, 1\}^{\mathbb{N}}$$

Dans les cas où l'ensemble $\Omega$ est fini ou dénombrable, la tribu que l'on considère est, en général, la tribu $\mathcal{P}(\Omega)$ des parties de $\Omega$ et la donnée d'une probabilité sur l'espace *discret* $(\Omega, \mathcal{P}(\Omega))$ équivaut à la donnée d'une famille $(p(\omega))_{\omega \in \Omega}$ de nombres positifs vérifiant :
$$\sum_{\omega \in \Omega} p(\omega) = 1$$
La probabilité d'un événement $A \in \mathcal{P}(\Omega)$ est alors définie par la formule :
$$P(A) = \sum_{\omega \in A} p(\omega)$$

Lorsque $\Omega$ est fini, des considérations de symétrie conduisent souvent à supposer que $p(\omega)$ ne dépend pas de $\omega \in \Omega$. On a alors :
$$p(\omega) = \frac{1}{\mathrm{Card}(\Omega)}$$

et

$$P(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)} = \frac{\text{nombres de cas favorables à } A}{\text{nombre de cas possibles}} \qquad \text{(loi uniforme)}$$

Lorsque l'espace $\Omega$ n'est pas dénombrable, par exemple $\Omega = \mathbb{R}$ ou $\Omega = \mathbb{R}^d$ ou même un espace métrique, on utilise souvent la notion suivante :

### 1.1.3 Definition 2

Soit $\mathcal{C} \subseteq P(\Omega)$, une famille quelconque de parties de $\Omega$ ; on appelle *tribu engendrée par* $\mathcal{C}$, la plus petite tribu sur $\Omega$ (au sens de l'inclusion) contenant $\mathcal{C}$ ; on la note $\sigma(\mathcal{C})$. L'intersection d'une famille quelconque de tribus étant une tribu, et $\mathcal{P}(\Omega)$ étant une tribu sur $\Omega$ contenant $\mathcal{C}$, on voit que :

$$\sigma(\mathcal{C}) = \bigcap_{\substack{\mathcal{T}\text{tribu} \\ \mathcal{C} \subseteq \mathcal{T}}} \mathcal{T}$$

### 1.1.4 Exemples associés à la définition 2

1. Si $\mathcal{C} = \{A\}$ où $A \subseteq \Omega$, $A \neq \phi$ alors $\sigma(\mathcal{C}) = \{A, A^c, \Omega, \phi\}$
   Si $\mathcal{C} = \{A_1, A_2, \cdots, A_n\}$, $A_i \subseteq \Omega$, $A_i \neq \phi$, déterminer $\sigma(\mathcal{C})$ (considérer d'abord le cas $n = 2$).
2. Si $\Omega$ est muni d'une structure d'espace métrique, on peut considérer la famille $\mathcal{C}$ de tous les ouverts de $\Omega$ ; la tribu engendrée par $\mathcal{C}$ est alors appelée *tribu borélienne de* $\Omega$ et notée souvent $\mathcal{B}_\Omega$. Remarquer que, si $\mathcal{C}'$ est la famille de tous les fermés de $\Omega$, on a aussi $\mathcal{B}_\Omega = \sigma(\mathcal{C}') = \sigma(\mathcal{C})$.
3. Si $\Omega = \mathbb{R}^d$ $(d \in \mathbb{N}^*)$ la tribu $\mathcal{B}_{\mathbb{R}^d}$ est engendrée par la famille $\mathcal{C}$ de tous les pavés de la forme

   $$\prod_{i=1}^{d} ]a_i, b_i[ \qquad \text{où} \qquad a_i < b_i \quad , \qquad a_i \in \mathbb{Q} \quad , \qquad b_i \in \mathbb{Q}$$

   Soit $(\Omega, \mathcal{a})$ un espace mesurable quelconque ; le problème de la construction d'une mesure de probabilité $P$ sur $\mathcal{a}$ peut être non trivial. On a cependant les exemples suivants.
4. Soient $(a_k)$ une suite fixée de points de $\Omega$ et $(\alpha_k)$ une suite fixée de nombres positifs tels que $\sum_k \alpha_k = 1$. Pour tout $A \in \mathcal{a}$, posons :

   $$P(A) = \sum_{k \geq 0} \alpha_k . \mathbf{1}_A(a_k) \qquad \textbf{(*)}$$

   où

   $$\mathbf{1}_A(a_k) = \begin{cases} 0 & \text{si} \quad a_k \notin A \\ 1 & \text{si} \quad a_k \in A \end{cases}$$

   On vérifie facilement qu'on définit ainsi une mesure de probabilité $P$ sur $\mathcal{a}$. Lorsque $a \in \Omega$ est fixé, la mesure de probabilité définie par :

   $$\forall A \in \mathcal{a} \qquad P(A) = \mathbf{1}_A(a)$$

est appelée *masse de Dirac au point a* et notée $\delta_{\{a\}}$. La probabilité $P$ définie par la formule (*) est donc donnée par

$$P = \sum_{k \geq 0} \alpha_k . \delta_{\{a_k\}}$$

5. On appelle *fonction de répartition sur* $\mathbb{R}$, toute application $F$ de $\mathbb{R}$ dans $[0,1]$, croissante, continue à droite, telle que

$$\lim_{t \to +\infty} F(t) = 1 \qquad \lim_{t \to -\infty} F(t) = 0$$

Il existe alors (*Théorème admis*) une unique probabilité $P$ sur l'espace mesurable $(\Omega = \mathbb{R}, \mathcal{B}_{\mathbb{R}})$ telle que

$$\forall s \leq t \qquad : \qquad P(\,]s,t]) = F(t) - F(s)$$

On a alors, pour tout $t \in \mathbb{R}$ :

$$P(\{t\}) = \lim_{\substack{s \to t \\ s < t}} P(\,]s,t]) = F(t) - \lim_{\substack{s \to t \\ s < t}} F(s) = F(t) - F(t_0) = \Delta F(t)$$

Si la fonction $F$ est continue en $t$, le « saut » de $F$ en $t$ : $\Delta F(t)$ est nul et $P(\{t\}) = 0$. On dira que $P$ est « diffuse » lorsque $P(\{t\}) = 0, \forall\, t$. Si pour tout $n \in \mathbb{N}$, $\Delta F(n) = \alpha_n$ où $\alpha_n \geq 0$, vérifie $\sum_n \alpha_n = 1$, on voit que $P$ est une probabilité discrète portée par $\mathbb{N}$ telle que :

$$\forall\, n \in \mathbb{N} \qquad P(\{n\}) = \alpha_n$$

Si, par contre, il existe une application $f \;:\; \mathbb{R} \to \mathbb{R}_+$ continue par morceaux (positive) telle que

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$$

et vérifiant

$$\forall\, t \in \mathbb{R} \qquad F(t) = \int_{-\infty}^{t} f(x)\,\mathrm{d}x$$

on voit que $F$ est une fonction de répartition continue et que la probabilité $P$ associée à $F$ vérifie :

$$\forall\, s \leq t \qquad P(\,]s,t]) = \int_{s}^{t} f(x)\,\mathrm{d}x$$

On dira que $P$ est la probabilité de densité $f$ et on notera $P(\mathrm{d}x) = f(x)\,\mathrm{d}x$. On peut, de la même façon, définir des probabilités de densité $f$ sur $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$.

### 1.1.5 Proposition 1

Soient $(\Omega, \mathcal{A}, P)$ un espace de probabilité et $A$, $B$, $A_n$ ($n \in \mathbb{N}$) des événements. Alors :

1)

$$A \supseteq B \Rightarrow P(A) \geq P(B)$$

2)

$$P\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n\in\mathbb{N}} P(A_n)$$

3) Si pour tout $n$, $A_n \subseteq A_{n+1}$, on a :

$$P\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} P(A_n) = \sup_n P(A_n)$$

Si pour tout $n$, $A_{n+1} \subseteq A_n$, on a :

$$P\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} P(A_n) = \inf_n P(A_n)$$

Démonstration :
1) On suppose $A \supseteq B$, alors : $A = B \cup (A\backslash B)$ avec $B \cap (A\backslash B) = \emptyset$ donc $P(A) = P(B) + P(A\backslash B) \geq P(B)$
3) Si pour tout $n$, $A_n \subseteq A_{n+1}$, on voit que les événements $A_0,\ A_1\backslash A_0,\ A_2\backslash A_1,\ \ldots,\ A_{n+1}\backslash A_n,\ \ldots$ sont disjoints de réunion $A = \underset{n}{\cup} A_n$ et donc

$$\begin{aligned}
P(A) &= P\left(A_0 + \sum_{n\geq 0}(A_{n+1}\backslash A_n)\right) \\
&= P\left(A_0\right) + \sum_{n\geq 0} P(A_{n+1}\backslash A_n) \\
&= P(A_0) + \lim_{n\to\infty} \left\{ \sum_{m=0}^{n-1} \left(P(A_{m+1}) - P(A_m)\right) \right\} \\
&= \lim_{n\to\infty} P(A_n)
\end{aligned}$$

(pour une union d'événements deux à deux disjoints, on remplacera, par commodité, le symbole $\cup$ par le symbole $+$).
Si pour tout $n$, $A_{n+1} \subseteq A_n$, on a alors

$$\begin{aligned}
P(\underset{n}{\cap} A_n) &= 1 - P\left((\underset{n}{\cap} A_n)^C\right) \\
&= 1 - P\left(\underset{n}{\cup} A_n^C\right) \\
&= 1 - \lim_{n\to\infty} P(A_n^C) \\
&= \lim_{n\to\infty} P(A_n)
\end{aligned}$$

2) Posons, pour tout $n$, $A_n^{'} = \overset{n}{\underset{m=0}{\cup}} A_m$ ; alors $A_n^{'} \subseteq A_{n+1}^{'}$ et $\underset{n}{\cup} A_n^{'} = \underset{n}{\cup} A_n$ donc :

$$P(\underset{n}{\cup} A_n) = \lim_{n\to\infty} P(A_n^{'}) \leq \lim_{n\to\infty} \sum_{m=0}^{n} P(A_m) = \sum_{n\geq 0} P(A_n)$$

car $A$ et $B$ étant deux événements quelconques,

$$P(A\cup B) = P\left(A + (B\backslash A)\right) = P(A) + P(B\backslash A) \leq P(A) + P(B)$$

et donc, par récurrence sur $n$

$$P(A_n^{'}) = P(\overset{n}{\underset{m=0}{\cup}} A_m) \leq \sum_{m=0}^{n} P(A_m)$$

## 1.2 Probabilités conditionnelles, indépendance

Supposons, qu'en étudiant une expérience aléatoire représentée par un espace $(\Omega, \mathcal{a}, P)$, on sache déjà qu'un événement $A \in \mathcal{a}$ (de probabilité $P(A) > 0$) s'est produit. Il est alors intuitif que les probabilités des événements $B \in \mathcal{a}$ doivent être modifiées.

### 1.2.1 Exemple

Soit $\Omega = \{\omega_0, \omega_1, \ldots, \omega_n\}$ un espace fini. On suppose que $P$ représente le tirage d'un point de $\Omega$ selon la loi uniforme. On a donc

$$P(\{\omega_k\}) = \frac{1}{\text{Card}(\Omega)} \qquad \text{et si} \qquad B \subseteq \Omega, \quad P(B) = \frac{\text{Card}(B)}{\text{Card}(\Omega)}$$

Soit $A \subseteq \Omega$.
Quelle est la probabilité qu'un point tiré soit dans $B$, sachant qu'il est dans $A$ ?
C'est :

$$\frac{\text{Card}(A \cap B)}{\text{Card}(A)} = \frac{P(A \cap B)}{P(A)}$$

### 1.2.2 Définition 3

Soient $(\Omega, \mathcal{a}, P)$ un espace de probabilité et $A \in \mathcal{a}$ tel que $P(A) > 0$. On appelle *probabilité conditionnelle de $B \in \mathcal{a}$ sachant $A$*, le nombre

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

### 1.2.3 Proposition 2 (Formule de Bayes)

Soit $(A_n)$ une *suite* finie ou infinie d'événements *deux à deux disjoints* tels que $\Omega = \underset{n}{\cup} A_n$ (partition de $\Omega$) et $P(A_n) > 0$, pour tout $n$, alors :

$$\forall B \in \mathcal{a}, \qquad \text{on a} \qquad P(B) = \sum_n P(B/A_n).P(A_n)$$

<u>Démonstration</u>
On a :

$$B = B \cap \Omega = B \cap (\underset{n}{\cup} A_n) = \sum_n (B \cap A_n)$$

donc

$$P(B) = P\left(\sum_n (B \cap A_n)\right) = \sum_n P(B \cap A_n) = \sum_n P(B/A_n).P(A_n)$$

On dira que deux événements $A$ et $B$ sont indépendants si $P(B/A) = P(B)$ et donc si $P(A \cap B) = P(A).P(B)$

### 1.2.4 Définition 4

Une suite finie $A_1$, $A_2$, $\cdots$, $A_n$ d'événements est dite *indépendante* si

$$P(A_{i_1}A_{i_2}\ldots A_{i_k}) = P(A_{i_1}).P(A_{i_2}).\cdots .P(A_{i_k})$$

pour *toute suite* $i_1 < i_2 < \ldots < i_k$ d'entiers $\{1, 2, \ldots, n\}$ deux à deux distincts. (N.B. on notera, par commodité, $A.B$ pour désigner l'intersection $A \cap B$ de deux évènements.)
Si $A \in \mathcal{a}$, soit $\sigma(A) = \{\emptyset, \Omega, A, A^C\}$ la tribu engendrée par $A$.

### 1.2.5 Proposition 3

Pour que la suite $A_1$, $A_2$, $\cdots$, $A_n$ d'événements dans l'espace de probabilité $(\Omega, \mathcal{a}, P)$ soit indépendante il faut et il suffit que

$$P(B_1B_2\ldots B_n) = P(B_1).P(B_2).\cdots .P(B_n)$$

quels que soient les $B_m$ appartenant à $\sigma(A_m), m = 1, 2, \ldots, n$
Démonstration
Pour tout $i \in \{1, 2, \ldots, n\}$, soit $A_i^{'} = A_i$ ou $A_i^C$.
On a, en fait, l'équivalence :
$(A_1, A_2, \cdots, A_n)$ indépendante $\iff$ $(A_1^{'}, A_2^{'}, \cdots, A_n^{'})$ indépendante.
Il suffit, par récurrence de vérifier que :
$(A_1, A_2, \cdots, A_n)$ indépendante $\Rightarrow$ $(A_1^{'}, A_2, \cdots, A_n)$ indépendante ce qui est facile et laissé en exercice.

### 1.2.6 Définition 4 bis

Soit $(A_i)_{i\in I}$ une famille d'événements dans l'espace $(\Omega, \mathcal{a}, P)$ où $I$ est un ensemble quelconque d'indices. On dit que *la famille* $(A_i)_{i\in I}$ *est indépendante* (ou que les $A_i$, $(i \in I)$ sont indépendants) si *pour tout sous-ensemble* $J$ *fini*, $J \subseteq I$, on a

$$P(\underset{i\in J}{\cap} A_i) = \prod_{i\in J} P(A_i)$$

Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements dans l'espace $(\Omega, \mathcal{a}, P)$. On définit :

$$\overline{\lim}\ A_n = \{\omega \in \Omega \text{ tels que, pour une infinité d'entiers } n : \omega \in A_n\}$$

On vérifie aisément que :

$$\overline{\lim}\ A_n = \{\omega \in \Omega\ :\ \sum_{n\geq 0} \mathbf{1}_{A_n}(\omega) = +\infty\} = \underset{n}{\cap}\ (\underset{m\geq n}{\cup} A_m) \in \mathcal{a}$$

### 1.2.7 Proposition 4 (Lemme de Borel-Cantelli)

Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements. On a

1. $$\left[\sum_{n\geq 0} P(A_n) < \infty\right] \Rightarrow P(\overline{\lim}\ A_n) = 0$$

2. Si la suite $(A_n)_{n\in\mathbb{N}}$ est indépendante et si $\sum_{n\geq 0} P(A_n) = +\infty$ alors $P(\overline{\lim}\ A_n) = 1$

Démonstration :

1. On a, lorsque $\sum_n P(A_n) < \infty$ :

$$\begin{aligned} P(\overline{\lim}\ A_n) &= P(\bigcap_n (\bigcup_{m\geq n} A_n)) \\ &= \lim_{n\to\infty} P(\bigcup_{m\geq n} A_m) \\ &\leq \lim_{n\to\infty} \Big\{ \sum_{m\geq n} P(A_n) \Big\} = 0 \end{aligned}$$

2. On suppose que $\sum_n P(A_n) = +\infty$ et que la suite $(A_n)$ est indépendante, alors :

$$\begin{aligned} P\big\{(\overline{\lim}\ A_n)^C\big\} &= P\big\{\bigcup_n (\bigcap_{m\geq n} A_m^C)\big\} \\ &= \lim_{n\to\infty} P\big\{\bigcap_{m\geq n} A_m^C\big\} = 0 \end{aligned}$$

car

$$\begin{aligned} P\big\{\bigcap_{m\geq n} A_m^C\big\} &= \lim_{k\to\infty} P\big\{\bigcap_{m=n}^{k} A_m^C\big\} \\ &= \lim_{k\to\infty} \prod_{m=n}^{k} (1 - P(A_m)) \\ &\leq \lim_{k\to\infty} e^{-\sum_{m=n}^{k} P(A_m)} = 0 \end{aligned}$$

( on a utilisé le fait que : $\forall x \geq 0 \quad (1-x) \leq e^{-x}$)

## 1.3 Variables aléatoires

### 1.3.1 Définition 5

Soient $(\Omega, \mathcal{A}, P)$ un espace de probabilité et $(F, \mathcal{B})$ un espace mesurable quelconque. On dit qu'une application $X$ de $(\Omega, \mathcal{A})$ dans $(F, \mathcal{B})$ est une *variable aléatoire* (en abrégé : v.a.) si pour tout $B \in \mathcal{B}$ alors $X^{-1}(B) \in \mathcal{A}$. En Analyse, on dit aussi que $X$, vérifiant la propriété précédente, est une application $\mathcal{A} - \mathcal{B}$ *mesurable* (ou, tout simplement, mesurable s'il n'y a pas d'ambiguité). Lorsque l'espace $F$ est *fini ou dénombrable*, la tribu $\mathcal{B}$ est, en général, égale à $\mathcal{P}(F)$ et on voit qu'une application $X$ de $(\Omega, \mathcal{A})$ dans $F$ est une variable aléatoire si, pour tout $x \in F$ on a $X^{-1}(\{x\}) = (X = x) \in \mathcal{A}$. On dit, dans ce cas, que $X$ est une variable aléatoire *discrète*. Lorsque $F = \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$, muni de sa tribu borélienne $\mathcal{B}_{\overline{\mathbb{R}}}$ (engendrée par les ouverts de $\mathbb{R}$ ainsi que par les points $\{+\infty\}$, $\{-\infty\}$), on dit que $X$ est une *variable aléatoire réelle* (v.a.r.).

### 1.3.2 Proposition 5

1. Pour qu'une application $X$ de $(\Omega, \mathcal{A})$ dans $(F, \mathcal{B})$ soit une variable aléatoire, il suffit que, $\mathcal{C}$ étant une famille de parties de $F$ qui engendre $\mathcal{B}$ ($\sigma(\mathcal{C}) = \mathcal{B}$), on ait : pour tout $C \in \mathcal{C}$ : $X^{-1}(C) \in \mathcal{A}$ (**).

2. Lorsque $\Omega$ et $F$ sont des espaces métriques, munis respectivement de leurs tribus boréliennes, toute application continue $X$, de $\Omega$ dans $F$, est mesurable.

Démonstration :

1. On suppose que $\sigma(\mathcal{C}) = \mathcal{B}$ et que la propriété (**) est satisfaite. Soit

$$\mathcal{B}^{'} = \{B \subseteq F \text{ tel que } X^{-1}(B) \in \mathcal{A}\}$$

$\mathcal{A}$ étant une tribu sur $\Omega$, on vérifie facilement que $\mathcal{B}^{'}$ est une tribu sur $F$. De plus, la propriété (**) signifie que $\mathcal{C} \subseteq \mathcal{B}^{'}$. Donc $\sigma(\mathcal{C}) = \mathcal{B} \subseteq \mathcal{B}^{'}$ et $X$ est bien une variable aléatoire.

2. On sait qu'une application $X$ de $\Omega$ dans $F$ est continue, si pour tout ouvert $C$ de $F$, $X^{-1}(C)$ est un ouvert de $\Omega$ . L'assertion 2) est donc une conséquence immédiate de 1) et de la définition d'une tribu borélienne.

### 1.3.3 Proposition 6

Soient $(\Omega, \mathcal{A})$, $(F, \mathcal{B})$ et $(T, \mathcal{S})$ trois espaces mesurables, $X$ une application mesurable de $(\Omega, \mathcal{A})$ dans $(F, \mathcal{B})$, f une application mesurable de $(F, \mathcal{B})$ dans $(T, \mathcal{S})$, alors $Y = f \circ X$ est une application mesurable de $(\Omega, \mathcal{A})$ dans $(T, \mathcal{S})$
Démonstration : elle est immédiate.
Remarques :

1. Soit $X$ une application de $(\Omega, \mathcal{A})$ dans $\overline{\mathbb{R}}$. Pour que $X$ soit une variable aléatoire réelle il suffit que, pour tout $t \in \mathbb{Q}$, l'ensemble $(X \leq t)$ [resp. $(X \geq t)$] appartienne à $\mathcal{A}$. En effet si $\mathcal{C} = \{[-\infty, t], t \in \mathbb{Q}\}$ et $\mathcal{C}^{'} = \{[t, +\infty], t \in \mathbb{Q}\}$, on vérifie facilement que $\sigma(\mathcal{C}) = \sigma(\mathcal{C}^{'}) = \mathcal{B}_{\overline{\mathbb{R}}}$ (tribu borélienne de $\overline{\mathbb{R}}$). On peut donc appliquer le 1) de la proposition 5.

2. Une conséquence des propositions 5 et 6 est que, si $X$ et $Y$ sont deux variables aléatoires réelles finies, alors, pour tous $\alpha$ et $\beta$ appartenant à $\mathbb{R}$, $\alpha X + \beta Y$, $X.Y$ et $\frac{1}{X}$ (si $X$ ne s'annule pas) sont aussi des variables aléatoires réelles.

Soit $(a_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\overline{\mathbb{R}}$. On définit les nombres $\overline{\lim}_{n\to\infty} a_n$ et $\underline{\lim}_{n\to\infty} a_n$ par :

$$\overline{\lim_{n\to\infty}} a_n = \lim_{n\to\infty} \{ \sup_{m\geq n} a_m \} = \inf_n \{ \sup_{m\geq n} a_m \}$$

$$\underline{\lim_{n\to\infty}} a_n = \lim_{n\to\infty} \{ \inf_{m\geq n} a_m \} = \sup_n \{ \inf_{m\geq n} a_m \}$$

Ces quantités sont donc bien définies (dans $\overline{\mathbb{R}}$), de plus :

$$\overline{\lim_{n\to\infty}} (-a_n) = - \underline{\lim_{n\to\infty}} (a_n) \quad \text{et} \quad \underline{\lim_{n\to\infty}} a_n \leq \overline{\lim_{n\to\infty}} a_n$$

On vérifie, d'autre part, que $l = \lim_{n\to\infty} a_n$ existe (dans $\overline{\mathbb{R}}$) si et seulement si $\overline{\lim_n} a_n = \underline{\lim_n} a_n = l$. Ces propriétés sont laissées en exercice.

### 1.3.4 Proposition 7

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles, alors $\sup_n X_n$, $\inf_n X_n$, $\overline{\lim\limits_{n\to\infty}} X_n$ et $\underline{\lim\limits_{n\to\infty}} X_n$ sont aussi des variables aléatoires réelles. En particulier, soit $X_\infty$ l'application de $\Omega$ dans $\overline{\mathbb{R}}$ définie par :

$$X_\infty(\omega) = \begin{cases} \lim\limits_{n\to\infty} X_n(\omega) & \text{si } \lim\limits_{n\to\infty} X_n(\omega) \text{ existe} \\ 0 & \text{sinon} \end{cases}$$

alors $X_\infty$ est une variable aléatoire réelle.
Démonstration :
Pour tout $t \in \mathbb{R}$, on a

$$(\sup_n X_n \leq t) = \bigcap_n (X_n \leq t) \in \mathcal{A}$$

$$(\inf_n X_n \geq t) = \bigcap_n (X_n \geq t) \in \mathcal{A}$$

donc $\sup_n X_n$ et $\inf_n X_n$ sont des variables aléatoires réelles. On en déduit que $\overline{\lim\limits_{n\to\infty}} X_n$ et $\underline{\lim\limits_{n\to\infty}} X_n$ sont aussi des variables aléatoires réelles puisque

$$\overline{\lim_{n\to\infty}} X_n = \inf_n \{\sup_{m\geq n} X_m\} \quad \text{et} \quad \underline{\lim_{n\to\infty}} X_n = \sup_n \{\inf_{m\geq n} X_m\}$$

Donc :

$$\{\omega \in \Omega \text{ tel que } \lim_{n\to\infty} X_n(\omega) \text{ existe}\} = \{\omega \in \Omega \ : \ \overline{\lim_{n\to\infty}} X_n(\omega) = \underline{\lim_{n\to\infty}} X_n(\omega)$$

est un événement $A$. On voit alors que :

$$X_\infty = (\overline{\lim_{n\to\infty}} X_n).\mathbf{1}_A$$

(avec la convention $0 \cdot \infty = 0$) est bien une variable aléatoire réelle. La proposition précédente montre que toute limite simple (si elle existe) d'une suite de variables aléatoires réelles est une variable aléatoire réelle.

### 1.3.5 Définition 6

Soit $(\Omega, \mathcal{A})$ un espace mesurable. On appelle variable aléatoire *étagée* toute variable aléatoire $X : (\Omega, \mathcal{A}) \to \mathbb{R}$ qui ne prend qu'un nombre fini de valeurs. On vérifie alors que $X$ est étagée si et seulement si elle est de la forme

$$X = \sum_{\text{finie}} \alpha_i.\mathbf{1}_{A_i} \quad \text{où} \quad \alpha_i \in \mathbb{R} \quad \text{et} \quad A_i \in \mathcal{A}$$

Remarquer que la décomposition précédente de $X$ n'est pas unique et que l'ensemble $\mathcal{E}$ des variables aléatoires étagées est un espace vectoriel sur $\mathbb{R}$.

### 1.3.6 Proposition 8

Soit $X$ une application de $(\Omega, \mathcal{A})$ dans $\overline{\mathbb{R}}_+ = [0, +\infty]$. Alors $X$ est une variable aléatoire si et seulement si il existe une suite croissante $(X_n)_{n\in\mathbb{N}}$ de

variables aléatoires étagées positives qui converge simplement vers $X$.
Démonstration :
S'il existe une suite $(X_n)$ d'éléments de $\mathcal{E}$ qui converge simplement vers $X$ on sait, d'après la proposition 7, que $X$ est une variable aléatoire réelle. Réciproquement supposons que $X : (\Omega, \mathcal{A}) \to \overline{\mathbb{R}}_+ = [0, +\infty]$ est une variable aléatoire réelle. Posons

$$X_n = \sum_{k=0}^{n^2-1} \frac{k}{n}.\mathbf{1}_{(X \in [\frac{k}{n}, \frac{k+1}{n}[)} + n.\mathbf{1}_{(X=+\infty)}$$

On vérifie alors facilement que, pour tout $n$, $X_n \in \mathcal{E}$, $X_n \leq X_{n+1}$ et que pour tout $\omega \in \Omega$ : $\lim_{n\to\infty} X_n(\omega) = X(\omega)$

## 1.4 Espérance des variables aléatoires réelles et intégration

Soit $X$ une variable aléatoire réelle *discrète* telle que

$$\sum_{x \in I_m(X)} |x|.P(X = x) < \infty \quad \text{où} \quad I_m(X) = X(\Omega) \subseteq \mathbb{R}$$

On sait alors (cf. le cours de DEUG) que l'espérance de $X$ est définie par :

$$E(X) = \sum_{x \in I_m(X)} x.P(X = x)$$

On se propose ici d'étendre la notion d'espérance à une classe beaucoup plus large de variables aléatoires réelles. L'espérance de $X$ (si elle existe) apparaîtra alors comme « l'intégrale de $X$ par rapport à la mesure de probabilité $P$ » :

$$E(X) = \int_\Omega X \,\mathrm{d}P$$

Il s'agit donc de donner un sens mathématique précis à cette écriture.

### 1.4.1 Définition 7

Soit $(\Omega, \mathcal{A})$ un espace mesurable. Une *mesure positive $\sigma$-finie* sur $\mathcal{A}$ est une application $\mu$ de $\mathcal{A}$ dans $\overline{\mathbb{R}}_+$, vérifiant :

1. si $(A_n)$ est une suite d'éléments de $\mathcal{A}$ deux à deux disjoints, alors

$$\mu\left(\sum_n A_n\right) = \sum_n \mu(A_n)$$

2. il existe une suite $\Omega_n$ ($n \in \mathbb{N}$) d'éléments de $\mathcal{A}$ telle que $\Omega = \underset{n}{\cup}\, \Omega_n$ et, pour tout $n$, $\mu(\Omega_n) < \infty$.

La mesure $\mu$ est dite *bornée* si sa masse totale $\mu(\Omega)$ est finie ; $\mu$ est une probabilité si $\mu(\Omega) = 1$ (cf. la définition 1). Le triplet $(\Omega, \mathcal{A}, \mu)$ s'appelle un *espace mesuré*. Remarquer que la définition précédente entraine que $\mu(\emptyset) = 0$, puisque, si $\Omega_n \in \mathcal{A}$ vérifie $\mu(\Omega_n) < \infty$, on a $\mu(\Omega_n) = \mu(\Omega_n \cup \emptyset) = \mu(\Omega_n) + \mu(\emptyset)$.

### 1.4.2 Proposition 9

Soient $\mu$ une mesure positive $\sigma$-finie sur $\mathcal{A}$ et $A$, $B$, $A_n$ $(n \in \mathbb{N})$ des événements. Alors :

1. $A \supseteq B \Rightarrow \mu(A) \geq \mu(B)$
2. $\mu(\underset{n}{\cup} A_n) \leq \sum_n \mu(A_n)$
3. si, pour tout $n$, $A_n \subseteq A_{n+1}$, on a : $\mu(\underset{n}{\cup} A_n) = \lim\limits_{n\to\infty} \mu(A_n) = \sup_n \mu(A_n)$
4. si, pour tout $n$, $A_{n+1} \subseteq A_n$ et *s'il existe* $n_0$ *tel que* $\mu(A_{n_0}) < \infty$, alors $\mu(\underset{n}{\cap} A_n) = \lim\limits_{n\to\infty} \mu(A_n) = \inf_n \mu(A_n)$

Démonstration : Cf. la preuve de la proposition 1.
Exemple important (mesure de Lebesgue)
En dehors des exemples de probabilités donnés précédemment on a le résultat suivant : Considérons $\mathbb{R}^d$ $(d \in \mathbb{N}^*)$ muni de sa tribu borélienne $\mathcal{B}$ ; on sait que $\mathcal{B}$ est engendrée par la famille $\mathcal{C}$ de tous les pavés $C$ de la forme

$$C = \prod_{i=1}^{d} ]a_i, b_i] \qquad \text{où} \qquad a_i < b_i, \qquad a_i, b_i \in \mathbb{R}$$

Il existe alors (nous admettrons ce théorème) une unique mesure positive $\sigma$-finie $\lambda$ sur $\mathcal{B}$ telle que

$$\lambda\left(\prod_{i=1}^{d} ]a_i, b_i]\right) = \prod_{i=1}^{d} (b_i - a_i) \qquad \text{pour tous} \qquad a_i < b_i, \qquad a_i, b_i \in \mathbb{R}$$

$\lambda$ est appelée *mesure de Lebesgue sur* $\mathbb{R}^d$.
Autre exemple (*mesure de comptage*)
Soit $\Omega$ un ensemble fini ou dénombrable, muni de la tribu $\mathcal{A} = \mathcal{P}(\Omega)$. Pour tout $B \subseteq \Omega$, posons $\mu(B) = \text{Card}(B)$. Il est facile de voir qu'on définit ainsi une mesure $\sigma$-finie sur $(\Omega, \mathcal{P}(\Omega))$ appelée « mesure de comptage ».

### 1.4.3 Définition 8

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Un ensemble $\mathcal{N} \subseteq \Omega$ est dit *négligeable* s'il existe $A \in \mathcal{A}$ tel que $\mathcal{N} \subseteq A$ et $\mu(A) = 0$. Une propriété $P(.)$ concernant les éléments $\omega \in \Omega$ est dite vraie *presque partout* (en abrégé p.p.) [*presque sûrement* (p.s.) lorsque $\mu$ est une probabilité] si $\{\omega \in \Omega$ tels que « non $P(\omega)$ »$\}$ est négligeable. On voit, grâce au 2) de la proposition 9 qu'une réunion dénombrable d'ensembles négligeables est négligeable. Si on considère la mesure de Lebesgue sur $\mathbb{R}^d$, un point est négligeable donc tout ensemble dénombrable est négligable ; par contre, pour la mesure de comptage sur un ensemble dénombrable $\Omega$, aucun point n'est négligeable.
Soient $X$, $Y$, $X_n$ $(n \in \mathbb{N})$ des applications de $\Omega$ dans $\overline{\mathbb{R}}$. On dira que :

- $X$ est finie presque partout si $\{\omega : |X(\omega)| = +\infty\}$ est négligeable.
- $X \leq Y$ presque partout si $\{\omega : X(\omega) > Y(\omega)\}$ est négligeable.
- $X_n \to X$ presque partout si $\{\omega : X_n(\omega) \not\to X(\omega)\}$ est négligeable.

La notion d'ensemble négligeable est liée à la mesure de référence $\mu$. S'il y a ambiguïté, on écrira donc « $\mu$-négligeable » ou « $\mu$-presque-partout » [$\mu$ presque-sûrement lorsque $\mu$ est une probabilité].

## 1.5 Intégration des variables aléatoires étagées positives

Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $X : (\Omega, \mathcal{A}) \to \mathbb{R}_+$ une variable aléatoire étagée positive. Il existe alors une décomposition de $X$ sous la forme

$$X = \sum_{i \in I} \alpha_i.\mathbf{1}_{A_i}$$

où $(A_i, i \in I)$ est une *partition* finie de $\Omega$, $A_i \in \mathcal{A}$ et $\alpha_i \in \mathbb{R}_+$. Le nombre $\sum_{i \in I} \alpha_i \mu(A_i)$ ne dépend pas de la décomposition choisie pour $X$ : en effet, si

$$X = \sum_{j \in J} \beta_j.\mathbf{1}_{B_j} = \sum_{i \in I} \alpha_i.\mathbf{1}_{A_i}$$

où $(B_j, j \in J)$ est une autre partition finie de $\Omega$, $B_j \in \mathcal{A}$ et $\beta_j \in \mathbb{R}_+$. On voit que $\alpha_i = \beta_j$ si $A_i \cap B_j \neq \emptyset$ et donc

$$\begin{aligned}
\sum_i \alpha_i \mu(A_i) &= \sum_i \alpha_i \Big\{ \sum_j \mu(A_i \cap B_j) \Big\} \\
&= \sum_{i,j} \alpha_i \mu(A_i \cap B_j) \\
&= \sum_{i,j} \beta_j \mu(A_i \cap B_j) \\
&= \sum_j \beta_j \mu(B_j)
\end{aligned}$$

(avec la convention $0 \cdot \infty = 0$).

### 1.5.1 Définition 9

Avec les notations précédentes, le nombre $\sum_{i \in I} \alpha_i \mu(A_i)$ s'appelle *l'intégrale de $X$ par rapport à $\mu$* et se note $\int_\Omega X \,\mathrm{d}\mu$. On a donc, pour tout $A \in \mathcal{A}$

$$\mu(A) = \int_\Omega \mathbf{1}_A \,\mathrm{d}\mu$$

### 1.5.2 Proposition 10

Soient $X$ et $Y$ deux variables aléatoires étagées positives, alors :

1. Pour tout $c \in \mathbb{R}_+$, $\quad \int cX \,\mathrm{d}\mu = c \int X \,\mathrm{d}\mu$
2. $\int (X + Y) \,\mathrm{d}\mu = \int X \,\mathrm{d}\mu + \int Y \,\mathrm{d}\mu$
3. Si $X \leq Y$, on a : $\int X \,\mathrm{d}\mu \leq \int Y \,\mathrm{d}\mu$

Démonstration :
L'assertion 1) est évidente.
Soient $(A_1, \ldots, A_n)$, $(B_1, \ldots, B_m)$ deux partitions de $\Omega$ formées d'éléments de $\mathcal{A}$ et $\alpha_i, \ldots, \alpha_n$, $\beta_1, \ldots, \beta_m \in \mathbb{R}_+$ tels que $X = \sum_i \alpha_i.\mathbf{1}_{A_i}$, $Y = \sum_j \beta_j.\mathbf{1}_{B_j}$. Alors

$$X + Y = \sum_{i,j} (\alpha_i + \beta_j).\mathbf{1}_{A_i \cap B_j}$$

et

$$\begin{aligned}
\int (X+Y)\,\mathrm{d}\mu &= \sum_{i,j} (\alpha_i+\beta_j)\mu(A_i\cap B_j) \\
&= \sum_{i,j} \alpha_i \mu(A_i\cap B_j) + \sum_{i,j} \beta_j \mu(A_i\cap B_j) \\
&= \sum_{i} \alpha_i \mu(A_i) + \sum_{j} \beta_j \mu(B_j) \\
&= \int X\,\mathrm{d}\mu + \int Y\,\mathrm{d}\mu
\end{aligned}$$

On suppose que $X \leq Y$. Remarquer que :

$$A_i \cap B_j \neq \emptyset \quad \Rightarrow \quad \alpha_i \leq \beta_j$$

en effet

$$\omega \in A_i \cap B_j \quad \Rightarrow \quad X(\omega) = \alpha_i \ \leq \ Y(\omega) = \beta_j$$

donc

$$\int X\,\mathrm{d}\mu = \sum_{i,j} \alpha_i \mu(A_i\cap B_j) \ \leq \ \sum_{i,j} \beta_j \mu(A_i\cap B_j) = \int Y\,\mathrm{d}\mu$$

## 1.6 Intégration des variables aléatoires réelles positives

### 1.6.1 Définition 10

Soit $X$ une variable aléatoire réelle positive (définie sur l'espace $(\Omega, \mathcal{A}, \mu)$). L'intégrale de $X$ par rapport à $\mu$, notée $\int_\Omega X\,\mathrm{d}\mu$, est donnée par la formule :

$$\int_\Omega X\,\mathrm{d}\mu = \sup\Big\{ \int_\Omega Y\,\mathrm{d}\mu, \quad Y \text{variable aléatoire étagée telle que} \quad 0 \leq Y \leq X \Big\}$$

Remarquer que $\int_\Omega X\,\mathrm{d}\mu$ peut être égale à $+\infty$.

### 1.6.2 Lemme 1

Soient $Z$ une variable aléatoire étagée positive et $(Y_n)_{n\in\mathbb{N}}$ une suite *croissante* de variables aléatoires étagées positives telle que :

$$\lim_n \nearrow Y_n \geq Z$$

Alors

$$\lim_n \nearrow \int Y_n\,\mathrm{d}\mu \ \geq \ \int Z\,\mathrm{d}\mu$$

Démonstration :
Soit $\alpha \in ]0,1[$ . Posons $B_n = \{Y_n \geq \alpha Z\}$, alors $B_n \subseteq B_{n+1}$, $B_n \in \mathcal{A}$ et $\underset{n}{\cup} B_n = \Omega$ en effet, pour tout $\omega \in \Omega$, on a :

$$Z(\omega) = 0 \quad \Rightarrow \quad \omega \in B_0$$

et

$$(Z(\omega) > 0) \quad \Rightarrow \quad (Z(\omega) > \alpha Z(\omega)) \quad \Rightarrow \quad \text{pour tout } n \text{ assez grand } (Y_n(\omega) > \alpha Z(\omega))$$

De plus, si

$$Z = \sum_{\text{finie}} \alpha_i . \mathbf{1}_{A_i}$$

on a

$$Y_n \geq \alpha Z . \mathbf{1}_{B_n} = \sum_{\text{finie}} \alpha \alpha_i . \mathbf{1}_{A_i \cap B_n}$$

Donc

$$\int Y_n \, \mathrm{d}\mu \ \geq \ \alpha \int Z . \mathbf{1}_{B_n} \, \mathrm{d}\mu \ = \ \alpha \sum_{\text{finie}} \alpha_i \mu(A_i \cap B_n)$$

et

$$\lim_n \nearrow \int Y_n \, \mathrm{d}\mu \ \geq \ \alpha \sum_{\text{finie}} \alpha_i \mu(A_i) \ = \ \alpha \int Z \, \mathrm{d}\mu$$

puisque

$$\mu(A_i \cap B_n) \ \nearrow \ \mu(A_i) \quad \text{quand} \quad n \to +\infty$$

### 1.6.3 Proposition 11

Soient $X$ une variable aléatoire réelle positive et $(Y_n)_{n \in \mathbb{N}}$ une suite croissante de variables aléatoires réelles étagées positives telle que :

$$\lim_n \nearrow Y_n = X$$

Alors

$$\int X \, \mathrm{d}\mu \ = \ \lim_n \nearrow \int Y_n \, \mathrm{d}\mu$$

<u>Démonstration</u> :
Pour tout $n \in \mathbb{N}$, on a : $0 \leq Y_n \leq Y_{n+1} \leq X$ donc

$$\int Y_n \, \mathrm{d}\mu \ \leq \ \int X \, \mathrm{d}\mu$$

par définition et

$$\lim_n \nearrow \int Y_n \, \mathrm{d}\mu \ \leq \ \int X \, \mathrm{d}\mu$$

Réciproquement, soit $Z$ une variable aléatoire réelle étagée telle que

$$0 \ \leq \ Z \ \leq \ X \ = \ \lim_n \nearrow Y_n$$

alors, d'après le lemme 1 :

$$\int Z \, \mathrm{d}\mu \ \leq \ \lim_n \nearrow \int Y_n \, \mathrm{d}\mu$$

d'où l'inégalité

$$\int X \, \mathrm{d}\mu \ \leq \ \lim_n \nearrow \int Y_n \, \mathrm{d}\mu$$

### 1.6.4 Proposition 12

Soient $X$ et $Y$ deux variables aléatoires réelles positives, alors :

1. pour tout $c \in \mathbb{R}_+$ : $\int cX \,\mathrm{d}\mu = c \int X \,\mathrm{d}\mu$
2. $\int (X+Y) \,\mathrm{d}\mu = \int X \,\mathrm{d}\mu + \int Y \,\mathrm{d}\mu$
3. si $X \leq Y$, on a : $\int X \,\mathrm{d}\mu \ \leq \ \int Y \,\mathrm{d}\mu$

Démonstration :
Le 3) résulte de la définition 10. On sait, d'autre part, d'après la proposition 8, qu'il existe $X_n$, $Y_n$ variables aléatoires réelles étagées positives telles que

$$\lim \nearrow X_n = X \qquad \lim \nearrow Y_n = Y$$

donc :

$$\begin{aligned}\int (X+Y) \,\mathrm{d}\mu &= \lim_n \nearrow \int (X_n + Y_n) \,\mathrm{d}\mu \\ &= \lim_n \nearrow \int X_n \,\mathrm{d}\mu + \lim_n \nearrow \int Y_n \,\mathrm{d}\mu \\ &= \int X \,\mathrm{d}\mu + \int Y \,\mathrm{d}\mu\end{aligned}$$

On montre de même que

$$\int cX \,\mathrm{d}\mu = c \int X \,\mathrm{d}\mu \qquad (c \in \mathbb{R}_+)$$

### 1.6.5 Définition 12

Une variable aléatoire réelle positive $X$ est dite *intégrable* si $\int_\Omega X \,\mathrm{d}\mu < \infty$. Donc si $Y$ est une variable aléatoire réelle telle que $0 \leq Y \leq X$, on voit, d'après la proposition 12, que $Y$ est intégrable dès que $X$ l'est.

## 1.7 Intégration des variables aléatoires réelles

### 1.7.1 Définition 13

Une variable aléatoire réelle $X$ (définie sur l'espace $(\Omega, \mathcal{A}, \mu)$) est dite *intégrable* si $|X|$ est intégrable. Son intégrale est définie par

$$\int_\Omega X \,\mathrm{d}\mu = \int_\Omega X^+ \,\mathrm{d}\mu - \int_\Omega X^- \,\mathrm{d}\mu$$

Remarquer que $X^+$ et $X^-$ sont intégrables puisque
- $0 \leq X^+ = \max\{X, 0\} \ \leq \ |X|$
- $0 \leq X^- = \max\{-X, 0\} \ \leq \ |X|$

### 1.7.2 Proposition 13

Soient $X$ et $Y$ deux variables aléatoires réelles intégrables, alors

1. pour tout $c \in \mathbb{R}$ : $\int cX \,\mathrm{d}\mu = c \int X \,\mathrm{d}\mu$

2. Si $X+Y$ a un sens : $\int (X+Y)\,\mathrm{d}\mu = \int X\,\mathrm{d}\mu + \int Y\,\mathrm{d}\mu$
3. si $X \leq Y$, $\int X\,\mathrm{d}\mu \;\leq\; \int Y\,\mathrm{d}\mu$
4. $|\int X\,\mathrm{d}\mu| \;\leq\; \int |X|\,\mathrm{d}\mu$

Démonstration :
On a : $(-X)^+ = X^-$ donc :

$$\begin{aligned}\int (-X)\,\mathrm{d}\mu &= \int (-X)^+\,\mathrm{d}\mu - \int (-X)^-\,\mathrm{d}\mu \\ &= \int X^-\,\mathrm{d}\mu - \int X^+\,\mathrm{d}\mu \\ &= -\int X\,\mathrm{d}\mu\end{aligned}$$

L'égalité 1) est évidente si $c \geq 0$ ;
si $c < 0$, on écrit

$$\int cX\,\mathrm{d}\mu = \int (-c)(-X)\,\mathrm{d}\mu = (-c)\int (-X)\,\mathrm{d}\mu = c\int X\,\mathrm{d}\mu$$

On a de plus : $|X+Y| \leq |X| + |Y|$, donc $X+Y$ est intégrable et :

$$(X+Y) = (X+Y)^+ - (X+Y)^- = (X^+ - X^-) + (Y^+ - Y^-)$$

donc

$$(X+Y)^+ + X^- + Y^- = (X+Y)^- + X^+ + Y^+$$

$$\int (X+Y)^+\,\mathrm{d}\mu + \int X^-\,\mathrm{d}\mu + \int Y^-\,\mathrm{d}\mu = \int (X+Y)^-\,\mathrm{d}\mu + \int X^+\,\mathrm{d}\mu + \int Y^+\,\mathrm{d}\mu$$

et, ces quantités étant finies, par hypothèse :

$$\int (X+Y)^+\,\mathrm{d}\mu - \int (X+Y)^-\,\mathrm{d}\mu = \left(\int X^+\,\mathrm{d}\mu - \int X^-\,\mathrm{d}\mu\right) + \left(\int Y^+\,\mathrm{d}\mu - \int Y^-\,\mathrm{d}\mu\right)$$

d'où l'égalité 2).
Pour le 3) on a :

$$\text{si} \quad X \leq Y, \quad Y - X \geq 0 \quad \text{et} \quad 0 \;\leq\; \int (Y-X)\,\mathrm{d}\mu = \int Y\,\mathrm{d}\mu - \int X\,\mathrm{d}\mu$$

Pour le 4) on a, puisque $|X| = X^+ + X^-$ :

$$\int |X|\,\mathrm{d}\mu = \int X^+\,\mathrm{d}\mu + \int X^-\,\mathrm{d}\mu \;\geq\; \left|\int X^+\,\mathrm{d}\mu - \int X^-\,\mathrm{d}\mu\right| = \left|\int X\,\mathrm{d}\mu\right|$$

### 1.7.3 Définition 14

On notera $\mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ l'ensemble des variables aléatoires réelles $X$ *finies* et intégrables. La proposition 13 montre que cet ensemble est muni, naturellement, d'une structure d'espace vectoriel sur $\mathbb{R}$, l'application

$$X \in \mathcal{L}^1 \longmapsto \int X\,\mathrm{d}\mu \in \mathbb{R}$$

étant une forme linéaire *positive* sur cet espace (i.e. que $X \geq 0 \Rightarrow \int X \,\mathrm{d}\mu \geq 0$). Une variable aléatoire $X$ *à valeurs dans* $\mathbb{C}$ est dite intégrable si $|X| \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$. On posera alors :

$$\int X \,\mathrm{d}\mu = \int \mathcal{R}\mathrm{e}(X) \,\mathrm{d}\mu + \imath \int \mathcal{I}\mathrm{m}(X) \,\mathrm{d}\mu$$

On désignera par $\mathcal{L}^1_{\mathbb{C}}(\Omega, \mathcal{A}, \mu)$ l'ensemble des variables aléatoires complexes et intégrables. Cet ensemble forme un espace vectoriel sur $\mathbb{C}$ ; de plus, on vérifie facilement que l'application :

$$X \in \mathcal{L}^1_{\mathbb{C}} \longmapsto \int X \,\mathrm{d}\mu \in \mathbb{C}$$

est une forme linéaire sur cet espace.
Soit $X \in \mathcal{L}^1_{\mathbb{C}}$. Alors $\int X \,\mathrm{d}\mu$ est un nombre complexe de la forme $\rho e^{\imath\theta}$ où $\rho \geq 0$ ; donc

$$\begin{aligned} \left| \int X \,\mathrm{d}\mu \right| &= \rho = e^{-\imath\theta} \int X \,\mathrm{d}\mu = \int e^{-\imath\theta} X \,\mathrm{d}\mu \\ &= \int \mathcal{R}\mathrm{e}(e^{-\imath\theta} X) \,\mathrm{d}\mu \leq \int |X| \,\mathrm{d}\mu \end{aligned}$$

On voit ainsi que pour tout $X \in \mathcal{L}^1_{\mathbb{C}}$ : $|\int X \,\mathrm{d}\mu| \leq \int |X| \,\mathrm{d}\mu$

### 1.7.4 Proposition 14

1. Soit $X$ une variable aléatoire à valeurs dans $\overline{\mathbb{R}}$ ou $\mathbb{C}$. S'il existe une variable aléatoire réelle $Y$ intégrable et positive telle que $|X| \leq Y$, alors $X$ est intégrable.
2. Si $X$ et $Y$ sont deux variables aléatoires à valeurs dans $\overline{\mathbb{R}}$ ou $\mathbb{C}$ égales presque partout (p.p.) et si $X$ est intégrable, alors $Y$ est intégrable ; de plus : $\int X \,\mathrm{d}\mu = \int Y \,\mathrm{d}\mu$.
3. Si $X$ est une variable aléatoire intégrable, alors, pour tout $\lambda \in \mathbb{R}^*_+$

$$\mu(|X| \geq \lambda) \leq \frac{1}{\lambda} \int |X| \,\mathrm{d}\mu \quad \text{et} \quad (|X| < \infty) \quad \text{presque partout}$$

<u>Démonstration</u> :
Le 1) est évident
Pour 2) si $S = \{X \neq Y\}$ ; alors $S \in \mathcal{A}$, de plus

$$|X| \leq (+\infty).\mathbf{1}_S + |Y|$$

$$|Y| \leq (+\infty).\mathbf{1}_S + |X|$$

et

$$\int (+\infty).\mathbf{1}_S \,\mathrm{d}\mu = \lim_{n \to +\infty} \int n.\mathbf{1}_S \,\mathrm{d}\mu = 0 \quad \text{puisque} \quad \mu(S) = 0$$

donc

$$\int |X| \,\mathrm{d}\mu = \int |Y| \,\mathrm{d}\mu < \infty$$

Pour 3) si $|X| \in \mathcal{L}^1$, on a, pour tout $\lambda > 0$ :

$$\lambda.\mathbf{1}_{(|X|\geq\lambda)} \leq |X|$$

donc

$$\int \lambda.\mathbf{1}_{(|X|\geq\lambda)}\,\mathrm{d}\mu = \lambda\mu(|X| \geq \lambda) \leq \int |X|\,\mathrm{d}\mu < \infty$$

De plus

$$\mu(|X| = +\infty) = \lim_{\substack{\lambda\to\infty \\ \lambda\in\mathbb{N}}} \mu(|X| \geq \lambda) \leq \lim_{\substack{\lambda\to\infty \\ \lambda\in\mathbb{N}}} \frac{1}{\lambda}\int |X|\,\mathrm{d}\mu = 0$$

donc $(|X| < \infty)$ presque partout.
Remarques :

1. La propostion 14 montre que, si $X$ est une variable aléatoire intégrable, alors $X$ et $X.\mathbf{1}_{(|X|<\infty)}$ sont égales presque partout et ont même intégrale.
2. Si $X$ est une variable aléatoire réelle *positive*, les définitions précédentes montrent que $\int X\,\mathrm{d}\mu$ a toujours un sens (et peut valoir $+\infty$). Si $X$ est une variable aléatoire réelle *quelconque* et si $\int X^+\,\mathrm{d}\mu < \infty$ ou bien $\int X^-\,\mathrm{d}\mu < \infty$ (par exemple : $X$ majorée ou minorée par une variable aléatoire intégrable), on peut, par extension, définir l'intégrale de $X$ en posant :

$$\int X\,\mathrm{d}\mu = \int X^+\,\mathrm{d}\mu - \int X^-\,\mathrm{d}\mu$$

qui appartient à $[-\infty, +\infty] = \overline{\mathbb{R}}$

### 1.7.5 Définition 15 (*espérance des variables aléatoires réelles*)

On suppose que la mesure $\mu$ est une probabilité $P$ sur l'espace mesurable $(\Omega, \mathcal{a})$. Soit $X \in \mathcal{L}^1(\Omega, \mathcal{a}, P)$. *L'espérance de la variable aléatoire réelle $X$ est alors définie par* :

$$E(X) = \int_\Omega X\,\mathrm{d}P$$

Remarquer que, si $X$ est une variable aléatoire réelle discrète ($\mathcal{I}\mathrm{m}(X) = X(\Omega)$ est fini ou dénombrable).

$$X = \sum_{x\in\mathcal{I}\mathrm{m}(X)} x.\mathbf{1}_{(X=x)}$$

$$|X| = \sum_{x\in\mathcal{I}\mathrm{m}(X)} |x|.\mathbf{1}_{(X=x)}$$

On vérifie facilement, dans ce cas, que $X$ est intégrable si et seulement si

$$E(|X|) = \int_\Omega |X|\,\mathrm{d}P = \sum_{x\in\mathcal{I}\mathrm{m}(X)} |x|.P(X=x) < \infty$$

et que

$$E(X) = \int_\Omega X\,\mathrm{d}P = \sum_{x\in\mathcal{I}\mathrm{m}(X)} x.P(X=x)$$

Remarquer que

$$\int \left( \sum_x |x|.\mathbf{1}_{(X=x)} \right) \mathrm{d}P = \lim_{n\to+\infty} \int \left( \sum_{|x|\leq n} |x|.\mathbf{1}_{(X=x)} \right) \mathrm{d}P = \lim_{n\to+\infty} \sum_{|x|\leq n} |x|.P(X=x)$$

### 1.7.6 Proposition 16 (loi d'une variable aléatoire)

Soit $X$ une variable aléatoire définie sur l'espace de probabilité $(\Omega, \mathcal{A}, P)$ et à valeurs dans un espace mesurable quelconque $(F, \mathcal{B})$. La formule suivante :

$$(*) \text{ pour tout } \quad B \in \mathcal{B} \quad : \quad \nu(B) = P\left(X^{-1}(B)\right)$$

définit alors une probabilité $\nu$ sur l'espace $(F, \mathcal{B})$. On dira que $\nu$ est *la loi (image) de la variable aléatoire* $X$ et on posera $\nu = P_X$. De plus, pour toute application mesurable $\varphi$ de $(F, \mathcal{B})$ dans $\overline{\mathbb{R}}$, on a $\varphi \in \mathcal{L}^1(F, \mathcal{B}, P_X)$ si et seulement si $\varphi \circ X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ et dans ce cas

$$(**) \qquad \int_F \varphi \,\mathrm{d}P_X = \int_\Omega \varphi \circ X \,\mathrm{d}P \qquad (\textit{formule des lois images})$$

Démonstration :
La formule $(*)$ montre que, si $B_n \quad (n \in \mathbb{N})$ est une suite d'éléments de $\mathcal{B}$ deux a deux disjoints, on a :

$$X^{-1}\left(\underset{n}{\cup} B_n\right) = \underset{n}{\cup} X^{-1}(B_n)$$

et les $X^{-1}(B_n)$ $(n \in \mathbb{N})$ étant deux à deux disjoints :

$$\begin{aligned} \nu\left(\underset{n}{\cup} B_n\right) &= P\left(X^{-1}\left(\underset{n}{\cup} B_n\right)\right) = P\left(\underset{n}{\cup} X^{-1}(B_n)\right) \\ &= \sum_n P\left(X^{-1}(B_n)\right) \\ &= \sum_n \nu(B_n) \end{aligned}$$

De plus $\nu(F) = P\left(X^{-1}(F)\right) = P(\Omega) = 1$.
$\nu$ est bien une probabilité sur l'espace $(F, B)$.
Pour démontrer la formule des lois images, remarquons d'abord, que si $\varphi = \mathbf{1}_B$, $(B \in \mathcal{B})$, on a

$$\begin{aligned} \int_F \varphi \,\mathrm{d}P_X = P_X(B) = P\left(X^{-1}(B)\right) &= \int_\Omega \mathbf{1}_{X^{-1}(B)} \,\mathrm{d}P \quad (\text{par définition}) \\ = \int_\Omega \mathbf{1}_B \circ X \,\mathrm{d}P &= \int_\Omega (\varphi \circ X) \,\mathrm{d}P \end{aligned}$$

La formule $(**)$ est donc vérifiée lorsque $\varphi = \mathbf{1}_B$ $(B \in \mathcal{B})$, et par linéarité, lorsque $\varphi$ est une variable aléatoire étagée de la forme $\varphi = \sum_{\text{finie}} \alpha_i.\mathbf{1}_{B_i}$ où $B_i \in \mathcal{B}$, $\alpha_i \in \mathbb{R}_+$. Soit $\varphi$ une application mesurable *positive* de $(F, \mathcal{B})$ dans $\overline{\mathbb{R}}$ ; il existe (proposition 8) une suite croissante $(\varphi_n)_{n\in\mathbb{N}}$ de variables aléatoires étagées positives telles que $\varphi = \lim_n \nearrow \varphi_n$ ; donc :

$$\int_F \varphi \,\mathrm{d}P_X = \lim \nearrow \int_F \varphi_n \,\mathrm{d}P_X = \lim \nearrow \int_\Omega (\varphi_n \circ X) \,\mathrm{d}P = \int_\Omega (\varphi \circ X) \,\mathrm{d}P$$

On voit alors que si $\varphi$ est une application mesurable quelconque de $(F, \mathcal{B})$ dans $\overline{\mathbb{R}}$, on a

$$|\varphi| \in \mathcal{L}^1(F, \mathcal{B}, P_X) \qquad \Longleftrightarrow \qquad |\varphi| \circ X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$$

et que, dans ce cas, en décomposant $\varphi$ :

$$\begin{aligned}\int_F \varphi \, \mathrm{d}P_X &= \int_F \varphi^+ \, \mathrm{d}P_X - \int_F \varphi^- \, \mathrm{d}P_X \\ &= \int_\Omega (\varphi^+ \circ X) \, \mathrm{d}P_X - \int_\Omega (\varphi^- \circ X) \, \mathrm{d}P_X \\ &= \int_\Omega (\varphi \circ X) \, \mathrm{d}P\end{aligned}$$

Remarques :

1. Avec les notations de la proposition 16, si $F$ est un ensembe fini ou dénombrable, on a $\mathcal{B} = \mathcal{P}(F)$. La loi $P_X$ d'une variable aléatoire discrète $X$ à valeurs dans $F$ s'identifie alors avec la famille dénombrable de nombres positifs de somme égale à 1 :

   $$P_X(x) = P(X = x) \qquad (x \in F)$$

   En effet, pour tout $B \subseteq F$ :

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

   et

   $$P_X(F) = \sum_{x \in F} P_X(x) = 1$$

   On dit aussi que $(P_X(x), x \in F)$ est *la densité discrète* de $X$. Pour toute application $\varphi$ de $F$ dans $\mathbb{R}$, on a : $\varphi \circ X$ est intégrable si et seulement si $\sum_{x \in F} |\varphi(x)| P(X = x) < \infty$ et dans ce cas :

   $$E\{\varphi \circ X\} = \int_F \varphi \, \mathrm{d}P_X = \sum_{x \in F} \varphi(x) P_X(x) = \sum_{x \in F} \varphi(x) P(X = x) = \int_\Omega \varphi \circ X \, \mathrm{d}P$$

2. Soit $X$ une variable aléatoire réelle intégrable définie sur l'espace mesuré $(\Omega, \mathcal{A}, \mu)$ ; on notera indifféremment, dans la suite, suivant les contextes :

   $$\int_\Omega X \, \mathrm{d}\mu = \int_\Omega X(\omega) \, \mathrm{d}\mu(\omega) = \int_\Omega X(\omega) \mu \, (\mathrm{d}\omega)$$

   Si $A \in \mathcal{A}$, on posera, par définition :

   $$\int_A X \, \mathrm{d}\mu = \int_\Omega X.\mathbf{1}_A \, \mathrm{d}\mu$$

## 1.8 Théorèmes de convergence

### 1.8.1 Théorème 1 (Beppo-Levi)

Soit $(X_n)_{n\in\mathbb{N}}$ une suite *croissante* de variables aléatoires réelles *positives* (définies sur l'espace mesuré $(\Omega, \mathcal{A}, \mu)$). Posons $X_\infty = \lim_n \nearrow X_n$. Alors :

$$\int X_\infty \, \mathrm{d}\mu = \lim_n \nearrow \int X_n \, \mathrm{d}\mu$$

<u>Démonstration</u> :
Pour tout $n \in \mathbb{N}$ soit $(X_{n,k})_{k\in\mathbb{N}}$ une suite croissante de variables aléatoires étagées positives telle que :

$$0 \leq X_{n,k} \leq X_{n,k+1} \nearrow X_n, \qquad \text{quand} \qquad k \to +\infty$$

Posons

$$Y_k = \max\{X_{m,k},\ 0 \leq m \leq k\}$$

Alors $Y_k$ est *étagée* et

$$0 \leq Y_k \leq Y_{k+1} \leq X_{k+1} \leq X_\infty \qquad (1)$$

de plus, pour tout $m \leq l$ dans $\mathbb{N}$ :

$$X_{m,l} \leq Y_l \leq \lim_k \nearrow Y_k$$

donc

$$X_m \leq \lim_{l\to\infty} X_{m,l} \leq \lim_k \nearrow Y_k$$

et

$$\lim_m \nearrow X_m = X_\infty \leq \lim_k \nearrow Y_k \leq X_\infty \qquad (2)$$

Des inégalités (1) et (2), on déduit que :

$$\int X_\infty \, \mathrm{d}\mu = \lim_k \nearrow \int Y_k \, \mathrm{d}\mu = \lim_k \nearrow \int X_k \, \mathrm{d}\mu$$

<u>Corollaire</u> :
Soit $(Z_n)_{n\in\mathbb{N}}$ une suite quelconque de variables aléatoires réelles positives alors :

$$\int \left(\sum_n Z_n\right) \mathrm{d}\mu = \sum_n \int Z_n \, \mathrm{d}\mu$$

<u>Démonstration</u> :
Pour tout $m \in \mathbb{N}$, soit $X_m = \sum_{n=0}^{m} Z_n$, alors

$$0 \leq X_m \leq X_{m+1} \nearrow X_\infty = \sum_n Z_n$$

Il suffit d'appliquer le théorème 1 .

### 1.8.2 Théorème 2 (Lemme de Fatou)

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles. S'il existe une variable aléatoire réelle intégrable $Y$ telle que, pour tout $n$ : $Y \leq X_n$ alors :

$$\int \underline{\lim}_{n\to\infty} X_n \,\mathrm{d}\mu \leq \underline{\lim}_{n\to\infty} \int X_n \,\mathrm{d}\mu$$

S'il existe une variable aléatoire réelle intégrable $Z$ telle que, pour tout $n$ : $X_n \leq Z$ alors :

$$\overline{\lim}_{n\to\infty} \int X_n \,\mathrm{d}\mu \leq \int \overline{\lim}_{n\to\infty} X_n \,\mathrm{d}\mu$$

<u>Démonstration</u> :
On suppose d'abord que, pour tout $n$, $X_n \geq 0$, alors

$$0 \leq Y_m = \inf\{X_n, n \geq m\} \leq Y_{m+1} \nearrow \underline{\lim}_{n\to\infty} X_n, \qquad \text{quand} \qquad m \to \infty$$

donc, d'après le théorème 1

$$\int \underline{\lim} X_n \,\mathrm{d}\mu = \lim_{m\to\infty} \nearrow \int Y_m \,\mathrm{d}\mu \leq \lim_{m\to\infty} \left( \inf_{n\geq m} \int X_n \,\mathrm{d}\mu \right) = \underline{\lim}_{n\to\infty} \int X_n \,\mathrm{d}\mu$$

puisque, pour tout $n \geq m$, $Y_m \leq X_n$ et donc $\int Y_m \,\mathrm{d}\mu \leq \int X_n \,\mathrm{d}\mu$.
Si, pour tout $n$ : $X_n \geq Y$ où $Y \in \mathcal{L}^1$, il suffit de poser $X_n^{'} = X_n - Y \geq 0$ et de remarquer que

$$\underline{\lim}_{n\to\infty} X_n^{'} = \underline{\lim}_{n\to\infty} X_n - Y \geq 0$$

donc

$$\begin{aligned}\int \underline{\lim}_{n\to\infty} X_n \,\mathrm{d}\mu &= \int \underline{\lim}_{n\to\infty} X_n^{'} \,\mathrm{d}\mu + \int Y \,\mathrm{d}\mu \\ &\leq \underline{\lim}_{n\to\infty} \int X_n^{'} \,\mathrm{d}\mu + \int Y \,\mathrm{d}\mu = \underline{\lim}_{n\to\infty} \int X_n \,\mathrm{d}\mu\end{aligned}$$

Pour démontrer la seconde assertion du théorème, il suffit de changer $X_n$ en $-X_n$ $(n \in \mathbb{N})$ .

### 1.8.3 Théorème 3 (Théorème de Lesbesgue ou de « convergence dominée »)

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles. S'il existe une variable aléatoire réelle *intégrable* $Y$ telle que, pour tout $n$ : $|X_n| \leq Y$ presque partout et si $(X_n)$ converge presque partout vers une variable aléatoire réelle $X_\infty$, quand $n \to \infty$ alors $X_\infty$ est intégrable et $\int |X_n - X_\infty| \,\mathrm{d}\mu \to 0$, quand $n \to \infty$ ; en particulier $\int X_n \,\mathrm{d}\mu \to \int X_\infty \,\mathrm{d}\mu$ quand $n \to \infty$.
<u>Démonstration</u> :
Commençons par remarquer que, pour tout $n$ :

**(*)**$|X_n - X_\infty| \leq |X_\infty| + Y$ presque partout puisque $|X_n| \leq Y$ presque partout

Quitte à modifier $Y$ sur un ensemble négligeable, on peut supposer que les inégalités précédentes sont valables partout et, en faisant tendre $n$ vers l'infini, on voit que $|X_\infty| \leq Y$ presque partout, donc $X_\infty \in \mathcal{L}^1$. D'après le lemme de Fatou et grâce à (*) :

$$\overline{\lim_{n\to\infty}} \int |X_n - X_\infty| \, \mathrm{d}\mu \leq \int \overline{\lim_{n\to\infty}} |X_n - X_\infty| \, \mathrm{d}\mu = 0$$

car

$$\overline{\lim} |X_n - X_\infty| = 0 \quad \text{presque partout et} \quad |X_\infty| + Y \in \mathcal{L}^1$$

donc

$$\begin{aligned} \left| \int X_n \, \mathrm{d}\mu - \int X_\infty \, \mathrm{d}\mu \right| &= \left| \int (X_n - X_\infty) \, \mathrm{d}\mu \right| \\ &\leq \int |X_n - X_\infty| \, \mathrm{d}\mu \to 0 \quad \text{quand} \quad n \to \infty \end{aligned}$$

Corollaire 1 :
Soit $(X_s)_{s\in S}$ une famille de variables aléatoires réelles où $S$ est un espace métrique. On suppose que :

1. il existe $l \in S$ tel que $\lim_{s\to l} X_s = X_l$ presque partout
2. pour tout $s$ variant dans une boule ouverte centrée en $l$ : $|X_s| \leq Y$ presque partout où $Y$ est une variable aléatoire réelle intégrable fixe, alors

$$\lim_{s\to l} \int X_s^{(\cdot)} \, \mathrm{d}\mu = \int X_l^{(\cdot)} \, \mathrm{d}\mu$$

Démonstration :
Il suffit de vérifier que, pour toute suite $(s_n)_{n\in\mathbb{N}}$ d'éléments de $S$ telle que $\lim_{n\to\infty} s_n = l$, on a

$$\lim_{n\to\infty} \int X_{s_n} \, \mathrm{d}\mu = \int X_l \, \mathrm{d}\mu$$

Appliquer le théorème de Lebesgue.
Corollaire 2 : (*Dérivation sous le signe intégral*)
Soit $(X_s)_{s\in S}$ une famille de variables aléatoires réelles intégrables où $S$ est un intervalle ouvert de $\mathbb{R}$. On suppose que, pour presque tout $\omega \in \Omega$,

1. la fonction $s \in S \longmapsto X_s(\omega)$ est dérivable sur $S$
2. $|\frac{\mathrm{d}}{\mathrm{d}s} X_s(\omega)| \leq Y(\omega)$ en tout point $s$ de $S$ où $Y(.)$ est une variable aléatoire réelle intégrable fixe.

Alors $\int X_s^{(\cdot)} \, \mathrm{d}\mu$ est dérivable en $s \in S$ et

$$\frac{\mathrm{d}}{\mathrm{d}s} \int X_s^{(\cdot)} \, \mathrm{d}\mu = \int \left( \frac{\mathrm{d}}{\mathrm{d}s} X_s^{(\cdot)} \right) \mathrm{d}\mu$$

Démonstration :
Soit $s_0 \in S$. Pour tout $h \neq 0$, assez petit, on a :

$$\frac{1}{h} \left( \int X_{s_0+h} \, \mathrm{d}\mu - \int X_{s_0} \, \mathrm{d}\mu \right) = \int \left( \frac{X_{s_0+h} - X_{s_0}}{h} \right) \mathrm{d}\mu$$

et, d'après le théorème des accroissements finis :

$$\left| \frac{X_{s_0+h} - X_{s_0}}{h} \right| = \left| X'_{\theta_h} \right| \leq Y \qquad \text{presque partout}$$

où $X_s^{'} = \frac{\mathrm{d}}{\mathrm{d}s} X_s$ et $\theta_h$ est compris entre $s_0$ et $s_0 + h$. On voit alors, d'après le corollaire 1, que :

$$\lim_{\substack{h \to 0 \\ h \neq 0}} \frac{1}{h} \left( \int X_{s_0+h} \, \mathrm{d}\mu - \int X_{s_0} \, \mathrm{d}\mu \right) = \int X_{s_0}^{'} \, \mathrm{d}\mu$$

### 1.8.4 Proposition 17 (lien avec l'intégrale de Riemann)

1. Soient $a$ et $b$ dans $\mathbb{R}$, $a < b$ et $f \in \mathcal{C}([a,b], \mathbb{R})$ (espace des fonctions continues de $[a,b]$ dans $\mathbb{R}$). La fonction $f.\mathbf{1}_{[a,b]}$ est alors intégrable pour la mesure de Lebesgue $\lambda$ sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ et l'intégrale de Riemann $\int_a^b f(x) \, \mathrm{d}x$ est égale à $\int f.\mathbf{1}_{[a,b]} \, \mathrm{d}\lambda$.
2. Soit $f$ une application de $\mathbb{R}$ dans $\mathbb{R}$ continue par morceaux, alors $f \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \lambda)$ si et seulement si l'intégrale $I = \int_{-\infty}^{+\infty} f(x) \, \mathrm{d}x$ est *absolument convergente* et dans ce cas $I = \int f \, \mathrm{d}\lambda$

Démonstration :
Assertion 1) : Soit $f \in \mathcal{C}([a,b], \mathbb{R})$ On sait qu'il existe une constante $M \geq 0$ telle que, pour tout $x \in [a,b]$ : $|f(x)| \leq M$, donc

$$\left| f.\mathbf{1}_{[a,b]} \right| \leq M.\mathbf{1}_{[a,b]}$$

et

$$\int |f|.\mathbf{1}_{[a,b]} \, \mathrm{d}\lambda \leq M.(b-a) < \infty \quad : \quad f.\mathbf{1}_{[a,b]} \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \lambda)$$

Soient $\Delta = \{t_0 = a, < t_1 < \ldots < t_{n_\Delta} = b\}$ une subdivision de l'intervalle $[a,b]$

$$|\Delta| = \max_i |t_{i+1} - t_i| \quad \text{et} \quad f_\Delta = \sum_{i=0}^{n_\Delta - 1} f(t_i).\mathbf{1}_{[t_i, t_{i+1}[}$$

On sait, de plus, que $f_\Delta \to f$, quand $|\Delta| \to 0$, uniformément sur $[a,b[$ (cf. le cours de DEUG) et que les sommes de Riemann

$$R(\Delta) = \sum_i f(t_i).(t_{i+1} - t_i)$$

convergent, quand $|\Delta| \to 0$, vers $\int_a^b f(x) \, \mathrm{d}x$ ; or $R(\Delta) = \int f_\Delta \, \mathrm{d}\lambda$ et $|f_\Delta| \leq M.\mathbf{1}_{[a,b]}$ donc (théorème de Lebesgue)

$$\lim_{|\Delta| \to 0} R(\Delta) = \lim_{|\Delta| \to 0} \int f_\Delta \, \mathrm{d}\lambda = \int f.\mathbf{1}_{[a,b[} \, \mathrm{d}\lambda = \int_a^b f(x) \, \mathrm{d}x$$

remarquer aussi que

$$\int f.\mathbf{1}_{[a,b]} \, \mathrm{d}\lambda = \int f.\mathbf{1}_{[a,b[} \, \mathrm{d}\lambda = \int f.\mathbf{1}_{]a,b[} \, \mathrm{d}\lambda$$

puisque $\lambda(\{b\}) = \lambda(\{a\}) = 0$.
L'assertion 2) est laissée en exercice.

Considérons l'espace discret $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$ où $\mu$ est la mesure de comptage (pour tout $B \subseteq \mathbb{N} : \mu(B) = \mathrm{Card}(B)$). Soit $U = (U_n)_{n\in\mathbb{N}}$ une suite de nombres réels ou complexes. On vérifie facilement que $U \in \mathcal{L}^1_{\mathbb{C}}(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$ si et seulement si la série $\sum_{n\geq 0} U_n$ est *absolument convergente* ($\sum_{n\geq 0} |U_n| < \infty$) et que dans ce cas

$$\int_{\mathbb{N}} U \,\mathrm{d}\mu = \sum_{n\geq 0} U_n$$

### 1.8.5 Proposition 18

Soit $V = (V_{n,m})_{(n,m)\in\mathbb{N}\times\mathbb{N}}$ une suite à deux indices $n$ et $m$ ($V_{n,m} \in \mathbb{C}$). On suppose que :

1. Pour tout $n$, $\lim\limits_{m\to\infty} V_{n,m}$ existe
2. $\sum_{n\geq 0}\{\sup_m |V_{n,m}|\} < \infty$

alors

$$\lim_{m\to\infty}\left\{\sum_{n\geq 0} V_{n,m}\right\} = \sum_{n\geq 0}\left\{\lim_{m\to\infty} V_{n,m}\right\} \qquad (*)$$

L'égalité (*) reste valable si, au lieu des conditions 1) et 2), on suppose 3) : pour tous $n$ et $m$ : $V_{n,m} \in \mathbb{R}_+$ et $V_{n,m} \leq V_{n,m+1}$.

Démonstration :

Considérons

$$f_m = (V_{n,m})_{n\in\mathbb{N}} \quad \text{et} \quad g = \left(\sup_m |V_{n,m}|\right)_{n\in\mathbb{N}}$$

Pour tout $n \in \mathbb{N}$, on a

$$|f_m(n)| = |V_{n,m}| \leq g(n)$$

La condition 1) signifie que $f_m$ converge simplement, quand $m \to \infty$ et 2) signifie que $g \in \mathcal{L}^1_{\mathbb{C}}(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$. Pour obtenir (*) il suffit d'appliquer le théorème de Lebesgue sous les conditions 1) et 2), ou bien le théorème de Beppo-Levi sous l'hypothèse 3).

# Chapitre 2

# Espaces produits, indépendance

## 2.1 Espaces produits

Soient $(F_i, \mathcal{B}_i)$, $i = 1, 2, \ldots, n$ des espaces mesurables. Posons :

$$F = F_1 \times F_2 \times \ldots \times F_n = \prod_{i=1}^{n} F_i$$

On munit l'ensemble $F$ de la tribu $\mathcal{B}$ engendrée par les pavés de la forme

$$A_1 \times A_2 \times \ldots \times A_n \quad \text{où} \quad A_i \in \mathcal{B}_i,\ i = 1, \ldots, n$$

$\mathcal{B}$ s'appelle la *tribu produit* et se note :

$$\mathcal{B} = \mathcal{B}_1 \otimes \mathcal{B}_2 \otimes \ldots \otimes \mathcal{B}_n = \bigotimes_{i=1}^{n} \mathcal{B}_i$$

Remarquer que $\mathcal{B}$ est la plus petite tribu sur $F$ rendant mesurables les applications coordonnées $\theta_i, (i = 1, \ldots, n)$, de $F$ dans $(F_i, \mathcal{B}_i)$ définies par :

$$\theta_i\left((x_1, x_2, \ldots, x_n)\right) = x_i \qquad \text{où} \qquad (x_1, \ldots, x_n) \in F$$

Si pour tout $i$, $F_i = \mathbb{R}$ et $\mathcal{B}_i = \mathcal{B}_{\mathbb{R}}$ (tribu borélienne de $\mathbb{R}$) on vérifie facilement que, pout tout $n \in \mathbb{N}^*$ :

$$\mathcal{B}_{\mathbb{R}^n} = \mathcal{B}_{\mathbb{R}} \otimes \ldots \otimes \mathcal{B}_{\mathbb{R}} \qquad n \text{ fois}$$

et que $\mathcal{B}_{\mathbb{R}^{n+m}}$ s'identifie à $\mathcal{B}_{\mathbb{R}^n} \otimes \mathcal{B}_{\mathbb{R}^m}$ $(m \in \mathbb{N}^*)$.

### 2.1.1 Théorème 1

Soient $\mu_1, \mu_2, \ldots, \mu_n$ des mesures $\geq 0$ $\sigma$-finies définies sur

$$(F_1, \mathcal{B}_1), (F_2, \mathcal{B}_2), \ldots, (F_n, \mathcal{B}_n)$$

respectivement. *Il existe une unique mesure* $\geq 0$ *$\sigma$-finie* $\mu$ sur l'espace produit

$$\left(F = \prod_{i=1}^{n} F_i, \mathcal{B} = \bigotimes_{i=1}^{n} \mathcal{B}_i\right)$$

telle que pour tous $A_1 \in \mathcal{B}_1$, $A_2 \in \mathcal{B}_2$, ..., $A_n \in \mathcal{B}_n$ :

$$\mu(A_1 \times A_2 \times \ldots \times A_n) = \mu_1(A_1)\mu_2(A_2)\ldots\mu_n(A_n)$$

La mesure $\mu$ s'appelle la *mesure produit* de $\mu_1, \mu_2, \ldots, \mu_n$ et se note :

$$\mu = \mu_1 \otimes \mu_2 \otimes \ldots \otimes \mu_n = \bigotimes_{i=1}^{n} \mu_i$$

Par exemple, si $\lambda_n$ désigne la mesure de Lebesgue sur $\mathbb{R}^n$ ($n \in \mathbb{N}^*$), on a

$$\lambda_n = \lambda_1 \otimes \lambda_1 \otimes \ldots \otimes \lambda_1 \qquad (n \text{ fois})$$

Remarquer, dans le théorème 1, que si les $\mu_i$, $(i = 1, \ldots, n)$ sont des probabilités, alors $\mu = \bigotimes_{i=1}^{n} \mu_i$ est une probabilité.

### 2.1.2 Théorème 2 (de Fubini)

Cas $n = 2$. Soit $f$ une application mesurable de $(F_1 \times F_2, \mathcal{B}_1 \otimes \mathcal{B}_2, \mu_1 \otimes \mu_2)$ dans $\overline{\mathbb{R}}$ alors :

1. Pour tout $x = (x_1, x_2) \in F_1 \times F_2$, les applications $f_{x_1}$ de $(F_2, \mathcal{B}_2)$ dans $\overline{\mathbb{R}}$ et $f_{x_2}$ de $(F_1, \mathcal{B}_1)$ dans $\overline{\mathbb{R}}$ définies par
$$f_{x_1}(y_2) = f(x_1, y_2) \qquad (y_2 \in F_2)$$
$$f_{x_2}(y_1) = f(y_1, x_2) \qquad (y_1 \in F_1)$$
sont mesurables.
2. Si $f$ est positive, les applications :
$$x_1 \in F_1 \longmapsto \int_{F_2} f(x_1, x_2)\mu_2(\mathrm{d}x_2)$$
$$x_2 \in F_2 \longmapsto \int_{F_1} f(x_1, x_2)\mu_1(\mathrm{d}x_1)$$
sont mesurables sur $(F_1, \mathcal{B}_1)$ et $(F_2, \mathcal{B}_2)$ respectivement, de plus :
$$\begin{aligned}\int_{F_1 \times F_2} f(x_1, x_2)\mu_1 \otimes \mu_2(\mathrm{d}x_1, \mathrm{d}x_2) &= \int_{F_1} \mu_1(\mathrm{d}x_1) \int_{F_2} f(x_1, x_2)\mu_2(\mathrm{d}x_2) \\ &= \int_{F_2} \mu_2(\mathrm{d}x_2) \int_{F_1} f(x_1, x_2)\mu_1(\mathrm{d}x_1) \quad (*)\end{aligned}$$
3. Si $f \in \mathcal{L}^1(F_1 \times F_2, \mathcal{B}_1 \otimes \mathcal{B}_2, \mu_1 \otimes \mu_2)$ alors
pour $\mu_1$ presque tout $x_1$ : $f_{x_1} \in \mathcal{L}^1(F_2, \mathcal{B}_2, \mu_2)$
pour $\mu_2$ presque tout $x_2$ : $f_{x_2} \in \mathcal{L}^1(F_1, \mathcal{B}_1, \mu_1)$
et l'intégrale de $f$ par rapport à $\mu_1 \otimes \mu_2$ est encore donnée par l'égalité (*).
Ce théorème s'étend facilement, par récurrence, au cas d'un produit fini d'espaces $(F_i, \mathcal{B}_i, \mu_i)$, $(i = 1, 2, \ldots, n)$.

La démonstration des théorèmes 1 et 2 est basée sur le théorème 3 suivant (appelé « *théorème de classe monotone* »).

### 2.1.3 Définition 1

Soient $\Omega$ un ensemble, $\mathcal{C}$ et $\mathcal{S}$ deux sous-ensembles de $\mathcal{P}(\Omega)$. On dira que $\mathcal{C}$ est un $\pi$-*système* si, pour tous $A$ et $B$ appartenant à $\mathcal{C}$, on a $A \cap B \in \mathcal{C}$.
$\mathcal{S}$ est un $\lambda$-*système* si,

1. pour toute suite $(A_n)$ d'éléments de $\mathcal{S}$ telle que $A_n \subseteq A_{n+1}, \forall n$, on a $\underset{n\in\mathbb{N}}{\cup} A_n \in \mathcal{S}$
2. pour tous $A$ et $B$ appartenant à $\mathcal{S}$, on a : $A \supseteq B \Rightarrow A \setminus B \in \mathcal{S}$

### 2.1.4 Théorème 3

1. Si $\mathcal{S}$ est un $\lambda$-système contenant un $\pi$-système $\mathcal{C}$ et si $\Omega \in \mathcal{S}$, alors $\mathcal{S}$ contient $\sigma(\mathcal{C})$ (la tribu engendrée par $\mathcal{C}$).
2. Soit $\mathcal{H}$ un espace vectoriel de fonctions numériques finies (resp. bornées) sur un espace mesurable $(\Omega, \mathcal{A})$ tel que :
   (a) $\mathbf{1} \in \mathcal{H}$ et $\mathbf{1}_A \in \mathcal{H}$, pour tout $A \in \mathcal{C}$ où $\mathcal{C}$ est un $\pi$-système tel que $\sigma(\mathcal{C}) = \mathcal{A}$
   (b) $\mathcal{H}$ est stable par passage à la limite croissante (i. e. : si $(f_n)$ est une suite d'éléments de $\mathcal{H}$ telle que $f_n \leq f_{n+1}$ et $f_n \underset{n\to\infty}{\nearrow} f$ finie (resp. bornée), on a $f \in \mathcal{H}$) alors $\mathcal{H}$ contient toutes les fonctions numériques $\mathcal{A}$-mesurables finies (resp. bornées).

Démonstration :

1. Soit $\mathcal{S}'$ le $\lambda$-système engendré par $\mathcal{C}$ et $\Omega$. $\mathcal{S}'$ est l'intersection de tous les $\lambda$-systèmes contenant $\mathcal{C}$ et $\Omega$. On a : $\mathcal{S}' \subseteq \mathcal{S}$ et $\mathcal{C} \subseteq \mathcal{S}' \subseteq \sigma(\mathcal{C})$. Si on montre que $\mathcal{S}'$ est stable par intersection finie, il est facile de voir que $\mathcal{S}'$ est, en fait, une tribu et, puisque $\mathcal{C} \subseteq \mathcal{S}'$, $\mathcal{S}' = \sigma(\mathcal{C}) \subseteq \mathcal{S}$. Soit
$$\mathcal{S}_1 = \{A \in \mathcal{S}' \quad \text{tel que} \quad A \cap B \in \mathcal{S}', \quad \forall\, B \in \mathcal{C}\}$$
alors $\mathcal{S}_1$ est un $\lambda$-système contenant $\mathcal{C}$ et $\Omega \in \mathcal{S}_1$ donc $\mathcal{S}_1 \subseteq \mathcal{S}' \Rightarrow \mathcal{S}_1 = \mathcal{S}'$.
Soit
$$\mathcal{S}_2 = \{A \in \mathcal{S}' \quad \text{tel que} \quad A \cap B \in \mathcal{S}', \quad \forall\, B \in \mathcal{S}'\}$$
$\mathcal{S}_2$ est un $\lambda$-système contenant $\mathcal{C}$ et $\Omega \in \mathcal{S}_2$ donc $\mathcal{S}_2 = \mathcal{S}'$, ce qui signifie que, pour tous $A$ et $B$ dans $\mathcal{S}'$, on a $A \cap B \in \mathcal{S}'$
2. Soit
$$\mathcal{S} = \{A \in \mathcal{A} \quad \text{tel que} \quad \mathbf{1}_A \in \mathcal{H}\} \quad \subseteq \mathcal{A}$$
Les hypothèses sur $\mathcal{H}$ montrent que $\mathcal{S}$ est un $\lambda$-système contenant $\mathcal{C}$ et $\Omega$, donc $\mathcal{S}$ contient $\sigma(\mathcal{C}) = \mathcal{A}$ (d'après 1)) et $\mathcal{S} = \mathcal{A}$.
L'espace vectoriel $\mathcal{H}$ contient donc toutes les fonctions étagées. Soit $f$ une fonction $\mathcal{A}$-mesurable finie (resp. bornée), on a $f = f^+ - f^-$ où $f^+$ et $f^-$ sont des fonctions mesurables positives qui sont limites de suites croissantes de fonctions étagées positives, donc $f^+ \in \mathcal{H}$, $f^- \in \mathcal{H}$ et $f = f^+ - f^- \in \mathcal{H}$.

Corollaire 1 :
Soient $\mu_1$ et $\mu_2$ deux mesures positives $\sigma$-finies sur $(\Omega, \mathcal{A})$ et $\mathcal{C} \subseteq \mathcal{A}$ un $\pi$-système tel que $\sigma(\mathcal{C}) = \mathcal{A}$. On suppose que :

1. Pour tout $C \in \mathcal{C}$ : $\mu_1(C) = \mu_2(C)$
2. Il existe une suite croissante $(C_n)$ d'éléments de $\mathcal{C}$ telle que $\Omega = \underset{n}{\cup} C_n$ et $\mu_i(C_n) < \infty$, pour tout $n \in \mathbb{N}$.

Alors $\mu_1 = \mu_2$.
Démonstration :
Soit $n \in \mathbb{N}$ fixé. Considérons :

$$\mathcal{S} = \{A \in \mathcal{a} \quad \text{tel que} \quad \mu_1(A \cap C_n) = \mu_2(A \cap C_n)\} \quad \subseteq \mathcal{a}$$

On voit que $\mathcal{S}$ est un $\lambda$-système, $\mathcal{S} \supseteq \mathcal{C}$ et $\Omega \in \mathcal{S}$, donc $\mathcal{S} \supseteq \sigma(\mathcal{C}) = \mathcal{a}$ et $\mathcal{S} = \mathcal{a}$.
On en déduit que pour tout $A \in \mathcal{a}$ et tout $n \in \mathbb{N}$

$$\mu_1(A \cap C_n) = \mu_2(A \cap C_n)$$

donc

$$\mu_1(A) = \lim_{n\to\infty} \mu_1(A \cap C_n) = \lim_{n\to\infty} \mu_2(A \cap C_n) = \mu_2(A)$$

### 2.1.5 Démonstration du théorème 1

Unicité :
Soient $\mu$ et $\tilde{\mu}$ deux mesures $\sigma$-finies sur $(F = \prod_{i=1}^n F_i, \mathcal{B} = \bigotimes_{i=1}^n \mathcal{B}_i)$ telles que, pour tous $A_1 \in \mathcal{B}_1, \ldots, A_n \in \mathcal{B}_n$

$$\mu(A_1 \times \ldots \times A_n) = \tilde{\mu}(A_1 \times \ldots \times A_n) = \prod_{i=1}^n \mu_i(A_i) \qquad (\mathcal{P})$$

On sait qu'il existe, pour tout $i \in \{1, 2, \ldots, n\}$, une suite croissante $(A_i^k)_{k\in\mathbb{N}}$ d'éléments de $\mathcal{B}_i$ telle que $\underset{k}{\cup} A_i^k = F_i$ et $\mu_i(A_i^k) < \infty$, pour tout $k \in \mathbb{N}$. Posons :

$$C_k = A_1^k \times A_2^k \times \ldots \times A_n^k$$

alors, pour tout $k$ :

$$\mu(C_k) = \tilde{\mu}(C_k) \ < \ \infty, \quad C_k \subseteq C_{k+1} \quad \text{et} \quad \underset{k}{\cup} C_k = F$$

Soit $\mathcal{C}$ la famille des pavés $C$ de la forme

$$C = A_1 \times A_2 \times \ldots \times A_n \quad \text{où} \quad A_i \in \mathcal{B}_i$$

$\mathcal{C}$ est un $\pi$-système tel que $\sigma(\mathcal{C}) = \mathcal{B}$ et la propriété $(\mathcal{P})$ montre que $\mu$ et $\tilde{\mu}$ coïncident sur $\mathcal{C}$, de plus $C_k \in \mathcal{C}$ pour tout $k$, donc $\mu = \tilde{\mu}$ (d'après le corollaire 1).
Existence :
Il suffit de se restreindre au cas $n = 2$. Pour tout $A \in \mathcal{B} = \mathcal{B}_1 \otimes \mathcal{B}_2$ et pour $x = (x_1, x_2) \in F_1 \times F_2$, considérons :

$$A_{x_1} = \{x_2 \in F_2 \quad \text{tel que} \quad (x_1, x_2) \in A\} \quad \subseteq F_2$$

$$A_{x_2} = \{x_1 \in F_1 \quad \text{tel que} \quad (x_1, x_2) \in A\} \quad \subseteq F_1$$

Remarquer que :

$$\mathbf{1}_{A_{x_1}}(x_2) = \mathbf{1}_{A_{x_2}}(x_1) = \mathbf{1}_A(x_1, x_2)$$

Considérons l'ensemble $\mathcal{S}$ défini par :
$\mathcal{S} = \{A \in \mathcal{B}$ vérifiant les propriétés suivantes :

1. pour tout $x = (x_1, x_2) \in F$ : $A_{x_1} \in \mathcal{B}_2$ et $A_{x_2} \in \mathcal{B}_1$
2. les applications $x_1 \in F_1 \longmapsto \mu_2(A_{x_1})$ et $x_2 \in F_2 \longmapsto \mu_1(A_{x_2})$ sont mesurables sur $(F_1, \mathcal{B}_1)$ et $(F_2, \mathcal{B}_2)$ respectivement
3. $$\int_{F_1} \mu_1(\mathrm{d}x_1)\mu_2(A_{x_1}) = \int_{F_2} \mu_2(\mathrm{d}x_2)\mu_1(A_{x_2})\}$$

On se propose de montrer que $\mathcal{S} = \mathcal{B}$.
Supposons d'abord que $\mu_1$ et $\mu_2$ sont bornées. On vérifie, dans ce cas, facilement que $\mathcal{S}$ est un $\lambda$-système contenant tous les pavés de la forme $C = A_1 \times A_2$ où $A_1 \in \mathcal{B}_1$ et $A_2 \in \mathcal{B}_2$, en particulier$F = F_1 \times F_2 \in \mathcal{S}$ Donc $\mathcal{S} = \mathcal{B} = \mathcal{B}_1 \otimes \mathcal{B}_2$ (théorème 3). Dans le cas général, on sait qu'il existe deux suites *croissantes* $(A_i^k)_{k\in\mathbb{N}}$, $(i = 1, 2)$ telles que $A_i^k \in \mathcal{B}_i$, $\bigcup\limits_k A_i^k = F_i$ et $\mu_i(A_i^k) < \infty$ pour tout $k \in \mathbb{N}$.
Posons alors, pour tous $B_1 \in \mathcal{B}_1$ et $B_2 \in \mathcal{B}_2$ :

$$\mu_1^k(B_1) = \mu_1(B_1 \cap A_1^k), \qquad \mu_2^k(B_2) = \mu_2(B_2 \cap A_2^k)$$

On voit que, pour tout $k$, $\mu_1^k$ et $\mu_2^k$ sont deux mesures bornées et que les propriétés 2) et 3) sont satisfaites pour tout $A \in \mathcal{B}$ lorsque l'on remplace $\mu_1$ par $\mu_1^k$ et $\mu_2$ par $\mu_2^k$ ; en faisant tendre $k$ vers l'infini, on vérifie alors, sans peine, que $\mathcal{S} = \mathcal{B} = \mathcal{B}_1 \otimes \mathcal{B}_2$. Posons ensuite, pour tout $A \in \mathcal{B}$ :

$$\mu(A) = \int_{F_1} \mu_1(\mathrm{d}x_1)\mu_2(A_{x_1}) = \int_{F_2} \mu_2(\mathrm{d}x_2)\mu_1(A_{x_2})$$

Il est immédiat de vérifier que $\mu$ est une mesure $\sigma$-finie sur $\mathcal{B}$ telle que

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

pour tous $A_1 \in \mathcal{B}_1$, $A_2 \in \mathcal{B}_2$

### 2.1.6 Démonstration du théorème 2

On désigne par $\mathcal{H}$ l'ensemble suivant :

$$\begin{aligned}\mathcal{H} = \{\ f : (F_1 \times F_2, \mathcal{B}_1 \otimes \mathcal{B}_2) \to \overline{\mathbb{R}}_+ = [0, \infty],\\ \text{mesurables vérifiant les assertions 1) et 2) du théorème 2. }\}\end{aligned}$$

On voit facilement que si $f$ et $g$ sont deux éléments de $\mathcal{H}$ alors $f + g$ appartient à $\mathcal{H}$ et $\alpha.f \in \mathcal{H}$ $(\alpha \in \mathbb{R}_+)$ et que si $0 \leq f_n \leq f_{n+1} \leq \ldots$ est une suite croissante d'éléments $f_n$ de $\mathcal{H}$ alors $f_\infty = \lim_n \nearrow f_n$ appartient à $\mathcal{H}$. La démonstration du théorème 1 montre, de plus, que $\mathcal{H}$ contient les fonctions indicatrices $\mathbf{1}_A$, pour tout $A \in \mathcal{B}_1 \otimes \mathcal{B}_2$ donc $\mathcal{H}$ contient les variables aléatoires réelles étagées positives et, par passage à la limite croissante, *toutes* les variables aléatoires réelles positives. Soit

$$f \in \mathcal{L}^1(F_1 \times F_2, \mathcal{B}_1 \otimes \mathcal{B}_2, \mu_1 \otimes \mu_2)$$

alors, d'après 1) et 2), on voit que :

$$\int_{F_1} \mu_1(\mathrm{d}x_1) \int_{F_2} |f|(x_1, x_2)\mu_2(\mathrm{d}x_2) = \int_{F_2} \mu_2(\mathrm{d}x_2) \int_{F_1} |f|(x_1, x_2)\mu_1(\mathrm{d}x_1) < \infty$$

donc, pour $\mu_1$ presque tout $x_1$ :

$$\int_{F_2} |f|(x_1,x_2)\mu_2(\mathrm{d}x_2) < \infty$$

et, pour $\mu_2$ presque tout $x_2$ :

$$\int_{F_1} |f|(x_1,x_2)\mu_1(\mathrm{d}x_1) < \infty$$

En décomposant $f$ sous la forme $f = f^+ - f^-$ où $f^+ = \max(f,0) \geq 0$ et $f^- = \max(-f,0) \geq 0$, on voit alors, par linéarité, que l'intégrale de $f$ par rapport à $\mu_1 \otimes \mu_2$ est encore donnée par l'égalité :

$$\begin{aligned}\int_{F_1\times F_2} f\mathrm{d}\mu_1\otimes\mu_2 &= \int_{F_1}\mu_1(\mathrm{d}x_1)\int_{F_2} f(x_1,x_2)\mu_2(\mathrm{d}x_2)\\ &= \int_{F_2}\mu_2(\mathrm{d}x_2)\int_{F_1} f(x_1,x_2)\mu_1(\mathrm{d}x_1) \quad (*)\end{aligned}$$

L'extension du théorème 2 au cas d'un produit fini d'espaces se démontre facilement par récurrence.
Soit $(F_i,\mathcal{B}_i)_{i\in I}$ une famille d'espaces mesurables où $I \neq \emptyset$ est un ensemble quelconque d'indices (par exemple $I = \mathbb{N}$ ou $\mathbb{R}_+$ ou $\mathbb{R}\ldots$) L'espace produit $F = \prod_{i\in I} F_i$ est, par définition, l'ensemble des applications $x$ de $I$ dans $\underset{i\in I}{\cup} F_i$ telles que, pour tout $i \in I$, $x_i \in F_i$ ; on posera $x = (x_i)_{i\in I}$. Les applications coordonnées $\theta_i$ de $F$ dans $F_i$ $(i \in I)$ sont alors définies par $\theta_i(x) = x_i$, pour tout $x = (x_i)_{i\in I}$ dans $F$. On munit l'espace $F = \prod_{i\in I} F_i$ de la tribu produit $\mathcal{B} = \bigotimes_{i\in I}\mathcal{B}_i$ qui est donnée par

$$\mathcal{B} = \sigma\{\theta_i^{-1}(B_i), \quad i\in I, \quad B_i \in \mathcal{B}_i\}$$

On vérifie facilement que $\mathcal{B}$ est la plus petite tribu sur $F$ qui rend mesurables toutes les applications $\theta_i$ $(i \in I)$. Soit $\mathcal{C}$ l'ensemble des pavés $C$ de $F$ de la forme $C = \underset{i\in J}{\cap} \theta_i^{-1}(B_i)$ où $J$ est fini, $J \subseteq I$ et $B_i \in \mathcal{B}_i$. Remarquer que $\mathcal{C}$ est un $\pi$-système tel que $\sigma(C) = \mathcal{B}$.

### 2.1.7 Proposition 1

Soit $(\Omega, \mathcal{a})$ un espace mesurable. Une application $X$ de $(\Omega, \mathcal{a})$ dans $(F = \prod_{i\in I} F_i,\ \mathcal{B} = \bigotimes_{i\in I}\mathcal{B}_i)$ est mesurable si et seulement si :

$$(\mathcal{P}) \quad \forall\ i\in I, \quad \theta_i \circ X \text{ est mesurable de } (\Omega, \mathcal{a}) \text{ dans } (F_i,\mathcal{B}_i)$$

Démonstration :
Puisque $\mathcal{B}$ est engendrée par la famille des $\theta_i^{-1}(B_i)$, $i \in I$, $B_i \in \mathcal{B}_i$, on voit que :

$$\begin{aligned} X \text{ est mesurable} &\iff \text{pour tous } i\in I \text{ et } B_i\in\mathcal{B}_i \quad X^{-1}\left(\theta_i^{-1}(B_i)\right) \in \mathcal{a}\\ &\iff \text{pour tout } i\in I \quad \theta_i\circ X \text{ est mesurable}\end{aligned}$$

car $X^{-1}(\theta_i^{-1}(B_i)) = (\theta_i \circ X)^{-1}(B_i)$.
Nous admettrons le théorème suivant, dû à Kolmogorov, qui assure l'existence, sous certaines conditions, d'un produit infini d'espaces de probabilités.

### 2.1.8 Théorème 4

Pour chaque $i \in I$, soit $\mu_x$ une *probabilité* sur l'espace mesurable $(F_i, \mathcal{B}_i)$ ; on suppose que $F_i = \mathbb{R}^{d_i}$, $(d_i \in \mathbb{N}^*)$ et que $\mathcal{B}_i = \mathcal{B}_{\mathbb{R}^{d_i}}$ (tribu borélienne de $\mathbb{R}^{d_i}$). *Il existe alors une unique probabilité* $\mu$ sur l'espace produit

$$\left( F = \prod_{i \in I} F_i, \quad \mathcal{B} = \bigotimes_{i \in I} \mathcal{B}_i \right)$$

telle que pour tout pavé $C = \underset{i \in J}{\cap} \theta_i^{-1}(B_i)$ où $J$ est fini, $J \subseteq I$ et $B_i \in \mathcal{B}_i$ on a

$$\mu(C) = \prod_{i \in J} \mu_i(B_i) \quad (*)$$

La probabilité $\mu$ est appelée *probabilité produit* des $\mu_i$ $(i \in I)$ et notée $\mu = \bigotimes_{i \in I} \mu_i$. Le théorème 4 est une extension directe du théorème 1 au cas d'un produit quelconque d'espaces. Remarquer que l'égalité (*) entraîne immédiatement l'unicité de $\mu$ puisque la famille $\mathcal{C}$ des pavés est un $\pi$-système qui engendre la tribu produit. Noter aussi que le théorème 4 reste valable si on suppose que, pour tout $i$, $F_i$ est un ensemble dénombrable et $\mathcal{B}_i = \mathcal{P}_{(F_i)}$.

## 2.2 Indépendance

### 2.2.1 Définition 1

Soient $X_1, X_2, \ldots, X_n$ $n$ variables aléatoires définies sur un espace de probablité $(\Omega, \mathcal{A}, P)$ et à valeurs dans des espaces mesurables $(F_1, \mathcal{B}_1)$, $(F_2, \mathcal{B}_2)$, ..., $(F_n, \mathcal{B}_n)$ respectivement. On dira que *la famille* $(X_i)_{i=1,2,\ldots,n}$ *est indépendante* (ou que les $X_i, i = 1, 2, \ldots, n$, sont indépendantes) si :
pour tous $B_1 \in \mathcal{B}_1$, $B_2 \in \mathcal{B}_2$, ..., $B_n \in \mathcal{B}_n$, on a

$$P\{X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n\} = \prod_{i=1}^{n} P(X_i \in B_i) \quad (*)$$

Soit $X$ l'application de $(\Omega, \mathcal{A}, P)$ dans l'espace produit

$$\left( F = \prod_{i=1}^{n} F_i, \quad \mathcal{B} = \bigotimes_{i=1}^{n} \mathcal{B}_i \right)$$

définie par :

$$X(\omega) = (X_1(\omega), \ldots, X_n(\omega)), \qquad \omega \in \Omega$$

On vérifie facilement que $X$ est bien une variable aléatoire. Notons $P_X$ la loi de $X$ sur l'espace produit $(F, \mathcal{B})$ et $P_{X_i}$ la loi de $X_i$ sur $(F_i, \mathcal{B}_i)$, $i = 1, 2, \ldots, n$. L'égalité (*), dans la définition 1, montre que les variables aléatoires $X_i, i = 1, 2, \ldots, n$ sont indépendantes *si et seulement si* la loi de $X$ est égale au produit des lois des $X_i$ :

$$P_X = P_{X_1} \otimes P_{X_2} \otimes \ldots \otimes P_{X_n} \quad \text{sur} \quad \left( F = \prod_{i=1}^{n} F_i, \quad \mathcal{B} = \bigotimes_{i=1}^{n} \mathcal{B}_i \right) \qquad (**)$$

### 2.2.2 Proposition 1

Soient $X_i$, $(i = 1, 2, \ldots, n)$ des variables aléatoires indépendantes à valeurs dans $(F_i, \mathcal{B}_i)$ et $\varphi_i$ des applications mesurables de $(F_i, \mathcal{B}_i)$ dans $(G_i, \mathcal{T}_i)$, alors les variables aléatoires $Y_i = \varphi_i \circ X_i$ $(i = 1, 2, \ldots, n)$ à valeurs dans $(G_i, \mathcal{T}_i)$, sont aussi indépendantes.
Démonstration :
Pour tous $C_1 \in \mathcal{T}_1$, $C_2 \in \mathcal{T}_2$, ... , $C_n \in \mathcal{T}_n$, on a

$$\begin{aligned} P\{Y_1 \in C_1, \ldots, Y_n \in C_n\} &= P\{\varphi_1 \circ X_1 \in C_1, \ldots, \varphi_n \circ X_n \in C_n\} \\ &= P\{X_1 \in \varphi_1^{-1}(C_1), \ldots, X_n \in \varphi_n^{-1}(C_n)\} \\ &= \prod_{i=1}^{n} P\{X_i \in \varphi_i^{-1}(C_i)\} \\ &= \prod_{i=1}^{n} P\{Y_i \in C_i\} \end{aligned}$$

donc $Y_1, Y_2, \ldots, Y_n$ sont indépendantes.

### 2.2.3 Proposition 2

Soient $X_i : (\Omega, \mathcal{A}, P) \to (F_i, \mathcal{B}_i), i = 1, 2, \ldots, n$ $n$ variables aléatoires. Pour tout $i \in \{1, 2, \ldots, n\}$, soit $\mathcal{C}_i$ un $\pi$-système contenu dans $\mathcal{B}_i$ et engendrant $\mathcal{B}_i$ $(\sigma(\mathcal{C}_i) = \mathcal{B}_i)$. Alors $(X_1, \ldots, X_n)$ sont indépendantes si et seulement si :

$$(*)' \left\{ \begin{array}{l} \text{pour tous } C_1 \in \mathcal{C}_1, C_2 \in \mathcal{C}_2, \ldots, C_n \in \mathcal{C}_n, \text{ on a :} \\ P\{X_1 \in C_1, X_2 \in C_2, \ldots, X_n \in C_n\} = \prod_{i=1}^n P\{X_i \in C_i\} \end{array} \right.$$

Démonstration :
La propriété $(*)'$ signifie que pour tout $C = C_1 \times C_2 \times \ldots \times C_n$ appartenant à $\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2 \times \ldots \times \mathcal{C}_n$ :

$$P_X(C) = P_{(X_1,\ldots,X_n)}(C_1 \times \ldots \times C_n) = \prod_{i=1}^{n} P_{X_i}(C_i) = P_{X_1} \otimes \ldots \otimes P_{X_n}(C)$$

où $P_X = P_{(X_1,\ldots,X_n)}$ est la loi de $X = (X_1, \ldots, X_n)$ sur $(\prod_{i=1}^n F_i, \bigotimes_{i\in I}^n \mathcal{B}_i)$. Or on vérifie facilement que $\mathcal{C}$ est un $\pi$-système qui engendre la tribu produit $\bigotimes_{i\in I}^n \mathcal{B}_i$, donc $P_X = P_{X_1} \otimes \ldots \otimes P_{X_n}$, d'après le corollaire 1 du théorème 3.
Par exemple, pour que $n$ variables aléatoires réelles $X_1, X_2, \ldots, X_n$ soient indépendantes, il suffit que :

$$\left\{ \begin{array}{l} \text{pour tout } (t_1, t_2, \ldots, t_n) \in \mathbb{R}^n, \\ P\{X_1 \leq t_1, X_2 \leq t_2, \ldots, X_n \leq t_n\} = \prod_{i=1}^n P\{X_i \leq t_i\} \end{array} \right.$$

### 2.2.4 Proposition 3

Soient $X_1, X_2, \ldots, X_n$ n variables aléatoires réelles indépendantes et intégrables. Alors $X = X_1.X_2 \ldots X_n$ (variable aléatoire produit) est intégrable, de plus :

$$E(X) = \prod_{i=1}^{n} E(X_i)$$

Démonstration :
On a :

$$\begin{aligned}
E(|X|) &= E(|X_1|.|X_2|\dots|X_n|) \\
&= \int_{\mathbb{R}^n} (|x_1|.|x_2|\dots|x_n|)(P_{X_1}\otimes\dots\otimes P_{X_n})(\mathrm{d}x_1,\mathrm{d}x_2,\dots,\mathrm{d}x_n) \\
&= \prod_{i=1}^{n}\left(\int_{\mathbb{R}} |x_i| P_{X_i}(\mathrm{d}x_i)\right) \quad \text{d'après le théorème de Fubini} \\
&= \prod_{i=1}^{n} E(|X_i|) \ < \ \infty
\end{aligned}$$

Donc $X \in \mathcal{L}^1$ et le même calcul montre que :

$$E(X) = \prod_{i=1}^{n} E(X_i)$$

Remarque :
Si $X_1$ et $X_2$ sont deux variables aléatoires réelles intégrables, le produit $X = X_1.X_2$ n'est pas intégrable en général (sans hypothèse d'indépendance). Par exemple, sur $(\Omega, \mathcal{A}, P) = ([0,1], \mathcal{B}_{[0,1]}, \lambda)$ où $\lambda$ est la mesure de Lebesgue restreinte à $[0,1]$, soient :

$$X_1 = \frac{1}{\sqrt{x}}.\mathbf{1}_{x>0} \quad \text{et} \quad X_2 = X_1$$

On voit que :

$$E(X_1) = \int_{0^+}^{1} \frac{\mathrm{d}x}{\sqrt{x}} = [2\sqrt{x}]_0^1 = 2 = E(X_2)$$

mais

$$E(X_1.X_2) = E(X_1^2) = \int_{0^+}^{1} \frac{\mathrm{d}x}{x} = +\infty$$

### 2.2.5 Proposition 4

Soient $X_1, X_2, \dots, X_n$ n variables aléatoires à valeurs dans $(F_1, \mathcal{B}_1)$, $(F_2, \mathcal{B}_2)$, ..., $(F_n, \mathcal{B}_n)$ respectivement. Alors $(X_1, X_2, \dots, X_n)$ sont indépendantes si et seulement si :
(1) pour tout $i \in \{1, 2, \dots, n\}$ et toute application $\varphi_i$ de $(F_i, \mathcal{B}_i)$ dans $\mathbb{R}$ mesurable et bornée on a :

$$E\{\varphi_1(X_1)\dots\varphi_n(X_n)\} = \prod_{i=1}^{n} E\{\varphi_i(X_i)\}$$

Démonstration :
Si $(X_1, \dots, X_n)$ sont indépendantes alors la propriété (1) est satisfaite d'après la proposition 3. Réciproquement, (1) entraîne que pour tous $B_1 \in \mathcal{B}_1$, $B_2 \in \mathcal{B}_2$, ..., $B_n \in \mathcal{B}_n$ :

$$E\{\mathbf{1}_{B_1}(X_1)\mathbf{1}_{B_2}(X_2)\dots\mathbf{1}_{B_n}(X_n)\} = \prod_{i=1}^{n} E\{\mathbf{1}_{B_i}(X_i)\}$$

ou bien que

$$P\{X_1 \in B_1, X_2 \in B_2, \dots X_n \in B_n\} = \prod_{i=1}^{n} P\{X_i \in B_i)\}$$

### 2.2.6 Définition 2

Soit $X$ une variable aléatoire à valeurs dans un espace mesurable $(F, \mathcal{B})$ muni d'une mesure $\mu$ positive $\sigma$-finie et soit $f$ une application mesurable et positive de $(F, \mathcal{B})$ dans $\mathbb{R}_+$. On dira que $X$ *a pour densité $f$ par rapport à $\mu$* si (R), pour tout $B \in \mathcal{B}$ :

$$P(X \in B) = \int_B f(x)\mu(\mathrm{d}x)$$

en particulier

$$P(X \in F) = 1 = \int_F f(x)\mu(\mathrm{d}x)$$

La propriété (R) est satisfaite *si et seulement si*, pour toute application $\varphi$ de $(F, \mathcal{B})$ dans $\mathbb{R}$ mesurable et bornée (ou positive) :

$$E\{\varphi(X)\} = \int_F \varphi(x) f(x)\mu(\mathrm{d}x) = \int_F \varphi(x) P_X(\mathrm{d}x) \quad \text{(R')}$$

où $P_X(\mathrm{d}x) = P_X$ est la loi de $X$. On notera alors :

$$P_X(\mathrm{d}x) = f(x)\mu(\mathrm{d}x)$$

ou bien

$$\frac{\mathrm{d}P_X}{\mathrm{d}\mu} = f$$

ou bien

$$\mathrm{d}P_X = f\mathrm{d}\mu$$

L'équivalence de (R) et (R') se démontre par linéarité et continuité en considérant, d'abord, le cas où $\varphi$ est une variable aléatoire étagée positive.
Si $\mathrm{d}P_X = f\mathrm{d}\mu$, on voit que, pour toute application mesurable $\varphi$ de $(F, \mathcal{B})$ dans $\overline{\mathbb{R}}$, la variable aléatoire réelle $\varphi(X)$ est intégrable si et seulement si :

$$\int_F |\varphi(x)|.f(x)\mu(\mathrm{d}x) < \infty$$

et, dans ce cas,

$$E(\varphi(X)) = \int_F \varphi(x) f(x)\mu(\mathrm{d}x) < \infty$$

Lorsque $(F, \mathcal{B}) = (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ est muni de la mesure de Lebesgue $\lambda(\mathrm{d}x)$, on écrira simplement, $P_X(\mathrm{d}x) = f(x)\mathrm{d}x$ et on dira que $X$ a pour densité $f$. Lorsque $F$ est un ensemble dénombrable, muni de la mesure de comptage $\mu$ sur $\mathcal{P}(F)$, on voit que toute variable aléatoire $X$ à valeurs dans $F$ a une densité $f$ par rapport à $\mu$ donnée par

$$f(x) = P(X = x) \quad \text{où} \quad x \in F$$

La densité discrète $f$ s'identifie à la loi $P_X$ de $X$.

### 2.2.7 Proposition 5

Soient $X_1, X_2, \dots, X_n$ n variables aléatoires à valeurs dans $(F_1, \mathcal{B}_1)$, $(F_2, \mathcal{B}_2)$, …, $(F_n, \mathcal{B}_n)$ respectivement. On suppose que, pour tout $i \in \{1, 2, \dots, n\}$ :

$$\mathrm{d}P_{X_i} = f_i \mathrm{d}\mu_i$$

où $P_{X_i}$ est la loi de $X_i$ et $\mu_i$ est une mesure positive $\sigma$-finie sur $(F_i, \mathcal{B}_i)$. Alors $X_1, X_2, \dots, X_n$ sont indépendantes *si et seulement si* la loi $P_X$ de

$$X = (X_1, X_2, \dots, X_n) \quad \text{sur} \quad \left( F = \prod_{i=1}^{n} F_i,\ \mathcal{B} = \bigotimes_{i=1}^{n} \mathcal{B}_i \right)$$

$$\text{vérifie} \quad \mathrm{d}P_X = (f_1 \otimes f_2 \otimes \dots \otimes f_n)\mathrm{d}\mu$$

$$\text{où} \quad \mu = \mu_1 \otimes \dots \otimes \mu_n$$

et $f_1 \otimes f_2 \otimes \dots \otimes f_n$ est définie par :

$$(f_1 \otimes f_2 \otimes \dots \otimes f_n)(x_1, x_2, \dots, x_n) = f_1(x_1) f_2(x_2) \dots f_n(x_n)$$

pour tout $(x_1, x_2, \dots, x_n) \in F_1 \times F_2 \times \dots \times F_n$.

Démonstration :

On a, si $\mathrm{d}P_{X_i} = f_i \mathrm{d}\mu_i$ $(i = 1, 2, \dots, n)$ :

$$\{X_1, X_2, \dots, X_n\} \quad \text{indépendantes} \quad \Longleftrightarrow$$

pour tous $B_1 \in \mathcal{B}_1, \dots, B_n \in \mathcal{B}_n$,

$$\begin{aligned} P\{X_1 \in \mathcal{B}_1, \dots, X_n \in \mathcal{B}_n\} &= \prod_{i=1}^{n} P(X_i \in \mathcal{B}_i) \\ &= \prod_{i=1}^{n} \int_{\mathcal{B}_i} f_i(x_i) \mu_i(\mathrm{d}x_i) \\ &= \int_{\mathcal{B}_1 \times \dots \times \mathcal{B}_n} f_1(x_1) \dots f_n(x_n) (\mu_1 \otimes \dots \otimes \mu_n)(\mathrm{d}x_1, \dots, \mathrm{d}x_n) \end{aligned}$$

$$\Longleftrightarrow$$

$$P_{(X_1, X_2, \dots, X_n)}(\mathrm{d}x_1, \mathrm{d}x_2, \dots, \mathrm{d}x_n) = f_1(x_1) f_2(x_2) \dots f_n(x_n) (\mu_1 \otimes \dots \otimes \mu_n)(\mathrm{d}x_1, \dots, \mathrm{d}x_n)$$

$$\text{sur} \quad \left( F = \prod_{i=1}^{n} F_i, \mathcal{B} = \bigotimes_{i=1}^{n} \mathcal{B}_i \right)$$

### 2.2.8 Définition 3

Soient $\mu$ et $\nu$ deux probabilités sur $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. Pour tout $A \in \mathcal{B}_{\mathbb{R}^d}$, posons :

$$\mu * \nu(A) = \iint \mathbf{1}_A(x+y) \mu(\mathrm{d}x) \nu(\mathrm{d}y) \quad (1)$$

On définit ainsi une probabilité sur $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, notée $\mu * \nu$ et appelée *produit de convolution de $\mu$ et $\nu$*. Noter que la définition de $\mu * \nu$ entraîne que, pour toute application $\varphi$ de $\mathbb{R}^d$ dans $\mathbb{R}$ mesurable et positive ou bornée, on a :

$$\int \varphi \, \mathrm{d}(\mu * \nu) = \iint \varphi(x+y) \mu(\mathrm{d}x) \nu(\mathrm{d}y) \quad (2)$$

### 2.2.9 Proposition 6

Soient $X$ et $Y$ deux variables aléatoires indépendantes à valeurs $\mathbb{R}^d$, alors la loi de $X+Y$ est égale au produit de convolution des lois de $X$ et de $Y$ :

$$P_{X+Y} = P_X * P_Y$$

Si $X$ a une densité $f$ (par rapport à la mesure de Lebesgue) alors $X+Y$ admet une densité $\tilde{f}$ donnée par :

$$\tilde{f}(z) = \int_{\mathbb{R}^d} f(z-y) P_Y(\mathrm{d}y) \qquad (z \in \mathbb{R}^d)$$

Si, de plus, $Y$ a une densité $g$, on voit que la densité $\tilde{f}$ de $X+Y$ est égale au produit de convolution $f * g$ défini par la formule :

$$f * g(z) = \int_{\mathbb{R}^d} f(z-y) g(y)\,\mathrm{d}y \qquad (z \in \mathbb{R}^d)$$

<u>Démonstration</u> :
Soit $\varphi$ une application mesurable et positive de $\mathbb{R}^d$ dans $\mathbb{R}$, on a :

$$\begin{aligned} E\{\varphi(X+Y)\} &= \iint \varphi(x+y) P_{(X,Y)}(\mathrm{d}x, \mathrm{d}y) \\ &= \iint \varphi(x+y) P_X(\mathrm{d}x) P_Y(\mathrm{d}y) \\ &= \int \varphi(z)(P_X * P_Y)(\mathrm{d}z) \end{aligned}$$

donc :

$$P_{(X+Y)}(\mathrm{d}z) = P_X * P_Y(\mathrm{d}z)$$

Si

$$P_X(\mathrm{d}x) = f(x)(\mathrm{d}x)$$

on voit que :

$$\begin{aligned} E\{\varphi(X+Y)\} &= \iint \varphi(x+y) f(x)(\mathrm{d}x) P_Y(\mathrm{d}y) \\ &= \int P_Y(\mathrm{d}y) \left( \int \varphi(x+y) f(x)\,\mathrm{d}x \right) \\ &= \int P_Y(\mathrm{d}y) \left( \int \varphi(z) f(z-y)\,\mathrm{d}z \right) \\ &= \int \varphi(z) \left( \int f(z-y) P_Y(\mathrm{d}y) \right) \mathrm{d}z \\ &= \int \varphi(z) P_{(X+Y)}(\mathrm{d}z) \end{aligned}$$

donc

$$P_{X+Y}(\mathrm{d}z) = \left( \int f(z-y) P_Y(\mathrm{d}y) \right) \mathrm{d}z$$

Si, de plus :

$$P_Y(\mathrm{d}y) = g(y)\,\mathrm{d}y$$

on voit que :

$$\begin{aligned} P_{(X+Y)}(\mathrm{d}z) &= \left( \int f(z-y) g(y)\, \mathrm{d}y \right) \mathrm{d}z \\ &= (f * g)(z)\, \mathrm{d}z \end{aligned}$$

### 2.2.10 Définition 4

Soit $(X_i)_{i \in I}$ une famille *quelconque* de variables aléatoires définies sur $(\Omega, \mathcal{a}, P)$ où, pout tout $i \in I$, $X_i$ est à valeurs dans $(F_i, \mathcal{B}_i)$. On dira que $(X_i)_{i \in I}$ est *indépendante* (ou que les $X_i$, $i \in I$, sont indépendantes) si toute sous famille finie $(X_i)_{i \in J}$, ($J$ fini $\subseteq I$) est indépendante i.e. si *pour tout $J$ fini* $\subseteq I$, on a :

$$P_{(X_i)_{i \in J}} = \bigotimes_{i \in J} P_{X_i} \quad \text{sur} \quad \left( \prod_{i \in J} F_i, \bigotimes_{i \in J} \mathcal{B}_i \right)$$

ce qui signifie aussi que, pour tout $A_i \in \mathcal{B}_i$ $(i \in J)$ on a :

$$P\Big\{ \underset{i \in J}{\cap} (X_i \in A_i) \Big\} = \prod_{i \in J} P\{X_i \in A_i\}$$

### 2.2.11 Proposition 6

Soit $(X_i)_{i \in I}$ une famille indépendante de variables aléatoires et soit $(I_k)_{k \in K}$ une partition de l'ensemble d'indices $I$ :

$$I = \underset{k}{\cup} I_k \quad \text{et} \quad I_k \cap I_l = \emptyset \quad \text{si} \quad l \neq k$$

Posons, pout tout $k \in K$ : $Y_k = (X_i)_{i \in I_k}$ ; alors la famille de variables aléatoires $(Y_k)_{k \in K}$ est aussi indépendante.
Démonstration : laissée en exercice.
Soient $I = \{1, 2, \ldots, n\}$ et pour tout $i \in I$, une probabilité $\mu_i$, donnée à l'avance, sur l'espace mesurable $(F_i, \mathcal{B}_i)$. Comment construit-on une famille $(X_i)_{i \in I}$ de variables aléatoires indépendantes telle que, pour tout $i \in I$, $X_i$ soit à valeurs $F_i$, de loi $P_{X_i}$ égale à $\mu_i$ sur $\mathcal{B}_i$ ?
Posons :

$$\Omega = \prod_{i=1}^{n} F_i, \qquad \mathcal{a} = \bigotimes_{i=1}^{n} \mathcal{B}_i, \qquad P = \bigotimes_{i=1}^{n} \mu_i$$

Sur l'espace de probabilité $(\Omega, \mathcal{a}, P)$, on vérifie immédatement que les applications coordonnées $\theta_i$ de $\Omega$ dans $F_i$ définies par

$$\theta_i(x_1, \ldots, x_n) = x_i \quad \text{si} \quad (x_1, \ldots, x_n) \in \Omega$$

forment une famille $(\theta_i)_{i=1,\ldots,n}$ de variables aléatoires indépendantes telles que, pour tout $i$ : $P_{\theta_i} = \mu_i$. Lorsque l'ensemble d'indices $I$ est infini il suffit d'utiliser, de la même façon, le théorème 4 (de Kolmogorov) pour construire une famille $(\theta_i)_{i \in I}$ de variables aléatoires indépendantes où, pour tout $i$, la loi $\mu_i$ de $\theta_i$ est donnée sur l'espace $(F_i, \mathcal{B}_i)$.

# Chapitre 3

# Calculs de lois, fonctions caractéristiques et variables aléatoires gaussiennes

## 3.1 Généralités

Soit $X$ une variable aléatoire réelle définie sur $(\Omega, \mathcal{A}, P)$, de loi notée $P_X$ sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Si $X$ est intégrable ($X \in \mathcal{L}^1$) et $E(X) = 0$, on dira que X est *centrée.* Pour tout $n \in \mathbb{N}^*$, si $|X|^n$ est intégrable ($X \in \mathcal{L}^n$), on définit *le moment d'ordre* $n$ :

$$E(X^n) = \int_{\mathbb{R}} x^n P_X(\mathrm{d}x)$$

*le moment centré d'ordre* $n$ :

$$E\left((X - E(X))^n\right) = \int_{\mathbb{R}} (x - m)^n P_X(\mathrm{d}x) \quad \text{où} \quad m = \int x P_X(\mathrm{d}x)$$

Pour tout $1 \leq p < \infty$, on désignera par $\mathcal{L}^p$ l'ensemble des variables aléatoires réelles $X$ telles que $|X|^p \in \mathcal{L}^1$. En utilisant l'inégalité

$$(|a| + |b|)^p \leq 2^p \max(|a|^p, |b|^p) \leq 2^p(|a|^p + |b|^p)$$

où $a$ et $b \in \mathbb{R}$, on vérifie que $\mathcal{L}^p$ est un espace vectoriel sur $\mathbb{R}$. Remarquer, de plus, que $1 \leq p \leq q \Rightarrow \mathcal{L}^q \subseteq \mathcal{L}^p$. En effet, si $1 \leq p \leq q$, on voit facilement qu'il existe une constante $C > 0$ telle que, pour tout $x \in \mathbb{R}$ :

$$|x|^p \leq C + |x|^q$$

L'espace vectoriel $\mathcal{L}^2$ est muni, naturellement, d'une forme bilinéaire symétrique positive ou nulle, donnée par :

$$< X, Y >= E\{X.Y\} = \int_{\Omega} X.Y \mathrm{d}P \quad \text{où} \quad X \text{ et } Y \in \mathcal{L}^2$$

Noter que :

$$(X \text{ et } Y \in \mathcal{L}^2) \Rightarrow Z = X.Y \in \mathcal{L}^1$$

puisque

$$|Z| = |X|.|Y| \leq \frac{1}{2}(X^2 + Y^2)$$

On voit facilement, grâce à la linéarité de l'intégrale, que $< X, Y >$ est linéaire en $X$ et $Y$ ; de plus :

$$< X, X >= \int_{\Omega} X^2 \, \mathrm{d}P \geq 0$$

On a donc l'inégalité de Schwarz : pour tous $X$ et $Y \in \mathcal{L}^2$

$$| < X, Y > | \leq \sqrt{< X, X >}.\sqrt{< Y, Y >} \qquad (1)$$

ou bien

$$|E(X.Y)| \leq \sqrt{E(X^2)}.\sqrt{E(Y^2)}$$

et, en prenant $Y = 1$ :

$$|E(X)| \leq E(|X|) \leq \sqrt{E(X^2)} \qquad (1')$$

Remarquer que si $X = 0$ *presque sûrement* alors $< X, X >= E(X^2) = 0$, on va voir que, réciproquement, $E(X^2) = 0 \Rightarrow X = 0$ presque sûrement. Lorsque $n = 2$, le moment centré d'ordre 2 de $X$ est appelé *variance de* $X$ :

$$\mathrm{Var}(X) = E\{(X - E(X))^2\} = E\{X^2\} - (E\{X\})^2 \qquad (X \in \mathcal{L}^2)$$

*L'ecart type* $\sigma(X)$ est donné par :

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}$$

### 3.1.1 Proposition 1

1. *Inégalité de Markov* :
   Soit $X \in \mathcal{L}^1$, alors pour tout $\lambda > 0$ :

$$P\{|X| \geq \lambda\} \leq \frac{1}{\lambda} E\{|X|\}$$

2. *Inégalité de Bienaymé-Chebichev* :
   Soit $X \in \mathcal{L}^2$, alors pour tout $\lambda > 0$ :

$$P\{|X - E(X)| \geq \lambda\} \leq \frac{\mathrm{Var}(X)}{\lambda^2} = \frac{\sigma^2(X)}{\lambda^2}$$

Démonstration :
Soit $X$ une variable aléatoire réelle quelconque. Pour tous $\lambda > 0$ et $1 \leq p < \infty$, on a :

$$\{|X| \geq \lambda\} = \{|X|^p \geq \lambda^p\}$$

et

$$\lambda.\mathbf{1}_{\{|X| \geq \lambda\}} \leq |X|$$

donc

$$\int \lambda.\mathbf{1}_{\{|X| \geq \lambda\}} \mathrm{d}P \leq \int |X| \mathrm{d}P$$

i. e. :

$$\lambda P\{|X| \geq \lambda\} \leq E\{|X|\}$$

On voit, de même, que :

$$\lambda^p P\{|X| \geq \lambda\} \leq E\{|X|^p\}$$

Pour obtenir 2), il suffit de remplacer $X$ par $|X - E(X)|$.
Corollaire 1 :
Soit $X$ une variable aléatoire réelle quelconque, on a $E\{|X|\} = 0$ si et seulement si $X = 0$ presque sûrement. Si $X \in \mathcal{L}^2$, on a :

$$\mathrm{Var}(X) = 0 \quad \Longleftrightarrow \quad X = E(X) \quad \text{presque sûrement}$$

Démonstration :
On a :

$$\{|X| > 0\} = \bigcup_{\substack{\lambda \in \mathbb{Q} \\ \lambda > 0}} \{|X| \geq \lambda\}$$

donc

$$E(|X|) = 0 \quad \Rightarrow \quad P\{|X| > 0\} \leq \sum_{\substack{\lambda \in \mathbb{Q} \\ \lambda > 0}} P\{|X| \geq \lambda\} = 0$$

puisque

$$P\{|X| \geq \lambda\} \leq \frac{1}{\lambda} E(|X|) = 0$$

donc

$$\mathrm{Var}(X) = E\left((X - E(X))^2\right) = 0 \quad \Rightarrow \quad (X - E(X))^2 = 0$$

presque sûrement.

### 3.1.2 Proposition 2

1. Soit $X \in \mathcal{L}^2$. Alors pour tous $\lambda \in \mathbb{R}$ et $a \in \mathbb{R}$ :

$$\mathrm{Var}(\lambda X) = \lambda^2 \mathrm{Var}(X)$$
$$\mathrm{Var}(X + a) = \mathrm{Var}(X)$$

2. Soient $X_1, X_2, \ldots, X_n$ $n$ variables aléatoires réelles *indépendantes* de carré intégrable, alors :

$$\mathrm{Var}(X_1 + X_2 + \ldots + X_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \ldots + \mathrm{Var}(X_n)$$

Démonstration :
Le 1) est évident
Pour 2) on a :

$$\begin{aligned}
\mathrm{Var}(X_1 + \ldots + X_n) &= E\Big\{ \Big( \sum_i (X_i - E(X_i)) \Big)^2 \Big\} \\
&= \sum_{i,j} E\big\{(X_i - E(X_i)).(X_j - E(X_j))\big\} \\
&= \sum_{i=1}^{n} E\big\{(X_i - E(X_i))^2\big\} \\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i)
\end{aligned}$$

car si $i \neq j$ :

$$E\big\{(X_i - E(X_i)).(X_j - E(X_j))\big\} = E\big\{(X_i - E(X_i))\big\}E\big\{(X_j - E(X_j))\big\} = 0$$

### 3.1.3 Proposition 3

1. Soit $\varphi$ une application de $\mathbb{R}$ dans $\mathbb{R}$ *convexe* positive alors, pour tout $X \in \mathcal{L}^1$, on a :
$$\varphi\{E(X)\} \leq E\{\varphi(X)\}$$

2. Soient $p$ et $q \in ]1, \infty[$ tels que $\frac{1}{p} + \frac{1}{q} = 1$, alors
$$[X \in \mathcal{L}^p, Y \in \mathcal{L}^q] \Rightarrow X.Y \in \mathcal{L}^1$$
et
$$|E\{X.Y\}| \leq \big(E\big\{|X|^p\big\}\big)^{\frac{1}{p}} . \big(E\big\{|Y|^q\big\}\big)^{\frac{1}{q}}$$
(*Inégalité de Hölder*).

Démonstration :

1. Soit $\mathcal{S}$ l'ensemble des fonctions *affines* $\Delta$ de $\mathbb{R}$ dans $\mathbb{R}$ telles que pour tout $x \in \mathbb{R}$ : $\Delta(x) \leq \varphi(x)$. Si $\varphi$ est convexe, on montre facilement que, pour tout $x \in \mathbb{R}$ :
$$\varphi(x) = \sup\{\Delta(x), \Delta \in \mathcal{S}\}$$
Soit $\Delta \in \mathcal{S}$, on a :
$$\Delta\{E(X)\} = E\{\Delta(X)\} \leq E\{\varphi(X)\}$$
donc
$$\varphi\{E(X)\} = \sup_{\Delta \in \mathcal{S}} E\{\Delta(X)\} \leq E\{\varphi(X)\}$$

2. La fonction exponentielle étant convexe, on sait que pour tous $x \leq y$ dans $\mathbb{R}$ et $\alpha \in [0, 1]$ :
$$e^{\alpha x + (1-\alpha) y} \leq \alpha e^x + (1 - \alpha) e^y$$
En posant :
$$a = e^{\frac{x}{p}}, \quad b = e^{\frac{y}{q}} \quad \alpha = \frac{1}{p}, \quad \text{et} \quad \frac{1}{q} = 1 - \alpha$$
on voit que, pour tous $a$ et $b \in ]0, \infty[$ :
$$ab \leq \frac{1}{p}\, a^p + \frac{1}{q}\, b^q$$
donc
$$E\{|X|.|Y|\} \leq \frac{1}{p} E\{|X|^p\} + \frac{1}{q} E\{|Y|^q\} < \infty \quad (*)$$
si $X \in \mathcal{L}^p$ et $Y \in \mathcal{L}^q$.
Posons :
$$||X||_p = (E\{|X|^p\})^{\frac{1}{p}}$$

Alors si $||X||_p > 0$ et $||Y||_q > 0$, on voit, d'après (*), que

$$E\Big\{\frac{|X|}{||X||_p}.\frac{|Y|}{||Y||_q}\Big\} \leq \frac{1}{p} + \frac{1}{q} = 1$$

donc

$$E\{|X|.|Y|\} \leq ||X||_p.||Y||_q$$

Si $||X||_p = 0$ ou $||Y||_q = 0$, l'inégalité précédente est triviale.

Corollaire 4 :
Soient $X$ et $Y$ appartenant à $\mathcal{L}^p$ où $1 < p < \infty$. Alors

$$(E\{|X+Y|^p\})^{\frac{1}{p}} \leq (E\{|X|^p\})^{\frac{1}{p}} + (E\{|Y|^p\})^{\frac{1}{p}} \qquad (**)$$

Démonstration :
On a

$$|X+Y|^p \leq |X|.|X+Y|^{p-1} + |Y|.|X+Y|^{p-1}$$

soit $q \in ]1, \infty[$ tel que $\frac{1}{p} + \frac{1}{q} = 1$ alors $\frac{1}{q} = \frac{p-1}{p}$ et, d'après l'inégalité de Hölder :

$$\begin{aligned} E\{|X+Y|^p\} &\leq E\{|X|.|X+Y|^{\frac{p}{q}}\} + E\{|Y|.|X+Y|^{\frac{p}{q}}\} \\ &\leq \Big[(E\{|X|^p\})^{\frac{1}{p}} + (E\{|Y|^p\})^{\frac{1}{p}}\Big] (E\{|X+Y|^p\})^{\frac{1}{q}} \end{aligned}$$

Donc si $E\{|X+Y|^p\} > 0$, on obtient (**) en divisant les deux termes de l'inégalité par $(E\{|X+Y|^p\})^{\frac{1}{q}}$

### 3.1.4 Variables aléatoires vectorielles

Soit $M$ une matrice $r \times d$ ($r, d \in \mathbb{N}^*$). On notera $M'$ la matrice transposée. Si

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix} \in \mathbb{R}^r$$

on identifie donc $x$ à une matrice colonne $r \times 1$ et $x'$ à une matrice ligne $1 \times r$ et alors si

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_r \end{pmatrix} \in \mathbb{R}^r$$

$$x'.y = \sum_{l=1}^{r} x_l.y_l = y'.x =< x, y >$$

$$x'.x = ||x||^2$$

$$x.y' = (x_l.y_k)_{l,k=1,\ldots,r} \quad \text{matrice} \quad r \times r$$

Soit

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

une variable aléatoire à valeurs $\mathbb{R}^r$. Posons

$$||X||^2 = X_1^2 + X_2^2 + \ldots + X_r^2$$

donc

$$\max_{l=1,\ldots,r}\{|X_l|\} \leq ||X|| \leq |X_1| + \ldots + |X_r| \leq r. \max_{l=1,\ldots,r}\{|X_l|\}$$

et on voit que, pour tout $p \in [1, \infty[$, on a : $||X|| \in \mathcal{L}^p$ si et seulement si

$$|X_l| \in \mathcal{L}^p \qquad \forall\, l \in \{1, \ldots, r\}$$

On dira que le vecteur aléatoire $X$ appartient à $\mathcal{L}^p$ si $||X|| \in \mathcal{L}^p$.
Soit

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_r \end{pmatrix}$$

les lois des composantes $X_l$ $(l = 1, \ldots, r)$ s'appellent les *lois marginales* de $X$.
Les variables aléatoires $\{X_l,\ l = 1, \ldots, r\}$ sont indépendantes si et seulement si la loi de $X$ est le produit des lois marginales. Si la loi de

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_r \end{pmatrix}$$

a pour densité la fonction $f(x_1, x_2, \ldots, x_r)$, alors la loi de $X_1$ la pour densité :

$$g(x_1) = \int_{\mathbb{R}^{r-1}} f(x_1, x_2, \ldots, x_r)\,\mathrm{d}x_2\,\mathrm{d}x_3 \ldots \mathrm{d}x_r \qquad \text{etc}\ldots$$

<u>Définition</u> :
Soit

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_r \end{pmatrix}$$

un vecteur aléatoire intégrable ($X \in \mathcal{L}^1$). Le vecteur

$$E(X) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_r) \end{pmatrix} \in \mathbb{R}^r$$

est appelé *espérance de* $X$. $E(X)$ est aussi appelé *barycentre de la loi* $P_X$ *de* $X$.
Si $X$ est de carré intégrable ($X \in \mathcal{L}^2$), la matrice $r \times r$ :

$$D(X) = E\big\{(X - E(X)).(X - E(X))^{'}\big\}$$

est appelée *matrice de dispersion* ou de *variance - covariance* de $X$.
Soient $Y$ et $Z$ deux variables aléatoires réelles de carré intégrable ; rappelons que *la covariance de* $Y$ *et* $Z$ est donnée par :

$$\mathrm{Cov}(Y, Z) = E\big\{(Y - E(Y)).(Z - E(Z))\big\} = E\{Y.Z\} - E(Y).E(Z)$$

On a donc $\mathrm{Cov}(Y,Y) = \mathrm{Var}(Y)$ et, d'après l'inégalité de Schwarz :

$$|\mathrm{Cov}(Y,Z)| \leq \sqrt{\mathrm{Var}(Y)}.\sqrt{\mathrm{Var}(Z)}$$

si $Y$ et $Z$ sont indépendantes on a , de plus, $\mathrm{Cov}(Y,Z) = 0$ (mais la réciproque est fausse, en général : $\mathrm{Cov}(Y,Z) = 0 \quad \not\Rightarrow Y$ et $Z$ sont indépendantes).
Soit

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_r \end{pmatrix}$$

un vecteur aléatoire de carré intégrable. On voit que :

$$\begin{aligned} D(X) &= E\big\{(X - E(X)).(X - E(X))^{'}\big\} \\ &= E\{X.X^{'}\} - E(X).E(X)^{'} \\ &= (\mathrm{Cov}(X_l, X_k))_{l,k=1,\ldots,r} \\ &= (D_{l,k}(X))_{l,k=1,\ldots,r} \end{aligned}$$

### 3.1.5 Propriétés de la matrice de dispersion

1. $D = D(X)$ est une matrice symétrique non négative i.e. pour tout

$$\lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_r \end{pmatrix} \in \mathbb{R}^r$$

$$\sum_{l,k=1}^{r} \lambda_l.\lambda_k D_{l,k} = \lambda^{'}.D.\lambda =< \lambda, D\lambda > \quad \geq 0$$

en effet

$$\begin{aligned} \lambda^{'}.D.\lambda &= \lambda^{'}.E\big\{(X - E(X)).(X - E(X))^{'}\big\}.\lambda \\ &= E\big\{\lambda^{'}.(X - E(X)).(X - E(X))^{'}.\lambda\big\} \\ &= E\big\{ < \lambda, (X - E(X)) >^2 \big\} \quad \geq 0 \end{aligned}$$

2. Si $\lambda \in \mathbb{R}^r \quad D(X + \lambda) = D(X)$ c'est évident.
3. Soient $M$ une matrice (non aléatoire) de type $d \times r$ et $Y = M.X$ (variable aléatoire de carré intégrable à valeurs $\mathbb{R}^d$), alors :

$$D(Y) = M.D(X).M^{'}$$

en effet :

$$\begin{aligned} D(Y) &= E\big\{(Y - E(Y)).(Y - E(Y))^{'}\big\} \\ &= E\big\{[M.(X - E(X))].[M.(X - E(X))]^{'}\big\} \\ &= M.E\big\{(X - E(X)).(X - E(X))^{'}\big\}.M^{'} \\ &= M.D(X).M^{'} \end{aligned}$$

4. Soient $X$ et $Y$ deux vecteurs aléatoires *indépendants* à valeurs $\mathbb{R}^r$, de carré intégrable, alors :

$$D(X+Y) = D(X) + D(Y)$$

en particulier, si $r = 1$, on retrouve le fait que

$$\sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y)$$

Pour le vérifier, on peut supposer, d'après 2), que $X$ et $Y$ sont centrées et alors :

$$\begin{aligned} D(X+Y) &= E\big\{(X+Y).(X^{'}+Y^{'})\big\} \\ &= E\{X.X^{'}\} + E\{Y.Y^{'}\} \\ &= D(X) + D(Y) \end{aligned}$$

car

$$E\{X.Y^{'}\} = E\{Y.X^{'}\} = 0 \qquad \text{(indépendance)}$$

## 3.2 Calculs de lois

### 3.2.1 Rappel de quelques lois usuelles

**Loi binômiale** $B(n,p)$

Soient $n \in \mathbb{N}^*$ et $p \in [0,1]$, on dira qu'une variable aléatoire $X \sim B(n,p)$ si $X \in \{0,1,\ldots,n\}$ et $P\{X=k\} = C_n^k.p^k(1-p)^{n-k}$ pour tout $k \in \{0,1,\ldots,n\}$. On vérifie facilement que $X$ a même loi que $X_1 + X_2 + \ldots + X_n$ où les variables aléatoires $X_i, (i = 1,2,\ldots,n)$ sont indépendantes, de même loi $B(1,p)$. Si $X \sim B(n,p)$, $E(X) = np$ et $\mathrm{Var}(X) = np(1-p)$.

**Loi de Poisson de paramètre $\theta$ :** $P(\theta)$

Soit $\theta > 0$. On dira qu'une variable aléatoire $X \sim P(\theta)$ si $X \in \mathbb{N}$ et

$$P\{X=k\} = e^{-\theta}.\frac{\theta^k}{k!}$$

pour tout $k \in \mathbb{N}$. On a alors $E(X) = \theta$, $\mathrm{Var}(X) = \theta$

**Loi géométrique de paramètre** $a : g(a)$

Soit $a \in ]0,1[$. On dira qu'une variable aléatoire $X \sim g(a)$ si $X \in \mathbb{N}$ et $P\{X=k\} = (1-a)a^k$, pour tout $k \in \mathbb{N}$. On a alors

$$E(X) = \frac{a}{1-a}, \qquad \mathrm{Var}(X) = \frac{a}{(1-a)^2}$$

**Loi uniforme** $U(a,b)$

Soient $a$ et $b \in \mathbb{R}$, $a < b$ ; on appelle loi uniforme de paramètres $a$ et $b$, la loi de densité

$$\frac{1}{b-a}.\mathbf{1}_{[a,b]}(x)$$

par rapport à la mesure de Lebesgue sur $\mathbb{R}$. Si

$$X \sim U(a,b), \qquad E(X) = \frac{a+b}{2}, \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

**Loi gamma** $G(\alpha, \beta)$

Rappelons que pour $a > 0$, on pose

$$\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} \, \mathrm{d}x$$

alors, si $a > 1$, $\Gamma(a) = (a-1)\Gamma(a-1)$, en particulier $\Gamma(n) = (n-1)!$, si $n \in \mathbb{N}^*$. On en déduit que :

$$\int_0^{+\infty} x^{\alpha-1} e^{-\frac{x}{\beta}} \, \mathrm{d}x = \Gamma(\alpha).\beta^\alpha \quad \text{si} \quad \alpha > 0 \quad \text{et} \quad \beta > 0$$

On peut donc considérer la loi de probabilité sur $\mathbb{R}_+$ de densité par rapport à la mesure de Lebesgue :

$$f_{\alpha,\beta}(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1}.e^{-\frac{x}{\beta}}.\mathbf{1}_{\mathbb{R}_+}(x)$$

où $\alpha > 0$ et $\beta > 0$ sont des paramètres. Cette loi s'appelle loi gamma $G(\alpha, \beta)$. Si

$$X \sim G(\alpha, \beta), \qquad E(X) = \alpha.\beta, \qquad \mathrm{Var}(X) = \alpha.\beta^2$$

La loi $G(1, \beta)$ est aussi appelée *loi exponentielle de paramètre* $\frac{1}{\beta}$.

**Loi de Gauss centrée réduite sur** $\mathbb{R}^d$ : $N_d(0, I_d)$

Si on considère sur $\mathbb{R}^d$, la fonction

$$f(x) = (2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2}|x|^2} \quad \text{où} \quad |x|^2 = x_1^2 + \ldots + x_d^2$$

si

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \quad \text{on a} \quad f > 0 \quad \text{et} \quad \int_{\mathbb{R}^d} f(x)\mathrm{d}x = 1$$

$f$ est donc la densité d'une loi de probabilité sur $\mathbb{R}^d$ appelée $N_d(0, I_d)$ (on justifiera cette notation plus tard). En particulier, lorsque $d = 1$, on a la loi sur $\mathbb{R}$, $N_1(0,1)$ et si

$$X \sim N_1(0,1) \quad : \quad E(X) = 0, \quad \mathrm{Var}(X) = 1$$

On vérifie facilement qu'une variable aléatoire :

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$$

à valeurs $\mathbb{R}^d$ suit la loi $N_d(0, I_d)$ *si et seulement si* les composantes $X_1, X_2, \ldots, X_d$ sont indépendantes, de même loi $N_1(0,1)$.

**Loi de Gauss** $N_1(m, \sigma^2)$

Soient $m \in \mathbb{R}$ et $\sigma^2 > 0$, deux paramètres. On dira qu'une variable aléatoire réelle $X \sim N_1(m, \sigma^2)$ si $X$ a pour densité la fonction :

$$f_{m,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\ e^{-\frac{1}{2\sigma^2}(x-m)^2} \qquad (x \in \mathbb{R})$$

On vérifie facilement que $X \sim N_1(m, \sigma^2)$ *si et seulement si* $Y = \frac{X-m}{\sigma} \sim N_1(0, 1)$, dans ce cas : $E(X) = m$, $\mathrm{Var}(X) = \sigma^2$.

### 3.2.2 Calculs de loi

On a souvent à calculer la loi d'une variable aléatoire. Ce calcul peut se faire à l'aide de la proposition suivante :

**Proposition 1**

Soit $X$ une variable aléatoire à valeurs $\mathbb{R}^d$. Pour qu'une probabilité $\mu$ sur $\mathbb{R}^d$ soit la loi de $X$, il faut et il suffit que, pour toute application $\varphi$ mesurable positive (ou bornée) de $\mathbb{R}^d$ dans $\mathbb{R}$, on ait :

$$E\{\varphi(X)\} = \int_{\mathbb{R}^d} \varphi(x)\mu(\mathrm{d}x)$$

Démonstration : Immédiate
Exemple 1 :
Si $X \sim N_1(0, 1)$, quelle est la loi de $Y = X^2$ ? Soit $\varphi$ mesurable : $\mathbb{R} \to \mathbb{R}_+$

$$\begin{aligned} E\{\varphi(Y)\} &= E\{\varphi(X^2)\} \\ &= \frac{1}{\sqrt{2\pi}} \int \varphi(x^2) e^{-\frac{x^2}{2}}\,\mathrm{d}x \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} \varphi(x^2) e^{-\frac{x^2}{2}}\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} \varphi(y) e^{-\frac{y}{2}} \frac{1}{\sqrt{y}}\,\mathrm{d}y \end{aligned}$$

Donc la loi de $Y$ a pour densité :

$$\frac{1}{\sqrt{2\pi y}} e^{-\frac{y^2}{2}}.\mathbf{1}_{\mathbb{R}_+}(y)$$

On reconnait ici la loi $G(\frac{1}{2}, 2)$.
On a souvent à utiliser la formule du changement de variables dans les intégrales multiples que nous rappelons sans démonstration.

**Proposition 2**

Soient $U$ et $V$ deux ouverts de $\mathbb{R}^d$ et $\varphi$ un difféomorphisme de $U$ sur $V$ (bijection de classe $C^1$ ainsi que son inverse) ; on a, pour toute fonction $f$ de $V$ dans $\mathbb{R}$ mesurable positive (ou intégrable) :

$$\int_V f(v)\,\mathrm{d}v = \int_U f(\varphi(u)).|J\varphi(u)|\,\mathrm{d}u$$

où $J\varphi$ désigne le Jacobien de $\varphi = (\varphi^1, \ldots, \varphi^d)$ :

$$J\varphi(u) = \det\left(\left(\frac{\partial \varphi^i}{\partial u_j}(u)\right)_{i,j=1,\ldots,d}\right)$$

Considérons maintenant une variable aléatoire $X$ presque sûrement à valeurs dans l'ouvert $U$, de densité $p(x)$ et soit $\varphi$ un difféomorphisme de l'ouvert $U$ sur l'ouvert $V$ ; soit $Y = \varphi(X)$. Cherchons la loi de $Y$ : on a, pour $\theta \geq 0$ mesurable :

$$\begin{aligned} E\{\theta(Y)\} &= E\{\theta\left(\varphi(X)\right)\} \\ &= \int_U \theta\left(\varphi(x)\right) p(x) \,\mathrm{d}x \\ &= \int_V \theta(y) p\left(\varphi^{-1}(y)\right).|J\varphi^{-1}(y)| \,\mathrm{d}y \end{aligned}$$

et donc la densité de $Y$ est donnée par :

$$p\left(\varphi^{-1}(y)\right).|J\varphi^{-1}(y)|.\mathbf{1}_V(y)$$

<u>Remarque</u> :
Il peut arriver que l'on ait à chercher la densité de $Y = \varphi(X)$ lorsque $\varphi$ n'est pas une bijection. Alors on décompose (si c'est possible) l'image de $X$ en régions où $\varphi$ est bijective et on applique, sur chaque morceau, la formule précédente.
<u>Exemple 2</u> :
Soit $X = \binom{X_1}{X_2}$ une variable aléatoire à valeurs $\mathbb{R}^2$ de loi $N_2(0, I_2)$. On pose $Y = \frac{X_1}{X_2}$ ($= 0$ si $X_2 = 0$). Quelle est la loi de $Y$ ? Remarquons, d'abord, que :

$$P\{X_2 = 0\} = \int_{\mathbb{R}^2} \mathbf{1}_{(x_2=0)} \frac{1}{2\pi} e^{-\frac{|x|^2}{2}} \,\mathrm{d}x = 0$$

On a, pour toute application $\theta$ de $\mathbb{R}$ dans $\mathbb{R}_+$, mesurable :

$$\begin{aligned} E\{\theta(Y)\} &= E\Big\{\theta\left(\frac{X_1}{X_2}).\mathbf{1}_{(X_2\neq 0)}\right)\Big\} \\ &= \int_{\mathbb{R}^2} \theta\left(\frac{x_1}{x_2}\right).\mathbf{1}_{(x_2\neq 0)}.\frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2+x_2^2)} \,\mathrm{d}x_1 \,\mathrm{d}x_2 \end{aligned}$$

On pose :

$$\frac{x_1}{x_2} = y, \qquad x_2 = z \qquad J = \begin{vmatrix} z & y \\ 0 & 1 \end{vmatrix} = z$$

l'ouvert $U = \{x_2 \neq 0\}$ se transforme en $U$ donc :

$$\begin{aligned} E\{\theta(Y)\} &= \frac{1}{2\pi} \int_{\mathbb{R}^2} \theta(y) e^{-\frac{1}{2}z^2(1+y^2)} |z| \,\mathrm{d}y \,\mathrm{d}z \\ &= \int_{\mathbb{R}} \theta(y) \Big\{\frac{1}{2\pi} \int_{\mathbb{R}} e^{-\frac{1}{2}z^2(1+y^2)}.|z| \,\mathrm{d}z\Big\} \,\mathrm{d}y \\ &= \frac{1}{\pi} \int \theta(y) \frac{\mathrm{d}y}{1+y^2} \end{aligned}$$

La loi de $Y$ a donc pour densité $\frac{1}{\pi(1+y^2)}$ ; cette loi s'appelle *la loi de Cauchy.* Remarquer que $Y \notin \mathcal{L}^1$ puisque

$$E\{|Y|\} = \frac{1}{\pi} \int \frac{|y|}{1+y^2} \,\mathrm{d}y = +\infty$$

Exemple 3 : Processus de Poisson de paramètre $\theta$

On suppose que les instants d'arrivée d'un phénomène aléatoire forment une suite strictement croissante de variables aléatoires réelles positives

$$0 < T_1(\omega) < T_2(\omega) < \ldots < T_n(\omega) < T_{n+1}(\omega) < \ldots \nearrow +\infty$$

vérifiant la condition :
(*) les variables aléatoires $T_1$, $T_2 - T_1$, ... , $T_{n+1} - T_n$, ... sont *indépendantes* de *même loi exponentielle* de paramètre $\theta > 0$
Pour tout $t \in \mathbb{R}_+$, soit

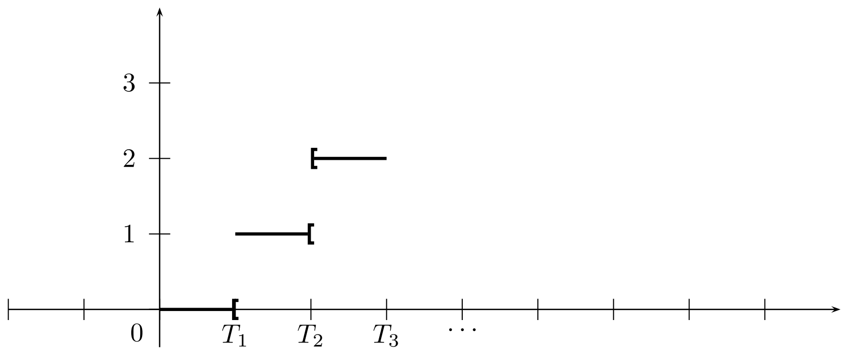
$$N_t(\omega) = \mathrm{Card}\{m \in \mathbb{N}^* : T_m(\omega) \leq t\} = \sum_{m \in \mathbb{N}^*} \mathbf{1}_{(T_m \leq t)}(\omega)$$

Sous les hypothèses précédentes la famille de variables aléatoires entières $(N_t)_{t \geq 0}$ est appelée *« Processus de Poisson de paramètre $\theta$ »*. Ce processus est *caractérisé* par les propriétés suivantes :

1. Pour presque tout $\omega \in \Omega$, la « trajectoire »

$$t \in \mathbb{R}+ \longmapsto N_t(\omega) \in \mathbb{N}$$

   est croissante, continue à droite et ne croît que par sauts égaux à $+1$ ; de plus $N_0(\omega) = 0$




2. Pour toute suite croissante de réels positifs :

$$0 \leq t_0 < t_1 < \ldots < t_n < t_{n+1} < \ldots$$

   Les variables aléatoires

$$N_{t_0}, N_{t_1} - N_{t_0}, \ldots, N_{t_{n+1}} - N_{t_n}, \ldots$$

   sont indépendantes. (on dit alors que le processus $(N_t)_{t \geq 0}$ est *à accroissements indépendants*)

3. Pour tous $0 \leq s < t$, les variables aléatoires $N_t - N_s$ et $N_{t-s}$ ont même loi de Poisson de paramètre $\theta(t-s)$.
Pour démontrer ces propriétés, cherchons, d'abord, la loi du vecteur $(T_1, T_2, \ldots, T_n)$ ($n \in \mathbb{N}^*$). En posant :

$$X_1 = T_1,\ X_2 = T_2 - T_1,\ \ldots,\ X_{n+1} = T_{n+1} - T_n,\ \ldots$$

l'hypothèse (*****) signifie que le vecteur aléatoire $(X_1, X_2, \ldots, X_n)$ a, pour tout $n \in \mathbb{N}^*$, la densité sur $\mathbb{R}^n$ :

$$\theta^n e^{-\theta(x_1+x_2+\ldots+x_n)}.\mathbf{1}_{(x_1\geq 0, x_2\geq 0, \ldots, x_n \geq 0)}$$

Soit $\varphi$ une application mesurable de $\mathbb{R}^n$ dans $\mathbb{R}_+$, on a :

$$E\{\varphi(T_1, T_2, \ldots, T_n)\} = E\{\varphi(X_1, X_1 + X_2, \ldots, X_1 + X_2 + \cdots + X_n)\}$$

$$= \theta^n \int_{\mathbb{R}^n} \varphi(x_1, x_1 + x_2, \ldots, x_1 + x_2 + \cdots + x_n) e^{-\theta(x_1+\cdots+x_n)}.\mathbf{1}_{(x_1\geq 0,\ldots,x_n\geq 0)} \mathrm{d}x_1 \ldots \mathrm{d}x_n$$
$$= \theta^n \int_{\mathbb{R}^n} \varphi(t_1, t_2, \ldots, t_n) e^{-\theta t_n}.\mathbf{1}_{(0\leq t_1 \leq t_2 \leq \ldots \leq t_n)} \mathrm{d}t_1 \mathrm{d}t_2 \ldots \mathrm{d}t_n$$

en effectuant le changement de variables :

$$x_1 = t_1, \quad x_1 + x_2 = t_2, \quad x_1 + \ldots + x_n = t_n$$

(de Jacobien égal à 1). La densité du vecteur $(T_1, T_2, \ldots, T_n)$ est donc égale à :

$$\theta^n.e^{-\theta t_n}.\mathbf{1}_{(0\leq t_1\leq t_2 \leq \ldots \leq t_n)}$$

Vérifions maintenant les propriétés 2) et 3), en cherchant la loi du vecteur $(N_{t_0}, N_{t_1} - N_{t_0}, \ldots, N_{t_{n+1}} - N_{t_n})$ où $0 \leq t_0 < t_1 < \ldots < t_n < t_{n+1}$. On fera le calcul pour $n = 2$, la démonstration du cas « $n$ quelconque » étant analogue.
Soient $0 < t_0 < t_1$ et $k, l \in \mathbb{N}$. on a :

$$\{\omega : N_{t_0}(\omega) = k, \quad N_{t_1}(\omega) - N_{t_0}(\omega) = l\} = \{\omega : N_{t_0}(\omega) = k, \quad N_{t_1}(\omega) = k + l\}$$
$$= \{\omega : T_k(\omega) \leq t_0 < T_{k+1}(\omega), \quad T_{k+l}(\omega) \leq t_1 < T_{k+l+1}(\omega)\}$$

Donc :

$$P\{N_{t_0} = k, \quad N_{t_1} - N_{t_0} = l\}$$
$$= \theta^{k+l+1} \int_{\mathbb{R}^{k+l+1}} e^{-\theta s_{k+l+1}}.$$
$$\mathbf{1}_{(0<s_1<\ldots<s_k\leq t_0 < s_{k+1} < \ldots < s_{k+l} \leq t_1 < s_{k+l+1})} \mathrm{d}s_1 \ldots \mathrm{d}s_k \, \mathrm{d}s_{k+1} \ldots \mathrm{d}s_{k+l} \, \mathrm{d}s_{k+l+1}$$
$$= \theta^{k+l}.e^{-\theta t_1} \int_{\mathbb{R}^k} \mathbf{1}_{(0<s_1<\ldots<s_k\leq t_0)} \mathrm{d}s_1 \ldots \mathrm{d}s_k. \int_{\mathbb{R}^l} \mathbf{1}_{(t_0<s_{k+1}<\ldots<s_{k+l}\leq t_1)} \mathrm{d}s_{k+1} \ldots \mathrm{d}s_{k+l}$$
$$= \theta^{k+l}.e^{-\theta t_1}.\frac{t_0^k}{k!}\frac{(t_1 - t_0)^l}{l!} \qquad \text{en utilisant le théorème de Fubini}$$
$$= \left[\frac{(\theta t_0)^k}{k!} e^{-\theta t_0}\right] \left[\frac{(\theta(t_1 - t_0))^l}{l!}.e^{-\theta(t_1-t_0)}\right] \qquad (1)$$

Si on somme en $l$ l'égalité (1) précédente, on voit que pour tout $k \in \mathbb{N}$

$$P\{N_{t_0} = k\} = \frac{(\theta t_0)^k}{k!}.e^{-\theta t_0}$$

et on voit de même (sommer en $k$) que pour tout $l \in \mathbb{N}$

$$P\{N_{t_1} - N_{t_0} = l\} = \frac{(\theta(t_1 - t_0))^l}{l!}.e^{-\theta(t_1-t_0)}$$

Donc :

$$P\{N_{t_0} = k, N_{t_1} - N_{t_0} = l\} \ = \ P\{N_{t_0} = k\}P\{N_{t_1} - N_{t_0} = l\}, \qquad \forall\ k, l \in \mathbb{N}$$

Les variables aléatoires $N_{t_0}$ et $N_{t_1} - N_{t_0}$ sont inépendantes,

$$N_{t_0} \sim \text{Poisson } \theta t_0, \qquad N_{t_1} - N_{t_0} \sim \text{Poisson } \theta(t_1 - t_0)$$

Le calcul précédent montre que, pour tout $t \geq 0$, $P(N_t < \infty) = 1$. La vérification de 1) est une conséquence immédiate de la définition de $N_t$.
Réciproquement, soit $(\tilde{N}_t)_{t\geq 0}$ un processus à valeurs dans $\mathbb{N}$ (famille de variables aléatoires) vérifiant les propriétés 1), 2) et 3). Montrons que la suite $(\tilde{T}_n)$ des « instants de saut » de la trajectoire $t \in \mathbb{R}_+ \longmapsto \tilde{N}_t(\omega) \in \mathbb{N}$ vérifie la propriété (*). On a

$$\tilde{T}_1 = \inf\{t > 0 : \tilde{N}_t = 1\}, \tilde{T}_2 = \inf\{t > 0 : \tilde{N}_t = 2\}, \ldots, \tilde{T}_n = \inf\{t > 0 : \tilde{N}_t = n\}$$

On voit facilement que $0 < \tilde{T}_1(\omega) < \tilde{T}_2(\omega) < \ldots < \tilde{T}_n(\omega) < \ldots$ et, par continuité, que $\tilde{N}_{\tilde{T}_n(\omega)}(\omega) = n$, pour tout $n$.
Nous allons vérifier que, pour tout $n \in \mathbb{N}^*$, les vecteurs aléatoires $(T_1, T_2, \ldots, T_n)$ et $(\tilde{T}_1, \tilde{T}_2, \ldots, \tilde{T}_n)$ ont même loi. Remarquons d'abord que ces deux vecteurs sont presque sûrement à valeurs dans l'ouvert :

$$\bigcirc = \left\{ t = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} \quad \text{tel que} \quad 0 < t_1 < t_2 < \ldots < t_n \right\} \ \subseteq \mathbb{R}^n$$

Soit $\mathcal{C}$ l'ensemble des pavés $C$ de la forme :

$$C = ]u_1, v_1] \times ]u_2, v_2] \times \ldots \times ]u_n, v_n] \quad \text{où} \quad 0 < u_1 < v_1 < u_2 < v_2 < \ldots < u_n < v_n$$

On voit facilement que la tribu engendrée par $\mathcal{C}$ : $\sigma(\mathcal{C})$ esr égale à la tribu borélienne de $\bigcirc$ (exercice). Pour tout $C \in \mathcal{C}$, $C = ]u_1, v_1] \times \ldots \times ]u_n, v_n]$, on a :

$$\begin{aligned}
&P\{(T_1, T_2, \ldots, T_n) \in C\} \\
&= P\{u_1 < T_1 \leq v_1, u_2 < T_2 \leq v_2, \ldots, u_n < T_n \leq v_n\} \\
&= P\{N_{u_1} = 0, N_{v_1} - N_{u_1} = 1, N_{u_2} - N_{v_1} = 0, N_{v_2} - N_{u_2} = 1, \ldots, N_{u_n} - N_{v_{n-1}} = 0, \\
&\qquad N_{v_n} - N_{u_n} = 1\} \\
&= \left[e^{-\theta u_1}.\theta(v_1 - u_1).e^{-\theta(v_1-u_1)}\right].\left[e^{-\theta(u_2-v_1)}.\theta(v_2 - u_2).e^{-\theta(v_2-u_2)}\right]\ldots \\
&\qquad \left[e^{-\theta(u_n-v_{n-1})}.\theta(v_n - u_n).e^{-\theta(v_n-u_n)}\right] \\
&= P\{\tilde{N}_{u_1} = 0, \tilde{N}_{v_1} - \tilde{N}_{u_1} = 1, \tilde{N}_{u_2} - \tilde{N}_{v_1} = 0, \tilde{N}_{v_2} - \tilde{N}_{u_2} = 1, \ldots, \tilde{N}_{u_n} - \tilde{N}_{v_{n-1}} = 0, \\
&\qquad \tilde{N}_{v_n} - \tilde{N}_{u_n} = 1\} \\
&= P\{(\tilde{T}_1, \tilde{T}_2, \ldots, \tilde{T}_n) \in C\}
\end{aligned}$$

Donc le vecteur aléatoire $(\tilde{T}_1, \tilde{T}_2, \ldots, \tilde{T}_n)$ a même densité que $(T_1, T_2, \ldots, T_n)$ et vérifie ainsi la propriété (*), comme on le voit grâce à un changement de variable évident.

## 3.3 Fonctions caractéristiques

### 3.3.1 Définition 1

Soit $\mu$ une mesure positive $\sigma$-finie, *bornée* sur$(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. Pour tout $t \in \mathbb{R}^d$, la fonction

$$x \in \mathbb{R}^d \longmapsto e^{\imath t' x} \in \mathbb{C} \quad \text{où} \quad t' x = \sum_{l=1}^{d} t_l x_l$$

est intégrable pour $\mu$, puisque $|e^{\imath t' x}| = 1$. Posons :

$$\widehat{\mu}(t) = \int_{\mathbb{R}^d} e^{\imath t' x} \mu(\mathrm{d}x)$$

L'application $\widehat{\mu} : t \in \mathbb{R}^d \longmapsto \widehat{\mu}(t) \in \mathbb{C}$ est appelée *transformée de Fourier de $\mu$*.
Si $\mu$ est la loi $P_X$ d'une variable aléatoire

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$$

à valeurs dans $\mathbb{R}^d$, définie sur $(\Omega, \mathcal{A}, P)$, on a :

$$\widehat{P}_X(t) = \int_{\Omega} e^{\imath t' X(\omega)} \mathrm{d}P(\omega) = E\{e^{\imath t' X}\} \quad \text{(formule des lois images)}$$

On notera alors $\widehat{P}_X(t) = \varphi_X(t)$ $(t \in \mathbb{R}^d)$.
$\varphi_X$ est appelée *fonction caractéristique de $X$*.

### 3.3.2 Proposition 1

$\varphi_X$ est une fonction uniformément continue sur $\mathbb{R}^d$, $|\varphi_X| \leq 1$ et $\varphi_X(0) = 1$.
Démonstration :
Pour tout $t \in \mathbb{R}^d$, on a :

$$|\varphi_X(t)| = |E\{e^{\imath t' X}\}| \leq E\{|e^{\imath t' X}|\} = E\{1\} = 1$$

et

$$\varphi_X(0) = E\{1\} = 1$$

Remarquons ensuite, que pour tout $x \in \mathbb{R}$ : $|e^{\imath x} - 1| \leq \min\{2, |x|\}$. Donc si $t$ et $s \in \mathbb{R}^d$ et $|t - s| \leq r$, on a :

$$\begin{aligned}
|\varphi_X(t) - \varphi_X(s)| &= |E\{e^{\imath t' X} - e^{\imath s' X}\}| \\
&\leq E\{|e^{\imath (t' - s') X} - 1|\} \\
&\leq E\Big\{\min\{2, |(t' - s')X|\}\Big\} \\
&\leq E\Big\{\min\{2, r|X|\}\Big\} \to 0 \quad \text{quand} \quad r \to 0
\end{aligned}$$

grâce au théorème de Lebesgue.

### 3.3.3 Théorème 1 fondamental

1. Soient $X$ et $\widetilde{X}$ deux variables aléatoires à valeurs dans $\mathbb{R}^d$ telles que, pour tout $t \in \mathbb{R}^d$
$$\varphi_X(t) = \varphi_{\widetilde{X}}(t)$$
alors $X$ et $\widetilde{X}$ ont même loi.
2. Si
$$\varphi_X(.) \in \mathcal{L}^1(\mathbb{R}^d, \mathrm{d}t)$$
où $\mathrm{d}t$ est la mesure de Lebesgue alors $X$ admet une densité $f$ (par rapport à $\mathrm{d}t$) et, pour tout $x \in \mathbb{R}^d$
$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-\imath t' x} \varphi_X(t) \mathrm{d}t$$
(formule d'inversion)

Nous donnons, en appendice, une démonstration de 1) ; pour une preuve de 2), consulter, par exemple, le livre de W. Rudin « Real and Complex Analysis » Mc Grow Hill (Ed.)

Corollaire 1 :

Soient $X_1, X_2, \ldots, X_k$ des variables aléatoires à valeurs $\mathbb{R}^{d_1}$, $\mathbb{R}^{d_2}$, …, $\mathbb{R}^{d_k}$ respectivement. Pour que $X_1, X_2, \ldots, X_k$ soient indépendantes, il faut et il suffit que, quels que soient $t_1 \in \mathbb{R}^{d_1}$, $t_2 \in \mathbb{R}^{d_2}$, …, $t_k \in \mathbb{R}^{d_k}$ :
$$E\{e^{\imath(t_1'.X_1+t_2'.X_2+\ldots+t_k'.X_k)}\} = \prod_{l=1}^{k} E\{e^{\imath t_l'.X_l}\}$$

Démonstration :

On a
$$(X_1, \ldots, X_k) \quad \text{indépendantes} \quad \Longleftrightarrow \quad P_{(X_1,\ldots,X_k)} = \bigotimes_{l=1}^{k} P_{X_l} \quad \text{sur} \quad (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$$
(où $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \ldots \times \mathbb{R}^{d_k}$)
$$\Longleftrightarrow \quad \forall\ t = \begin{pmatrix} t_1 \\ \vdots \\ t_k \end{pmatrix} \in \mathbb{R}^{d_1} \times \ldots \times \mathbb{R}^{d_k} \ : \ \widehat{P}_{(X_1,\ldots,X_k)}(t) = \widehat{\bigotimes_{l=1}^{k} P_{X_l}}(t)$$
mais
$$\widehat{P}_{(X_1,\ldots,X_k)}(t) = E\left\{e^{\imath \sum_{l=1}^{k} t_l'.X_l}\right\}$$
et
$$\begin{aligned}
\widehat{\bigotimes_{l=1}^{k} P_{X_l}}(t) &= \int_{\mathbb{R}^{d_1} \times \ldots \times \mathbb{R}^{d_k}} e^{\imath \sum_{l=1}^{k} t_l'.x_l} P_{X_1}(\mathrm{d}x_1) \ldots P_{X_k}(\mathrm{d}x_k) \\
&= \prod_{l=1}^{k} \int_{\mathbb{R}^{d_l}} e^{\imath t_l'.X_l} P_{x_l}(\mathrm{d}x_l) \quad \text{(Fubini)} \\
&= \prod_{l=1}^{k} E\left\{e^{\imath t_l'.X_l}\right\}
\end{aligned}$$

### 3.3.4 Proposition 2

Soient $X$ et $Y$ deux variables aléatoires *indépendantes* à valeurs $\mathbb{R}^d$. Alors, pour tout $t \in \mathbb{R}^d$ :

$$\varphi_{X+Y}(t) = \varphi_X(t).\varphi_Y(t) \quad (*)$$

Démonstration :
On a

$$\begin{aligned} E\{e^{\imath t'(X+Y)}\} &= E\{e^{\imath t' X}.e^{\imath t' Y}\} \\ &= E\{e^{\imath t' X}\}.E\{e^{\imath t' Y}\} \qquad \text{(indépendance)} \end{aligned}$$

Remarquer que l'égalité (*) signifie que :

$$\widehat{P}_{X+Y} = \widehat{P_X * P_Y} \quad \text{i.e.} \quad P_{X+Y} = P_X * P_Y$$

### 3.3.5 Proposition 3

Soit :

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$$

une variable aléatoire à valeurs $\mathbb{R}^d$, de fonction caractéristique $\varphi_X = \varphi$.

1. Si $X$ est intégrable, alors $\varphi$ est de classe $\mathcal{C}^1$ et
$$E\{X_k\} = -\imath \frac{\partial \varphi}{\partial t_k}(0), \qquad k = 1, \ldots, d$$
2. Si $X$ est de carré intégrable, alors $\varphi$ est de classe $\mathcal{C}^2$ et
$$E\{X_k X_l\} = -\frac{\partial^2 \varphi}{\partial t_k \partial t_l}(0), \qquad k, l \in \{1, \ldots, d\}$$

Plus généralement si $X \in \mathcal{L}^p$ ($p \in \mathbb{N}^+$), alors $\varphi$ est de classe $\mathcal{C}^p$ et

$$\frac{\partial^r}{\partial t_1^{\alpha_1} \ldots \partial t_d^{\alpha_d}} \varphi(0) = (\imath)^r.E\{X_1^{\alpha_1} X_2^{\alpha_2} \ldots X_d^{\alpha_d}\}$$

où $\alpha_1, \ldots, \alpha_d \in \mathbb{N}$ et $\alpha_1 + \alpha_2 + \ldots + \alpha_d = r \leq p$
Démonstration :
Le 1) est une conséquence facile du théorème de dérivation sous le signe intégral :

$$\varphi(t) = \int_\Omega e^{\imath \sum_{l=1}^d t_l X_l(\omega)} \mathrm{d}P(\omega)$$

et

$$\frac{\partial}{\partial t_k} \varphi(t) = \imath \int_\Omega X_k(\omega) e^{\imath \sum_{l=1}^d t_l X_l(\omega)} \mathrm{d}P(\omega)$$

car

$$|\imath X_k e^{\imath \sum_l t_l X_l}| = |X_k| \leq ||X||$$

qui est intégrable par hypothèse.
Le fait que $\frac{\partial}{\partial t_k} \varphi(.)$ soit continue est une conséquence du théorème de Lebesgue ; de plus

$$\frac{\partial}{\partial t_k} \varphi(0) = \imath \int_\Omega X_k(\omega) \mathrm{d}P(\omega) = \imath E\{X_k\}$$

On démontre, de la même façon, l'assertion 2).

### 3.3.6 Proposition 3'

Si la fonction caractéristique d'une variable aléatoire réelle $X$ est de classe $\mathcal{C}^2$ alors $X$ est de carré intégrable.

Démonstration :
La formule de Taylor au voisinage de zéro montre que :

$$\varphi(h) = 1 + \varphi^{'}(0).h + \frac{1}{2}\varphi^{"}(0).h^2 + h^2\varepsilon(h)$$

$$\varphi(-h) = 1 - \varphi^{'}(0).h + \frac{1}{2}\varphi^{"}(0).h^2 + h^2\varepsilon(-h)$$

où $\varepsilon(h)$ et $\varepsilon(-h) \to 0$, quand $h \to 0$, donc

$$(\varphi(h) + \varphi(-h) - 2) \cdot \frac{1}{h^2} \to \varphi^{"}(0) \quad \text{quand} \quad (h \neq 0) \to 0$$

mais

$$\begin{aligned}\varphi(h) + \varphi(-h) - 2 &= E\{e^{\imath hX} + e^{-\imath hX} - 2\} \\ &= E\big\{[e^{\imath \frac{h}{2} X} - e^{-\imath \frac{h}{2} X}]^2\big\} \\ &= -4E\left\{\sin^2\left(\frac{hX}{2}\right)\right\}\end{aligned}$$

On en déduit que, d'après le lemme de Fatou

$$\begin{aligned}E\{X^2\} &= E\left\{\lim_{\substack{h\to 0\\ h\neq 0}} 4\frac{\sin^2(\frac{hX}{2})}{h^2}\right\} \\ &\leq \underline{\lim}_{\substack{h\to 0\\ h\neq 0}} E\left\{4\frac{\sin^2(\frac{hX}{2})}{h^2}\right\} \\ &\leq -\varphi^{"}(0) < \infty\end{aligned}$$

Exemples :

1. Soit $X \sim B(p,n)$ (loi binômiale de paramètres $p \in [0,1]$ et $n \in \mathbb{N}^*$). Alors :

$$\varphi_X(t) = [pe^{\imath t} + (1-p)]^n \qquad (t \in \mathbb{R})$$

En effet

$$\begin{aligned}\varphi_X(t) &= \sum_{r=0}^{n} C_n^r e^{\imath tr} p^r (1-p)^{n-r} \\ &= \sum_{r=0}^{n} C_n^r \left(pe^{\imath t}\right)^r (1-p)^{n-r}\end{aligned}$$

si, de plus, $Y \sim B(p,m)$ est une variable aléatoire *indépendante* de $X$ :

$$\begin{aligned}\varphi_{X+Y}(t) &= \varphi_X(t).\varphi_Y(t) \\ &= [pe^{\imath t} + (1-p)]^n.[pe^{\imath t} + (1-p)]^m \\ &= [pe^{\imath t} + (1-p)]^{n+m}\end{aligned}$$

donc $X + Y \sim B(p, m+n)$.

2. Soit $X \sim N_1(0,1)$, alors
$$\varphi_X(t) = e^{-\frac{t^2}{2}}$$
En effet
$$\varphi_X(t) = \frac{1}{\sqrt{2\pi}} \int e^{\imath tx - \frac{x^2}{2}} \mathrm{d}x$$
et
$$\begin{aligned} \sqrt{2\pi}\varphi^{'}(t) &= \imath \int e^{\imath tx} x e^{-\frac{x^2}{2}} \mathrm{d}x \\ &= -\int t e^{\imath tx} e^{-\frac{x^2}{2}} \mathrm{d}x \qquad \text{(intégration par parties)} \\ &= -t\varphi(t)\sqrt{2\pi} \end{aligned}$$
donc
$$\varphi^{'}(t) = -t\varphi(t) \qquad \text{et} \qquad \varphi(t) = Ce^{-\frac{t^2}{2}}$$
de plus $C = 1$ car $\varphi(0) = 1$
3. Soit $X \sim N_1(m, \sigma^2)$, alors
$$\varphi_X(t) = e^{\imath tm - \frac{t^2\sigma^2}{2}}$$
On sait que, si $Z \sim N_1(0,1)$, $X = m + \sigma Z \sim N_1(m, \sigma^2)$ donc
$$\begin{aligned} E\{e^{itX}\} &= E\{e^{\imath t(m+\sigma Z)}\} \\ &= e^{\imath tm} E\{e^{\imath t\sigma Z}\} \\ &= e^{\imath tm} e^{-\frac{t^2\sigma^2}{2}} \\ &= e^{\imath tm - \frac{t^2\sigma^2}{2}} \end{aligned}$$
Si, de plus, $Y \sim N_1(\tilde{m}, \tilde{\sigma}^2)$ est une variable aléatoire *indépendante* de $X$ alors
$$\begin{aligned} \varphi_{X+Y}(t) &= \varphi_X(t).\varphi_Y(t) \\ &= e^{\imath tm - \frac{t^2}{2}\sigma^2}.e^{\imath t\tilde{m} - \frac{t^2}{2}\tilde{\sigma}^2} \\ &= e^{\imath t(m+\tilde{m}) - \frac{t^2}{2}(\sigma^2 + \tilde{\sigma}^2)} \end{aligned}$$
donc $X + Y \sim N_1(m + \tilde{m}, \sigma^2 + \tilde{\sigma}^2)$.
4. Soit
$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} \sim N_d(0, I_d)$$
alors
$$\begin{aligned} \varphi(t) &= e^{-\frac{1}{2}(t_1^2 + \ldots + t_d^2)} \\ &= e^{-\frac{1}{2}|t|^2} \end{aligned}$$

si

$$t = \begin{pmatrix} t_1 \\ \vdots \\ t_d \end{pmatrix} \in \mathbb{R}^d$$

En effet $(X_1, X_2, \dots, X_d)$ sont indépendantes, de même loi $N_1(0,1)$ donc :

$$\begin{aligned} E\{e^{\imath t' X}\} &= E\{e^{\imath \sum_{l=1}^d t_l X_l}\} \\ &= \prod_{l=1}^d E\{e^{\imath t_l X_l}\} \\ &= \prod_{l=1}^d e^{-\frac{1}{2}t_l^2} \\ &= e^{-\frac{1}{2}|t|^2} \end{aligned}$$

5. Soit $X \sim$ Poisson de paramètre $\lambda$, alors

$$\varphi_X(t) = e^{\lambda(e^{\imath t}-1)}$$

en effet

$$\begin{aligned} \varphi(t) &= \sum_{k\in\mathbb{N}} e^{\imath tk}\frac{\lambda^k}{k!}e^{-\lambda} \\ &= \sum_{k\in\mathbb{N}} \frac{(\lambda e^{\imath t})^k}{k!}e^{-\lambda} \\ &= e^{\lambda e^{\imath t}-\lambda} \end{aligned}$$

Si, de plus, $Y \sim$ Poisson de paramètre $\mu$ est une variable indépendante de $X$ :

$$\begin{aligned} \varphi_{X+Y}(t) &= \varphi_X(t)\varphi_Y(t) \\ &= e^{\lambda(e^{\imath t}-1)}e^{\mu(e^{\imath t}-1)} \\ &= e^{(\lambda+\mu)(e^{\imath t}-1)} \end{aligned}$$

donc $X+Y \sim$ Poisson de paramètre $\lambda+\mu$.

6. Soit $X \sim G(\alpha,\beta)$ (loi Gamma de paramètres $\alpha > 0$ et $\beta > 0$ alors

$$\varphi_X(t) = (1-\imath t\beta)^{-\alpha}$$

en effet :

$$\varphi(t) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_{\mathbb{R}_+} e^{\imath tx}e^{-\frac{x}{\beta}}x^{\alpha-1}\mathrm{d}x$$

et

$$\begin{aligned} \varphi'(t) &= \frac{\imath}{\Gamma(\alpha)\beta^\alpha}\int_{\mathbb{R}_+} e^{\imath tx}e^{-\frac{x}{\beta}}x^{\alpha}\mathrm{d}x \\ &= \frac{-\imath\alpha\beta}{\Gamma(\alpha)\beta^\alpha(\imath t\beta-1)}\int e^{\imath tx}e^{-\frac{x}{\beta}}x^{\alpha-1}\mathrm{d}x \quad \text{(intégration par parties)} \\ &= \frac{\imath\alpha\beta}{(1-\imath t\beta)}\varphi(t) \end{aligned}$$

donc

$$\varphi(t) = C(1 - \imath t\beta)^{-\alpha} \quad \text{et} \quad C = 1 \quad \text{car} \quad \varphi(0) = 1$$

Si, de plus, $Y \sim G(\alpha', \beta)$ est une variable aléatoire indépendante de $X$ alors

$$\begin{aligned}\varphi_{X+Y}(t) &= \left[(1 - \imath t\beta)^{-\alpha}\right]\left[(1 - \imath t\beta)^{-\alpha'}\right] \\ &= (1 - \imath t\beta)^{-(\alpha+\alpha')}\end{aligned}$$

donc $X + Y \sim G(\alpha + \alpha', \beta)$.

### 3.3.7 Transformée de Laplace

Pour des variables aléatoires *positives*, la notion suivante est naturelle et utile.

### 3.3.8 Définition 2

Soit :

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$$

une variable aléatoire à valeurs $\mathbb{R}_+^d$ ; on pose, pour tout

$$s = \begin{pmatrix} s_1 \\ \vdots \\ s_d \end{pmatrix} \in \mathbb{R}_+^d$$

$$\begin{aligned}\Psi_X(s) &= E\{e^{-s'.X}\} \\ &= E\{e^{-\sum_{l=1}^{d} s_l.X_l}\}\end{aligned}$$

La fonction $\Psi_X(.)$ (définie sur $\mathbb{R}_+^d$) est appelée *Transformée de Laplace de la variable aléatoire* $X$. Noter que $\Psi_X(.)$ est bien définie ; une application facile du théorème de dérivation sous le signe intégral montre, de plus, que $\Psi_X(.)$ est de classe $\mathcal{C}^\infty$ sur $(\;]0, \infty[\;)^d = W$ et que, pour tout

$$s = \begin{pmatrix} s_1 \\ \vdots \\ s_d \end{pmatrix} \in W \quad : \quad \begin{aligned} &\frac{\partial}{\partial s_k}\Psi_X(s) = -E\{X_k e^{-s'.X}\} \\ &\frac{\partial^2}{\partial s_k \partial s_l}\Psi_X(s) = E\{X_k X_l e^{-s'.X}\} \quad k, l \in \{1, \ldots, d\}\end{aligned}$$

plus généralement :

$$\frac{\partial^{|\alpha|}}{\partial s_1^{\alpha_1}\partial s_2^{\alpha_2}\ldots\partial s_d^{\alpha_d}}\Psi_X(s) = (-1)^{|\alpha|}E\{X_1^{\alpha_1}.X_2^{\alpha_2}\ldots X_d^{\alpha_d}e^{-s'.X}\}$$

pour tout

$$\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_d \end{pmatrix} \in \mathbb{N}^d, \qquad |\alpha| = \sum_{l=1}^{d} \alpha_l$$

Si $X \in \mathcal{L}^p$ ($p \in \mathbb{N}^*$), on peut donc calculer, connaissant $\Psi_X(.)$, tous les moments de $X$ :

$$E\{X_1^{\alpha_1}.X_2^{\alpha_2} \dots X_d^{\alpha_d}\} = (-1)^{|\alpha|} \lim_{\substack{s \to 0 \\ s \in W}} \frac{\partial^{|\alpha|}}{\partial s_1^{\alpha_1} \dots \partial s_d^{\alpha_d}} \Psi_X(s)$$

où $\alpha_1, \dots, \alpha_d \in \mathbb{N}$ et $|\alpha| = \alpha_1 + \alpha_2 + \dots + \alpha_d \leq p$

### 3.3.9 Proposition 4

Soient $X$ et $\tilde{X}$ deux variables aléatoires à valeurs $\mathbb{R}_+^d$ telles que, pour tout $s \in \mathbb{R}_+^d$ :

$$\Psi_X(s) = \Psi_{\tilde{X}}(s)$$

alors $X$ et $\tilde{X}$ ont même loi : $P_X = P_{\tilde{X}}$

Idée de démonstration :

L'espace vectoriel $\mathcal{L}$ sur $\mathbb{R}$ des fonctions $f$ de la forme :

$$f(x) = \sum_{\text{finie}} C_i e^{-s_i'.x}$$

où $C_i \in \mathbb{R}$, $s_i \in (\ ]0, \infty[\ )^d$, $x \in \mathbb{R}_+^d$ est une algèbre de fonctions, séparant les points de $\mathbb{R}_+^d$, contenue dans $\mathcal{C}_0(\mathbb{R}_+^d, \mathbb{R})$ (espace des fonctions continues sur $\mathbb{R}_+^d$ tendant vers zéro à l'infini). On en conclut (théorème de Stone Weirstrass) qu'il est dense dans $\mathcal{C}_0(\mathbb{R}_+^d, \mathbb{R})$. L'hypothèse que $X$ et $\tilde{X}$ ont même transformée de Laplace entraîne que, pour tout $f \in \mathcal{L}$ :

$$\int_{\mathbb{R}^d} f \mathrm{d}P_X = \int_{\mathbb{R}^d} f \mathrm{d}P_{\tilde{X}}$$

On montre alors facilement, par un argument de densité et continuité, que $P_X = P_{\tilde{X}}$.

Corollaire 5 :

Soient $X_1, X_2, \dots, X_k$ des variables aléatoires à valeurs dans $\mathbb{R}_+^{d_1}$, $\mathbb{R}_+^{d_2}$, ..., $\mathbb{R}_+^{d_k}$ respectivement. Pour que $X_1, X_2, \dots, X_k$ soient indépendantes, il faut et il suffit que, quels que soient $s_1 \in \mathbb{R}_+^{d_1}$, ..., $s_k \in \mathbb{R}_+^{d_k}$ :

$$E\{e^{-\sum_{l=1}^{k} s_l'.X_l}\} = \prod_{l=1}^{k} E\{e^{-s_l'.X_l}\}$$

Démonstration :

Appliquer la proposition 4 (cf. le corollaire 1 du théorème 1) [exercice].

Exemple :

Soit $X$ une variable aléatoire réelle de loi $G(\alpha, \beta)$ ($\alpha > 0$, $\beta > 0$)

sa loi a pour densité :

$$\frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}.\mathbf{1}_{\mathbb{R}^+}(x)$$

et sa transformée de Laplace $\Psi(s)$ vaut

$$\begin{aligned}\Psi(s) &= \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\int_0^{+\infty} e^{-sx}e^{-\frac{x}{\beta}}x^{\alpha-1}\mathrm{d}x \\ &= \frac{\Gamma(\alpha)}{\beta^{\alpha}\Gamma(\alpha)(s+\frac{1}{\beta})^{\alpha}} \\ &= \frac{1}{(1+\beta s)^{\alpha}} \qquad (s \geq 0)\end{aligned}$$

Et on a :

$$\Psi^{'}(s) = -\alpha\beta(1+\beta s)^{-(\alpha+1)}, \qquad E(X) = -\Psi^{'}(0) = \alpha\beta$$

$$\Psi^{"}(s) = \alpha(\alpha+1)\beta^2(1+\beta s)^{-(\alpha+2)}, \qquad E(X^2) = +\Psi^{"}(0) = \alpha^2\beta^2 + \alpha\beta^2$$

$$\sigma^2(X) = E(X^2) - (E(X))^2 = \alpha\beta^2$$

Enfin on retrouve que si $X$ et $Y$ sont indépendantes de loi $G(\alpha,\beta)$ et $G(\alpha^{'},\beta)$ respectivement, alors $X+Y$ suit une loi $G(\alpha+\alpha^{'},\beta)$, il suffit d'appliquer la proposition suivante.

### 3.3.10 Proposition 6

La transformée de Laplace de la somme de deux variables aléatoires positives indépendantes est le produit des transformées de Laplace.
Démonstration :
Immédiate (cf. la proposition 2).

### 3.3.11 Fonctions génératrices

Pour des variables aléatoires *entières* positives, la notion de fonction génératrice est très utile.

### 3.3.12 Définition 3

Soit

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$$

une variable aléatoire à valeurs $\mathbb{N}^d$,
on pose, pour tout

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_d \end{pmatrix} \in (\ ]0,1[\ )^d \qquad g_X(u) = E\{u_1^{X_1}.u_2^{X_2}\dots u_d^{X_d}\}$$

La fonction $g_X(.)$ (définie sur le cube $(\ ]0,1[\ )^d = C$) est appelée *fonction génératrice de la variable aléatoire $X$*.
Remarque : Pour tout

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_d \end{pmatrix} \in C$$

il existe un unique

$$s = \begin{pmatrix} s_1 \\ \vdots \\ s_d \end{pmatrix} \in (\ ]0, \infty[\ )^d$$

tel que

$$u_1 = e^{-s_1}, \ldots, u_d = e^{-s_d}$$

et on voit que

$$g_X(u) = E\{e^{-(s_1X_1+\ldots+s_dX_d)}\} = \Psi_X(s)$$

Les propriétés de la fonction génératrice se déduisent donc des propriétés de la transformée de Laplace ; en particulier, on voit que $g_X$ est de classe $\mathcal{C}^\infty$ sur le cube $C = (\ ]0, 1[\ )^d$. Plus précisément, $g_X$ est une série entière en $u_1, u_2, \ldots, u_d$ :

$$g_X(u) = \sum_{n_1,n_2,\ldots,n_d \in \mathbb{N}} u_1^{n_1}.u_2^{n_2} \ldots u_d^{n_d} P(X_1 = n_1, X_2 = n_2, \ldots, X_d = n_d)$$

La connaissance de $g_X(.)$ détermine entièrement la loi du vecteur :

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$$

pour tous $n_1, n_2, \ldots, n_d \in \mathbb{N}$ :

$$\frac{\partial^{n_1+\ldots+n_d}}{\partial u_1^{n_1} \partial u_2^{n_2} \ldots \partial u_d^{n_d}} g_X(0) = n_1!n_2! \ldots n_d! P(X_1 = n_1, X_2 = n_2, \ldots, X_d = n_d)$$

de plus :

$$\frac{\partial^{n_1+\ldots+n_d}}{\partial u_1^{n_1} \partial u_2^{n_2} \ldots \partial u_d^{n_d}} g_X(1) = E\{(X_1(X_1-1)\ldots(X_1-n_1+1))\ldots(X_d(X_d-1)\ldots(X_d-n_d+1))\}$$

si $E\{||X||^{n_1+\ldots+n_d}\} < \infty$ ce qui permet de calculer, connaissant $g_X(.)$, tous les moments de $X$ lorsqu'ils existent. Lorsque $d = 1$, on voit que, si $X \in \mathcal{L}^2$

$$E\{X\} = \lim_{u \to 1} g^{'}(u) = g^{'}(1), \qquad E\{X(X-1)\} = \lim_{u \to 1} g^{"}(u) = g^{"}(1)$$

donc

$$\mathrm{Var}(X) = g^{"}(1) + g^{'}(1) - \left(g^{'}(1)\right)^2$$

### 3.3.13 Proposition 7

Soit

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$$

une variable aléatoire à valeurs $\mathbb{N}^d$. Alors les composantes $X_1, X_2, \ldots, X_d$ sont indépendantes si et seulement si, pour tout

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_d \end{pmatrix} \in (\ ]0, 1[\ )^d$$

on a :

$$g_X(u) = E\{u_1^{X_1} \dots u_d^{X_d}\} = E\{u_1^{X_1}\} \dots E\{u_d^{X_d}\} = \prod_{l=1}^{d} g_{X_l}(u_l) \quad (1)$$

<u>Démonstration</u> :
Cf. le corollaire 5. Remarquer aussi que l'égalité (1) signifie que $\forall\ u$ :

$$\sum_{n_1,\dots,n_d \in \mathbb{N}} u_1^{n_1} \dots u_d^{n_d} P\{X_1 = n_1, \dots, X_d = n_d\} = \prod_{l=1}^{d} \Big\{ \sum_{n \in \mathbb{N}} u^n P(X_l = n) \Big\}$$

$$= \sum_{n_1,\dots,n_d \in \mathbb{N}} u_1^{n_1} \dots u_d^{n_d} P\{X_1 = n_1\} \dots P\{X_d = n_d\}$$

ou bien que :

$$\forall\ n_1, \dots, n_d \in \mathbb{N} \qquad P\{X_1 = n_1, \dots, X_d = n_d\} = \prod_{l=1}^{d} P\{X_l = n_l\}$$

## 3.4 Vecteurs aléatoires gaussiens

On dit qu'une variable aléatoire réelle $\xi$ est *gaussienne* si :
ou bien $\xi$ a une densité de la forme :

$$\frac{2}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-m)^2} \quad \text{où} \quad \sigma^2 > 0 \quad \text{et} \quad m \in \mathbb{R} \quad (\xi \sim N_1(m, \sigma^2))$$

on bien $\xi$ est presque sûrement constante :

$$\xi = m \quad (m \in \mathbb{R}) \quad (P_\xi = \delta_m; \quad \xi \sim N_1(m, 0))$$

### 3.4.1 Définition 1

Soit

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$$

un vecteur aléatoire à valeurs $\mathbb{R}^d$, on dit que $X$ est *un vecteur aléatoire gaussien* si, pour tout

$$t = \begin{pmatrix} t_1 \\ \vdots \\ t_d \end{pmatrix} \in \mathbb{R}^d$$

la variable aléatoire réelle

$$t^{'}.X = \sum_{l=1}^{d} t_l X_l = < t, X >$$

est gaussienne.

Exemples :

1. Soient $X_1, X_2, \ldots, X_d$ $d$ variables aléatoires réelles gaussiennes *indépendantes* alors le vecteur aléatoire
$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$$
est gaussien ; en effet, pour tous
$$t = \begin{pmatrix} t_1 \\ \vdots \\ t_d \end{pmatrix} \in \mathbb{R}^d$$
et $u \in \mathbb{R}$ :
$$\begin{aligned} \varphi_{<t,X>}(u) &= E\{e^{\imath u \sum_{l=1}^d t_l X_l}\} \\ &= \prod_{l=1}^d E\{e^{\imath u t_l X_l}\} \\ &= \prod_{l=1}^d e^{(\imath u t_l m_l - \frac{1}{2}u^2 t_l^2 \sigma_l^2)} \quad \text{si} \quad X_l \sim N_1(m_l, \sigma_l) \\ &= e^{\imath u \sum_{l=1}^d t_l m_l - \frac{1}{2}u^2 \sum_{l=1}^d t_l^2 \sigma_l^2} \end{aligned}$$
donc
$$< t, X > \sim N_1\left(\sum_l t_l m_l, \sum_l t_l^2 \sigma_l^2\right)$$
2. Soit
$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$$
un vecteur gaussien et soient $B$ une matrice $r \times d$, $m \in \mathbb{R}^r$ (déterministes) alors le vecteur aléatoire $Y = m + B.X$, à valeurs $\mathbb{R}^r$ est gaussien ; car toute combinaison linéaire des composantes $Y_1, \ldots, Y_r$ de $Y$ est une combinaison affine des composantes de $X$ qui, par hypothèse, suit une loi gaussienne sur $\mathbb{R}$.

Remarque :
Soit
$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$$
un vecteur gaussien, alors, pour tout $l \in \{1, \ldots, d\}$, $X_l$ suit une loi gaussienne mais la réciproque est fausse en général ; contre exemple : si $X_1 \sim N_1(0,1)$ , posons $X_2 = X_1.\mathbf{1}_{|X_1| \leq a} - X_1.\mathbf{1}_{|X_1| > a}$ où $a > 0$ est fixé. Alors $X_2 \sim N_1(0,1)$ (vérification facile) mais
$$X_1 + X_2 = 2X_1.\mathbf{1}_{|X_1| \leq a} \in [-2a, +2a] \quad \text{presque sûrement}$$

donc $X_1+X_2$ est une variable aléatoire réelle non constante telle que $P\{X_1+X_2>2a\}=0$ ; $X_1+X_2$ ne peut être gaussienne et le vecteur $X=\binom{X_1}{X_2}$ n'est pas gaussien.

### 3.4.2 Proposition 1

Soit

$$X=\begin{pmatrix}X_1\\ \vdots\\ X_d\end{pmatrix}$$

un vecteur gaussien à valeurs $\mathbb{R}^d$, alors $X\in\mathcal{L}^p$ pour tout $1\leq p<\infty$. Posons :

$$m=E(X)=\begin{pmatrix}E(X_1)\\ \vdots\\ E(X_d)\end{pmatrix}\qquad D=D(X)=(\mathrm{Cov}(X_l,X_k))_{l,k=1,\dots,d}$$

la fonction caractéristique de $X$ est donnée par :

$$\varphi_X(t)=e^{\imath t'm-\frac{1}{2}t'Dt}\quad\text{pour tout}\quad t=\begin{pmatrix}t_1\\ \vdots\\ t_d\end{pmatrix}\in\mathbb{R}^d$$

Il en résulte que la loi de $X$ est entièrement déterminée par sa moyenne $m$ et sa matrice de dispersion $D$.
Démonstration :
Pour montrer que $E\{||X||^p\}<\infty$ $(1\leq p<\infty)$ il suffit de vérifier que

$$E\{|X_l|^p\}<\infty,\quad l=1,\dots,d$$

c'est évident si $X_l$ est constante sinon $X_l\sim N_1(m_l,\sigma_l)$ $(\sigma_l>0)$ et alors

$$E\{|X_l|^p\}=\frac{1}{\sqrt{2\pi\sigma_l^2}}\int_{\mathbb{R}}|x|^p e^{-\frac{1}{2\sigma_l^2}(x-m_l)^2}\mathrm{d}x<\infty$$

Posons donc $m=E(X)$, $D=D(X)$ ; on sait que pour tout $t\in\mathbb{R}^d$, la variable aléatoire réelle $t'X=<t,X>$ est gaussienne de moyenne $t'.m=<t,m>$ et de variance $t'.D.t=<t,DT>$ ; donc, pour tout $u\in\mathbb{R}$ :

$$\varphi_{<t,X>}(u)=E\{e^{\imath u<t,X>}\}=e^{\imath u<t,X>-\frac{1}{2}u^2<t,Dt>}$$

en particulier :

$$\varphi_X(t)=E\{e^{\imath<t,X>}\}=e^{\imath<t,m>-\frac{1}{2}<t,Dt>}$$

On notera $N_d(m,D)$ la loi de tout vecteur gaussien à valeurs $\mathbb{R}^d$, de moyenne $m$ et de matrice de dispersion $D$. Il reste à montrer qu'il en existe une lorsque $m$ et $D$ sont données à l'avance.

### 3.4.3 Lemme 1

Soit $D$ une matrice $d \times d$ symétrique, positive ou nulle, de rang $1 \leq r \leq d$. Il existe alors une matrice $B$, $d \times r$ de rang $r$ telle que $B.B' = D$.

Démonstration :

Soient $\lambda_1 > 0, \lambda_2 > 0, \ldots, \lambda_r > 0, \lambda_{r+1} = 0, \ldots, \lambda_d = 0$ les valeurs propres de $D$. On sait (résultat d'algèbre linéaire), qu'il existe une matrice *orthogonale* $O$ telle que $ODO' = \Delta$ où

$$\Delta = \begin{pmatrix} \lambda_1 & \ldots & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_r & \ldots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & \ldots & 0 \end{pmatrix} \qquad (O' = O^{-1})$$

Soit $\Delta_1$ la matrice $d \times r$, de rang $r$, donnée par :

$$\Delta_1 = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \ldots & 0 \\ 0 & \sqrt{\lambda_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sqrt{\lambda_r} \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix}$$

on a $\Delta = \Delta_1.\Delta_1^{'}$ et $O.D.O' = \Delta_1.\Delta_1^{'}$ donc $D = O'.\Delta_1.\Delta_1^{'}.O = B.B'$ où $B = O'.\Delta_1$.

### 3.4.4 Proposition 2

Soient $m \in \mathbb{R}^d$, $D$ une matrice symétrique positive ou nulle, $d \times d$ de rang $r$, $B$ une matrice $d \times r$ telle que $B.B' = D$ et

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_r \end{pmatrix}$$

un vecteur aléatoire gaussien à valeurs $\mathbb{R}^r$ de loi $N_r(0, I_r)$, alors $Y = m + B.X$ suit la loi $N_d(m, D)$.

Démonstration :

On sait que $B$ existe d'après le lemme 1 et que $X_1, X_2, \ldots, X_r$ sont $r$ variables aléatoires réelles indépendantes de même loi de densité

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

donc $Y = m + B.X$ est un vecteur gaussien à valeurs $\mathbb{R}^d$ (cf. l'exemple 2), de plus :

$$E(Y) = m + BE(X) = m, \qquad D(Y) = B.I_r.B' = B.B' = D$$

On dira que $X \sim N_d(m, D)$ est non dégénérée si $D$ est inversible : $\det(D) \neq 0$.

### 3.4.5 Proposition 3

Si $X \sim N_d(m, D)$ avec $\det(D) \neq 0$, la loi de $X$ a pour densité sur $\mathbb{R}^d$ :

$$(2\pi)^{-\frac{d}{2}} \left(\det(D)\right)^{-\frac{1}{2}} e^{-\frac{1}{2}(x-m)'.D^{-1}.(x-m)}$$

Démonstration :
Soit $Y \sim N_d(0, I_d)$ et soit $B$, matrice $d \times d$ telle que $B.B' = D$, $(\det(B) \neq 0)$ ; on sait que $X$ a même loi que $m + B.Y$ ; donc, si $\varphi : \mathbb{R}^d \to \mathbb{R}$ est mesurable positive ou nulle :

$$E\{\varphi(X)\} = E\{\varphi(m + BY)\} = (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} \varphi(m + By) e^{-\frac{1}{2}|y|^2} \mathrm{d}y$$

on pose $m + By = x$, $y = B^{-1}(x - m)$, $\frac{D(y)}{D(x)} = \det(B^{-1})$ ce qui donne :

$$E\{\varphi(X)\} = (2\pi)^{-\frac{d}{2}} \det(B^{-1}) \int_{\mathbb{R}^d} \varphi(x) e^{-\frac{1}{2}(x-m)'.(B^{-1})'.B^{-1}.(x-m)} \mathrm{d}x$$

comme $B.B' = D$, $\det(D) = \left(\det(B)\right)^2$ et $D^{-1} = (B.B')^{-1} = (B^{-1})'.B^{-1}$, on obtient l'expression cherchée.
Une propriété importante des variables aléatoires gaussiennes est que, pour une variable aléatoire gaussienne centrée, l'indépendance des composantes équivaut à leur orthogonalité dans $\mathcal{L}^2$.

### 3.4.6 Proposition 4

Soit

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} \sim N_d(m, D)$$

alors les variables aléatoires $X_1, \ldots, X_d$ sont indépendantes si et seulement si $D$ est diagonale.
Démonstration :
On a déjà vu que si $X_1, \ldots, X_d$ sont indépendantes alors $D$ est diagonale. Réciproquement, si

$$t = \begin{pmatrix} t_1 \\ \vdots \\ t_d \end{pmatrix} \in \mathbb{R}^d \quad \text{et} \quad \sigma_l^2 (l = 1, \ldots, d)$$

désignent les termes diagonaux de $D$, on a :

$$\begin{aligned} \varphi_X(t) &= e^{\imath <t,m> - \frac{1}{2} <t,Dt>} \\ &= e^{\imath(t_1 m_1 + \ldots + t_d m_d) - \frac{1}{2}(t_1^2 \sigma_1^2 + \ldots + t_d^2 \sigma_d^2)} \\ &= \prod_{l=1}^{d} e^{\imath t_l m_l - \frac{1}{2} t_l^2 \sigma_l^2} \\ &= \prod_{l=1}^{d} \varphi_{X_l}(t_l) \end{aligned}$$

ce qui implique l'indépendance des composantes $X_1, \ldots, X_d$ d'après le corollaire 1 du Théorème 1 (III).
Généralisons ce résultat.

### 3.4.7 Proposition 5

Soit

$$X_k = \begin{pmatrix} X_k^1 \\ \vdots \\ X_k^{p_k} \end{pmatrix} \qquad k = 1, \ldots, n$$

des variables aléatoires à valeurs $\mathbb{R}^{p_k}$ telles que le vecteur

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

à valeurs $\mathbb{R}^d = \mathbb{R}^{p_1} \times \ldots \times \mathbb{R}^{p_n}$, $(d = p_1 + \ldots + p_n)$ soit gaussien. Les variables aléatoires $X_1, X_2, \ldots, X_n$ sont indépendantes si et seulement si, quels que soient $k \neq l$ : $\mathrm{Cov}(X_k^r, X_l^s) = 0$ pour tous $r \in \{1, \ldots, p_k\}$, $s \in \{1, \ldots, p_l\}$.
Démonstration :
D'après le corollaire 1 du Théorème 1 (III), il suffit de montrer que quels que soient $u_k \in \mathbb{R}^{p_k}$ :

$$E\{e^{\imath(u_1' X_1 + \ldots + u_n' X_n)}\} = \prod_{k=1}^{n} E\{^{\imath u_k' X_k}\}$$

soit encore l'indépendance des variables aléatoires réelles $u_1' X_1, \ldots, u_n' X_n$. Comme le vecteur $(u_1' X_1, \ldots, u_n' X_n)$ est gaussien (fonction linéaire de $X$), il suffit de vérifier que si $k \neq l$ : $\mathrm{Cov}(u_k' X_k, u_l' X_l) = 0$ mais :

$$\mathrm{Cov}(u_k' X_k, u_l' X_l) = \sum_{r,s} u_k^r . u_l^s \, \mathrm{Cov}(X_k^r, X_l^s) = 0$$

### 3.4.8 Appendice

Démonstration du théorème 1
On note, pour $\sigma > 0$,

$$g_\sigma(u) = (2\pi\sigma^2)^{-\frac{d}{2}} e^{-\frac{1}{2\sigma^2}|u|^2}, \qquad u \in \mathbb{R}^d$$

on sait que

$$e^{-\frac{1}{2}|t|^2} = \int_{\mathbb{R}^d} e^{\imath t' u} g_1(u) \mathrm{d}u \qquad t \in \mathbb{R}^d$$

d'où :

$$\begin{aligned} e^{-\frac{1}{2\sigma^2}|x-y|^2} &= \int_{\mathbb{R}^d} e^{\imath \frac{(x-y)'.u}{\sigma}} g_1(u) \mathrm{d}u \\ &= \sigma^d \int_{\mathbb{R}^d} g_1(\sigma u) e^{\imath (x-y)'.u} \mathrm{d}u \end{aligned}$$

Lemme 1 : Soit $\mu$ une mesure bornée sur $\mathbb{R}^d$, alors :

$$\int g_\sigma(x-y)\mu(\mathrm{d}x) = (2\pi)^{-\frac{d}{2}} \int \hat{\mu}(u) g_1(\sigma u) e^{-\imath y' u} \mathrm{d}u$$

Démonstration :

$$\begin{aligned}\int g_\sigma(x-y)\mu(\mathrm{d}x) &= (2\pi)^{-\frac{d}{2}} \int \mu(\mathrm{d}x) \int g_1(\sigma u) e^{+\imath (x-y)' u} \mathrm{d}u \\ &= (2\pi)^{-\frac{d}{2}} \int \left[\int e^{+\imath x' u} \mu(\mathrm{d}x)\right] g_1(\sigma u) e^{-\imath y' u} \mathrm{d}u\end{aligned}$$

d'après le théorème de Fubini, puisque

$$|g_1(\sigma u) e^{-\imath (x-y)' u}| \leq e^{-\frac{\sigma^2 |u|^2}{2}}$$

qui est intégrable pour $\mu(\mathrm{d}x)\mathrm{d}u$

Lemme 2 :

L'espace vectoriel engendré par les applications $x \longmapsto g_\sigma(x-y)$ où $\sigma$ parcourt $\mathbb{R}_+^*$ et $y \in \mathbb{R}^d$ est dense dans $\mathcal{C}_0(\mathbb{R}^d)$ (espace des fonctions continues sur $\mathbb{R}^d$ qui tendent vers zéro à l'infini).

Démonstration :

On vérifie, en effet, facilement que cet espace est une sous algèbre de $\mathcal{C}_0(\mathbb{R}^d)$ qui sépare les points (y compris le point à l'infini). On conclut en appliquant le théorème de Stone-Wierstrass.

Les lemmes 1 et 2 impliquent le théorème 1 car, d'après le premier, si $\mu_1$ et $\mu_2$ désignent les lois de $X$ et $\tilde{X}$ respectivement, $\int f \mathrm{d}\mu_1$ et $\int f \mathrm{d}\mu_2$ coïncident pour $f(x) = g_\sigma(x-y)$ donc, d'après le second, par densité et continuité, pour toute $f \in \mathcal{C}_0$.

# Chapitre 4

# Lois des grands nombres et théorème central limite

## 4.1 Différents modes de convergence des variables aléatoires

Soient $X_n$ ($n \in \mathbb{N}$) et $X_\infty$ des variables aléatoires définies sur $(\Omega, \mathcal{A}, P)$ et à valeurs $\mathbb{R}^d$.

1. On dit que $X_n \to X_\infty$ quand $n \to \infty$ *presque sûrement* s'il existe $\mathcal{N} \in \mathcal{A}$ tel que $P(\mathcal{N}) = 0$ et $\forall\, \omega \in \Omega$, $\omega \notin \mathcal{N} \Rightarrow X_n(\omega) \to X_\infty(\omega)$ quand $n \to \infty$ dans $\mathbb{R}^d$.
2. $X_n \to X_\infty$ quand $n \to \infty$ *dans* $\mathcal{L}^p$ ($1 \leq p < \infty$) si pour tout n, $X_n$ et $X_\infty$ appartiennent à $\mathcal{L}^p$ et $E\{|X_n - X_\infty|^p\} \to 0$ quand $n \to \infty$.
3. $X_n \to X_\infty$ quand $n \to \infty$ *en probabilité* si pour tout $\varepsilon > 0$, $P\{|X_n - X_\infty| > \varepsilon\} \to 0$ quand $n \to \infty$.
4. $X_n \to X_\infty$ quand $n \to \infty$ en loi si pour tout $f \in \mathcal{C}_b(\mathbb{R}^d) = \{$applications continues et bornées de $\mathbb{R}^d$ dans $\mathbb{R}\}$, on a $E\{f(X_n)\} \to E\{f(X_\infty)\}$ quand $n \to \infty$, en d'autres termes :

$$\int_{\mathbb{R}^d} f(x) P_{X_n}(\mathrm{d}x) \to \int_{\mathbb{R}^d} f(x) P_{X_\infty}(\mathrm{d}x) \quad \text{quand} \quad n \to \infty$$

   On dit alors que les lois $P_{X_n}$ *convergent étroitement* vers $P_{X_\infty}$ quand $n \to \infty$.

### 4.1.1 Proposition 1

On a les relations suivantes entre ces différents modes de convergence :
1) $\Rightarrow$ 3) et 2) $\Rightarrow$ 3) $\Rightarrow$ 4)
les implications en sens inverse (sans conditions supplémentaires) étant fausses en général.

Démonstration :

1. Cas 1) $\Rightarrow$ 3) ci dessus. On suppose que $X_n \to X_\infty$ presque sûrement.
   Soient $\varepsilon > 0$ et $A_n = \{|X_n - X_\infty| > \varepsilon\}$. On a :

   $\overline{\lim} A_n = \{\omega \in \Omega \quad \text{tel que} \quad |X_n(\omega) - X_\infty(\omega)| > \varepsilon \quad \text{pour une infinité d'entiers } n\}$

   On voit que :

   $$\overline{\lim} A_n \subseteq \{\omega : X_n(\omega) \not\to X_\infty(\omega)\} \subseteq \mathcal{N} \quad \text{où} \quad P(\mathcal{N}) = 0$$

   et

   $$0 = P(\overline{\lim} A_n) = P\{\bigcap_n (\bigcup_{m \geq n} A_m)\} = \lim_{n\to\infty} P\{\bigcup_{m\geq n} A_m\} \geq \overline{\lim}_{n\to\infty} P(A_n) \geq 0$$

   donc

   $$\lim_{n\to\infty} P\{A_n\} = \lim_{n\to\infty} P\{|X_n - X_\infty| > \varepsilon\} = 0$$

2. Cas 2) $\Rightarrow$ 3) ci dessus. Soit $\varepsilon > 0$. On a :

   $$P\{|X_n - X_\infty| > \varepsilon\} \leq \frac{1}{\varepsilon^p} E\{|X_n - X_\infty|^p\} \to 0 \quad \text{quand} \quad n \to \infty$$

3. Cas 3) $\Rightarrow$ 4) ci dessus. On suppose que $\forall\, \varepsilon > 0$ $P\{|X_n - X_\infty| > \varepsilon\} \to 0$.
   Premier cas : $f \in \mathcal{C}_k(\mathbb{R}^d) = \{$fonctions continues à support compact dans $\mathbb{R}^d\}$ ; on sait alors que $f$ est uniformément continue : pour tout $\varepsilon > 0$, il existe $\alpha_\varepsilon > 0$ tel que :

   $$\forall\, x, y \in \mathbb{R}^d, \quad |x - y| \leq \alpha_\varepsilon \Rightarrow |f(x) - f(y)| < \varepsilon$$

   donc

   $$\begin{aligned}
   &|E\{f(X_n) - f(X_\infty)\}| \leq E\{|f(X_n) - f(X_\infty)|\} \\
   &\leq E\{|f(X_n) - f(X_\infty)|.\mathbf{1}_{(|X_n - X_\infty| \leq \alpha_\varepsilon)}\} + E\{|f(X_n) - f(X_\infty)|.\mathbf{1}_{(|X_n - X_\infty| > \alpha_\varepsilon)}\} \\
   &\leq \varepsilon P\{|X_n - X_\infty| \leq \alpha_\varepsilon\} + 2||f||\ P\{|X_n - X_\infty| > \alpha_\varepsilon\} \\
   &\leq \varepsilon + 2||f||\ P\{|X_n - X_\infty| > \alpha_\varepsilon\} \\
   &\leq 2\varepsilon
   \end{aligned}$$

   pour tout $n$ assez grand ($||f|| = \sup_x |f(x)| < \infty$).
   Cas général Soient $\varepsilon > 0$ et $C_\varepsilon > 0$ tels que $P\{|X_\infty| > C_\varepsilon\} < \varepsilon$. Il existe alors $\varphi \in \mathcal{C}_k(\mathbb{R}^d)$ telle que pour tout $x \in \mathbb{R}^d$ : $0 \leq \varphi(x) \leq 1$ et $(|x| \leq C_\varepsilon) \Rightarrow \varphi(x) = 1$.
   Soit $f \in \mathcal{C}_k(\mathbb{R}^d)$. On a :

   $$\begin{aligned}
   &|E\{f(X_n)\} - E\{f(X_\infty)\}| \\
   &\leq E\{|f(X_n) - f\varphi(X_n)|\} + E\{|f\varphi(X_n) - f\varphi(X_\infty)|\} + E\{|f\varphi(X_\infty) - f(X_\infty)|\} \\
   &\leq ||f|| \Big\{ (1 - E\{\varphi(X_n)\}) + (1 - E\{\varphi(X_\infty)\}) \Big\} + E\{|f\varphi(X_n) - f\varphi(X_\infty)|\} \ (*)
   \end{aligned}$$

   Or :

   $$0 \leq 1 - \varphi \leq \mathbf{1}_{(|x| > C_\varepsilon)}$$

$\varphi$ et $f\varphi$ appartiennent à $\mathcal{C}_k(\mathbb{R}^d)$ donc :

$$|1 - E\{\varphi(X_\infty)| \leq P\{|X_\infty| > C_\varepsilon\} < \varepsilon$$

et

$$\lim_{n\to\infty} |1 - E\{\varphi(X_n)| = |1 - E\{\varphi(X_\infty)| \leq \varepsilon$$

$$\lim_{n\to\infty} E\{|f\varphi(X_n) - f\varphi(X_\infty)|\} = 0$$

L'inégalité (*) montre alors que, pour tout $n$ assez grand :

$$|E\{f(X_n)\} - E\{f(X_\infty)\}| \leq ||f||(3\varepsilon) + \varepsilon$$

$\varepsilon > 0$ étant arbitraire, on voit que :

$$E\{f(X_n)\} \to E\{f(X_\infty)\}$$

### 4.1.2 Proposition 2

Si $X_n \to X_\infty$ en probabilité, il existe une sous suite $n_k \nearrow \infty$ telle que $X_{n_k} \to X_\infty$ presque sûrement.
Démonstration :
On suppose que :

$$\forall\ \varepsilon > 0 \quad : \quad P\{|X_n - X_\infty| > \varepsilon\} \to \infty \quad \text{quand} \quad n \to \infty$$

Il est facile de construire, par récurrence, une sous-suite $n_k \in \mathbb{N}$ telle que, pour tout $k \in \mathbb{N}$ :

$$P\left\{|X_{n_k} - X_\infty| > \frac{1}{2^k}\right\} < \frac{1}{2^k}$$

et $n_k < n_{k+1} \nearrow \infty$. Soit :

$$A_k = \left\{|X_{n_k} - X_\infty| > \frac{1}{2^k}\right\}$$

On a alors :

$$\sum_k P(A_k) < \sum_k \frac{1}{2^k} < \infty$$

donc

$$P\{\overline{\lim} A_k\} = 0$$

d'après le lemme de Borel-Cantelli.
Posons $\mathcal{N} = \overline{\lim} A_k$. On voit que $P(\mathcal{N}) = 0$ et que, pour tout $\omega \in \Omega$ :
$(\omega \notin \mathcal{N}) \Rightarrow \omega \notin A_k$ pour tout $k$ assez grand donc $(\omega \notin \mathcal{N}) \Rightarrow$ il existe un entier $K(\omega)$ tel que, pour tout $k \geq K(\omega)$

$$|X_{n_k}(\omega) - X_\infty(\omega)| \leq \frac{1}{2^k}$$

## 4.2 Lois des grands nombres

### 4.2.1 Théorème 1

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles, de carré intégrable, vérifiant les conditions suivantes :

1. $E\{X_n\} \to m$ quand $n \to \infty$, où $m \in \mathbb{R}$
2. $\forall\, l \neq k$ $\mathrm{Cov}(X_l, X_k) = 0$
3. $\sup_n \mathrm{Var}(X_n) < \infty$

alors

$$\overline{X_n} = \frac{X_1 + \ldots + X_n}{n} \to m \quad \text{quand} \quad n \to \infty$$

dans $\mathcal{L}^2$ et presque sûrement.

Remarque :
Si les variables aléatoires $X_n$ $(n \in \mathbb{N}^*)$ sont indépendantes, de même loi, de carré intégrable alors les conditions 1), 2) et 3) sont satisfaites.

Démonstration du théorème 1 :
Posons $X_n^{'} = X_n - E(X_n)$. On voit que $(X_n^{'})$ vérifie les conditions 2) et 3) et que $E(X_n^{'}) = 0$,

$$\overline{X_n} = \overline{X_n'} + \frac{E(X_1) + \ldots + E(X_n)}{n}$$

$$\frac{E(X_1) + \ldots + E(X_n)}{n} \to m \quad \text{quand} \quad n \to \infty$$

Il suffit de montrer que $\overline{X_n'} \to 0$ dans $\mathcal{L}^2$ et presque sûrement. Soit

$$C = \sup_n \mathrm{Var}(X_n)$$

On a, d'après 2) et 3) :

$$\begin{aligned} E\{(\overline{X_n'})^2\} &= \frac{1}{n^2}\mathrm{Var}\{X_1 + \ldots + X_n\} \\ &= \frac{1}{n^2}\left[\mathrm{Var}\{X_1\} + \ldots + \mathrm{Var}\{X_n\}\right] \\ &\leq \frac{C.n}{n^2} = \frac{C}{n} \to 0 \quad \text{quand} \quad n \to \infty \end{aligned}$$

donc $\overline{X_n'} \to 0$ dans $\mathcal{L}^2$. Pour tout $n$, posons $S_n = X_1^{'} + \ldots + X_n^{'}$. On vient de voir que :

$$E\Big\{\left(\frac{S_n}{n}\right)^2\Big\} \leq \frac{C}{n}$$

donc :

$$E\left\{\sum_{n\geq 1}\left(\frac{S_{n^2}}{n^2}\right)^2\right\} = \sum_{n\geq 1} E\left\{\left(\frac{S_{n^2}}{n^2}\right)^2\right\} \leq \sum_{n\geq 1}\frac{C}{n^2} < \infty$$

La variable aléatoire réelle

$$\sum_{n\geq 1}\left(\frac{S_{n^2}}{n^2}(\omega)\right)^2$$

est intégrable, donc *finie presque sûrement* et

$$\frac{S_{n^2}}{n^2}(\omega) \to 0 \quad \text{quand} \quad n \to \infty \quad \text{presque sûrement}$$

Pour tout $n \in \mathbb{N}^*$, soit $m_n \in \mathbb{N}^*$ tel que :

$$m_n^2 \leq n < (m_n + 1)^2$$

alors

$$1 \leq \frac{n}{m_n^2} < \left(1 + \frac{1}{m_n}\right)^2 \to 1 \quad \text{quand} \quad n \to \infty$$

et

$$0 \leq n - m_n^2 < 1 + 2m_n \leq 1 + 2\sqrt{n}$$

On a donc :

$$(1) \qquad \frac{S_{m_n^2}}{n} = \frac{S_{m_n^2}}{m_n^2} \times \frac{m_n^2}{n} \to 0 \quad \text{presque sûrement}$$

de plus

$$E\left\{\left(\frac{S_n - S_{m_n^2}}{n}\right)^2\right\} \leq \frac{1}{n^2} 2C.\sqrt{n} = \frac{2C}{n^{\frac{3}{2}}}$$

donc

$$E\left\{\sum_{n\geq 1}\left(\frac{S_n - S_{m_n^2}}{n}\right)^2\right\} \leq \sum_{n\geq 1} \frac{2C}{n^{\frac{3}{2}}} < \infty$$

et

$$(2) \qquad \frac{S_n - S_{m_n^2}}{n} \to 0 \quad \text{quand} \quad n \to \infty \quad \text{presque sûrement}$$

En regroupant (1) et (2), on voit que :

$$\frac{S_n}{n} = X_n^{'} \to 0 \quad \text{quand} \quad n \to \infty \quad \text{presque sûrement}$$

### 4.2.2 Théorème 2 (Loi forte des grands nombres)

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles *indépendantes*, *de même loi*, *intégrables*, alors :

$$\overline{X_n} = \frac{X_1 + X_2 + \ldots + X_n}{n} \to E(X_1) \quad \text{quand} \quad n \to \infty \quad \text{presque sûrement}$$

et dans $\mathcal{L}^1$ Nous admettrons la démonstration (un peu délicate) de ce théorème [Cf. : Breiman « Probability » Ed. SIAM]
<u>Corollaire</u> : <u>Théorème fondamental de la statistique</u>
<u>$(X_n)_{n\in\mathbb{N}}$</u> étant une suite de variables aléatoires indépendantes de même loi <u>inconnue</u> $\mu$, la loi des grands nombres montre que, pour toute fonction $f$ mesurable et bornée, on a

$$M_n(f, \omega) = \frac{f(X_1) + \ldots + f(X_n)}{n}(\omega) \to E\{f(X_1)\} = \int f(x)\mu(\mathrm{d}x) \quad \text{quand} \quad n \to \infty$$

presque sûrement.
On voit donc que la connaissance, pour tout $n$ assez grand, des moyennes empiriques (i.e. *accessibles à l'experience*) $M_n(f,\omega)$ fournit une bonne approximation de la véritable moyenne

$$\int f(x)\mu(\mathrm{d}x) = \mu(f)$$

Le théorème Central limite (convergence vers la loi de Gauss) permet d'estimer « l'erreur » $[M_n(f,\omega) - \mu(f)]$ ainsi que sa vitesse de convergence vers zéro. On a aussi :

### 4.2.3 Théorème 3 (Loi du logarithme itéré)

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes, de même loi et bornées, alors :

$$\overline{\lim}_{n\to\infty} \frac{(X_1 + \ldots + X_n) - nE(X_1)}{[2n\log(\log(n))]^{\frac{1}{2}}} = \sigma(X_1) \quad \text{presque sûrement}$$

$$\underline{\lim}_{n\to\infty} \frac{(X_1 + \ldots + X_n) - nE(X_1)}{[2n\log(\log(n))]^{\frac{1}{2}}} = -\sigma(X_1) \quad \text{presque sûrement}$$

Démonstration : Admise [Cf. : Breiman « Probability » Ed. SIAM]
Ce théorème montre que $\overline{X_n} - E(X_1)$ converge presque sûrement vers zéro « aussi vite » que

$$\left(\frac{2\log(\log(n))}{n}\right)^{\frac{1}{2}}$$

quand $n \to \infty$.

## 4.3 Convergence en loi

Soient $X_n (n \in \mathbb{N})$ et $X_\infty$ des variables aléatoires à valeurs $\mathbb{R}^d$. Si $X_n \to X_\infty$ en loi, alors pour tout $t \in \mathbb{R}^d$ :

$$\varphi_{X_n}(t) = E\{e^{\imath t X_n}\} \to \varphi_{X_\infty}(t) = E\{e^{\imath t X_\infty}\} \quad \text{quand} \quad n \to \infty$$

puisque

$$f(x) = e^{\imath t x} \quad \text{appartient à} \quad \mathcal{C}_b(\mathbb{R}^d)$$

Réciproquement on a le :

### 4.3.1 Théorème de Paul Lévy

Si, pour tout $t \in \mathbb{R}^d$, $\varphi_{X_n}(t) \to \varphi(t)$ (simplement) et si $\varphi$ est continue en zéro, alors $\varphi$ est la fonction caractéristique d'une variable aléatoire $X_\infty$ à valeurs $\mathbb{R}^d$ et $X_n \to X_\infty$ en loi. On va démontrer une propriété plus simple.

### 4.3.2 Théorème 4

Si, pour tout $t \in \mathbb{R}^d$, $\varphi_{X_n}(t) \to \varphi_{X_\infty}(t)$, alors $X_n \to X_\infty$ en loi.
Démonstration (succinte).
Soit

$$g_\sigma(u) = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^d}\ e^{-\frac{1}{2\sigma^2}|u|^2}$$

où $\sigma > 0$ et $u \in \mathbb{R}^d$. On sait [ Cf. l'appendice du chapitre 3 ] que :

$$\int g_\sigma(x-y)P_{X_n}(\mathrm{d}x) = \frac{1}{(2\pi)^{\frac{d}{2}}} \int \varphi_{X_n}(u) g_1(\sigma u) e^{-\imath uy}\,\mathrm{d}u$$

donc, puisque $\varphi_{X_n}(u) \to \varphi_{X_\infty}(u)$, on voit, grâce au théorème de Lebesgue, que :

$$\begin{aligned}(*)\ :\ \int g_\sigma(x-y)P_{X_n}(\mathrm{d}x) &\to \frac{1}{(2\pi)^{\frac{d}{2}}} \int \varphi_{X_\infty}(u) g_1(\sigma u) e^{-\imath uy}\mathrm{d}u \\ &= \int g_\sigma(x-y)P_{X_\infty}(\mathrm{d}x)\end{aligned}$$

Soit $\mathcal{L}$ le sous espace vectoriel de $\mathcal{C}_0(\mathbb{R}^d) = \{$fonctions continues qui tendent vers zéro à l'infini$\}$ engendré par les fonctions $x \longmapsto g_\sigma(x-y)$ où $\sigma > 0$ et $y \in \mathbb{R}^d$. Une application du théorème de Stone-Weirstrass montre que $\mathcal{L}$ est dense dans $\mathcal{C}_0(\mathbb{R}^d)$ et la propriété (*) entraîne, par continuité, que

$$\forall\ f \in \mathcal{C}_0(\mathbb{R}^d)\ :\ \int f(x)P_{X_n}(\mathrm{d}x) \to \int f(x)P_{X_\infty}(\mathrm{d}x)$$

On vérifie alors [Cf; la démonstration de la proposition 1 ], que la convergence précédente a lieu pour tout $f \in \mathcal{C}_b(\mathbb{R}^d)$

### 4.3.3 Proposition 3

Soient $X_n (n \in \mathbb{N})$ et $X_\infty$ des variables aléatoires à valeurs $\mathbb{R}^d$, alors si $X_n \to X_\infty$ en loi, on a pour tout $A \in \mathcal{B}_{\mathbb{R}^d}$, tel que $P\{X_\infty \in \partial A\} = 0$ ( où $\partial A$ désigne la frontière de A)

$$P\{X_n \in A\} \to P\{X_\infty \in A\} \quad \text{quand} \quad n \to \infty$$

En particulier, si les variables aléatoires $X_n$ et $X_\infty$ sont à valeurs réelles, pour tout $t \in \mathbb{R}$, tel que $P\{X_\infty = t\} = 0$ (i.e. en tout point de continuité de la fonction de répartition $F(t) = P\{X_\infty \leq t\}$), on a

$$P\{X_n \leq t\} \to P\{X_\infty \leq t\}$$

Réciproquement, si les fonctions de répartition $F_n(t) = P\{X_n \leq t\}$ convergent vers $P\{X_\infty \leq t\} = F(t)$ en tout point $t \in D$ où $D$ est dense dans $\mathbb{R}$, alors $X_n \to X_\infty$ en loi.
Démonstration : elle est admise.

### 4.3.4 Théorème 5 (Théorème Central Limite)

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes, de même loi, de carré intégrable, à valeurs dans $\mathbb{R}^d$. Posons $m = E(X_1) \in \mathbb{R}^d$, $D = D(X_1)$ est la matrice de dispersion de $X_1$ et $S_n = X_1 + X_2 + \ldots + X_n$, alors $\sqrt{n}\{\frac{S_n}{n} - m\}$ converge en loi, quand $n \to \infty$, vers la loi de Gauss $N_d(0, D)$. En particulier, si $d = 1$, en posant $\sigma^2 = \mathrm{Var}(X_1)$ (on suppose $\sigma^2 > 0$) :

$$P\Big\{\alpha \leq \frac{\sqrt{n}}{\sigma}\{\frac{S_n}{n} - m\} \leq \beta\Big\} \to \frac{1}{\sqrt{2\pi}}\int_\alpha^\beta e^{-\frac{x^2}{2}}\,\mathrm{d}x \quad \text{quand} \quad n \to \infty$$

pour tous $\alpha \leq \beta$ dans $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$.
Démonstration (succinte) :
Soit

$$\Phi(t) = E\{e^{\imath t(X_1 - m)}\} \qquad (t \in \mathbb{R}^d)$$

alors $\Phi$ est de classe $\mathcal{C}^2$ et :

$$\frac{\partial \Phi}{\partial t_l}(0) = 0, \quad \frac{\partial^2 \Phi}{\partial t_l \partial t_k}(0) = -D_{l,k}, \quad \text{pour tous} \quad l, k \in \{1, \ldots, d\}$$

d'où le développement limité :

$$\Phi(t) = 1 - \frac{1}{2} < t, Dt > + |t|^2 \varepsilon(t) \quad \text{où} \quad \varepsilon(t) \to 0 \quad \text{quand} \quad t \to 0$$

On a donc, pour tout $t \in \mathbb{R}^d$ :

$$\begin{aligned}
\Phi_n(t) &= E\big\{e^{\imath t(\frac{S_n - nm}{\sqrt{n}})}\big\} \\
&= \left[\Phi\left(\frac{t}{\sqrt{n}}\right)\right]^n \\
&= \left[1 - \frac{1}{2n} < t, Dt > + \frac{|t|^2}{n}\varepsilon\left(\frac{t}{\sqrt{n}}\right)\right]^n \\
&= e^{n\log(1 - \frac{1}{2n} < t, Dt > + \frac{|t|^2}{n}\varepsilon(\frac{t}{\sqrt{n}}))} \\
&\to e^{-\frac{1}{2}<t,Dt>} \quad \text{quand} \quad n \to \infty \quad \text{puisque} \quad \varepsilon\left(\frac{t}{\sqrt{n}}\right) \to 0
\end{aligned}$$

(le log est bien défini pour tout n assez grand).
On sait de plus que $t \longmapsto e^{-\frac{1}{2}<t,Dt>}$ est la fonction caractéristique de toute variable aléatoire de loi $N_d(0, D)$ ; il suffit donc d'appliquer le théorème 4.

# Chapitre 5

# Espérance conditionnelle

Soient $(\Omega, \mathcal{a}, P)$ un espace de probabilité associé à une expérience aléatoire et $\mathcal{B}$ une sous-tribu de $\mathcal{a}$. $\mathcal{a}$ est la tribu de tous les événements « possibles » et $\mathcal{B}$ représente une famille d'événements remarquables : par exemple famille des événements du « passé », du « futur », du « présent » ou bien famille des événements qui s'expriment en fonction d'une variable aléatoire $T$ fixée de $(\Omega, \mathcal{a}, P)$ dans un espace $(F, \mathcal{F})$ :

$$\mathcal{B} = \sigma(T) = \{T^{-1}(S); \quad S \in \mathcal{F}\} \quad \subseteq \mathcal{a}$$

Rappelons que si $Y$ est une variable aléatoire réelle, on dira que $Y$ est $\mathcal{B}$-mesurable si :

$$\sigma(Y) = \{Y^{-1}(C); \quad C \text{ borélien de } \mathbb{R}\} \quad \subseteq \mathcal{B}$$

en d'autres termes, $Y$ « ne dépend que des événements de $\mathcal{B}$ ».
$X$ étant une variable aléatoire réelle donnée sur $(\Omega, \mathcal{a}, P)$ un problème naturel consiste à rechercher la « meilleur approximation », en un sens à préciser, de $X$ par une variable aléatoire réelle $Y$ $\mathcal{B}$-mesurable.

## 5.1 Théorème 1 (fondamental)

Soient $X$ une variable aléatoire réelle intégrable définie sur $(\Omega, \mathcal{a}, P)$ et $\mathcal{B}$ une sous-tribu de $\mathcal{a}$. Il existe alors une variable aléatoire réelle $Y$ $\mathcal{B}$-mesurable et intégrable, unique à l'égalité presque sûre près, telle que :

$$(*) \quad \text{pour tout} \quad B \in \mathcal{B} \quad : \quad \int_B X \mathrm{d}P = \int_B Y \mathrm{d}P$$

On dira que $Y$ est (une version de) l'espérance conditionnelle de $X$ sachant $\mathcal{B}$ et on notera : $Y = E\{X/\mathcal{B}\}$ presque sûrement.
Nous admettrons (pour le moment) l'existence de $Y$, basée sur le théorème de Radon-Nicodym, [Cf. W. Rudin « Real and Complex Analysis » MC Graw H (ed)]. Montrons l'unicité de $Y$, à l'égalité presque sûre près.

## 5.2 Proposition 1

Soit $Z$ une variable aléatoire réelle $\mathcal{B}$-mesurable et intégrable alors $Z \geq 0$ presque sûrement si et seulement si :

$$(1) \quad \text{pour tout} \quad B \in \mathcal{B} \quad : \quad \int_B Z \mathrm{d}P \geq 0$$

Démonstration :
Il suffit, évidemment, de montrer que (1) $\Rightarrow Z \geq 0$ presque sûrement. Soient $\varepsilon > 0$ et $B = \{Z \leq -\varepsilon\}$. Alors $B \in \mathcal{B}$, puisque $Z$ est $\mathcal{B}$-mesurable, de plus

$$(1) \quad \Rightarrow \quad 0 \leq \int_B Z \mathrm{d}P = \int_{Z \leq -\varepsilon} Z \mathrm{d}P \leq -\varepsilon P\{Z \leq \varepsilon\} \leq 0$$

donc, pour tout

$$\varepsilon > 0 \ : \ P\{Z \leq -\varepsilon\} = 0 \quad \text{et} \quad P\{Z < 0\} = P\left\{ \bigcup_{\substack{\varepsilon > 0 \\ \varepsilon \in \mathbb{Q}}} \{Z \leq -\varepsilon\} \right\} = 0$$

Corollaire 1 :
Soient $Y_1$ et $Y_2$ deux variables aléatoires réelles $\mathcal{B}$-mesurables et intégrables telles que :

$$\text{pour tout} \quad B \in \mathcal{B} \quad : \quad \int_B Y_1 \mathrm{d}P = \int_B Y_2 \mathrm{d}P \quad \text{alors} \quad Y_1 = Y_2$$

presque sûrement.
Démonstration : Immédiate : poser $Z = Y_1 - Y_2$.

## 5.3 Proposition 2

Soient $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ et $\mathcal{B}$ une sous-tribu de $\mathcal{A}$ alors $Y = E\{X/\mathcal{B}\}$ presque sûrement si et seulement si
(1) $Y$ est presque sûrement égale à une variable aléatoire $\mathcal{B}$-mesurable et intégrable
(2) pour toute variable aléatoire réelle $Z$ $\mathcal{B}$-mesurable et bornée, on a : $E\{X.Z\} = E\{Y.Z\}$
Démonstration :
Si $Y$ vérifie les propriétés (1) et (2), on a immédiatement $Y = E\{X/\mathcal{B}\}$ presque sûrement (poser $Z = \mathbf{1}_B$, $B \in \mathcal{B}$). Réciproquement supposons que $Y = E\{X/\mathcal{B}\}$ presque sûrement ; on a (1) par définition, de plus, pour toute variable aléatoire réelle $Z$ $\mathcal{B}$-mesurable étagée :

$$Z = \sum_{\text{finie}} \alpha_i \mathbf{1}_{B_i} \quad \text{où} \quad \alpha_i \in \mathbb{R} \quad \text{et} \quad B_i \in \mathcal{B}$$

on a :

$$\begin{aligned} E\{X.Z\} &= \sum_{\text{finie}} \alpha_i E\{X.\mathbf{1}_{B_i}\} \\ &= \sum_{\text{finie}} \alpha_i E\{Y.\mathbf{1}_{B_i}\} \\ &= E\{Y.Z\} \end{aligned}$$

Si $Z$ est une variable aléatoire réelle $\mathcal{B}$-mesurable et bornée, posons :

$$Z_n = \sum_{k\in\mathbb{Z}} \frac{k}{2^n}.\mathbf{1}_{(Z\in[\frac{k}{2^n},\frac{k+1}{2^n}[)} \qquad (n \in \mathbb{N}^*)$$

alors $Z_n$ est $\mathcal{B}$-mesurable étagée,

$$|Z - Z_n| \leq \frac{1}{2^n}$$

$$|Z_n| \leq \sup_\omega |Z(\omega)| + 1 < \infty$$

$$E\{X.Z_n\} = E\{Y.Z_n\}$$

il suffit de faire tendre $n$ vers l'infini et d'appliquer le théorème de Lebesgue pour obtenir l'égalité :

$$E\{X.Z\} = E\{Y.Z\}$$

## 5.4 Propriétés de l'espérance conditionnelle

Soient $X$ et $Y$ appartenant à $\mathcal{L}^1(\Omega, \mathcal{A}, P)$ et $\mathcal{B}$ une sous-tribu de $\mathcal{A}$ alors :

1. $E\{X+Y/\mathcal{B}\} = E\{X/\mathcal{B}\} + E\{Y/\mathcal{B}\}$ presque sûrement
2. Pour tout $\alpha \in \mathbb{R}$ $\quad E\{\alpha X/\mathcal{B}\} = \alpha E\{X/\mathcal{B}\}$ presque sûrement
3. Si $X \leq Y$ presque sûrement, on a : $E\{X/\mathcal{B}\} \leq E\{Y/\mathcal{B}\}$ presque sûrement, en particulier : $X \geq 0$ presque sûrement $\Rightarrow E\{X/\mathcal{B}\} \geq 0$ presque sûrement
4. $E\big\{E\{X/\mathcal{B}\}\big\} = E\{X\}$
5. Si $X$ et $\mathcal{B}$ sont indépendantes, on a $E\{X/\mathcal{B}\}$ est constante presque sûrement et égale à $E\{X\}$, en particulier, pour toute constante $\alpha \in \mathbb{R}$ $E\{\alpha/\mathcal{B}\} = \alpha$ presque sûrement
6. Soit $\varphi$ une fonction convexe minorée de $\mathbb{R}$ dans $\mathbb{R}$ alors

   $$\varphi\big\{E\{X/\mathcal{B}\}\big\} \leq E\{\varphi(X)/\mathcal{B}\}$$

   presque sûrement si $\varphi(X) \in \mathcal{L}^1$, en particulier

   $$|E\{X/\mathcal{B}\}| \leq E\{|X|/\mathcal{B}\}$$

   et

   $$|E\{X/\mathcal{B}\}|^p \leq E\{|X|^p/\mathcal{B}\}$$

   presque sûrement si $p \in [1, \infty[$ et $X \in \mathcal{L}^p$.
7. Si $Y$ est $\mathcal{B}$-mesurable et $X.Y \in \mathcal{L}^1$ alors

   $$E\{X.Y/\mathcal{B}\} = Y E\{X/\mathcal{B}\}$$

   presque sûrement, en particulier

   $$E\{Y/\mathcal{B}\} = Y$$

   presque sûrement

8. Si $\mathcal{B}^{'}$ est une sous-tribu de $\mathcal{B}$, on a

$$E\big\{E\{X/\mathcal{B}\}/\mathcal{B}^{'}\big\} = E\{X/\mathcal{B}^{'}\} = E\big\{E\{X/\mathcal{B}^{'}\}/\mathcal{B}\big\}$$

presque sûrement
9. Si $X_n$ $(n \in \mathbb{N})$ est une suite de variables aléatoires réelles positives telles que $0 \leq X_n \leq X_{n+1}$ et $X_n \nearrow X$ presque sûrement alors

$$E\{X_n/\mathcal{B}\} \nearrow E\{X/\mathcal{B}\}$$

presque sûrement

Démonstration :
Pour démontrer les propriétés précédentes on utilise les assertions (1) et (2) de la proposition 2 qui caractérisent l'espérance conditionnelle.

1. Soient $Y_1 = E\{X/\mathcal{B}\}, Y_2 = E\{Y/\mathcal{B}\}, Y_3 = E\{X+Y/\mathcal{B}\}$ presque sûrement et $Z$ une variable aléatoire réelle $\mathcal{B}$-mesurable bornée. On a, d'après (2) :

$$\begin{aligned} E\{(Y_1+Y_2).Z\} &= E\{Y_1.Z\} + E\{Y_2.Z\} \\ &= E\{X.Z\} + E\{Y.Z\} \\ &= E\{(X+Y).Z\} \\ &= E\{(Y_3.Z\} \end{aligned}$$

donc $Y_1 + Y_2 = Y_3$ presque sûrement puisque $Y_1 + Y_2$ est $\mathcal{B}$-mesurable et intégrable.
2. On démontre de même le 2).
3. On suppose que $X \leq Y$ presque sûrement, alors pour tout $B \in \mathcal{B}$ :

$$\int_B E\{(Y-X)/\mathcal{B}\}\,\mathrm{d}P = \int_B (Y-X)\,\mathrm{d}P \geq 0$$

et puisque $E\{(Y-X)/\mathcal{B}\}$ est $\mathcal{B}$-mesurable intégrable, on a, d'après le proposition 1 :

$$E\{Y/\mathcal{B}\} - E\{X/\mathcal{B}\} = E\{(Y-X)/\mathcal{B}\} \geq 0$$

presque sûrement
4. On a : $\Omega \in \mathcal{B}$, donc :

$$\int_\Omega X\,\mathrm{d}P = \int_\Omega E\{X/\mathcal{B}\}\,\mathrm{d}P$$

5. Dire que $X$ et $\mathcal{B}$ sont indépendantes signifie que, pour tout $B \in \mathcal{B}$ les variables aléatoires $X$ et $\mathbf{1}_B$ sont indépendantes. Soit $B \in \mathcal{B}$ alors

$$\int_B X\,\mathrm{d}P = E\{X.\mathbf{1}_B\} = E\{X\}P(B) = \int_B E(X)\,\mathrm{d}P$$

donc $E(X) = E\{X/\mathcal{B}\}$ presque sûrement puisque la variable aléatoire constante $E(X)$ est évidemment $\mathcal{B}$-mesurable, intégrable.

6. Soit $\mathcal{S}$ l'ensemble des fonctions affines $\Delta$ telles que, pour tout $x \in \mathbb{R}$ : $\Delta(x) \leq \varphi(x)$. On sait que $\varphi(x) = \sup_{\Delta \in \mathcal{S}} \Delta(x)$. Si $\Delta \in \mathcal{S}$, on a d'après 1), 2), 3) et 5) :

$$\Delta\big\{E\{X/\mathcal{B}\}\big\} = E\{\Delta(X)/\mathcal{B}\} \leq E\{\varphi(X)/\mathcal{B}\}$$

presque sûrement donc

$$\varphi\big\{E\{X/\mathcal{B}\}\big\} \leq E\{\varphi(X)/\mathcal{B}\}$$

presque sûrement

7. On suppose que $Y$ est $\mathcal{B}$-mesurable et $X.Y \in \mathcal{L}^1$. Posons $Y_1 = Y.E\{X/\mathcal{B}\}$ et $Y_2 = E\{X.Y/\mathcal{B}\}$. Soit $Z$ une variable aléatoire réelle $\mathcal{B}$-mesurable et bornée. Pour tout $n \in \mathbb{N}$, on a :

$$\begin{aligned} E\{(Y_2.Z.\mathbf{1}_{(|Y|\leq n)}\} &= E\{X.Y.Z.\mathbf{1}_{(|Y|\leq n)}\} \\ &= E\big\{E\{X/\mathcal{B}\}.Y.Z.\mathbf{1}_{(|Y|\leq n)}\big\} \end{aligned}$$

(car la variable aléatoire $Y.Z.\mathbf{1}_{(|Y|\leq n)}$ est $\mathcal{B}$-mesurable bornée) donc

$$\begin{aligned} Y_2.\mathbf{1}_{(|Y|\leq n)} &= E\{X/\mathcal{B}\}.Y.\mathbf{1}_{(|Y|\leq n)} \\ &= Y_1.\mathbf{1}_{(|Y|\leq n)}\} \end{aligned}$$

presque sûrement (d'après le corollaire 1). En faisant tendre $n$ vers l'infini dans l'égalité précédente, on voit que $Y_1 = Y_2$ presque sûrement.

8. Soit $\mathcal{B}^{'}$ une sous-tribu de $\mathcal{B}$. Pour tout $B \in \mathcal{B}^{'}$, on a, par définition :

$$\begin{aligned} \int_B E\big\{E\{X/\mathcal{B}\}/\mathcal{B}^{'}\big\}\,\mathrm{d}P &= \int_B E\{X/\mathcal{B}\}\,\mathrm{d}P \\ &= \int_B X\,\mathrm{d}P \\ &= \int_B E\{X/\mathcal{B}^{'}\}\,\mathrm{d}P \end{aligned}$$

donc :

$$E\big\{E\{X/\mathcal{B}\}/\mathcal{B}^{'}\big\} = E\{X/\mathcal{B}^{'}\} \quad \text{presque sûrement}$$

d'après le corollaire 1, de plus

$$E\{X/\mathcal{B}^{'}\} = E\big\{E\{X/\mathcal{B}^{'}\}/\mathcal{B}\big\} \quad \text{presque sûrement}$$

d'après 7) puisque toute variable aléatoire $\mathcal{B}^{'}$ mesurable est $\mathcal{B}$-mesurable.

9. Pour tout $n \in \mathbb{N}$ : $0 \leq X_n \leq X_{n+1} \nearrow X$ presque sûrement et $X \in \mathcal{L}^1$ donc :

$$0 \leq E\{X_n/\mathcal{B}\} \leq E\{X_{n+1}/\mathcal{B}\} \leq E\{X/\mathcal{B}\}$$

presque sûrement. Soit $H = \lim \nearrow E\{X_n/\mathcal{B}\}$, alors $H$ est $\mathcal{B}$-mesurable et $0 \leq H \leq E\{X/\mathcal{B}\}$ presque sûrement, donc $H \in \mathcal{L}^1$. De plus, pour tout

$B \in \mathcal{B}$, on a, d'après le théorème de Beppo-Levi :

$$\begin{aligned}\int_B H \,\mathrm{d}P &= \lim_n \nearrow \int_B E\{X_n/\mathcal{B}\} \,\mathrm{d}P \\ &= \lim_n \nearrow \int_B X_n \,\mathrm{d}P \\ &= \int_B X \,\mathrm{d}P \\ &= \int_B E\{X/\mathcal{B}\} \,\mathrm{d}P\end{aligned}$$

donc $H = E\{X/\mathcal{B}\}$ presque sûrement.

## 5.5 Théorème 2

Soit $X$ une variable aléatoire réelle de carré intégrable, définie sur $(\Omega, \mathcal{A}, P)$ et $\mathcal{B}$ une sous-tribu de $\mathcal{A}$, alors $Y = E\{X/\mathcal{B}\}$ presque sûrement si et seulement si $Y$ vérifie les conditions 1) et 2) suivantes :
1) $Y$ est presque sûrement égale à une variable aléatoire réelle $\mathcal{B}$-mesurable, de carré intégrable.
2) Pour toute variable aléatoire réelle $Z$ $\mathcal{B}$-mesurable, de carré intégrable, on a : $E\{(X - Y)^2\} \leq E\{(X - Z)^2\}$
De plus, sous la condition 1), 2) est équivalente à 3) :
3) Pour toute variable aléatoire réelle $Z$ $\mathcal{B}$-mesurable, de carré intégrable, on a : $E\{(X - Y).Z\} = 0$
Remarque :
Le théorème précédent montre que l'espérance conditionnelle de $X$ sachant $\mathcal{B}$, $E\{X/\mathcal{B}\}$, est « la meilleure approximation en moyenne quadratique » (dans $\mathcal{L}^2$, à l'égalité presque sûre près) de $X$ par une variable aléatoire réelle $Y$ $\mathcal{B}$-mesurable.
Démonstration : On suppose que $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ ; on a alors, d'après les propriétés de l'espérance conditionnelle :

$$(E\{X/\mathcal{B}\})^2 \leq E\{X^2/\mathcal{B}\} \quad \text{presque sûrement}$$

et

$$E\big\{(E\{X/\mathcal{B}\})^2\big\} \leq E\big\{E\{X^2/\mathcal{B}\}\big\} = E\{X^2\} < \infty$$

donc $E\{X/\mathcal{B}\}$ vérifie 1).
Soit $Z$ une variable aléatoire réelle $\mathcal{B}$-mesurable, de carré intégrable. Pour tout $n \in \mathbb{N}$ on sait, d'après la proposition 2, que

$$E\{X.Z.\mathbf{1}_{(|Z| \leq n)}\} = E\big\{E\{X/\mathcal{B}\}.Z.\mathbf{1}_{(|Z| \leq n)}\big\}$$

Si $n \to \infty$ on voit, en appliquant le théorème de Lebesgue, que :

$$E\{X.Z\} = E\big\{E\{X/\mathcal{B}\}.Z\big\}$$

ou bien

$$E\big\{(X - E\{X/\mathcal{B}\}).Z\big\} = 0$$

donc $E\{X/\mathcal{B}\}$ vérifie 3) ; de plus

$$\begin{aligned}
E\{(X-Z)^2\} &= E\big\{[(X-E\{X/\mathcal{B}\})+(E\{X/\mathcal{B}\}-Z)]^2\big\} \\
&= E\big\{(X-E\{X/\mathcal{B}\})^2\big\}+E\big\{(E\{X/\mathcal{B}\}-Z)^2\big\} \\
&\quad + 2\ E\big\{(X-E\{X/\mathcal{B}\}).(E\big\{(X/\mathcal{B}\}-Z)\big\} \\
&= E\big\{(X-E\{X/\mathcal{B}\})^2\big\}+E\big\{(E\{X/\mathcal{B}\}-Z)^2\big\} \qquad (1) \\
&\geq E\big\{(X-E\{X/\mathcal{B}\})^2\big\} \qquad (*)
\end{aligned}$$

Car $\tilde{Z}=E\{X/\mathcal{B}\}-Z$ est $\mathcal{B}$-mesurable de carré intégrable donc

$$E\big\{(X-E\{X/\mathcal{B}\}).\tilde{Z}\big\}=0$$

L'inégalité (*) montre donc que : $E\{X/\mathcal{B}\}$ vérifie 2).
Réciproquement, si $Y$ est une variable aléatoire réelle vérifiant 1) et 2), on voit, d'après (1), en remplaçant $Z$ par $Y$, que :

$$\begin{aligned}
&E\big\{(X-E\{X/\mathcal{B}\})^2\big\}+E\big\{(E\{X/\mathcal{B}\}-Y)^2\big\} \\
&\quad = E\big\{(X-Y)^2\big\} \\
&\quad \leq E\big\{(X-E\{X/\mathcal{B}\})^2\big\}
\end{aligned}$$

donc

$$E\big\{(E\{X/\mathcal{B}\}-Y)^2\big\}=0 \quad \text{et} \quad Y=E\{X/\mathcal{B}\} \quad \text{presque sûrement}$$

Si $Y$ est une variable aléatoire réelle vérifiant 1) et 3), il est immédiat de voir, d'après la proposition 2, que $Y=E\{X/\mathcal{B}\}$ presque sûrement. On a donc montré que :

$$\begin{aligned}
&Y=E\{X/\mathcal{B}\} \quad \text{presque sûrement} \\
&\iff Y \text{ vérifie 1) et 2)} \\
&\iff Y \text{ vérifie 1) et 3)}
\end{aligned}$$

C.Q.F.D.

## 5.6 Proposition 3

Soient $X_1$ et $X_2$ appartenant à $\mathcal{L}^2(\Omega,\mathcal{A},P)$ et $\mathcal{B}$ une sous-tribu de $\mathcal{A}$, alors :

$$|E\{X_1.X_2/\mathcal{B}\}| \leq \big(E\{X_1^2/\mathcal{B}\}\big)^{\frac{1}{2}}\big(E\{X_2^2/\mathcal{B}\}\big)^{\frac{1}{2}} \quad \text{presque sûrement}$$

en particulier :

$$|E\{X_1/\mathcal{B}\}| \leq E\{|X_1|/\mathcal{B}\} \leq \big(E\{X_1^2/\mathcal{B}\}\big)^{\frac{1}{2}} \quad \text{presque sûrement}$$

Démonstration :
Exercice : utiliser le fait que, pour tout $\lambda \in \mathbb{R}$ :

$$E\{(X_1+\lambda X_2)^2/\mathcal{B}\} \geq 0 \quad \text{presque sûrement}$$

## 5.7 Théorème 3

Soient $T$ une variable aléatoire définie sur $(\Omega, \mathcal{A})$ à valeurs dans un espace mesurable $(F, \mathcal{F})$ et $Y$ une application de $\Omega$ dans $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Notons $\sigma(T)$ la tribu engendrée par $T$ :

$$\sigma(T) = \{T^{-1}(S);\ S \in \mathcal{F}\} \ \subseteq \ \mathcal{A}$$

Alors $Y$ est $\sigma(T)$-mesurable si et seulement si il existe une application mesurable $h$ de $(F, \mathcal{F})$ dans $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ telle que $Y = h \circ T$. On a donc le schéma suivant :

$$\begin{tikzcd}
(\Omega, \sigma(T)) \arrow[r, "T"] \arrow[dr, "Y = h \circ T"'] & (F, \mathcal{F}) \arrow[d, "h"] \\
& (\mathbb{R}, \mathcal{B}_{\mathbb{R}})
\end{tikzcd}$$

Démonstration :
Si $Y = h \circ T$, avec $h$ mesurable de $(F, \mathcal{F})$ dans $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ il est évident que :

$$\sigma(Y) = \{T^{-1}\left(h^{-1}(C)\right);\ C \in \mathcal{B}_{\mathbb{R}}\} \ \subseteq \ \sigma(T)$$

Réciproquement, supposons d'abord, que $Y$ est une variable aléatoire réelle étagée $\sigma(T)$-mesurable

$$Y = \sum_{\text{finie}} \alpha_i.\mathbf{1}_{D_i} \quad \text{où} \quad \alpha_i \in \mathbb{R} \quad \text{et} \quad D_i \in \sigma(T)$$

pour tout $i$ :

$$D_i = T^{-1}(S_i) \quad \text{où} \quad S_i \in \mathcal{F} \quad \text{donc} \quad \mathbf{1}_{D_i} = \mathbf{1}_{S_i} \circ T$$

et

$$Y = \sum_{\text{finie}} \alpha_i.\mathbf{1}_{S_i} \circ T = h \circ T \quad \text{avec} \quad h = \sum_{\text{finie}} \alpha_i.\mathbf{1}_{S_i}$$

Si maintenant $Y$ est une variable aléatoire réelle positive $\sigma(T)$-mesurable, il existe une suite croissante $Y_n$ de variables aléatoires réelles étagées, $\sigma(T)$-mesurables positives telles que : $Y = \lim_n \nearrow Y_n$ ; on peut donc écrire $Y_n = h_n \circ T$ où $h_n$ est une application mesurable de $(F, \mathcal{F})$ dans $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Soit

$$h = \overline{\lim_n}\ h_n.\mathbf{1}_{(\overline{\lim}\ h_n < \infty)}$$

on voit que

$$Y = \lim_n \nearrow h_n \circ T = h \circ T$$

Si $Y$ est de signe quelconque, on décompose : $Y = Y^+ - Y^-$.

## 5.8 Définition 1

Soient $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ et $T$ une variable aléatoire définie sur $(\Omega, \mathcal{A})$ à valeur dans $(F, \mathcal{F})$. L'espérance conditionnelle de $X$ sachant $T$, notée $E\{X/T\}$ est définie par :

$$E\{X/T\} = E\{X/\sigma(T)\}$$

où $\sigma(T) \subseteq \mathcal{A}$ est la sous-tribu engendrée par $T$.
En combinant la proposition 2 et le théorème 3, on voit que : $Y = E\{X/T\}$ presque sûrement est caractérisée par (1) et (2) :
(1) Il existe $g$ application mesurable de $(F, \mathcal{F})$ dans $\mathbb{R}$ telle que : $Y = g(T)$ presque sûrement et $g(T) \in \mathcal{L}^1$.
(2) Pour toute application $h$ mesurable et bornée de $(F, \mathcal{F})$ dans $\mathbb{R}$ on a :

$$E\{X.h(T)\} = E\{Y.h(T)\}$$

Le théorème 2 montre, de plus, que si $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$, alors $E\{X/T\}$ est la meilleure approximation en moyenne quadratique de $X$ par une fonction mesurable de $T$, plus précisément :

1. Il existe $g$ application mesurable de $(F, \mathcal{F})$ dans $\mathbb{R}$ telle que $E\{X/T\} = g(T)$ presque sûrement et $g(T) \in \mathcal{L}^2$
2. Pour toute application mesurable $h$ de $(F, \mathcal{F})$ dans $\mathbb{R}$ telle que $h(T) \in \mathcal{L}^2$, on a :
$$E\{(X - g(T))^2\} \leq E\{(X - h(T))^2\}$$

Remarque :
Dans la pratique, pour calculer l'espérance conditionnelle $E\{X/T\}$, le problème qui se pose est de trouver une fonction $g$ telle que $E\{X/T\} = g \circ T$ presque sûrement. Connaissant la loi du couple $(X, T)$, les conditions (1) et (2) précédentes permettent, en général, de déterminer la restriction de $g$ à un sous-ensemble mesurable $S$ de $F$ tel que $P\{T \in S\} = 1$, ce qui est suffisant car la variable aléatoire $g \circ T$ est alors bien définie presque sûrement.
Exemples importants :
Exemple A : Si $T$ est une variable aléatoire discrète à valeurs dans un ensemble dénombrable $F$, posons :

$$S = \{t \in F \ : \ P(T = t) > 0\}$$

alors $P\{T \in S\} = 1$ et pour toute variable aléatoire réelle $X \in \mathcal{L}^1$, on a : $E\{X/T\} = g \circ T$ presque sûrement où :

$$\forall\ t \in S \ : \ g(t) = \frac{1}{P(T = t)} \int_{\{T=t\}} X \, \mathrm{d}P$$

Démonstration :
Soit $t \in S$. Posons $h = \mathbf{1}_{\{t\}}$, alors, d'après (2), si $E\{X/T\} = g \circ T$ on a

$$\begin{aligned} E\{X.(\mathbf{1}_{\{t\}} \circ T)\} &= E\{(g \circ T).(\mathbf{1}_{\{t\}} \circ T)\} \\ &= E\{X.\mathbf{1}_{(T=t)}\} \\ &= g(t).P\{T = t\} \end{aligned}$$

donc

$$g(t) = \frac{1}{P(T = t)} E\{X.\mathbf{1}_{(T=t)}\}$$

Exemple B : Si $(X, T)$ est à valeurs $\mathbb{R}^n \times \mathbb{R}^m$, de densité $f = (f(x, t))$, posons

$$S = \{t \in \mathbb{R}^m \ ; \ \int_{\mathbb{R}^n} f(x, t) \, \mathrm{d}x \ > \ 0\}$$

alors $P\{T \in S\} = 1$ et pour toute fonction $\theta$ mesurable et bornée de $\mathbb{R}^n$ dans $\mathbb{R}$, on a

$$E\{\theta(X)/T\} = g \circ T \quad \text{presque sûrement}$$

où pour presque tout $t \in S$ :

$$g(t) = \int_{\mathbb{R}^n} \theta(x) f(x,t)\, \mathrm{d}x \times \frac{1}{\int_{\mathbb{R}^n} f(x,t)\, \mathrm{d}x} \quad (*)$$

Démonstration :
On sait que la densité $k$ de $T$ est donnée par :

$$k(t) = \int_{\mathbb{R}^n} f(x,t)\, \mathrm{d}x \quad (t \in \mathbb{R}^m)$$

donc $S = \{t \ : \ k(t) > 0\}$ et on a bien $P\{T \in S\} = 1$.
Soit $\theta \ : \ \mathbb{R}^n \to \mathbb{R}$ mesurable et bornée, alors, d'après (2), si $E\{\theta(X)/T\} = g(T)$, pour toute application $h$ mesurable et bornée de $\mathbb{R}^n$ dans $\mathbb{R}$

$$\begin{aligned}
E\{\theta(X).h(T)\} &= E\{g(T).h(T)\} \\
&= \iint \theta(x) h(t) f(x,t)\, \mathrm{d}x\, \mathrm{d}t \\
&= \iint g(t) h(t) f(x,t)\, \mathrm{d}x\, \mathrm{d}t \\
&= \int_{\mathbb{R}^m} h(t) \left[ \int_{\mathbb{R}^n} \theta(x) f(x,t)\, \mathrm{d}x \right] \mathrm{d}t \\
&= \int_{\mathbb{R}^m} h(t) g(t) k(t)\, \mathrm{d}t
\end{aligned}$$

l'égalité précédente étant valable pour tout $h$, on voit que

$$g(t)k(t) = \int_{\mathbb{R}^n} \theta(x) f(x,t)\, \mathrm{d}x$$

presque partout en $t$. Si $t \in S$, on obtient donc la formule (*).
Exemple C : Soit $(X,T)$ une variable aléatoire à valeurs dans $(F_1 \times F_2, \mathcal{B}_1 \otimes \mathcal{B}_2)$. Si $X$ et $T$ sont indépendantes, alors pour toute application $\varphi$ mesurable et bornée de $F_1 \times F_2$ dans $\mathbb{R}$, on a :

$$E\{\varphi(X,T)/T\} = g \circ T \quad \text{presque sûrement}$$

où

$$g(t) = E\{\varphi(X,t)\} = \int_{F_1} \varphi(x,t) P_X(\mathrm{d}x)$$

pour presque tout $t \in F_2$ (relativement à la loi de $T$).
Démonstration :
Si $t \in F_2$, posons $g(t) = E\{\varphi(X,t)\}$. Pour toute application $h$ mesurable et

bornée de $F_2$ dans $\mathbb{R}$, on a, d'après le théorème de Fubini :

$$\begin{aligned}
E\{\varphi(X,T).h(T)\} &= \iint_{F_1\times F_2} \varphi(x,t)h(t)P_X(\mathrm{d}x)P_T(\mathrm{d}t) \\
&= \int_{F_2} h(t)\left[\int_{F_1} \varphi(x,t)P_X(\mathrm{d}x)\right] P_T(\mathrm{d}t) \\
&= \int_{F_2} h(t)E\{\varphi(X,t)\}P_T(\mathrm{d}t) \\
&= \int_{F_2} h(t)g(t)P_T(\mathrm{d}t) \\
&= E\{g(T).h(T)\} \\
&= E\{E\{\varphi(X,T)/T\}.h(T)\}
\end{aligned}$$

donc

$$E\{\varphi(X,T)/T\} = g(T) \quad \text{presque sûrement}$$

## 5.9 Définition 2

Soient $X$ et $T$ deux variables aléatoires définies sur $(\Omega, \mathcal{A}, P)$ à valeurs dans des espaces mesurables $(F_1, \mathcal{B}_1)$ et $(F_2, \mathcal{B}_2)$ respectivement. On appelle loi conditionnelle de $X$ sachant $T$ une famille de probabilités sur l'espace $(F_1, \mathcal{B}_1)$ indicée par $t \in F_2$ : $(N(t, \mathrm{d}x))_{t\in F_2}$ vérifiant les conditions suivantes :

1. Pour tout $B_1 \in \mathcal{B}_1$, l'application $t \in F_2 \longmapsto N(t, B_1) \in [0,1]$ est mesurable.
2. Pour toute application $\theta$ mesurable et bornée de $(F_1, \mathcal{B}_1)$ dans $\mathbb{R}$ on a :

   $$E\{\theta(X)/T\}(\omega) = \int_{F_1} \theta(x)N\left(T(\omega), \mathrm{d}x\right) \quad \text{presque sûrement}$$

   En particulier

   $$P\{X \in B_1/T\} \stackrel{\text{def}}{=} E\{\mathbf{1}_{B_1} \circ X/T\} = N(T, B_1) \quad \text{presque sûrement}$$

   pour tout $B_1 \in \mathcal{B}_1$.

On dit alors que $N(t, \mathrm{d}x)$ est la loi conditionnelle de $X$ sachant que $(T = t)$.
Remarques :

1. Il suffit de connaître la loi conditionnelle de $X$ sachant que $(T = t)$, pour tout $t \in S$ où $S$ est un sous-ensemble mesurable de $F_2$ tel que $P\{T \in S\} = 1$.
2. Pour toute application mesurable et bornée $\varphi$ de $F_1 \times F_2$ dans $\mathbb{R}$, on peut montrer (facilement) que

   $$(*) \quad E\{\varphi(X,T)/T\}(\omega) = \int_{F_1} \varphi(x, T(\omega))N\left(T(\omega), \mathrm{d}x\right) \quad \text{presque sûrement}$$

   On utilisera souvent la notation suivante :

   $$E\{\varphi(X,T)/(T = t)\} \stackrel{\text{def}}{=} \int_{F_1} \varphi(x,t)N(t, \mathrm{d}x)$$

(même si $P(T=t)=0!$).
La formule (*) entraîne, de plus, que :

$$\begin{aligned}E\{\varphi(X,T)\} &= E\big\{E\{\varphi(X,T)/T\}\big\}\\ &= \int_{F_1\times F_2}\varphi(x,t)P_{(X,T)}(\mathrm{d}x,\mathrm{d}t)\\ &= \int_{F_2}P_T(\mathrm{d}t)\int_{F_1}\varphi(x,t)N(t,\mathrm{d}x)\end{aligned}$$

(formule des lois images)
On a donc l'égalité formelle :

$$P_{(X,T)}(\mathrm{d}x,\mathrm{d}t)=P_T(\mathrm{d}t)N(t,\mathrm{d}x)$$

où $N(t,\mathrm{d}x)$ est la loi conditionnlle de $X$ sachant que $(T=t)$.

3. Les variables aléatoires $X$ et $T$ sont indépendantes si et seulement si $N(t,\mathrm{d}x)$ ne dépend pas de $t$ : $N(t,\mathrm{d}x)=P_X(\mathrm{d}x)$ pour tout $t\in F_2$.
4. Calculons la loi conditionnelle de $X$ sachant $T$ dans chacun des exemples précédents :
   Exemple A : ($T$ variable aléatoire discrète)
   Pour toute application mesurable et bornée $\theta$ à valeurs réelles :

$$\begin{aligned}E\{\theta(X)/(T=t)\} &= \frac{1}{P(T=t)}\int_{\{T=t\}}\theta(X)\,\mathrm{d}P\\ &= \int_\Omega \theta(X)(\omega)P_t(\mathrm{d}\omega) \qquad \text{où}\quad P_t(\mathrm{d}\omega)=P\{\mathrm{d}\omega/(T=t)\}\\ &\qquad\qquad\qquad\qquad\qquad \text{si}\quad t\in F\quad\text{et}\quad P(T=t)>0\\ &= \int_{\mathbb{R}}\theta(x)N(t,\mathrm{d}x)\end{aligned}$$

   donc $N(t,\mathrm{d}x)$ est la loi de $X$ sous $P_t$.
   Exemple B :
   $(X,T)$ admet une densité $(f(x,t))$ sur $\mathbb{R}^n\times\mathbb{R}^m$. On voit immédiatement que :

$$N(t,\mathrm{d}x)=f(x,t)\,\mathrm{d}x.\frac{1}{\int_{\mathbb{R}^n}f(x,t)\,\mathrm{d}x}\quad \text{si}\quad t\in\mathbb{R}^m\quad\text{et}\quad \int_{\mathbb{R}^n}f(x,t)\,\mathrm{d}x>0$$

   Exemple C :
   $X$ et $T$ indépendantes à valeurs dans $(F_1,\mathcal{B}_1)$ et $(F_2,\mathcal{B}_2)$ respectivement. Soit $\Phi$ une application mesurable de $(F_1\times F_2,\mathcal{B}_1\otimes\mathcal{B}_2)$ dans $(F_3,\mathcal{B}_3)$. Posons $\tilde X=\Phi(X,T)$. Alors la loi conditionnelle de $\tilde X$ sachant que $(T=t)$, $\tilde N(t,\mathrm{d}y)$ est égale à la loi de la variable aléatoire $\Phi(X,t)$. En effet, pour toute application mesurable et bornée $\theta$ de $(F_3,\mathcal{B}_3)$ dans $\mathbb{R}$ on a :

$$E\{\theta(\tilde X)/(T=t)\}=E\{\theta\left(\Phi(X,T)\right)/(T=t)\}=g(t)$$

   où

$$g(t)=E\{\theta\left(\Phi(X,t)\right)\}=\int_{F_3}\theta(y)\tilde N(t,\mathrm{d}y)\quad (t\in F_2)$$

## 5.10 Espaces $L^P$

Soit $1 \leq p < \infty$, rappelons (cf. le chapitre 3 I) que $\mathcal{L}^p(\Omega, \mathcal{a}, P) = \mathcal{L}^p$ désigne l'ensemble des variables aléatoires réelles finies $X$ telles que $E\{|X|^p\} < \infty$. $\mathcal{L}^p$ est un espace vectoriel sur $\mathbb{R}$. Pour tout $X \in \mathcal{L}^p$, posons

$$N_p(X) = (E\{|X|^p\})^{\frac{1}{p}}$$

L'application $N_p(.)$ vérifie les propriétés :

1. $N_p(X) = 0 \iff X = 0$ presque sûrement
2. $\forall\ \lambda \in \mathbb{R}, \quad \forall\ X \in \mathcal{L}^p, \quad N_p(\lambda X) = |\lambda|.N_p(X)$.
3. $\forall\ X$ et $Y$ dans $\mathcal{L}^p$, $N_p(X+Y) \leq N_p(X) + N_p(Y)$

A cause de la propriété 1), $N_p(.)$ n'est pas une norme sur $\mathcal{L}^p$. Pour obtenir un « bon » espace normé, on effectue la construction suivante :

$$\forall\ X \in \mathcal{L}^p, \quad \text{soit} \quad \dot{X} = \{Y \in \mathcal{L}^p \quad \text{tel que} \quad X = Y \quad \text{presque sûrement}\} \subseteq \mathcal{L}^p$$

On définit : $L^p = \{\dot{X} \quad \text{où} \quad X \quad \text{varie dans} \quad \mathcal{L}^p\}$
Remarquer que pour tous $X_1, Y_1, X_2, Y_2$ dans $\mathcal{L}^p$ et $\lambda \in \mathbb{R}$, on a :

$$[\dot{X}_1 = \dot{Y}_1 \quad \text{et} \quad \dot{X}_2 = \dot{Y}_2] \Rightarrow \dot{\widehat{X_1 + X_2}} = \dot{\widehat{Y_1 + Y_2}}$$

$$[\dot{X}_1 = \dot{Y}_1] \Rightarrow \dot{\widehat{\lambda X_1}} = \dot{\widehat{\lambda Y_1}}$$

On munit l'ensemble $L^p$ d'une addition et d'une multiplication par les scalaires en posant : si $C_1$ et $C_2$ appartiennent à $L^p$, $C_1 = \dot{X}_1$, $C_2 = \dot{X}_2$ où $X_1$ et $X_2 \in \mathcal{L}^p$

$$C_1 \dot{+} C_2 \stackrel{\text{def}}{=} \dot{\widehat{X_1 + X_2}}$$

et pour tout $\lambda \in \mathbb{R}$ :

$$\lambda.C_1 \stackrel{\text{def}}{=} \dot{\widehat{\lambda X_1}}$$

La remarque précédente montre que ces définitions ont bien un sens et on vérifie facilement que $(L^p, \dot{+}, .)$ est un espace vectoriel sur $\mathbb{R}$. Soient $X$ et $Y$ dans $\mathcal{L}^p$, noter que

$$\dot{X} = \dot{Y} \Rightarrow N_p(X) = N_p(Y)$$

Si $C$ appartient à $L_p$, on peut donc définir sans ambiguïté $N_p(C)$ par la formule :

$$(1) \qquad N_p(C) = N_p(X)$$

où $X \in \mathcal{L}^p$ est un représentant quelconque de $C$ : $\dot{X} = C$.
On a alors :

$$N_p(C) = N_p(\dot{X}) = 0 \Rightarrow X = 0 \quad \text{presque sûrement} \quad \Rightarrow C = \dot{0}$$

où $\dot{0}$ est l'élément neutre de $L^p$. Il est immédiat de vérifier que la formule (1) définit une véritable norme $N_p(.)$ sur l'espace vectoriel $L^p$. On identifiera, dans la suite, un élément $C$ de $L^p$ à l'un quelconque de ses représentants $X$ dans $\mathcal{L}^p$ : $C = \dot{X}$. Une variable aléatoire réelle $X$ telle que $E\{|X|^p\} < \infty$ sera donc considérée (pour simplifier les notations) indifferemment, suivant les contextes, comme un élément de $L^p$ ou de $\mathcal{L}^p$. Si $X \in L^p$, on notera de plus :

$$N_p(X) = (E\{|X|^p\})^{\frac{1}{p}} = ||X||_p$$

## 5.11 Théorème 4

Soit $1 \leq p < \infty$. L'espace $L^p$, muni de la norme $||.||_p$, est un espace de Banach.
Démonstration :
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de Cauchy dans $L^p$. On construit facilement, par récurrence, une sous-suite $1 \leq n_1 < n_2 < \ldots < n_k \nearrow +\infty$ telle que, pour tout k :

$$||X_{n_{k+1}} - X_{n_k}||_p \leq \frac{1}{2^k}$$

soit

$$0 \leq Y_k = \sum_{l=1}^{k-1} |X_{n_{l+1}} - X_{n_l}| \quad (k \geq 2)$$

On voit que, pour tout $k \geq 2$ :

$$||Y_k||_p = (E\{(Y_k)^p\})^{\frac{1}{p}} \leq \sum_{l=1}^{k-1} ||X_{n_{l+1}} - X_{n_l}||_p \leq \sum_{l=1}^{k-1} \frac{1}{2^l} < 1$$

Le lemme de Fatou montre alors que :

$$E\left\{\underline{\lim}_{k\to\infty} (Y_k)^p\right\} \leq \underline{\lim}_{k\to\infty} E\{Y_k^p\} \leq 1$$

donc

$$\underline{\lim}_{k\to\infty} Y_k(\omega) = \sum_{l\geq 1} |X_{n_{l+1}} - X_{n_l}|(\omega) < \infty \quad \text{presque sûrement}$$

Il est facile d'en déduire que la sous-suite $(X_{n_l})_{l\geq 1}$ converge presque sûrement vers une variable aléatoire $X_\infty$, de plus

$$|X_\infty| = \lim_{k\to\infty} |X_{n_k}| \leq |X_{n_1}| + \lim_{k\to\infty} Y_k$$

donc $X_\infty \in L^p$ et

$$\begin{aligned}
(E\{|X_\infty - X_{n_k}|^p\})^{\frac{1}{p}} &\leq \underline{\lim}_{l\to\infty} (E\{|X_{n_l} - X_{n_k}|^p\})^{\frac{1}{p}} \\
&\leq \lim_{l\to\infty} \sum_{i=k}^{l-1} ||X_{n_{i+1}} - X_{n_i}||_p \\
&\leq \sum_{i\geq k} \frac{1}{2^i} = \frac{2}{2^k} \to 0 \quad \text{quand} \quad k \to \infty
\end{aligned}$$

d'après le lemme de Fatou et l'inégalité triangulaire donc $X_{n_k} \to X_\infty$ dans $L^p$, or la suite $(X_n)_{n\in\mathbb{N}}$ étant de Cauchy, on voit que :

$$||X_n - X_\infty||_p \leq ||X_n - X_{n_k}||_p + ||X_{n_k} - X_\infty||_p \to 0$$

quand $k$ et $n \to \infty$ et on a bien $\lim_{n\to\infty} X_n = X_\infty$ dans $L^p$.
Corollaire 1 :
Toute suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires réelles qui convergent dans $L^p$ vers une variable aléatoire réelle $X_\infty$, admet une sous-suite $(X_{n_k})_{k\in\mathbb{N}}$ qui converge

presque sûrement vers $X_\infty$.
Démonstration : Immédiate, compte tenu de la démonstration du théorème 4. (Cf. aussi la proposition 2 du chapitre 4).
Corollaire 2 :
Pour tous $X$ et $Y$ dans $L^2$, posons : $< X, Y >= E\{X.Y\}$. On définit ainsi un produit scalaire $< ., . >$ sur $L^2$ et l'espace $L^2$, muni de ce produit scalaire est un espace de Hilbert.
Démonstration : Immédiate, d'après le théorème 4 : remarquer que

$$||X||_2 = (< X, X >)^{\frac{1}{2}}$$

## 5.12 Espérance conditionnelle dans $L^2$

Commençons par un rappel sur les espaces de Hilbert. Soient $(H, < ., . >)$ un espace de Hilbert et $F \subseteq H$ un sous-espace fermé. Pour tout $x \in H$, il existe $y \in F$, unique projection orthogonale de $x$ sur $F$, vérifiant l'une des conditions équivalentes suivantes :

1. $\forall\ z \in F, \quad (x - y) \perp z \quad$ i.e. $\quad < x, z >=< y, z >$
2. $\forall\ z \in F, \quad ||x - y|| \leq ||x - z||$

Soient $(\Omega, \mathcal{A}, P)$ un espace de probabilité et $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. Considérons l'espace de Hilbert $H = L^2(\Omega, \mathcal{A}, P)$ et $F$ le sous-espace fermé de $H$ constitué des éléments ayant un représentant $\mathcal{B}$-mesurable.
Si $X$ est une variable aléatoire réelle de carré intégrable, il est naturel de considérer, comme nous l'avons déjà souligné, la meilleure approximation en moyenne quadratique de $X$ par une variable aléatoire réelle $Y$ ayant un représentant $\mathcal{B}$-mesurable (Cf. le théorème 2). D'après 1) et 2), cette variable aléatoire est la projection orthogonale de $X$ sur $F$. elle est caractérisée par :
① $Y$ est $\mathcal{B}$-mesurable (presque surement), de carré intégrable.
② $\forall\ Z \in L^2$, $\mathcal{B}$-mesurable : $E\{|X - Y|^2\} \leq E\{|X - Z|^2\}$
Compte tenu de ①, on a :
② $\iff$ ②' : $\forall\ Z \in L^2$, $\mathcal{B}$-mesurable : $E\{X.Z\} = E\{Y.Z\}$
②' $\iff$ ②" : $\forall\ B \in \mathcal{B}, \quad E\{X.\mathbf{1}_B\} = E\{Y.\mathbf{1}_B\}$
$Y$ est ainsi déterminée, de manière unique (à l'égalité presque sûre près) par les conditions { ①, ②} ou { ①, ②'} ou { ①, ②"}.
Nous savons déjà, d'après le théorème 2, que $Y$ est l'espérance conditionnelle de $X$ sachant $\mathcal{B}$ : $Y = E\{X/\mathcal{B}\}$ presque sûrement.
Remarque :
L'approche précédente permet de démontrer le théorème 1 au moins si $X \in L^2$. Lorsque $X$ est seulement intégrable, on posera si $X \geq 0$ presque sûrement :

$$E\{X/\mathcal{B}\} = \lim_n \nearrow E\{X.\mathbf{1}_{(X \leq n)}/\mathcal{B}\} \quad \text{presque sûrement}$$

et dans le cas général :

$$E\{X/\mathcal{B}\} = E\{X^+/\mathcal{B}\} - E\{X^-/\mathcal{B}\} \quad \text{presque sûrement}$$

Nous n'entrerons pas dans les détails ...

## 5.13 Cas des vecteurs gaussiens

Soit $(X, T) = (X, T_1, T_2, \ldots, T_d)$ un vecteur gaussien à valeurs $\mathbb{R} \times \mathbb{R}^d$, alors la meilleure approximation en moyenne quadratique de $X$ par une fonction mesurable de $T = (T_1, T_2, \ldots, T_d)$, $E\{X/T\}$ est, en fait, la meilleure approximation en moyenne quadratique par une fonction affine de $T$. Plus précisement :

### 5.13.1 Théorème 1

Soit $(X, T)$ un vecteur gaussien à valeurs $\mathbb{R} \times \mathbb{R}^d$, alors

1. On peut trouver $a, b_1, b_2, \ldots, b_d \in \mathbb{R}$ tels que

$$E\{X/T\} = a + \sum_{j=1}^{d} b_j T_j$$

2. On a la décomposition :

$$X = \left( a + \sum_{j=1}^{d} b_j T_j \right) + Z$$

où $Z$ est une variable aléatoire réelle gaussienne centrée indépendante de $T$

3. La loi conditionnelle de $X$ sachant que $T = t = (t_1, t_2, \ldots, t_d)$ est gaussienne :

$$N(t, \mathrm{d}x) = N_1 \left( a + \sum_{j=1}^{d} b_j t_j, \sigma^2 \right)$$

où $\sigma^2 = \mathrm{Var}(Z)$.

Démonstration :
Soit :

$$G = \mathrm{Vect}\{1, T_1, T_2, \ldots, T_d\} \subseteq L^2(\Omega, \mathcal{A}, P)$$

(sous-espace de dimension finie engendré par $1, T_1, \ldots, T_d$) alors la projection orthogonale de $X$ sur $G$ : $\Pi_G(X)$ est la meilleure approximation dans $L^2$ de $X$ par une fonction affine de $T$ :

1. $\exists a, b_1, \ldots, b_d \in \mathbb{R}$ tels que

$$\Pi_G(X) = a + \sum_{j=1}^{d} b_j T_j$$

2. $X - \Pi_G(X)$ est orthogonal à $G$, donc :

$$E\{(X - \Pi_G(X)) .\mathbf{1}\} = 0$$

$$E\{(X - \Pi_G(X)) .T_l\} = 0 \quad \text{pour tout} \quad l \in \{1, 2, \ldots, d\}$$

On a donc le système linéaire suivant dont les inconnues sont $a, b_1, b_2, \dots, b_d$ :

$$\begin{cases} E\{\Pi_G(X)\} = a + \sum_{j=1}^d b_j E\{T_j\} = E\{X\} \\ E\{\Pi_G(X).T_l\} = aE\{T_l\} + \sum_{j=1}^d b_j E\{T_j T_l\} = E\{X.T_l\} \end{cases}$$

où $l \in \{1, 2, \dots, d\}$
ce qui permet de calculer les constantes $a, b_1, \dots, b_d$
Considérons maintenant le vecteur gaussien :

$$\begin{pmatrix} X - \left(a + \sum_{j=1}^d b_j T_j\right) \\ T \end{pmatrix}$$

Remarquons que :

$$\mathrm{Cov}\left( X - \left( a + \sum_{j=1}^{d} b_j T_j \right), T_l \right) = E\{(X - \Pi_G(X)) . T_l\} = 0$$

pour tout $l \in \{1, 2, \dots, d\}$ . Donc, si on pose

$$Z = X - \left( a + \sum_{j=1}^{d} b_j T_j \right)$$

la variable aléatoire gaussienne $\underline{Z \text{ est indépendante de } T}$, de plus :

$$E\{Z\} = E\{X - \Pi_G(X)\} = 0$$

donc

$$E\Big\{ \left( X - \left( a + \sum_{j=1}^{d} b_j T_j \right) \right) /T \Big\} = E\{Z/T\} = E\{Z\} = 0 \quad \text{presque sûrement}$$

et :

$$E\{X/T\} = E\{a + \sum_{j=1}^{d} b_j T_j /T\} = a + \sum_{j=1}^{d} b_j T_j \quad \text{presque sûrement}$$

On a donc démontré les assertions 1) et 2) du théorème 1 ; l'assertion 3) est une conséquence de la décomposition :

$$X = \left( a + \sum_{j=1}^{d} b_j T_j \right) + Z$$

et du fait que $Z$ et $T$ sont indépendantes (Cf. l'exemple C, page 92)
Remarque :

$$\sigma = (\mathrm{Var}(Z))^{\frac{1}{2}} = ||Z||_2$$

est égal à « l'erreur »

$$\min_{h \in \mathcal{S}} \{||X - h(T)||_2\}$$

où $\mathcal{S}$ désigne l'ensemble des applications mesurables $h$ de $\mathbb{R}^d$ dans $\mathbb{R}$ telles que $h(T) \in L^2$.