# Table des matières

# Chapitre I

# Notions préalables

## Plan du chapitre



## 1 Rappels sur les ensembles

Soit E un ensemble, $A, B, C$ des sous ensembles de E, $I \subset \mathbb{N}$ un ensemble d'indice et $A_i$ des sous ensembles de E.

On a les définitions suivantes :

**Définition 1.** *Opérations sur les ensembles.*

- $A \cup B = \{x, x \in A \ \ ou \ x \in B\}$
- $A \cap B = \{x, x \in A \ \ et \ x \in B\}$
- $A \setminus B = \{x, x \in A \ \ et \ x \notin B\}$
- $\overline{A} = A^c = \{x, x \notin A\}$
- $\bigcup_{i \in I} A_i = A^c = \{x, \exists i_0 \in I \ \ t.q. \ x \in A_{i_0}\}$
- $\bigcap_{i \in I} A_i = A^c = \{x, \forall i \in I, x \in A_i\}$

ainsi que les propriétés :

**Proposition 1.** *Opérations sur les ensembles.*

- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $\overline{A \cup B} = \overline{A} \cap \overline{B}$

- $\overline{A \cap B} = \overline{A} \cup \overline{B}$
- $\overline{\overline{A}} = A$
- $A \subset B \iff \overline{B} \subset \overline{A}$

**Exercice** :

Soit $E$ un ensemble de $A$ et $B$ deux sous-parties de E. Montrer par preuve cyclique l'équivalence entre ces 8 propriétés :

$$\begin{array}{llll} (i) A \subset B & (ii) A \cap B = A & (iii) \overline{A} \cup \overline{B} = \overline{A} & (iv) A \cap \overline{B} = \emptyset \\ (v) \overline{A} \cup B = E & (vi) \overline{B} \subset \overline{A} & (vii) \overline{A} \cap \overline{B} = \overline{B} & (viii) A \cup B = B \end{array}$$

# 2 Limites d'ensembles

**Définition 2.** *Limites de suites monotones d'ensembles.*

- *Une suite d'ensembles $(E_i)_{i\geq 0}$ est dite croissante si pour tout $i$, on a $E_i \subset E_{i+1}$. On a alors par définition*
$$\lim_{i\to\infty} E_i = \bigcup_{i\geq 0} E_i$$
- *Une suite d'ensembles $(E_i)_{i\geq 0}$ est dite décroissante si pour tout $i$, on a $E_{i+1} \subset E_i$. On a par définition*
$$\lim_{i\to\infty} E_i = \bigcap_{i\geq 0} E_i$$

La relation d'inclusion n'est pas une relation d'ordre totale, i.e., on a pas soit $A \subset B$ soit $B \subset A$. On peut toutefois definir les limites inferieurs et supérieures d'une suite d'ensemble $(E_i)_i$ quelconque.

**Définition 3.** *Limites inférieures et supérieurs d'ensembles.*
*Soit une suite d'ensembles $(E_i)_{i\geq 0}$. On définit :*

- *la limite supérieure :* $\limsup_{i\to\infty} E_i = \bigcap_{i\geq 0} \bigcup_{j>i} E_j$.
- *la limite inférieure :* $\liminf_{i\to\infty} E_i = \bigcup_{i\geq 0} \bigcap_{j>i} E_j$.

**Exemple.** *x Prenons la suite $(E_i)_{i\geq 0}$ definie par*

$$E_{2k} = \Omega$$
$$E_{2k+1} = \emptyset$$

*On a immédiatement que*

- $\forall i \bigcup_{j>i} E_j = \Omega$ *et donc* $\limsup_{i\to\infty} E_i = \Omega$

- $\forall i \bigcap_{j>i} E_j = \emptyset$ *et donc* $\liminf_{i\to\infty} E_i = \emptyset$

**Proposition 2.**
*Soit une suite d'ensembles* $(E_i)_{i\geq 0}$*. On a*

1. $\overline{\liminf_i E_i} = \limsup_i \overline{E_i}$
2. $\liminf_i E_i \subset \limsup_i E_i$

**Preuve**

1. Decoule des proprietes $\overline{A\cup B} = \overline{A} \cap \overline{B}$ et $\overline{A\cap B} = \overline{A} \cup \overline{B}$
2. $\forall q > n$, $f$, $\bigcap_{k>n} E_k \subset E_q$, donc $\forall p$, $\bigcap_{k>n} E_k \subset \bigcup_{q>p} E_q$. On a donc $\forall n$

$$\bigcap_{k>n} E_k \subset \bigcap_{p\geq 0} \bigcup_{q>p} E_q = \limsup_{p\to\infty} E_p$$

Donc

$$\liminf_{n\to\infty} E_n = \bigcup_{n\geq 0} \bigcap_{k>n} E_k \subset \limsup_{p\to\infty} E_p$$

**Définition 4.** *Suite convergente d'ensembles.*
*Une suite d'ensembles* $(E_i)_{i\geq 0}$ *est dite convergente si*

$$\liminf_{n\to\infty} E_n = \limsup_{n\to\infty} E_n$$

**Proposition 3.** *Soit une suite d'ensembles* $(E_i)$ *et* $\omega \in \Omega$

1. *A partir d'un certain rang,* $\omega$ *est dans tous les* $E_j$ *s'écrit*

$$\omega \in \bigcup_{i\geq 0} \bigcap_{j>i} E_j = \liminf_{n\to\infty} E_n$$

2. $\omega$ *est dans une infinité de* $E_j$ *s'écrit*

$$\omega \in \bigcap_{i\geq 0} \bigcup_{j>i} E_j = \limsup_{n\to\infty} E_n$$

**Preuve**

1. Par définition de $\cup$ et $\cap$ :

$$\begin{aligned} \omega \in \bigcup_{i\geq 0} \bigcap_{j>i} E_j &\iff \exists i_0 \text{ t.q. } \omega \in \bigcap_{j>i_0} E_j \\ &\iff \exists i_0 \text{ t.q. } \forall j > i_0, \omega \in E_j \\ &\iff \text{ a partir d'un rang } i_0, \omega \text{ est dans tous les } E_j. \end{aligned}$$

2. Par définition de $\cup$ et $\cap$ :

$$\begin{aligned}\omega \in \bigcap_{i\geq 0}\bigcup_{j>i} E_j &\iff \forall i \geq 0, \omega \in \bigcup_{j>i} E_j \\ &\iff \forall i \geq 0, \exists j > i \text{ tq } \omega \in E_j \\ &\iff \omega \text{ est dans une infinité de } E_j.\end{aligned}$$

# 3 Dénombrabilité

On ajoute un rappel sur la notion de dénombrabilité qui quantifie un certain type d'infini. Les infinis que l'on peut *compter* sont dis dénombrables, et il existe des infinis *plus gros* que l'on ne peut pas *compter*.

**Définition 5.** *Un ensemble $E$ est dit dénombrable lorsqu'il est fini ou lorsqu'il est infini et que ses éléments peuvent être listés sans omission ni répétition dans une suite indexée par les entiers $n \in \mathbb{N}$.*
*Mathématiquement, un ensemble $E$ est dit dénombrable quand il est en bijection avec une partie de $\mathbb{N}$.*

**Proposition 4.** *On a les propriétés suivantes sur les ensembles dénombrables :*
- *Toute partie $A$ de $\mathbb{N}$.*
- *Le produit cartésien d'une famille finie non vide d'ensembles dénombrables est dénombrable.*
- *Une réunion dénombrable d'ensembles dénombrables est dénombrable.*

Une illustration simple est l'exemple de l'hotel de Hilbert.

**Exemple.** *Il y a un hôtel avec un infinité de chambre numérotées $1, 2, 3, \ldots$*
- *Un bus contenant une infinité de passagers numérotés $1, 2, 3, \ldots$ arrive devant l'hotel. Comment loger tous les passagers dans les chambres ?*
- *Deux bus contenant chacun une infinité de passagers numérotés $1, 2, 3, \ldots$ arrivent devant l'hotel. Comment loger tous les passagers dans les chambres ?*
- *$N$ bus contenant chacun une infinité de passagers numérotés $1, 2, 3, \ldots$ arrivent devant l'hotel. Comment loger tous les passagers dans les chambres ?*
- *Une infinité de bus numérotés $1, 2, 3, \ldots$, contenant chacun une infinité de passagers numérotés $1, 2, 3, \ldots$, arrivent devant l'hotel. Comment loger tous les passagers dans les chambres ?*

**Proposition 5.** *L'ensemble $[0, 1[$ n'est pas dénombrable.*

**Preuve.** Il suffit de démontrer que pour toute partie dénombrable $D$ de $[0,1[$, un élément de $[0,1[$ n'appartient pas à $D$.

Soit donc une partie dénombrable de $[0,1[$ énumérée à l'aide d'une suite $(u_n)_{n\in\mathbb{N}} = u_1, u_2, u_3, \ldots$. Chaque terme de cette suite a une écriture décimale avec une infinité de chiffres après la virgule, avec éventuellement une infinité de zéros pour un nombre décimal. On a donc

$$u_i = 0, u_{i,1}u_{i,2}\ldots u_{i,n}\ldots = \sum_{k=1}^{\infty} \frac{u_{i,k}}{10^k}$$

On construit maintenant un nombre réel $x \in [0,1[$ en utilisant $u_{n,n}$, le n-ième chiffre après la virgule (i.e. la n-ième décimale) de $u_n$. Par exemple, on a les écritures décimales suivantes pour les premiers termes de la suite

$$\begin{aligned}
u_1 &= 0,\mathbf{0}405440\ldots\\
u_2 &= 0,1\mathbf{4}23012\ldots\\
u_3 &= 0,83\mathbf{1}5036\ldots\\
u_4 &= 0,322\mathbf{0}436\ldots\\
u_5 &= 0,1407\mathbf{3}16\ldots\\
u_6 &= \ldots
\end{aligned}$$

Le nombre réel $x$ est construit par la donnée de ses décimales suivant par exemple la règle suivante. Si la n-ième décimale de $u_n$ est différente de 4, alors la n-ième décimale de x est 4, sinon la n-ième est 3. Avec l'exemple ci-dessus, la règle donne $x = 0,43444\ldots$.

Le nombre $x$ est dans l'intervalle $[0,1[$ mais ne peut pas être dans la suite $(u_n)_n$ car il n'est égal à aucun des nombres de la suite. En effet, il a au moins une décimales différente de chacun des $u_n$, a savoir la n-ieme.

Un nombre réel n'a pas une écriture décimale unique en général. En effet, pour les nombre décimaux non nuls, deux écritures sont possibles pour ces nombres, l'une avec toutes les décimales valant 0 sauf un nombre fini, l'autre avec toutes les décimales valant 9 sauf un nombre fini. Le nombre $x$ construit ci-dessus n'est pas décimal, puisque son écriture décimale est infinie et ne comporte que les chiffres 3 et 4. Il a donc une écriture décimale unique. On en conclut que $x \notin D$.

$[0,1[$ n'est donc pas dénombrable. A fortiori, $\mathbb{R}$, ou tout intervalle de $\mathbb{R}$ n'est pas dénombrable.

# Chapitre II

# Généralités sur les probabilités

## Plan du chapitre



## 1 Expériences et événements aléatoires

On cherche à modéliser les phénomènes aléatoires concrets. Une expérience aléatoire est une expérience dont on ne peut donner l'issue à l'avance.

**Définition 6.** *On appelle expérience aléatoire une expérience qui reproduite dans des conditions identiques, peut conduire a plusieurs résultats possibles, et dont on ne peut prévoir le résultat a l'avance. L'espace de tous les résultats possibles est noté $\Omega$, et un résultat possible est noté $\omega \in \Omega$.*

**Exemple.**

- *un lancer de pièce,* $\Omega = \{P, F\}$
- *un lancer de dé,* $\Omega = \{1, 2, 3, 4, 5, 6\}$

- *tirage de carte*

$$\Omega = \{1\ Pique, \ldots, K\ Pique, 1\ Coeur, \ldots, K\ Coeur,\\ 1\ Carreau, \ldots, K\ Carreau, 1\ Trèfle, \ldots, K\ Trèfle\}$$

- *un lancer de deux pièces,* $\Omega = \{PP, PF, FP, FF\}$
- *envoi d'une fléchette sur une cible de 30cm de diamètre*

$$\Omega = \{(0,) \bigcup (1, x, y),\ t.q.\ x^2 + y^2 \leq 15^2\}$$

**Définition 7.** *On appelle évènement aléatoire un sous ensemble de $\Omega$, dont on peut dire, au vu du résultat de l'expérience aléatoire, s'il est réalisé ou non.*

**Exemple.** *Supposons qu'on lance deux dés l'un après l'autre et que le résultat de l'expérience soir les deux chiffres des dés sans préciser l'ordre.*
- *L'ensemble $A = \{La\ somme\ des\ dés\ est\ supérieure\ a\ 6\}$ est un évènement aléatoire.*
- *L'ensemble $B = \{La\ le\ premier\ dé\ est\ égal\ a\ 4\}$ est un évènement aléatoire.*

Les évènements aléatoires sont donc des ensembles. On utilise donc les opérations élémentaire sur les ensembles pour décrire des évènements.

**Définition 8.** *Rappel de définitions sur les ensembles.*
- ***Intersection*** *: $A \bigcap B = \{c\ t.q.\ c \in A\ et\ c \in B\}$. L'intersection est associative et commutative.*
- ***Union*** *: $A \bigcup B = \{c\ t.q.\ c \in A\ ou\ c \in B\}$. L'union est associative et commutative.*
- ***Ensemble vide*** *: ensemble ne contenant aucun point, noté $\emptyset$.*
- ***Ensembles disjoints*** *: A et B tels que $A \bigcap B = \emptyset$.*
- ***Différence*** *: A
  $B = \{c\ t.q.\ c \in A\ et\ c \notin B\}$.*
- ***Complémentaire*** *: $A^c = \Omega$
  $A = \{a' \notin A\}$, aussi noté $\overline{A}$.*
- ***Partition*** *: Famille d'ensemble $A_i$ disjoints deux a deux, i.e. $A_i \bigcap A_j = \emptyset$, $i \neq j$ et qui recouvre $\Omega$, i.e. $\bigcup_i A_i = \Omega$.*

On peut donner aux opérations sur les ensembles définies ci-dessus une interprétation en terme d'évènement aléatoire en définissant les opérations suivantes.

**Définition 9.** *Rappel de définitions sur les ensembles.*
- ***NON*** *: l'évènement NON A correspond au complémentaire $A^c$ de A*
- ***OU*** *: l'évènement A OU B correspond a l'union $A \bigcup B$.*
- ***IMPLIQUE*** *: le fait que l'évènement A IMPLIQUE B signifie que $A \subset B$.*
- ***INCOMPATIBILITÉ*** *: A et B sont dits incompatibles si $A \bigcap B = \emptyset$.*
- ***TOUJOURS VRAI*** *: l'évènement $\Omega$ est le seul évènement qui soit toujours vrai.*
- ***IMPOSSIBLE*** *: a l'inverse, l'évènement $\emptyset$ est impossible.*

# 2 Espaces probabilisés

**Définition 10.** *Soit $\Omega$ un ensemble et $\mathcal{A}$ un ensemble de parties de $\Omega$. $\mathcal{A}$ est une tribu si*
- $\emptyset \in \mathcal{A}$ *et* $\Omega \in \mathcal{A}$.
- $\mathcal{A}$ *est stable par passage au complémentaire*

$$\text{Si } A \in \mathcal{A}, A^c \in \mathcal{A}$$

- $\mathcal{A}$ *est stable par intersection et union dénombrables*

$$A_n \text{ suite d'éléments de } \mathcal{A}, \bigcap_n A_n \in \mathcal{A}, \bigcup_n A_n \in \mathcal{A}$$

On comprend la nécessite d'avoir une tribu pour pouvoir appliques a notre ensemble d'évènements les opération NON, ET, OU, etc. définies précédemment.

**Exemple.**

- $\{\emptyset, \Omega\}$ *est la tribu triviale, la plus petite tribu de* $\Omega$
- $P(\Omega)$, *l'ensemble des parties de* $\Omega$, *est une tribu. Lorsque* $\Omega$ *est un ensemble dénombrable, on choisira cette tribu.*

**Définition 11.** *Si $C \subset \mathcal{P}(\Omega)$, on appelle tribu engendrée par $C$ la plus petite tribu contenant $C$. Elle existe toujours car $\mathcal{P}(\Omega)$ est une tribu.*

**Définition 12.** *On appelle tribu borelienne la tribu engendrée par l'ensemble des intervalles ouverts de $\mathbb{R}$*

**Exemple.** *Montrer que la tribu borélienne de $\mathbb{R}$ est engendrée par l'ensemble des intervalles de la forme $]-\infty, a]$ avec $a \in \mathbb{Q}$.*

On peut maintenant définir une probabilité

**Définition 13.** *Une probabilité sur l'espace $(\Omega, \mathcal{A})$ est une application $P$ de $\mathcal{A}$ dans $[0, 1]$ vérifiant*
- $P(\Omega) = 1$
- *Pour toute suite dénombrable $(A_n)_n$ d'éléments* ***deux à deux disjoints***

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

*Remarque.* Du point de vue de la théorie de la mesure, une probabilité est une mesure abstraite de masse 1 sur l'espace mesurable $(\Omega, \mathcal{A})$. Nous sommes donc dans le cadre mathématique développé par la théorie de la mesure, mais nous ne ferons pas appel aux résultats généraux de cette théorie.

Par application de la définition, on les propriétés suivantes.

**Proposition 6.**

- $P(\emptyset) = 0$
- $P(A) + P(A^c) = 1$
- $P(A) = P(A \setminus B) + P(A \cap B)$
- $P(A \bigcup B) = P(A \setminus B) + P(B \setminus A) + P(A \bigcap B)$
- $P(A \bigcup B) + P(A \bigcap B) = P(A) + P(B)$
- $P(A) \leq P(B)$ *si* $A \subset B$.

Cela nous amène a la définition d'espace abstrait de probabilité, ou espace probabilisé, que l'on doit au mathématicien russe Kolmogorov en 1933.

**Définition 14.** *On appelle espace (abstrait) de probabilité un triplet* $(\Omega, \mathcal{F}, \mathbb{P})$ *où*

- $\Omega$ *est un ensemble abstrait (univers)*
- $\mathcal{F}$ *est une tribu d'ensembles mesurables de* $\Omega$ *(les événements)*
- $\mathbb{P}$ *est une mesure positive sur* $(\Omega, \mathcal{F})$ *telle que* $\mathbb{P}(\Omega) = 1$ *(la probabilité)*

**Exemples.**

1. *Pour* $(\Omega, \mathcal{F})$ *mesurable et* $x \in \Omega$ *fixé,* $\delta_x$ *est une mesure de probabilité où* $\forall A \in \mathcal{A}$

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

   *En effet :*
   - $\delta_{x_n}$ *est bien a valeurs dans* $[0, 1]$
   - $\delta_{x_n}(\Omega) = 1$
   - *pour toute suite dénombrable* $(A_n)_n$ *d'éléments deux a deux disjoints :*
     - $*$ *soit* $\exists n_0$ *t.q.* $x \in A_{n0}$ *et* $P(\bigcup_n A_n) = 1 = \sum_n P(A_n)$.
     - $*$ *soit* $\forall n,\ x \notin A_n$ *et* $P(\bigcup_n A_n) = 0 = \sum_n P(A_n)$

2. *Pour* $(\Omega, \mathcal{F})$ *mesurable,* $(x_n)_n$ *une suite de points distincts de* $\Omega$ *et* $(p_n)_{n\in\mathbb{N}}$ *telles que* $p_n \geq 0$, $\sum\limits_{n\in\mathbb{N}} p_n = 1$, $\mathbb{P} = \sum\limits_{n\in\mathbb{N}} p_n \delta_{x_n}$ *est une mesure de probabilité.*
   *En effet :*
   - *La somme de deux mesure est un somme d'applications au sens*

$$(\mathbb{P}_1 + \mathbb{P}_2)(F) = \mathbb{P}_1(F) + \mathbb{P}_2(F)$$

   - $\forall A \in \mathcal{F}$, *la série* $\sum\limits_{n\in\mathbb{N}} p_n \delta_{x_n}(A)$ *est absolument convergente.*

- *On a*

$$\begin{aligned}\mathbb{P}(\Omega) &= \sum_{n\in\mathbb{N}} p_n \delta_{x_n}(\Omega)\\ &= \sum_{n\in\mathbb{N}} p_n \\ &= 1\end{aligned}$$

- *Pour tout suite $A_m$ deux a deux disjoints :*

$$\mathbb{P}(\bigcup_m A_m) = \sum_n p_n \delta_{x_n}(\bigcup_m A_m) = \sum_n \sum_m p_n \delta(A_m) = \sum_m (\sum_n p_n \delta_{x_n}(A_m)) = \sum_m \mathbb{P}(A_m)$$

*L'interversion des sommations sur $n$ et $m$ se fait par le théorème de Fubini pour les séries absolument convergentes.*

**Proposition 7** (Propriétés élémentaires)**.** *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, $A, B, A_n \in \mathcal{F}$.*

1. *Continuité monotone : Soit $(A_n)_{n\in\mathbb{N}}$ une suite monotone d'événements pour la relation d'inclusion :*

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

2. *Sous-additivité dénombrable :*

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

**Preuve. 1)**

a) Si $(A_n)_{n\in\mathbb{N}}$ est croissante $A_0 \subset A_1 \subset ...$
On pose $B_n = A_n \backslash A_{n-1}$ : $A_n = A_0 \bigcup \left(\bigcup\limits_{k=1}^{n} B_k\right)$ avec les $B_k$ disjoints.

$$\lim_{n\to+\infty} A_n = \bigcup_{n\in\mathbb{N}} A_n = A_0 \bigcup (\bigcup_{k=1}^{+\infty} B_k)$$

et

$$\begin{aligned}
\mathbb{P}(\lim_{n\to+\infty} A_n) &= \mathbb{P}(A_0) + \sum_{k=1}^{+\infty} \mathbb{P}(B_k) \quad (\text{ par definition de } \mathbb{P}) \\
&= \mathbb{P}(A_0) + \lim_{n\to+\infty} \sum_{k=1}^{n} \mathbb{P}(B_k) \\
&= \mathbb{P}(A_0) + \lim_{n\to+\infty} \sum_{k=1}^{n} (\mathbb{P}(A_k) - \mathbb{P}(A_{k-1})) \\
&= \lim_{n\to+\infty} \mathbb{P}(A_n).
\end{aligned}$$

b) Si $(A_n)_{n\in\mathbb{N}}$ est décroissante, $(A_n^c)_{n\in\mathbb{N}}$ est croissante.

$$\mathbb{P}(\lim_{n\to+\infty} A_n^c) = \lim_{n\to+\infty} \mathbb{P}(A_n^c) = 1 - \lim_{n\to+\infty} \mathbb{P}(A_n)$$

et

$$\mathbb{P}(\lim_{n\to+\infty} A_n^c) = \mathbb{P}(\bigcup_{n\in\mathbb{N}} A_n^c) = \mathbb{P}\left((\bigcap_{n\in\mathbb{N}} A_n)^c\right) = 1 - \mathbb{P}(\bigcap_{n\in\mathbb{N}} A_n) = 1 - \mathbb{P}(\lim_{n\to+\infty} A_n)$$

de sorte que

$$\mathbb{P}(\lim_{n\to+\infty} A_n) = \lim_{n\to+\infty} \mathbb{P}(A_n)$$

**2)**

$$\begin{aligned}
\mathbb{P}(\bigcup_{n\in\mathbb{N}} A_n) &= \mathbb{P}(\lim_{n\to+\infty} \bigcup_{k=1}^{n} A_n) \text{ par def. de la limite d'une suite croissance d'enemble} \\
&= \lim_{n\to+\infty} \mathbb{P}(\bigcup_{k=1}^{n} A_n) \text{ par continuité monotone} \\
&\leq \lim_{n\to+\infty} \sum_{k=1}^{n} \mathbb{P}(A_k) \text{ par sous-additivité de l'union finie} \\
&= \sum_{k=1}^{+\infty} \mathbb{P}(A_k)
\end{aligned}$$

**Définition 15.** *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace probabilisé.* $A \in \mathcal{F}$ *est dit avoir lieu* $\mathbb{P}$*-presque sûrement si* $\mathbb{P}(A) = 1$.

**Théorème 8** (Borel-Cantelli)**.** *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace probabilisé et* $(A_i)_{i\in\mathbb{N}}$ *une suite d'événements de* $\mathcal{F}$.

*i) Si* $\sum_{j=1}^{+\infty} \mathbb{P}(A_i) < +\infty$*, alors*

$$\mathbb{P}(\limsup_{n\to\infty} A_n) = 0.$$

*ii) Si* $\sum_{j=1}^{+\infty} \mathbb{P}(A_i) = +\infty$ *et si les événements* $A_j$ *sont indépendants, alors*

$$\mathbb{P}(\limsup_{n\infty} A_n) = 1.$$

**Preuve**

i) $\limsup_{n\to\infty} A_n = \bigcap_{n\in\mathbb{N}} \bigcup_{i\geq n} A_i$ donc

$$\begin{aligned}\mathbb{P}(\limsup_{n\to\infty} A_n) &= \lim_{n\to\infty} \mathbb{P}(\bigcup_{i\geq n} A_i) \text{ car } (\bigcup_{i\geq n} A_i)_n \text{ est décroissante}\\ &\leq \lim_{n\infty} \sum_{i\geq n} \mathbb{P}(A_i)\\ &= 0 \text{ si } \sum_{i=1}^{\infty} \mathbb{P}(A_i) < +\infty.\end{aligned}$$

ii) Si les $A_i$ sont indépendantes et si $\sum_{j=1}^{+\infty} \mathbb{P}(A_i) = +\infty$,

$$\mathbb{P}((\limsup_{n\infty} A_n)^c) = \mathbb{P}(\liminf_{n\infty} A_n^c) = \lim_{n\to\infty} \mathbb{P}(\bigcap_{i\geq n} A_i^c) = \lim_{n\to\infty} \lim_{m\to\infty} \mathbb{P}(\bigcap_{i=n}^{m} A_i^c)$$

puis

$$\begin{aligned}\mathbb{P}(\bigcap_{i=n}^{m} A_i^c) &= \prod_{i=n}^{m} \mathbb{P}(A_i^c) \text{ par indépendance}\\ &= \prod_{i=n}^{m} (1 - \mathbb{P}(A_i))\\ &\leq \prod_{i=n}^{m} e^{-\mathbb{P}(A_i)} \qquad \text{car } \forall x, 1 + x \leq e^x\\ &= e^{-\sum_{i=n}^{m} \mathbb{P}(A_i)}.\end{aligned}$$

On a donc $\lim_{m\to\infty} e^{-\sum_{i=n}^{m} \mathbb{P}(A_i)} = 0$ si $\sum_{j=1}^{+\infty} \mathbb{P}(A_i) = +\infty$. Ainsi $\lim_{m\to\infty} \mathbb{P}(\bigcap_{i=n}^{m} A_i^c) = 0$ et

$$\mathbb{P}((\limsup_{n\to\infty} A_n)^c) = 0 \text{ donc } \mathbb{P}(\limsup_{n\to\infty} A_n) = 1.$$

**Probabilités sur un espace dénombrable** Lorsque l'espace $\Omega$ est dénombrable, nous considérerons systématiquement, sauf exception, la tribu $\mathcal{P}(\Omega)$, qui est l'ensemble des parties de $\Omega$.

**Proposition 9.** *Soit $\Omega = (x_n)_n$ un ensemble dénombrable. Une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ est caractérisée par ses valeur sur sur les singletons. Ces valeurs $p_n = \mathbb{P}(x_n)$ constituent le terme d'une série positive de somme 1.*

**Preuve**

- Comme nous l'avons vu, si on note $\Omega = (x_n)_n$ et que l'on se donne $(p_n)_n$ positifs de somme 1, $P = \sum_n p_n \delta_{x_n}$ définit une probabilité.
- Inversement, soit une probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{P}(\Omega))$. Toute partie $A = (x_n), n \in \mathbb{N}$ de $\Omega$ est réunion disjointe des singletons $\{x_n, n \in \mathbb{N}\}$, donc $\mathbb{P}(A)$ est caractérisée par $p_n = \mathbb{P}(x_n) \geq 0$. De plus $\sum_n p_n = \mathbb{P}(\Omega) = 1$.

## 3 Variables aléatoires

En théorie moderne des probabilités, on préfère utiliser un point de vue fonctionnel plutôt qu'ensembliste en utilisant les variables aléatoires. **Une variable aléatoire est une grandeur qui dépend du résultat de l'expérience que l'on regarde.**

**Définition 16.** *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Une variable aléatoire $X$ est une application de $\Omega$ dans un ensemble $E$.*

$$X : \omega \in \Omega \mapsto X(\omega) \in E$$

*Remarques.*

- Le terme variable dans variable aléatoire est assez mal adapte, car une variable aléatoire est une fonction de la variable $\omega \in \Omega$
- L'espace E est mieux connu dans la pratique et nous nous interesserons aux chances de réalisation de $X$ plutot que les chances de réalisations d'évenements aléatoires
- La variable aléatoire $X$ nous permet de transporter la structure $(\Omega, \mathcal{F}, \mathbb{P})$ dans l'ensemble $E$ en posant, pour $B \bigcap E$

$$X^{-1}(B) = \{\omega \in \Omega \text{ t.q. } X(\omega) \in B\}$$

et en utilisant la probabilite

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$$

**Proposition 10.**

- *L'ensemble* $\mathcal{E} = \{B \bigcap E \text{ t.q. } X^{-1}(B) \in \mathcal{F}\}$ *est une tribu de E.*
- *L'application* $\mathbb{P}_X$ *definie par*

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$$

*est une probabilite definie sur la tribu* $\mathcal{E}$.

**Preuve** Il suffit de remarquer que les operations elementaires sur les ensembles sont transportees par $X^{-1}$

- $X^{-1}(\emptyset) = \emptyset$
- $X^{-1}(E) = \Omega$
- $X^{-1}(B^c) = (X^{-1}(B))^c$
- $X^{-1}(\bigcap_i A_i) = \bigcap_i X^{-1}(A_i)$
- $X^{-1}(\bigcup_i A_i) = \bigcup_i X^{-1}(A_i)$

Du fait que $\mathcal{F}$ est une tribu,$\mathcal{E}$ en est un aussi. Idem pour la probabilité $\mathbb{P}_X$.

**Définition 17.** *La probabilité* $\mathbb{P}_X$ *est appelée loi de la variable aléatoire* $X$*, ou distribution de probabilité de* $X$*. Du point de vue de la théorie de la mesure, c'est la mesure image de* $\mathbb{P}$ *par* $X$*.*

*Remarque.* L'interet majeur est que $\mathbb{P}_X$ sera plus facile a caractériser que $\mathbb{P}$ car l'espace $E$ ($\mathbb{R}$ par ex.) est mieux connu que l'espace abstrait $\Omega$ en général.

# 4 Fonction de répartition

Nous avons vu que lorsque l'espace $\Omega$ est dénombrable, ou lorsque l'image $X(\Omega)$ est dénombrable, la loi est de $X$ est caractérisée par les probabilités sur les singletons, qui forme le terme général d'une série convergente.

Dans le cas général d'une variable aléatoire réelle, l'image $X(\Omega)$ est $\mathbb{R}$ qui n'est pas dénombrable, et on ne peut pas caratériser la loi par les probabilités sur les singletons. Dans ce cas, la fonction de répartition est un outil de caractérisation des lois de variables aléatoires réelles.

**Définition 18.** *Soit* $X$ *une v.a.r. sur* $(\Omega, \mathcal{F}, \mathbb{P})$*. On appelle* **fonction de répartition** *de* $X$ *(ou de* $\mathbb{P}_X$*), la fonction*

$$\begin{aligned} F_X : \mathbb{R} &\to \mathbb{R} \\ t &\mapsto \mathbb{P}_X(]-\infty, t]) = \mathbb{P}(X \leq t) \end{aligned}$$

**Proposition 11** (Propriétés de $F_X$)**.** .

1. $F_X(t) \in [0, 1]$, $\forall t \in \mathbb{R}$

2. *$F_X$ est croissante*
3. $\lim_{t\to-\infty} F_X(t) = 0$ *et* $\lim_{t\to+\infty} F_X(t) = 1$
4. *$F_X$ est continue à droite :* $\forall t \in \mathbb{R}$, $\lim_{h\to0^+} F_X(t+h) = F_X(t)$
5. *$F_X$ a une limite à gauche :* $\forall t \in \mathbb{R}$, $F_X(t^-) = \lim_{h\to0^+} F_X(t-h)$ *existe et*

$$F_X(t) - F_X(t^-) = \mathbb{P}(X = t).$$

**Preuve de 3.** On pose $A_n = \{X \le -n\}$ : $(A_n)_{n\in\mathbb{N}}$ est décroissante et $\lim_{n\to+\infty} A_n = \varnothing$ : $\lim_{n\to+\infty} \mathbb{P}(A_n) = 0$. Ainsi $\lim_{n\to+\infty} F_X(-n) = 0$ d'où $\lim_{t\to-\infty} F_X(t) = 0$ car $F_X$ est croissante.

Puis on pose $B_n = \{X \le n\}$ : $(B_n)_{n\in\mathbb{N}}$ est croissante et $\lim_{n\to+\infty} B_n = \Omega$ : $\lim_{t\to-\infty} F_X(t) = \lim_{n\to+\infty} F_X(n) = 1$.

**Preuve de 4.** Soit $t \in \mathbb{R}$ et $h > 0$.

$$0 \le \mathcal{F}_X(t+h) - \mathcal{F}_X(t) = \mathbb{P}_X(]t, t+h]) \le \mathbb{P}_X(]t, t+\frac{1}{n}]) \text{ pour } h \le \frac{1}{n}.$$

$(]t, t+\frac{1}{n}])_n$ est décroissante donc $\lim_{n\to+\infty} \mathbb{P}_X(]t, t+\frac{1}{n}]) = 0$.

**Preuve de 5.** On pose $B_n = \{X \le t - \frac{1}{n}\} \subset B_{n+1}$. $\lim_{n\to+\infty} \mathbb{P}(B_n) = \mathbb{P}(\lim_{n\to+\infty} B_n) = \mathbb{P}(X < t)$ donc

$$\lim_{h\to0^+} F_X(t-h) = \lim_{n\to+\infty} F_X(t - \frac{1}{n}) = \mathbb{P}(X < t)(= F_X(t^-))$$

et $\lim_{h\to0^+} F_X(t) - F_X(t-h) = F_X(t) - F_X(t^-) = \mathbb{P}(X = t)$.

**Réciproquement :**

**Théorème 12.** *Soit $F : \mathbb{R} \to \mathbb{R}$ une fonction croissante, continue à droite, vérifiant* $\lim_{x\to-\infty} F(x) = 0$ *et* $\lim_{x\to+\infty} F(x) = 1$.

*Alors il existe un espace $(\Omega, \mathcal{F}, \mathbb{P})$ et $X$ une variable aléatoire définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ telle que $F_X = F$.*

**Preuve** On pose

$$(\Omega, \mathcal{F}, \mathbb{P}) = (]0,1[, \mathcal{B}(]0,1[), \lambda)$$

et

$$\forall \omega \in ]0,1[, X(\omega) = \inf\{x \in \mathbb{R} : \omega \le F(x)\}.$$

Comme $F$ est croissante continue à droite, on a

$$X(\omega) \leq t \Leftrightarrow \omega \leq F(t)$$

donc $\mathbb{P}(X(\omega) \leq t) = \lambda(\omega \leq F(t)) = \lambda([0, F(t)]) = F(t)\square$

**Exemples.** *Fonctions de répartition :*

1. $F_X(t) = \mathbf{1}_{[x,+\infty[}(t)$ *si* $\mathbb{P}_X = \delta_x$
2. $$F_X(t) = \begin{cases} 0 & si \quad t \leq 0 \\ t & si \quad 0 \leq t \leq 1 \\ 1 & si \quad t \geq 1 \end{cases}$$

   *si* $\mathbb{P}_X \sim U[0,1]$.

# 5 Variables aléatoires à densité

La fonction de répartition est un outil général de caractérisation de la loi des variables aléatoires, mais peut être un peu limitant pour les calculs, notamment lorsqu'on veux calculer la loi de $f(X)$ pour $f$ une fonction donnée.

Les densités de probabilités permettent de calculer des probabilités via des intégrales su des intervalles bornés $[a, b]$ ou sur $\mathbb{R}$ tout entier, c'est à dire qu'il s'agit d'intégrales généralisées. Cela signifie que

**Définition 19.** *Une fonction réelle $f : \mathbb{R} \to \mathbb{R}$ est une densité de probabilité si elle est positive, intégrable et vérifie*

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Si $f$ est une densité de probabilité, elle vérifie les propriétés d'une fonction de répartition. C'est donc la fonction de répartition d'une probabilité.

**Définition 20.** *Une variable aléatoire $X$ est de densité $f$ si la loi $\mathbb{P}_X$ admet la densité de probabilité $f$, c'est a dire que*

$$\mathbb{P}(x \leq x) = \int_{-\infty}^{x} f(y)dy$$

Les variable aléatoires réelles ont les propriétés suivantes.

**Proposition 13.** *Soit $X$ une v.a. de densité $f$*

- *La fonction de répartition est continue, et donc*

$$\forall x \in \mathbb{R}, \mathbb{P}(X = x) = 0$$

- *Sa fonction de répartition $F$ est dérivable en tout point et*

$$F'(x) = f(x)$$

- *Inversement, si la fonction de répartition $F$ est dérivable par morceau, alors la v.a. $X$ admet la densité $f(x) = F'(x)$*

**Exemple.** *Une variable aléatoire de loi normale $\mathcal{N}(m, \sigma^2)$, $m \in \mathbb{R}$, $\sigma^2 > 0$ admet pour densité*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \text{ et } F(t) = \int_{-\infty}^{t} f(x)dx.$$

# 6 Espérance

## 6 .1 Espérance

**Définition 21** (Espérance)**.** *Pour $X$ une v.a.r sur $(\Omega, \mathcal{F}, \mathbb{P})$, on dit que $X$ est d'espérance finie si*

$$\int_{\Omega} |X(\omega)| d\mathbb{P}(\omega) < +\infty$$

*(i.e. $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. On pose alors*

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

*Remarques.*

- On dit que $X$ est centrée si $\mathbf{E}[X] = 0$.
- Pour $p \geq 0$, le moment absolu d'ordre $p$ est $\mathbf{E}[|X|^p] = \int_{\Omega} |X(\omega)|^p d\mathbb{P}(\omega)$.
- Sous réserve d'existence, le moment d'ordre $n \in \mathbb{N}$ est

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

**Théorème 14** (de transfert)**.** *Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{E})$ une variable aléatoire et $\varphi : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ mesurable. Nous savons que $Z = \varphi \circ X$ est une variable aléatoire. Supposons que $Z \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, c'est à dire que*

$$\int_{E} |\varphi(x)| d\mathbb{P}_X(x) < +\infty.$$

*On a alors égalité des espérances de h relativement à la probabilité $\mathbb{P}_X$ et de $h(X)$ relativement à la probabilité $\mathbb{P}$, c'est a dire que*

$$\mathbf{E}[\varphi(X)] = \int_\Omega \varphi(X(\omega))d\mathbb{P}(\omega) = \int_E \varphi(x)d\mathbb{P}_X(x).$$

**Preuve**

a) Pour $\varphi = \alpha \mathbf{1}_B$ :

$$\begin{aligned}
\int_\Omega \varphi(X(\omega))d\mathbb{P}(\omega) &= \int_\Omega \alpha \mathbf{1}_B(X(\omega))d\mathbb{P}(\omega) \\
&= \alpha \int_\Omega \mathbf{1}_{X^{-1}(B)}(\omega)d\mathbb{P}(\omega) \\
&= \alpha \mathbb{P}(X^{-1}(B)) \\
&= \alpha \mathbb{P}_X(B) \\
&= \alpha \int_E \mathbf{1}_B(x)d\mathbb{P}_X(x) \\
&= \int_E \varphi(x)d\mathbb{P}_X(x).
\end{aligned}$$

b) Par linéarité de l'intégrale, c'est vrai pour $\varphi$ étagée.

c) Par passage à la limite, c'est vrai pour $\varphi$ positive (donc pour $|\varphi|$ d'où l'équivalence).

d) Pour $\phi \in \mathcal{L}^1(\mathbb{P}_X)$, $\varphi = \varphi_+ - \varphi_-$ et la formule est vrai par linéarité.

*Remarque.* Ce théorème de transfert signifie que, si nous connaissons la loi de la v.a. $X$, nous pouvons calculer $\mathbf{E}(h(X))$ quelque soit la fonction $h$ intégrable.

**Exemple.** *Si $X \in \{x_1, ..., x_n\}$,*

$$\mathbf{E}[X] = \int_\Omega X(\omega)d\mathbb{P}(\omega) = \int_E x d\mathbb{P}_X(x) = \sum_{i=1}^n x_i \mathbb{P}_X(x_i) = \sum_{i=1}^n x_i \mathbb{P}(X = x_i)$$

*C'est la moyenne pondérée des valeurs de $X$. On a également*

$$\mathbf{E}[X^2] = \int_E x^2 d\mathbb{P}_X(x) = \sum_{i=1}^n x_i^2 \mathbb{P}(X = x_i).$$

**Cas de v.a. à densité**

**Proposition 15.** *Soit $X$ une v.a. réelle admettant la densité $f$ et $g$ une fonction mesurable de $\mathbb{R}$ dans $\mathbb{R}$. Alors $g(X) \in L^1$ si et seulement si*

$$\int_{-\infty}^{+\infty} |g(x)|f(x)dx < \infty$$

*et alors*

$$\mathbf{E}(g(X)) = \int_{-\infty}^{+\infty} g(x)f(x)dx$$

*Remarque.* On voit ici l'intérêt des variables aléatoires a densité, qui ramènent les calculs d'espérance à des calculs d'intégrales, qui peuvent être traités efficacement numériquement.

## 6 .2 Variance

**Définition 22** (Variance)**.** *Pour $X$ v.a.r. (%$(\Omega, \mathcal{F}, \mathbb{P})$) telle que $\mathbf{E}[|X|] < +\infty$, on appelle* **variance** *de $X$ la quantité*

$$\mathbf{Var}(X) = \mathbf{E}\left[(X - \mathbf{E}[X])^2\right] \in [0, +\infty].$$

**Proposition 16.**

1. *Pour $(a, b) \in \mathbb{R}^2$et $Y = aX + b$,*

$$\mathbf{E}[Y] = a\mathbf{E}[X] + b$$

*et*

$$\mathbf{Var}(Y) = \mathbf{E}\left[(aX + b - a\mathbf{E}[X] - b)^2\right] = a^2\mathbf{Var}(X).$$

2.

$$\mathbf{Var}(X) = 0 \Leftrightarrow \int_\Omega (X(\omega) - \mathbf{E}[X])^2 d\mathbb{P}(\omega) = 0 \Leftrightarrow X(\omega) = \mathbf{E}[X] \textit{ presque sûrement}$$

$$\Leftrightarrow X = \textit{ Cste p.s.}$$

3. $\mathbf{Var}(X) < +\infty \Leftrightarrow X - \mathbf{E}[X] \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P}) \Leftrightarrow X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$
4. **Formules** *:*
   a) $\mathbf{Var}(X) = \int_E (x - \mathbf{E}[X])^2 d\mathbb{P}_X(x)$
   b) *Dans le cas ou $X$ admet une densité $f$,*

$$Var(X) = \int_{-\infty}^{+\infty} (x - \mathbf{E}(x))^2 f(x)dx$$

*c)* $\mathbf{Var}(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2$

**Proposition 17.** *Soient $X$ et $Y$ des v.a.r $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. On a*

$$\mathbf{Var}(X+Y) = \mathbf{Var}(X) + \mathbf{Var}(Y) + 2\mathbf{cov}(X,Y)$$

*où* $\mathbf{cov}(X,Y) = \mathbf{E}\left[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])\right]$ *est la* **covariance** *de $X$ et $Y$.*

**Preuve** : $(X + Y - \mathbf{E}[X+Y])^2 = ((X - \mathbf{E}[X]) + (Y - \mathbf{E}[Y]))^2 \square$

*Remarque.* On a $\mathbf{cov}(X,X) = \mathbf{Var}(X)$

**Interprétation de $\mathbf{cov}(X,Y)$** Si $\mathbf{cov}(X,Y) > 0$, alors "$X > \mathbf{E}[X] \Rightarrow Y > \mathbf{E}[Y]$" et "$X < \mathbf{E}[X] \Rightarrow Y < \mathbf{E}[Y]$" en moyenne : $X$ et $Y$ sont liées dans le même sens.

**Définition 23** (Variance de vecteur aléatoire)**.** *Si $X = (X_1, ..., X_d)$ est un vecteur aléatoire de $(\Omega, \mathcal{F}, \mathbb{P})$, et si $\max_{i=1...d} \mathbf{E}[|X_i|^p] < +\infty$, on dit que $X$ a un moment d'ordre $p$. L'espérance de $X$ est le vecteur de $\mathbb{R}^d$ $\mathbf{E}[X] = (\mathbf{E}[X_1], ..., \mathbf{E}[X_d])$.*

*La variance de $X$ est la matrice de variance-covariance*

$$\Gamma_X = \left(\mathbf{E}[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])]\right)_{i,j=1,...,d}.$$

## 6 .3 Inégalités portant sur l'espérance

**Proposition 18.**

1. **Inégalité de Markov** *:Pour $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, $\forall t > 0$,*

$$\mathbb{P}(|X| \geq t) \leq \frac{\mathbf{E}[|X|]}{t}.$$

2. **Inégalité de Tchebychev** *: Pour $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ et $\varepsilon > 0$,*

$$\mathbb{P}(|X - \mathbf{E}[X]| \geq \varepsilon) \leq \frac{\mathbf{Var}(X)}{\varepsilon^2}.$$

**Preuve**

1.

$$\begin{aligned}
\mathbf{E}[|X|] &= \int\limits_{\Omega} |X(\omega)| d\mathbb{P}(\omega) = \int\limits_{\Omega} |X(\omega)| \mathbf{1}_{|X(\omega)|\geq t} d\mathbb{P}(\omega) + \int\limits_{\Omega} |X(\omega)| \mathbf{1}_{|X(\omega)|<t} d\mathbb{P}(\omega) \\
&\geq \int\limits_{\Omega} |X(\omega)| \mathbf{1}_{|X(\omega)|\geq t} d\mathbb{P}(\omega) \\
&\geq t \int\limits_{\Omega} \mathbf{1}_{|X(\omega)|\geq t} d\mathbb{P}(\omega) \\
&= t\mathbb{P}(|X| \geq t).
\end{aligned}$$

2. $\mathbb{P}(|X - \mathbf{E}[X]| \geq \varepsilon) = \mathbb{P}(|X - \mathbf{E}[X]|^2 \geq \varepsilon^2) \leq \frac{\mathbf{E}[|X - \mathbf{E}[X]|^2]}{\varepsilon^2}$ avec 1).

## 6 .4 Théorème pour l'identification de la densité

**Théorème 19** (Identification de la densité)**.** *Si il existe une fonction $f \geq 0$ telle que pour toute fonction continue bornée $h$ on ait*

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

*alors la v.a. $X$ admet la densité $f$.*

**Preuve** Si l'on prend pour $h$ la fonction $\mathbf{1}_{]-\infty,y]}$, on a

$$\mathbb{P}(X \leq y) = \int_{-\infty}^{y} f(x) dx$$

c'est a dire la fonction de répartition associée a la densité $f$. $\mathbf{1}_{]-\infty,y]}$ n'est pas continue bornée mais on peut l'approcher uniformément par une suite de fonction continue bornées.

**Conseil pratique** : Pour trouver la densité $f_X$ de $X$, calculer $\mathbf{E}[\varphi(X)]$ pour toute fonction borélienne positive (ou bornée) et mettre sous la forme $\int_{\mathbb{R}^d} \varphi(x) h(x) dx$. $h$ est alors la densité de $X$.

# Chapitre III

# Conditionnement et indépendance

## Plan du chapitre



Les notions de conditionnement et d'indépendance sont des notions propres a la theorie des probabilites et qui differencie cette theorie de la theorie de la mesure. L'idee de base de cette notion est la suivante : une information supplementaire concernant l'experience modifie la probabilite de l'evenement que l'on etudie.

Par exemple, lorque l'on lance deux des, la probabilité que la somme des faces soit superieure ou egale a 11 est de 1/12. Cette probabilité est exactement nulle si l'on sait que l'un de des a une face egale a 4, et elle est de 1/3 si l'un de des est egal a 6.

## 1 Conditionnement

**Définition 24.** *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *espace probabilise et* $A, B \in \mathcal{F}$ *avec* $\mathbb{P}(B) > 0$. *On appelle probabilité conditionnelle de* $A$ *sachant* $B$ *le nombre*

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

*Remarques.*

- Si l'on revient a l'intuition des frequences empiriques pour les probabilites ($\mathbb{P}(A) = \lim_n f_n(A)$), la frequence de realisation de l'evenement $A$ sachant que $B$ est realise est la frequence de realisation des evenements $A$ **ET** $B$ divise par la frequence de realisation de $B$

$$f_n(A|B) = \frac{f_n(A \cap B)}{f_n(B)}$$

et on retrouve la definition.

- Au delà de l'impossibilité de diviser par 0, cette notion n'a pas de sens si un évènement n'est jamais réalisé.
- Si $\mathbb{P}(A) > 0, \mathbb{P}(B) > 0,$, on a

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

**Proposition 20.** *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ espace probabilise et $B \in \mathcal{F}$ un evenement avec $\mathbb{P}(B) > 0$. L'application*

$$\mathbb{P}(.|B) : A \in \mathcal{F} \mapsto = \mathbb{P}(A|B)$$

*definit une probabilite.*

**Preuve**

- $\mathbb{P}(.|B)$ est positive et $\mathbb{P}(.|B) \leq 1$ car $A \cap B \subset B$ donc $\mathbb{P}(A \cap B) \leq \mathbb{P}(B)$
- $\mathbb{P}(\Omega|B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$
- On a $(\cup_n A_n) \cap B = \cup_n (A_n \cap B)$, d'ou $\mathbb{P}(\cup_n A_n|B) = \sum_n \mathbb{P}(A_n|B)$ pour $A_n$ disjoints

**Proposition 21. *Formule de bayes***
*Soit $B_i \in \mathcal{F}$ une partition denombrable de $\Omega$ telle que $P(B_i) > 0$. Pour tout $A \in \mathcal{F}$,*

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

**Proposition 22. *Formule des probabilités composées***
*Pour $A_1, ..., A_n \in \mathcal{F}$,*

$$\mathbb{P}(A_1 \cap A_2 \cap ... \cap A_n) = \mathbb{P}(A_n|A_1 \cap ... \cap A_{n-1})...\mathbb{P}(A_2|A_1)\mathbb{P}(A_1).$$

**Preuve** Pa recurrence.

- n=2 par defintion de la proba conditionnelle
- On pose $B = A_0 \cap \ldots \cap A_{n-1}$, on applique la propriété pour $n = 2$ puis l'hypothèse de reccurence.

**Exemple** (Schéma des urnes)**.** *$\mathcal{E}$ : "On tire successivement* 3 *boules d'une urne contenant* 5 *noires,* 3 *blanches,* 2 *rouges, sans remise". $T_i$ : $i^{ème}$ tirage.*

$$\begin{aligned}
\mathbb{P}("NNR") &= \mathbb{P}(T_3 = R|T_1 = N, T_2 = N)\mathbb{P}(T_2 = N|T_1 = N)\mathbb{P}(T_1 = N) \\
&= \frac{2}{8} \times \frac{4}{9} \times \frac{5}{10} \\
&= \frac{1}{18}.
\end{aligned}$$

**Proposition 23.** ***Formule des probabilités totales***
*Soit $B_i \in \mathcal{F}$ une partition denombrable de $\Omega$ telle que $P(B_i) > 0$. Pour tout $A \in \mathcal{F}$,*

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap B_i) = \sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

**Preuve** $A = \cup_i (A \cap B_i)$ et les $(A \cap B_i)$ sont deux a deux disjoints (car les $B_i$ le sont), et $P(A \cap B_i) = \mathbb{P}(A|B_i)\mathbb{P}(B_i)$. Puis On applique la définition d'une probabilité.

**Exemple.** *Un individu se fait tester pour une maladie sachant que la proportion de la population infectee est $10^{-4}$. On lui dit que si il est effectivement infecte, il y a 99% de chance que le test soit positif, et si il n'est pas infecte, il y a 0.1% de chances d'avoir un test positif. Le test est positif, quelle est la probabilite pour que l'individu soit infecte ?*

*Soit A l'evenement "le test est positif" et B "l'individu est infecte". On veut connaitre ici $P(B|A)$.*

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

- *$P(B) = 10^{-4}$ et $P(A|B) = 0.99$ donc $P(A|B)P(B) = 9.9.10^{-5}$*
- *Par formule de probabilite totale*

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c) = 9.910^{-5} + 0.001 \times 9.999.10^{-4}$$

*Au total $P(B|A) = 1/11$, ce qui modère l'inquiétude du patient.*

# 2 Independance

## 2 .1 Evenements

**Définition 25** (Indépendance). *Pour $A, B \in \mathcal{F}$, on dit que $A$ et $B$ sont indépendants si*

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

**Proposition 24.**

- *Si $\mathbb{P}(B) > 0$, on a alors $\mathbb{P}(A|B) = \mathbb{P}(A)$.*
- *l'independance est stable par passage au complementaire*
- *l'independance est stable par applicaion d'une variable aleatoire.*

**Définition 26.** *Pour une famille quelconque d'événements $A_i \in \mathcal{F}$, $i \in I$, les $A_i$ sont indépendants si*

$$\forall J \text{ finie } \subset I, \quad \mathbb{P}(\bigcap_{j\in J} A_j) = \prod_{j \in J} \mathbb{P}(A_j).$$

**Propriétés** Si les $(A_i)_{i\in I}$ sont indépendants,

i) Pour $J_1, J_2$ finies $\subset I$, $J_1 \cap J_2 = \varnothing$, les événements $A_{J_1} = \bigcap_{j\in J_1} A_j$ et $A_{J_2} = \bigcap_{j\in J_2} A_j$ sont indépendants.

ii) On peut remplacer certains $A_i$ par $A_i^c$ sans perdre l'indépendance.

## 2 .2 Variables aléatoires

**Définition 27.** *Une famille quelconque de variables aléatoires sur* $(\Omega, \mathcal{F}, \mathbb{P})$ $X_i : (\Omega, \mathcal{F}) \to (E_i, \mathcal{E}_i)$ *est indépendante si la famille des tribus engendrées par les* $X_i$ *l'est :*

$$\forall J \text{ finie } \subset I, \quad \forall B_i \in \mathcal{E}_i, i \in J, \quad \mathbb{P}(\bigcap_{j\in J}\{X_i \in B_i\}) = \prod_{j\in J} \mathbb{P}(X_j \in B_i).$$

$$(\Leftrightarrow \forall B_i \in \mathcal{E}_i, i \in I, \qquad \text{les événements } X_i^{-1}(B_i) \text{ sont indépendants .})$$

**Propriétés**

i) Pour $(X_i)_{i\in I}$ indépendantes et $\varphi_i : (E_i, \mathcal{E}_i) \to (\tilde{E}_i, \tilde{\mathcal{E}}_i)$ mesurables, $(\varphi_i(X_i))_{i\in I}$ est une famille de variables indépendantes.

ii) Pour $A_1, ..., A_n \in \mathcal{F}$,

$$(A_i)_{i=1,...,n} \text{ indépendants } \Leftrightarrow (\mathbf{1}_{A_i})_{i=1,...,n} \text{ indépendantes .}$$

**Théorème 25** (Fondamental)**.** *Pour* $(X_i)_{i=1,...,n}$ *variables aléatoires, on pose*

$$Z = (X_1, ..., X_n) : (\Omega, \mathcal{F}) \to (\prod_{i=1}^{n} E_i, \mathcal{E}_1 \otimes ... \otimes \mathcal{E}_n).$$

$$\text{Les variables } X_i \text{ sont indépendantes } \Leftrightarrow \mathbb{P}_Z = \mathbb{P}_{X_1} \otimes ... \otimes \mathbb{P}_{X_n}.$$

**Preuve du théorème fondamental** On utilise le fait que $\forall B_i \in \mathcal{E}_i$,

$$\mathbb{P}_Z(B_1 \times ... \times B_n) = \mathbb{P}(Z^{-1}(B_1 \times ... \times B_n)) = \mathbb{P}(\bigcap_{i=1}^{n} X_i^{-1}(B_i)).$$

En cas d'indépendance, $\mathbb{P}_Z$ et $\mathbb{P}_{X_1} \otimes ... \otimes \mathbb{P}_{X_n}$ coïncident sur un $\pi$-système qui engendre $\mathcal{E}_1 \otimes ... \otimes \mathcal{E}_n$.

**Corollaire 26.** *Soit* $(X_1, ..., X_d)$ *un vecteur aléatoire réel sur* $(\Omega, \mathcal{F}, \mathbb{P})$. *Les variables aléatoires* $X_1, ..., X_d$ *sont indépendantes si et seulement si pour toute famille de fonctions* $\varphi_i : \mathbb{R} \to \mathbb{R}$ *mesurables telles que* $\varphi_i(X_i) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, $\forall J \subset \{1, ..., d\}$,

$$\mathbf{E}\left[\prod_{j\in J} \varphi_j(X_j)\right] = \prod_{j\in J} \mathbf{E}\left[\varphi_j(X_j)\right].$$

**Preuve du corollaire** Si les $X_i$ sont indépendantes, quitte à renuméroter, $J = \{1, ..., j\}$ et

$$\begin{aligned}\mathbf{E}\left[\prod_{i\in J}\varphi_i(X_i)\right] &= \int\prod_{i=1}^{j}\varphi_i(x_i)d\mathbb{P}_{(X_1,...,X_j)}(x_1,...,x_j) \text{ (Théorème de transfert)}\\ &= \int\prod_{i=1}^{j}\varphi_i(x_i)d\mathbb{P}_{X_1}(x_1)...d\mathbb{P}_{X_j}(x_j) \text{ (par indépendance)}\\ &= \prod_{i=1}^{j}\int\varphi_i(x_i)d\mathbb{P}_{X_i}(x_i)\\ &= \prod_{i\in J}\mathbf{E}\left[\varphi_i(X_i)\right].\end{aligned}$$

Pour la réciproque, en prenant $\varphi_j = \mathbf{1}_{B_j}$, la formule donne directement que $\mathbb{P}_Z = \mathbb{P}_{X_1} \otimes ... \otimes \mathbb{P}_{X_d}$ coïncident sur $B_1 \times ... \times B_d$.

**Corollaire 27.** *Si les variables $X_1, ..., X_n$ sont indépendantes et intégrables,*

$$\mathbf{E}[X_1X_2...X_n] = \mathbf{E}[X_1]\mathbf{E}[X_2]...\mathbf{E}[X_n].$$

**Théorème 28** (Fondamental, Cas discret)**.** *Si les $(X_i)_{i\in I}$ sont discrètes, à valeurs dans $E$,*

$$\mathbb{P}_Z = \mathbb{P}_{X_1} \otimes ... \otimes \mathbb{P}_{X_n} \Leftrightarrow \forall (x_j)_{j=1,...,n} \in E^n, \mathbb{P}(\bigcap_{j=1}^{n}\{X_j = x_j\}) = \prod_{j=1}^{n}\mathbb{P}(X_j = x_j).$$

**Théorème 29** (Fondamental, Cas à densité)**.** *Si $Z = (X_1, ..., X_n)$ possède la densité $f_Z$, on note $f_{X_j}$ la densité marginale de $X_j$ et on a*

$$\begin{aligned}\mathbb{P}_Z = \mathbb{P}_{X_1} \otimes ... \otimes \mathbb{P}_{X_n} &\Leftrightarrow f_Z(x_1,...,x_n)dx_1...dx_n = f_{X_1}(x_1)dx_1...f_{X_n}(x_n)dx_n\\ &\Leftrightarrow f_Z(x_1,...,x_n) = f_{X_1}(x_1)...f_{X_n}(x_n) \quad \lambda - \textit{presque partout.}\end{aligned}$$

**Proposition 30.** *Si $X$ et $Y$ sont indépendantes, alors $\mathbf{cov}(X, Y) = 0$ ($X$ et $Y$ ne sont pas corrélées). La réciproque est fausse.*

**Preuve**

- $\Rightarrow$ $\mathbf{cov}(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$
- $\Leftarrow$ Soit $X \sim \mathcal{N}(0, 1) \sim f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, \forall x \in \mathbb{R}$ et $Y = X^2$.

$$\mathbf{cov}(X, Y) = \mathbf{E}[X^3] - \mathbf{E}[X]\mathbf{E}[X^2] = \mathbf{E}[X^3] = \int_{\mathbb{R}} x^2 f_X(x)dx = 0$$

car $f_X$ est paire. Mais

$$\mathbb{P}(X \geq 1 \cap Y \geq 1) = \mathbb{P}(X \geq 1) \neq \mathbb{P}(X \geq 1)\mathbb{P}(Y \geq 1)$$

car $\mathbb{P}(X \geq 1) > 0$ et $\mathbb{P}(Y \geq 1) < 1$.

**Proposition 31.** *Si $X_1, ..., X_n$ sont des variables aléatoires 2 à 2 décorrélées et de carrés intégrables, et si $S_n = X_1 + ... + X_n$ alors*

$$\mathbf{Var}(S_n) = \mathbf{Var}(X_1) + ... + \mathbf{Var}(X_n).$$

**Preuve**

i) Cas $\forall j, \mathbf{E}[X_j] = m_j = 0$ et $X_j \in L^2$ :
On a $\mathbf{E}[S_n] = 0$ et

$$\begin{aligned}\mathbf{Var}(S_n) &= \mathbf{E}[S_n^2] \\ &= \mathbf{E}\left[\sum_{j=1}^n X_j^2 + \sum_{j=1}^n \sum_{i\neq j} X_j X_i\right] \\ &= \sum_{j=1}^n \mathbf{E}[X_j^2] + \sum_{j=1}^n \sum_{i\neq j} \mathbf{E}[X_j]\mathbf{E}[X_i] \text{ par décorrélation} \\ &= \sum_{j=1}^n \mathbf{E}[X_j^2] \\ &= \sum_{j=1}^n \mathbf{Var}(X_j).\end{aligned}$$

ii) Cas $X_j \in L^2$ :
On a $\mathbf{E}[S_n] = \sum\limits_{j=1}^n m_j$ et

$$\begin{aligned}\mathbf{Var}(S_n) &= \mathbf{Var}(S_n - \mathbf{E}[S_n]) \\ &= \mathbf{Var}(\sum_{j=1}^n (X_j - m_j)) \\ &= \sum_{j=1}^n \mathbf{Var}(X_j - m_j) \\ &= \sum_{j=1}^n \mathbf{Var}(X_j).\end{aligned}$$

iii) Cas $\exists j : \mathbf{E}[X_j^2] = +\infty$ :
$\sum\limits_{j=1}^n \mathbf{Var}(X_j) = +\infty$ et $\mathbf{Var}(S_n) = +\infty$ car $\left|\sum\limits_{j=1}^n \sum_{i\neq j} \mathbf{E}[X_j]\mathbf{E}[X_i]\right| < +\infty$.

## 2 .3 Somme de v.a. indépendantes a densité

**Proposition 32.** *Si $U$ et $V$ sont deux variables aléatoires indépendantes de densité respectives $f_U$ et $f_V$, alors la densité $f_Z$ de la variable aléatoire somme $Z = U + V$ est le produit de convolution de $f_U$ et $f_V$*

$$f_Z(z) = \int_{-\infty}^{+\infty} f_U(u)f_V(z-u)du = \int_{-\infty}^{+\infty} f_U(z-v)f_V(v)dv$$

# Chapitre IV

# Notions fortes de convergences et lois des grands nombres

### Plan du chapitre



## 1 Convergence en probabilité et loi faible des grands nombres

**Définition 28.** *Soit $X$, $X_n$, $n \geq 1$ des variables aléatoires réelles.*
*On dit que $(X_n)_{n\geq 1}$* **converge en probabilité** *vers $X$ si*

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

*On note $X_n \xrightarrow{\mathbb{P}} X$.*

**Loi faible des grands nombres** Supposons $X_i$ suite de v.a. i.i.d telles que $L^2$ : $\mathbf{E}[X_1^2] < +\infty$. On note

$$\mu = \mathbf{E}[X_1]$$
$$\sigma^2 = \mathbf{Var}(X_1)$$

l'espérance et la variance commune des variables $X_i$ qui ont même loi. Par linéarité de l'espérance, on a

$$\mathbf{E}[M_n] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[X_i] = \mu.$$

Comme les $X_i$ sont indépendantes, $\mathbf{Var}(\sum X_i) = \sum \mathbf{Var}(X_i)$ et donc

$$\mathbf{Var}(M_n) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{Var}(X_i) = \frac{\sigma^2}{n} \longrightarrow 0.$$

La variance de $M_n$ tend vers, 0, i.e., ne varie plus à l'infini.

**Théorème 33.** *Loi faible des grands nombres.*

$$\forall \varepsilon > 0, \quad \mathbb{P}(|M_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

*On a donc, par définition de la convergence en probabilités,*

$$M_n \xrightarrow{\mathbb{P}} \mu.$$

**Preuve** Par inégalité de Tchebychev,

$$\mathbb{P}(|M_n - \mu| > \varepsilon) \leq \frac{\mathbf{Var}(M_n)}{\varepsilon^2}.$$

# 2 Convergence en moyenne d'ordre p

**Définition 29.** *Soit $X$, $X_n$, $n \geq 1$ v.a. ayant des moments d'ordre $p > 0$. On dit que $(X_n)_{n\geq 1}$ converge vers $X$* **en moyenne d'ordre** *$p$ si*

$$\lim_{n\to+\infty}\mathbf{E}[|X_n - X|^p] = 0$$

*i.e.*

$$X_n \overset{L^p(\Omega)}{\to} X.$$

# 3 Convergence presque sûre

**Définition 30.** *On dit que la suite $(X_n)_{n\geq 1}$ de variables aléatoires converge* **presque sûrement** *vers $X$ si*

$$\mathbb{P}(\omega : \lim_{n\to+\infty} X_n(\omega) = X(\omega)) = 1.$$

*On note $X_n \overset{p.s.}{\to} X$.*

**Proposition 34.** *Soit $X$, $X_n$ des v.a.r.*

$$X_n \overset{p.s.}{\to} X \Leftrightarrow \forall \varepsilon > 0, \quad \mathbb{P}(\limsup_{n\to+\infty}\{|X_n - X| > \varepsilon\}) = 0.$$

**Preuve**

$$\begin{aligned}
\lim_{n\to+\infty} X_n(\omega) \neq X(\omega) &\Leftrightarrow \exists \varepsilon > 0, \forall n_0, \exists n \geq n_0, |X_n(\omega) - X(\omega)| > \varepsilon \\
&\Leftrightarrow \exists k \in \mathbb{N}^*, \forall n_0, \exists n \geq n_0, |X_n(\omega) - X(\omega)| > \frac{1}{k} \\
&\Leftrightarrow \omega \in \bigcup_{k\in\mathbb{N}^*} \bigcap_{n_0\in\mathbb{N}^*} \bigcup_{n\geq n_0} \{|X_n - X| > \frac{1}{k}\} \\
&\Leftrightarrow \omega \in \bigcup_{k\in\mathbb{N}^*} \limsup_{n\to+\infty} \{|X_n - X| > \frac{1}{k}\} =: N
\end{aligned}$$

• Supposons $X_n \overset{p.s.}{\to} X$ : $\mathbb{P}(N) = 0$ par définition.
$\forall \varepsilon > 0$, $\limsup_{n\to+\infty}\{|X_n - X| > \varepsilon\} \subset N$ donc $\mathbb{P}(\limsup_{n\to+\infty}\{|X_n - X| > \varepsilon\}) = 0$.
• Supposons $\forall \varepsilon > 0, \quad \mathbb{P}(\limsup_{n\to+\infty}\{|X_n - X| > \varepsilon\}) = 0$.
$\mathbb{P}(N) \leq \sum_{k\in\mathbb{N}^*} \mathbb{P}(\limsup_{n\to+\infty}\{|X_n - X| > \frac{1}{k}\}) = 0$ ■

**Exemple.** *Soit $(X_n)_{n\geq 1}$ v.a.i.i.d. tq $\mathbf{E}[|X_1|] < +\infty$. Alors*

$$\lim_{n\to+\infty} \frac{X_n}{n} = 0 \ p.s.$$

En effet, soit $\varepsilon > 0$,

$$\mathbb{P}(|\frac{X_n}{n}| > \varepsilon) = \mathbb{P}(\frac{|X_1|}{\varepsilon} > n).$$

Comme $\mathbf{E}[\frac{|X_1|}{\varepsilon}] < +\infty$, $\sum_n \mathbb{P}(\frac{|X_1|}{\varepsilon} > n) < +\infty$ donc par Borel-Cantelli,

$$\mathbb{P}(\limsup_{n\to+\infty}\{|\frac{X_n}{n}| > \varepsilon\}) = 0.$$

**Critère de convergence presque sûre**

**Proposition 35.** *Si $\forall \varepsilon > 0$, $\sum_n \mathbb{P}(|X_n - X| > \varepsilon) < +\infty$, alors $\lim_{n\to+\infty} X_n = X$ p.s.*

**Preuve** Borel-Cantelli.

# 4 Liens entre les convergences

**Proposition 36.** .

1. $X_n \overset{p.s.}{\to} X \quad \Rightarrow \quad X_n \overset{\mathbb{P}}{\to} X.$
2. $X_n \overset{L^p}{\to} X \quad \Rightarrow \quad X_n \overset{\mathbb{P}}{\to} X.$
3. *Les réciproques sont fausses (en général).*

**Exemple de suite convergeant p.s. mais pas en moyenne**

$$(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1[, \mathcal{B}_{[0,1[}, \lambda_{[0,1[})$$

On pose $X_n(\omega) = n\mathbf{1}_{[0,\frac{1}{n}]}(\omega)$. (Tracer $X_n$ en fonction de $\omega$ pour mieux visualiser)
$\forall \omega \in \Omega \backslash \{0\}$, $\lim\limits_{n\to+\infty} X_n(\omega) = 0$ donc $\lim\limits_{n\to+\infty} \frac{X_n}{n} = 0$ p.s. Mais $\mathbf{E}[|X_n - 0|] = 1$.

**Exemple de suite convergeant en moyenne mais pas p.s.**

$$(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1[, \mathcal{B}_{[0,1[}, \lambda_{[0,1[})$$

Pour $n \geq 1$, on pose $l = [\frac{\log n}{\log 2}]$ et $k = n - 2^l$ pour avoir $n = 2^l + k$, $k \in \{0, ..., 2^l - 1\}$. On pose

$$X_n = \mathbf{1}_{[\frac{k}{2^l}, \frac{k+1}{2^l}[}.$$

(Tracer $X_1$ jusqu'à $X_5$ en fonction de $\omega$ pour mieux visualiser)
$\lim\limits_{n\to+\infty} \mathbf{E}[|X_n|] = \lim\limits_{n\to+\infty} \frac{1}{2^l} = 0$. Mais $\forall \omega \in \Omega, \limsup\limits_{n\infty} X_n(\omega) = 1$ et $\liminf\limits_{n\infty} X_n(\omega) = 0$.

**Preuve**

1. Supposons $X_n \overset{p.s.}{\to} X$ ($\Leftrightarrow \quad \mathbb{P}(\limsup\limits_{n\infty}\{|X_n - X| > \varepsilon\}) = 0$)
   Soit $\varepsilon > 0$.
   $$\begin{aligned}\limsup_{n\infty} \mathbb{P}(|X_n - X| > \varepsilon) &\leq \mathbb{P}(\limsup_{n\infty}\{|X_n - X| > \varepsilon\})(\text{voir Exercice 1 Feuille 1})\\ &\leq 0\end{aligned}$$
   donc $\lim\limits_{n\to+\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$.
2. Supposons $X_n \overset{L^p}{\to} X$ (i.e. $\lim\limits_{n\to+\infty} \mathbf{E}[|X_n - X|^p] = 0$)
   Soit $\varepsilon > 0$.
   $$\mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{\mathbf{E}[|X_n - X|^p]}{\varepsilon^p}(\text{par inégalité de Markov})$$
   donc $\lim\limits_{n\to+\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$.

3. On reprend les deux exemples précédents.

**Bilan des convergences**

$$\begin{array}{ccc} \text{p.s.} & & L^p \\ & \searrow \quad \swarrow & \\ & \text{proba} & \end{array}$$

**Proposition 37** (Propriétés de la convergence en probabilité)**.** .

1. *Si* $X_n \xrightarrow{\mathbb{P}} X_1$ *et* $X_n \xrightarrow{\mathbb{P}} X_2$ *alors* $X_1 = X_2$ *p.s.*
2. *Si* $X_n \xrightarrow{\mathbb{P}} X$*, il existe une sous-suite* $(X_{n_k})_k$ *telle que* $\lim\limits_{k\to+\infty} X_{n_k} = X$ *p.s.*

**Preuve**

1. Soit $\varepsilon > 0$.
$$\mathbb{P}(|X_1 - X_2| > \varepsilon) \leq \mathbb{P}(|X_n + X_1| > \frac{\varepsilon}{2}) + \mathbb{P}(|X_n - X_2| > \frac{\varepsilon}{2})$$
donc $\mathbb{P}(|X_1 - X_2| > \varepsilon) = 0$, $\forall \varepsilon > 0$.
2. a) On montre d'abord par récurrence sur $k$ :
$$\exists (n_k)_k \text{ croissante tq } \mathbb{P}(|X_{n_k} - X| \geq \frac{1}{2^k}) \leq \frac{1}{2^k}.$$
Pour $k = 0$, pas de soucis puis $\lim\limits_{n\to+\infty} \mathbb{P}(|X_n - X| \geq \frac{1}{2^{k+1}}) = 0$ : $\exists n_{k+1} > n_k$ tq $\mathbb{P}(|X_{n_{k+1}} - X| \geq \frac{1}{2^{k+1}}) \leq \frac{1}{2^{k+1}}$.
b) On montre $\lim\limits_{k\to+\infty} X_{n_k} \overset{p.s.}{=} X$.
Soit $\varepsilon > 0$. $\exists k_0$ tq $\forall k \geq k_0$, $\frac{1}{2^k} \leq \varepsilon$ :
$$\sum_{k\geq k_0} \mathbb{P}(|X_{n_k} - X| > \varepsilon) \leq \sum_{k \geq k_0} \mathbb{P}(|X_{n_k} - X| > \frac{1}{2^k}) < +\infty$$
donc on peut utiliser le critère de convergence p.s.

**Définition 31.** *Pour* $X$*,* $X_n$*,* $n \geq 1$ *vecteurs de* $\mathbb{R}^d$*, on définit les convergences de* $X_n$ *vers* $X$ *en remplaçant* $|.|$ *par* $\|.\|_d$ *norme de* $\mathbb{R}^d$*.*

**Proposition 38.**

$$X_n = (X_n^{(1)}, ..., X_n^{(d)}) \to X = (X^{(1)}, ..., X^{(d)}) \Leftrightarrow \forall i = 1, ..., d, X_n^{(i)} \to X^{(i)}.$$

**Preuve** : Exercice.

**Proposition 39** (Continuous mapping theorem)**.** *Soit $(X_n)_{n\geq 1}$ v.a. de $\mathbb{R}^d$ convergeant en probabilité vers $X \in \mathbb{R}^p$. Si $g$ est une fonction continue, $g : \mathbb{R}^d \to \mathbb{R}^p$, alors*

$$g(X_n) \xrightarrow{\mathbb{P}} g(X).$$

**Preuve** Supposons $X_n \xrightarrow{\mathbb{P}} X$ et $g(X_n)$ ne converge pas vers $g(X)$ en probabilité :

$$\exists \varepsilon > 0, \quad \exists \eta > 0, \quad \exists (n_k)_k \quad \text{tq} \quad \mathbb{P}(\|g(X_{n_k}) - g(X)\| > \varepsilon) \geq \eta, \quad \forall k.$$

Comme $X_{n_k} \xrightarrow{\mathbb{P}} X$, on peut extraire $(n_{k_l})_l$ tq $\lim\limits_{l\to+\infty} X_{n_{k_l}} = X$ p.s. On a alors $\lim\limits_{l\to+\infty} g(X_{n_{k_l}}) = g(X)$ p.s. donc $\lim\limits_{l\to+\infty} g(X_{n_{k_l}}) = g(X)$ en probabilité d'où $\lim\limits_{l\to+\infty} \mathbb{P}(\|g(X_{n_{k_l}}) - g(X)\| > \varepsilon) = 0$ : absurde.

**Corollaire 40.** *Si $X_n \xrightarrow{\mathbb{P}} X$ et $Y_n \xrightarrow{\mathbb{P}} Y$ alors $X_n + Y_n \xrightarrow{\mathbb{P}} X + Y$ et $X_n Y_n \xrightarrow{\mathbb{P}} XY$ (pour $d = 1$).*

En effet, $g_1(x, y) = x + y$ et $g_2(x, y) = xy$ sont continues.

# 5 Loi forte des grands nombres

## 5 .1 Cas $L^2$ : $\mathbf{E}[X_1^2] < +\infty$

**Théorème 41.** *Soit $(X_n)_{n\geq 1}$ une suite de v.a. i.i.d. telle que $\mathbf{E}[X_1^2] < +\infty$. On a*

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{j=1}^{n} X_j = \mathbf{E}[X_1] \quad \textit{presque sûrement.}$$

**Preuve**

a) On pose $X_n = X_n^+ - X_n^-$ avec $X_n^+ = \max\{X_n, 0\} \geq 0$ et $X_n^- = \max\{-X_n, 0\} \geq 0$. On a $M_n = M_n^+ - M_n^-$ avec

$$M_n^+ = \frac{X_1^+ + ... + X_n^+}{n}$$
$$M_n^- = \frac{X_1^- + ... + X_n^-}{n}$$

$\mathbf{E}[X_1] = \mathbf{E}[X_1^+] - \mathbf{E}[X_1^-]$ donc il suffit de considérer des variables positives.

b) Montrons que
$$M_{n^2} \underset{n\ to\infty}{\overset{p.s.}{\rightarrow}} \mu$$
En appliquant l'inegalite de Bienayme-Tchebyshev a $M_n$ avec $p = 2$ et $a = 1/q, q \in \mathbb{N}^*$, on a
$$\forall q \in \mathbb{N}^*, \quad \mathbb{P}(|M_n - \mu| > \frac{1}{q}) \leq \frac{\sigma^2 q^2}{n}$$
On extrait la sous suite de terme $(M_{n^2})_n$ et on a
$$\mathbb{P}(|M_{n^2} - \mu| > \frac{1}{q}) \leq \frac{\sigma^2 q^2}{n^2}$$

On définit les évènements
$$A_{n,q} = \{\omega : |M_{n^2}(\omega) - \mu| > \frac{1}{q}\}$$
$$C_q = \limsup_{n\to\infty} A_{n,q}$$

Par le lemme de Borel-Cantelli comme la serie de terme general $\mathbb{P}(A_{n,q})$ est convegente, la probabilité de la lmite sup est nulle
$$\forall q, \mathbb{P}(C_q) = 0$$
Soit $\omega$ tq
$$M_{n^2}(\omega) \underset{n\to\infty}{\nrightarrow} \mu$$
Alors il existe $\epsilon_\omega > 0$ et une infinité d'indices i tels que
$$|M_{i^2}(\omega) - \mu| > \epsilon_\omega$$
Soit $q_\omega$ tel que $\epsilon_\omega > 1/q_\omega$. $\omega$ est dans une infinité de $A_{i,q_\omega}$ donc dans la limsup (par propriété de la limsup) :
$$\omega \in \limsup_{n\to\infty} A_{n,q_\omega} = C_{q_\omega}$$
Donc
$$\omega \in \cup_{q\geq 1} C_q$$
Or
$$\mathbb{P}(\cup_{q\geq 1} C_q) \leq \sum_{q\geq 1} \mathbb{P}(C_q) = 0$$
Donc
$$\mathbb{P}\big(\{\omega : M_{n^2}(\omega) \underset{n\to\infty}{\nrightarrow} \mu\}\big) = 0$$
C'est a dire que $M_{n^2}$ converge presque sûrement vers $\mu$.

c) On a montré que

$$M_{n^2} \underset{n\ to\infty}{\overset{p.s.}{\to}} \mu$$

Montrons que

$$M_n \underset{n\ to\infty}{\overset{p.s.}{\to}} \mu$$

On pose $k_n = [\sqrt{n}]$ (Partie entière). On a

$$k_n \leq \sqrt{n} < k_n + 1$$

d'où

$$k_n^2 \leq n < (k_n+1)^2$$
$$n - 2k_n - 1 < k_n^2 \leq n$$

et donc par le theoreme des gendarmes

$$\lim_{n\to+\infty} \frac{k_n^2}{n} = 1.$$

Comme on a supposé que $X_n \geq 0$, on a

$$\frac{k_n^2}{n} M_{k_n^2} \leq M_n \leq M_{(k_n+1)^2} \frac{(k_n+1)^2}{n}$$

Soit

$$E = \{\omega : M_{k_n^2} \nrightarrow \mu\} \cup \{\omega : M_{(k_n+1)^2} \nrightarrow \mu\}$$

On a $P(E) \leq 0 + 0 = 0$. Par théorème des gendarmes, $\forall \omega \notin E$, c'est a dire avec probabilité 1, $M_n(\omega) \to \mu$. Ainsi

$$M_n \underset{n\ to\infty}{\overset{p.s.}{\to}} \mu$$

## 5 .2 Cas $L^1$ : $\mathbf{E}[|X_1|] < +\infty$

**Théorème 42** (Loi Forte des Grands Nombres)**.** .
*Soit $(X_n)_{n\geq 1}$ une suite de v.a. i.i.d. telle que $\mathbf{E}[|X_1|] < +\infty$. On a*

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{j=1}^{n} X_j = \mathbf{E}[X_1] \qquad \textit{presque sûrement.}$$

**Preuve** Pas de preuve ici.

# Chapitre V

# Fonctions caractéristiques, convergence en loi et théorème de la limite centrale

## Plan du chapitre



## 1 Fonction caractéristique

### 1 .1 Définition et propriétés élémentaires

Nous introduisons ici un outil important en calcul des probabilités : **la fonction caracteristique** d'une variable aléatoire. On retrouve cet outil dans d'autres branches des mathématiques ou on la nomme **transformée de Fourier**.

**Définition 32.** *Soit $X$ une v.a. réelle. On définit la* **fonction caractéristique** *de $X$, notée $\varphi_X$, par*

$$\begin{array}{rcl} \varphi_X : \mathbb{R} & \to & \mathbb{C} \\ \theta & \mapsto & \mathbf{E}\left[e^{i\theta X}\right] = \mathbf{E}\left[cos(\theta X)\right] + i\mathbf{E}\left[sin(\theta X)\right] \end{array}$$

*Remarques.* • Si $X$ est à densité $f_X$,

$$\varphi_X(\theta) = \int_{\mathbb{R}} e^{i\theta x} f_X(x)dx.$$

C'est la transformée de Fourier de $f_X$ (à $2\pi$ près).

• On peut étendre cette défintion a un vecteur aléatoire $X = (X_1, \ldots, X_d)$ à valeur dans $\mathbb{R}^d$. Dans ce cas

$$\begin{array}{rcl} \varphi_X : \mathbb{R}^d & \to & \mathbb{C} \\ \theta & \mapsto & \mathbf{E}\left[e^{i\langle\theta,X\rangle}\right] = E\left[e^{i\sum_{j=1}^d \theta_j X_j \langle\theta,X\rangle}\right] \end{array}$$

**Proposition 43.** *On retrouve pour la fonction caractéristique les propriétés de la transformée de Fourier.*

- $\varphi_X$ *est continue (et même uniformément continue)*
- $X$ *est une* ***variable aléatoire réelle***
  - $*$ *Si* $\mathbf{E}[|X|] < +\infty$ *, alors* $\varphi_X$ *est de classe* $\mathcal{C}^1$ *sur* $\mathbb{R}$ *et*

    $$\varphi'_X(\theta) = i\mathbf{E}\left[Xe^{i\theta X}\right]$$

    *en particulier*

    $$\varphi'_X(0) = i\mathbf{E}\left[X\right]$$

  - $*$ *Si* $\mathbf{E}[|X|^2] < +\infty$ *alors* $\varphi_X$ *est de classe* $\mathcal{C}^2$ *sur* $\mathbb{R}$ *et*

    $$\varphi''_X(\theta) = -\mathbf{E}\left[X^2e^{i\theta X}\right]$$

    *en particulier*

    $$\varphi''_X(0) = -\mathbf{E}\left[X^2\right]$$

  - $*$ *On a, avec la formule de Taylor-Young, le développement limité en 0 pour* $\varphi_X$

    $$\varphi_X(\theta) = 1 + i\mathbf{E}[X]\theta - \frac{1}{2}\mathbf{E}[X^2]\theta^2 + \theta^2\epsilon(\theta). \tag{V.1}$$

  - $*$ *Si* $\mathbf{E}[|X|^k] < +\infty,\ k \in \mathbb{N}$*, alors* $\varphi_X$ *est de classe* $\mathcal{C}^k$ *et*

    $$\varphi_X^{(k)}(\theta) = i^k\mathbf{E}[X^k e^{i\theta X}].$$

    *On a alors*

    $$\mathbf{E}[X^k] = \frac{\varphi_X^{(k)}(0)}{i^k}.$$

- $X$ *est un* ***vecteur aléatoire*** *de* $\mathbb{R}^d$*.*

$*$ *Si* $\mathbf{E}[\|X\|_1] < +\infty$ *alors* $\varphi_X$ *est de classe* $\mathcal{C}^1$ *sur* $\mathbb{R}^d$ *et*

$$\frac{\partial \varphi_X}{\partial \theta_j}(\theta) = \mathbf{E}\left[\frac{\partial}{\partial \theta_j} e^{i\sum\limits_{l=1}^{d} \theta_l X_l}\right] = i\mathbf{E}\left[X_j e^{i<\theta,X>}\right]$$

$*$ *Si* $\mathbf{E}[\|X\|_2^2] < +\infty$ *alors* $\varphi_X$ *est de classe* $\mathcal{C}^2$ *sur* $\mathbb{R}^d$ *et*

$$\frac{\partial^2 \varphi_X}{\partial \theta_j \partial \theta_k}(\theta) = \mathbf{E}\left[\frac{\partial^2}{\partial \theta_j \partial \theta_k} e^{i\sum\limits_{l=1}^{d} \theta_l X_l}\right] = -\mathbf{E}\left[X_j X_k e^{i<\theta,X>}\right]$$

$*$ *On a, avec la formule de Taylor-Young, le développement limité en 0*

$$\varphi_X(\theta) = 1 + i\sum_{j=1}^{d} \mathbf{E}[X_j]\theta_j - \frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d} \mathbf{E}[X_j X_k]\theta_j\theta_k + \|\theta\|^2\varepsilon(\theta). \tag{V.2}$$

**Preuve** Il suffit de montrer $\frac{d}{d\theta}\mathbf{E}[e^{i\theta X}] = \mathbf{E}[\frac{d}{d\theta}e^{i\theta X}]$ et d'itérer. On a

$$\begin{aligned}\frac{\varphi_X(\theta+\varepsilon) - \varphi_X(\theta)}{\varepsilon} &= \mathbf{E}\left[\frac{e^{i(\theta+\varepsilon)X} - e^{i\theta X}}{\varepsilon}\right] \\ &= \int_\Omega \frac{e^{i(\theta+\varepsilon)X(\omega)} - e^{i\theta X(\omega)}}{\varepsilon} d\mathbb{P}(\omega)\end{aligned}$$

On a ensuite

- $\lim\limits_{\varepsilon\to 0} \frac{e^{i(\theta+\varepsilon)X(\omega)} - e^{i\theta X(\omega)}}{\varepsilon} = iX(\omega)e^{i\theta X(\omega)}$ $\mathbb{P}$ p.p.
- $$\begin{aligned}\left|\frac{e^{i(\theta+\varepsilon)X(\omega)} - e^{i\theta X(\omega)}}{\varepsilon}\right| &\leq \frac{|e^{i\varepsilon X(\omega)} - 1|}{\varepsilon} \\ &\leq \frac{|\cos(\varepsilon X(\omega)) - 1|}{\varepsilon} + \frac{|\sin(\varepsilon X(\omega))|}{\varepsilon} \\ &\leq 2|X(\omega)| \quad \mathbb{P} \text{ p.p.}\end{aligned}$$

  donc on conclut par convergence dominée.

**Exemples.**

*1)* $X \sim \mathcal{N}(0,1)$
*Comme* $E[|X|] < +\infty$,

$$\forall t \in \mathbb{R}, \quad \varphi_X'(t) = i\mathbf{E}[Xe^{itX}] = \frac{i}{\sqrt{2\pi}} \int_{\mathbb{R}} x e^{itx} e^{-\frac{x^2}{2}} dx$$

*puis par Intégration Par Parties*

$$\varphi'_X(t) = \frac{i}{\sqrt{2\pi}} \left( \left[ -e^{-\frac{x^2}{2}} e^{itx} \right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} ite^{itx} e^{-\frac{x^2}{2}} dx \right) = -t\varphi_X(t)$$

*donc* $\varphi'_X(t) + t\varphi_X(t) = 0$, $\forall t \in \mathbb{R}$ *et* $\varphi_X(0) = 1$ *: finalement* $(\varphi_X(t)e^{\frac{t^2}{2}})' = 0$ *donc* $\varphi_X(t) = Ke^{-\frac{t^2}{2}} = e^{-\frac{t^2}{2}}$ *:*

$$\forall t \in \mathbb{R}, \quad \varphi_X(t) = e^{-\frac{t^2}{2}}.$$

*2)* $X \sim \mathcal{N}(\mu, \sigma^2)$
*On a* $Y = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$ *donc*

$$\forall t \in \mathbb{R}, \quad \varphi_X(t) = \mathbf{E}[e^{itX}] = \mathbf{E}[e^{it(\sigma Y + \mu)}] = e^{it\mu}\varphi_Y(\sigma t) = e^{it\mu}e^{-\frac{\sigma^2 t^2}{2}}.$$

**Théorème 44** (Somme et indépendance)**.** *Soient* $X$ *et* $Y$ *des variables aléatoires réelles indépendantes. Alors la fonction caractéristique* $\varphi_{X+Y}$ *de* $X+Y$ *est le produit des fonctions caractéristiques de* $X$ *et* $Y$ *:*

$$\varphi_{X+Y}(\theta) = \varphi_X(\theta) \times \varphi_Y(\theta).$$

**Preuve** $\forall \theta \in \mathbb{R}$, on a :

$$\begin{aligned} \varphi_{X+Y}(\theta) &= \mathbf{E}[e^{i(X+Y)}] \text{ par définition} \\ &= \mathbf{E}[e^{iX}e^{iY}] \\ &= \mathbf{E}[e^{iX}]\mathbf{E}[e^{iY}] \text{ par indépendance} \\ &= \varphi_X(\theta)\varphi_Y(\theta) \end{aligned}$$

*Remarque.* Ce théorème s'étend a une combinaison affine de $n$ vecteurs aléatoires de $R^d$,a vec une preuve similaire. :
Soient $X_1, ..., X_n$ des v.a. indépendantes à valeurs dans $\mathbb{R}^d$, $a_1, ..., a_n$ des réels et $b \in \mathbb{R}^d$. On pose $Y = a_1X_1 + ... + a_nX_n + b$. On a $\forall \theta \in \mathbb{R}^d$,

$$\varphi_Y(\theta) = \varphi_{X_1}(a_1\theta) \times ... \times \varphi_{X_n}(a_n\theta)e^{i<\theta,b>}.$$

**Exemple.** *Pour* $X_i \sim \mathcal{N}(0, \sigma^2)$ *indépendantes, on pose* $Y = (X_1, ..., X_d)$*. On a*

$$d\mathbb{P}_Y(x) = \frac{1}{(\sqrt{2\pi\sigma^2})^d} e^{-\sum_{j=1}^{d} \frac{x_j^2}{2\sigma^2}} dx_1...dx_d = (2\pi\sigma^2)^{-\frac{d}{2}} e^{-\frac{\|x\|^2}{2\sigma^2}} \lambda(dx)$$

*et* $\varphi_Y(\theta) = \mathbf{E}[e^{i<\theta,Y>}] = \prod_{j=1}^{d} \varphi_{X_j}(\theta_j) = \prod_{j=1}^{d} e^{-\frac{\sigma^2\theta^2}{2}}$ *d'où*

$$\varphi_Y(\theta) = e^{-\frac{\sigma^2\|\theta\|^2}{2}}.$$

**Théorème 45** (d'injectivité)**.** *Soient $X$ et $Y$ deux variables aléatoires de $\mathbb{R}^d$. On a*

$$\mathbb{P}_X = \mathbb{P}_Y \Leftrightarrow \varphi_X = \varphi_Y.$$

**Preuve**

- $\Rightarrow$ Ok
- $\Leftarrow$ On suppose que $\varphi_X = \varphi_Y$.

a) $\forall \sigma > 0, \forall \mu \in \mathbb{R}^d$, on a la formule

$$\int\limits_{\mathbb{R}^d} e^{-\frac{\sigma^2}{2}\|x-\mu\|^2} d\mathbb{P}_X(x) = (2\pi\sigma^2)^{-\frac{d}{2}} \int\limits_{\mathbb{R}^d} e^{-\frac{\|u\|^2}{2\sigma^2}} e^{-i<\mu,u>} \varphi_X(u) du.$$

En effet,

$$\begin{aligned} I_1 &:= \int\limits_{\mathbb{R}^d} e^{-\frac{\|u\|^2}{2\sigma^2}} e^{-i<\mu,u>} \varphi_X(u) du \\ &= \int\limits_{\mathbb{R}^d} e^{-\frac{\|u\|^2}{2\sigma^2}} e^{-i<\mu,u>} \int_{\mathbb{R}^d} e^{i<u,x>} d\mathbb{P}_X(x) du \\ &= \int\limits_{\mathbb{R}^d} (\int_{\mathbb{R}^d} e^{-\frac{\|u\|^2}{2\sigma^2}} e^{i<u,x-\mu>} du) d\mathbb{P}_X(x) \\ &= \int_{\mathbb{R}^d} (2\pi\sigma^2)^{d/2} e^{-\frac{\sigma^2}{2}\|x-\mu\|^2} d\mathbb{P}_X(x) \end{aligned}$$

d'après l'exemple précédent.

b) Comme $\varphi_X = \varphi_Y$, on a avec a), $\forall \sigma > 0, \forall \mu \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} e^{-\frac{\sigma^2}{2}\|x-\mu\|^2} d\mathbb{P}_X(x) = \int_{\mathbb{R}^d} e^{-\frac{\sigma^2}{2}\|x-\mu\|^2} d\mathbb{P}_Y(x).$$

Soit $\varphi : \mathbb{R}^d \to \mathbb{R}$ continue à support compact. On a

$$I_2 := \int_{\mathbb{R}^d} \varphi(\mu) \frac{\sigma^d}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-\frac{\sigma^2}{2}\|x-\mu\|^2} d\mathbb{P}_X(x) d\lambda(\mu) = \int_{\mathbb{R}^d} \varphi(\mu) \frac{\sigma^d}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-\frac{\sigma^2}{2}\|x-\mu\|^2} d\mathbb{P}_Y(x) d\lambda(\mu).$$

Or

$$I_2 = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(x - \frac{z}{\sigma}) \frac{1}{(2\pi)^{d/2}} e^{-\frac{\|z\|^2}{2}} d\mathbb{P}_X(x) d\lambda(\mu)$$

donc $\lim\limits_{\sigma \to +\infty} I_2 = \int\limits_{\mathbb{R}^d} \varphi(x) d\mathbb{P}_X(x)$. Finalement, pour toute fonction $\varphi$ continue à support compact,

$$\int\limits_{\mathbb{R}^d} \varphi(x) d\mathbb{P}_X(x) = \int\limits_{\mathbb{R}^d} \varphi(x) d\mathbb{P}_Y(x).$$

c) Soit $(\varphi_j)_j$ une suite de fonctions continues à support compact qui converge vers $\mathbf{1}_{[a,b]}$ (pour $[a,b] = \prod_{l=1}^{d} [a_l, b_l]$ pavé de $\mathbb{R}^d$) avec $0 \leq \varphi_j \leq 1$.
On a

$$\int\limits_{\mathbb{R}^d} \varphi_j(x) d\mathbb{P}_X(x) = \int\limits_{\mathbb{R}^d} \varphi_j(x) d\mathbb{P}_Y(x) \to \int\limits_{[a,b]} d\mathbb{P}_X(x) = \int\limits_{[a,b]} d\mathbb{P}_Y(x).$$

Finalement, $\mathbb{P}_X([a,b]) = \mathbb{P}_Y([a,b])$ donc $\mathbb{P}_X$ et $\mathbb{P}_Y$ coïncident sur les pavés de $\mathbb{R}^d$ donc aussi sur un $\pi$-système qui engendre $\mathcal{B}(\mathbb{R}^d)$ donc $\mathbb{P}_X = \mathbb{P}_Y$ $\square$

**Exemple d'application : Loi de Cauchy** $C(\alpha), \alpha > 0$ .
Si $X$ a pour densité $f_X(x) = \frac{\alpha}{\pi(\alpha^2+x^2)}, x \in \mathbb{R}$ alors $X \sim C(\alpha)$.

*Remarque.* $E[|X|] = \int_{\mathbb{R}} \frac{\alpha|x|}{\pi(\alpha^2+x^2)} = +\infty$ : il n'y a pas d'espérance ni de variance.

Pour $X_1 \sim C(\alpha_1)$ et $X_2 \sim C(\alpha_2)$ indépendantes, on veut la loi de $X_1 + X_2$.

**Calcul de $\varphi_X$ :** Soit $f(t) = e^{-\alpha|t|} \in L^1(\mathbb{R})$.

$$\begin{aligned}
\hat{f}(x) &= \int_{\mathbb{R}} e^{-itx} e^{-\alpha|t|} dt = \int_{-\infty}^{0} e^{-itx} e^{\alpha t} dt + \int_{0}^{+\infty} e^{-itx} e^{-\alpha t} dt \\
&= \int_{-\infty}^{0} e^{t(\alpha - ix)} dt + \int_{0}^{+\infty} e^{-t(ix+\alpha)} dt \\
&= \frac{1}{\alpha - ix} + \frac{1}{\alpha + ix} = \frac{2\alpha}{\alpha^2 + x^2} \in L^1(\mathbb{R})
\end{aligned}$$

Par formule d'inversion de Fourier, $\lambda$ presque partout,

$$\frac{1}{2\pi} \int_{\mathbb{R}} \frac{2\alpha}{\alpha^2 + x^2} e^{ixt} dx = f(t) = e^{-\alpha|t|}$$

donc $\varphi_X(t) = e^{-\alpha|t|}$ $\lambda$ presque partout.
Par continuité,

$$\forall t \in \mathbb{R}, \quad \varphi_X(t) = e^{-\alpha|t|}.$$

**Loi de $X_1 + X_2$ :** On a $\forall t \in \mathbb{R}$, $\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t)\varphi_{X_2}(t)$ par indépendance donc

$$\varphi_{X_1+X_2}(t) = e^{-\alpha_1|t|} e^{-\alpha_2|t|} = e^{-(\alpha_1+\alpha_2)|t|}.$$

C'est la fonction caractéristique de la loi de Cauchy $C(\alpha_1 + \alpha_2)$. Par théorème d'injectivité,

$$X_1 + X_2 \sim C(\alpha_1 + \alpha_2).$$

## 1 .2 A propos de la transformée de Fourier

La transformée de Fourier apparaît souvent en mathématiques. Pourquoi ?

**Proposition 46.** *Soient $A, B \in \mathcal{M}_n(K)$ deux matrices diagonalisables. $A$ et $B$ sont codiagonalisables si, et seulement si, $A$ et $B$ commutent.*

**Preuve**

"$\Rightarrow$" On suppose qu'il existe $P$ inversible et $D_1, D_2$ diagonales telles que

$$A = P^{-1}D_1P, \quad B = P^{-1}D_2P$$

Puisque les matrices diagonales commutent

$$AB = (P^{-1}D_1P)(P^1D_2P) = P^{-1}D_1D_2P = P^{-1}D_2D_1P = (P^{-1}D_2P)(P^{-1}D_1P) = BA$$

"$\Leftarrow$" Soient $A, B$ deux matrices diagonalisables. Notons u et v les endomorphismes canoniquement associes a A et B. Notons $(_i)_{i=1...r}$ les valeurs propres associées à $u$ et $(E_i)_{i=1...r}$ les sous-espaces propres associes. Comme u est diagonalisable alors

$$K^n = \bigoplus_{i=1}^{r} E_i$$

$\forall e \in E_i$

$$u(v(e_i)) = v(u(e_i) = v(\lambda_i e_i) = \lambda_i v(e_i)$$

c.à.d $v(e_i) \in E_i$, $E_i$ est stable par $v$. Comme $v$ est diagonalisable, il l'est en particulier sur l'espace $E_i$. Soit $B_i$ une base de diagonalisation de $v$ sur $E_i$. $B_i$ est aussi une base de diagonalisation de $u$ sur $E_i$, car sur $E_i$, $u = \lambda_i I_{E_i}$. La base

$$B = \bigcup_{i=1}^{r} B_i$$

est donc une base de codiagonalisation de $u$ et $v$.

Fondamentalement, la transformée de Fourier c'est un changement de base dans l'espace $L^1$ qui est de dimension infinie. On représente une fonction de $L^1(\mathbb{R})$ dans sa base "canonique", c'est à dire en donant les valeurs $\{f(x), x \in \mathbb{R}\}$. La base de Fourier est l'ensemble

$$\{x \mapsto e^{i\theta x}, \theta \in \mathbb{R}$$

et la transformée de Fourier

$$\hat{f}(\theta) = \int_{\mathbb{R}} f(x)e^{i\theta x}dx$$

est simplement un produit scalaire qui nous donne le coefficient de $f$ sur le vecteur de base $x \mapsto e^{i\theta x}$. La formule d'inversion de Fourier est simplement l'expansion de $f$ sur la base de Fourier.

Pourquoi la base de Fourier est-elle intéressante ? Parce qu'elle diagonalise l'operation de translation

$$\tau : (f : x \mapsto f(x)) \mapsto (f_\tau : x \mapsto f(x - \tau)\}$$

. La base de Fourier diagonalise également les operateurs linéaires qui commutent avec la translation. Par exemple les operateurs différentiels usuels (dérivée, laplacien,...) sont des opérateurs linéaires covariant par translation :

$$(f_\tau)' = f'_\tau$$

La transformée de Fourier diagonalise donc ces opérateurs et facilite donc la résolution d'équations différentielles (en éléctricité, en électromagnétisme, en mécanique des fluides ...)

De façon générale, les opérateurs linéaire covariants par translation sont les opérateur de convolution. On admet que la masse de dirac en 0 est l'élément neutre pour la convolution, c'est à dire que

$$f(x) = (f * \delta_0)(x) = \int_{\mathbb{R}} f(x-u)\delta_0(u)du$$

Il faut pour cela développer mathématiquement la théorie des distributions. Soit $L$ un opérateur linéaire qui commute avec la translation, c'est à dire tel que

$$L(f_\tau) = L(f)_\tau$$

ou bien $\forall x \in \mathbb{R}$

$$L(f(x-\tau)) = L(f)(x-\tau)$$

On a

$$\begin{aligned}
L(f)(x) &= L(\int_{\mathbb{R}} f(u)\delta_0(x-u)du) \\
&= \int_{\mathbb{R}} f(u)L(\delta_0(x-u))du) \text{ par linéarité} \\
&= L\int_{\mathbb{R}} f(u)L(\delta_0)(x-u)du \text{ par commutation avec la translation} \\
&= (f * L(\delta))(x)
\end{aligned}$$

$L(f)$ est la convolution de $f$ avec $L(\delta)$. On comprend pourquoi la transformée de Fourier apparaît en traitement du signal, ou on effectue beaucoup de convolutions pour filtrer le son, les images, les signaux en général.

En théorie des probabilités, la convolution apparaît lorsque l'on somme deux v.a. $U, V$ indépendantes de densité $f_U, f_V$. La somme $U + V$ a pour densité la convlution $f_U * f_V$, qui est une simple multiplication en Fourier. Ce sera un outil indispensable pour caractériser la loi de

$$\frac{(X_1 + \ldots + X_n}{n}$$

pour $X_i$ v.a. i.i.d. de même densité $f$.

## 2 Convergence en loi

La notion de convergence de loi est plus faible que les notions de convergence presque sure, en moyenne et en probabilité vues précédemment. C'est la convergence des lois $\mathbb{P}_{X_n}$ et non celle des tirages $X_n(\omega)$.

*Remarque.* Si on définit, pour $\mu_n$ probabilité, $\mu_n \to \mu$ par

$$\forall A \in \mathcal{B}(\mathbb{R}), \quad \lim_{n\to+\infty} \mu_n(A) = \mu(A)$$

alors pour $\mu_n = \delta_{\frac{1}{n}}$ et $\mu = \delta_0$,

$$\mu_n(\{0\}) = \delta_{\frac{1}{n}}(\{0\}) = 0 \nrightarrow \mu(\{0\}) = \delta_0(\{0\}) = 1 : \text{ c'est pas top.}$$

On pose

- $\mathcal{C}_b(\mathbb{R}^d) = \{f$ continue bornée $\}$
- $\mathcal{C}_0(\mathbb{R}^d) = \{f$ continue qui tend vers 0 à l'infini $\}$
- $\mathcal{C}_c(\mathbb{R}^d) = \{f$ continue à support compact $\}$

**Définition 33** (Convergence étroite, faible et vague de mesures)**.** *Soit $\mu$, $\mu_n$, $n \geq 1$ des mesures de $\mathbb{R}^d$. On dit que $(\mu_n)_{n\geq 1}$ converge* **étroitement** *vers $\mu$ (resp. faiblement, vaguement) si*

$$\forall f \in \mathcal{C}_b(\mathbb{R}^d), \quad \lim_{n\to+\infty} \int f d\mu_n = \int f d\mu$$

*(resp. $\forall f \in \mathcal{C}_0(\mathbb{R}^d), \forall f \in \mathcal{C}_c(\mathbb{R}^d)$).*

**Proposition 47.** *Si $\mu_n$ et $\mu$ sont des mesures de probabilité, les trois définitions coïncident.*

On note $\lim\limits_{n\to+\infty} \mu_n = \mu$.

Il y a unicité de la limite si les mesures sont des probabilités : si $\int f d\mu = \int f d\mu'$, $\forall f \in \mathcal{C}_c(\mathbb{R}^d)$ alors $\mu([a,b]) = \mu'([a,b])$ et donc (si $\mu$ et $\mu'$ sont $\sigma$-finies) $\mu = \mu'$.

**Exemples.**

1. *Si $x_n \in \mathbb{R}^d$ avec $x_n \to x \in \mathbb{R}^d$ alors $\lim\limits_{n\to+\infty} \delta_{x_n} = \delta_x$*

   *($\forall f \in \mathcal{C}_c(\mathbb{R}^d), \int f d\delta_{x_n} = f(x_n) \to f(x) = \int f \delta_x$)*

2. *Si $x_n \in \mathbb{R}^d$ avec $\|x_n\| \to +\infty$ alors $\lim\limits_{n\to+\infty} \delta_{x_n} = 0$*

   *($\forall f \in \mathcal{C}_c(\mathbb{R}^d), f(x_n) \to 0$)*

3. *Soit $d = 1$ et $\mu_n = \frac{1}{n}\sum\limits_{k=0}^{n-1} \delta_{\frac{k}{n}}$ : $\lim\limits_{n\to+\infty} \mu_n = \lambda_{|[0,1]}$ (mesure de Lebesgue).*

   $\forall f \in \mathcal{C}_c(\mathbb{R}), \int f d\mu_n = \frac{1}{n}\sum\limits_{k=0}^{n-1} f(\frac{k}{n}) \to \int_0^1 f(x)dx = \int f d\lambda_{|[0,1]}$.

**Théorème 48.** *Soit $\mu$ et $\mu_n$ des mesures de probabilité sur $\mathbb{R}$, de fonctions de répartitions $F$ et $F_n$, $n \geq 1$ ($\forall t \in \mathbb{R}, F(t) = \mu(]-\infty, t])$). On a l'équivalence entre :*

*i)* $\lim\limits_{n\to+\infty} \mu_n = \mu$

*ii)* $\forall t \in \mathbb{R}$ *tq* $F(t) = F(t^-)$ *(i.e.* $\mu(\{t\}) = 0$*),*

$$\lim_{n\to+\infty} F_n(t) = F(t).$$

**Preuve** On commence par $i) \Rightarrow ii)$

Soit $t$ tel que $\mu(\{t\}) = 0$. On approche $\mathbf{1}_{]-\infty,t])}(x)$ par $g_k(x)$ et $f_k(x)$ avec $g_k \in \mathcal{C}_b(\mathbb{R})$, $f_k \in \mathcal{C}_b(\mathbb{R})$,

$$0 \leq g_k(x) \leq \mathbf{1}_{]-\infty,t])}(x) \leq f_k(x)$$

et

$$\forall x \neq t, \lim_{k\to+\infty} g_k(x) = \mathbf{1}_{]-\infty,t])}(x) = \lim_{k\to+\infty} f_k(x).$$

**Faire dessin pour mieux visualiser $f_k$ et $g_k$**

Comme $\mu(\{t\}) = 0$, $\lim g_k = \lim f_k$ $\mu$-presque partout. On a ensuite

$$\int g_k d\mu_n \leq F_n(t) \leq \int f_k d\mu_n.$$

On passe à la limite en $n$ :

$$\int g_k d\mu \leq \liminf_{n\infty} F_n(t) \leq \limsup_{n\infty} F_n(t) \leq \int f_k d\mu.$$

On passe à la limite en $k$ :

$$F(t) \leq \liminf_{n\infty} F_n(t) \leq \limsup_{n\infty} F_n(t) \leq F(t).$$

On montre ensuite : $ii) \Rightarrow i)$. Soit $f \in \mathcal{C}_c(\mathbb{R})$.

- 1er cas : $f \in \mathcal{C}^1$

  On a $f(x) = \int\limits_{-\infty}^{x} f'(y)dy$ donc

$$\begin{aligned}\int_{\mathbb{R}} f(x)d\mu(x) &= \int_{\mathbb{R}} \int_{-\infty}^{x} f'(y)dyd\mu(x)\\ &= \int_{\mathbb{R}} f'(y) \int_{y}^{+\infty} d\mu(x)dy \text{ (Fubini)}\\ &= \int_{\mathbb{R}} f'(y)(1 - F(y))dy.\end{aligned}$$

On a donc

$$\int f d\mu_n = \int_{\mathbb{R}} f'(y)(1 - F_n(y))dy \to \int_{\mathbb{R}} f'(y)(1 - F(y))dy = \int f d\mu$$

par convergence dominée : $F$ est croissante donc $\{y : F(y) \neq F(y^-)\}$ est dénombrable donc avec $ii)$, $\lim\limits_{n\to+\infty} F_n(y) = F(y)$, $\lambda$ pp.

- 2ème cas : $\mathcal{C}_c(\mathbb{R}) \cap \mathcal{C}^1$ est dense dans $\mathcal{C}_c(\mathbb{R})$ pour la norme infinie.
  *[On peut utiliser la convolution avec une identité approchée à support compact et $\mathcal{C}^1$]*

*Remarque.* Si ii) a lieu et si $F$ est continue, la convergence est uniforme. En effet, $t \mapsto F_n(t)$ est croissante. Le deuxième théorème de Dini donne

$$\lim_{n\to+\infty} \sup_{t\in\mathbb{R}} |F_n(t) - F(t)| = 0.$$

**Définition 34.** *Soient $X$, $X_n$, $n \geq 1$ v.a. de $\mathbb{R}^d$. On dit que $(X_n)_n$* **converge en loi** *vers $X$ si*

$$\lim_{n\to+\infty} \mathbb{P}_{X_n} = \mathbb{P}_X$$

*(i.e. $\forall f \in \mathcal{C}_{b,o,c}(\mathbb{R}^d)$, $\lim\limits_{n\to+\infty} \mathbf{E}[f(X_n)] = \mathbf{E}[f(X)]$.)*

*On note $X_n \xrightarrow{\mathcal{L}} X$.*

**Exemple.** $\Omega = \{0, 1\}$, $\mathbb{P} = \frac{1}{2}(\delta_0 + \delta_1)$, $X_n = X = \mathbf{1}_{\{0\}}$, $Y = \mathbf{1}_{\{1\}}$.

*On a $\mathbf{E}[f(X_n)] = \frac{1}{2}f(0) + \frac{1}{2}f(1) = \mathbf{E}[f(X)] = \mathbf{E}[f(Y)]$ donc $X_n \xrightarrow{\mathcal{L}} X$, $X_n \xrightarrow{\mathcal{L}} Y$, $X(\omega) \neq Y(\omega)$ mais $X \stackrel{\mathcal{L}}{=} Y$ ($\mathbb{P}_X = \mathbb{P}_Y$).*

**Proposition 49.** *Soient $X$, $X_n$, $n \geq 1$ des v.a. de $\mathbb{R}^d$.*

- *Si $X_n \xrightarrow{\mathbb{P}} X$ alors $X_n \xrightarrow{\mathcal{L}} X$.*
- *La réciproque est fausse en général.*
- *Si $X_n \xrightarrow{\mathcal{L}} a \in \mathbb{R}^d$ alors $X_n \xrightarrow{\mathbb{P}} a$.*

**Preuve**

- On suppose $\forall\varepsilon > 0$,

$$\lim_{n\to+\infty} \mathbb{P}(\|X_n - X\| > \varepsilon) = 0.$$

Soit $f \in \mathcal{C}_c(\mathbb{R}^d)$.

$$\begin{aligned}|\mathbf{E}[f(X_n)] - \mathbf{E}[f(X)]| &\leq \int_\Omega |f(X_n) - f(X)| d\mathbb{P} \\ &= \int_{\|X_n - X\| > \varepsilon} |f(X_n) - f(X)| d\mathbb{P} + \int_{\|X_n - X\| \leq \varepsilon} |f(X_n) - f(X)| d\mathbb{P} \\ &\leq 2\|f\|_\infty \mathbb{P}(\|X_n - X\| > \varepsilon) + \Delta(\varepsilon)\mathbb{P}(\|X_n - X\| \leq \varepsilon) \\ &\leq 2\|f\|_\infty \mathbb{P}(\|X_n - X\| > \varepsilon) + \Delta(\varepsilon)\end{aligned}$$

où $\Delta(\varepsilon) = \sup\{|f(x) - f(y)| : x, y \in \mathbb{R}^d, \|x - y\| \leq \varepsilon\}$. Comme $f \in \mathcal{C}_c(\mathbb{R}^d)$, $f$ est uniformément continue donc $\lim_{\varepsilon \to 0} \Delta(\varepsilon) = 0$.

- Contre-exemple :
  On considère $X_0, X_1, ..., X_n \sim \mathcal{N}(0, 1)$ indépendantes. On pose $Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$ indép de $X_0$. On sait que $Y_n \sim \mathcal{N}(0, 1)$ i.e. $Y_n \overset{\mathcal{L}}{=} X_0$ donc $Y_n \xrightarrow{\mathcal{L}} X_0$.
  Or pour $\varepsilon > 0$, $\mathbb{P}(\||Y_n - X_0|\| > \varepsilon) = \mathbb{P}(|Z| > \varepsilon)$ où $Z \sim \mathcal{N}(0, 2)$.
- On suppose $X_n \xrightarrow{\mathcal{L}} a \in \mathbb{R}^d$. Soit $\varepsilon > 0$.

$$\begin{aligned}\mathbb{P}(\|X_n - a\| > \varepsilon) &= \mathbf{E}[\mathbf{1}_{\frac{\|x-a\|}{\varepsilon} > 1}(X_n)] \\ &\leq \mathbf{E}[1 \wedge \frac{\|x - a\|}{\varepsilon}(X_n)].\end{aligned}$$

Or $x \mapsto 1 \wedge \frac{\|x-a\|}{\varepsilon} \in \mathcal{C}_b(\mathbb{R}^d)$ donc $\mathbf{E}[1 \wedge \frac{\|x-a\|}{\varepsilon}(X_n)] \to \mathbf{E}[1 \wedge \frac{\|x-a\|}{\varepsilon}(a)] = 0$ donc

$$\lim_{n \to +\infty} \mathbb{P}(\|X_n - a\| > \varepsilon).$$

# 3 Théorème de Paul Levy

Le théorème de Paul Levy caractérise la convergence en loi a l'aide de la convergence simple des fonctions caractéristiques. Ce critère est très utile, et nous servira entre autres à démontrer le théorème de la limite centrale.

**Théorème 50** (Paul Levy). *Soit $X_n$ une suite de variables aléatoires*

*a) Si la suite $X_n$ converge en loi vers $X$, alors $\varphi_(X_n)$ converge simplement vers $\phi(X)$.*

*b) Si la suite $\varphi_{X_n}$ converge simplement vers une fonction $\varphi$ et si cette fonction est continue en 0, alors cette fonction $\varphi$ est la fonction caractéristique d'une v.a. $X$ et la suite $X_n$ converge en loi vers $X$*

**Preuve** Pas de preuve ici.

# 4 Théorème de la Limite Centrale

**Théorème 51** (TLC)**.** *Soit* $(X_n)_{n\geq 1}$ *une suite de v.a.i.i.d. telle que* $\mathbf{E}[X_1^2] < +\infty$*. On note* $\mu = \mathbf{E}[X_1]$ *et* $\sigma^2 = \mathbf{Var}(X_1) > 0$*. On a la convergence en loi*

$$\frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{j=1}^{n} X_j - \mu\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

*Remarques.*

- $\frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{j=1}^{n} X_j - \mu\right) = \sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{X_j-\mu}{\sigma}\right)$ : on centre et on réduit chaque variable $X_j$, ou encore
- $\mathbf{E}[\frac{\sqrt{n}}{\sigma}(M_n-\mu)] = 0$ et $\mathbf{Var}(\frac{\sqrt{n}}{\sigma}(M_n-\mu)) = 1$ : on centre et on réduit $M_n$ à chaque $n$.
- La loi limite universelle est la loi Gaussienne si $0 < \mathbf{Var}(X_1) < +\infty$.

**Preuve**

- Avec les remarques, on peut supposer $\mu = 0$ et $\sigma = 1$, et montrer $\frac{S_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$.
- Soit $t \in \mathbb{R}$ et $\varphi_n$ la fonction caractéristique de $\frac{S_n}{\sqrt{n}}$ :

  $$\varphi_n(t) = \mathbf{E}[e^{\frac{it}{\sqrt{n}}\sum_{j=1}^{n} X_j}] = \left(\varphi_{X_1}(\frac{t}{\sqrt{n}})\right)^n$$

  par indépendance et identique distribution.
- Développement limité en 0 de $\varphi_{X_1}$ :

  $$\varphi_{X_1}(\theta) = 1 - \frac{\theta^2}{2} + \theta^2\varepsilon(\theta)$$

  avec $\lim\limits_{\theta\to 0}\varepsilon(\theta) = 0$ (voir Partie II).
- On pose $Log(1+u) = \sum_{n\geq 1}(-1)^{n+1}\frac{u^n}{n}$ pour $u \in \mathbb{C}$, $|u| < 1$. On a alors $\forall |u| < 1$,

  $$\exp(Log(1+u)) = 1 + u$$

  et au voisinage de 0, $Log(1+u) \sim u$.
- Soit $t \in \mathbb{R}$. Pour $n$ assez grand, on a

  $$\left|-\frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\left(\frac{t}{\sqrt{n}}\right)\right| < 1$$

  L'expression

  $$Log\left(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\left(\frac{t}{\sqrt{n}}\right)\right)$$

a un sens et on peut ecrire

$$\varphi_n(t) = \left(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon(\frac{t}{\sqrt{n}})\right)^n = \exp(nLog\left(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\left(\frac{t}{\sqrt{n}}\right)\right)).$$

- Comme

$$Log(1+u) \underset{u\to 0}{\sim} u$$

On a

$$Log\left(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\left(\frac{t}{\sqrt{n}}\right)\right) \underset{n\to\infty}{\sim} -\frac{t^2}{2n}$$

donc

$$\lim_{n\to+\infty} nLog\left(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\left(\frac{t}{\sqrt{n}}\right)\right) = -\frac{t^2}{2}$$

donc par continuité de exp

$$\lim_{n\to+\infty} \varphi_n(t) = e^{-\frac{t^2}{2}}$$

Or un variable aléatoire $Z \sim \mathcal{N}(0,1)$ a pour fonction caractéristique

$$\varphi_Z(t) = e^{-\frac{t^2}{2}}$$

On a donc convergence **simple** des $\varphi_n$ vers $\varphi_Z$ avec $Z \sim \mathcal{N}(0,1)$.
- On conclut, par le théorème de Paul Lévy, que la suite de variables aléatoires $\frac{S_n}{\sqrt{n}}$ converge en loi vers $\mathcal{N}(0,1)$.