Rappels utiles au cours de statistique mathématique

O. Wintenberger

Les résultats d'algèbre linéaire et de probabilités présentés ici sont utiles à la compréhension des méthodes développées dans le cours de statistique mathématique. Dans ce fascicule sont rappelé des résultats connus et un résultat probabiliste nouveau : le théorème de Cochran. Ce théorème sera détaillé durant le cours magistral de statistique mathématique et non dans le polycopié de ce cours.

Je remercie Jean Marc Bardet et Vincent Rivoirard pour avoir mis à ma disposition leurs notes de cours.

Chapitre 1

Rappels d'algèbre linéaire

Ce premier chapitre a pour objet de rappeler les principaux éléments d'algèbre linéaire nécessaires pour l'étude des méthodes développées dans le cours de statistique mathématique.

1.1 Rappels généraux sur les matrices

On ne considère que des matrices réelles. Pour toute matrice A, on note A^T la matrice transposée de A. On note I la matrice identité.

Proposition 1 *Si A et B sont deux matrices telles que le nombre de colonnes de B correspond au nombre de lignes de A, on a : $(AB)^T = B^T A^T$.*

Définition 1 *On dit qu'une matrice carrée A est inversible s'il existe une matrice B telle que*

$$AB = BA = I.$$

Dans ce cas, on note $B = A^{-1}$.

Proposition 2 *Une matrice carrée A est inversible si et seulement si* $\det(A) \neq 0$. *Dans ce cas, on a :*

$$A^{-1} = \frac{1}{\det(A)}\tilde{A}^T,$$

où $\tilde{A}$ est la comatrice de A. Par ailleurs, si A et B sont deux matrices inversibles de même taille,

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Définition 2 *Soit A une matrice carrée. λ est une valeur propre de A si et seulement si* $\det(A - \lambda I) = 0$. *Un vecteur propre associé à la valeur propre λ est un vecteur x non nul satisfaisant $Ax = \lambda x$.*

Définition 3 *Soit A une matrice. Le rang de A est la plus petite des dimensions des deux sous-espaces engendrés par les lignes et par les colonnes de A.*

Proposition 3 *Soit A une matrice possédant n lignes et p colonnes. On a :*

1. $0 \leq rang(A) \leq \min(n; p)$
2. $rang(A) = rang(A^T)$
3. $rang(A) + dim(Ker(A)) = p$ *(formule du rang)*
4. $rang(AA^T) = rang(A^T A) = rang(A)$
5. *pour $p \leq n$, si $rang(A) = p$ ($\iff$ A injective), alors $A^T A$ est inversible*
6. *pour B matrice de p lignes et q colonnes, $rang(AB) \leq \min(rang(A); rang(B))$*

Définition 4 *Soit A une matrice carrée. On note n le nombre de ses lignes (ou de ses colonnes). La trace de A est définie par :*

$$Tr(A) = \sum_{i=1}^{n} A_{ii}.$$

Proposition 4 *Soient A et B deux matrices carrées de taille $n \times n$. On a :*

1. $Tr(A + B) = Tr(A) + Tr(B)$
2. $Tr(AB) = Tr(BA)$
3. $Tr(AA^T) = Tr(A^T A) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2$
4. $\det(AB) = \det(A) \det(B)$

1.2 Matrices symétriques réelles

Définition 5 *Une matrice carrée A est dite symétrique si $A^T = A$. On rappelle également que A est orthogonale si $A^T = A^{-1}$.*

Théorème 1 *Les valeurs propres d'une matrice symétrique sont toutes réelles. De plus elle est diagonalisable dans une base orthonormale (BON) de vecteurs propres : il existe U matrice orthogonale (formée par les vecteurs propres de A) telle que*

$$U^T AU = D,$$

où D est une matrice diagonale, la diagonale étant formée des valeurs propres de A. En particulier, le rang d'une matrice symétrique correspond au nombre de valeurs propres non-nulles.

Définition 6 *Une matrice carrée A est dite semi-définie positive si et seulement pour tout vecteur x,*

$$x^T A x \geq 0.$$

Elle est dite définie positive si et seulement si pour tout vecteur x non nul,

$$x^T A x > 0.$$

Proposition 5 *Les valeurs propres d'une matrice semi-définie positive sont toutes positives ou nulles, celle d'une matrice définie positive sont toutes strictement positives.*

Théorème 2 *Soit A une matrice symétrique semi-définie positive. Alors il existe une matrice Γ symétrique telle que*

$$A = \Gamma^2 = \Gamma^T \Gamma.$$

Preuve : Il suffit d'écrire $A = U^T D U$ et de poser $\Gamma = U^T \sqrt{D} U$ où si

$$D = \begin{pmatrix} {}_1 & & & & & \\ & \ddots & & & & \\ & & {}_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix},$$

$$\sqrt{D} = \begin{pmatrix} \sqrt{{}_1} & & & & & \\ & \ddots & & & & \\ & & \sqrt{{}_r} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}.$$

1.3 Matrices de projections orthogonales

Définition 7 *Une matrice P est une matrice de projection orthogonale si elle vérifie $P^2 = P$ et $P^T = P$. On rappelle que P est la projection sur $Im(P)$ (parallèlement à $Ker(P)$).*

On munit $\mathbb{R}^p$ de la structure d'un espace de Hilbert avec le produit scalaire $< X, Y >= X^T Y$ qui induit la norme $\|X\| = \sqrt{X^T X}$. On vérifie qu'une matrice de projection orthogonale caractérise bien une projection orthogonale :

Proposition 6 *Une matrice de projection orthogonale P se caractérise par la propriété :*

$$(Pu)^T(y - Py) = 0, \qquad \textit{pour tout } u \textit{ et tout } y.$$

De plus, $I - P$ est la matrice de projection sur Ker(P).

Proposition 7 *L'ensemble des valeurs propres de P est contenu dans $\{0, 1\}$.*

En particulier P est symétrique positive donc diagonalisable dans une BON $U^TPU = D$ avec D diagonale contenant que des 1 ou des 0. Le nombre de 1 correspond au rang de P.

Proposition 8 *Pour toute projection orthogonale P de rang p il existe une BON tel que $P(X) = (x_1, \ldots, x_p, 0, \ldots, 0)$ pour tout $X = (x_1, \ldots, x_p)$ exprimé dans cette BON.*

Chapitre 2

Rappels de probabilités

Ce chapitre a pour objet de rappeler des notions de probabilités utiles pour le cours de statistique mathématique.

2.1 Rappels sur la théorie de la mesure

2.1.1 Mesures

Une mesure est une fonction positive de l'ensemble (tribu) des événements $\mathcal{A}$ d'un espace mesurable $(\Omega, \mathcal{A})$.

• **<u>Tribus</u>**

Dans toute la suite nous adoptons les notations standards :
- Ω est un ensemble (fini ou infini).
- $\mathcal{P}(\Omega)$ est l'ensemble de tous les sous-ensembles (parties) de Ω.

Rappel 1 (Dénombrabilité) *Soit E un ensemble. E est dit dénombrable s'il existe une bijection entre E et $\mathbf{N}$ ou un sous-ensemble de $\mathbf{N}$. Par exemple, un ensemble fini, $\mathbf{Z}$, $\mathbf{D}$, $\mathbf{Z} \times \mathbf{Z}$, $\mathbf{Q}$ sont dénombrables. En revanche, $\mathbf{R}$ n'est pas dénombrable.*

Définition 8 *Soit une famille $\mathcal{F}$ de parties de Ω (donc $\mathcal{F} \subset \mathcal{P}(\Omega)$). On dit que $\mathcal{F}$ est une <u>algèbre</u> si :*
- $\Omega \in \mathcal{F}$;
- *lorsque $A \in \mathcal{F}$ alors le <u>complémentaire</u> $A^c = \overline{A} = (\Omega \setminus A)$ appartient à $\mathcal{F}$;*
- *pour tout $n \in \mathbf{N}^T$, lorsque $(A_1, \cdots, A_n) \in \mathcal{F}^n$ alors la <u>réunion</u> $A_1 \cup \cdots \cup A_n \in \mathcal{F}$.*

Définition 9 *Soit une famille $\mathcal{A}$ de parties de Ω (donc $\mathcal{A} \subset \mathcal{P}(\Omega)$). On dit que $\mathcal{A}$ est une tribu (ou σ-algèbre) sur Ω si :*
- *$\Omega \in \mathcal{A}$;*
- *lorsque $A \in \mathcal{A}$ alors $A^c \in \mathcal{A}$;*
- *pour $I \subset \mathbf{N}$, lorsque $(A_i)_{i \in I} \in \mathcal{A}^I$ alors $\bigcup_{i \in I} A_i \in \mathcal{A}$.*

Tout sous ensemble A de Ω élément de la tribu $\mathcal{A}$ est appelé un événement.

Propriété 1 *Avec les notations précédentes :*
1. *$\emptyset \in \mathcal{A}$;*
2. *si A et B sont dans la tribu $\mathcal{A}$, alors $A \cap B$ est dans $\mathcal{A}$;*
3. *si $\mathcal{A}_1$ et $\mathcal{A}_2$ sont deux tribus sur Ω, alors $\mathcal{A}_1 \cap \mathcal{A}_2$ est une tribu sur Ω . Plus généralement, pour $I \subset \mathbf{N}$, si $(\mathcal{A}_i)_{i \in I}$ ensemble de tribus sur Ω, alors $\bigcap_{i \in I} \mathcal{A}_i$ est une tribu sur Ω ;*
4. *si $\mathcal{A}_1$ et $\mathcal{A}_2$ sont deux tribus sur Ω, alors $\mathcal{A}_1 \cup \mathcal{A}_2$ n'est pas forcément une tribu sur Ω.*

Définition 10 *Si $\mathcal{E}$ est une famille de parties de Ω (donc $\mathcal{E} \subset \mathcal{P}(\Omega)$), alors on appelle tribu engendrée par $\mathcal{E}$, notée $\sigma(\mathcal{E})$, la tribu engendrée par l'intersection de toutes les tribus contenant $\mathcal{E}$ (on peut faire la même chose avec des algèbres).*

La tribu engendrée est la "plus petite" tribu (au sens de l'inclusion) contenant la famille $\mathcal{E}$.

Rappel 2 (Topologie)
- *Un ensemble ouvert U dans un espace métrique X (muni d'une distance d) est tel que pour tout $x \in U$, il existe $r > 0$ tel que $B(x, r) \subset U$ ou $B(x, r) = \{y \in X;\ d(x, y) < r\}$.*
- *On dit qu'un ensemble dans un espace métrique X est fermé si son complémentaire dans X est ouvert.*

Une tribu naturelle est celle engendrée par les ensembles ouverts (et donc fermés) :

Définition 11 *Soit Ω un espace métrique. On appelle tribu borélienne sur Ω, notée, $\mathcal{B}(\Omega)$, la tribu engendrée par les ouverts de Ω. Un ensemble de $\mathcal{B}(\Omega)$ est appelé borélien.*

• Espace mesurable

Lorsque Ω est un ensemble et $\mathcal{A}$ une tribu sur Ω on dit que $(\Omega, \mathcal{A})$ est un espace mesurable. Quand on s'intéressera aux probabilités, on dira de même que $(\Omega, \mathcal{A})$ est un espace probabilisable.

Propriété 2 *Si $(\Omega_i, \mathcal{A}_i)_i$ sont n espaces mesurables, alors un ensemble élémentaire de $\Omega = \Omega_1 \times \cdots \times \Omega_n$ est une réunion finie d'ensembles $A_1 \times \cdots \times A_n$ appelés* <u>*pavés*</u> *ou* <u>*cylindres*</u>*, où chaque $A_i \in \mathcal{A}_i$. L'ensemble des ensembles élémentaires est une algèbre et on note $\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n = \bigotimes_{i=1}^n \mathcal{A}_i$ (on dit $\mathcal{A}_1$ tensoriel $\mathcal{A}_2$ ... tensoriel $\mathcal{A}_n$) la tribu sur Ω engendrée par ces ensembles élémentaires.*

Définition 12 *On appelle* <u>*espace mesurable produit*</u> *des $(\Omega_i, \mathcal{A}_i)_i$ l'espace mesurable $\left(\prod_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{A}_i\right)$.*

Il et désormais temps de définir ce qu'est une mesure :

Définition 13 *Soit $(\Omega, \mathcal{A})$ un espace mesurable. L'application $\mu \ : \mathcal{A} \to [0, +\infty]$ est une* <u>*mesure*</u> *si :*
- $\mu(\emptyset) = 0$.
- *Pour tout $I \subset \mathbf{N}$ et pour $(A_i)_{i \in I}$ famille disjointe de $\mathcal{A}$ (telle que $A_i \cup A_j = \emptyset$ pour $i \neq j$), alors $\mu\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i)$ (propriété dite de σ-additivité).*

Une mesure pondère les événements. Avec les notations précédentes :
- Si $\mu(\Omega) < +\infty$, on dit que μ est finie.
- Si $\mu(\Omega) < M$ avec $M < +\infty$, on dit que μ est bornée.
- Si $\mu(\Omega) = 1$, on dit que μ est une mesure de probabilité.

Définition 14 *Si $(\Omega, \mathcal{A})$ est un espace mesurable (resp. probabilisable) alors $(\Omega, \mathcal{A}, \mu)$ est un* <u>*espace mesuré*</u> *(resp.* <u>*probabilisé*</u> *quand μ est une probabilité).*

Sur $(\Omega, \mathcal{A})$, on peut définir une infinité de mesures. Elles ont toutes les propriétés suivantes :

Propriété 3 *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $(A_i)_{i \in \mathbf{N}}$, une famille de $\mathcal{A}$.*
1. *Si $A_1 \subset A_2$, alors $\mu(A_1) \leq \mu(A_2)$.*
2. *Si $\mu(A_1) < +\infty$ et $\mu(A_2) < +\infty$, alors $\mu(A_1 \cup A_2) + \mu(A_1 \cap A_2) = \mu(A_1) + \mu(A_2)$.*
3. *Pour tout $I \subset \mathbf{N}$, on a $\mu\left(\bigcup_{i \in I} A_i\right) \leq \sum_{i \in I} \mu(A_i)$.*
4. *Si $A_i \subset A_{i+1}$ pour tout $i \in \mathbf{N}$ (suite croissante en sens de l'inclusion), alors $(\mu(A_n))_{n \in \mathbf{N}}$ est une suite croissante et $\mu\left(\bigcup_{i \in \mathbf{N}} A_i\right) = \lim_{i \to +\infty} \mu(A_i)$ (même si cette limite est $+\infty$).*

5. *Si* $A_{i+1} \subset A_i$ *pour tout* $i \in \mathbf{N}$ *(suite décroissante en sens de l'inclusion) et* $\mu(A_0) < +\infty$*, alors* $(\mu(A_n))_{n\in\mathbf{N}}$ *est une suite décroissante convergente telle que* $\mu\left(\bigcap_{i\in\mathbf{N}} A_i\right) = \lim_{i\to+\infty} \mu(A_i)$.

Définition 15 *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espace mesuré et* $(A_i)_{i\in\mathbf{N}}$ *une famille de* $\mathcal{A}$.

1. *On définit la* <u>*limite supérieure*</u> *des événements* $\limsup(A_n)_n = \bigcap_{n\in\mathbf{N}} \bigcup_{m\geq n} A_m$.
   Intuitivement, $\limsup(A_n)_n$ *est l'ensemble des* $\omega \in \Omega$ *tels que* ω *appartienne à une infinité de* A_n.
2. *On définit la* <u>*limite inférieure*</u> *des événements* $\liminf(A_n)_n = \bigcup_{n\in\mathbf{N}} \bigcap_{m\geq n} A_m$.
   Intuitivement, $\liminf(A_n)_n$ *est l'ensemble des* $\omega \in \Omega$ *tels que* ω *appartienne à tous les* A_n *sauf à un nombre fini d'entre eux.*

On vérifie que $\limsup$ et $\liminf$ sont des événements vérifiant $\mu(\limsup(A_n)_n) = \lim \mu(\bigcup_{m\geq n} A_m)$ et $\mu(\liminf(A_n)_n) = \lim \mu(\bigcap_{m\geq n} A_m)$ les limites étant bien définies car portant sur des suites décroissantes et croissantes d'événements. Ci-dessous un résultat utile à la définition de la mesure de Lebesgue sur $\mathbf{R}$, $\mathbf{R}^n$,...

Théorème 3 (Théorème d'extension de Hahn - Caratheodory) *Si* Ω *est un ensemble,* $\mathcal{F}$ *une algèbre sur* Ω*, et* ν *une application de* $\mathcal{F}$ *dans* $[0, +\infty]$ *additive (telle que* $\nu(A\cup B) = \nu(A)+\nu(B)$ *pour* $A\cup B = \emptyset$*), alors si* $\mathcal{A}$ *est la tribu engendrée par* $\mathcal{F}$*, il existe une mesure* $\widehat{\nu}$ *sur la tribu* $\mathcal{A}$ *qui coïncide avec* ν *sur* $\mathcal{F}$ *(c'est-à-dire que pour tout* $F \in \mathcal{F}$*,* $\widehat{\nu(F)} = \nu(F)$*). On dit que* $\widehat{\nu}$ *prolonge* ν *sur la tribu* $\mathcal{A}$.

Définition 16 *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espace mesuré.*

1. *Pour* $A \in \mathcal{A}$*, on dit que* A *est* <u>μ*-négligeable*</u> *si* $\mu(A) = 0$.
2. *Soit une propriété* $\mathcal{P}$ *dépendant des éléments* ω *de* Ω*. On dit que* $\mathcal{P}$ *est vraie* <u>μ*-presque partout*</u> *(*<u>μ*-presque sûrement*</u> *(p.s.) sur un espace probabilisé) si l'ensemble des* ω *pour laquelle elle n'est pas vérifiée est* μ*-négligeable.*

Ainsi la propriété “ la suite de fonction $f_n(x) = x^n$ converge vers la fonction $f(x) = 0$” est vraie λ-presque partout sur $[0, 1]$.

• <u>Fonctions mesurables</u>

Soit $f : E \mapsto F$, où E et F sont 2 espaces métriques.

- Pour $I \subset F$, on appelle ensemble réciproque de I par f, l'ensemble $f^{-1}(I) = \{x \in E, f(x) \in I\}$.

– (f continue) $\Longleftrightarrow$ (pour tout ouvert U de F alors $f^{-1}(U)$ est un ouvert de E).
– On note $f^{-1}(\mathcal{I})$ l'ensemble de sous-ensembles de Ω tel que $f^{-1}(\mathcal{I}) = \{f^{-1}(I), I \in \mathcal{I}\}$.
– Soit $(\Omega', \mathcal{A}')$ un espace mesurable et soit $f : \Omega \mapsto \Omega'$. Alors $f^{-1}(\mathcal{A})$ est une tribu sur Ω appelée tribu engendrée par f.

Définition 17 *Soit $(\Omega, \mathcal{A})$ et $(\Omega', \mathcal{A}')$ deux espaces mesurables. Une fonction $f : \Omega \mapsto \Omega'$ est dite mesurable pour les tribus $\mathcal{A}$ et $\mathcal{A}'$ si et seulement si $f^{-1}(\mathcal{A}') \subset \mathcal{A}$ (donc si et seulement si $\forall A' \in \mathcal{A}'$, alors $f^{-1}(A') \in \mathcal{A}$).*

Par exemple, les fonctions indicatrices d'événements et les combinaisons linéaires de telles fonctions indicatrices sont mesurables. Dans le cas où $(\Omega, \mathcal{A})$ est un espace probabilisable, et si $f : \Omega \mapsto \mathbf{R}$, alors si f est une fonction mesurable pour $\mathcal{A}$ et $\mathcal{B}(\mathbf{R})$, alors f est une variable aléatoire. Dans le cas où $(\Omega, \mathcal{A})$ est un espace mesurable, et si $f : \Omega \mapsto (\Omega', \mathcal{B}(\Omega'))$, où Ω' est un espace métrique et $\mathcal{B}(\Omega')$ l'ensemble des boréliens ou tribu borélienne de Ω', si f est une fonction mesurable sur $\mathcal{A}$ et $\mathcal{B}(\Omega')$, alors f est dite fonction borélienne.

Proposition 9 *Soit $(\Omega, \mathcal{A})$ et $(\Omega', \mathcal{A}')$ deux espaces mesurables et $f : \Omega \mapsto \Omega'$. Soit $\mathcal{F}$ une famille de sous-ensembles de Ω' telle que $\sigma(\mathcal{F}) = \mathcal{A}'$. Alors*

1. *$f^{-1}(\mathcal{F})$ engendre la tribu $f^{-1}(\mathcal{A})$.*
2. *(f mesurable) $\Longleftrightarrow$ ($f^{-1}(\mathcal{F}) \subset \mathcal{A}$)*

En particulier si $(\Omega, \mathcal{A})$ et $(\Omega', \mathcal{A}')$ sont deux espaces mesurables boréliens, alors toute application continue de $\Omega \mapsto \Omega'$ est mesurable. De plus, pour montrer qu'une fonction $f : \Omega \mapsto \mathbf{R}$ est mesurable, il suffit de montrer que la famille d'ensemble $(\{\omega \in \Omega, f(\omega) \leq a\})_{a \in \mathbf{R}} \in \mathcal{A}$.

Propriété 4

1. *Soit f mesurable de $(\Omega, \mathcal{A})$ dans $(\Omega', \mathcal{A}')$ et g mesurable de $(\Omega', \mathcal{A}')$ dans $(\Omega'', \mathcal{A}'')$. Alors $g o f$ est mesurable pour $\mathcal{A}$ et $\mathcal{A}'$.*
2. *Soit f_1 mesurable de $(\Omega, \mathcal{A})$ dans $(\Omega_1, \mathcal{A}_1)$ et f_2 mesurable de $(\Omega, \mathcal{A})$ dans $(\Omega_2, \mathcal{A}_2)$. Alors $h : \Omega \mapsto \Omega_1 \times \Omega_2$ telle que $h(\omega) = (f_1(\omega), f_2(\omega))$ est mesurable pour $\mathcal{A}$ et $\mathcal{A}_1 \otimes \mathcal{A}_2$.*
3. *Soit $(f_n)_{n \in \mathbf{N}}$ une suite de fonctions mesurables de $(\Omega, \mathcal{A})$ dans $(\Omega', \mathcal{B}(\Omega'))$, où Ω' est un espace métrique, telle qu'il existe une fonction f limite simple de (f_n) (donc $\forall \omega \in \Omega$, $\lim\limits_{n \to \infty} f_n(\omega) = f(\omega)$). Alors f est mesurable pour $\mathcal{A}$ et $\mathcal{B}(\Omega')$.*

Définition 18 *Soit f mesurable de $(\Omega, \mathcal{A}, \mu)$ dans $(\Omega', \mathcal{A}')$ et soit $\mu_f : \mathcal{A}' \mapsto [0, +\infty]$ telle que pour tout $A' \in \mathcal{A}'$, on ait $\mu_f(A') = \mu(f^{-1}(A'))$. Alors μ_f est une mesure sur $(\Omega', \mathcal{A}')$ appelée mesure image de μ par f.*

Si μ est une mesure de probabilité et si X est une variable aléatoire alors μ_X est la mesure (loi) de probabilité de la variable aléatoire X.

• **Cas des fonctions réelles mesurables**

Propriété 5 *1. Soit f et g deux fonctions réelles mesurables (de $(\Omega, \mathcal{A}, \mu)$ dans $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$). Alors $\alpha.f$, $f+g$, $\min(f,g)$ et $\max(f,g)$ sont des fonctions réelles mesurables.*

2. Soit $(f_n)_{n\in\mathbf{N}}$ une suite de fonctions réelles mesurables. Alors $\inf(f_n)$ et $\sup(f_n)$ sont des fonctions réelles mesurables.

Définition 19 *Soit $f : \Omega \to \mathbf{R}$. Alors f est dite étagée s'il existe une famille d'ensembles disjoints $(A_i)_{1\leq i\leq n}$ de Ω et une famille de réels $(\alpha_i)_{1\leq i\leq n}$ telles que pour tout $\omega \in \Omega$, on ait $f(\omega) = \sum_{i=1}^{n} \alpha_i \mathbf{I}_{A_i}(\omega)$.*

Si les A_i sont tous dans $\mathcal{A}$ tribu sur Ω, alors f est $\mathcal{A}$-mesurable.

Théorème 4 *Toute fonction réelle mesurable à valeurs dans $[0, +\infty]$ est limite simple d'une suite croissante de fonctions étagées.*

On en déduit que si f une fonction réelle mesurable. Alors f est limite simple de fonctions étagées. Pour des calculs sur les fonctions mesurables, il suffit donc de calculer sur les fonctions étagées puis de passer à la limite. C'est la méthode de Lebesgue dans le cas de l'intégration.

2.1.2 Intégration de Lebesgue

Dans toute la suite, on considère $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. On procède par étape pour définir l'intégrale de Lebesgue. L'intégrale de Lebesgue d'une fonction positive est définie à partir de l'intégrale des fonctions indicatrices et du passage à la limite via les fonctions étagées :

1. Soit $f = \mathbf{I}_A$, où $A \in \mathcal{A}$. Alors :

$$\int f \, d\mu = \int_\omega f(\omega) d\mu(\omega) = \mu(A).$$

2. Soit $f = \mathbf{I}_A$, où $A \in \mathcal{A}$ et soit $B \in \mathcal{A}$. Alors :
$$\int_B f\, d\mu = \int_B f(\omega) d\mu(\omega) = \int \mathbf{I}_B \mu(A)(\omega) f(\omega) d\mu(\omega) = \mu(A \cap B).$$
3. Soit f une fonction étagée positive telle que $f = \sum_{i=1}^{n} \alpha_i \mathbf{I}_{A_i}$, où les $A_i \in \mathcal{A}$ et $\alpha_i > 0$ et soit $B \in \mathcal{A}$. Alors :
$$\int_B f\, d\mu = \int_B f(\omega) d\mu(\omega) = \int \mathbf{I}_B(\omega) f(\omega) d\mu(\omega) = \sum_{i=1}^{n} \alpha_i \mu(A_i \cap B).$$

Définition 20 *Soit f une fonction $\mathcal{A}$-mesurable positive et soit $B \in \mathcal{A}$. Alors l'intégrale de Lebesgue de f par rapport à μ sur B est :*
$$\int_B f\, d\mu = \int \mathbf{I}_B(\omega) f(\omega) d\mu(\omega) = \sup \left\{ \int_B g\, d\mu, \text{ pour } g \text{ étagée positive avec } g \leq f \right\}.$$

Propriété 6 *Soit f une fonction $\mathcal{A}$-mesurable positive et soit A et $B \in \mathcal{A}$. Alors :*
1. *Pour $c \geq 0$, $\int_B cf\, d\mu = c \int_B f\, d\mu$.*
2. *Si $A \subset B$, alors $\int_A f\, d\mu \leq \int_B f\, d\mu$.*
3. *Si g est $\mathcal{A}$-mesurable telle que $0 \leq f \leq g$ alors $0 \leq \int_B f\, d\mu \leq \int_B g\, d\mu$.*
4. *Si $\mu(B) = 0$ alors $\int_B f\, d\mu = 0$.*

Théorème 5 (Théorème de convergence monotone (Beppo-Lévi)) *Si $(f_n)_n$ est une suite croissante de fonctions mesurables positives convergeant simplement vers f sur Ω, alors :*
$$\lim_{n\to\infty} \left(\int f_n d\mu \right) = \int f\, d\mu = \int \lim_{n\to\infty} f_n d\mu.$$

En particulier pour les séries de fonctions mesurables positives, on peut toujours appliquer le Théorème de convergence monotone et donc inverser la somme et l'intégrale.

Lemme 1 (Lemme de Fatou) *Soit $(f_n)_n$ est une suite de fonctions mesurables positives alors :*
$$\int \left(\liminf_{n\to\infty} f_n \right) d\mu \leq \liminf_{n\to\infty} \int f_n d\mu.$$

• **Intégrale de Lebesgue d'une fonction réelle et propriétés**

Définition 21 *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espace mesuré,* $B \in \mathcal{A}$ *et soit* f *une fonction mesurable à valeurs réelles telle que* $f = f^+ - f^-$ *avec* $f^+ = \max(f, 0)$ *et* $f^- = \max(-f, 0)$*. On dit que* f *est* *μ-intégrable* *sur* B *si* $\int_B |f|\, d\mu < +\infty$*. On a alors*

$$\int_B f\, d\mu = \int_B f^+\, d\mu - \int_B f^-\, d\mu.$$

Lorsque f est μ-intégrable sur Ω, soit $\int |f|\, d\mu < +\infty$, on note $f \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ (on dit que f est $\mathcal{L}^1$). On a les propriétés de linéarité et de monotonie de l'intégrale : soient f et $g \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$. Alors :

1. $\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu$ pour $(\alpha, \beta) \in \mathbf{R}^2$.
2. Si $f \leq g$ alors $\int f\, d\mu \leq \int g\, d\mu$.

Théorème 6 (Théorème de convergence dominée de Lebesgue) *Soit* $(f_n)_n$ *une suite de fonctions de* $\mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ *telles que pour tout* $n \in \mathbf{N}$, $|f_n| \leq g$ *avec* $g \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$*. Si on suppose que* (f_n) *converge simplement vers* f *sur* Ω *alors :*

$$\lim_{n\to\infty} \int f_n d\mu = \int f\, d\mu.$$

On peut étendre le Théorème de Lebesgue dans le cas où $(f_n)_n$ converge presque partout vers f.

Théorème 7 (Inégalité de Jensen) *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé, soit* $\phi : \mathbf{R} \mapsto \mathbf{R}$ *une fonction convexe et soit* $f : \Omega \mapsto \mathbf{R}$ *mesurable telle que* $\phi(f)$ *soit une fonction intégrable par rapport à* P*. Alors :*

$$\phi\left(\int f\, dP\right) \leq \int \phi(f)\, dP.$$

• **Mesures induites et densités**

Théorème 8 (Théorème du Transport) *Soit f une fonction mesurable de $(\Omega, \mathcal{A}, \mu)$ dans $(\Omega', \mathcal{A}')$ telle que μ_f soit la mesure induite par f (donc $\mu_f(A') = \mu(f^{-1}(A'))$ pour $A' \in \mathcal{A}'$) et soit ϕ une fonction mesurable de $(\Omega', \mathcal{A}')$ dans $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. Alors, si $\phi \circ f \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$,*

$$\int_{\Omega'} \phi \, d\mu_f = \int_{\Omega} \phi \circ f \, d\mu.$$

Définition 22 *Soit μ et ν deux mesures sur $(\Omega, \mathcal{A})$. On dit que μ domine ν (ou ν est dominée par μ) et que ν est absolument continue par rapport à μ lorsque pour tout $A \in \mathcal{A}$, $\mu(A) = 0 \Longrightarrow \nu(A) = 0$.*

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et f une fonction définie sur $(\Omega, \mathcal{A})$ mesurable et positive. On suppose que pour $A \in \mathcal{A}$, $\nu(A) = \int_A f \, d\mu$. Alors, ν est une mesure sur $(\Omega, \mathcal{A})$, dominée par μ. De plus, pour toute fonction g définie sur $(\Omega, \mathcal{A})$ mesurable et positive,

$$\int g d\nu = \int g.f d\mu.$$

Enfin, g est ν intégrable si et seulement si $g.f$ est μ intégrable.

Définition 23 *On dit que μ mesure sur $(\Omega, \mathcal{A})$ est σ-finie lorsqu'il existe une famille $(A_i)_{i \in I}$, avec I dénombrable, d'ensembles de $\mathcal{A}$ telle que $\bigcup A_i = \Omega$ et $\mu(A_i) < +\infty$ pour tout $i \in I$.*

Théorème 9 (Théorème de Radon-Nikodym) *On suppose que μ et ν sont deux mesures σ-finies sur $(\Omega, \mathcal{A})$ telles que μ domine ν. Alors il existe une fonction f définie sur $(\Omega, \mathcal{A})$ mesurable et positive, appelée densité de ν par rapport à μ, telle que pour tout $A \in \mathcal{A}$, $\nu(A) = \displaystyle\int_A f \, d\mu$.*

Théorème 10 (Théorème de Fubini) *Soit $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ et $\mu = \mu_1 \otimes \mu_2$ (mesures σ finies), où $(\Omega_1, \mathcal{A}_1, \mu_1)$ et $(\Omega_2, \mathcal{A}_2, \mu_2)$ sont des espaces mesurés. Soit une fonction $f : \Omega \mapsto \mathbf{R}$, $\mathcal{A}$-mesurable et μ-intégrable. alors :*

$$\int_\Omega f d\mu = \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \right) d\mu_1(\omega_1) = \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \right) d\mu_2(\omega_2).$$

• Espaces $\mathcal{L}^p$ et L^p pour $0 < p \leq \infty$

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. On appelle espace $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$, où $p > 0$, l'ensemble des fonctions $f : \Omega \mapsto \mathbf{R}$, mesurables et telles que $\int |f|^p d\mu < +\infty$. Pour $f \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$, où $p > 0$, on note $\|f\|_p = \left(\int |f|^p d\mu\right)^{1/p}$. Pour $p = +\infty$ on définit $\|f\|_\infty = \mathrm{essup}_\Omega |f|$ le supremum essentiel de f tel que $\|f\|_\infty \geq |f(\omega)|$ pour presque tout $\omega \in \Omega$. Alors pour tout f mesurable, $f \in \mathcal{L}^\infty(\Omega, \mathcal{A}, \mu)$ lorsque $\|f\|_\infty < \infty$.

Propriété 7 (Inégalité de Hölder) *Soit* $p \geq 1$ *et* $q \geq 1$ *tels que* $\dfrac{1}{p} + \dfrac{1}{q} = 1$, *et soit* $f \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ *et* $g \in \mathcal{L}^q(\Omega, \mathcal{A}, \mu)$. *Alors,* $f\, g \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ *et*

$$\|f\, g\|_1 \leq \|f\|_p . \|g\|_q .$$

Propriété 8 (Inégalité de Minkowski) *Soit* $p \geq 1$ *et soit* f *et* $g \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$. *Alors,* $f + g \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ *et*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p .$$

Pour $p \geq 1$, $\|.\|_p$ définit ainsi sur une semi-norme sur $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$. Pour obtenir une norme, on se place dans l'espace $L^p(\Omega, \mathcal{A}, \mu)$ obtenu en “quotientant” $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ par la relation d'équivalence $f = g$ μ-presque partout (c'est-à-dire que dans $L^p(\Omega, \mathcal{A}, \mu)$ on dira que $f = g$ lorsque $f = g$ μ-presque partout).

Pour f et $g \in L^2(\Omega, \mathcal{A}, \mu)$, on définit le produit scalaire $< f, g >= \int f.g\, d\mu$. On muni ainsi $L^2(\Omega, \mathcal{A}, \mu)$ d'une structure d'espace de Hilbert. On dira que f est orthogonale à g lorsque $< f, g >= 0$. Si A est un sous-espace vectoriel fermé de $L^2(\Omega, \mathcal{A}, \mu)$ (par exemple un sous-espace de dimension finie), alors pour tout $f \in L^2(\Omega, \mathcal{A}, \mu)$, il existe un unique projeté orthogonal de f sur A, noté $P_A(f)$, qui vérifie $P_A(f) = \operatorname{Arginf}_{g \in A} \|g - f\|_2$.

2.2 Applications de la théorie de la mesure et de l'intégration en Probabilités

2.2.1 Espérance de variables aléatoires

Soit X une variable aléatoire (v.a.) sur $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Pour tout $0 < p < \infty$, $X \in L^p(\Omega, \mathcal{A}, P)$ lorsque $\int |X|^p dP < \infty$ c'est à dire lorsque $\int |x|^p dP_X(x)$ est fini. Alors si $X \in L^1(\Omega, \mathcal{A}, P)$, on définit l'espérance de X par le nombre $\mathbb{E}X = \int X dP = \int x dP_X(x)$. Plus généralement, si $\phi : \mathbf{R} \mapsto \mathbf{R}$ est borélienne et si $\phi(X) \in L^1(\Omega, \mathcal{A}, P)$, on définit l'espérance de $\phi(X)$ par $\mathbb{E}(\phi(X)) = \int \phi(X)\, dP = \int \phi(x)\, dP_X(x)$.

Si X est une variable aléatoire sur $(\Omega, \mathcal{A}, P)$, si $\phi : \mathbf{R} \mapsto \mathbf{R}$ est borélienne telle que $\phi(X) \in L^1(\Omega, \mathcal{A}, P)$, et si P_X est la mesure de probabilité de X alors :

$$\mathbb{E}(\phi(X)) = \int_{\mathbf{R}} \phi(x)\, dP_X(x).$$

Si P_X est absolument continue par rapport à la mesure de Lebesgue (donc X est une v.a. dite continue) alors elle admet une densité f_X, alors $\mathbb{E}(\phi(X)) = \int_{\mathbf{R}} \phi(x) f_X(x) dx$. Si P_X est absolument continue par rapport à la mesure de comptage sur $\mathbf{N}$ (donc X est une v.a. dite discrète), de densité p_X, alors $\mathbb{E}(\phi(X)) = \sum_{k=0}^{\infty} p_X(x_k)\, \phi(x_k)$.

Propriété 9

1. *Soit X et Y des variables aléatoires telles que X et $Y \in L^1(\Omega, \mathcal{A}, P)$. Alors pour tout $(a,b) \in \mathbf{R}^2$, $aX + bY \in L^1(\Omega, \mathcal{A}, P)$ et*
$$\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y.$$
2. *Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, P)$, et soit $A \in \mathcal{A}$. Alors $\mathbb{E}(\mathbf{I}_A(X)) = P(X \in A)$.*
3. *Soit X et Y des variables aléatoires telles que $X \in L^p(\Omega, \mathcal{A}, P)$ et $Y \in L^q(\Omega, \mathcal{A}, P)$ avec $1/p + 1/q = 1$ et $p \geq 1$, $q \geq 1$. Alors $XY \in L^1(\Omega, \mathcal{A}, P)$ et*
$$\mathbb{E}|XY| \leq (\mathbb{E}|X|^p)^{1/p} (\mathbb{E}|Y|^q)^{1/q}.$$
4. *Soit X et Y des variables aléatoires telles que X et $Y \in L^p(\Omega, \mathcal{A}, P)$, avec $p \geq 1$. Alors $X + Y \in L^p(\Omega, \mathcal{A}, P)$ et*
$$(\mathbb{E}|X+Y|^p)^{1/p} \leq (\mathbb{E}|X|^p)^{1/p} + (\mathbb{E}|Y|^p)^{1/p}.$$
5. *Soit X une variable aléatoire telle que $X \in L^p(\Omega, \mathcal{A}, P)$ pour $p > 0$. Alors pour tout $0 < r \leq p$, $X \in L^r(\Omega, \mathcal{A}, P)$ et*
$$(\mathbb{E}|X|^r)^{1/r} \leq (\mathbb{E}|X|^p)^{1/p}.$$
6. *Si X est une variable aléatoire sur $(\Omega, \mathcal{A}, P)$, si $\phi : \mathbb{R} \mapsto \mathbb{R}$ est une fonction borélienne convexe telle que X et $\phi(X) \in L^1(\Omega, \mathcal{A}, P)$, alors*
$$\mathbb{E}(\phi(X)) \geq \phi(\mathbb{E}X).$$

Définition 24 *Pour X et Y des variables aléatoires telles que X et $Y \in L^2(\Omega, \mathcal{A}, P)$, on définit la covariance de X et Y par*
$$Cov(X,Y) = \mathbb{E}\left[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\right];$$
On appelle variance de X, $Var(X) = Cov(X,X) = \mathbb{E}\left[(X - \mathbb{E}X)^2\right]$. D'après la formule de Huygens, $Var(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$.

Sur $L^2(\Omega, \mathcal{A}, P)$, $\mathrm{Cov}\,(.,.)$ définit un produit scalaire. De plus
$$|\mathrm{Cov}\,(X,Y)|^2 \leq \mathrm{Var}\,(X)\mathrm{Var}\,(Y).$$

Il est souvent utile d'utiliser une espérance pour contrôler une probabilité via le résultat :

Proposition 10 *Soit g une fonction paire, positive et croissante sur* $[0,+\infty[$. *Alors pour toute v.a. X et $\epsilon > 0$*

$$P(|X| > \epsilon) \leq \frac{\mathbb{E}g(X)}{g(\epsilon)}.$$

Les inégalités suivantes sont des conséquences immédiates de la Proposition 10 :

- Inégalité de Markov : Soit X une v.a. et $\epsilon > 0$. Alors pour $r > 0$

$$P(|X| > \epsilon) \leq \frac{\mathbb{E}|X|^r}{\epsilon^r}.$$

- Inégalité de Chebychev : Soit X une v.a. d'espérance finie μ et de variance finie σ^2, et $\epsilon > 0$. Alors

$$P(|X-\mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

2.2.2 Fonction de répartition et quantiles d'une loi de probabilité

Il y a une correspondance bijective entre la connaissance de P_X et celle de la fonction de répartition $F_X(x) = P_X(]-\infty, x])$. La fonction de répartition permet également de définir les quantiles qui sont essentiels à la construction d'intervalles de confiance et de test.

Soit $\alpha \in [0,1]$. Des propriétés de la fonction de répartition, on en déduit qu'il existe $x_\alpha \in \mathbf{R}$, tel que :

$$\lim_{x \to x_\alpha^-} F_X(x) \leq \alpha \leq F_X(x_\alpha). \tag{2.1}$$

Soit $I_\alpha = \{x_\alpha \in \mathbf{R} \text{ tel que } x_\alpha \text{ vérifie (2.1)}\}$. On appelle quantile (ou fractile, ou percentile en anglais) d'ordre α de la loi P_X, noté q_α, le milieu de l'intervalle I_α. Evidemment, lorsque X admet une distribution absolument continue par rapport à la mesure de Lebesgue, $q_\alpha = F_X^{-1}(\alpha)$, où F_X^{-1} désigne la fonction réciproque de F_X.

Trois cas particuliers sont à connaître :

1/ pour $\alpha = 0.5$, $q_{0.5}$ est appelé la médiane de P_X ;

2/ pour $\alpha = 0.25$ et $\alpha = 0.75$ (respectivement), $q_{0.25}$ et $q_{0.25}$ sont appelés premier et troisième quartile (respectivement) de P_X.

3/ pour $\alpha = 0.1, \ldots, 0.9$, on parlera de décile de P_X.

2.2.3 Indépendance

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

- Soit $(A_i)_{i\in I}$ une famille dénombrable d'événements de $\mathcal{A}$. On dit que les événements $(A_i)_{i\in I}$ sont indépendants si et seulement si pour tous les sous-ensembles finis $K \subset I$,

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

- Soit $(\mathcal{A}_i)_{i\in I}$ une famille de sous-tribus de $\mathcal{A}$ (donc pour tout $i \in I$, $\mathcal{A}_i \subset \mathcal{A}$). On dit que les tribus $(\mathcal{A}_i)_{i\in I}$ sont indépendantes si et seulement si pour tous les sous-ensembles finis $K \subset I$, et pour tous les événements $A_k \in \mathcal{A}_k$ avec $k \in K$, les A_k sont indépendants.
- Soit $(X_i)_{i\in I}$ des variables aléatoires sur $(\Omega, \mathcal{A})$ à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On dit que les v.a. $(X_i)_{i\in I}$ sont indépendantes si et seulement si les tribus engendrées $(X_i^{-1}(\mathcal{B}(\mathbb{R})))_{i\in I}$ sont indépendantes.

Proposition 11 *Si $(X_1, \cdots, X_n)$ sont des variables aléatoires sur $(\Omega, \mathcal{A}, P)$. Alors les (X_i) sont indépendantes si et seulement si $P_{(X_1,\cdots,X_n)} = \bigotimes_{i=1}^{n} P_{X_i}$.*

Proposition 12 *Si $(X_i)_{i\in I}$ sont des variables aléatoires indépendantes sur $(\Omega, \mathcal{A}, P)$. Alors les (X_i) sont indépendantes si et seulement si pour tout $J \subset I$, J fini, pour toutes fonctions boréliennes $(g_j)_{j\in J}$ telles que $g_j(X_j)$ soit intégrable, alors*

$$\mathbb{E}\left(\prod_{j\in J} g_j(X_j)\right) = \prod_{j\in J} \mathbb{E}(g_j(X_j)).$$

Lemme 2 (Lemme de Borel-Cantelli) *Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements sur $(\Omega, \mathcal{A}, P)$.*

1. *Si $\sum P(A_n) < +\infty$ alors $P(\limsup A_n) = 0$.*
2. *Si les (A_n) sont indépendants, $\sum P(A_n) = +\infty$ implique que $P(\limsup A_n) = 1$.*

2.2.4 Principales lois de probabilités

On rappelle que la fonction Gamma est telle que $\Gamma(a) = \int_0^\infty x^{a-1} \cdot e^{-x}$ pour $a \geq 0$.

• **Loi uniforme discrète**

C'est la loi de probabilité discrète à valeurs dans $\{x_1, \ldots, x_n\}$ telle que

$$P(X = x_i) = \frac{1}{n}.$$

On alors : $\mathbb{E}X = \frac{1}{n}(x_1 + \ldots + x_n)$ et $Var(X) = \frac{1}{n}(x_1^2 + \ldots + x_n^2) - (\mathbb{E}X)^2$.

• **Loi de Bernoulli**

C'est la loi de probabilité discrète notée $\mathcal{B}(p)$ à valeurs dans $\{0, 1\}$ telle que

$$P(X = 1) = p \quad \text{et} \quad P(X = 0) = 1 - p.$$

On alors : $\mathbb{E}X = p$ et $Var(X) = p(1 - p)$.

• **Loi Binomiale**

C'est la loi de probabilité discrète notée $\mathcal{B}(n, p)$ à valeurs dans $\{0, 1, \ldots, n\}$ telle que

$$P(X = k) = C_n^k \cdot p^k \cdot (1 - p)^{n-k} \quad \text{pour } k \in \{0, 1, \ldots, n\}.$$

On alors : $X = X_1 + \cdots + X_n$, où (X_i) est une suite de v.a. iid de loi $\mathcal{B}(p)$, d'où $\mathbb{E}X \cdot p$ et $Var(X) \cdot p(1 - p)$.

• **Loi de Poisson**

C'est la loi de probabilité discrète notée $\mathcal{P}(\theta)$ à valeurs dans $\mathbb{N}$ telle que

$$P(X = k) = \frac{\theta^k}{k!} \cdot e^{-\theta} \quad \text{pour } k \in \mathbb{N}.$$

On alors $\mathbb{E}X = \theta$ et $Var(X) = \theta$.

• **Loi Uniforme sur** $[a, b]$

Cette loi est généralement notée $\mathcal{U}([a, b])$, où $-\infty < a < b < \infty$. C'est la loi de probabilité à valeurs dans $[a, b]$ de densité par rapport à la mesure de Lebesgue :

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

On a alors $\mathbb{E}X = \frac{b + a}{2}$ et $Var(X) = \frac{(b - a)^2}{12}$.

• **Loi Gamma**

Cette loi est généralement notée $\gamma(p,\theta)$, où $p>0$ et $\theta>0$. C'est la loi de probabilité à valeurs dans $\mathbb{R}_+$ de densité par rapport à la mesure de Lebesgue :

$$f_X(x)=\frac{\theta^p}{\Gamma(p)}\cdot e^{-\theta\cdot x}\cdot x^{p-1}\cdot \mathbf{1}_{x\in\mathbb{R}_+}.$$

On a alors $\mathbb{E}X=\dfrac{p}{\theta}$ et $Var(X)=\dfrac{p}{\theta^2}$.
Si $X\sim\gamma(p,\theta)$ et $Y\sim\gamma(q,\theta)$ avec X et Y indépendantes et $p>0$ et $q>0$, alors $X+Y\sim\gamma(p+q,\theta)$.

• **Loi Exponentielle**

Pour $p=1$, la loi $\gamma(p,\theta)$ est appelée loi exponentielle de paramètre $\theta>0$ et est notée $\mathcal{E}(\theta)$.

• **Loi Béta**

Cette loi est généralement notée $\beta(p,q)$, où $p>0$ et $q>0$. C'est la loi de probabilité à valeurs dans $[0,1]$ de densité par rapport à la mesure de Lebesgue :

$$f_X(x)=\frac{x^{p-1}(1-x)^{q-1}}{B(p,q)}\cdot\mathbf{1}_{x\in[0,1]},\quad \text{où}\quad B(p,q)=\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

On a alors $\mathbb{E}X=\dfrac{B(p+1,q)}{B(p,q)}$ et $Var(X)=\dfrac{p\cdot q}{(p+q)^2(p+q+1)}$.
Si $X\sim\gamma(p,\theta)$ et $Y\sim\gamma(q,\theta)$ avec X et Y indépendantes et $p>0$ et $q>0$, alors $\dfrac{X}{X+Y}\sim\beta(p,q)$.

• **Loi Normale (ou Gaussienne) centrée réduite**

Cette loi est généralement notée $\mathcal{N}(0,1)$. C'est la loi de probabilité à valeurs dans $\mathbb{R}$ de densité par rapport à la mesure de Lebesgue :

$$f_X(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right).$$

On a :

$$\mathbb{E}(X)=0\quad\text{et}\quad \mathrm{Var}\,(X)=1.$$

• **Loi Normale (ou Gaussienne) de moyenne m et de variance σ^2**

Si Z suit la loi $\mathcal{N}(0,1)$, $X = m + \sigma Z$ suit par définition la loi $\mathcal{N}(m, \sigma^2)$, loi normale d'espérance m et de variance σ^2. La densité de X est donnée par :

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

A partir de la loi gaussienne, on peut en déduire les lois suivantes.

• **Loi du χ^2 à n degrés de libertés**

Soit $X_1, \cdots, X_n$, n variables aléatoires indépendantes de loi $\mathcal{N}(0,1)$, alors

$$S = X_1^2 + \cdots + X_n^2$$

suit une loi du χ^2 à n degrés de libertés, loi notée $\chi^2(n)$. Cette loi est à valeurs dans $\mathbb{R}_+$, d'espérance n et de variance $2n$. C'est aussi la loi Gamma $\gamma(n/2, 1/2)$, c'est-à-dire que $X \sim \chi^2(n)$ admet pour densité par rapport à la mesure de Lebesgue :

$$f_X(x) = \frac{1}{2^{n/2} \cdot \Gamma(n/2)} x^{n/2-1} \exp\left(-\frac{x}{2}\right) \cdot \mathbf{1}_{\{x \geq 0\}},$$

• **Loi de Student à n degrés de libertés**

La loi de Student à n degrés de liberté, notée $T(n)$, est la loi du quotient

$$T = \frac{N}{\sqrt{S/n}}$$

où N suit une loi $\mathcal{N}(0,1)$ et S suit une loi $\chi^2(n)$, N et S étant deux variables aléatoires indépendantes. Il est également possible de déterminer la densité d'une telle loi par rapport à la mesure de Lebesgue, à savoir,

$$f_X(x) = \frac{1}{\sqrt{n} \cdot B(1/2, n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2},$$

où la fonction Beta est telle que $B(a,b) = \dfrac{\Gamma(a) \cdot \Gamma(b)}{\Gamma(a+b)}$ pour $a > 0$ et $b > 0$.

Remarque : Par la loi des grands nombres, plus n est grand, plus S est proche de son espérance qui vaut n. Le dénominateur est donc proche de 1. Il s'ensuit que la loi $T(n)$ est d'autant plus proche d'une loi normale que n est grand.

• **Loi de Fisher à n_1 et n_2 degrés de liberté**

Soit S_1 et S_2 deux variables aléatoires indépendantes de loi respectives $\chi^2(n_1)$ et $\chi^2(n_2)$. Alors par définition :

$$F = \frac{S_1/n_1}{S_2/n_2}$$

suit une loi de Fisher à n_1 et n_2 degrés de liberté, notée $F(n_1, n_2)$. Si $F \sim F(n_1, n_2)$, alors $\mathbb{E}(F) = \dfrac{n_2}{n_2 - 2}$ lorsque $n_2 > 2$ et $\mathrm{Var}\,(F) = \dfrac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 4)(n_2 - 2)^2}$ lorsque $n_2 > 4$. De plus, si T suit une loi de Student $T(n)$, alors T^2 suit une loi de Fisher $F(1, n)$

Remarque : Par les mêmes considérations que précédemment, la loi F est d'autant plus proche de 1 que les degrés de liberté n_1 et n_2 sont grands.

2.2.5 Caractérisations d'une densité

• **Méthode de la fonction muette**

Soit l'ensemble C_c des fonctions continues à support compact. Alors on vérifie que C_c est dense dans $L^p(\Omega, \mathcal{A}, P)$. On en déduit

Proposition 13 *La loi P de X est caractérisée par l'ensemble $\{\mathbb{E}(\phi(X)) \ / \ \phi \in C_c\}$.*

On en déduit la méthode de la fonction muette : si X admet une densité f_X alors on identifie pour tout $\phi \in C_c$ la densité f_Y de $Y = g(X)$ dans l'expression de $\mathbb{E}(\phi(Y))$ en faisant le changement de variable $y = g(x)$:

$$\mathbb{E}(\phi(g(X)) = \int_{\mathbb{R}} \phi(g(x)) f_X(x) dx = \int_{g(\mathbb{R})} \phi(y) f_X(g^{-1}(y)) |{g^{-1}}'(y)| dy,$$

valable lorsque g est un C^1-difféomorphisme. On obtient ainsi

$$f_Y(y) = f_X(g^{-1}(y))|{g^{-1}}'(y)| \qquad \text{pour } y \in f(\mathbb{R}).$$

• **Support, modes et symétrie**

Définition 25 *Le support S d'une densité f (≥ 0) est l'ensemble des réelles où la densité est non nulle :*

$$S = \{x \in \mathbb{R} \ / \ f(x) > 0\}.$$

Définition 26 *L'ensemble des modes d'une densité f vaut*

$$\arg\max_{x \in \mathbb{R}} f(x) = \{x \in \mathbb{R} \ / \ \forall y \in \mathbb{R}, f(x) \geq f(y)\}.$$

Cet ensemble peut être vide. Si f est telle que son support S soit un compact ou $\mathbb{R}$ en entier, alors l'ensemble des modes est non vide. Si f n'a qu'un mode, i.e. $\exists! x_0 \in \mathbb{R}$ tel que $\forall y \neq x_0$ on ait $f(x_0) > f(y)$, alors la loi correspondante est unimodale. Si f a plusieurs modes, la loi est multimodale.

Exemple 1 *Les lois Gamma (avec $p > 1$), Béta (avec $p > 1$, $q > 1$), Normale et Student sont unimodales. La loi Uniforme sur $[a, b]$ est multimodale. La loi Exponentielle n'a pas de mode.*

Définition 27 *Une variable X est symétrique lorsque X et $-X$ ont même loi.*

Lorsque X admet une densité f_X, alors f_X est symétrique par rapport à l'axe des ordonnées, i.e. c'est une fonction paire, ssi X est symétrique. Si X est symétrique, alors sa fonction de répartition satisfait $F_X(x) = 1 - F_X(-x)$ pour tout $x \in \mathbb{R}$ et ses quantiles satisfont $q_\alpha = -q_{1-\alpha}$ pour tout $0 \leq \alpha \leq 1$.

2.2.6 Vecteurs aléatoires

Lois isotropes

Définition 28 *On dit que X est un vecteur aléatoire sur $(\Omega, \mathcal{A}, P)$, un espace probabilisé, si X est une fonction mesurable de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.*

Lorsque X un vecteur aléatoire sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbb{R}^d$, la loi (ou mesure) de probabilité de X, P_X, est définie de façon univoque à partir de la fonction de répartition de X, telle que pour $x = (x_1, \cdots, x_d)$,

$$F_X(x) = P_X(\prod_{i=1}^{d}]-\infty, x_i]) = P(X \in \prod_{i=1}^{d}]-\infty, x_i]).$$

Propriété 10 *Soit X un vecteur aléatoire sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbb{R}^d$. On suppose que $X = (X_1, \cdots, X_d)$. Alors les X_i sont des variables aléatoires sur $(\Omega, \mathcal{A}, P)$, de fonction de répartition*

$$F_{X_i}(x_i) = \lim_{\substack{x_j \to +\infty \\ j \neq i}} F_X(x_1, \cdots, x_i, \cdots, x_d).$$

Les mesures de probabilités P_{X_i} déterminées de façon univoque à partir des F_{X_i} sont appelées lois marginales de X.

On se place maintenant dans la base canonique orthonormale de $\mathbb{R}^d$. Si Z est un vecteur aléatoire à valeurs sur $\mathbb{R}^d$, on définit $\mathbb{E}(Z)$, le vecteur dont les coordonnées sont les espérances des coordonnées de Z. Ainsi, si dans la base canonique de $\mathbb{R}^d$, $Z = (Z_1, \cdots, Z_d)^T$,

$$\mathbb{E}(Z) = \mathbb{E}\begin{pmatrix} Z_1 \\ \vdots \\ Z_d \end{pmatrix} = \begin{pmatrix} \mathbb{E}(Z_1) \\ \vdots \\ \mathbb{E}(Z_d) \end{pmatrix}.$$

De la même manière, on définira l'espérance d'une matrice dont les coordonnées sont des variables aléatoires par la matrice dont les coordonnées sont les espérances de chacune de ces variables aléatoires.

Ceci nous permet de définir la <u>matrice de variance-covariance</u> de Z de la manière suivante :

$$\mathrm{Var}\,(Z) = \mathbb{E}\left[(Z - \mathbb{E}(Z))(Z - \mathbb{E}(Z))^T\right]$$

donc si $Z = (Z_1, \cdots, Z_d)^T$,

$$\mathrm{Var}\begin{pmatrix} Z_1 \\ \vdots \\ Z_d \end{pmatrix} = \begin{pmatrix} \mathrm{Var}\,(Z_1) & \mathrm{Cov}(Z_1, Z_2) & \cdots & \mathrm{Cov}(Z_1, Z_d) \\ \mathrm{Cov}(Z_1, Z_2) & \mathrm{Var}\,(Z_2) & \cdots & \mathrm{Cov}(Z_2, Z_d) \\ \vdots & \vdots & \cdots & \vdots \\ \mathrm{Cov}(Z_1, Z_d) & \mathrm{Cov}(Z_2, Z_d) & \cdots & \mathrm{Var}\,(Z_d) \end{pmatrix}$$

matrice (d, d) dont les éléments diagonaux sont les variances et les éléments non diagonaux sont les covariances des coordonnées de Z (remarquons que la variance de Z_1 est aussi la covariance de Z_1 et de Z_1).

On vérifie également le résultat suivant : si C est une matrice (p, d) à coordonnées constituées de réels constants et si Z est un vecteur aléatoire à valeurs dans $\mathbb{R}^d$, alors CZ est un vecteur de taille p de matrice de variance-covariance

$$\mathrm{Var}\,(CZ) = C\,\mathrm{Var}\,(Z)C^T.$$

En particulier, si p vaut 1, alors $C = h^T$ où h est un vecteur de taille d, et :

$$\mathrm{Var}\,(h^T Z) = h^T\,\mathrm{Var}\,(Z)h.$$

Notez que cette dernière quantité est un scalaire.

D'après ce qui précède, la loi d'un vecteur gaussien X peut dépendre de la BON dans lequel on l'exprime (autrement dit PX n'a pas la même loi que X pour P orthogonal). Le choix de la BON étant arbitraire, pour éviter cette difficulté (qui n'apparaît que dans le cas multidimensionnel) nous étudions

Définition 29 *Un vecteur aléatoire suit une loi isotrope si sa loi est la même quelque soit la BON choisie.*

Vecteurs gaussiens et théorème de Cochran

Définition 30 *Un vecteur gaussien Y à valeurs dans $\mathbb{R}^d$ (non dégénéré), d'espérance $\mu \in \mathbb{R}^d$ et de matrice de variance-covariance Σ^2 (symétrique définie positive) suit la loi notée $\mathcal{N}_d(\mu, \Sigma^2)$ et admet la densité*

$$f_Y(y) = \frac{(2\pi)^{-d/2}}{\det(\Sigma)} \exp\left(-\frac{1}{2}(y-\mu)^T \left(\Sigma^2\right)^{-1} (y-\mu)\right),$$

pour $y \in \mathbb{R}^d$, et avec $\det(\Sigma)$ le déterminant de la matrice $\Sigma = \sqrt{\Sigma^2}$.

Remarque 1 *L'espérance et la variance définissent complètement la loi de probabilité d'un vecteur gaussien.*

A partir des propriétés générales sur les vecteurs aléatoires, on obtient le fait que :

Propriété 11 *Soit $X \sim \mathcal{N}_d(\mu, \Sigma^2)$. Soit A une matrice réelle de taille (p, d) où $p \in \mathbb{N}^*$ et $B \in \mathbb{R}^p$. Alors $A\,X + B$ est un vecteur gaussien tel que :*

$$A\,X + B \sim \mathcal{N}_p(A\,\mu + B,\ A\,\Sigma\,A^T)$$

En particulier, si X est un vecteur gaussien d'espérance μ et de matrice de variance Σ^2 et si h un vecteur de $\mathbb{R}^d$, alors $h^T X$ est une combinaison linéaire des coordonnées de X et $h^T X$ suit la loi unidimensionnelle $\mathcal{N}(h^T \mu, h^T \Sigma^2 h)$.
Nous pouvons caractériser toutes les lois gaussiennes isotropes :

Proposition 14 *Les lois gaussiennes isotropes sont de la forme $\mathcal{N}_d(0, \sigma^2 I_d)$ avec $\sigma^2 \geq 0$.*

Démonstration : $\Longleftarrow$: soit $X \sim \mathcal{N}_d(0, \sigma^2 I_d)$ et P une matrice de changement de BON. Alors P est une matrice orthogonale vérifiant $P^T P = I_q$ et $PX \sim \mathcal{N}_d(P0, P^T \sigma^2 I_d P) = \mathcal{N}_d(O, \sigma^2 I_q)$.
$\Longrightarrow$: Réciproquement, pour que $\mathcal{N}_d(\mu, \Sigma^2)$ soit isotrope il suffit que $U\mu = \mu$ et $U^T \Sigma^2 U = \Sigma^2$ pour toute matrice orthogonale U. D'où $\mu = 0$ et Σ^2 étant symétrique il existe P orthogonale telle que $P^T \Sigma^2 P = D$ donc $\Sigma^2 = D$ est diagonale.

Si ses éléments diagonaux $d_{ii} \neq d_{jj}$ pour $i \neq j$, alors considérant le changement de BON $(e_1, \ldots, e_d)$ vers la BON $((e'_1, \ldots, e'_d))$ où $e'_k = e_k$ et $(e'_i, e'_j) = (e_j, e_i)$ on obtient une contradiction. $\square$

Remarque 2 *Soit $Y_1, \ldots, Y_d$ des variables aléatoires indépendantes de même loi $\mathcal{N}(0, \sigma^2)$, indépendantes (ce qui, dans le cas gaussien, est équivalent à $Cov\,(Y_i, Y_j) = 0$ pour $i \neq j$), alors le vecteur $Y = (Y_1, \ldots, Y_d)^T \sim \mathcal{N}_d(0, \sigma^2 I_d)$. On obtient ainsi tous les vecteurs gaussiens isotropes.*

Il est possible de déterminer la loi de la norme d'un vecteur gaussien isotrope grâce à la proposition suivante :

Théorème 11 (Théorème de Cochran) *Soit E_1 et E_2, deux sous-espaces vectoriels orthogonaux de $E = \mathbb{R}^d$ de dimensions respectives k_1 et k_2 et soit Y un vecteur aléatoire de $\mathbb{R}^d$ de loi normale centrée isotrope de variance σ^2. Alors $P_{E_1}(Y)$ et $P_{E_2}(Y)$ sont deux variables aléatoires gaussienne centrées indépendantes et $\|P_{E_1}(Y)\|^2/\sigma^2$ (resp. $\|P_{E_2}(Y)\|^2/\sigma^2$) est une loi $\chi^2(k_1)$ (resp. $\chi^2(k_2)$). Ce théorème se généralise naturellement pour $2 < m \leq d$ sous-espaces vectoriels orthogonaux $(E_i)_{1 \leq i \leq m}$ de $E = \mathbb{R}^d$.*

Démonstration : Soit $(e_1, \cdots, e_{k_1})$ et $(e_{k_1+1}, \cdots, e_{k_1+k_2})$ deux BON de E_1 et E_2 (respectivement). L'ensemble de ces deux bases peut être complété en

$$(e_1, \cdots, e_{k_1}, e_{k_1+1}, \cdots, e_{k_1+k_2}, e_{k_1+k_2+1}, \cdots, e_d)$$

pour former une BON de $\mathbb{R}^d$ (du fait que E_1 et E_2 sont orthogonaux).

Soit $(Y_1, \cdots, Y_d)$, les coordonnées de Y dans cette base ; elles sont indépendantes de loi $\mathcal{N}(0, \sigma^2)$ car le changement de base est orthonormal et nous avons vu que la distribution de Y était conservé par transformation isométrique. Comme

$$P_{E_1}(Y) = Y_1 e_1 + \cdots + Y_{k_1} e_{k_1} \Longrightarrow$$
$$\|P_{E_1}(Y)\|^2 = \sigma^2 \left(\left(\frac{Y_1}{\sigma}\right)^2 + \cdots + \left(\frac{Y_{k_1}}{\sigma}\right)^2 \right),$$
$$P_{E_2}(Y) = Y_{k_1+1} e_{k_1+1} + \cdots + Y_{k_1+k_2} e_{k_1+k_2} \Longrightarrow$$
$$\|P_{E_2}(Y)\|^2 = \sigma^2 \left(\left(\frac{Y_{k_1+1}}{\sigma}\right)^2 + \cdots + \left(\frac{Y_{k_1+k_2}}{\sigma}\right)^2 \right).$$

On voit bien ainsi l'indépendance entre les deux projections et le fait que la loi de $\|P_{E_1}(Y)\|^2/\sigma^2$ (resp. $\|P_{E_2}(Y)\|^2/\sigma^2$) est une loi $\chi^2(k_1)$ (resp. $\chi^2(k_2)$). $\square$

2.2.7 Fonctions caractéristiques

Définition 31 *Soit X un vecteur aléatoire sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbb{R}^d$. La fonction caractéristique de X est la fonction $\phi_X : \mathbb{R}^d \mapsto \mathbb{C}$ telle que*

$$\phi_X(t) = \mathbb{E}\left[\exp(i\, t^T X)\right] = \int_{\mathbb{R}^d} e^{ii\, t^T X} dP_X(x).$$

Corollaire 1 *$(X_1, \cdots, X_n)$ sont des variables aléatoires indépendantes si et seulement si pour tout $(t_1, \cdots, t_n) \in \mathbb{R}^n$,*

$$\phi_{(X_1, \cdots, X_n)}(t_1, \cdots, t_n) = \prod_{j=1}^{n} \phi_{X_j}(t_j).$$

La fonction caractéristique existe sur $\mathbb{R}$ et $\phi_X(0) = 1$. ϕ_X est aussi la transformée de Fourier de la mesure P_X. Elle caractérise complément la loi de X :

Théorème 12 *Soit X et Y des vecteurs aléatoires sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbb{R}^d$, de lois P_X et P_Y. Alors $P_X = P_Y$ si et seulement si $\phi_X = \phi_Y$.*

Théorème 13 (Théorème d'inversion) *Si X est un vecteur aléatoire sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbb{R}^d$ et si ϕ_X est une fonction intégrable par rapport à la mesure de Lebesgue λ_d sur $\mathbb{R}^d$, alors X admet une densité f_X par rapport à λ_d telle que pour $x \in \mathbb{R}^d$,*

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\, t^T x} \phi_X(t) dt.$$

Si X est une variable aléatoire sur $(\Omega, \mathcal{A}, P)$ de fonction génératrice ϕ_X. Alors si $\mathbb{E}(|X|^n) < +\infty$ (ou $X \in L^n(\Omega, \mathcal{A}, P)$), ϕ_X est n fois dérivable et $\phi_X^{(n)}(t) = i^n \mathbb{E}(X^n e^{itX})$. On a alors $i^n \mathbb{E}(X^n) = \phi_X^{(n)}(0)$. On retrouve ainsi $\mathbb{E}(X^n)$ le moment d'ordre n de X noté m_n.

2.2.8 Convergence de suites de variables aléatoires

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires sur $(\Omega, \mathcal{A}, P)$. On dit que

- (X_n) converge en probabilité vers X, noté $X_n \xrightarrow{P} X$, lorsque pour tout $\varepsilon > 0$,
$$\lim_{n \to \infty} P(|X_n - X| > \varepsilon) = 0.$$
- (X_n) converge dans $L^p(\Omega, \mathcal{A}, P)$ vers X, noté $X_n \xrightarrow{L^p} X$, avec $p > 0$, lorsque
$$\lim_{n \to \infty} \|X_n - X\|_p = 0.$$

– (X_n) converge en loi vers X, noté $X_n \xrightarrow{\mathcal{L}} X$, lorsque,

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \quad \text{pour tout } x \in \mathbb{R} \text{ tel que } F_X \text{ continue en } x.$$

– (X_n) converge presque sûrement vers X, noté $X_n \xrightarrow{p.s.} X$, lorsque

$$P(E) = 1 \quad \text{où} \quad E = \{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}.$$

ou encore lorsque pour tout $\varepsilon > 0$,

$$\lim_{n\to\infty} P(\sup_{m\geq n} |X_m - X| > \varepsilon) = 0.$$

Remarque : Les définitions de la convergence en loi, en probabilité et de la convergence presque sûre se généralisent facilement pour les vecteurs aléatoires de dimension $d > 1$. Pour la convergence en probabilité par exemple, on remplacera $\lim_{n\to+\infty} P(|X_n - X| > \epsilon) = 0,\ \forall\, \epsilon > 0$ par $\lim_{n\to+\infty} P(\|X_n - X\| > \epsilon) = 0,\ \forall\, \epsilon > 0$ où $\|\cdot\|$ est une norme quelconque sur $\mathbb{R}^d$ puisque toutes les normes sont équivalentes sur $\mathbb{R}^d$.

Propriété 12

1. *p.s. et $L^p \longrightarrow \mathcal{P} \longrightarrow \mathcal{L}$.*
2. *pour $q \geq p$, $L^q \longrightarrow L^p$ et $L^\infty \longrightarrow$ p.s.*
3. *La convergence en loi n'entraîne pas la convergence en probabilité. Mais $(X_n \xrightarrow{P} C) \Longleftrightarrow (X_n \xrightarrow{\mathcal{L}} C)$ pour C une constante.*
4. *Si g est une fonction continue alors $(X_n \xrightarrow{p.s.} X) \Longrightarrow (g(X_n) \xrightarrow{p.s.} g(X))$, $(X_n \xrightarrow{P} X) \Longrightarrow (g(X_n) \xrightarrow{P} g(X))$ et $X_n \xrightarrow{\mathcal{L}} X \Longrightarrow g(X_n) \xrightarrow{\mathcal{L}} g(X)$.*
5. *Si pour tout $\varepsilon > 0$, $\displaystyle\sum_{n=0}^{\infty} P(|X_n - X| > \varepsilon) < +\infty$ alors $X_n \xrightarrow{p.s.} X$ (application du Lemme de Borel-Cantelli).*
6. *Si il existe $r > 0$ tel que $\mathbb{E}(|X_n|^r) < +\infty$ et $\displaystyle\sum_{n=0}^{\infty} \mathbb{E}(|X_n - X|^r) < +\infty$ alors $X_n \xrightarrow{p.s.} X$.*
7. *Si $X_n \xrightarrow{p.s.} X$ et $|X_n|^r \leq Z$ telle que Z est une v.a. positive telle que $E(Z) < +\infty$, alors $X_n \xrightarrow{L_r} X$ (application du Théorème de la convergence dominée).*

Théorème 14 (Loi faible des Grands Nombres) *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. iid Alors si $\mathbb{E}(|X_i|) < +\infty$,*

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n} \xrightarrow{P} m = \mathbb{E}X_i.$$

Théorème 15 (Loi forte des Grands Nombres) *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de v.a. iid Alors si* $\mathbb{E}(|X_i|) < +\infty$,

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n} \xrightarrow{p.s.} m = \mathbb{E}X_i.$$

Théorème 16 (Théorème de la limite centrale) *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de v.a. iid Alors si* $\sigma^2 = \mathbb{E}X_i^2 < +\infty$, *et* $m = \mathbb{E}X_i$,

$$\sqrt{n}\frac{\overline{X}_n - m}{\sigma} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Théorème 17 (Loi forte des Grands Nombres multidimensionnelle) *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de vecteurs aléatoires iid à valeurs dans* $\mathbb{R}^d$. *Alors si* $\mathbb{E}(\|X_i\|) < +\infty$ *(pour* $\|.\|$ *une norme sur* $\mathbb{R}^d$*)*,

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n} \xrightarrow{p.s.} m = \mathbb{E}X_i.$$

Théorème 18 (Théorème de la limite centrale multidimensionnel) *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de vecteurs aléatoires iid à valeurs dans* $\mathbb{R}^d$. *Alors si* Σ *matrice de covariance de chaque* X_i *existe, et* $m = \mathbb{E}X_i$,

$$\sqrt{n}\Big(\overline{X}_n - m\Big) \xrightarrow{\mathcal{L}} \mathcal{N}_d(0,\Sigma).$$

On donne aussi les conditions de convergences d'un couple de v.a. :

- Si $X_n \xrightarrow{P} X$ et $Y_n \xrightarrow{P} Y$ alors $(X_n, Y_n) \xrightarrow{P} (X,Y)$.
- Si $X_n \xrightarrow{p.s.} X$ et $Y_n \xrightarrow{p.s.} Y$ alors $(X_n, Y_n) \xrightarrow{p.s.} (X,Y)$.
- Si $X_n \xrightarrow{L^p} X$ et $Y_n \xrightarrow{L^p} Y$ alors $(X_n, Y_n) \xrightarrow{L^p} (X,Y)$.
- Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} Y$ <u>et</u> que les suites (X_n) et (Y_n) sont indépendantes alors $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X,Y)$ où X et Y sont indépendants.

En général $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} Y$ n'implique pas la convergence du couple (X_n, Y_n) sauf dans le cas utile suivant :

Théorème 19 (Théorème de Slutsky) *Soit* (A_n, B_n, X_n, X) *des vecteurs aléatoires de* $\mathbb{R}^d$ *et* A *une matrice* $q \times q$ *et* B *un vecteur de* $\mathbb{R}^d$. *Supposons que* $A_n \xrightarrow{\mathcal{L}} A$, $B_n \xrightarrow{\mathcal{L}} B$ *et* $X_n \xrightarrow{\mathcal{L}} X$. *Alors*

$$A_nX_n + B_n \xrightarrow{\mathcal{L}} AX + B.$$

Théorème 20 (δ-méthode) *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$, indépendants et identiquement distribués, telle que Σ matrice de covariance de chaque X_i existe, et $m = \mathbb{E}X_i$. Soit $g : \mathbb{R}^d \to \mathbb{R}^p$ une fonction de classe $\mathcal{C}^1$ sur un voisinage autour de m, de matrice Jacobienne $J_g(m)$ en m. Alors,*

$$\sqrt{n}\Big(g(\overline{X}_n) - g(m)\Big) \xrightarrow{\mathcal{L}} \mathcal{N}_d(0, J_g(m) \cdot \Sigma \cdot J'_g(m)).$$

2.2.9 Espérance conditionnelle

Définition 32 *Soit Y une variable aléatoire sur $(\Omega, \mathcal{A}, P)$. Si $\mathcal{B}$ est une sous-tribu de $\mathcal{A}$ et si $Y \in L^2(\Omega, \mathcal{A}, P)$. Alors on note $\mathbb{E}(Y \mid \mathcal{B})$ la projection orthogonale de Y sur $L^2(\Omega, \mathcal{B}, P)$, appelée espérance conditionnelle de Y sachant $\mathcal{B}$. Ainsi :*

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

Par extension, si $Y \in L^1(\Omega, \mathcal{A}, P)$, on définit l'espérance conditionnelle par rapport à $\mathcal{B}$, comme l'unique (p.s.) variable aléatoire, $\mathcal{B}$-mesurable vérifiant p.s. :

$$\int_B \mathbb{E}(Y \mid \mathcal{B})\, dP = \int_B Y\, dP, \quad \text{pour tout } B \in \mathcal{B}.$$

Par convention, si X un vecteur aléatoire à valeurs dans $\mathbb{R}^n$ sur $(\Omega, \mathcal{A}, P)$ et si Y une variable aléatoire sur $(\Omega, \mathcal{A}, P)$, on note $\mathbb{E}(Y \mid X) = \mathbb{E}(Y \mid X^{-1}(\mathcal{B}(\mathbb{R})))$.

Propriété 13

1. *Lemme de Doob : Pour $Y \in L^1(\Omega, \mathcal{A}, P)$, et X une v.a. de$(\Omega, \mathcal{A}, P)$, alors p.s. $\mathbb{E}(Y \mid X) = h(X)$, avec h une fonction borélienne.*
2. *Pour Y_1 et Y_2 deux variables aléatoires sur $(\Omega, \mathcal{A}, P)$, et $(a, b, c) \in \mathbb{R}^3$, alors*
$$\mathbb{E}(aY_1 + bY_2 + c \mid \mathcal{B}) = a\mathbb{E}(Y_1 \mid \mathcal{B}) + b\mathbb{E}(Y_2 \mid \mathcal{B}) + c.$$
3. *Si $Y_1 \leq Y_2$, alors $\mathbb{E}(Y_1 \mid \mathcal{B}) \leq \mathbb{E}(Y_2 \mid \mathcal{B})$.*
4. *Le Lemme de Fatou, les théorèmes de Beppo-Levi, Lebesgue et Jensen s'appliquent avec l'espérance conditionnelle.*
5. *Si $Y \in L^2(\Omega, \mathcal{B}, P)$, alors $\mathbb{E}(Y \mid \mathcal{B}) = Y$; ainsi $\mathbb{E}(g(X) \mid X) = g(X)$ pour g une fonction mesurable réelle.*
6. *On a $\mathbb{E}(\mathbb{E}(Y \mid \mathcal{B})) = \mathbb{E}Y$.*
7. *Si $Y^{-1}(\mathcal{B}(\mathbb{R}))$ et $\mathcal{B}$ sont indépendantes alors $\mathbb{E}(Y \mid \mathcal{B}) = \mathbb{E}Y$; ainsi, si X et Y sont indépendantes, $\mathbb{E}(Y \mid X) = \mathbb{E}Y$.*

8. *Si (X, Y) est un couple de v.a. à valeurs dans $\mathbb{R}^2$ possédant une densité $f_{(X,Y)}$ par rapport à la mesure de Lebesgue, alors si X est intégrable ,*

$$\mathbb{E}(Y \mid X = x) = \frac{\int_{\mathbb{R}} y\, f_{(X,Y)}(x, y)\, dy}{\int_{\mathbb{R}} f_{(X,Y)}(x, y)\, dy},$$

pour tout x tel que $\int_{\mathbb{R}} f_{(X,Y)}(x, y)\, dy > 0$.

Dans le cas Gaussien, on a la propriété suivante :

Proposition 15 *Si $(Y, X_1, \cdots, X_n)$ est un vecteur gaussien, alors on a*

$$\mathbb{E}(Y \,|\, (X_1, \cdots, X_n)) = a_0 + a_1 X_1 + \cdots + a_n X_n,$$

où les a_i sont des réels.