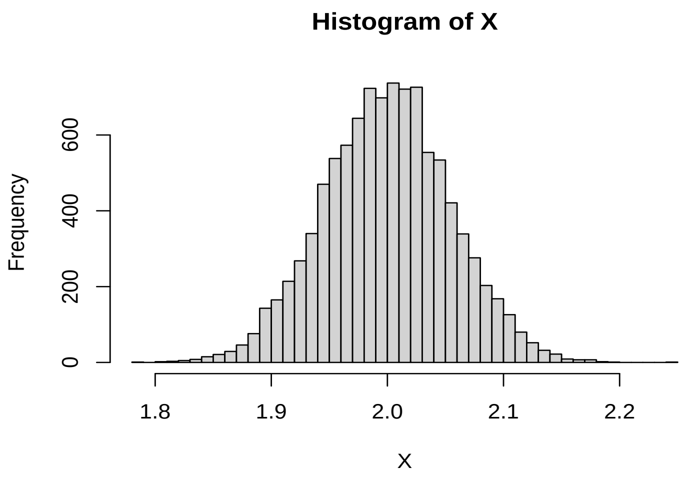
Université Paris I, Panthéon - Sorbonne

LICENCE M.I.A.S.H.S. TROISIÈME ANNÉE

Cours de Statistique 2

JEAN-MARC BARDET (UNIVERSITÉ PARIS 1, SAMM)



U.F.R. 27 et SAMM (Statistique, Analyse et Modélisation Multidisciplinaire)
Université Panthéon-Sorbonne, 90 rue de Tolbiac, 75013 Paris.

Plan du cours

Introduction

1. Variables aléatoires et espérance
2. Vecteurs aléatoires et indépendance
3. Vecteurs gaussiens
4. Convergence et théorèmes limite
5. Estimation paramétrique
6. Tests paramétriques et non paramétriques

References

[1] Barbe et Ledoux, *Probabilités*, Belin.

[2] Dacunha-Castelle et Duflo, *Probabilités et Statistiques (I)*, Masson

[3] Dauxois, J. et Hassenforder, C. (2004). Toutes les probabilités et Statistiques. Cours et Exercices corrigés. Ellipses.

[4] Garet, O. et Kuntzmann, A., *De l'intégration aux probabilités*, Ellipses.

[5] Leboeuf, C., Guegand, J., Roque, J.L. et Landry, P. Cours de Probabilités et de statistiques, Ellipses

[6] Leboeuf, C.,Guegand, J.,Roque, J.L. et Landry, P. Exercices corrigés de probabilités, Ellipses

[7] Ross, S.M (2007). Initiation aux probabilités, Enseignement des Mathématiques. Presses polytechniques et universitaires romandes.

[8] Saporta, G. Probabilités, analyse des données et statistique (2nd édition), éditions Technip.

Documents accessibles librement sur internet

- Des cours sur le site STAFAV: `https://www.math.u-psud.fr/ stafav/`.
- Un site de Paris V: `http://helios.mi.parisdescartes.fr/ glaunes/Proba3/CoursProbasStats.pdf`.
- Le site du LPSM de Paris VI: `https://www.lpsm.paris/formation/supports-de-cours/`. Notamment le cours de J. Bertoin.
- Un poly de maths générales (avec probas) très bien fait: `https://www.math.univ-toulouse.fr/ barthe/L1mat`
- Le premier cours de probabilités de l'Ecole Polytechnique: `http://josselin-garnier.org/wp-content/uploa`

Introduction

Il demeure des choses inconnues à partir des connaissances antérieures en probabilités :

- Qu'est-ce qu'un événement et l'ensemble de tous les événements ?
- Que se passe-t-il pour des probabilités d'événements moins classiques (par exemple l'ensemble des décimaux) ?
- Comment traiter une variable aléatoire qui est continue et discrète à la fois (par exemple le nombre de minutes passées devant la TV) ?

Rappels: Mesures

Tribus

Notation.
- Ω est un ensemble (fini ou infini).
- $\mathcal{P}(\Omega)$ est l'ensemble de tous les sous-ensembles (parties) de Ω.

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

Définition. Soit une famille $\mathcal{F}$ de parties de Ω (donc $\mathcal{F} \subset \mathcal{P}(\Omega)$). On dit que $\mathcal{F}$ est une algèbre si:

- $\Omega \subset \mathcal{F}$;
- lorsque $A \in \mathcal{F}$ alors $(\Omega \setminus A) \in \mathcal{F}$;
- pour tout $n \in \mathbb{N}^*$, lorsque $(A_1, \cdots, A_n) \in \mathcal{F}^n$ alors $A_1 \cup \cdots \cup A_n \in \mathcal{F}$.

Définition. 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 :

- $\Omega \subset \mathcal{F}$;
- lorsque $A \in \mathcal{F}$ alors $(\Omega \setminus A) \in \mathcal{F}$;
- pour $I \subset \mathbb{N}$, lorsque $(A_i)_{i \in I} \in \mathcal{F}^I$ alors $\bigcup_{i \in I} A_i \in \mathcal{A}$.

Exemple.

- Cas du Pile ou Face.
- Cas où Ω est infini : $\Omega = \mathbb{N}$ par exemple.

Propriété. *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 $I \subset \mathbb{N}$, 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. 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).

Remarque.

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

Rappel.

- Un ensemble ouvert U dans un espace métrique X est telle que pour tout $x \in U$, il existe $r > 0$ tel que $B(x, r) \subset U$.
- On dit qu'un ensemble dans un espace métrique X est fermé si son complémentaire dans X est ouvert.

Définition. 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.

Exemple.

- Boréliens sur $\mathbb{R}$, sur $]0, 1[$.
- Boréliens sur $\mathbb{R}^2$.

Espace mesurable

Définition. Soit Ω un ensemble et soit $\mathcal{A}$ une tribu sur Ω. On dit que $(\Omega, \mathcal{A})$ est un espace mesurable.

Corollaire. *Quand on s'intéressera aux probabilités, on dira que $(\Omega, \mathcal{A})$ est un espace probabilisable.*

Propriété. *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 $A_1 \times \cdots \times A_n$ où chaque $A_i \in \mathcal{A}_i$. L'ensemble des ensembles élémentaires est une algèbre et on note $\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n$ (on dit $\mathcal{A}_1$ tensoriel $\mathcal{A}_2$ ... tensoriel $\mathcal{A}_n$) la tribu sur Ω engendrée par ces ensembles élémentaires.*

Exemple.

Pavés de $\mathbb{R}^d$.

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

Exemple.

Pile / Face 2 fois.

Définitions et Propriétés d'une mesure

Définition. Soit $(\Omega, \mathcal{A})$ un espace mesurable. L'application $\mu : \mathcal{A} \to [0, +\infty]$ est une mesure si :

- $\mu(\emptyset) = 0$.
- Pour tout $I \subset \mathbb{N}$ 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).

Définition. 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.

Exemple.

Cas de $\Omega = \mathbb{R}$, de $\mathbb{N}$, ou $\mathbb{R}^2$.

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

Remarque.

Sur $(\Omega, \mathcal{A})$, on peut définir une infinité de mesures.

Propriété. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $(A_i)_{i \in \mathbb{N}}$, 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 $I \subset \mathbb{N}$, 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 $i \in \mathbb{N}$ (suite croissante en sens de l'inclusion), alors $(\mu(A_n))_{n \in \mathbb{N}}$ est une suite croissante convergente telle que* $\mu\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \lim_{i \to +\infty} \mu(A_i)$ *(même si cette limite est $+\infty$).*
5. *Si $A_{i+1} \subset A_i$ pour tout $i \in \mathbb{N}$ (suite décroissante en sens de l'inclusion) et $\mu(A_0) < +\infty$, alors $(\mu(A_n))_{n \in \mathbb{N}}$ est une suite décroissante convergente telle que $\mu\left(\bigcap_{i \in \mathbb{N}} A_i\right) = \lim_{i \to +\infty} \mu(A_i)$.*

Exemple.

1. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. On définit $\nu(A) = \mu(A \cap B)$ où $B \in \mathcal{A}$. ν mesure ?
2. Si μ_1 et μ_2 mesures sur $(\Omega, \mathcal{A})$, $\mu_1 + \mu_2$ et $\alpha\mu$ sont-elles des mesures ?

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

1. On définit $\limsup(A_n)_n = \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m$ (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 $\liminf(A_n)_n = \bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} A_m$ (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).

Exemple.

Cas des suites croissantes et décroissantes d'ensembles.

Théorème (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 $\mathcal{A}$.*

Exemple.

Définition de la mesure de Lebesgue sur $\mathbb{R}$, $\mathbb{R}^n$,...

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

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

Exemple.

- Mesure de Lebesgue sur $\mathbb{N}$ ou $\mathbb{Q}$.
- La propriété " la suite de fonction $f_n(x) = x^n$ converge vers la fonction $f(x) = 0$" est vraie λ-presque partout sur $[0,1]$.
- Soit $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ et soit F la fonction définie par $F(x) = \mu(]-\infty, x])$ pour $x \in \mathbb{R}$.

Fonctions mesurables

Rappel. 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).

Définition. Soit $f : E \mapsto F$ et soit $\mathcal{I}$ une tribu sur F. On note $f^{-1}(\mathcal{I})$ l'ensemble de sous-ensembles de Ω tel que $f^{-1}(\mathcal{I}) = \{f^{-1}(I), I \in \mathcal{I}\}$.

Propriété. *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. 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 $f^{-1}(A') \in \mathcal{A}$).

Exemple.

- Fonction indicatrice.
- Combinaison linéaire de fonctions indicatrices.

Remarque.

Dans le cas où $(\Omega, \mathcal{A})$ est un espace probabilisable, et si $f : \Omega \mapsto \mathbb{R}$, alors si f est une fonction mesurable sur $\mathcal{A}$ et $\mathcal{B}(\mathbb{R})$, alors f est une variable aléatoire.

Exemple.

Nombre de Piles dans un jeu de Pile/Face.

Remarque.

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 de Ω', si f est une fonction mesurable sur $\mathcal{A}$ et $\mathcal{B}(\Omega')$, alors f est dite fonction borélienne.

Proposition. *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) $\Longleftrightarrow$ ($f^{-1}(\mathcal{F}) \subset \mathcal{A}$)*

Conséquence.

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

Propriété.

- *Soit f mesurable de $(\Omega, \mathcal{A})$ dans $(\Omega', \mathcal{A}')$ et g mesurable de $(\Omega', \mathcal{A}')$ dans $(\Omega'', \mathcal{A}'')$. Alors $g_0 f$ est mesurable dans $\mathcal{A}$ et $\mathcal{A}'$.*
- *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 dans $\mathcal{A}$ et $\mathcal{A}_1 \otimes \mathcal{A}_2$.*
- *Soit $(f_n)_{n \in \mathbb{N}}$ 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 $\forall \omega \in \Omega$, $\lim_{n \to \infty} f_n(\omega) = f(\omega)$). Alors f est mesurable dans $\mathcal{A}$ et $\mathcal{B}(\Omega')$.*

Définition. 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.

Cas particulier.

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

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

Définition. Soit $f : \Omega \to \mathbb{R}$. 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 $f(\omega) = \sum_{i=1}^{n} \alpha_i \mathbb{I}_{A_i}(\omega)$.

Remarque.

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

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

Conséquence. Soit f une fonction réelle mesurable. Alors f est limite simple de fonctions étagées.

Intégration de Lebesgue

Dans toute la suite, on considère $(\Omega, \mathcal{A}, \mu)$ un espace mesuré.

Intégrale de Lebesgue d'une fonction positive

Définition. 1. Soit $f = \mathbb{I}_a$, où $A \in \mathcal{A}$. Alors :

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

2. Soit $f = \mathbb{I}_a$, où $A \in \mathcal{A}$ et soit $B \in \mathcal{A}$. Alors :

$$\int_B f\, d\mu = \int_B f(\omega) d\mu(\omega) = \int \mathbb{I}_B \mu(A)(\omega) f(\omega) d\mu(\omega) = \mu(A \cap B).$$

3. Soit f une fonction étagée positive telle que $f = \sum_{i=1}^{n} \alpha_i \mathbb{I}_{A_i}$, où les $A_i \in \mathcal{A}$ et $\alpha_i > 0$ et soit $B \in \mathcal{A}$. Alors :

$$\int_B f\, d\mu = \int_B f(\omega) d\mu(\omega) = \int \mathbb{I}_B(\omega) f(\omega) d\mu(\omega) = \sum_{i=1}^{n} \alpha_i \mu(A_i \cap B).$$

Exemple.

Fonction $\mathbb{I}_{\mathbb{Q}}$, fonctions en escalier,...

Définition. 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 :

$$\int_B f\, d\mu = \int \mathbb{I}_B(\omega) f(\omega) d\mu(\omega) = \sup\left\{ \int_B g\, d\mu,\ \text{pour } g \text{ étagée positive telle que } g \leq f \right\}.$$

Propriété. *Soit f une fonction $\mathcal{A}$-mesurable positive et soit A et $B \in \mathcal{A}$. Alors :*

1. *Pour* $c \geq 0$, $\displaystyle\int_B cf\,d\mu = c\int_B f\,d\mu$.

2. *Si* $A \subset B$, *alors* $\displaystyle\int_A f\,d\mu \leq \int_B f\,d\mu$.

3. *Si* g *est une fonction* $\mathcal{A}$*-mesurable positive telle que* $0 \leq f \leq g$ *alors* $0 \leq \displaystyle\int_B f\,d\mu \leq \int_B g\,d\mu$.

4. *Si* $\mu(B) = 0$ *alors* $\displaystyle\int_B f\,d\mu = 0$.

Théorème (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 :*

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

Conséquence. 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 (Lemme de Fatou)**.** *Soit* $(f_n)_n$ *est une suite de fonctions mesurables positives alors :*

$$\int \left(\liminf_{n\to\infty} f_n\right) d\mu \leq \liminf_{n\to\infty} \int f_n d\mu.$$

Exemple.

Appliquer Fatou à (f_n) telle que $f_{2n} = \mathbb{I}_A$ et $f_{2n+1} = \mathbb{I}_B$.

Intégrale de Lebesgue d'une fonction réelle et propriétés

Définition. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, $B \in \mathcal{A}$ et soit f une fonction $\mathcal{A}$-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 $\displaystyle\int_B |f|\,d\mu < +\infty$. On a alors

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

Notation. Lorsque f est μ-intégrable sur B, soit $\displaystyle\int |f|\,d\mu < +\infty$, on note $f \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ (on dit que f est $\mathcal{L}^1$).

Exemple.

Intégrale de Riemann et intégrale de Lebesgue.
Cas de la masse de Dirac.

Propriété. *On suppose que* f *et* $g \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$. *Alors :*

1. $\displaystyle\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu$ *pour* $(\alpha, \beta) \in \mathbb{R}^2$.

2. *Si* $f \leq g$ *alors* $\displaystyle\int f\,d\mu \leq \int g\,d\mu$.

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

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

Extension.

Le Théorème de Lebesgue s'applique également dans le cas où $(f_n)_n$ converge presque partout vers f.

Exemple.

Convergence d'intégrale dépendant d'un paramètre : par exemple $\displaystyle\int_0^\infty \frac{f(x)}{1+x^n} dx$.

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

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

Exemple.

Soit X une v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$. Alors $\phi\left(\mathbb{E}[X]\right) \leq \mathbb{E}\left(\phi(X)\right)$.

Mesures induites et densités

Théorème (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 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors, si $\phi_0 f \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$,*

$$\int_{\Omega'} \phi\, d\mu_f = \int_\Omega \phi_0 f\, d\mu.$$

Définition. 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 $A \in \mathcal{A}$, $\mu(A) = 0 \Longrightarrow \nu(A) = 0$.

Propriété. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et f une fonction définie sur $(\Omega, \mathcal{A})$ mesurable et positive. On suppose que pour $A \in \mathcal{A}$, $\nu(A) = \displaystyle\int_A f\, d\mu$. 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. 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 (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 (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 $f : \Omega \mapsto \mathbb{R}$, $\mathcal{A}$-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 $\mathcal{L}^p$

Définition. 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 $f : \Omega \mapsto \mathbb{R}$, mesurables et telles que $\int |f|^p d\mu < +\infty$.

Définition. Pour $f \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$, où $p > 0$, on note $\| f \|_p = \left(\displaystyle\int |f|^p d\mu \right)^{1/p}$.

Propriété (Inégalité de Hölder). *Soit $p > 1$ et $q > 1$ 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 (Inégalité de Minkowski). *Soit $p > 1$ 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*

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

Remarque.

Pour $p > 1$, $\| . \|_p$ définie ainsi sur une semi-norme sur $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$. Pour obtenir une norme, il faut se placer dans l'espace $\mathbb{L}^p(\Omega, \mathcal{A}, \mu)$ 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 $\mathbb{L}^p(\Omega, \mathcal{A}, \mu)$ on dira que $f = g$ lorsque $f = g$ μ-presque partout).

Définition. Pour f et $g \in \mathbb{L}^2(\Omega, \mathcal{A}, \mu)$, on définit le produit scalaire $< f, g > = \displaystyle\int f.g\, d\mu$. On muni ainsi $\mathbb{L}^2(\Omega, \mathcal{A}, \mu)$ d'une structure d'espace de Hilbert. On dira que f est orthogonale à g lorsque $< f, g > = 0$.

Conséquence. Si A est un sous-espace vectoriel fermé de $\mathbb{L}^2(\Omega, \mathcal{A}, \mu)$ (par exemple un sous-espace de dimension finie), alors pour tout $f \in \mathbb{L}^2(\Omega, \mathcal{A}, \mu)$, il existe un unique projeté orthogonal de f sur A, noté f_A, qui vérifie $f_A = \operatorname*{Arginf}_{g \in A} \| g - f \|_2$.

1 Variables aléatoires et espérance

1.1 Variables aléatoires

Définition. On dit que X est une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité si: $X : \Omega \to \mathbb{R}$ est une application mesurable de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Pourquoi demander à X d'être mesurable? Parce que l'on veut pouvoir définir une fonction de répartition pour X, soit $F_X(x) = \mathbb{P}(\text{"}X \leq x\text{"}) = \mathbb{P}\big(\{\omega \in \Omega,\ X(\omega) \leq x\}\big) = X^{-1}(]-\infty, x])$: il faut donc que l'ensemble $X^{-1}(]-\infty, x])$ soit un évenement de $\mathcal{A}$ pour tout $x \in \mathbb{R}$.

Dans la suite, pour une variable aléatoire X définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ on définit $F_X(x) = \mathbb{P}(X \leq x)$ et la mesure de probabilité $\mathbb{P}_X$ de X: pour $B \in \mathcal{B}(\mathbb{R})$, $\mathbb{P}_X(B) = \mathbb{P}(X \in B)$.
On va dans la propriété suivante utiliser pleinement le fait qu'une variable aléatoire est une fonction mesurable:

Propriété. *Soit X et Y deux v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$ et $g : \mathbb{R}^2 \to \mathbb{R}$ une fonction borélienne. Alors $Z = g(X, Y)$ est une v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$.*

Par itération du procédé, la propriété est aussi vraie pour une fonction g à n variables. Ma aussi pour une fonction à 1 variable.
De la même manière, toujours du fait qu'une variable aléatoire est une fonction mesurable:

Propriété. *Soit X et Y deux v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$ telles que $\mathbb{P}_X = \mathbb{P}_Y$, ce que l'on peut aussi noter $X \overset{\mathcal{L}}{\sim} Y$. Alors pour et $g : \mathbb{R} \to \mathbb{R}$ une fonction borélienne, $g(X) \overset{\mathcal{L}}{\sim} g(Y)$.*

Proof. Pour tout $B \in \mathcal{B}(\mathbb{R})$, $\mathbb{P}_{g(X)}(B) = \mathbb{P}(g(X) \in B) = \mathbb{P}(X \in g^{-1}(B)) = \mathbb{P}_X(g^{-1}(B))$. Mais on a aussi $\mathbb{P}_{g(Y)}(B) = \mathbb{P}_Y(g^{-1}(B)) = \mathbb{P}_X(g^{-1}(B))$ puisque $\mathbb{P}_X = \mathbb{P}_Y$. D'où le résultat. □

Définition. Si X est une v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$ et $0 < p < 1$, le quantile d'ordre p de X est : $q_X(p) = \inf \big\{y \in \mathbb{R},\ F_X(y) \geq p\big\}$.

Cas particulier: Si $p = 1/2$, alors $q_X(p)$ est la médiane (théorique) de X.

Propriété. *Si X est une v.a. continue, $q_X(p) = \widetilde{F}_X^{-1}(p)$ où $\widetilde{F}_X : x \in X(\Omega) \mapsto F_X(x)$.*

Exemple: Si X suit une distribution $\mathcal{E}(\lambda)$ (exponentielle),

$$F_X(x) = (1 - e^{-\lambda x}) \mathbb{I}_{x \geq 0} \quad \text{and} \quad q_X(1/2) = \frac{\ln(2)}{\lambda} \neq \mathbb{E}[X] = \frac{1}{\lambda}.$$

1.2 Espérance de variables aléatoires

Définition. Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Alors si $X \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ (donc si $\int_\Omega |X(\omega)|\, d\mathbb{P}(\omega) < \infty$), on définit l'espérance de X par le réel:

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

Plus généralement, si $\phi : \mathbb{R} \mapsto \mathbb{R}$ est borélienne et si $\phi(X) \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, on définit l'espérance de $\phi(X)$ par

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

Propriété. *Si X est une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$, si $\phi : \mathbb{R} \mapsto \mathbb{R}$ est borélienne telle que $\phi(X) \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, alors :*

$$\mathbb{E}\left[\phi(X)\right] = \int_{\mathbb{R}} \phi(x)\, d\mathbb{P}_X(x).$$

Proof. Théorème du transport... □

Conséquence.
- Si $\mathbb{P}_X$ est absolument continue par rapport à la mesure de Lebesgue (donc X est une v.a. dite absolument continue), de densité f_X, alors $\mathbb{E}[\phi(X)] = \displaystyle\int_{\mathbb{R}} \phi(x)\, f_X(x)\, dx$.
- Si $\mathbb{P}_X$ est absolument continue par rapport à une mesure de comptage sur $\{x_i\}_{i \in I}$ avec $I \subset \mathbb{N}$ (donc X est une v.a. dite discrète), de densité p_X avec $p_X(i) = \mathbb{P}(X = x_i)$, alors $\mathbb{E}[\phi(X)] = \displaystyle\sum_{i \in I} p_X(i)\, \phi(x_i)$.

Propriété.
1. *Soit X et Y des variables aléatoires telles que X et $Y \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$. Alors pour tout $(a, b) \in \mathbb{R}^2$, $aX + bY \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ et*
$$\mathbb{E}[a\, X + b\, Y] = a\, \mathbb{E}[X] + b\, \mathbb{E}[Y].$$
2. *Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$, et soit $B \in \mathcal{B}(\mathbb{R})$. Alors $\mathbb{E}[\mathbb{I}_B(X)] = \mathbb{P}(X \in B)$.*
3. *Si X est une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$, si $\phi : \mathbb{R} \mapsto \mathbb{R}$ est une fonction borélienne* **convexe** *telle que X et $\phi(X) \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, alors*
$$\mathbb{E}[\phi(X)] \geq \phi(\mathbb{E}[X]) \quad \textit{(Inégalité de Jensen).}$$

4. *Soit X et Y des variables aléatoires telles que $X \in \mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ et $Y \in \mathbb{L}^q(\Omega, \mathcal{A}, \mathbb{P})$ avec $\frac{1}{p} + \frac{1}{q} = 1$ où $p > 1$, $q > 1$. Alors $X\,Y \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ et*

$$\mathbb{E}[|X\,Y|] \leq (\mathbb{E}[|X|^p])^{1/p}(\mathbb{E}[|Y|^q])^{1/q} \quad \textit{(Inégalité de Hölder).}$$

5. *Soit X et Y des variables aléatoires telles que X et $Y \in \mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$, avec $p \geq 1$. Alors $X + Y \in \mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ et*

$$(\mathbb{E}[|X+Y|^p])^{1/p} \leq (\mathbb{E}[|X|^p])^{1/p} + (\mathbb{E}[|Y|]^p)^{1/p} \quad \textit{(Inégalité triangulaire de Minkowski).}$$

Proof. 1. On a d'abord $aX + bY$ qui est une v.a., puis $\mathbb{E}[|aX + bY|] \leq |a|\mathbb{E}[|X|] + |b|\mathbb{E}[|Y|] < \infty$. On utilise ensuite la linéarité de l'intégrale.

2. La v.a. $\mathbb{I}_B(X)$ est une v.a. de Bernoulli de paramètre $\mathbb{P}(X \in B)$, et donc également d'espérance $\mathbb{P}(X \in B)$.

3. Vue en "Intégration et probabilités".

4. Plus généralement, on a $\int |fg| d\mu \leq \|f\|_p \, \|g\|_q$ avec $\|f\|_p = \big(\int |f|^p d\mu\big)^{1/p}$. En effet, la fonction $x \in]0, \infty[\mapsto -\ln(x)$ est convexe, donc pour tout $a > 0$, $b > 0$ et $\theta \in [0,1]$, $-\log(a\theta + b(1-\theta)) \leq -\theta \log(a) - (1-\theta)\log b$. En passant à l'exponentielle, on en déduit que $a^\theta b^{1-\theta} \leq a\theta + b(1-\theta)$.
Pour tout $x \in \Omega$, choisissons $a = \frac{|f(x)|^p}{\|f\|_p^p}$, $b = \frac{|g(x)|^q}{\|f\|_q^q}$ et $\theta = 1/p$. Alors pour tout $x \in \Omega$ et avec $1 - \theta = 1/q$,

$$\frac{|f(x)|}{\|f\|_p} \, \frac{|g(x)|}{\|g\|_q} \leq \frac{1}{p} \, \frac{|f(x)|^p}{\|f\|_p^p} + \frac{1}{q} \, \frac{|g(x)|^q}{\|g\|_q^q}.$$

Intégrons des 2 côtés (on peut et on garde le sens de l'inégalité), on obtient:

$$\frac{\int |fg| d\mu}{\|f\|_p \|g\|_q} \leq \frac{1}{p} \, \frac{\int |f|^p d\mu}{\|f\|_p^p} + \frac{1}{q} \, \frac{\int |g|^q d\mu}{\|g\|_q^q} \leq \frac{1}{p} + \frac{1}{q} = 1,$$

d'où le résultat.

5. On va montrer plus généralement que $\|f + g\|_p \leq \|f\|_p + \|g\|_p$ pour $p \geq 1$. Il est clair que si $\|f + g\|_p = 0$, le résultat est vrai. De même si $p = 1$. Pour $p > 1$,

$$\begin{aligned}
\|f+g\|_p^p &\leq \int (|f| + |g|)|f+g|^{p-1} d\mu \\
&\leq \int |f| \, |f+g|^{p-1} d\mu + \int |g| \, |f+g|^{p-1} d\mu \\
&\leq \|f\|_p \Big(\int |f+g|^{q(p-1)} d\mu\Big)^{1/q} + \|g\|_p \Big(\int |f+g|^{q(p-1)} d\mu\Big)^{1/q} \quad \text{(Inégalité de Hölder)} \\
&\leq \big(\|f\|_p + \|g\|_p\big) \, \|f+g\|_p^{p-1},
\end{aligned}$$

d'où l'inégalité.

□

Conséquence. Soit X une variable aléatoire telle que $X \in \mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ pour $p \geq 1$. Alors $\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}$ est une norme sur $\mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$.

Conséquence. Soit X une variable aléatoire telle que $X \in \mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ pour $p > 0$. Alors pour tout $0 < r \leq p$, $X \in \mathbb{L}^r(\Omega, \mathcal{A}, \mathbb{P})$ et

$$(\mathbb{E}[|X|^r])^{1/r} \leq (\mathbb{E}[|X|^p])^{1/p}.$$

Définition. Pour X et Y des variables aléatoires telles que X et $Y \in \mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$, on définit:

- la variance de X, $\mathrm{var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.
- la covariance de X et Y par

$$\mathrm{cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] = \mathbb{E}[X\,Y] - \mathbb{E}[X]\,\mathbb{E}[Y].$$

Remarque: On a $\mathrm{var}(X) = \mathrm{cov}(X, X)$.

Propriété. *Sur l'espace vectoriel* $E = \Big\{X \in \mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P}),\ \mathbb{E}[X] = 0\Big\}$*, on définit* $< X, Y >= cov(X, Y)$ *pour* $X, Y \in E$*. Alors* $< \cdot, \cdot >$ *définit un produit scalaire.*

Conséquence. Pour X et Y deux v.a. sur $\mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$, alors

$$|\mathrm{cov}(X, Y)|^2 \leq \mathrm{var}(X)\,\mathrm{var}(Y).$$

Propriété. *Pour* X *et* Y *deux v.a. sur* $\mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$*, on peut définir le coefficient de corrélation entre* X *et* Y *par:*

$$cor(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)\,var(Y)}} \quad et \quad \big|cor(X, Y)\big| \leq 1.$$

1.3 Fonction génératrice et fonction caractéristique

Définition. Si X est une v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{N}$, on définit la fonction génératrice de X par :

$$g(z) = \mathbb{E}\big[z^X\big] \quad \text{pour tout } z \in [-1, 1].$$

Remarque: $g(z) = \displaystyle\sum_{k=0}^{\infty} \mathbb{P}(X = k)\, z^k$ est une série entière de rayon de convergence $R \geq 1$.

Propriété. *La fonction* g *est une fonction* $\mathcal{C}^\infty((-1, 1))$ *et* $\mathbb{P}(X = k) = g^{(k)}(0)/k!$.

Proof. Fonction $\mathcal{C}^\infty((-1, 1))$: propriété d'une série entière. De plus, pour $|z| < 1$, $g^{(k)}(z) = \sum_{j=k}^{\infty} \mathbb{P}(X = j)\, j(j-1) \times \cdots \times (j-k+1)\, z^{j-k} = \sum_{\ell=0}^{\infty} \frac{(\ell+k)!}{\ell!}\, z^\ell$. □

Propriété. *Si* $\mathbb{E}[|X|] < \infty$*, la fonction* g *est telle que* $g'(1) = \mathbb{E}[X]$.

Proof. Pour $|z| < 1$, $g'(z) = \sum_{j=1}^{\infty} \mathbb{P}(X = j)\, j\, z^{j-1}$. L'hypothèse $\mathbb{E}[X] < \infty$ implique que cette série entière est normalement convergente sur $[-1, 1]$ et ainsi $g'(1) = \mathbb{E}[X]$. □

Exemple: Si $X \sim \mathcal{P}(\lambda)$ (loi de Poisson) avec $\lambda > 0$. Alors $g(z) = e^{\lambda(z-1)}$ pour $z \in \mathbb{R}$.

A quoi sert la fonction génératrice?

- A remplacer une mesure de probabilité par une fonction (dans le cas des v.a. à valeurs entières positives);
- Mais surtout, à obtenir la loi ou montrer la convergence de sommes de variables indépendantes.

Une extension de la fonction génératrice à toute variable aléatoire, discrète, continue, ou autre, est obtenue par la fonction caractéristique:

Définition. Si X est une v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$, on définit la **fonction caractéristique** de X par:

$$\phi_X(u) = \mathbb{E}\big[e^{i\,u\,X}\big] = \mathbb{E}[\cos(u\,X)] + i\,\mathbb{E}[\sin(u\,X)] \quad \text{pour tout } u \in \mathbb{R}.$$

Remarque: ϕ_X est défine sur $\mathbb{R}$ du fait que $|e^{i\,u\,X}| \leq 1$ pour tout $u \in \mathbb{R}$ et en appliquant ensuite le Théorème de convergence dominée de Lebesgue.

Théorème. $\phi_X = \phi_Y \iff$ *X et Y ont la même distribution de probabilité.*

Proof. On peut prouver que pour tout $a \leq b$ tels que $\mathbb{P}_X(\{a,b\}) = 0$, alors

$$\mathbb{P}(X \in [a,b]) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \, \phi_X(t) \, dt.$$

Pour commencer, on montre que pour $c \in \mathbb{R}$, $\int_{-\infty}^{\infty} \frac{\sin(c\,t)}{t}\, dt = \pi$ si $c > 0$ et $= -\pi$ si $c < 0$, $= 0$ si $c = 0$. Tous ces cas se ramène au calcul de $I(0) = \int_0^\infty \frac{\sin(t)}{t}\, dt$. Cette intégrale existe en tant qu'intégrale de Riemann (mais pas en tant qu'intégrale de Lebesgue!). En effet, la fonction $t \to \frac{\sin(t)}{t}$ est prolongeable par continuité en 0 et en $+\infty$, on utilise une intégration par parties

$$\int_1^T \frac{\sin(t)}{t}\, dt = \big[- \frac{\cos(t)}{t} \big]_1^T - \int_1^T \frac{\cos(t)}{t^2}\, dt = \cos(1) - \int_1^T \frac{\cos(t)}{t^2}\, dt.$$

Cette dernière intégrale existe car la fonction est majorée par $\frac{1}{t^2}$ qui est intégrable en $+\infty$.
Par ailleurs, on considère pour $t \geq 0$, $I(x) = \int_0^\infty \frac{\sin(t)}{t}\, e^{-xt}\, dt$. Cette fonction existe et pour $x > 0$ elle est dérivable et $I'(x) = -\int_0^\infty \sin(t)\, e^{-xt}\, dt$ pour $x > 0$. Or

$$\int_0^\infty \sin(t)\, e^{-xt}\, dt = \int_0^\infty \mathcal{I}m(e^{it})\, e^{-xt}\, dt = \mathcal{I}m\Big(\int_0^\infty e^{it-xt}\, dt\Big) = \mathcal{I}m\Big(-\frac{1}{i-x}\Big) = \frac{1}{1+x^2}.$$

On en déduit que $I'(x) = -\frac{1}{1+x^2}$ soit $I(x) = -\arctan(x) + K$ où $K \in \mathbb{R}$. Comme on sait que $\lim_{x\to\infty} I(x) = 0$, donc $K = \pi/2$. Il suffit de montrer que la fonction I est continue en 0, d'où $I(0) = \int_0^\infty \frac{\sin(t)}{t}\, dt = \pi/2$.

Revenons à la preuve et partons de $\lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\, \phi_X(t)\, dt$. Soit:

$$\begin{aligned}
\frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\, \phi_X(t)\, dt &= \frac{1}{2\pi} \int_{-T}^{T} \int_{-\infty}^{\infty} \frac{e^{it(x-a)} - e^{it(x-b)}}{it}\, d\mathbb{P}_X(x)\, dt \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it}\, dt\, d\mathbb{P}_X(x) \quad \text{(Fubini)} \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-T}^{T} \Big(\frac{\sin(t(x-a))}{t} - \frac{\sin(t(x-b))}{t}\Big)\, dt\, d\mathbb{P}_X(x) \quad \text{(Parité)}.
\end{aligned}$$

Mais si on s'intéresse à $\lim_{T\to\infty} \int_{-T}^{T} \Big(\frac{\sin(t(x-a))}{t} - \frac{\sin(t(x-b))}{t}\Big)\, dt = \int_{-\infty}^{\infty} \Big(\frac{\sin(t(x-a))}{t} - \frac{\sin(t(x-b))}{t}\Big)\, dt = \phi_{a,b}(x)$, alors $\phi_{a,b}(x) = 0$ si $x > b$ ou $x < a$ et $\phi_{a,b}(x) = 2\pi$ pour $a < x < b$. Pour $x = a$ et $x = b$, on obtient $\pm\pi$, mais de l'hypothèse $\mathbb{P}_X(\{a,b\}) = 0$ cela n'interviendra pas. D'où:

$$\lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\, \phi_X(t)\, dt = \frac{1}{2\pi} \int_a^b 2\pi\, d\mathbb{P}_X(x) = \mathbb{P}(X \in [a,b]).$$

□

Remarque: Si X est une v.a. discrète à valeurs entières, alors $\phi_X(u) = g(e^{iu})$.

Propriété. *Si X est une v.a. définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ telle que $\mathbb{E}[|X|^k] < \infty$, alors ϕ_X est une fonction $\mathcal{C}^k(\mathbb{R})$ et $\phi_X^{(k)}(u) = i^k\, \mathbb{E}[X^k e^{iuX}]$ pour tout $u \in \mathbb{R}$.*

Proof. Pour $k = 1$, $u \in \mathbb{R}$ et $h \neq 0$, nous avons $\frac{\phi_X(u+h) - \phi_X(u)}{h} = \mathbb{E}\Big[e^{iuX}\Big(\frac{e^{ihX}-1}{h}\Big)\Big]$. Il est clair que $e^{iuX}\Big(\frac{e^{ihX}-1}{h}\Big) \to i\,X\,e^{iuX}$ pour $h \to 0$ puisque $\frac{e^{ihx}-1}{h} \to i\,x$ lorsque $h \to 0$ pour tout $x \in \mathbb{R}$ (on peut considérer $g(h) = e^{ihx}$ et $(g(h)-g(0))/h \to g'(0)$ lorsque $h \to 0$). De plus, $\Big|e^{iuX}\Big(\frac{e^{iuh}-1}{h}\Big)\Big| \leq |X|$ pour tout $u \in \mathbb{R}$ et $h \neq 0$, et $\mathbb{E}[|X|] < \infty$ par hypothèse. Le théorème de Lebesgue implique alors que $\frac{\phi_X(u+h) - \phi_X(u)}{h} \to \mathbb{E}[iXe^{iuX}] = \phi'_X(u)$. Même type de preuve pour $k \geq 2$. □

Remarque: $\phi'_X(0) = i\,\mathbb{E}[X]$ et $\phi''_X(0) = -\mathbb{E}[X^2]$.

Voici une autre propriété de la fonction caractéristique, spécifique aux v.a. "continues":

Propriété. *Si X est une v.a. de loi continue par rapport à la mesure de Lebesgue, de densité f_X et de fonction caractéristique ϕ_X telle que $\int_{\mathbb{R}} |\phi_X(u)|\, du < \infty$, alors*

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_X(u)\, e^{-i\,u\,x} du \quad \textit{pour tout } x \in \mathbb{R}.$$

Proof. Dans le cas présent, on a: $\phi_X(u) = \displaystyle\int_{-\infty}^{\infty} f_X(x)\, e^{-i\,u\,x} du$. On peut reprendre la preuve d'identification de la loi mais en passant directement à la limite car la fonction caractéristique est alors de module intégrable. D'où pour tout $a < b$,

$$\mathbb{P}(X \in [a,b]) = \int_a^b f_X(x)\,dx = \frac{1}{2\,\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it}\, \phi_X(t)\,dt = \frac{1}{2\,\pi} \int_{-\infty}^{\infty} \int_a^b e^{-itx}\,dx\, \phi_X(t)\,dt = \int_a^b \frac{1}{2\,\pi} \int_{-\infty}^{\infty} e^{-itx}\, \phi_X(t)\,dt\,dx,$$

par Fubini. D'où le résultat par identification puisqu'il est vrai pour tout $a < b$. □

A quoi sert la fonction caractéristique?

- Fonction caractéristique de la somme de v.a. indépendantes;
- Fonction caractéristique pour caractériser l'indépendance (voir un peu plus loin);
- Convergence d'une suite de fonctions caractéristiques vers une fonction caractéristique $\Longleftrightarrow$ convergence en loi (Théorème de Lévy, voir un peu plus loin).

2 Vecteurs aléatoires

2.1 Définitions et premières propriétés

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

Définition. Soit X un vecteur aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$. Alors 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 $x = (x_1, \cdots, x_d)$,

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

Propriété. *Soit X un vecteur aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$. On suppose que $X = (X_1, \cdots, X_d)$. Alors les X_i sont des variables aléatoires sur $(\Omega, \mathcal{A}, \mathbb{P})$, 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 $\mathbb{P}_{X_i}$ déterminées de façon univoque à partir des F_{X_i} sont appelées lois marginales de X.

Remarque: Les lois marginales ne permettent pas d'identifier la loi du vecteur, sauf à spécifier d'autres propriétés (indépendance par exemple).

Corollaire. *Si $F_X(x)$ est continue sur $\mathbb{R}^d$ et si $\frac{\partial^d}{\partial x_1 \cdots \partial x_d} F_X(x_1, \ldots, x_d)$ existe presque partout sur $\mathbb{R}^d$, alors la densité de la loi $\mathbb{P}_X$ par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ vaut:*

$$f_X(x_1, \ldots, x_d) = \frac{\partial^d}{\partial x_1 \cdots \partial x_d} F_X(x_1, \ldots, x_d) \quad \textit{pour presque tout } (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

Propriété. *Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$ dont la mesure de probabilité est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ et de densité f_X. Alors les X_i sont des v.a. à loi absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$ de densité $f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(z_1, \ldots, z_{i-1}, x_i, z_{i+1}, \ldots, z_d)\, dz_1 \ldots dz_{i-1} dz_{i+1} \ldots dz_d$.*

Proof. Si F_X est la fonction de répartition de X, alors pour tout $x_i \in \mathbb{R}$, $F_{X_i}(x_i) = \lim_{x_j \to \infty, j \neq i} F_X(x_1, \ldots, x_d)$. Mais comme la mesure est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$, il existe f_X tel que $F_X(x_1, \ldots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f_X(z_1, \ldots, z_d)\, dz_1 \ldots dz_d$. Donc en utilisant Fubini

$$\begin{aligned} F_{X_i}(x_i) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{x_i} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(z_1, \ldots, z_{i-1}, z_i, z_{i+1}, \ldots, z_d)\, dz_1 \ldots dz_d \\ &= \int_{-\infty}^{x_i} \Big(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(z_1, \ldots, z_{i-1}, z_i, z_{i+1}, \ldots, z_d)\, dz_1 \ldots dz_{i-1} dz_{i+1} \ldots dz_d \Big)\, dz_i \\ &= \int_{-\infty}^{x_i} f_{X_i}(z_i)\, dz_i, \end{aligned}$$

avec $f_{X_i}(z_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(z_1, \ldots, z_{i-1}, z_i, z_{i+1}, \ldots, z_d)\, dz_1 \ldots dz_{i-1} dz_{i+1} \ldots dz_d$, fonction mesurable positive: la loi de X_i est bien absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$ et sa densité est f_{X_i}. □

Remarque: Si X est à valeurs dans $\{x_i\}_{i \in I}$, où $x_i \in \mathbb{R}^d$ et $I \subset \mathbb{N}$, alors X est discrète et sa densité par rapport à la mesure de comptage sur $\{x_i\}_{i \in I}$ est donnée par $\mathbb{P}(X = x_i) = \mathbb{P}\big(X_1 = x_{i,1} \cap \cdots, X_1 = x_{i,d}\big)$.

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

$$\phi_X(t) = \mathbb{E}\left[\exp(i < t, X >)\right] = \int_{\mathbb{R}^d} e^{i<t,x>}\, d\mathbb{P}_X(x),$$

où $< . >$ désigne le produit scalaire euclidien sur $\mathbb{R}^d$ vérifiant $< t, x >= \sum_{i=1}^d t_i x_i$ pour $t = (t_1, \cdots, t_d)$ et $x = (x_1, \cdots, x_d)$.

Remarque: La fonction caractéristique existe sur $\mathbb{R}$ et $\phi_X(0) = 1$. ϕ_X est aussi la transformée de Fourier de la mesure $\mathbb{P}_X$.

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

Proof. Même type de preuve que dans $\mathbb{R}$ sauf que dans le cas multidimensionnel on choisit des $a_j \leq b_j$ où $j = 1, \ldots, n$ vérifiant $\mathbb{P}_X\big(\{a_1, b_1\} \times \cdots \times \{a_d, b_d\}\big) = 0$, et alors on montre que

$$\mathbb{P}\big(X \in [a_1, b_1] \times \cdots \times [a_d, b_d]\big) = \lim_{T \to \infty} \prod_{j=1}^d \Big(\frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-it_j a_j} - e^{-it_j b_j}}{it_j} \Big) \phi_X(t_1, \ldots, t_d)\, dt_1 \ldots dt_d.$$

□

Théorème (Théorème d'inversion). *Si X est un vecteur aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ 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 $x \in \mathbb{R}^d$,*

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

Proof. Même type de preuve que dans $\mathbb{R}$. □

Corollaire. *Si X est un vecteur aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$ alors connaître la loi des variables aléatoires $< u, X >$, pour tous les $u \in \mathbb{R}^d$ (ensemble des combinaisons linéaires issues de X) caractérise la loi de X.*

Définition. Pour X un vecteur aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$ telle que $\mathbb{E}\big[\|X\|^r\big] < \infty$ avec $r > 0$.

1. Si $r=1$, alors on définit le vecteur $\mathbb{E}[X]=\big(\mathbb{E}[X_1],\dots,\mathbb{E}[X_d]\big)^\top$, qui est appelé **espérance** de X;

2. Si $r=2$, alors on définit la matrice $\mathrm{cov}(X)=\mathbb{E}\big[\big(X-\mathbb{E}[X]\big)\,\big(X-\mathbb{E}[X]\big)^\top\big]=\big(\mathrm{cov}(X_i,X_j)\big)_{1\le i,j\le d}$ appelée **matrice de variance-covariance** de X.

Propriété. *Soit X et Y deux vecteurs aléatoires définis sur $(\Omega,\mathcal{A},\mathbb{P})$ à valeurs dans $\mathbb{R}^d$.*

1. *Si A et B sont des matrices de réels de taille (ℓ,d) et C un vecteur de $\mathbb{R}^d$, alors:*
$$\mathbb{E}\big[A\,X+B\,Y+C\big]=A\,\mathbb{E}[X]+B\,\mathbb{E}[Y]+C.$$

2. *Si A est une matrice de réels de taille (ℓ,d) et C un vecteur de $\mathbb{R}^d$,*
$$cov\big(A\,X+C\big)=A\,cov(X)\,A^\top.$$

Proof.
1. Voir la dimension 1.
2. On a $\mathbb{E}\big[A\,X+C\big]=A\,\mathbb{E}[X]+C$ d'où:
$$\mathrm{cov}\big(A\,X+C\big)=\mathbb{E}\big[\big((A\,X+C)-\mathbb{E}[A\,X+C]\big)\,\big((A\,X+C)-\mathbb{E}[A\,X+C]\big)^\top\big]=\mathbb{E}\big[A\,(X-\mathbb{E}[X])\,{}^t\big(A\,(X-\mathbb{E}[X])\big)^\top\big]=A\,\mathrm{cov}(X)\,A^\top.$$

□

2.2 Indépendance

Définition. Soit $(\Omega,\mathcal{A},\mathbb{P})$ un espace de probabilité et $I\subset\mathbb{N}$.

- Soit $(A_i)_{i\in I}$ une famille 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 $K\subset I$,
$$\mathbb{P}\Big(\bigcap_{i\in K}A_i\Big)=\prod_{i\in K}\mathbb{P}(A_i).$$

- 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 $(X_i)_{i\in I}$ une famille de variables aléatoires sur $(\Omega,\mathcal{A},\mathbb{P})$ 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, ou encore pour tous les sous-ensembles finis $K\subset I$:
$$\mathbb{P}\big(\bigcap_{i\in K}X_i\in B_i\big)=\prod_{i\in K}\mathbb{P}(X_i\in B_i)\quad\text{pour tous } B_i\in\mathcal{B}(\mathbb{R}).$$

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

Proof. Par le théorème du transport et Fubini, il est clair que pour toute famille $(B_i)_{1\le i\le n}$ de boréliens de $\mathcal{B}(\mathbb{R})$ alors:
$$\mathbb{P}(X_1\in B_1,\dots,X_n\in B_n)=\int_{B_1\times\cdots\times B_n}d\mathbb{P}_{(X_1,\dots,X_n)}(x_1,\dots,x_n)=\prod_{i=1}^n\int_{B_i}d\mathbb{P}_{X_i}(x_i)=\prod_{i=1}^n\mathbb{P}(X_i\in B_i).$$

□

Proposition. *Si $(X_i)_{i\in I}$ sont des variables aléatoires indépendantes sur $(\Omega,\mathcal{A},\mathbb{P})$. 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*

$$\mathbb{E}\Big[\prod_{j\in J} g_j(X_j)\Big]=\prod_{j\in J}\mathbb{E}[g_j(X_j)].$$

Proof. $\Longrightarrow$ Par le théorème du transport, puis par Fubini,

$$\mathbb{E}\Big[\prod_{j\in J} g_j(X_j)\Big]=\int_{\mathbb{R}^{|J|}}\prod_{j\in J} g_j(x_j)\,d\mathbb{P}_{(X_j)_{j\in J}}((dx_j)_{j\in J})=\prod_{j\in J}\int_{\mathbb{R}} g_j(x_j)\,d\mathbb{P}_{X_j}(dx_j)=\prod_{j\in J}\mathbb{E}[g_j(X_j)].$$

$\Longleftarrow$ On prenant le cas particulier $g_j(x_j)=\mathbb{I}_{x_j\in B_j}$, avec (B_j) des boréliens de $\mathcal{B}(\mathbb{R})$ on retombe sur la définition de l'indépendance. $\square$

Corollaire. *$(X_1,\cdots,X_d)$ sont des variables aléatoires indépendantes si et seulement si pour tout $(t_1,\cdots,t_d)\in\mathbb{R}^d$,*

$$\phi_{(X_1,\cdots,X_d)}(t_1,\cdots,t_d)=\prod_{j=1}^d\phi_{X_j}(t_j).$$

Proof. $\Longrightarrow$ Comme $\phi_{(X_1,\cdots,X_d)}(t_1,\cdots,t_d)=\mathbb{E}\big[e^{i\,\sum_{i=1}^d t_jX_j}\big]=\mathbb{E}\big[\prod_{j=1}^d e^{i\,t_jX_j}\big]$, on utilise la caractérisation précédente de l'indépendance avec les $g_j(x)=e^{i\,x}$.
$\Longleftarrow$ D'après la formule de caractérisation de la loi par la fonction caractéristique,

$$\begin{aligned}\mathbb{P}\big(X\in[a_1,b_1]\times\cdots\times[a_d,b_d]\big)&=\lim_{T\to\infty}\prod_{j=1}^d\Big(\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-it_ja_j}-e^{-it_jb_j}}{it_j}\Big)\phi_X(t_1,\ldots,t_d)\,dt_1\ldots dt_d\\&=\lim_{T\to\infty}\prod_{j=1}^d\Big(\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-it_ja_j}-e^{-it_jb_j}}{it_j}\phi_{X_j}(t_j)dt_j\Big)=\prod_{j=1}^d\mathbb{P}\big(X_j\in[a_j,b_j]\big),\end{aligned}$$

soit la caractérisation de l'indépendance pour presque tous les pavés fermés. $\square$

Corollaire. *Soit $X=(X_1,\ldots,X_n)$ un vecteur aléatoire défini sur $(\Omega,\mathcal{A},\mathbb{P})$ à valeurs dans $\mathbb{R}^d$ dont la mesure de probabilité $\mathbb{P}_X$ est absolument continue par rapport à la mesure de Lebesgue λ_d sur $\mathbb{R}^d$ avec pour densité f_X. Alors:*

$$(X_1,\cdots,X_n)\ \textit{v.a. indépendantes}\iff f_X(x_1,\ldots,x_n)=\prod_{i=1}^d f_{X_i}(x_i)\ \textit{pour tout}\ (x_1,\ldots,x_n)\in\mathbb{R}^n.$$

Proof. On utilise la formule d'inversion pour la fonction caractéristique:

$$f_X(x_1,\ldots,x_n)=\frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\phi_X(u_1,\ldots,u_n)\,e^{-i(u_1x_1+\cdots+u_nx_n)}du_1\ldots du_n=\frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\prod_{j=1}^n\phi_{X_j}(u_j)\,e^{-i\,u_jx_j}du_1\ldots du_n$$

d'après la caractérisation précédente. Il ne reste plus qu'à utiliser Fubini et à réécrire l'intégrale multiple comme un produit d'intégrales simples. $\square$

Corollaire. *Deux vecteurs aléatoires définis sur $(\Omega,\mathcal{A},\mathbb{P})$ sont indépendants si et seulement si toute combinaison linéaire de l'un est indépendante de toute combinaison linéaire de l'autre.*

Corollaire. *Soit $X=(X_1,\ldots,X_n)$ un vecteur aléatoire défini sur $(\Omega,\mathcal{A},\mathbb{P})$ à valeurs dans $\mathbb{R}^d$. On suppose que $\mathbb{E}\big[\|X\|^2\big]<\infty$ et $(X_1,\ldots,X_n)$ mutuellement indépendantes. Alors:*

$cov(X)$ est une matrice diagonale avec les $var(X_i)$ sur sa diagonale.

Proof. On utilise ici le fait que $\mathrm{cov}(X)=\big(\mathrm{cov}(X_i,X_j)\big)_{1\le i,j\le n}$ et $\mathrm{cov}(X_i,X_j)=0$ si $i\neq j$. $\square$

Remarque: Il est bien connu que $\mathrm{cov}(X_1, X_2) = 0$ n'entraîne pas que X_1 et X_2 soient indépendantes. Exemple: $X_1 \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1)$ et $X_2 = X_1^2$.

Propriété. *Soit X et Y deux v.a.* **indépendantes** *et de loi absolument continues par rapport à la mesure de Lebesgue λ_1, de densités f_X et f_Y. Alors $Z = X + Y$ est une v.a. absolument continue par rapport à la mesure de Lebesgue λ_1 de densité f_Z et*

$$f_Z(z) = \int_{\mathbb{R}} f_X(t)\, f_Y(z-t)\, dt = \int_{\mathbb{R}} f_Y(t)\, f_X(z-t)\, dt \quad \textit{pour tout } z \in \mathbb{R}.$$

f_Z est appelée **produit de convolution de f_X et f_Y**.

Proof. On peut écrire en utilisant Funbini et des changements de variables, que pour tout $z \in \mathbb{R}$,

$$F_{X+Y}(z) = \mathbb{P}(X+Y \leq z) = \int_{-\infty}^{\infty} f_Y(y) \int_{-\infty}^{z-y} f_X(x)\, dx\, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{z} f_Y(y) f_X(x'+y)\, dx'\, dy = \int_{-\infty}^{z} \Big(\int_{-\infty}^{\infty} f_Y(y'-x') f_X(x')\, dx' \Big)\, dy',$$

d'où le résultat. $\square$

Remarque: Attention l'hypothèse d'indépendance est nécessaire pour obtenir le produit de convolution.