

THEORIE DES PROBABILITES

Cours rédigé par Samuel Herrmann

Table des matières

1	Les bases du calcul des probabilités	5
1.1	Le triplet fondamental	5
1.2	Indépendance et conditionnement	7
1.3	Variables aléatoires	8
1.4	Lois discrètes classiques	9
1.5	Lois continues classiques	12
1.6	Somme de deux variables aléatoires indépendantes	18
1.7	Convergences stochastiques	21
2	Fonctions caractéristiques	25
2.1	Fonctions caractéristiques : définition et propriétés	25
2.2	Application aux théorèmes limites	32
3	Vecteurs aléatoires gaussiens	39
3.1	Variables gaussiennes à valeurs réelles	39
3.2	Espérance et covariance de v.a. à valeurs vectorielles	42
3.3	Vecteurs aléatoires gaussiens	44
3.4	Convergence vers la loi normale	48
3.5	Loi du chi-deux	49
3.6	Application à la statistique	52
4	Conditionnement	59
4.1	Espérance conditionnelle	59
4.2	Loi conditionnelle	67
4.3	Le cas gaussien	73
4.4	Introduction aux martingales	76

Chapitre 1

Les bases du calcul des probabilités

L'idée principale de la théorie des probabilités est de maîtriser la notion de hasard.

1.1 Le triplet fondamental

La théorie des probabilités repose sur une brique de base : le triplet fondamental $(\Omega, \mathcal{A}, \mathbb{P})$. Les trois éléments qui le composent sont :

*Ω est un ensemble non vide, appelé **ensemble fondamental** et dont les éléments ω sont les épreuves (on utilisera aussi le terme scenario)*

Les épreuves représentent les résultats d'une expérience aléatoire.

Exemples :

- lancer de 3 pièces de monnaie de valeurs différentes (10, 20 et 50 centimes). L'ensemble fondamental est donc $\{10, 20, 50\} \times \{P, F\}$.
- La température de ce matin à 8 heures, $\Omega = \mathbb{R}$.

$\mathcal{A}$: la σ -algèbre ou tribu est une classe de sous-ensembles de Ω , appelés évènements.

L'évènement A se réalise en l'épreuve ω si $\omega \in A$. Il s'agit de l'information concernant l'expérience aléatoire qui est disponible.

Exemple : si on imagine qu'une note à l'examen a un caractère aléatoire, alors l'ensemble fondamental est $\{0, 1, 2, \dots, 20\}$. Sur le tableau d'affichage n'apparaît que la mention *admis* ou *non admis* et c'est la seule info disponible. La tribu est alors $\mathcal{A} = \{\emptyset, \{0, 1, \dots, 9\}, \{10, 11, \dots, 20\}, \Omega\}$. Dans ce cas, connaître la note précisément n'est pas une information disponible.

Par définition, la tribu vérifie :

1. $\Omega \in \mathcal{A}$
2. $A \in \mathcal{A} \rightarrow A^c \in \mathcal{A}$

3. Soit une suite $(A_n)_{n \geq 1}$ d'évènements alors

$$\bigcup_{n \geq 1} A_n \in \mathcal{A}.$$

De nombreuses tribus sont donc associées à un ensemble fondamental donné mais certaines d'entre elles jouent un rôle important : ce sont les *tribus engendrée par une classe de parties*. En effet on s'intéresse à la plus petite tribu qui contienne un ensemble de parties.

Soit $\mathcal{C}$ un ensemble de parties de Ω . Alors il existe une et une seule tribu $\sigma(\mathcal{C})$ telle que

- $\sigma(\mathcal{C}) \supset \mathcal{C}$
- si σ' est une tribu contenant $\mathcal{C}$ alors elle contient aussi $\sigma(\mathcal{C})$.

Exemples :

- Le compteur de la voiture ne fonctionne plus, seule info disponible : le PV que l'on reçoit par la poste (pas d'amende, ≤ 3 points (40 km/h), > 3 points)
- si la classe de parties est déjà une tribu alors $\sigma(\mathcal{C}) = \mathcal{C}$
- la tribu borélienne $\mathcal{B}(\mathbb{R})$ est engendrée par la classe des ouverts ou celle des fermés, ou celle des intervalles ouverts, ou celle des intervalles fermés,...

$\mathbb{P}$ est une pondération de l'ensemble des évènements de la tribu $\mathcal{A}$. Il s'agit d'une mesure de probabilité.

Cette pondération est souvent inspirée de l'expérimentation. Par exemple si on considère la tribu $\mathcal{P}(\Omega)$ avec Ω est un ensemble fini, alors pour chaque épreuve ω_k on calcule le rapport $\frac{N_k}{N}$ où N_k est le nombre de fois où ω_k apparaît au cours des N expériences répétées dans des conditions similaires. La limite quand N tend vers l'infini correspond à la notion intuitive de probabilité.

En d'autres termes, $\mathbb{P}$ doit satisfaire aux conditions suivantes :

- $0 \leq \mathbb{P}(A) \leq 1$ pour tout $A \in \mathcal{A}$
- $\mathbb{P}(\Omega) = 1$
- Soit $(A_n)_{n \geq 1}$ une suite d'évènements disjoints alors la propriété de σ -additivité est satisfaite

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mathbb{P}(A_n).$$

On remarque de suite que la probabilité n'est définie que sur l'information disponible donnée par la tribu $\mathcal{A}$. **Pour connaître la probabilité d'un évènement il faut donc impérativement que cet évènement appartienne à la tribu.**

On rappelle alors une propriété essentielle d'une mesure de probabilité : la **continuité**. Si (A_n) est une suite monotone d'évènements appartenant à $\mathcal{A}$ alors

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

1.2 Indépendance et conditionnement par un événement non négligeable

Si on dispose d'une information particulière (l'évènement B s'est réalisé) alors il est possible d'actualiser la probabilité que l'on donne à tout autre évènement A .

Définition 1 Soient A et B deux évènements de la tribu $\mathcal{A}$. La probabilité conditionnelle de A sachant B est notée $\mathbb{P}(A|B)$ et est donnée par :

$$\mathbb{P}(A|B) = \begin{cases} \mathbb{P}(A \cap B)/\mathbb{P}(B) & \text{si } \mathbb{P}(B) > 0. \\ \mathbb{P}(A) & \text{sinon.} \end{cases}$$

On sait que l'évènement B s'est réalisé, il est donc naturel d'associer à l'évènement A un poids proportionnel à $\mathbb{P}(A \cap B)$. Le quotient vient juste d'une renormalisation pour rester dans le cadre des probabilités.

Exemple : Dans une famille qui comporte deux enfants, l'un est une fille, la probabilité que l'autre soit un garçon est alors de $2/3$.

Parfois on connaît la probabilité conditionnelle, on utilisera alors plutôt la forme :

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

Le résultat à retenir est la probabilité conditionnelle dans le cadre d'une partition de l'ensemble fondamental. C'est la formule de Bayes.

Proposition 1 Soit $B_1, \dots, B_k$ une partition de Ω incluse dans $\mathcal{A}$ et soit $A \in \mathcal{A}$ alors pour tout $1 \leq i \leq k$ on a

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{j=1}^k \mathbb{P}(A|B_j)\mathbb{P}(B_j)}$$

Introduisons la notion d'indépendance. Un évènement A sera indépendant d'un autre évènement B si la probabilité conditionnelle n'est autre que la probabilité de départ. En d'autres termes, l'information supplémentaire concernant la réalisation de l'évènement B ne sert absolument à rien.

Définition 2 Deux évènements A et B de $\mathcal{A}$ sont dits indépendants si

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

On peut évidemment prolonger la notion d'indépendance aux classes d'évènements. Ainsi deux classes $\mathcal{C}_1$ et $\mathcal{C}_2$ sont indépendantes si pour tout $A \in \mathcal{C}_1$ et pour tout $B \in \mathcal{C}_2$, les évènements A et B sont indépendants. Finalement cette notion peut se généraliser aux tribus. Deux tribus pourront donc être indépendantes.

1.3 Variables aléatoires

A partir de l'expérience aléatoire, on peut construire des variables dont la valeur dépend du résultat de l'expérience. Il s'agit donc tout simplement d'une fonction des épreuves.

Définition 3 Une variable aléatoire est une fonction $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$, où $(E, \mathcal{E})$ est un espace mesurable (i.e. $\mathcal{E}$ est une tribu pour l'ensemble E), qui vérifie

$$\forall A \in \mathcal{E}, \quad X^{-1}(A) \in \mathcal{A}.$$

Exemples :

— lancer des 3 dés $\Omega = \{1, 2, 3, 4, 5, 6\}^3$, X_1 est le résultat du premier dé c'est-à-dire

$$X_1((\omega_1, \omega_2, \omega_3)) = \omega_1.$$

X_1 est alors une projection.

— Dans le même cadre, on peut considérer $X_2((\omega_1, \omega_2, \omega_3)) = \omega_1 + \omega_2 + \omega_3$.

Pour ces deux exemples, il faut déterminer $(E, \mathcal{E})$ pour que les variables présentées soient bien des variables aléatoires au sens de la définition ci-dessus.

Dans la plupart des cas, $(E, \mathcal{E})$ correspondra à $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ ou à $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Jusqu'ici nous n'avons pas eu besoin de la probabilité $\mathbb{P}$ pour définir une variable aléatoire. Par contre pour bien la décrire, nous allons nous en servir.

Proposition 2 Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (E, \mathcal{E})$ une variable aléatoire. Alors l'application $\mathbb{P}_X : \mathcal{E} \rightarrow [0, 1]$ définie pour tout ensemble B de $\mathcal{E}$ par :

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

est une mesure de probabilité. On l'appelle **loi de probabilité**.

Notons que nous utiliserons :

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

En fait la loi de probabilité est l'image de la probabilité $\mathbb{P}$ par l'application X . Plusieurs variables aléatoires peuvent en fait avoir la même loi de probabilité dans des contextes très différents. Pour pouvoir identifier ces lois, nous introduisons un outil fondamental : la fonction de répartition.

Définition 4 Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$. La **fonction de répartition** de X est donnée par

$$F_X(x) = \mathbb{P}_X(-\infty, x] = \mathbb{P}(X \leq x).$$

Quelques propriétés :

- $0 \leq F(x) \leq 1$ pour tout $x \in \mathbb{R}$.
- F est une fonction croissante (au sens large), continue à droite en tout point x de $\mathbb{R}$.
- $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.

Notons que la réciproque est vraie : si les trois propriétés précédentes sont satisfaites par une fonction F alors il s'agit de la fonction de répartition d'une certaine probabilité.

La fonction de répartition caractérise la loi : deux variables aléatoires ayant la même loi ont la même fonction de répartition.

Il devient alors essentiel de pouvoir calculer la fonction de répartition pour une loi donnée mais cet exercice peut s'avérer compliqué voire inextricable dans certaines situations (v.a. gaussiennes par exemple).

Notons également la propriété suivante :

$$\lim_{h \rightarrow 0} F_X(x+h) - F_X(x-h) = \mathbb{P}(X=x), \quad \text{pour tout } x \in \mathbb{R}.$$

Par conséquent, F_X est continue ssi $\mathbb{P}(X=x) = 0$ pour tout $x \in \mathbb{R}$. Et si F_X n'est pas continue en x , le saut de discontinuité est égal à $\mathbb{P}(X=x)$.

Comme dans le paragraphe précédent, il fut question d'évènements indépendants, il est maintenant possible de définir la notion d'indépendance pour des variables aléatoires.

Définition 5 Deux variables aléatoires X et Y définies sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans $(E, \mathcal{E})$ respectivement $(F, \mathcal{F})$ sont indépendantes, si

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \quad \text{pour tout } A \in \mathcal{E}, B \in \mathcal{F}.$$

En d'autres termes, la loi de probabilité du couple (X, Y) notée $\mathbb{P}_{(X,Y)}$ est définie par le produit des mesures :

$$\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y.$$

Une conséquence du lemme des classes monotones est que deux variables aléatoires réelles X et Y sont indépendantes ssi leurs fonctions de répartition satisfont :

$$\mathbb{P}(X \leq x, Y \leq y) = F_X(x)F_Y(y).$$

1.4 Lois discrètes classiques

Rappel sur les séries

Soient F et G deux ensembles dénombrables.

— si $(a_{(x,y)})_{(x,y) \in F \times G}$ est une famille de nombres positifs alors

$$\sum_{(x,y) \in F \times G} a_{(x,y)} = \sum_{x \in F} \left(\sum_{y \in G} a_{(x,y)} \right) = \sum_{y \in G} \left(\sum_{x \in F} a_{(x,y)} \right).$$

— si $(a_{(x,y)})_{(x,y) \in F \times G}$ est sommable au sens où $\sum_{(x,y) \in F \times G} |a_{(x,y)}| < +\infty$, alors l'égalité précédente reste vraie.

Dans la théorie des probabilités discrètes, il suffit bien souvent d'utiliser la théorie des séries numériques puisque la plupart des lois usuelles concernent des variables aléatoires à valeurs dans $\mathbb{N}$. Toutefois on ne peut se restreindre à cette situation pour un cadre général, il est alors nécessaire d'introduire la notion de famille sommable comme indiqué ci-dessus.

Variables aléatoires discrètes

Dans un premier temps, on définit les variables discrètes et dans un deuxième temps on présente les lois classiques discrètes.

Définition 6 Une variable aléatoire discrète est une application $X : \Omega \rightarrow F$ tel qu'il existe $E \subset F$ où E est un ensemble fini ou dénombrable vérifiant $\mathbb{P}(X \in E) = 1$. Par abus de langage la famille $(\mathbb{P}(X = x))_{x \in E}$ s'appelle la loi de X et on dira que X est à valeurs dans E .

A ne pas oublier : évidemment pour que la loi soit bien définie il faut que l'évènement $\{X = x\}$, dont on calcule la probabilité, soit dans la tribu $\mathcal{A}$ du triplet fondamental.

Pour caractériser les lois discrètes, il suffit donc de déterminer la famille $(\mathbb{P}(X = x))_{x \in E}$, il n'est pas forcément utile d'essayer de calculer la fonction de répartition qui n'apportera rien de plus. De même, pour vérifier l'**indépendance** de deux variables aléatoires discrètes X à valeurs dans E et Y à valeurs dans F , il suffira de montrer que

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y), \quad \forall x \in E, \forall y \in F.$$

Certains paramètres permettent de décrire plus précisément les lois de probabilité. Deux paramètres ont une importance primordiale : l'**espérance mathématique** et la **variance**. L'espérance permet pour une variable aléatoire de trouver une valeur moyenne.

Soit X une variable aléatoire discrète de Ω dans E . Elle est dite intégrable si

$$\sum_{x \in E} |x| \mathbb{P}(X = x) < +\infty$$

et alors l'espérance mathématique notée $\mathbb{E}[X]$ est définie par

$$\mathbb{E}[X] = \sum_{x \in E} x \mathbb{P}(X = x).$$

Pour toute fonction $f : E \rightarrow F$, il est possible de munir F d'une tribu adaptée de sorte que $f(X)$ soit une variable aléatoire. Il devient alors possible de calculer son espérance :

La variable aléatoire $f(X)$ est intégrable si et seulement si $\sum_{x \in E} |f(x)| \mathbb{P}(X = x) < +\infty$ et ainsi

$$\mathbb{E}[f(X)] = \sum_{x \in E} f(x) \mathbb{P}(X = x).$$

La fonction particulière $f(x) = x^2$ permet de définir la variance d'une loi de probabilité :

Une variable aléatoire X est dite de carré intégrable si X^2 est intégrable. Dans ce cas on définit la variance par

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

La propriété d'indépendance des variables aléatoires discrètes, décrite ci-dessus, combinée avec le résultat concernant les fonctions de variables aléatoires permet d'assurer le résultat suivant :

Proposition 3 Soient X et Y deux variables aléatoires discrètes. Si X et Y sont indépendantes alors pour toutes fonctions f et g telles que $f(X)$ et $g(Y)$ soient intégrables, $f(X)g(Y)$ est intégrable et satisfait :

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].$$

La réciproque est vraie pour toutes fonctions f et g bornées.

Par conséquent, si X et Y sont deux variables aléatoires indépendantes de carré intégrable alors $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ et ainsi

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Quelques lois discrètes

Loi binomiale. Soit p un nombre réel tel que $0 \leq p \leq 1$ et $n \in \mathbb{N}$. La mesure de probabilité $\mathcal{B}(n, p)$ définie sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ ou $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ par

$$\mathcal{B}(n, p) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \delta_k,$$

où $q = 1 - p$, est appelée *loi binomiale* de paramètres (n, p) . La loi $\mathcal{B}(1, p)$ est appelée *loi de Bernoulli* de paramètre p .

La loi binomiale intervient dans le tirage avec remise, si la proportion de boules blanches dans l'urne est égale à p et si le nombre de tirages est égale à n , alors la variable aléatoire *nombre de boules blanches tirées* est une variable aléatoire binomiale de paramètres (n, p) .

Les valeurs typiques de la loi binomiale sont :

X suit une loi binomiale de paramètres (n, p) alors $\mathbb{E}[X] = np$, $\text{Var}(X) = npq$.

Loi géométrique. Soit p un réel compris entre 0 et 1 et $q = 1 - p$. La mesure de probabilité définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ par

$$\sum_{k=1}^{\infty} pq^{k-1} \delta_k$$

est appelée *loi géométrique de paramètre p* et notée $\mathcal{G}(p)$.

X suit une loi géométrique de paramètre p alors $\mathbb{E}[X] = \frac{1}{p}$, $\text{Var}(X) = \frac{q}{p^2}$.

Application : un joueur procède à une suite de parties indépendantes de *pile* ou *face* et décide de s'arrêter dès la première apparition de *pile*. Le nombre de parties qu'il faudra jouer, noté X , suit alors une loi géométrique.

Loi de Poisson. Soit λ un nombre réel strictement positif. La mesure de probabilité π_λ définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ par

$$\pi_\lambda = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k,$$

est appelée *loi de Poisson de paramètre λ* . Cette loi est notée $\mathcal{P}(\lambda)$.

On peut noter que, pour tout entier $k \geq 1$ fixé, on a

$$\lim_{n \rightarrow \infty} \binom{k}{n} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Ainsi si X suit une loi binomiale de paramètres (n, p) de sorte que n et p soient liés entre eux par la relation $np = \lambda > 0$, alors pour n assez grand la probabilité que X soit égale à k est approximativement égale à $e^{-\lambda} \lambda^k / k!$ qui est la probabilité pour qu'une variable aléatoire de Poisson de paramètre λ soit égale à k . La loi de Poisson est donc une loi des *événements rares*.

X suit une loi de Poisson de paramètre λ alors $\mathbb{E}[X] = \lambda$, $\text{Var}(X) = \lambda$.

1.5 Lois continues classiques

Les lois discrètes ne permettent pas de décrire toutes les expériences aléatoires. De nombreuses situations font intervenir des variables dont l'ensemble des valeurs atteintes n'est ni fini ni dénombrable. Par exemple, votre voiture montre des signes de fatigue. Vous décidez de continuer à rouler car vous n'avez pas les moyens d'en acheter une neuve. La distance que vous parcourrez alors est une variable aléatoire à valeurs dans $\mathbb{R}_+$ qui n'est pas une valeur discrète car $\mathbb{P}(\text{distance} = x) = 0$ pour tout x .

Le cadre des variables à densité est adapté à ce cadre, les sommes manipulées dans le cadre des variables aléatoires discrètes sont alors remplacées par des intégrales.

Rappels concernant les intégrales multiples

Le premier résultat essentiel est le **théorème de Fubini**. On considère $f : \mathbb{R}^2 \rightarrow \mathbb{R}$.

— si f est une fonction positive alors

$$\int_{\mathbb{R}^2} f(x, y) dx dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dy \right) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dx \right) dy.$$

— si f est intégrable au sens où $\int_{\mathbb{R}^2} |f(x, y)| dx dy < \infty$ alors l'égalité ci-dessus est encore satisfaite.

Ce résultat de permutation des intégrales doit être complété par un résultat de changement de variables.

Soit φ une bijection continuellement différentiable ainsi que son inverse φ^{-1} d'un ouvert O de $\mathbb{R}^d$ à un ouvert O' de $\mathbb{R}^d$, $f : \mathbb{R}^d \rightarrow \mathbb{R}$ bornée et $g : \mathbb{R}^d \rightarrow \mathbb{R}$ intégrable. Alors

$$\int_O f(\varphi(x))g(x)dx = \int_{O'} f(y)g(\varphi^{-1}(y))|\text{Jac } \varphi^{-1}(y)| dy$$

où

$$\text{Jac } \varphi(y) = \text{Det} \left(\frac{\partial \varphi_i}{\partial y_j}(y), 1 \leq i, j \leq d \right).$$

Ces deux résultats essentiels issus de la théorie de l'intégration vont permettre d'appréhender les variables aléatoires à densité.

Variables continues

Il s'agit de définir les variables à densité et d'en donner les premières propriétés.

Définition 7 *La variable aléatoire $X : \Omega \rightarrow \mathbb{R}$ est continue et possède la densité $f : \mathbb{R} \rightarrow \mathbb{R}_+$ si pour tout a et b appartenant à $\mathbb{R}$ et vérifiant $a < b$ on a*

$$\mathbb{P}(a < X \leq b) = \int_a^b f(x) dx.$$

Il faut évidemment exiger que tout événement de la forme $\{a < X \leq b\}$ appartienne à la tribu $\mathcal{A}$.

On remarque très facilement que la densité satisfait les propriétés suivantes :

- $f(x) \geq 0$ pour tout $x \in \mathbb{R}$
- $\int_{\mathbb{R}} f(x)dx = 1$.

La réciproque est vraie également : si une fonction vérifie les deux propriétés précédemment citées alors il s'agit d'une densité de probabilité.

La définition met en évidence que la loi de probabilité associée à la variable X est la mesure

$$f(x) dx$$

où dx représente la mesure de Stieltjes-Lebesgue.

Proposition 4 *Si la fonction de répartition F_X de la variable aléatoire X est continue et $\mathcal{C}^1$ par morceaux alors X possède la densité F'_X .*

Il est possible de caractériser une variable aléatoire à densité par la *méthode de la fonction muette*. Pour ce faire on introduit la notion d'espérance mathématique pour une variable aléatoire générale : soit X une v.a. définie sur $(\Omega, \mathcal{A}, \mathbb{P})$. Si X est intégrable, on appelle espérance de X le nombre réel :

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

On dit que X est centrée si elle est intégrable et $\mathbb{E}[X] = 0$. Le théorème de transport, qui dit la chose suivante : si X est une variable aléatoire réelle à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et si ϕ est une fonction borélienne de $\mathbb{R}$ dans $\mathbb{R}$ à valeurs positives alors

$$\mathbb{E}[\phi(X)] = \int_{\Omega} \phi \circ X(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} \phi(x) d\mathbb{P}_X(x)$$

Si ϕ est à valeurs quelconque et si $\phi(X)$ est intégrable alors l'égalité précédente reste vraie. Ce sera notamment le cas pour les fonctions ϕ bornées.

En appliquant le théorème de transport aux variables à densité, on obtient donc :

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

Il est possible alors d'utiliser cette formule pour caractériser les variables à densité :

Proposition 5 *La variable aléatoire X possède la densité f si et seulement si, pour toute fonction ϕ bornée,*

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

Notons finalement que la caractérisation de l'indépendance en utilisant les fonctions muettes présentée dans la Proposition 3 est vraie également pour les variables continues.

Théorème 1 *Soient X_1 et X_2 deux v.a. Alors X_1 et X_2 sont indépendantes si et seulement si*

$$\mathbb{E}(h_1(X_1)h_2(X_2)) = \mathbb{E}(h_1(X_1))\mathbb{E}(h_2(X_2)). \quad (1.1)$$

Pour toutes fonctions h_1, h_2 boréliennes positives (ou bornées).

Preuve : 1) Supposons X_1, X_2 soient indépendantes. Soient h_1, h_2 deux fonctions boréliennes positives et $h(x, y) = h_1(x)h_2(y)$. On a,

$$\mathbb{E}[h_1(X_1)h_2(X_2)] = \mathbb{E}[h(X_1, X_2)] = \int h(x, y) d\mathbb{P}_{(X_1, X_2)}(x, y).$$

Mais X_1, X_2 sont indépendantes, donc $\mathbb{P}_{(X_1, X_2)} = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$ et,

$$\mathbb{E}[h_1(X_1)h_2(X_2)] = \int h(x, y) d\mathbb{P}_{X_1}(x) d\mathbb{P}_{X_2}(y) = \int h_1(x)h_2(y) d\mathbb{P}_{X_1}(x) d\mathbb{P}_{X_2}(y).$$

On applique le théorème de Fubini, il vient,

$$\mathbb{E}[h_1(X_1)h_2(X_2)] = \left(\int h_1(x) d\mathbb{P}_{X_1}(x) \right) \left(\int h_2(y) d\mathbb{P}_{X_2}(y) \right) = \mathbb{E}[h_1(X_1)] \mathbb{E}[h_2(X_2)].$$

2) Réciproquement, on choisit $h_1 = \mathbb{1}_{A_1}$ et $h_2 = \mathbb{1}_{A_2}$ où A_1 et A_2 sont deux boréliens, et on applique (1.1) :

$$\begin{aligned}\mathbb{E}[\mathbb{1}_{A_1}(X_1)\mathbb{1}_{A_2}(X_2)] &= \mathbb{P}(X_1 \in A_1; X_2 \in A_2) = \mathbb{E}[\mathbb{1}_{A_1}(X_1)] \mathbb{E}[\mathbb{1}_{A_2}(X_2)] \\ &= \mathbb{P}(X_1 \in A_1) \mathbb{P}(X_2 \in A_2).\end{aligned}$$

Ce qui signifie que X_1 et X_2 sont deux v.a. indépendantes. $\square$

L'indépendance de deux variables aléatoires X et Y absolument continues peut aussi se caractériser sur l'expression de la densité du couple (X, Y) . Pour cela nous présentons un premier résultat sur les lois marginales d'un couple de variables aléatoires.

Proposition 6 *Si $Z = (X, Y)$ admet une densité f alors X (resp. Y) admet une densité f_1 (resp. f_2), avec*

$$f_1(x) = \int_{\mathbb{R}} f(x, y) dy; \quad f_2(y) = \int_{\mathbb{R}} f(x, y) dx. \quad (1.2)$$

Preuve : Par symétrie il suffit de montrer que f_1 est la densité de X .

D'après le théorème de Fubini, l'application f_1 est mesurable (ou borélienne) et positive. Soit A un borélien de $\mathbb{R}$. Alors,

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}((X, Y) \in A \times \mathbb{R}) = \int_{\mathbb{R}^2} \mathbb{1}_A(x) f(x, y) dx dy.$$

On applique à nouveau le théorème de Fubini, il vient :

$$\mathbb{P}_X(A) = \int_{\mathbb{R}} \mathbb{1}_A(x) \left(\int_{\mathbb{R}} f(x, y) dy \right) dx = \int_{\mathbb{R}} \mathbb{1}_A(x) f_1(x) dx.$$

$\square$

Proposition 7 *On suppose que (X, Y) admet f (fonction à deux variables) comme densité. Les v.a. X et Y sont indépendantes si et seulement si,*

$$f(x, y) = g_1(x)g_2(y), \quad \text{presque partout,} \quad (1.3)$$

où g_1 et g_2 sont deux fonctions boréliennes positives.

Preuve : a) Supposons les deux v.a. X et Y indépendantes. D'après la Proposition 6, X (resp. Y) admet une densité f_1 (resp. f_2). Soient A et B deux boréliens. On utilise l'indépendance de X et Y , il vient,

$$\begin{aligned}\mathbb{P}((X, Y) \in (A \times B)) &= \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B) \\ &= \left(\int \mathbb{1}_A(x) f_1(x) dx \right) \left(\int \mathbb{1}_B(y) f_2(y) dy \right)\end{aligned}$$

Une application du théorème de Fubini conduit à,

$$\mathbb{P}((X, Y) \in (A \times B)) = \mathbb{P}_{(X, Y)}(A \times B) = \int \mathbb{1}_{A \times B}(x, y) f_1(x) f_2(y) dx dy.$$

$\mathbb{P}_{(X, Y)}$ désignant la loi de (X, Y) .

Pour tout borélien C on pose

$$\mathbb{Q}(C) = \int \mathbb{1}_C(x, y) f_1(x) f_2(y) dx dy.$$

Il est facile de vérifier que $\mathbb{Q}$ est une probabilité. Donc $\mathbb{P}_{(X, Y)}$ et $\mathbb{Q}$ sont égales sur les pavés $A \times B$. L'utilisation du théorème d'unicité des probabilités assure que deux probabilités qui coïncident sur les pavés sont égales partout :

$$\mathbb{P}((X, Y) \in C) = \int \mathbb{1}_C(x, y) f_1(x) f_2(y) dx dy, \text{ pour tout borélien } C.$$

D'après l'unicité de la densité, on a l'égalité : $f(x, y) = f_1(x) f_2(y)$, presque sûrement.

b) Réciproquement, supposons la relation (1.3) réalisée. Soient A et B deux boréliens.

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}((X, Y) \in (A \times B)) = \int \mathbb{1}_{A \times B}(x, y) g_1(x) g_2(y) dx dy.$$

On applique à nouveau le théorème de Fubini, il vient,

$$\mathbb{P}(X \in A, Y \in B) = \left(\int \mathbb{1}_A(x) g_1(x) dx \right) \left(\int \mathbb{1}_B(y) g_2(y) dy \right). \quad (1.4)$$

En particulier si $A = B = \mathbb{R}$, l'égalité précédente devient $1 = \lambda_1 \lambda_2$, où l'on a posé,

$$\lambda_1 = \int_{\mathbb{R}} g_1(x) dx \text{ et } \lambda_2 = \int_{\mathbb{R}} g_2(y) dy.$$

On introduit : $f_1 = g_1/\lambda_1$ et $f_2 = g_2/\lambda_2$. On divise (1.4) par le facteur $1 = \lambda_1 \lambda_2$,

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B) &= \left[\left(\int \mathbb{1}_A(x) g_1(x) dx \right) / \lambda_1 \right] \left[\left(\int \mathbb{1}_B(y) g_2(y) dy \right) / \lambda_2 \right] \\ &= \left(\int \mathbb{1}_A(x) f_1(x) dx \right) \left(\int \mathbb{1}_B(y) f_2(y) dy \right) \end{aligned}$$

En particulier si $B = \mathbb{R}$,

$$\mathbb{P}(X \in A) = \int \mathbb{1}_A(x) f_1(x) dx.$$

D'une manière analogue, $\mathbb{P}(Y \in B) = \int \mathbb{1}_B(y) f_2(y) dy$. On peut donc écrire (1.4) de la façon suivante :

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B).$$

Ce qui signifie que les v.a. X et Y sont indépendantes. □

Quelques lois continues

Loi uniforme. Une variable X à valeurs réelles, suit une loi uniforme sur $[a, b]$ avec $a < b$, notée $\mathcal{U}([a, b])$, si sa densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ est

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

Si X suit une loi uniforme sur $[a, b]$ alors $\mathbb{E}[X] = \frac{a+b}{2}$ et $\text{Var}(X) = \frac{(b-a)^2}{12}$.

Loi Gamma. Une variable aléatoire X à valeurs positives suit une loi Gamma de paramètres $p > 0$ et $\theta > 0$ notée $\gamma(p, \theta)$, si sa densité est :

$$f(x) = \frac{\theta^p}{\Gamma(p)} e^{-\theta x} x^{p-1} 1_{[0, \infty[}(x),$$

où l'intégrale $\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx$. La loi $\gamma(1, \theta)$ est appelée **loi exponentielle** de paramètre θ et notée $\mathcal{E}(\theta)$.

Si X suit une loi Gamma de paramètres (p, θ) alors $\mathbb{E}[X] = p/\theta$ et $\text{Var}(X) = p/\theta^2$. Par ailleurs la loi Gamma est stable par convolution : $\gamma(p, \theta) \star \gamma(q, \theta) = \gamma(p+q, \theta)$. Autrement dit, si X et Y sont deux v.a. indépendantes de lois respectives $\gamma(p, \theta)$ et $\gamma(q, \theta)$ alors $X+Y$ suit la loi $\gamma(p+q, \theta)$.

Loi de Gauss. Une variable aléatoire X à valeurs réelles suit une loi normale $\mathcal{N}(m, \sigma^2)$ de moyenne m et de variance σ^2 si sa densité est

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

La loi normale est stable par convolution : $\mathcal{N}(m_1, \sigma_1^2) \star \mathcal{N}(m_2, \sigma_2^2) = \mathcal{N}(m_1+m_2, \sigma_1^2+\sigma_2^2)$. Autrement dit, si X et Y sont deux variables aléatoires indépendantes de lois respectives $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$ alors $X+Y$ suit la loi $\mathcal{N}(m_1+m_2, \sigma_1^2+\sigma_2^2)$.

Loi de Cauchy. Une variable aléatoire à valeurs réelles suit une loi de Cauchy si sa densité est égale à

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}.$$

Il s'agit d'un exemple de loi telle que $|X|$ est non intégrable. L'espérance mathématique n'est donc pas définie ni la variance.

Il existe de nombreuses autres lois continues, citons notamment : la Loi **log-normale**, la loi de **Weibull**, la loi **inverse-gaussienne**, la loi **Beta**, la loi de **Fisher-Snedecor**, la loi de **Student**...

1.6 Somme de deux variables aléatoires indépendantes

Nous avons décrit dans les deux sections précédentes différentes lois de probabilité continues ou discrètes. Il est alors possible de combiner ces différentes lois en considérant la somme de deux variables aléatoires indépendantes. Le **produit de convolution** fait alors son apparition.

Définition :

Soient μ et ν deux mesures positives sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On appelle **convolée** de μ et ν la mesure positive $\mu * \nu$, qui est l'image de $\mu \otimes \nu$ par l'application qui au couple (x, y) de $\mathbb{R} \times \mathbb{R}$ associe $x + y$:

$$\int_{\mathbb{R}} f(t) d\mu * \nu(t) = \iint_{\mathbb{R}^2} f(x + y) d\mu(x) d\nu(y). \quad (1.5)$$

On vérifie que $\mu * \nu = \nu * \mu$ et que $\mu * \nu$ est une probabilité lorsque μ et ν sont deux probabilités.

Proposition 8 Soient X et Y deux v.a.r. indépendantes, de loi respectivement μ et ν . Alors $X + Y$ a pour loi $\mu * \nu$.

Preuve : Soient f une fonction borélienne, positive et $A = \mathbb{E}[f(X + Y)]$. On a

$$A = \iint_{\mathbb{R}^2} f(x + y) \mu(dx) \nu(dy) = \int_{\mathbb{R}} f(t) \mu * \nu(dt).$$

□

On examine à présent deux cas particuliers : X et Y sont discrètes, X et Y admettent une densité.

Proposition 9 Soient X et Y deux v.a.r. indépendantes, à valeurs dans $\mathbb{N}$. Alors la loi de $X + Y$ est donnée par la formule :

$$\mathbb{P}(X + Y = n) = \sum_{i=0}^n \mathbb{P}(X = i) \mathbb{P}(Y = n - i). \quad (1.6)$$

Preuve : On a : $\{X + Y = n\} = \bigcup_{i=0}^n \{X = i; X + Y = n\} = \bigcup_{i=0}^n \{X = i; Y = n - i\}$.

Les ensembles étant deux à deux disjoints et les v.a. X et Y indépendantes, on a,

$$\mathbb{P}(X + Y = n) = \sum_{i=0}^n \mathbb{P}(X = i; Y = n - i) = \sum_{i=0}^n \mathbb{P}(X = i) \mathbb{P}(Y = n - i).$$

□

Définition :

Soient f et g deux fonctions boréliennes positives, définies sur $\mathbb{R}$. On note $f * g$ la **convolée**

de f et g :

$$f * g(x) = \int_{\mathbb{R}} f(x-t)g(t)dt = \int_{\mathbb{R}} g(x-t)f(t)dt = g * f(x). \quad (1.7)$$

Remarques :

1) On vérifie que la convolution est une opération associative :

$$(f * g) * h = f * (g * h). \quad (1.8)$$

2) La convolution est une application "régularisante". Si f et g sont deux densités (i.e. f mesurables, positives et d'intégrale 1) bornées, la fonction $f * g$ est une densité et une fonction continue .

3) Le lien avec la convolution des mesures est le suivant :

$$\mu * \nu(dx) = (f * g)(x)dx, \text{ avec } \mu(dx) = f(x)dx \text{ et } \nu(dx) = g(x)dx. \quad (1.9)$$

Proposition 10 Soient X et Y deux v.a.r. indépendantes, X (resp. Y) de densité f (resp. g), alors $X + Y$ admet pour densité $f * g$.

Preuve : On calcule $A = \mathbb{E}[\varphi(X + Y)]$, où φ est une fonction borélienne positive. On sait (voir Proposition 7) que le couple (X, Y) a pour densité $f(x)g(y)$, par conséquent,

$$A = \iint_{\mathbb{R} \times \mathbb{R}} \varphi(x + y)f(x)g(y)dx dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \varphi(x + y)f(x)dx \right) g(y)dy.$$

On pose $z = x + y$ (y étant fixé),

$$A = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \varphi(z)f(z - y)dz \right) g(y)dy.$$

On applique le théorème de Fubini,

$$A = \int_{\mathbb{R}} \varphi(z) \left(\int_{\mathbb{R}} f(z - y)g(y)dy \right) dz = \int_{\mathbb{R}} \varphi(z)f * g(z)dz.$$

□

Nous donnons à présent des exemples de convolutions. Nous commençons par le cas discret.

Proposition 11 X et Y désignent deux v.a.r. indépendantes.

X de loi $\mathcal{B}(n, p)$ et Y de loi $\mathcal{B}(m, p)$, alors $X + Y$ a pour loi $\mathcal{B}(n + m, p)$. (1.10)

X de loi $\mathcal{P}(\lambda)$ et Y de loi $\mathcal{P}(\mu)$, alors $X + Y$ a pour loi $\mathcal{P}(\lambda + \mu)$. (1.11)

Preuve : 1) En ce qui concerne les lois binomiales on peut montrer directement la propriété (1.10). Il est plus rapide d'utiliser la représentation probabiliste d'une somme de v.a. de Bernoulli indépendantes.

2) On suppose que X et Y sont indépendantes, X de loi $\mathcal{P}(\lambda)$ et Y de loi $\mathcal{P}(\mu)$. On évalue :

$$\begin{aligned}\mathbb{P}(X + Y = n) &= \sum_{i=0}^n \mathbb{P}(X = i) \mathbb{P}(Y = n - i) = \sum_{i=0}^n \frac{\lambda^i}{i!} e^{-\lambda} \frac{\mu^{n-i}}{(n-i)!} e^{-\mu} \\ &= \frac{1}{n!} e^{-(\lambda+\mu)} \left(\sum_{i=0}^n \frac{n!}{i!(n-i)!} \lambda^i \mu^{n-i} \right) = \frac{1}{n!} e^{-(\lambda+\mu)} \left(\sum_{i=0}^n C_n^i \lambda^i \mu^{n-i} \right) \\ &= \frac{1}{n!} e^{-(\lambda+\mu)} (\lambda + \mu)^n.\end{aligned}$$

L'égalité précédente signifie que $X + Y$ a pour loi $\mathcal{P}(\lambda + \mu)$. □

Théorème 2 X et Y désignent deux v.a.r. gaussiennes indépendantes, X de loi $\mathcal{N}(m, \sigma^2)$ et Y de loi $\mathcal{N}(n, \tau^2)$, alors

$$X + Y \text{ a pour loi } \mathcal{N}(m + n, \sigma^2 + \tau^2). \quad (1.12)$$

Remarque. Supposons que X et Y soient deux v.a.r. gaussiennes. Dès que l'on sait que $X + Y$ est une v.a.r. gaussienne, on calcule les paramètres sans difficulté :

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = m + n; \quad \text{Var}(X + Y) = \text{Var}X + \text{Var}Y = \sigma^2 + \tau^2.$$

Preuve du théorème

a) D'après la représentation des v.a. gaussiennes,

$$X = \sigma U + m; \quad Y = \tau V + n,$$

où U et V sont deux v.a.r. gaussiennes réduites et centrées. De plus $U = \frac{X-m}{\sigma}$ et $V = \frac{Y-n}{\tau}$, donc U et V sont indépendantes. Il suffit de montrer que $\sigma U + \tau V$ est une v.a.r. gaussienne (ce qui revient à supposer $m = n = 0$).

b) Mais σU (resp. τV) a pour densité :

$$f_{0,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} \quad (\text{resp. } f_{0,\tau^2}(x) = \frac{1}{\tau\sqrt{2\pi}} e^{-x^2/2\tau^2})$$

donc $\sigma U + \tau V$ a pour densité $\varphi = f_{0,\sigma^2} * f_{0,\tau^2}$.

$$\varphi(x) = \frac{1}{2\pi\sigma\tau} \int_{\mathbb{R}} \exp \left\{ -\frac{(x-y)^2}{2\sigma^2} - \frac{y^2}{2\tau^2} \right\} dy.$$

Posons $\rho^2 = \sigma^2 + \tau^2$, $A = \frac{(x-y)^2}{\sigma^2} + \frac{y^2}{\tau^2}$. On met A sous forme canonique (par rapport à la variable y , x étant considéré comme fixé) :

$$A = \frac{\rho^2}{\sigma^2\tau^2} y^2 - \frac{2xy}{\sigma^2} + \frac{x^2}{\sigma^2} = \frac{\rho^2}{\sigma^2\tau^2} \left(y^2 - \frac{2\tau^2}{\rho^2} xy \right) + \frac{x^2}{\sigma^2},$$

$$A = \frac{\rho^2}{\sigma^2 \tau^2} \left(\left(y - \frac{\tau^2}{\rho^2} x \right)^2 - \frac{\tau^4}{\rho^4} x^2 \right) + \frac{x^2}{\sigma^2} = \frac{\rho^2}{\sigma^2 \tau^2} \left(y - \frac{\tau^2}{\rho^2} x \right)^2 + \frac{x^2}{\rho^2}.$$

Par conséquent,

$$\varphi(x) = \frac{1}{\rho \sqrt{2\pi}} \left(\int_{\mathbb{R}} f_{z,u^2}(y) dy \right) e^{-x^2/2\rho^2}; \quad z = \frac{\tau^2}{\rho^2} x, \quad u = \frac{\sigma^2 \tau^2}{\rho^2},$$

où f_{z,u^2} désigne la densité de la loi gaussienne de moyenne z et de variance u^2 . Mais alors,

$$\varphi(x) = \frac{1}{\rho \sqrt{2\pi}} e^{-x^2/2\rho^2} = f_{0,\rho^2}(x).$$

Ce qui montre que $\sigma U + \tau V$ est une v.a.r. gaussienne de moyenne nulle et de variance ρ^2 . $\square$

1.7 Convergences stochastiques

Considérons une suite de variables aléatoires $(X_n)_{n \geq 1}$ à valeurs dans $\mathbb{R}^d$, définies sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et étudions son comportement asymptotique lorsque n tend vers l'infini.

Convergence en loi

Définition 8 La suite de variables aléatoires $(X_n)_{n \geq 1}$ converge en loi vers une variable aléatoire X si, pour toute fonction test $\varphi \in \mathcal{C}_b^0(\mathbb{R}^d)$, on a

$$\mathbb{E}[\varphi(X_n)] \rightarrow \mathbb{E}[\varphi(X)] \quad \text{lorsque } n \rightarrow \infty.$$

Dans cette définition, $\mathcal{C}_b^0$ désigne l'ensemble des fonctions continues bornées.

Notation.

$$X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$$

Remarque.

1) Comme la convergence en loi ne dépend pas précisément de X mais juste de sa loi $\mathbb{P}_X$, on parlera aussi de convergence en loi vers une mesure de probabilité μ sur $\mathbb{R}^d$:

$$\text{pour tout } \varphi \in \mathcal{C}_b^0(\mathbb{R}^d), \quad \mathbb{E}(\varphi(X_n)) \rightarrow \int_{\mathbb{R}^d} \varphi(x) d\mu(x), \quad \text{lorsque } n \rightarrow \infty.$$

2) Supposons que $X_n \xrightarrow{\mathcal{L}} X$ et que, pour tout $n \in \mathbb{N}^*$, $\mathbb{E}[|X_n|] < \infty$ et $\mathbb{E}[|X|] < \infty$. Il n'en résulte pas que la suite $(\mathbb{E}[X_n])_{n \geq 1}$ converge, ni que, si elle converge, elle ait pour limite $\mathbb{E}[X]$.

Exemple : On considère une suite de réels $(a_n)_{n \geq 1}$ strictement positifs et une suite de variables aléatoires dont la loi est définie par : $\mathbb{P}(X_n = a_n) = \frac{1}{n}$ et $\mathbb{P}(X_n = 0) = 1 - \frac{1}{n}$, alors

$X_n \xrightarrow{\mathcal{L}} X (= 0)$. Il est intéressant de comparer alors $\mathbb{E}[X_n]$ avec $\mathbb{E}[X]$ pour les suites de réels suivantes : $a_n = \sqrt{n}, n, n^2$ et $n(2 + (-1)^n)$.

Proposition 12 (pour $d = 1$) Soit F_n la fonction de répartition de X_n et F celle de X . Alors $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si, en tout point x de continuité de F , on a $F_n(x) \rightarrow F(x)$ lorsque $n \rightarrow \infty$.

Exemple.

On définit une suite de variables aléatoires réelles : X_n suit une loi gaussienne $\mathcal{N}_1(0, \frac{1}{n})$. Alors $X_n \xrightarrow{\mathcal{L}} 0$ lorsque $n \rightarrow \infty$. En effet, pour tout $n \in \mathbb{N}^*$ et $x \in \mathbb{R}$,

$$F_n(x) = \frac{n}{\sqrt{2\pi}} \int_{-\infty}^x \exp -\frac{(un)^2}{2} du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{xn} \exp -\frac{v}{2} dv.$$

$$\text{Ainsi } F_n(x) \xrightarrow[n \rightarrow \infty]{} \tilde{F}(x) = \begin{cases} 0, & \text{si } x < 0, \\ 1/2, & \text{si } x = 0, \\ 1, & \text{si } x > 0. \end{cases}$$

Par ailleurs,

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 & \text{si } x \geq 0. \end{cases}$$

Il résulte donc que $\tilde{F}(x) = F(x)$ pour tout $x \in \mathbb{R}^*$. Or 0 est un point de discontinuité de F , on obtient bel et bien $F_n \rightarrow F$ en tout point de continuité de F .

Convergence en probabilité

Définition 9 La suite de variables aléatoires X_n converge en probabilité vers la variable aléatoire X si

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

Notation.

$$X_n \xrightarrow[n \rightarrow \infty]{p} X$$

La seconde remarque établie pour la convergence en loi est valable également pour la convergence en probabilité. Notons que la convergence en probabilité est compatible avec les opérations algébriques élémentaires :

Proposition 13 Si la suite de variables aléatoires $((X_n, Y_n))_{n \geq 1}$ converge en probabilité vers le point aléatoire (X, Y) alors

- $X_n + Y_n \xrightarrow[n \rightarrow \infty]{p} X + Y$
- $\lambda X_n \xrightarrow[n \rightarrow \infty]{p} \lambda X \quad (\lambda \in \mathbb{R})$

- $X_n Y_n \xrightarrow[n \rightarrow \infty]{p} XY$
- $X_n / Y_n \xrightarrow[n \rightarrow \infty]{p} X/Y$ si $\mathbb{P}(Y = 0) = 0$.

Théorème 3 (Critère de convergence)

S'il existe $r > 0$ tel que $\mathbb{E}[|X_n|^r] \xrightarrow[n \rightarrow \infty]{} 0$ alors $X_n \xrightarrow[n \rightarrow \infty]{p} 0$.

Preuve : On utilise l'inégalité de Markov :

$$\mathbb{P}(|X_n| > \varepsilon) \leq \frac{\mathbb{E}[|X_n|^r]}{\varepsilon^r}.$$

□

Convergence presque-sûre

Définition 10 *La suite de variables aléatoires X_n converge presque-sûrement vers X s'il existe un ensemble $A \subset \Omega$ tel que*

$$\forall \omega \in A \quad \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \text{ et } \mathbb{P}(A) = 1.$$

Notation : $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$

Proposition 14 (critère) *Posons, pour tout $\varepsilon > 0$, $E_n(\varepsilon) = \{|X_n - X| > \varepsilon\}$ et $E(\varepsilon) = \limsup_{n \rightarrow \infty} E_n(\varepsilon) = \bigcap_{n \geq 1} \bigcup_{k \geq n} E_k(\varepsilon)$ alors X_n converge p.s. vers X si, pour tout $\varepsilon > 0$, on a $\mathbb{P}(E(\varepsilon)) = 0$.*

Proposition 15 (critère de convergence)

- Si, pour tout $\varepsilon > 0$, la série de terme général $\mathbb{P}(|X_n - X| > \varepsilon)$ converge alors $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$.
- Si $\exists r > 0$ tel que la série de terme général $\mathbb{E}[|X_n - X|^r]$ soit convergente alors $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$.

Autres convergences

Définition 11 *La suite de variables aléatoires X_n converge en moyenne vers X si*

$$\mathbb{E}(\|X_n - X\|) \xrightarrow[n \rightarrow \infty]{} 0.$$

Notation : $X_n \xrightarrow[n \rightarrow \infty]{L^1} X$

Définition 12 *La suite de variables aléatoires X_n converge en moyenne quadratique si*

$$\mathbb{E}(\|X_n - X\|^2) \xrightarrow[n \rightarrow \infty]{} 0.$$

Notation : $X_n \xrightarrow[n \rightarrow \infty]{L^2} X$

Comparaison des types de convergence

convergence en moyenne quadratique

$\Downarrow$

convergence en moyenne

convergence presque-sûre

$\Downarrow$

$\Downarrow$

convergence en probabilité

$\Downarrow$

convergence en loi

Proposition 16 *Soit une suite $(X_n)_{n \geq 1}$ qui converge en probabilité vers X , alors il existe une suite partielle (X_{n_k}) extraite de $(X_n)_{n \geq 1}$ qui converge presque-sûrement vers X .*

Chapitre 2

Fonctions caractéristiques

2.1 Fonctions caractéristiques : définition et propriétés

Pour déterminer la loi de probabilité d'une variable aléatoire, il existe plusieurs outils à notre disposition. Nous avons déjà vu que la fonction de répartition déterminait de façon unique la loi de probabilité. Malheureusement cette fonction n'est pas facilement *calculable* de façon explicite pour certaines lois comme la loi de Gauss (c'est donc problématique!!). Dans ce chapitre nous nous intéressons donc à un autre outil, appelé *fonction caractéristique*, qui joue le même rôle que la fonction de répartition mais qui est à valeurs complexes. L'analyse complexe nous permet alors de faire beaucoup plus de calculs explicites.

Définition 13 Soit X une variable aléatoire réelle. On appelle **fonction caractéristique** de X (ou de la loi de X) la fonction de la variable réelle t définie par :

$$\varphi_X(t) = \mathbb{E}[e^{itX}].$$

En d'autres termes, la fonction caractéristique est la transformée de Fourier de la loi $\mathbb{P}_X$. En effet,

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx} d\mathbb{P}_X(x).$$

Notons que cette fonction est toujours bien définie puisque le module de l'exponentielle complexe vaut 1.

Remarque 1 1. Si X est une variable aléatoire de loi discrète $\sum_{k \geq 0} p_{x_k} \delta_k$ alors

$$\varphi_X(t) = \sum_{k \geq 0} p_k e^{itx_k}$$

2. Si X est une variable absolument continue, de densité f , alors

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx} f(x) dx.$$

Ainsi φ_X correspond à la transformée de Fourier de la densité f . La transformée de Fourier est bien définie puisque f , la densité, est une fonction intégrable (au sens de Lebesgue).

Dans un premier temps, nous présentons quelques propriétés élémentaires des fonctions caractéristiques :

Proposition 17 Désignons par φ la fonction caractéristique d'une variable aléatoire réelle X . Alors

1. φ est une fonction définie et continue pour tout t réel
2. φ est bornée et pour tout t , on a $|\varphi(t)| \leq \varphi(0) = 1$.
3. φ est une fonction hermitique $\varphi(-t) = \overline{\varphi(t)}$.
4. pour tous réels a et b , on a

$$\varphi_{aX+b}(t) = e^{ibt} \varphi_X(at).$$

5. si la loi de X est symétrique alors φ est une fonction réelle paire
6. toute combinaison convexe de fonctions caractéristiques est une fonction caractéristique.
7. si φ est intégrable, au sens où $\int_{\mathbb{R}} |\varphi(t)| dt < +\infty$, alors X possède la densité obtenue par inversion de Fourier :

$$x \in \mathbb{R} \mapsto f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi(t) dt.$$

Nous n'allons pas détailler la démonstration de ces propriétés élémentaires. Certaines d'entre elles seront abordées en exercice. La dernière propriété repose sur les propriétés de la transformée de Fourier (voir le cours d'Analyse Fonctionnelle). Donnons juste quelques idées pour la continuité de la fonction caractéristique lorsque X a une densité f . Soient u et v deux réels alors

$$|\varphi(u) - \varphi(v)| = \left| \int_{\mathbb{R}} (e^{iux} - e^{ivx}) f(x) dx \right|.$$

Comme $e^{iux} - e^{ivx} = e^{ivx}(e^{i(u-v)x} - 1)$, on obtient

$$|\varphi(u) - \varphi(v)| \leq \int_{\mathbb{R}} |e^{i(u-v)x} - 1| |f(x)| dx.$$

D'après le théorème de la convergence dominée, le membre de droite de l'inégalité précédente tend vers 0 lorsque $u - v$ tend vers 0. La fonction caractéristique est donc même uniformément continue. Lorsque la variable X n'est pas à densité le même raisonnement s'applique, il suffit de remplacer $f(x) dx$ par $d\mathbb{P}(x)$.

Le tableau suivant donne les fonctions caractéristiques des lois de probabilité usuelles.

Loi	Expression analytique	Fonction caractéristique
Bernoulli $\mathcal{B}(p)$	$p_0 = 1 - p, p_1 = p$	$(1 - p) + pe^{it}$
Loi binomiale	$p_k = \binom{n}{k} p^k (1 - p)^{n-k}$	$\left((1 - p) + pe^{it}\right)^n$
Loi géométrique $\mathcal{G}(p)$	$p_k = (1 - p)^{k-1} p, \quad (k \geq 1)$	$\frac{pe^{it}}{1 - (1 - p)e^{it}}$
Loi de Poisson $\mathcal{P}(\lambda)$	$p_k = e^{-\lambda} \frac{\lambda^k}{k!}$	$\exp\{\lambda(e^{it} - 1)\}$
Loi uniforme $\mathcal{U}([a, b])$	$f(x) = 1_{[a, b]}(x)$	$\frac{\sin\{(b - a)t/2\}}{(b - a)t/2} e^{it(a+b)/2}$
Loi exponentielle $\mathcal{E}(\lambda)$	$f(x) = \lambda e^{-\lambda x} 1_{\{x \geq 0\}}$	$\frac{\lambda}{\lambda - it}$
Loi de Cauchy $C(0, 1)$	$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$	$e^{- t }$
Loi gaussienne $\mathcal{N}(\mu, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$	$\exp\left\{i\mu t - \frac{\sigma^2 t^2}{2}\right\}$

Détaillons juste les calculs pour la loi gaussienne. Supposons que $X \sim \mathcal{N}(\mu, \sigma^2)$ alors en utilisant un changement de variable, on peut remarquer que

$$G = \frac{1}{\sigma} (X - \mu) \sim \mathcal{N}(0, 1).$$

En utilisant les propriétés élémentaires des fonctions caractéristiques, on a donc

$$\varphi_X(t) = e^{it\mu} \varphi_G(t\sigma).$$

Il suffit ainsi de calculer $\varphi_G(t)$ c'est-à-dire

$$\varphi_G(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{itx} e^{-\frac{x^2}{2}} dx$$

Par dérivation sous le signe intégral, on en déduit

$$\begin{aligned}
\varphi'_G(t) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} ix e^{itx} e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \left(-i \int_{\mathbb{R}} (it - x) e^{itx} e^{-\frac{x^2}{2}} dx - t \int_{\mathbb{R}} e^{itx} e^{-\frac{x^2}{2}} dx \right) \\
&= \frac{1}{\sqrt{2\pi}} \left(-i \left[e^{itx - \frac{x^2}{2}} \right]_{-\infty}^{+\infty} - t \int_{\mathbb{R}} e^{itx} e^{-\frac{x^2}{2}} dx \right) \\
&= -t \varphi_G(t).
\end{aligned}$$

On résout donc l'équation différentielle $\varphi'_G(t) = -t\varphi_G(t)$ avec la condition initiale $\varphi_G(0) = 1$. On obtient alors $\varphi_G(t) = e^{-\frac{t^2}{2}}$ puis on en déduit φ_X . Une autre démonstration est proposée dans le chapitre sur les vecteurs aléatoires gaussiens.

Théorème 4 (Théorème d'unicité) *La fonction caractéristique d'une variable aléatoire détermine sa loi de probabilité.*

Ainsi deux variables aléatoires ayant la même fonction caractéristique ont la même loi. L'idée repose sur le fait que les fonctions exponentielles complexes $x \mapsto e^{itx}$ forment un ensemble *suffisamment gros* de fonctions bornées. Ainsi, si l'égalité

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

est vérifiée pour toute fonction f dans cet ensemble, elle le sera pour toute fonction f bornée. Pour démontrer ce résultat, nous allons utiliser la propriété *ancipitale* de certaines densités de probabilité.

Définition 14 *On dit qu'une variable aléatoire continue Y a la propriété ancipitale si sa fonction caractéristique φ_Y et sa densité f_Y sont égales à un facteur multiplicatif près.*

On peut noter que la loi normale $\mathcal{N}(0, 1)$ a la propriété ancipitale.

Lemme 1 *Soient X une variable aléatoire réelle et Y une variable aléatoire réelle ancipitale indépendante de X . Alors $X + Y$ admet une densité qui ne dépend de la loi de X qu'à travers la fonction caractéristique φ_X . Cette propriété est aussi vérifiée pour $X + \epsilon Y$ avec $\epsilon \in \mathbb{R}^*$.*

Preuve : D'après l'hypothèse du Lemme, il existe une constante $c > 0$ telle que $f_Y = c\varphi_Y$. Ainsi la densité de $X + Y$, qui est obtenue par le produit de convolution, satisfait

$$f_{X+Y}(t) = \int_{\mathbb{R}} f_Y(t - x) d\mathbb{P}_X(x) = c \int_{\mathbb{R}} \varphi_Y(t - x) d\mathbb{P}_X(x)$$

Comme φ_Y est une fonction caractéristique, on a

$$f_{X+Y}(t) = c \int_{\mathbb{R}^2} e^{i(t-x)y} f_Y(y) dy d\mathbb{P}_X(x)$$

L'intégrale double est absolument convergente : on peut donc appliquer le théorème de Fubini et ainsi

$$f_{X+Y}(t) = c \int_{\mathbb{R}} e^{ity} f_Y(y) \varphi_X(-y) dy$$

On constate que cette densité ne dépend de la loi de X que par φ_X . $\square$

Nous en venons donc à la preuve du théorème d'unicité.

Preuve : Considérons la loi de $X + Y$, où Y est une variable aléatoire continue indépendante de X . Pour cela, on étudie la fonction de répartition $F_{X+Y}(z)$ donnée par

$$F_{X+Y}(z) = \mathbb{P}(X + Y \leq z) = \int_{-\infty}^z f_{X+Y}(u) du.$$

En appliquant le théorème de Fubini, on a

$$\begin{aligned} F_{X+Y}(z) &= \int_{\mathbb{R}} \left(\int_{-\infty}^z f_Y(u-t) du \right) d\mathbb{P}_X(t) = \int_{\mathbb{R}} \left(\int_{-\infty}^{z-t} f_Y(y) dy \right) d\mathbb{P}_X(t) \\ &= \int_{\mathbb{R}} F_Y(z-t) d\mathbb{P}_X(t). \end{aligned}$$

Ici F_Y désigne la fonction de répartition de la variable aléatoire Y . Considérons maintenant le cas particulier où $Y_n = \epsilon_n G$ avec $G \sim \mathcal{N}(0, 1)$ indépendante de X et $\epsilon_n > 0$ un petit paramètre que nous allons faire tendre vers 0. Il est clair que G est une variable aléatoire ancipitale. Lorsque ϵ_n tend vers 0, la suite de variables aléatoires $\epsilon_n G$ converge en loi vers 0 (voir l'exemple qui suit la Proposition 12). On en déduit en particulier que

$$\lim_{n \rightarrow \infty} F_{Y_n}(z-t) = 1_{\{t \leq z\}} \quad \text{pour tout } t \neq z.$$

Si on choisit z tel que $\mathbb{P}(X = z) = 0$ alors on peut appliquer le théorème de convergence dominée pour obtenir :

$$\lim_{n \rightarrow \infty} F_{X+Y_n}(z) = \int_{\mathbb{R}} 1_{\{t \leq z\}} d\mathbb{P}_X(t) = F_X(z)$$

Comme le membre de gauche de l'inégalité précédente ne dépend de la loi de X que par φ_X , le membre de droite satisfait la même propriété. On en déduit que la fonction de répartition de X (en dehors de ses points de discontinuité) ne dépend que de la fonction caractéristique φ_X . Pour conclure il suffit de se rappeler que F_X est une fonction continue à droite qui n'admet au plus qu'un nombre dénombrable de discontinuités ce qui permet de définir de manière unique la fonction de répartition qui elle-même détermine de façon unique la loi de probabilité. $\square$

La preuve du résultat d'unicité des fonctions caractéristiques établit un lien fort entre les fonctions caractéristiques et les fonctions de répartition en dehors de leurs points de discontinuité. Or nous avons déjà caractérisé la convergence en loi d'une suite de variables aléatoires à l'aide des fonctions de répartition (Proposition 12). Il est aussi possible de le faire par les fonctions caractéristiques.

Théorème 5 (de Paul Lévy) Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires de fonctions caractéristiques φ_n et Z une variable aléatoire de fonction caractéristique φ_Z . Si φ_n converge vers φ_Z au sens de la convergence simple alors X_n converge en loi vers Z .

La preuve de ce théorème ne sera pas démontrée en détail ici. L'idée est la suivante : on peut montrer que si $\varphi_{X_n} \rightarrow \varphi_Z$ alors $X_n + \epsilon G$ converge en loi vers $Z + \epsilon G$ lorsque n tend vers l'infini et cela, pour G une gaussienne centrée réduite indépendante de la suite (X_n) . Il suffit alors de faire tendre ϵ vers 0. On en déduit que X_n converge en loi vers Z .

Le résultat suivant illustre que la régularité de la fonction caractéristique est directement reliée à son intégrabilité.

Proposition 18 Soit X une variable aléatoire de carré intégrable (satisfaisant $\mathbb{E}[X^2] < \infty$). Sa fonction caractéristique est $\mathcal{C}^2(\mathbb{R})$ et admet le DL suivant en 0 :

$$\varphi_X(t) = 1 + it\mathbb{E}[X] - \frac{t^2}{2} \mathbb{E}[X^2] + o(t^2).$$

Ce résultat est tout à fait primordial, il permet en effet de calculer les premiers moments de la variable aléatoire à partir de sa fonction caractéristique. Cela ne pouvait se faire aussi simplement avec la densité ou la fonction de répartition.

Notons alors que

$$\text{Var}(X) = -\varphi_X''(0) + \varphi_X'(0)^2.$$

On se propose maintenant de démontrer la Proposition 18.

Preuve : Soit $t \in \mathbb{R}$ alors

$$\frac{1}{h} (e^{i(t+h)X} - e^{itX}) \rightarrow iX e^{itX}$$

presque sûrement lorsque h tend vers 0. Comme pour tout a et b on a $|e^{ia} - e^{ib}| \leq |b - a|$ on en déduit

$$\left| \frac{1}{h} (e^{i(t+h)X} - e^{itX}) \right| \leq |X|.$$

La variable X étant de carré intégrable, $|X|$ est intégrable et le théorème de convergence dominée implique

$$\lim_{h \rightarrow 0} \mathbb{E} \left(\frac{1}{h} (e^{i(t+h)X} - e^{itX}) \right) = \mathbb{E}[iX e^{itX}].$$

On en déduit que

$$\lim_{h \rightarrow 0} \frac{1}{h} (\varphi_X(t+h) - \varphi_X(t)) = \mathbb{E}[iX e^{itX}].$$

La fonction φ_X est donc dérivable et on peut calculer sa dérivée. En appliquant à nouveau le théorème de convergence dominée, on peut montrer que φ_X' est continue et même $\mathcal{C}^1$ avec $\varphi_X''(t) = -\mathbb{E}[X^2 e^{itX}]$. Il suffit alors d'appliquer la formule de Taylor avec reste intégral pour obtenir le résultat annoncé. $\square$

Ce résultat peut évidemment se généraliser aux ordres supérieurs.

Ainsi si X satisfait $\mathbb{E}[|X|^n] < \infty$ pour un certain ordre n , alors

- φ_X est continûment différentiable jusqu'à l'ordre n inclus
- pour tout $0 \leq k \leq n$, on a $\varphi_X^{(k)}(0) = i^k \mathbb{E}[X^k]$
- φ_X est représenté par la formule de Taylor :

$$\varphi_X(t) = \sum_{k=0}^n \frac{(it)^k}{k!} \mathbb{E}[X^k] + o(|t|^n)$$

lorsque $t \rightarrow 0$.

Corollaire 1 *Soit Z une v.a. de loi gaussienne réduite et centrée, alors Z admet des moments de tous ordres, les moments impairs sont nuls et*

$$\mathbb{E}[Z^{2k}] = \frac{(2k)!}{2^k k!}; \quad k \geq 1. \quad (2.1)$$

Preuve : On a démontré que la fonction caractéristique de Z vaut :

$$\varphi_Z(t) = \mathbb{E}[e^{itZ}] = e^{-t^2/2} = \sum_{k \geq 0} \frac{(-1)^k}{2^k k!} t^{2k}; \quad \forall t \in \mathbb{R}.$$

Mais φ_Z est analytique et admet pour développement : $\varphi_Z(t) = \sum_{n \geq 0} \frac{\varphi_Z^{(n)}(0)}{n!} t^n$. Il suffit ensuite d'identifier. □

Le lien entre l'indépendance et les fonctions caractéristiques est le suivant,

Proposition 19 *Si X et Y sont deux v.a. indépendantes, alors,*

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t); \quad \forall t \in \mathbb{R}^n. \quad (2.2)$$

Preuve : En effet,

$$\varphi_{X+Y}(t) = \mathbb{E}[e^{it(X+Y)}] = \mathbb{E}[e^{itX} e^{itY}] = \mathbb{E}[e^{itX}] \mathbb{E}[e^{itY}] = \varphi_X(t) \varphi_Y(t).$$

□

Il est finalement possible de définir des fonctions caractéristiques pour des couples de variables aléatoires. Soit (X, Y) un vecteur aléatoire alors la fonction de deux variables

$$\varphi_{(X,Y)}(t, u) = \mathbb{E}[e^{itX + iuY}]$$

est la fonction caractéristique du couple. En particulier cette fonction détermine de façon unique la loi jointe du couple (X, Y) .

2.2 Application aux théorèmes limites

Dans cette section, on souhaite utiliser les fonctions caractéristiques pour démontrer le théorème de la limite centrale. Pour ce faire, nous nous intéresserons dans un premier temps à la loi des grands nombres puis nous en viendrons au TCL.

Loi des grands nombres

Nous rappelons les lois faible et forte des grands nombres. Nous considérons une suite de variables aléatoires $(X_n)_{n \geq 1}$ et construisons sa somme partielle :

$$S_n = X_1 + X_2 + \dots + X_n.$$

Lorsque les variables aléatoires $(X_n)_{n \geq 1}$ ont la même loi, on notera $\mathbb{E}[X]$ leur espérance commune et $\text{Var}(X) = \sigma_X^2$ leur variance commune.

Théorème 6 (Loi faible des grands nombres).

On suppose que $(X_n)_{n \geq 1}$ est formée de variables aléatoires indépendantes et de même loi possédant un moment d'ordre deux, c'est-à-dire $\mathbb{E}[X^2] < \infty$. Alors la suite $(\frac{S_n}{n})_{n \geq 1}$ converge en probabilité vers $\mathbb{E}[X]$ lorsque n tend vers l'infini.

Preuve : La preuve de ce résultat repose sur l'inégalité de Bienaymé-Tchebitchef. En effet en considérant la variable aléatoire $Z_n = S_n/n$ on a les paramètres suivants : $\mathbb{E}[Z_n] = \mathbb{E}[X]$ et comme les variables aléatoires X_n sont indépendantes entre elles,

$$\text{Var}(Z_n) = \frac{1}{n^2} \text{Var}(S_n) = \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) = \frac{1}{n} \sigma_X^2.$$

L'inégalité de Bienaymé-Tchebitchef nous donne alors, pour tout $\epsilon > 0$,

$$\mathbb{P}(|Z_n - \mathbb{E}[Z_n]| > \epsilon) \leq \frac{\text{Var}(Z_n)}{\epsilon^2} = \frac{\sigma_X^2}{n\epsilon^2}.$$

Comme $\mathbb{E}[Z_n] = \mathbb{E}[X]$, on en déduit que

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}[X]\right| > \epsilon\right) = 0$$

ce qui signifie que S_n/n converge en probabilité vers $\mathbb{E}[X]$ (voir la section dans le premier chapitre traitant des convergences stochastiques). $\square$

La démonstration de ce résultat est particulièrement sympathique mais il arrive que les variables aléatoires considérées ne soient hélas pas de carré intégrable. Nous présentons alors une version plus générale de la loi faible des grands nombres.

Théorème 7 *On suppose que $(X_n)_{n \geq 1}$ est formée de variables aléatoires intégrables i.e. $\mathbb{E}[|X_n|] < \infty$, indépendantes et de même loi. Alors la suite $(\frac{S_n}{n})_{n \geq 1}$ converge en probabilité vers $\mathbb{E}[X]$ lorsque n tend vers l'infini.*

La seule condition sur les variables aléatoires est l'intégrabilité. Ce résultat repose sur un lemme concernant la convergence en probabilité.

Lemme 2 *Si une suite de variables aléatoires $(X_n)_{n \geq 1}$ converge en loi vers une constante a alors $(X_n)_{n \geq 1}$ converge vers a en probabilité.*

Preuve : On considère l'intervalle $A =]a - \epsilon, a + \epsilon[$. Comme A est un intervalle ouvert A^c est un fermé de $\mathbb{R}$, on en déduit que 1_{A^c} est une fonction borélienne. Ainsi comme X_n converge vers X en loi, on a

$$\lim_{n \rightarrow \infty} \mathbb{E}[1_{A^c}(X_n)] = \mathbb{E}[1_{A^c}(a)] = 1_{A^c}(a) = 0.$$

En réécrivant la limite on obtient :

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - a| > \epsilon) = 0,$$

ce qui correspond à la convergence en probabilité. $\square$

On peut alors démontrer la loi faible des grands nombres pour des variables juste intégrables.

Preuve : (loi faible) Soit φ_n la fonction caractéristique de S_n/n et φ celle de X_1 . Les variables aléatoires sont indépendantes et de même loi, ainsi

$$\varphi_n(t) = \mathbb{E}[e^{itS_n/n}] = \left(\varphi(t/n)\right)^n.$$

Puisque X_1 est intégrable, la fonction φ est dérivable et $\varphi'(0) = i\mathbb{E}[X]$. Par conséquent, au voisinage de 0, on a $\varphi(t) = 1 + it\mathbb{E}[X] + o(t)$. On en déduit

$$\varphi_n(t) = \left(1 + i\frac{t}{n}\mathbb{E}[X] + o\left(\frac{1}{n}\right)\right)^n = \exp \left\{ n \log \left(1 + i\frac{t}{n}\mathbb{E}[X] + o\left(\frac{1}{n}\right)\right) \right\}$$

Sachant que $\log(1 + z)$ est équivalent à z au voisinage de 0, on en déduit :

$$\lim_{n \rightarrow \infty} \varphi_n(t) = e^{it\mathbb{E}[X]}.$$

Ce qui signifie que S_n/n converge en loi vers la constante $\mathbb{E}[X]$ lorsque n tend vers l'infini. Il suffit alors d'appliquer le Lemme 2 pour démontrer la loi faible des grands nombres. $\square$

Théorème 8 (loi forte des grands nombres) *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et de même loi.*

1. *La suite $(S_n/n)_{n \geq 1}$ converge presque sûrement si et seulement si les variables aléatoires $(X_n)_{n \geq 1}$ sont intégrables.*
2. *Si les variables aléatoires $(X_n)_{n \geq 1}$ sont intégrables alors S_n/n converge presque sûrement et dans $L^1(\Omega)$ vers $\mathbb{E}[X]$.*

Preuve : Nous allons montrer juste une partie du résultat. Montrons que si les variables X_n admettent un moment d'ordre deux, alors S_n/n converge presque sûrement vers $\mathbb{E}[X]$.

On suppose que $\mathbb{E}[X^2] < \infty$. En écrivant $X_n = X_n^+ - X_n^-$ où X_n^+ (resp. X_n^-) désigne la partie positive de X_n (resp. la partie négative), on peut supposer que les variables X_n sont positives et de carré intégrable.

Considérons $Z_n = S_{n^2}/n^2$. On peut alors calculer l'espérance et la variance de Z_n . On a $\mathbb{E}[Z_n] = \mathbb{E}[S_{n^2}]/n^2 = \mathbb{E}[X]$ et $\text{Var}(Z_n) = \text{Var}(S_{n^2})/n^4 = \sigma_X^2/n^2$. Ainsi l'inégalité de Bienaymé-Tchebitchef nous assure :

$$\alpha_n := \mathbb{P}\left(\left|\frac{S_{n^2}}{n^2} - \mathbb{E}[X]\right| > \epsilon\right) \leq \frac{\sigma_X^2}{n^2}.$$

Ainsi α_n est le terme général d'une série convergente ce qui implique d'après le critère de convergence (Proposition 15) que S_{n^2}/n^2 converge presque sûrement vers $\mathbb{E}[X]$.

Les variables X_n étant positives, la suite $n \mapsto S_n$ est croissante. Par conséquent, pour tout k entier vérifiant $n^2 \leq k \leq (n+1)^2$, on a $S_{n^2} \leq S_k \leq S_{(n+1)^2}$. D'où

$$\frac{S_k}{k} \leq \frac{S_{(n+1)^2}}{n^2} = \left(\frac{n+1}{n}\right)^2 \frac{S_{(n+1)^2}}{(n+1)^2},$$

$$\frac{S_k}{k} \geq \frac{S_{n^2}}{(n+1)^2} = \left(\frac{n}{n+1}\right)^2 \frac{S_{n^2}}{n^2}.$$

On en déduit aisément la convergence presque sûre de S_k/k vers $\mathbb{E}[X]$ lorsque $k \rightarrow \infty$. $\square$

Applications de la loi forte des grands nombres

La loi forte des grands nombres a deux **applications** particulièrement importantes :

- la première application concerne l'estimation d'une probabilité. Imaginons que l'on observe une suite de réalisations indépendantes d'une expérience aléatoire et qu'on souhaite estimer la probabilité d'un événement A . On introduit alors la variable aléatoire X_n associée à chaque expérience et qui vaut 1 lorsque A est réalisée et 0 sinon (une variable de Bernoulli de paramètre inconnu p). En considérant la fréquence d'apparition de A au cours des n premières expériences, définie par

$$f_n = \frac{X_1 + \dots + X_n}{n},$$

la loi forte des grands nombres nous assure que f_n converge presque sûrement vers $\mathbb{E}[X_1] = p$. Il s'en suit donc que f_n est une bonne approximation de la valeur inconnue p .

- La seconde application importante est l'estimation de la valeur de certaines intégrales par la méthode de Monte-Carlo. En effet considérons f une fonction bornée et son intégrale $I(f)$ définie par :

$$I(f) = \int_0^1 f(t) dt.$$

On peut remarquer que l'intégrale s'exprime sous la forme d'une espérance de variable aléatoire, notamment $I(f) = \mathbb{E}[f(U_1)]$ où U_1 est une variable aléatoire uniforme sur l'intervalle $[0, 1]$. En choisissant une suite de variables aléatoires uniformes indépendantes $(U_n)_{n \geq 1}$ on obtient alors que I_n définie par

$$I_n = \frac{f(U_1) + \dots + f(U_n)}{n}$$

converge presque sûrement vers I lorsque $n \rightarrow \infty$. Cela permet d'avoir une approximation de I différente des approximations classiques données par exemple par la méthode des rectangles ou des trapèzes. Les méthodes de Monte Carlo seront développées dans un chapitre ultérieur.

Le théorème de la limite centrale ou "central limit"

L'utilité du théorème de la limite centrale est de donner une vitesse de convergence dans la loi des grands nombres.

Théorème 9 *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes, de même loi admettant un moment d'ordre deux, c'est-à-dire $\mathbb{E}[X_n^2] < \infty$ pour tout $n \geq 1$. En notant S_n/n la moyenne empirique, on observe la convergence suivante lorsque n tend vers l'infini :*

$$\frac{\sqrt{n}}{\sigma_X} \left(\frac{S_n}{n} - \mathbb{E}[X] \right) \xrightarrow{\mathcal{L}} Y \sim \mathcal{N}(0, 1).$$

Ce résultat explique la place fondamentale des variables aléatoires gaussiennes en probabilités et en statistique. En effet la loi commune des variables X_n n'influence pas du tout celle de la limite qui restera toujours gaussienne.

Une conséquence immédiate du théorème de la limite centrale est :

$$\mathbb{P} \left(\lambda_1 \leq \sqrt{n} \left(\frac{X_1 + \dots + X_n}{n} - \mathbb{E}[X] \right) \leq \lambda_2 \right) \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{\lambda_1}^{\lambda_2} e^{-t^2/2} dt.$$

Preuve : On va utiliser le rôle essentiel de la fonction caractéristique. On pose

$$Y_n := \frac{\sqrt{n}}{\sigma_X} \left(\frac{S_n}{n} - \mathbb{E}[X] \right).$$

Soit $t \in \mathbb{R}$, alors la fonction caractéristique satisfait

$$\begin{aligned} \varphi_{Y_n}(t) &= \mathbb{E} \left[e^{it \frac{\sqrt{n}}{\sigma_X} \left(\frac{S_n}{n} - \mathbb{E}[X] \right)} \right] = \mathbb{E} \left[e^{it \frac{1}{\sigma_X \sqrt{n}} \sum_{k=1}^n (X_k - \mathbb{E}[X])} \right] \\ &= \prod_{k=1}^n \mathbb{E} \left[e^{it \frac{1}{\sigma_X \sqrt{n}} (X_k - \mathbb{E}[X])} \right] \quad \text{par indépendance des } X_k \\ &= \left(\mathbb{E} \left[e^{it \frac{1}{\sigma_X \sqrt{n}} (X_k - \mathbb{E}[X])} \right] \right)^n \quad \text{car toutes les variables ont la même loi,} \\ &= \left(\varphi_{X_1 - \mathbb{E}[X]} \left(\frac{t}{\sigma_X \sqrt{n}} \right) \right)^n. \end{aligned}$$

Comme $X_1 - \mathbb{E}[X]$ est une variable aléatoire de moyenne nulle et d'écart-type σ_X , la Proposition 18 nous assure de développement limité suivant : $\varphi_{X_1 - \mathbb{E}[X]}(t) = 1 - \frac{\sigma_X^2}{2} t^2 + o(t^2)$. On obtient donc

$$\varphi_{X_1 - \mathbb{E}[X]} \left(\frac{t}{\sigma_X \sqrt{n}} \right) = 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right).$$

Ainsi

$$\varphi_{Y_n}(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \right)^n \xrightarrow{n \rightarrow \infty} e^{-\frac{t^2}{2}} = \varphi_Y(t).$$

La convergence des fonctions caractéristiques entraîne la convergence en loi. $\square$

Applications du théorème de la limite centrale

En appliquant le théorème de la limite centrale à différentes lois de probabilité, on obtient alors les résultats suivants :

Corollaire 2 (théorème de De Moivre-Laplace) Soit $(X_n)_{n \geq 1}$ une suite de variables indépendantes de loi de Bernoulli de paramètre p et soit $S_n = X_1 + \dots + X_n$. Alors la suite des variables aléatoires $(S_n - np)/\sqrt{np(1-p)}$ converge en loi, lorsque n tend vers l'infini, vers une v.a. gaussienne centrée réduite.

On peut également s'intéresser à la loi de Poisson :

Corollaire 3 Soit $(X_n)_{n \geq 1}$ une suite de variables indépendantes de loi de Poisson de paramètre 1 et soit $S_n = X_1 + \dots + X_n$. Alors la suite des variables aléatoires $(S_n - n)/\sqrt{n}$ converge en loi, lorsque n tend vers l'infini, vers une v.a. gaussienne centrée réduite.

On peut finalement s'intéresser à l'approximation de la loi gamma par une gaussienne :

Corollaire 4 Soit $(X_n)_{n \geq 1}$ une suite de variables indépendantes de loi Gamma $\gamma(a, b)$ et soit $S_n = X_1 + \dots + X_n$. Alors la suite des variables aléatoires $(S_n - nab)/b\sqrt{an}$ converge en loi, lorsque n tend vers l'infini, vers une v.a. gaussienne centrée réduite.

Chapitre 3

Vecteurs aléatoires gaussiens

Les variables aléatoires gaussiennes apparaissent naturellement comme limite de sommes re-normalisées, de v.a. indépendantes ; pour plus de précisions on peut se reporter à l'énoncé du théorème central limite dans la section 4. Ainsi la somme cumulée de petites fluctuations au niveau microscopique donne naissance à une fluctuation macroscopique gaussienne. En plus de cette propriété de “normalité”, les v.a. gaussiennes à valeurs multidimensionnelles sont très utilisées dans la modélisation de phénomènes physiques, car elles se prêtent extrêmement bien au calcul.

3.1 Définitions et propriétés des variables aléatoires gaussiennes à valeurs réelles

Définition 15 Une variable aléatoire à valeurs réelles X est dite gaussienne réduite et centrée si sa loi de probabilité admet la densité :

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp -\frac{x^2}{2}, \quad x \in \mathbb{R}.$$

On note $\mathcal{N}_1(0, 1)$ cette loi. Rappelons que

$$\mathbb{E}[f(X)] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) \exp \left(-\frac{x^2}{2} \right) dx,$$

pour toute fonction borélienne bornée ou positive. En particulier,

$$\int_{\mathbb{R}} \exp \left(-\frac{x^2}{2} \right) dx = \sqrt{2\pi}.$$

Remarque : On introduit la fonction d'erreur erf définie par $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$. Si X suit une loi $\mathcal{N}_1(0, 1)$ alors $\mathbb{P}(|X| \leq \sqrt{2}x) = \operatorname{erf}(x)$.

Proposition 20 Soit X une v.a.r. gaussienne réduite et centrée.

1) Pour tout $z \in \mathbb{C}$, $\mathbb{E}[e^{zX}] < \infty$ et

$$\mathbb{E}[\exp zX] = \exp z^2/2. \quad (3.1)$$

En particulier

$$\mathbb{E}[\exp(itX)] = e^{-t^2/2}, \quad \forall t \in \mathbb{R}. \quad (3.2)$$

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

$$\mathbb{E}(X^n) = \begin{cases} 0 & \text{si } n \text{ est impair,} \\ \frac{(2m)!}{m!2^m}, & \text{si } n \text{ est pair, } n = 2m. \end{cases} \quad (3.3)$$

Preuve : On vérifie que l'intégrale $\int_{\mathbb{R}} \exp(zx - \frac{1}{2}x^2)dx$ est absolument convergente pour tout z complexe. Par conséquent la quantité $\varphi(z) = \mathbb{E}(e^{zX})$ est bien définie et,

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(zx - \frac{1}{2}x^2)dx.$$

Supposons z réel. On écrit $zx - \frac{1}{2}x^2 = -\frac{1}{2}(x - z)^2 + \frac{z^2}{2}$ et l'on fait le changement de variable $y = x - z$ dans l'intégrale, il vient $\varphi(z) = e^{z^2/2}$. Remarquons que φ et $z \rightarrow e^{z^2/2}$ sont deux fonctions entières ; puisque ces deux fonctions coïncident sur $\mathbb{R}$, elles sont égales sur $\mathbb{C}$. En particulier, si $z = it$ avec $t \in \mathbb{R}$, on a : $\mathbb{E}(\exp(itX)) = e^{-t^2/2}$, $t \in \mathbb{R}$.

Soit $n \geq 1$. Sachant que $\lim_{|x| \rightarrow \infty} \frac{|x|^n}{\text{ch}(x)} = 0$, il existe une constante c_n telle que

$$|x|^n \leq c_n(e^x + e^{-x}), \quad \forall x \in \mathbb{R}. \quad (3.4)$$

Puisque $\mathbb{E}(\exp(aX))$ existe pour tout réel a , on déduit de (3.4) que $\mathbb{E}(|X|^n) < \infty$. Par conséquent, $\mathbb{E}(X^n)$ existe pour tout $n \geq 1$.

Par ailleurs, en utilisant le développement en série entière de $x \rightarrow e^{itx}$, on peut affirmer que, presque sûrement,

$$e^{itX} = \lim_{n \rightarrow \infty} S_n, \quad S_n = \sum_{k=0}^n \frac{(it)^k}{k!} X^k.$$

Mais $|S_n| \leq Y$ avec

$$Y = \sum_{k=0}^{\infty} \frac{|t|^k |X|^k}{k!} = e^{|tX|} \leq e^{tX} + e^{-tX}.$$

En utilisant à nouveau le fait que $\mathbb{E}(\exp(aX)) < \infty$, on en déduit que Y est intégrable. Une application du théorème de Lebesgue, conduit à,

$$\mathbb{E}(\exp(itX)) = \mathbb{E} \left(\sum_{n \geq 0} \frac{(itX)^n}{n!} \right) = \sum_{n \geq 0} \frac{i^n t^n}{n!} \mathbb{E}(X^n).$$

Par identification on en déduit (3.3). □

Notons en particulier $\mathbb{E}(X) = 0$, $\text{Var}(X) = 1$. Ce qui justifie les termes *réduit* et *centré*.

Définition 16 Une variable aléatoire réelle Y est dite gaussienne s'il existe une v.a. X gaussienne réduite et centrée, et deux réels a et b tels que $Y = aX + b$.

On peut identifier a et b à l'aide de l'espérance et la variance de Y , plus précisément,

$$\mathbb{E}(Y) = b, \quad \text{Var}Y = a^2 \times \text{Var}X = a^2.$$

Posons $\sigma = |a|$ et $m = b$, et supposons $a \geq 0$, alors

$$Y = \sigma X + m, \quad X \text{ de loi gaussienne réduite et centrée.} \quad (3.5)$$

Si $a < 0$, on écrit $Y = (-a)(-X) + b$, on peut se ramener au cas précédent en observant que $-X$ suit une loi gaussienne réduite et centrée.

La relation précédente permet d'exprimer X à l'aide de Y . En effet, soient Y une v.a. gaussienne, $m = \mathbb{E}(Y)$ et $\sigma = \sqrt{\text{Var}Y}$. On suppose $\sigma > 0$. Alors la v.a. $Y_* = \frac{Y-m}{\sigma}$ est une v.a. de loi gaussienne réduite et centrée et

$$Y = \sigma Y_* + m. \quad (3.6)$$

On note $\mathcal{N}_1(m, \sigma^2)$ la loi d'une v.a. de loi gaussienne de moyenne m et de variance σ^2 . Un calcul aisé montre :

$$\frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{(x-m)^2}{2\sigma^2} \text{ est la densité de } \mathcal{N}_1(m, \sigma^2). \quad (3.7)$$

Soit Y de loi $\mathcal{N}_1(m, \sigma^2)$, on déduit de (3.6)

$$\mathbb{E}[\exp(itY)] = e^{itm} \mathbb{E}[\exp i(t\sigma)Y_*].$$

Une application directe de (3.1) et (3.2) conduit à,

$$\mathbb{E}[\exp(itY)] = \exp\left(itm - \frac{t^2}{2}\sigma^2\right), \quad t \in \mathbb{R}. \quad (3.8)$$

En utilisant l'injectivité de la transformée de Fourier, on montre que si X est une v.a. telle que la fonction caractéristique (c'est-à-dire la fonction $t \rightarrow \mathbb{E}(\exp(itX))$) est égale à $t \rightarrow \exp(ita - t^2b^2)$ où $a \in \mathbb{R}$ et $b \in \mathbb{R}$, X est une v.a. gaussienne.

Proposition 21 Soient Y_1 et Y_2 deux v.a. gaussiennes, indépendantes, Y_1 de loi $\mathcal{N}_1(m_1, \sigma_1^2)$, Y_2 de loi $\mathcal{N}_1(m_2, \sigma_2^2)$. Alors $Y_1 + Y_2$ est une v.a. gaussienne de loi $\mathcal{N}_1(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Preuve : Les v.a. Y_1 et Y_2 étant indépendantes, on a pour tout t réel :

$$\varphi(t) = \mathbb{E}[\exp(it(Y_1 + Y_2))] = \mathbb{E}[\exp(itY_1)] \mathbb{E}[\exp(itY_2)].$$

De plus Y_1 et Y_2 sont gaussiennes, d'après (3.8), on a :

$$\begin{aligned} \varphi(t) &= \exp\left(itm_1 - \frac{t^2\sigma_1^2}{2}\right) \exp\left(itm_2 - \frac{t^2\sigma_2^2}{2}\right) \\ &= \exp\left(it(m_1 + m_2) - \frac{t^2(\sigma_1^2 + \sigma_2^2)}{2}\right). \end{aligned}$$

□

Remarques

- 1) Le résultat peut être faux si l'on ne suppose plus Y_1 et Y_2 indépendantes.
2) Plaçons-nous sous les hypothèses de la proposition 21. Si l'on sait que $Y_1 + Y_2$ est gaussienne, de loi $\mathcal{N}_1(m, \sigma^2)$, il est aisé d'identifier les deux paramètres :

$$m = \mathbb{E}(Y_1 + Y_2) = \mathbb{E}(Y_1) + \mathbb{E}(Y_2) = m_1 + m_2,$$

$$\sigma^2 = \text{Var}(Y_1 + Y_2) = \text{Var}Y_1 + \text{Var}Y_2 = \sigma_1^2 + \sigma_2^2.$$

La deuxième égalité a lieu car Y_1 et Y_2 sont deux v.a. indépendantes.

- 3) Le résultat se généralise sans difficulté au cas de n v.a. : soient $Y_1, Y_2, \dots, Y_n$, n v.a.r., indépendantes, Y_i de loi $\mathcal{N}_1(m_i, \sigma_i^2)$, $1 \leq i \leq n$, alors

$$Y_1 + Y_2 + \dots + Y_n \text{ suit une loi } \mathcal{N}_1 \left(\sum_{i=1}^n m_i, \sum_{i=1}^n \sigma_i^2 \right). \quad (3.9)$$

3.2 Espérance et covariance de v.a. à valeurs vectorielles

Avant de considérer les vecteurs gaussiens, il est bon de rappeler les définitions et propriétés des v.a. à valeurs vectorielles.

Définitions et notations

- 1) Si A est une matrice d'ordre, A^* désigne la matrice transposée. En particulier si $x \in \mathbb{R}^n$ est considéré comme un vecteur unicolonne, x^* est une matrice uniligne. Si x et y sont deux vecteurs de $\mathbb{R}^n$, leur produit scalaire est noté :

$$\langle x, y \rangle = x^* y = y^* x = \sum_{i=1}^n x_i y_i, \quad x^* = (x_1, \dots, x_n), \quad y^* = (y_1, \dots, y_n).$$

- 2) Une v.a. X à valeurs dans $\mathbb{R}^n$ est la donnée de n v.a. à valeurs réelles $X_1, X_2, \dots, X_n$. On note X la matrice **unicolonne** de coordonnées $X_1, X_2, \dots, X_n$:

$$X^* = (X_1, X_2, \dots, X_n) \text{ ou encore } X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_n \end{pmatrix}. \quad (3.10)$$

Pour d'évidentes raisons typographiques, on choisira la première écriture.

- 3) Soit X une v.a. de coordonnées $X_1, X_2, \dots, X_n$, on note $\mathbb{E}(X)$ le vecteur unicolonne de coordonnées $\mathbb{E}(X_1), \mathbb{E}(X_2), \dots, \mathbb{E}(X_n)$:

$$\mathbb{E}(X)^* = (\mathbb{E}(X_1), \mathbb{E}(X_2), \dots, \mathbb{E}(X_n)). \quad (3.11)$$

On suppose bien sûr que chaque v.a. X_i admet une espérance.

4) Soit X (resp. Y) une v.a. à valeurs dans $\mathbb{R}^n$ (resp. $\mathbb{R}^m$), $K_{X,Y}$ est la matrice d'ordre $n \times m$ (n lignes et m colonnes) définie par

$$K_{X,Y} = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))^*]. \quad (3.12)$$

Remarquons que $X - \mathbb{E}(X)$ est une matrice $n \times 1$ et $(Y - \mathbb{E}(Y))^*$ une matrice $1 \times m$, le produit $(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))^*$ est une matrice $n \times m$. Soit $K_{X,Y}(i, j)$ l'élément de $K_{X,Y}$ situé à la i ème ligne et j ème colonne, alors :

$$K_{X,Y}(i, j) = \text{Cov}(X_i, Y_j); \quad 1 \leq i \leq n, \quad 1 \leq j \leq m, \quad (3.13)$$

où $\text{Cov}(U, V) = \mathbb{E}(UV) - \mathbb{E}(U)\mathbb{E}(V)$. Rappelons que si U et V sont deux v.a. à valeurs réelles,

$$\text{Cov}(U, V) = \mathbb{E}[(U - \mathbb{E}(U))(V - \mathbb{E}(V))] = \mathbb{E}[UV] - \mathbb{E}[U]\mathbb{E}[V].$$

Si l'on choisit $Y = X$, la matrice $K_{X,X}$ est appelée matrice de covariance de X , on note pour simplifier $K_X = K_{X,X}$. Cette matrice est carrée d'ordre n , et :

$$K_X = \mathbb{E}[(X - \mathbb{E}(X))(X - \mathbb{E}(X))^*]. \quad (3.14)$$

$$K_X(i, j) = \text{Cov}(X_i, X_j); \quad 1 \leq i \leq n, \quad 1 \leq j \leq n. \quad (3.15)$$

Nous allons à présent donner quelques propriétés utiles en pratique. La démonstration en est laissée au lecteur.

Propriétés

1) Soient X une v.a. à valeurs $\mathbb{R}^n$, A une matrice d'ordre $m \times n$, u un vecteur de $\mathbb{R}^m$. Alors

$$\mathbb{E}[u + AX] = u + A\mathbb{E}(X). \quad (3.16)$$

2) Comme dans le cas unidimensionnel, on a deux formules pour calculer la covariance :

$$K_{X,Y} = \mathbb{E}(XY^*) - \mathbb{E}(X)\mathbb{E}(Y)^*. \quad (3.17)$$

En particulier,

$$K_X = \mathbb{E}(XX^*) - \mathbb{E}(X)\mathbb{E}(X)^*. \quad (3.18)$$

3) Soit A une matrice d'ordre $m \times n$, X une v.a. à valeurs dans $\mathbb{R}^n$ et $u \in \mathbb{R}^m$, alors

$$K_{u+AX} = K_{AX} = AK_X A^*. \quad (3.19)$$

3.3 Vecteurs aléatoires gaussiens

Nous conservons bien sûr les notations de la section précédente, les v.a. vectorielles seront représentées par des matrices unicolones.

Définition 17 Une v.a. X à valeurs dans $\mathbb{R}^n$, est dite gaussienne (on dit aussi que X est un vecteur aléatoire gaussien) si et seulement si pour tout $\lambda \in \mathbb{R}^n$, $\lambda^* = (\lambda_1, \dots, \lambda_n)$,

$$\langle \lambda, X \rangle = \lambda^* X = \sum_{i=1}^n \lambda_i X_i \text{ est une v.a. gaussienne réelle.} \quad (3.20)$$

Remarques :

- 1) Il est clair que si X est un vecteur gaussien, alors chaque composante X_i de X est une v.a.r. gaussienne.
- 2) L'exemple clé de vecteur gaussien est celui où $X_1, \dots, X_n$ sont n v.a. gaussiennes, indépendantes. Il suffit en effet d'appliquer directement (3.9). Attention ! L'hypothèse d'indépendance est essentielle. Il est facile de construire un exemple où X_1 et X_2 sont deux v.a.r. gaussiennes telles que (X_1, X_2) ne soit pas un vecteur gaussien.

Proposition 22 Soit K la matrice de covariance d'un vecteur gaussien X . Alors, pour tout $u \in \mathbb{R}^n$,

$$\mathbb{E}(\exp(i \langle u, X \rangle)) = e^{i \langle u, \mathbb{E}(X) \rangle - \frac{1}{2} u^* K u}, \quad (3.21)$$

où u est représenté sous la forme d'une matrice unicolonne.

Preuve : La v.a. $Y = \langle u, X \rangle = u^* X$ est une v.a. gaussienne et $\mathbb{E}(Y) = \langle u, \mathbb{E}(X) \rangle$. De plus d'après (3.19) : $\text{Var} Y = u^* K u$. Par conséquent, une application directe de (3.8) conduit à : $\mathbb{E}[\exp(i \langle u, X \rangle)] = e^{i \mathbb{E}(Y) - \frac{1}{2} \text{Var} Y}$. $\square$

Remarques :

- 1) Si $n = 1$, on retrouve exactement la formulation de (3.8). On peut remarquer d'ailleurs que pour établir la proposition 22, on se ramène au cas unidimensionnel.
- 2) La proposition 22 signifie que la loi d'un vecteur gaussien X est caractérisée par sa moyenne m et sa matrice de covariance K : si X et Y sont deux vecteurs gaussiens, ayant même moyenne et même matrice de covariance, ils ont même loi. Attention cette propriété concerne les vecteurs gaussiens, elle n'est pas vraie en général, il n'y a aucune raison pour que deux v.a. à valeurs réelles qui ont même moyenne et même variance aient même loi.

Proposition 23 Soient X un vecteur gaussien à valeurs dans $\mathbb{R}^n$, m sa moyenne et K_X sa matrice de covariance, A une matrice $p \times n$ (p lignes et n colonnes) et z un vecteur de $\mathbb{R}^p$. On pose $Y = AX + z$. Alors Y est un vecteur gaussien et

$$\mathbb{E}(Y) = z + Am, \quad K_Y = AK_X A^*, \quad (3.22)$$

K_Y désignant la matrice de covariance de Y .

Preuve : Soit $u \in \mathbb{R}^p$ représenté sous forme unicolonne. Alors

$$u^*Y = u^*z + u^*AX = u^*z + v^*X,$$

où l'on a posé $v = A^*u$.

X étant un vecteur gaussien, alors v^*X est une v.a.r. gaussienne. Y est bien un vecteur gaussien. De plus, $\mathbb{E}(Y) = z + A\mathbb{E}(X) = z + Am$; $K_Y = AK_XA^*$. $\square$

Proposition 24 *La matrice de covariance d'un vecteur aléatoire est une matrice symétrique et positive.*

Preuve : Soit K la matrice de covariance d'un vecteur aléatoire X à valeurs dans $\mathbb{R}^n$. Puisque $K(i, j) = \text{Cov}(X_i, X_j)$, il est clair que K est symétrique : $K(i, j) = K(j, i)$ pour tout i et j de $\{1, 2, \dots, n\}$. Montrons que K est une matrice positive : $u^*Ku \geq 0$, pour tout vecteur u de $\mathbb{R}^n$. En effet, posons $Y = u^*X$, alors $\text{Var}(Y) = u^*Ku \geq 0$. $\square$

Théorème 10 *Soient m un vecteur de $\mathbb{R}^n$ et Γ une matrice carrée symétrique positive d'ordre n . Alors il existe un vecteur gaussien d'espérance m et de matrice de covariance Γ .*

Afin de prouver le théorème on commence par établir un résultat d'algèbre linéaire :

Lemme 3 *Soit Γ une matrice d'ordre n , symétrique et positive. Alors il existe une matrice A , carré d'ordre n telle que $\Gamma = AA^*$.*

Preuve : (Lemme). Γ étant symétrique, il existe une matrice U orthogonale ($UU^* = U^*U = \text{Id}$) telle que $D_1 = U^*\Gamma U$ soit une matrice diagonale. On note $\lambda_1, \lambda_2, \dots, \lambda_n$ les éléments situés sur la diagonale de D_1 . Γ étant de plus positive, $\lambda_i \geq 0$ pour tout $i \in \{1, 2, \dots, n\}$. Soit D la matrice diagonale, dont les éléments situés sur la diagonale sont $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n}$. On pose $A = UD$. Rappelons que $U^{-1} = U^*$, donc $\Gamma = UD_1U^*$. Mais $D_1 = D^2 = DD^*$, d'où $\Gamma = UDD^*U^* = UD(UD)^* = AA^*$. $\square$

Preuve : (Théorème). Soient $Y_1, Y_2, \dots, Y_n$, n v.a. gaussiennes réelles, indépendantes, équidistribuées et de lois $\mathcal{N}_1(0, 1)$. On pose $Y^* = (Y_1, \dots, Y_n)$. Y est un vecteur gaussien. On définit X par :

$$X = m + AY, \tag{3.23}$$

où A est une matrice $n \times n$, telle que $\Gamma = AA^*$. D'après la proposition 23, X est un vecteur gaussien

$$\mathbb{E}(X) = m + \mathbb{E}(AY) = m + A\mathbb{E}(Y) = m,$$

$$K_X = AK_YA^* = A\text{Id}A^* = AA^* = \Gamma.$$

$\square$

Notation. $\mathcal{N}_n(m, \Gamma)$ désigne la loi d'un vecteur gaussien dans $\mathbb{R}^n$ de moyenne m et de matrice

de covariance Γ .

Remarque. Lorsque $n = 1$, la formule (3.23) est analogue à celle que l'on a vue en dimension 1 : $X = m + \sigma Y$ avec $Y \sim \mathcal{N}_1(0, 1)$ et σ l'écart-type de X .

Il est bon de retenir le procédé pour engendrer un vecteur gaussien de moyenne m et de matrice de covariance Γ : on considère

- A une "racine carrée positive" de Γ , c'est-à-dire une matrice carrée telle que $\Gamma = AA^*$.
- $Y_1, Y_2, \dots, Y_n$ n v.a.r., indépendantes de loi gaussienne réduite et centrée.

Alors

$$X = m + AY \text{ suit la loi } \mathcal{N}_n(m, \Gamma). \quad (3.24)$$

Théorème 11 Soit X un vecteur gaussien, $X^* = (X_1, \dots, X_n)$. Alors les v.a. $X_1, X_2, \dots, X_n$ sont indépendantes si et seulement si la matrice de covariance K de X est diagonale.

Preuve : Il est clair que si $X_1, \dots, X_n$ sont indépendantes alors $K(i, j) = \text{cov}(X_i, X_j) = 0$ si $i \neq j$. Donc K est diagonale.

Etudions la réciproque. Nous avons montré :

$$\mathbb{E}(\exp(i \langle u, X \rangle)) = e^{i \langle u, \mathbb{E}(X) \rangle - \frac{1}{2} u^* K u}, \quad u \in \mathbb{R}^n. \quad (3.25)$$

Mais K étant diagonale, on a :

$$u^* K u = \sum_{l=1}^n u_l^2 K(l, l) = \sum_{l=1}^n u_l^2 \text{Var}(X_l). \quad (3.26)$$

Notons ϕ_{X_l} la fonction caractéristique de X_l : $\phi_{X_l}(s) = \mathbb{E}[e^{isX_l}]$, $s \in \mathbb{R}$. Si l'on choisit u tel que $u_i = 0$, pour tout $i \neq l$, (3.25) et (3.26) impliquent

$$\phi_{X_l}(u_l) = \mathbb{E}(\exp(iu_l X_l)) = e^{iu_l \mathbb{E}(X_l) - \frac{1}{2} u_l^2 \text{Var}(X_l)}.$$

On peut donc ré-écrire (3.25) sous la forme suivante :

$$\phi_{X_1}(u_1) \dots \phi_{X_n}(u_n) = \mathbb{E} \left[\exp \left(i \sum_{l=1}^n u_l X_l \right) \right] = \mathbb{E}[\exp(i \langle u, X \rangle)],$$

pour tout $u = (u_1, u_2, \dots, u_n)$. Ce qui signifie que les v.a. $X_1, X_2, \dots, X_n$ sont indépendantes. $\square$

Théorème 12 Soit X un vecteur gaussien à valeurs dans $\mathbb{R}^n$, de matrice de covariance K .

1) X admet une densité si et seulement si K est inversible.

2) Si K est inversible, la densité de X est donnée par :

$$f(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{(\det K)^{1/2}} \exp - \left(\frac{1}{2} (x - m)^* K^{-1} (x - m) \right) \quad (3.27)$$

avec $m = \mathbb{E}(X)$.

Preuve : 1) a) On commence par établir un résultat préliminaire. Soit H un sous-espace vectoriel de $\mathbb{R}^n$, de dimension strictement plus petite que n . Si ξ est une v.a. à densité alors $\mathbb{P}(\xi \in H) = 0$. En effet soit H' un hyperplan contenant H . Quitte à changer les coordonnées on peut supposer que $H' = \{(x_1, x_2, \dots, x_n); x_n = 0\}$. Notons φ la densité de ξ , on a :

$$\mathbb{P}(\xi \in H) = \mathbb{P}(\xi \in H') = \int_{\mathbb{R}^n} \varphi(x_1, x_2, \dots, x_n) \mathbb{1}\{x_n = 0\} dx_1 dx_2 \dots dx_n = 0.$$

b) On a vu (cf. formule (3.24)) que X a même loi que $m + AY$, où $AA^* = K$ et Y est un vecteur dont toutes les composantes admettent des densités. Si A n'est pas inversible, A considéré comme application linéaire a une image H strictement incluse dans $\mathbb{R}^n$. Donc $AY \in H$. Raisonnons par l'absurde : si X admet une densité, $X - m$ aussi ; d'après ce qui précède on a $\mathbb{P}(X - m \in H) = 0$ et $\mathbb{P}(X - m \in H) = P(AY \in H) = 1$. Par conséquent X n'a pas de densité.

Lorsque A est inversible, l'application $y \mapsto m + Ay$ étant bijective et de classe $\mathcal{C}^1$, la v.a. $m + AY$ admet une densité.

Puisque $AA^* = K$, $\det(A) \det(A^*) = (\det(A))^2 = \det(K)$, on a l'équivalence :

$$A \text{ inversible} \iff K \text{ est inversible.}$$

2) Soit Y un vecteur gaussien centré de matrice de covariance identité. Il est clair que Y admet pour densité :

$$g(y) = \frac{1}{(2\pi)^{n/2}} \exp - \frac{1}{2} \langle y, y \rangle.$$

Posons $X' = AY + m$ donc $Y = A^{-1}(X' - m)$. Le jacobien de $x \rightarrow A^{-1}(x - m)$ est A^{-1} , son déterminant vaut $\det(A^{-1}) = \frac{1}{\det(A)}$. Mais

$(\det(A))^2 = \det(K)$ et $K^{-1} = (AA^*)^{-1} = (A^*)^{-1}A^{-1}$. De plus,

$$\begin{aligned} \langle y, y \rangle &= \langle A^{-1}(x - m), A^{-1}(x - m) \rangle = (x - m)^* (A^{-1})^* A^{-1} (x - m) \\ &= (x - m)^* K^{-1} (x - m). \end{aligned}$$

Donc X' admet pour densité la fonction f définie par (3.27). Puisque X et X' ont même loi, X admet pour densité f . $\square$

3.4 Convergence vers la loi normale

On rappelle la loi forte des grands nombres :

Théorème 13 Soit $(X_n; n \geq 1)$ une suite de v.a. à valeurs dans $\mathbb{R}^k$, indépendantes, de même loi et telles que $\mathbb{E}[\|X_1\|] < \infty$, $\mathbb{E}[X_1] = m \in \mathbb{R}^k$. Alors

$$\frac{X_1 + X_2 + \dots + X_n}{n}(\omega) \longrightarrow m, \quad \text{quand } n \rightarrow \infty, \quad \text{pour tout } \omega \in \Omega$$

sauf sur un ensemble de probabilité nulle.

Le théorème de la limite centrale précise un peu plus cette convergence :

Théorème 14 Soit $(X_n; n \geq 1)$ une suite de v.a. à valeurs dans $\mathbb{R}^k$, indépendantes et de même loi. On suppose que ces v.a. admettent un moment d'ordre 2. On note m leur espérance commune et Γ la covariance de X_i . Alors

$$\frac{X_1 + X_2 + \dots + X_n - nm}{\sqrt{n}}$$

converge en loi, lorsque $n \rightarrow \infty$ vers une v.a. de loi $\mathcal{N}_k(0, \Gamma)$.

Preuve : Pour démontrer la convergence en loi d'une suite de variables aléatoires T_n vers une variable aléatoire T il suffit de montrer la convergence des fonctions caractéristiques (voir Appendice).

On pose $T_n = \frac{X_1 + X_2 + \dots + X_n - nm}{\sqrt{n}}$.

On calcule alors la fonction caractéristique, qui peut s'écrire comme le produit de n fonctions caractéristiques car les variables aléatoires X_i sont indépendantes.

$$\Phi_{T_n}(t) = \mathbb{E}[\exp i \langle t, T_n \rangle] = e^{-i \langle t, \sqrt{n}m \rangle} \mathbb{E} \left[\exp i \left\langle \frac{t}{\sqrt{n}}, X_1 \right\rangle \right]^n.$$

Par une simple translation, on se ramène à considérer des variables aléatoires centrées.

On pose $\tilde{X}_1 = X_1 - m$. Alors $\Phi_{T_n}(t) = \mathbb{E} \left(\exp \frac{i}{\sqrt{n}} \langle t, \tilde{X}_1 \rangle \right)^n$. Comme $\tilde{X}_1$ admet un moment d'ordre 2, sa fonction caractéristique, c'est-à-dire sa transformée de Fourier, admet un développement de Taylor d'ordre 2, $\Phi_{X_1}(t) = 1 - \frac{1}{2} t^* K t + o(|t|^2)$. Ainsi, quand $n \rightarrow \infty$,

$$\Phi_{T_n}(t) = \left(1 - \left(\frac{t}{\sqrt{n}} \right)^* K \left(\frac{t}{\sqrt{n}} \right) + o \left(\frac{|t|^2}{n} \right) \right)^n \longrightarrow \exp -\frac{1}{2} t^* K t.$$

□

Corollaire 5 Soit $(X_n; n \geq 1)$ une suite de v.a.r. de Bernoulli, indépendantes et de même

loi, $\mathbb{P}(X_n = 1) = p$, $\mathbb{P}(X_n = 0) = 1 - p = q$. Alors

$$\frac{X_1 + \dots + X_n - np}{\sqrt{npq}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_1(0, 1).$$

Remarque. En pratique dès que $np > 15$, on approxime la loi de $\frac{X_1 + \dots + X_n - np}{\sqrt{npq}}$ par $\mathcal{N}_1(0, 1)$. Notons que $X_1 + \dots + X_n$ suit $\mathcal{B}(n, p)$.

3.5 Loi du chi-deux

Pour tout $\lambda > 0$ et $a > 0$, on note $\gamma(\lambda, a)$ la loi sur $\mathbb{R}$ de densité :

$$\frac{x^{\lambda-1}}{a^\lambda} \frac{1}{\Gamma(\lambda)} \left(\exp - \frac{x}{a} \right) \mathbb{1}_{\{x>0\}}, \quad (3.28)$$

où $\Gamma(\lambda) = \int_0^\infty x^{\lambda-1} e^{-x} dx$. Le but de cette section est de montrer qu'il existe un lien étroit entre les lois gamma et gaussiennes. On commence par un résultat préliminaire.

Proposition 25 1) Si X et Y sont deux v.a. indépendantes, la loi de X (resp. Y) a pour densité f (resp. g) alors $X + Y$ a pour densité $f * g$ où

$$f * g(x) = \int_{\mathbb{R}} f(x-t)g(t)dt = \int_{\mathbb{R}} g(x-t)f(t)dt; \quad x \in \mathbb{R}.$$

2) En particulier si, X a pour loi $\gamma(\lambda, a)$ et Y a pour distribution $\gamma(\mu, a)$ alors $X + Y$ suit la loi $\gamma(\lambda + \mu, a)$.

3) Si X a pour loi $\gamma(\lambda, a)$ et $\rho > 0$ alors ρX suit une loi $\gamma(\lambda, \rho a)$.

Preuve : 1) Le premier résultat est classique, on refait brièvement la démonstration. Soit φ une fonction borélienne positive et $\Delta = \mathbb{E}[\varphi(X + Y)]$. Puisque le couple (X, Y) a pour densité $f(x)g(y)$, on a,

$$\Delta = \iint_{\mathbb{R}^2} \varphi(x+y)f(x)g(y) dx dy.$$

On fait le changement de variables : $u = x$, $v = x + y$, d'où

$$\Delta = \iint_{\mathbb{R}^2} \varphi(v)f(u)g(v-u)dudv = \int_{\mathbb{R}} \varphi(v)(f * g)(v)dv.$$

2) Etudions le cas particulier où $X \sim \gamma(\lambda, a)$ et $Y \sim \gamma(\mu, a)$. $X + Y$ a pour densité h , avec

$$h(x) = \int_0^x \frac{1}{a^\lambda} (x-t)^{\lambda-1} \frac{1}{\Gamma(\lambda)} e^{-\frac{x-t}{a}} \frac{1}{a^\mu} t^{\mu-1} \frac{1}{\Gamma(\mu)} e^{-\frac{t}{a}} dt; \quad x \geq 0,$$

et $h(x) = 0$ si $x < 0$. On en déduit,

$$h(x) = c \left(\int_0^x (x-t)^{\lambda-1} t^{\mu-1} dt \right) e^{-x/a}, \quad \text{avec } c = \frac{1}{a^{\lambda+\mu}} \frac{1}{\Gamma(\lambda)\Gamma(\mu)}. \quad (3.29)$$

On fait le changement de variable $t = ux$, on obtient

$$h(x) = c' x^{\lambda+\mu-1} e^{-x/a}; \quad c' = c \int_0^1 (1-u)^{\lambda-1} u^{\mu-1} du. \quad (3.30)$$

Mais on sait que h est une densité, d'où

$$\int_{\mathbb{R}} h(x) dx = c' \int_0^{\infty} x^{\lambda+\mu-1} e^{-x/a} dx = c' a^{\lambda+\mu} \Gamma(\lambda + \mu) = 1. \quad (3.31)$$

On a montré que $X + Y$ suit une loi $\gamma(\lambda + \mu, a)$.

3) Si $X \sim \gamma(\lambda, a)$ et $\rho > 0$, alors un changement de variable immédiat montre que ρX suit une loi $\gamma(\lambda, a\rho)$. $\square$

Remarque.

En utilisant (3.29), (3.30) et (3.31) on a montré :

$$\int_0^1 (1-t)^{\lambda-1} t^{\mu-1} dt = \frac{\Gamma(\lambda)\Gamma(\mu)}{\Gamma(\lambda + \mu)}; \quad \lambda > 0, \quad \mu > 0. \quad (3.32)$$

Définition 18 On appelle loi du chi-deux à n degrés de liberté, la loi de v.a. $\sum_{i=1}^n X_i^2$, où les v.a. $X_1, X_2, \dots, X_n$ sont indépendantes et ont pour loi commune la loi gaussienne réduite et centrée. On note $\chi^2(n)$ cette distribution.

Proposition 26 La loi du $\chi^2(n)$ coïncide avec la loi $\gamma(n/2, 2)$. En particulier la loi de $\sum_{i=1}^n X_i^2$ est $\gamma(n/2, 2)$, où les v.a. $X_1, X_2, \dots, X_n$ sont indépendantes et de loi $\mathcal{N}_1(0, 1)$.

Preuve : 1) On commence par calculer la loi de X_1^2 . Posons $\Delta = \mathbb{E}[f(X_1^2)]$ où f est une fonction borélienne positive. On a :

$$\Delta = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x^2) \exp\left(-\frac{x^2}{2}\right) dx = \int_0^{\infty} f(x^2) \exp\left(-\frac{x^2}{2}\right) dx.$$

On fait le changement de variable $y = x^2$, il vient

$$\Delta = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} f(y) y^{-1/2} \exp(-y/2) dy.$$

Donc $\chi^2(1) = \gamma(1/2, 2)$.

2) Posons $\xi_n = \sum_{i=1}^n X_i^2$. Montrons par récurrence sur $n \geq 1$: $\xi_n \sim \gamma(n/2, 2)$.

On a vu à l'étape précédente que $\xi_1 \sim \gamma(1/2, 2)$, la propriété est réalisée pour $n = 1$. Explicitons $n \implies n + 1$. On remarque que X_{n+1}^2 est indépendante de ξ_n , $\xi_n \sim \gamma(n/2, 2)$ et $X_{n+1}^2 \sim \gamma(1/2, 2)$. D'après le 2) de la proposition 25, $\xi_{n+1} \sim \gamma(n/2 + 1/2, 2) = \gamma(\frac{n+1}{2}, 2)$. $\square$

On s'intéresse à une suite de n variables aléatoires gaussiennes. On introduit tout d'abord la notion de moyenne et de variance empirique.

Soient $X_1, X_2, \dots, X_n$, n variables aléatoires indépendantes et de même loi à valeurs réelles. On définit deux nouvelles v.a. $\bar{X}_n$ et S_n^2 la **moyenne** et la **variance empirique** :

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n); \quad S_n^2 = \frac{1}{n-1} \left(\sum_{k=1}^n (X_k - \bar{X})^2 \right). \quad (3.33)$$

En fait, la renormalisation étonnante de la variance empirique provient du fait que la moyenne et la variance empirique sont des estimateurs sans biais de l'espérance et de la variance de la loi de X_1 : si $m \in \mathbb{R}$ est l'espérance de X_1 et σ^2 sa variance, on a (exercice) : $\mathbb{E}[\bar{X}_n] = m$ et $\mathbb{E}[S_n^2] = \sigma^2$. Par ce procédé on peut donc obtenir une bonne approximation de la variance d'une variable aléatoire par la variance empirique. Il est alors important de connaître l'erreur commise dans un problème donné en prenant à la place de la variance (inconnue) la variance empirique (qui est aléatoire). On s'intéresse donc à la loi de S_n^2 qui précisera sa distribution autour de son espérance.

Théorème 15 Soient $X_1, X_2, \dots, X_n$, n v.a. réelles, indépendantes, de même loi gaussienne $\mathcal{N}_1(m, \sigma^2)$. Alors

- 1) $\bar{X}_n$ et S_n^2 sont indépendantes.
- 2) $\bar{X}_n$ suit une loi $\mathcal{N}_1(m, \frac{\sigma^2}{n})$ et $\frac{n-1}{\sigma^2} S_n^2$ a pour loi $\chi^2(n-1)$.

Preuve : 1) La première étape consiste à vérifier que l'on peut se ramener au cas où $m = 0$ et $\sigma = 1$. On pose

$$X'_i = \frac{X_i - m}{\sigma} \iff X_i = \sigma X'_i + m \quad 1 \leq i \leq n. \quad (3.34)$$

Les v.a. $X'_1, X'_2, \dots, X'_n$ sont indépendantes, chacune de loi $\mathcal{N}_1(0, 1)$. On note $\bar{X}'_n$ et S'^2_n la moyenne et la variance empirique de cette nouvelle suite. On déduit de (3.34),

$$\bar{X}_n = \sigma \bar{X}'_n + m. \quad (3.35)$$

Par conséquent, $X_i - \bar{X}_n = \sigma(X'_i - \bar{X}'_n)$. On en déduit

$$S_n^2 = \sigma^2 S'^2_n. \quad (3.36)$$

Donc si le théorème 15 est réalisé lorsque $m = 0$ et $\sigma = 1$, d'après (3.35) et (3.36) ce théorème l'est encore dans le cas général.

2) On suppose que chaque v.a. X_i est de loi $\mathcal{N}_1(0, 1)$.

Soit u_1 le vecteur de $\mathbb{R}^n$ de coordonnées $\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}$. Ce vecteur est de norme 1, il existe donc $u_2, u_3, \dots, u_n$ de $\mathbb{R}^n$ tels que $\mathcal{B} = \{u_1, u_2, \dots, u_n\}$ soit une base orthonormée de $\mathbb{R}^n$.

Soit A la matrice carrée d'ordre n , dont les colonnes sont $u_1, u_2, \dots, u_n$. Par construction A est une matrice orthogonale : $AA^* = A^*A = \text{Id}$. On note $Y = A^*X$, et $Y^* = (Y_1, Y_2, \dots, Y_n)$. D'après la proposition 23, Y est un vecteur gaussien centré de matrice de covariance K_Y :

$$K_Y = A^*K_X(A^*)^* = A^*\text{Id}.A = A^*A = \text{Id},$$

les v.a. $X_1, X_2, \dots, X_n$ étant indépendantes, réduites et centrées la matrice de covariance K_X de X est l'identité. Puisque la première ligne de A^* est $u_1^* = \left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right)$, on a :

$$Y_1 = \frac{1}{\sqrt{n}}(X_1 + X_2 + \dots + X_n) = \sqrt{n}\bar{X}_n. \quad (3.37)$$

On va exprimer S_n^2 à l'aide de Y .

$$\begin{aligned} (n-1)S_n^2 &= \sum_{k=1}^n (X_k - \bar{X}_n)^2 = \sum_{k=1}^n (X_k^2 - 2X_k\bar{X}_n + \bar{X}_n^2) \\ &= \left(\sum_{k=1}^n X_k^2\right) - 2\bar{X}_n \left(\sum_{k=1}^n X_k\right) + n\bar{X}_n^2. \end{aligned}$$

Par conséquent,

$$(n-1)S_n^2 = \left(\sum_{k=1}^n X_k^2\right) - 2\bar{X}_n(n\bar{X}_n) + n\bar{X}_n^2 = \left(\sum_{k=1}^n X_k^2\right) - n\bar{X}_n^2. \quad (3.38)$$

Mais $Y = A^*X$, A^* est une matrice orthogonale, elle conserve la norme :

$$\sum_{k=1}^n Y_k^2 = \sum_{k=1}^n X_k^2.$$

On déduit de (3.37) et (3.38) :

$$(n-1)S_n^2 = \sum_{k=1}^n Y_k^2 - Y_1^2 = \sum_{k=2}^n Y_k^2.$$

Y est un vecteur gaussien de matrice de covariance identité, les v.a. $Y_1, Y_2, \dots, Y_n$ sont donc indépendantes et ont pour loi $\mathcal{N}_1(0, 1)$, on en déduit l'indépendance de $\bar{X}_n$ et S_n^2 . De plus $\bar{X}_n \sim \mathcal{N}_1(0, 1/n)$, et d'après la proposition 26, la loi de S_n^2 est $\chi^2(n-1)$. $\square$

3.6 Application à la statistique

Nous terminerons ce chapitre en donnant une application du théorème 15 à la théorie des tests statistiques.

On s'intéresse à l'approximation d'une loi discrète sur $E = \{1, 2, \dots, k\}$, notée $p = (p_1, p_2, \dots, p_k)$:

$$p_i > 0, \quad \forall i \in \{1, 2, \dots, k\} \quad \text{et} \quad \sum_{i=1}^k p_i = 1. \quad (3.39)$$

On considère $Y_1, Y_2, \dots, Y_n$, un échantillon de loi p ; ce qui signifie que les v.a. $Y_1, Y_2, \dots, Y_n$ sont indépendantes et ont même loi p . On introduit :

$$N_n(i) = \#\{j; 1 \leq j \leq n, Y_j = i\} = \sum_{j=1}^n \mathbb{1}_{\{Y_j=i\}}; \quad i \in E. \quad (3.40)$$

$N_n(i)$ est le nombre de fois où $Y_j = i$, lorsque j varie de 1 à n , $\frac{N_n(i)}{n}$ est donc la fréquence d'apparition de i .

Proposition 27 Soit $\sqrt{p}$ le vecteur de $\mathbb{R}^k$ de coordonnées $(\sqrt{p_1}, \sqrt{p_2}, \dots, \sqrt{p_k})$. On note

$$T_n := \left(\frac{N_n(1) - np_1}{\sqrt{np_1}}, \frac{N_n(2) - np_2}{\sqrt{np_2}}, \dots, \frac{N_n(k) - np_k}{\sqrt{np_k}} \right); \quad n \geq 1.$$

Alors T_n converge en loi, lorsque $n \rightarrow \infty$, vers une v.a. de loi gaussienne $\mathcal{N}_k(0, \Gamma)$ avec $\Gamma = \text{Id} - \sqrt{p}(\sqrt{p})^*$.

Preuve : Soit $(e_1, e_2, \dots, e_k)$ la base canonique de $\mathbb{R}^k$. On introduit :

$$X_j = \sum_{l=1}^k \mathbb{1}_{\{Y_j=l\}} e_l; \quad 1 \leq j \leq n. \quad (3.41)$$

$(X_j; j \geq 1)$ est une famille de vecteurs aléatoires, indépendants, de même loi, et à valeurs dans $\mathbb{R}^k$. Les X_j ont un moment d'ordre 2, d'après le théorème central limite,

$$Z_n = \frac{1}{\sqrt{n}} \left(\sum_{j=1}^n X_j - n\mathbb{E}(X) \right), \quad \text{converge en loi vers} \quad \mathcal{N}_k(0, A),$$

où A désigne la matrice de covariance de X_1 . D'après (3.41),

$$Z_n = \frac{1}{\sqrt{n}} \left(\sum_{l=1}^k \left\{ \sum_{j=1}^n \mathbb{1}_{\{Y_j=l\}} - np_l \right\} e_l \right) = \frac{1}{\sqrt{n}} \left(\sum_{l=1}^k (N_n(l) - np_l) e_l \right).$$

Donc

$$Z_n^* = \left(\frac{N_n(1) - np_1}{\sqrt{n}}, \frac{N_n(2) - np_2}{\sqrt{n}}, \dots, \frac{N_n(k) - np_k}{\sqrt{n}} \right).$$

Il reste donc à évaluer la matrice $A = (a_{i,j})$.

$$a_{i,j} = \text{cov}(\mathbb{1}_{\{Y_1=i\}}, \mathbb{1}_{\{Y_1=j\}}) = \mathbb{P}(Y_1 = i, Y_1 = j) - \mathbb{P}(Y_1 = i)\mathbb{P}(Y_1 = j).$$

D'où $a_{i,j} = p_i\delta_{i,j} - p_ip_j$, avec $\delta_{i,j} = \mathbb{1}_{\{i=j\}}$. Soit u un vecteur de $\mathbb{R}^k$, $u^* = (u_1, u_2, \dots, u_k)$. On lui associe $v \in \mathbb{R}^k$, $v^* = (u_1/\sqrt{p_1}, u_2/\sqrt{p_2}, \dots, u_k/\sqrt{p_k})$. On a :

$$\langle u, T_n \rangle = u^* T_n = v^* Z_n = \langle v, Z_n \rangle.$$

Soit Φ_n la fonction caractéristique de T_n ; alors

$$\Phi_n(u) = \mathbb{E}[\exp(i \langle u, T_n \rangle)] = \mathbb{E}[\exp(i \langle v, Z_n \rangle)].$$

Sachant que Z_n converge en loi, vers $\mathcal{N}_k(0, A)$,

$$\lim_{n \rightarrow \infty} \Phi_n(u) = \exp -\frac{1}{2} v^* A v.$$

Un calcul immédiat conduit à :

$$\begin{aligned} v^* A v &= \sum_{i=1}^k a_{ii} v_i^2 + \sum_{1 \leq i \neq j \leq k} a_{i,j} v_i v_j = \sum_{i=1}^k (1 - p_i) u_i^2 - \sum_{1 \leq i \neq j \leq k} \sqrt{p_i} \sqrt{p_j} u_i u_j \\ &= \sum_{i=1}^k u_i^2 - \left(\sum_{i=1}^k \sqrt{p_i} u_i \right) \left(\sum_{j=1}^k \sqrt{p_j} u_j \right) = u^* \Gamma u. \end{aligned}$$

□

Considérons alors

$$d_n^2 = \|T_n\|^2 = \sum_{i=1}^k \left(\frac{N_n(i) - np_i}{\sqrt{np_i}} \right)^2 = \sum_{i=1}^k \frac{n}{p_i} \left(\frac{N_n(i)}{n} - p_i \right)^2. \quad (3.42)$$

$\frac{d_n^2}{n}$ correspond au carré d'une pseudo-distance dans $\mathbb{R}^k$, entre la fréquence empirique $\left(\frac{N_n(i)}{n}; 1 \leq i \leq k \right)$ et $(p_i; 1 \leq i \leq k)$.

Proposition 28 d_n^2 converge en loi, lorsque $n \rightarrow \infty$, vers la loi $\chi^2(k-1)$.

Avant de démontrer ce résultat, introduisons un lemme préliminaire :

Lemme 4 Soit $\Gamma = \text{Id.} - (\sqrt{p})(\sqrt{p})^*$. Alors il existe une matrice orthogonale B telle que

$$B\Gamma B^* = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}$$

Preuve : Soit u un vecteur de $\mathbb{R}^k$, alors

$$\sqrt{p}(\sqrt{p})^*u = \langle \sqrt{p}, u \rangle \sqrt{p}.$$

Par conséquent $u \rightarrow \sqrt{p}(\sqrt{p})^*u$ est l'opérateur de projection orthogonale sur le vecteur normé. Donc Γ est la projection orthogonale sur F , F désignant le sous-espace vectoriel de $\mathbb{R}^k$ orthogonal à $\sqrt{p}$: $F = \{v \in \mathbb{R}^k : \langle v, \sqrt{p} \rangle = 0\}$. Il existe une base orthonormée $\mathcal{B}_1$ pour laquelle Γ soit représenté par une matrice diagonale dont la diagonale est formée uniquement de 1 sauf le dernier élément qui est nul. On choisit B^* , la matrice de changement de base, de la base canonique à $\mathcal{B}_1$. B est une matrice orthogonale, le lemme 4 est alors vérifié. $\square$

Preuve : (Proposition 28)

Puisque $x \in \mathbb{R}^k \rightarrow \|x\|^2$ est continue, d'après la proposition 27, la v.a.r d_n^2 converge en loi vers $\|Z\|^2$, Z de loi $\mathcal{N}_k(0, \Gamma)$.

Posons $Z' = BZ$, où B est la matrice définie par le lemme 4. Puisque B est une matrice orthogonale, $\|Z'\|^2 = \|Z\|^2$. De plus Z' est un vecteur gaussien centré, de matrice de covariance $K_{Z'} = B\Gamma B^*$. Ainsi la loi de $\|Z'\|^2$ est $\chi^2(k-1)$. $\square$

Application : le test du chi-deux

La théorie des tests consiste à formuler des hypothèses particulières sur les paramètres ou sur les lois qui interviennent dans les problèmes étudiés, puis à apporter un jugement sur ces hypothèses. Ce jugement est basé d'une part sur les résultats obtenus sur un ou plusieurs échantillons extraits de la population concernée et d'autre part, sur l'acceptation d'un certain *risque* dans la prise de décision. Le test du chi-deux est un test de conformité de deux distributions, ou encore un *test d'ajustement* entre une distribution théorique et une distribution expérimentale.

On veut tester si une v.a. Y a pour loi p , c'est-à-dire si

$$\mathbb{P}(Y = i) = p_i; \quad 1 \leq i \leq k.$$

On définit deux hypothèses possibles :

$$H_0 : Y \text{ a pour loi } p, \quad H_1 : Y \text{ n'a pas la loi } p,$$

et on aimerait choisir entre les deux suivant les données fournies par un échantillon. Pour établir la crédibilité de l'hypothèse H_0 , des règles très précises doivent être énoncées pour permettre

de conclure au rejet ou à l'acceptation de H_0 . Cependant pour des événements dans lesquels intervient le hasard, il est impossible de prendre une bonne décision sans risque de se tromper. Il faut donc mettre en oeuvre une règle conduisant à rejeter H_0 , si elle est vraie, que dans une faible proportion des cas. Cette décision a un caractère probabiliste, toute décision comporte un risque qu'elle soit erronée. Ce risque, noté α , qui est le risque de rejeter à tort l'hypothèse H_0 alors qu'elle est vraie et qui favorise donc l'hypothèse H_1 s'appelle *seuil de signification* ou *risque de première espèce*.

$$\alpha = \mathbb{P}\{\text{rejeter } H_0 \mid H_0 \text{ vraie}\} = \mathbb{P}\{\text{choisir } H_1 \mid H_0 \text{ vraie}\}.$$

La probabilité α , appelée aussi *risque du client*, est choisie a priori par l'utilisateur. Les valeurs les plus utilisées pour α sont 0,05 et 0,01.

La règle de décision comporte un risque β ou *risque de deuxième espèce*, c'est le risque de ne pas rejeter H_0 alors que H_1 est vraie :

$$\beta = \mathbb{P}\{\text{ne pas rejeter } H_0 \mid H_1 \text{ vraie}\}.$$

Cette probabilité, appelée également *risque du fournisseur* (celui de voir, par exemple, une bonne production refusée), dépend de l'hypothèse alternative H_1 .

Élaboration d'un test et démarche à suivre :

- 1) formuler de façon précise les hypothèses H_0 et H_1 ,
- 2) fixer, avant l'expérience, le risque α de première espèce, c'est-à-dire le risque de rejeter l'hypothèse H_0 alors qu'elle est vraie,
- 3) choisir la statistique la mieux adaptée en fonction des caractéristiques de la population étudiée,
- 4) déterminer la région critique ou région de rejet de l'hypothèse H_0 au profit de l'hypothèse H_1 et en déduire la règle de décision,
- 5) calculer effectivement la valeur numérique de la variable de décision en utilisant les données de l'échantillon,
- 6) donner les conclusions du test.

Dans le cas qui nous intéresse (le cadre du test du chi-deux), la variable de décision est la pseudo-distance entre les fréquences empiriques et p . En effet, on suppose que l'on peut reproduire des copies indépendantes de $Y : Y_1, Y_2, \dots, Y_n$, qui sera notre échantillon.

On définit alors d_n^2 comme dans (3.42). Si d_n^2 est grand, on choisira de préférence l'hypothèse H_1 , par contre, si d_n^2 est petit, les fréquences empiriques sont proches de la probabilité attendue p et donc on choisira l'hypothèse H_0 . Il faut donc trouver un seuil θ_α , dépendant du risque α , qui définira la région de rejet.

$$\mathbb{P}(d_n^2 > \theta_\alpha) = \alpha.$$

En appliquant la proposition 28, on obtient, pour n suffisamment grand

$$\mathbb{P}(d_n^2 > \theta_\alpha) \sim \mathbb{P}(\xi > \theta_\alpha),$$

où ξ suit une loi $\chi^2(k-1)$. Le test du chi-deux consiste donc à définir le seuil θ_α par l'équation

$$\mathbb{P}(\xi > \theta_\alpha) = \alpha.$$

Dans la pratique, n assez grand veut dire $np_i \geq 10$ pour tout $i \in \{1, 2, \dots, k\}$.

Chapitre 4

Conditionnement

4.1 Espérance conditionnelle

On commence par un rappel sur les espaces de Hilbert. Soit H un espace de Hilbert et F un sous-espace fermé. Pour tout x de H il existe un unique $y \in F$, appelé projection orthogonale de x sur F , vérifiant l'une des conditions équivalentes :

$$\forall z \in F, \quad \langle x - y, z \rangle = 0. \quad (4.1)$$

$$\forall z \in F, \quad \langle x, z \rangle = \langle y, z \rangle. \quad (4.2)$$

$$\forall z \in F, \quad \|x - y\| \leq \|x - z\|. \quad (4.3)$$

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. On supposera que $\mathcal{A}$ est une tribu complète, c'est-à-dire qu'elle contient tous les ensembles $\mathbb{P}$ -négligeables : pour tout A tel que $\mathbb{P}(A) = 0$, alors $A \in \mathcal{A}$. On note par $\mathcal{N}$ l'ensemble des ensembles $\mathbb{P}$ -négligeables.

On rappelle que $L^2(\Omega, \mathcal{A}, \mathbb{P})$ désigne l'ensemble des classes d'équivalence de v.a. : deux v.a. Y et Y' sont égales dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ si $Y - Y'$ est presque sûrement nulle.

On considère une sous tribu $\mathcal{B}$ de $\mathcal{A}$. Si $\mathcal{N} \in \mathcal{B}$, alors l'espace $L^2(\Omega, \mathcal{B}, \mathbb{P})$ est inclus dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$. Dans le cas général, on pose $\mathcal{B}' = \mathcal{B} \cup \mathcal{N}$, par définition $\mathcal{B}' \subset \mathcal{A}$. On vérifie facilement que $\mathcal{B}'$ est une tribu. $\mathcal{B}'$ est la plus petite tribu complète contenant $\mathcal{B}$ et incluse dans $\mathcal{A}$. De plus $L^2(\Omega, \mathcal{B}', \mathbb{P}) \subset L^2(\Omega, \mathcal{A}, \mathbb{P})$. Par abus, on identifiera $L^2(\Omega, \mathcal{B}, \mathbb{P})$ et $L^2(\Omega, \mathcal{B}', \mathbb{P})$.

Proposition 1 *Pour toute v.a. X de $L^2(\Omega, \mathcal{A}, \mathbb{P})$, il existe une unique (classe de) v.a. Y telle que :*

$$Y \in L^2(\Omega, \mathcal{B}, \mathbb{P}), \quad (4.4)$$

$$\mathbb{E}[XZ] = \mathbb{E}[YZ], \text{ pour tout } Z \in L^2(\Omega, \mathcal{B}, \mathbb{P}). \quad (4.5)$$

Notation.

On note $Y = \mathbb{E}(X|\mathcal{B})$ ou encore $Y = \mathbb{E}^{\mathcal{B}}(X)$. Donc si $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, $\mathbb{E}(X|\mathcal{B})$ est caractérisé par les deux propriétés :

$$\mathbb{E}(X|\mathcal{B}) \in L^2(\Omega, \mathcal{B}, \mathbb{P}), \quad (4.6)$$

$$\mathbb{E}[Z\mathbb{E}(X|\mathcal{B})] = \mathbb{E}[ZX]; \quad \forall Z \in L^2(\Omega, \mathcal{B}, \mathbb{P}). \quad (4.7)$$

Preuve : (Proposition 1)

On choisit $H = L^2(\Omega, \mathcal{A}, \mathbb{P})$ et $F = L^2(\Omega, \mathcal{B}, \mathbb{P})$. On utilise le rappel précédent, les propriétés (4.2) et (4.5) étant équivalentes. $\square$

Exemples :

Exemple 1 : On commence par l'exemple le plus simple : $\mathcal{B} = \{\emptyset, \Omega\}$. Les v.a. $\mathcal{B}$ -mesurables sont les constantes, donc $\mathbb{E}(X|\mathcal{B}) = c$. Si Z est $\mathcal{B}$ -mesurable, $Z = a$, la propriété (4.7) devient :

$$\mathbb{E}[Z \mathbb{E}(X|\mathcal{B})] = \mathbb{E}[ac] = ac = \mathbb{E}[aX] = a\mathbb{E}(X), \quad \forall a \in \mathbb{R}.$$

Par conséquent $c = \mathbb{E}(X)$. En conclusion : $\mathbb{E}(X) = \mathbb{E}(X|\{\emptyset, \Omega\})$.

Exemple 2 : Soit $\{A_1, A_2, \dots, A_n\}$ une partition de Ω , telle que

$$A_i \in \mathcal{A}, \text{ et } 0 < \mathbb{P}(A_i) < 1, \text{ pour tout } 1 \leq i \leq n.$$

On note $\mathcal{B}$ la tribu engendrée par $A_1, A_2, \dots, A_n$: $\mathcal{B} = \sigma(A_1, A_2, \dots, A_n)$. Puisque les A_i forment une partition, on a l'équivalence :

$$B \in \mathcal{B} \iff B = \emptyset \text{ ou } B = \bigcup_{j=1}^k A_{i_j}. \quad (4.8)$$

Afin de calculer l'espérance conditionnelle de X sachant $\mathcal{B}$, il est nécessaire auparavant de caractériser les v.a. $\mathcal{B}$ -mesurables. Le résultat est le suivant :

$$\text{Une v.a. } X \text{ est } \mathcal{B} \text{ - mesurable } \iff X \text{ est constante sur chaque } A_i. \quad (4.9)$$

Il est clair que la condition est suffisante. Il s'agit de montrer le caractère nécessaire : supposons X est $\mathcal{B}$ -mesurable. A_i étant non vide ($\mathbb{P}(A_i) > 0$), on choisit $\omega_i \in A_i$. On note $x_i = X(\omega_i)$. Mais $\{X = x_i\}$ appartient à $\mathcal{B}$, et a une intersection non vide avec A_i , d'après (4.8), $\{X = x_i\}$ contient A_i . Par conséquent X vaut x_i sur A_i .

On peut à présent calculer $\mathbb{E}(X|\mathcal{B})$. D'après ce qui précède, $\mathbb{E}(X|\mathcal{B}) = \sum_{i=1}^n x_i \mathbb{1}_{A_i}$. On prend une v.a. test Z , également $\mathcal{B}$ -mesurable :

$$Z = \sum_{i=1}^n z_i \mathbb{1}_{A_i}.$$

On a :

$$\mathbb{E}[Z \mathbb{E}(X|\mathcal{B})] = \mathbb{E} \left[\sum_{i=1}^n x_i z_i \mathbb{1}_{A_i} \right] = \sum_{i=1}^n x_i z_i \mathbb{P}(A_i), \quad (4.10)$$

et

$$\mathbb{E}[ZX] = \mathbb{E} \left[\sum_{i=1}^n z_i X \mathbb{1}_{A_i} \right] = \sum_{i=1}^n z_i \mathbb{E}(X \mathbb{1}_{A_i}). \quad (4.11)$$

L'égalité entre (4.10) et (4.11) a lieu pour tout $z_1, z_2, \dots, z_n$, si et seulement si :

$$x_i = \frac{1}{\mathbb{P}(A_i)} \mathbb{E}(X \mathbb{1}_{A_i}) = \mathbb{E}(X|A_i), \quad 1 \leq i \leq n.$$

En conclusion,

$$\mathbb{E}(X|\mathcal{B}) = \sum_{i=1}^n \mathbb{E}(X|A_i) \mathbb{1}_{A_i}. \quad (4.12)$$

Proposition 2 1) *L'opérateur d'espérance conditionnelle est linéaire de $L^2(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans $L^2(\Omega, \mathcal{B}, \mathbb{P})$:*

$$\mathbb{E}(aX + bY|\mathcal{B}) = a\mathbb{E}(X|\mathcal{B}) + b\mathbb{E}(Y|\mathcal{B}) ; \forall (a, b) \in \mathbb{R}^2. \quad (4.13)$$

2) *Si X appartient à $L^2(\Omega, \mathcal{A}, \mathbb{P})$ et $X \geq 0$ alors $\mathbb{E}(X|\mathcal{B}) \geq 0$ p.s.*

Preuve : 1) Puisque la projection orthogonale sur un sous espace fermé d'un espace de Hilbert est un opérateur linéaire, il est clair que $X \rightarrow \mathbb{E}(X|\mathcal{B})$ est linéaire.

2) Supposons $X \geq 0$. On pose $A = \{\mathbb{E}(X|\mathcal{B}) < 0\}$. Puisque $\mathbb{E}(X|\mathcal{B})$ est $\mathcal{B}$ -mesurable, alors $A \in \mathcal{B}$. On choisit $Z = \mathbb{1}_A$ et on applique (4.5) :

$$\mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(\mathbb{E}(X|\mathcal{B})\mathbb{1}_A).$$

Mais

$$X\mathbb{1}_A \geq 0 \text{ et } \mathbb{E}(X|\mathcal{B})\mathbb{1}_A \leq 0 \implies \mathbb{E}(X\mathbb{1}_A) \geq 0 \text{ et } \mathbb{E}(\mathbb{E}(X|\mathcal{B})\mathbb{1}_A) \leq 0.$$

Par conséquent $\mathbb{E}(\mathbb{E}(X|\mathcal{B})\mathbb{1}_A) = 0$.

Rappelons que si U est une v.a. de signe constant, d'espérance nulle alors U est presque sûrement nulle. Donc $\mathbb{E}(X|\mathcal{B})\mathbb{1}_A = 0$ p.s. Mais $\mathbb{E}(X|\mathcal{B}) < 0$ sur A . Ainsi $\mathbb{P}(A) = 0$. $\square$

Remarque.

Si X et Y sont deux v.a. de $L^2(\Omega, \mathcal{A}, \mathbb{P})$ telles que $X \geq Y$, alors

$$\mathbb{E}(X|\mathcal{B}) \geq \mathbb{E}(Y|\mathcal{B}). \quad (4.14)$$

Il suffit en effet de remarquer que $X - Y \geq 0$. Ainsi

$$\mathbb{E}(X - Y|\mathcal{B}) = \mathbb{E}(X|\mathcal{B}) - \mathbb{E}(Y|\mathcal{B}) \geq 0.$$

Le cadre des espaces L^2 est très agréable pour définir l'espérance conditionnelle. Il est toutefois trop restrictif.

Théorème 1 *Soit X une v.a. $\mathcal{A}$ -mesurable.*

1) *Si X est intégrable, il existe une unique v.a. (à une égalité presque sûre près) intégrable, notée $\mathbb{E}(X|\mathcal{B})$ telle que*

$$\mathbb{E}(X|\mathcal{B}) \text{ est } \mathcal{B}\text{-mesurable.} \quad (4.15)$$

$$\mathbb{E}(ZX) = \mathbb{E}(Z\mathbb{E}(X|\mathcal{B})), \quad (4.16)$$

pour toute v.a. Z , $\mathcal{B}$ -mesurable bornée.

2) *Si X est positive, il existe une "unique" v.a. $\mathbb{E}(X|\mathcal{B}) \geq 0$ vérifiant (4.15) et (4.16) pour toute v.a. Z , $\mathcal{B}$ -mesurable, positive.*

Preuve : a) On commence par montrer l'existence de $Y = \mathbb{E}(X|\mathcal{B})$, lorsque $X \geq 0$. On se ramène au cadre L^2 , en tronquant X , on pose,

$$X_n = X \wedge n = \inf\{X, n\} ; n \geq 1.$$

X_n étant une v.a. bornée, X_n appartient à $L^2(\Omega, \mathcal{A}, \mathbb{P})$, on peut définir :

$$Y_n = \mathbb{E}(X_n|\mathcal{B}) ; n \geq 1.$$

Sachant que la suite X_n est croissante et positive, on déduit de (4.14) que Y_n est également croissante et positive. En particulier $(Y_n)_{n \geq 1}$ converge presque sûrement vers Y , Y est à valeurs dans $[0, +\infty]$.

Mais comme Y_n est $\mathcal{B}$ -mesurable, Y l'est aussi ; Y vérifie ainsi (4.15).

On vérifie à présent que Y satisfait (4.16), lorsque Z est une v.a. positive, bornée, $\mathcal{B}$ -mesurable. Les deux suites $(ZY_n ; n \geq 0)$ et $(ZX_n ; n \geq 0)$ sont positives et croissantes. D'après le théorème de convergence croissante,

$$\lim_{n \rightarrow \infty} \mathbb{E}(ZY_n) = \mathbb{E}(ZY) ; \lim_{n \rightarrow \infty} \mathbb{E}(ZX_n) = \mathbb{E}(ZX).$$

Mais $\mathbb{E}(ZY_n) = \mathbb{E}(Z \mathbb{E}(X_n|\mathcal{B})) = \mathbb{E}(ZX_n)$, $n \geq 1$. On passe à la limite, $n \rightarrow \infty$: $\mathbb{E}(ZX) = \mathbb{E}(ZY)$.

Supposons de plus X intégrable. On prend $Z = 1$, d'après (4.16) on a :

$$\mathbb{E}[\mathbb{E}(X|\mathcal{B})] = \mathbb{E}(X) < \infty, \text{ si } X \geq 0 \text{ et } X \in L^1(\Omega, \mathcal{A}, \mathbb{P}). \quad (4.17)$$

b) On montre que $\mathbb{E}(X|\mathcal{B})$ existe lorsque X est intégrable. On va se ramener au cas précédent, pour ce faire on écrit :

$$X = X_+ - X_-$$

où $x_+ = \sup(x, 0)$, et $x_- = \sup(0, -x)$. Puisque X est intégrable, X_+ et X_- le sont aussi. De plus $X_+ \geq 0$ et $X_- \geq 0$. On peut appliquer le résultat de l'étape précédente et (4.17) : il existe deux v.a. Y_1 et Y_2 , positives, $\mathcal{B}$ -mesurables, intégrables telles que

$$\mathbb{E}(ZY_1) = \mathbb{E}(ZX_+), \quad \mathbb{E}(ZY_2) = \mathbb{E}(ZX_-),$$

pour toute v.a. Z positive, bornée, $\mathcal{B}$ -mesurable. On pose $Y = Y_1 - Y_2$. Y est intégrable. Il est aisé de vérifier, grâce à la propriété de linéarité (4.13) que l'on a :

$$\begin{aligned} \mathbb{E}(ZX) &= \mathbb{E}(Z(X_+ - X_-)) = \mathbb{E}(ZX_+) - \mathbb{E}(ZX_-) = \mathbb{E}(ZY_1) - \mathbb{E}(ZY_2) \\ &= \mathbb{E}(Z(Y_1 - Y_2)) = \mathbb{E}(ZY). \end{aligned}$$

c) On établit l'unicité. On commence par considérer le cas où $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. Supposons qu'il existe une v.a. $Y' \in L^1(\Omega, \mathcal{B}, \mathbb{P})$ vérifiant :

$$\mathbb{E}(Y'Z) = \mathbb{E}(XZ), \quad (4.18)$$

pour toute v.a. Z , $\mathcal{B}$ -mesurable, bornée. On en déduit : $\mathbb{E}(UZ) = 0$, avec $U = Y - Y'$. Soient $a < 0$ et $Z = \mathbb{1}_{\{U < a\}}$. Il est clair que Z est une v.a. $\mathcal{B}$ -mesurable, bornée. Donc

$$0 = \mathbb{E}[UZ] = \mathbb{E}[U\mathbb{1}_{\{U < a\}}] \leq a\mathbb{P}(U < a) \leq 0.$$

On en déduit : $a\mathbb{P}(U < a) = 0$ implique $\mathbb{P}(U < a) = 0$.

On choisit $a = -\frac{1}{n}$, la suite d'ensembles $\{U < -\frac{1}{n}\}$ étant croissante,

$$\lim_{n \rightarrow \infty} \mathbb{P}(U < -\frac{1}{n}) = \mathbb{P}(U < 0).$$

Sachant que $\mathbb{P}(U < -\frac{1}{n}) = 0$, on a montré : $\mathbb{P}(U < 0) = 0$.

Mais $(-U)$ vérifie la même propriété : $\mathbb{E}[(-U)Z] = 0$, donc $\mathbb{P}(-U < 0) = \mathbb{P}(U > 0) = 0$. En conclusion U est presque sûrement nulle, ce qui signifie que $Y = Y'$ p.s.

Le cas où $X \geq 0$, se traite d'une manière analogue.

On part de l'égalité : $\mathbb{E}(YZ) = \mathbb{E}(Y'Z)$ et on choisit $Z = \mathbb{1}_{\{Y \leq a < b < Y'\}}$, on montre $\mathbb{P}(Y < Y') = 0$, puis $\mathbb{P}(Y > Y') = 0$; soit finalement : $Y = Y'$ p.s. $\square$

Remarque.

a) Lorsque $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$; il est équivalent de dire :

(4.16) a lieu pour toute v.a. Z , $\mathcal{B}$ -mesurable bornée,

(4.16) a lieu pour toute v.a. Z , $\mathcal{B}$ -mesurable bornée et positive.

b) lorsque $X \geq 0$ ou $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, les deux assertions suivantes sont équivalentes :

(4.16) a lieu pour toute v.a. Z , $\mathcal{B}$ -mesurable bornée et positive.

(4.16) a lieu pour toute v.a. $Z = \mathbb{1}_A$, A appartenant à $\mathcal{B}$.

En ce qui concerne la première équivalence, il suffit de décomposer Z sous la forme : $Z = Z_+ - Z_-$. Pour la seconde on utilise le résultat d'approximation suivant : si Z est une v.a. $\mathcal{B}$ -mesurable, il existe une suite (Z_n) de v.a. étagées qui converge presque sûrement vers Z (une v.a. T est dite étagée si $T = \sum_{i=1}^n t_i \mathbb{1}_{A_i}$ avec $A_i \in \mathcal{B}$).

Cas particulier : σ -algèbre engendrée par une v.a.

Comme l'a montré le cas où $\mathcal{B} = \sigma(A_1, A_2, \dots, A_n)$ il est parfois difficile de caractériser les v.a. $\mathcal{B}$ -mesurables, et donc de calculer une espérance conditionnelle sachant $\mathcal{B}$. Cette tâche est facilitée lorsque $\mathcal{B}$ est engendrée par une v.a. T . Soit T une application mesurable $T : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$. La σ -algèbre engendrée par T , notée $\sigma(T)$ est définie de la manière suivante :

$$\sigma(T) = \{A \in \mathcal{A} ; \exists C \in \mathcal{E}, A = T^{-1}(C)\}. \quad (4.19)$$

Soit X une application mesurable $X : (\Omega, \mathcal{A}) \rightarrow (F, \mathcal{F})$. On dit que X est $\sigma(T)$ -mesurable si $X^{-1}(C) \in \sigma(T)$ pour tout $C \in \mathcal{F}$. Ces applications sont caractérisées de la manière suivante :

$$X \text{ est } \sigma(T)\text{-mesurable} \iff \text{il existe } f : (E, \mathcal{E}) \rightarrow (F, \mathcal{F}) \text{ mesurable, telle que } X = f \circ T. \quad (4.20)$$

Pour les exemples pratiques : $E = \mathbb{R}^d$ et $F = \mathbb{R}$.

Revenons à la définition de $\mathbb{E}(X|\mathcal{B})$, lorsque $\mathcal{B} = \sigma(T)$. On note

$$\mathbb{E}(X|T) = \mathbb{E}(X|\sigma(T)). \quad (4.21)$$

La condition (4.15) est aisée : on cherche une fonction mesurable $f : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que

$$\mathbb{E}(X|T) = f(T) \quad (4.22)$$

$$\mathbb{E}[g(T)X] = \mathbb{E}[g(T)f(T)], \quad (4.23)$$

pour toute fonction mesurable $g : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, bornée positive. On peut se restreindre (voir remarque) à $g = \mathbb{1}_C$, $C \in \mathcal{E}$

$$\mathbb{E}[X\mathbb{1}_{\{T \in C\}}] = \mathbb{E}[f(T)\mathbb{1}_{\{T \in C\}}] ; \quad \forall C \in \mathcal{E}. \quad (4.24)$$

Examinons un exemple, celui où T prend un nombre fini de valeurs : $t_1, t_2, \dots, t_n$. On supposera que les t_i sont 2 à 2 distincts et $\mathbb{P}(T = t_i) > 0$. Alors $\mathcal{B} = \sigma(T) = \sigma(A_1, A_2, \dots, A_n)$ où l'on a posé $A_i = \{T = t_i\}$. On a vu :

$$\mathbb{E}(X|\mathcal{B}) = \sum_{i=1}^n \mathbb{E}(X|T = t_i) \mathbb{1}_{\{T=t_i\}}.$$

Posons

$$f(t) = \mathbb{E}(X|T = t), \quad t \in \{t_1, t_2, \dots, t_n\}.$$

Alors $\mathbb{E}(X|\mathcal{B}) = f(T)$.

On remarque sur cet exemple que f n'est pas unique, f est uniquement déterminée sur $\{t_1, t_2, \dots, t_n\}$.

Rappelons que

$$\mathbb{E}(X|T = t) = \frac{1}{\mathbb{P}(T = t)} \mathbb{E}(X\mathbb{1}_{\{T=t\}}).$$

Cette quantité étant bien définie si $\mathbb{P}(T = t) > 0$, condition satisfaite si $t \in \{t_1, t_2, \dots, t_n\}$. Lorsque T est une v.a. quelconque, on a : $\mathbb{E}(X|T) = f(T)$. On peut noter $f(t) = \mathbb{E}(X|T = t)$, mais il s'agit d'une convention d'écriture, et non d'un conditionnement par l'événement $\{T = t\}$, qui peut être de probabilité nulle. Il est conseillé de se tenir à la notation (4.22).

Exemples :

Exemple 3) Soient X et Y deux v.a., X à valeurs réelles, X intégrable ou positive. On suppose que X et Y sont deux v.a. indépendantes. Alors $\mathbb{E}(X|Y) = \mathbb{E}(X)$.

Preuve : Il s'agit de montrer :

$$\mathbb{E}(Xg(Y)) = \mathbb{E}(X)\mathbb{E}[g(Y)],$$

pour toute fonction g borélienne, positive et bornée. Mais X et Y sont deux v.a. indépendantes, cette égalité est réalisée. $\square$

Exemple 4) Soit (X, Y) un vecteur gaussien centré à valeurs dans $\mathbb{R}^2$. On suppose que Y est non dégénérée, i.e. Y est presque sûrement non nulle. Alors $\mathbb{E}(X|Y) = aY$.

Preuve : On procède comme dans le cadre des espaces euclidiens, en remplaçant l'orthogonalité par l'indépendance. On cherche un réel a tel que $X' = X - aY$ soit une v.a. indépendante de Y . Puisque (X, Y) est un vecteur gaussien, $(X', Y) = (X - aY, Y)$ l'est aussi. Par conséquent les v.a. X' et Y sont indépendantes si et seulement si :

$$\text{Cov}(X', Y) = \mathbb{E}(X'Y) = 0.$$

Mais

$$\mathbb{E}(X'Y) = \mathbb{E}(XY) - a\mathbb{E}(Y^2) = \mathbb{E}(XY) - a \text{Var}Y.$$

Sachant que $\text{Var}Y > 0$, a est unique et

$$a = \frac{\mathbb{E}(XY)}{\text{Var}Y},$$

a étant ainsi fixé, d'après la linéarité de l'espérance conditionnelle et le résultat de l'exemple 3, on a,

$$\mathbb{E}(X'|Y) = \mathbb{E}(X|Y) - a\mathbb{E}(Y|Y) = \mathbb{E}(X|Y) - aY = 0.$$

D'où

$$\mathbb{E}(X|Y) = aY = \frac{\mathbb{E}(XY)}{\text{Var}Y} Y. \quad (4.25)$$

□

L'application $X \rightarrow \mathbb{E}(X|Y)$ est une application de projection orthogonale sur le sous-espace vectoriel de dimension 1 engendrée par Y . Cet exercice sera repris dans un cadre plus général.

Proposition 3 Soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$.

1) Soient X et Y deux variables aléatoires intégrables ou positives, alors

$$\mathbb{E}(X + Y|\mathcal{B}) = \mathbb{E}(X|\mathcal{B}) + \mathbb{E}(Y|\mathcal{B}).$$

2) Pour tout $a \in \mathbb{R}$ si X est intégrable et pour $a \geq 0$ si X est positif, $\mathbb{E}(aX|\mathcal{B}) = a\mathbb{E}(X|\mathcal{B})$.

3) Si X est $\mathcal{B}$ -mesurable, alors $\mathbb{E}(XY|\mathcal{B}) = X\mathbb{E}(Y|\mathcal{B})$ lorsque $X \geq 0$ et $Y \geq 0$ ou Y et XY intégrables.

4) Si $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ alors $\mathbb{E}(X|\mathcal{B}) \in L^2(\Omega, \mathcal{B}, \mathbb{P})$.

5) Si $\mathcal{B}_1$ et $\mathcal{B}_2$ sont deux sous-tribus de $\mathcal{A}$, $\mathcal{B}_1 \subset \mathcal{B}_2$ alors

$$\mathbb{E}(\mathbb{E}(X|\mathcal{B}_2)|\mathcal{B}_1) = \mathbb{E}(X|\mathcal{B}_1), \quad (4.26)$$

pour $X \geq 0$ ou intégrable.

6) Si $X \leq Y$ p.s. alors $\mathbb{E}(X|\mathcal{B}) \leq \mathbb{E}(Y|\mathcal{B})$, lorsque $X \geq 0$ ou X et Y intégrables.

7) $|\mathbb{E}(X|\mathcal{B})| \leq \mathbb{E}(|X||\mathcal{B})$, X positive ou intégrable.

Preuve : a) Montrons 1) lorsque X et Y sont intégrables. Soit $U = \mathbb{E}(X|\mathcal{B}) + \mathbb{E}(Y|\mathcal{B})$. U est une v.a. intégrable et $\mathcal{B}$ -mesurable. Il s'agit de montrer,

$$\mathbb{E}[UZ] = \mathbb{E}[(X + Y)Z],$$

pour toute v.a. Z , $\mathcal{B}$ -mesurable, bornée. Mais

$$\mathbb{E}((X + Y)Z) = \mathbb{E}(XZ) + \mathbb{E}(YZ) = \mathbb{E}[\mathbb{E}(X|\mathcal{B})Z] + \mathbb{E}[\mathbb{E}(Y|\mathcal{B})Z] = \mathbb{E}[UZ].$$

Le 2) se montre d'une manière analogue.

b) Montrons 3), lorsque cette fois $X \geq 0$ et $Y \geq 0$. On pose $V = X\mathbb{E}(Y|\mathcal{B})$. V est une v.a. positive, $\mathcal{B}$ -mesurable. On considère une v.a. Z , $\mathcal{B}$ -mesurable, positive et bornée. XZ étant $\mathcal{B}$ -mesurable, on a :

$$\mathbb{E}[Z(XY)] = \mathbb{E}[(ZX)Y] = \mathbb{E}[ZX\mathbb{E}(Y|\mathcal{B})] = \mathbb{E}[ZV].$$

On en déduit 3).

c) La propriété 4) est une conséquence immédiate de la proposition 1.

d) La preuve de 5) est aisée, plaçons-nous dans le cas positif. Comme dans b), Z désigne une v.a. test positive, $\mathcal{B}_1$ -mesurable. Posons $W = \mathbb{E}(X|\mathcal{B}_1)$. Par définition W est $\mathcal{B}_1$ -mesurable. De plus, Z étant $\mathcal{B}_1$ -mesurable est $\mathcal{B}_2$ -mesurable, donc,

$$\mathbb{E}[Z\mathbb{E}(X|\mathcal{B}_2)] = \mathbb{E}(ZX) = \mathbb{E}(Z\mathbb{E}(Y|\mathcal{B}_1)) = \mathbb{E}(ZW).$$

e) On peut écrire $Y = X + T$ avec T v.a. positive. D'après la propriété 1), $\mathbb{E}(Y|\mathcal{B}) = \mathbb{E}(X|\mathcal{B}) + \mathbb{E}(T|\mathcal{B})$, mais $T \geq 0$ implique $\mathbb{E}(T|\mathcal{B}) \geq 0$.

f) Si $X \geq 0$, $\mathbb{E}(X|\mathcal{B}) \geq 0$ et $\mathbb{E}(|X||\mathcal{B}) = \mathbb{E}(X|\mathcal{B})$, l'inégalité est une égalité. Supposons X élément de $L^1(\Omega, \mathcal{A}, \mathbb{P})$. En utilisant successivement l'inégalité : $-|X| \leq X \leq |X|$ et 6) on a : $-\mathbb{E}(|X||\mathcal{B}) \leq \mathbb{E}(X|\mathcal{B}) \leq \mathbb{E}(|X||\mathcal{B})$. $\square$

Proposition 4 Soient $\mathcal{B}_1$ et $\mathcal{B}_2$ deux sous-tribus de $\mathcal{A}$. Alors $\mathcal{B}_1$ et $\mathcal{B}_2$ sont indépendantes si et seulement si pour toute v.a. X , $\mathcal{B}_2$ -mesurable et bornée, $\mathbb{E}(X|\mathcal{B}_1) = \mathbb{E}(X)$.

Preuve : On rappelle que deux tribus $\mathcal{B}_1$ et $\mathcal{B}_2$ sont indépendantes si et seulement si deux événements quelconques B_1 de $\mathcal{B}_1$ et B_2 de $\mathcal{B}_2$ sont indépendants, i.e. :

$$\mathbb{P}(B_1 \cap B_2) = \mathbb{P}(B_1)\mathbb{P}(B_2).$$

1) Supposons les deux tribus $\mathcal{B}_1$ et $\mathcal{B}_2$ indépendantes. Soit Z une v.a. bornée, $\mathcal{B}_1$ -mesurable. On a : $\mathbb{E}(XZ) = \mathbb{E}(X)\mathbb{E}(Z)$, d'où $\mathbb{E}(X|\mathcal{B}_1) = \mathbb{E}(X)$.

2) Réciproquement, on choisit $Z = \mathbb{1}_{B_1}$, $X = \mathbb{1}_{B_2}$ avec $B_1 \in \mathcal{B}_1$ et $B_2 \in \mathcal{B}_2$. Alors

$$\mathbb{E}(XZ) = \mathbb{P}(B_1 \cap B_2) = \mathbb{E}(X)\mathbb{E}(Z) = \mathbb{P}(B_2)\mathbb{P}(B_1).$$

Les deux tribus $\mathcal{B}_1$ et $\mathcal{B}_2$ sont indépendantes. $\square$

Remarques.

a) Si $\mathcal{B}_2 = \sigma(Y)$ et $\mathcal{B}_1 = \sigma(X)$, l'exemple apparaît comme un cas particulier de la proposition 4.

b) Soit Y une v.a. fixée. On sait que les deux v.a. X et Y sont indépendantes si et seulement si les tribus $\sigma(X)$ et $\sigma(Y)$ sont indépendantes. La proposition 4 affirme que X et Y sont indépendantes si et seulement si :

$$\mathbb{E}[f(X)|Y] = \mathbb{E}[f(X)], \quad (4.27)$$

pour toute fonction f borélienne, bornée et positive. En particulier $\mathbb{E}(X|Y) = \mathbb{E}(X)$. Mais cette relation n'implique pas, en général (4.27), sauf dans le cas gaussien !

4.2 Loi conditionnelle

Afin d'introduire cette nouvelle notion nous commençons par deux exemples :

Exemples :

Exemple 5) Ce premier exemple est relatif aux lois discrètes, il ne nécessite aucune théorie, tous les calculs peuvent être faits "à la main". Soient X et Y deux v.a. indépendantes, de loi Poisson de paramètre respectivement λ et μ . On pose $S = X + Y$. On veut calculer la loi de X sachant S . Il s'agit d'évaluer :

$$\mathbb{P}(X = n|S = m) = \frac{\mathbb{P}(X = n, X + Y = m)}{\mathbb{P}(X + Y = m)} = \frac{\mathbb{P}(X = n, Y = m - n)}{\mathbb{P}(X + Y = m)},$$

où $m \geq n \geq 0$. Mais $X + Y$ suit une loi de Poisson de paramètre $\lambda + \mu$, de plus X et Y sont indépendantes, d'où :

$$\begin{aligned} \mathbb{P}(X = n|S = m) &= \frac{\mathbb{P}(X = n)\mathbb{P}(Y = m - n)}{\mathbb{P}(X + Y = m)} = \frac{\lambda^n}{n!} e^{-\lambda} \frac{\mu^{m-n}}{(m-n)!} e^{-\mu} \frac{m!}{(\lambda + \mu)^m e^{-(\lambda + \mu)}} \\ &= C_m^n \rho^n (1 - \rho)^{m-n}, \end{aligned}$$

avec $\rho = \frac{\lambda}{\lambda + \mu}$. Par conséquent, conditionnellement à $\{S = m\}$, X suit une loi binomiale $\mathcal{B}(m, \rho)$.

Exemple 6) Le second exemple a trait aux lois continues. On se donne deux v.a. indépendantes X et Y , de loi exponentielle de paramètre 1. On pose $S = X + Y$. On va calculer $H(S) = \mathbb{E}(h(X)|S)$ pour toute fonction h borélienne bornée. Soient g une fonction borélienne bornée et $\Delta = \mathbb{E}(h(X)g(S))$. La fonction H est caractérisée par : $\Delta = \mathbb{E}[H(S)g(S)]$. Calculons Δ . On a :

$$\Delta = \mathbb{E}[h(X)g(X + Y)] = \iint_{x>0, y>0} h(x)g(x + y)e^{-(x+y)} dx dy.$$

x étant fixé, on fait le changement de variable $u = x + y$, il vient :

$$\Delta = \iint_{x>0, u \geq x} h(x)g(u)e^{-u} dx du = \int_0^\infty g(u)e^{-u} \left(\int_0^u h(x) dx \right) du.$$

On choisit $h = 1$. Alors $\Delta = \int_0^\infty g(u)ue^{-u}du$. Par conséquent la v.a. S admet pour densité $\mathbb{1}_{\{u>0\}}ue^{-u}$. On pose : $H(u) = \frac{1}{u} \int_0^u h(x)dx$, alors

$$\Delta = \int_0^\infty g(u)H(u)ue^{-u}du = \mathbb{E}[H(S)g(S)].$$

Mais $U_u(dx) = \frac{1}{u} \mathbb{1}_{[0,u]}(x)dx$ est la loi uniforme sur $[0, u]$. Nous dirons que conditionnellement à $S = u$, X suit la loi uniforme sur $[0, u]$,

$$\mathbb{E}[h(X)g(S)] = \int_0^\infty g(u) \left(\int_0^u h(x)U_u(dx) \right) \mu(du), \quad \forall f \geq 0, \quad \forall g \geq 0, \quad (4.28)$$

où $\mu(du) = \mathbb{1}_{\{u>0\}}ue^{-u}du$ est la loi de S .

Définition 1 Soient $(E, \mathcal{E})$ et $(F, \mathcal{F})$ deux espaces mesurables. On appelle noyau (positif) une application N définie sur $E \times \mathcal{F}$ à valeurs dans $\mathbb{R}_+$ telle que :

- (i) pour tout $x \in E$, $A \in \mathcal{F} \rightarrow N(x, A)$ est une mesure positive sur $(F, \mathcal{F})$,
- (ii) pour tout $A \in \mathcal{F}$, $x \rightarrow N(x, A)$ est mesurable de $(E, \mathcal{E})$ sur $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$.

On dit que le noyau N est une probabilité de transition de E vers F si pour tout x , $N(x, \cdot)$ est une probabilité. Lorsque $E = F$, N est appelé probabilité de transition sur E .

Remarques.

- 1) Une probabilité de transition est une famille "mesurable" de probabilités sur $(F, \mathcal{F})$ indexée par $E : (N(x, \cdot) ; x \in E)$.
- 2) Si g est une application mesurable définie sur F , positive alors

$$x \in E \rightarrow N(x, f) = \int f(y)N(x, dy) \text{ est } \mathcal{E}\text{-mesurable.} \quad (4.29)$$

- 3) Pour l'exemple 5, on a $E = F = \mathbb{N}$, $N(k, \cdot)$ est la loi binomiale $\mathcal{B}(k, \rho)$:

$$N(k, \cdot) = \sum_{i=0}^k C_k^i \rho^i (1 - \rho)^{k-i} \delta_i,$$

δ_i désignant la mesure de Dirac en i .

Quant à l'exemple 6, $E = F = \mathbb{R}_+$ et

$$N(u, dx) = \frac{1}{u} \mathbb{1}_{[0,u]} dx.$$

Exemple 7) :

Soient $f : (E \times F, \mathcal{E} \times \mathcal{F}) \rightarrow \mathbb{R}_+$ mesurable, μ une probabilité sur $(F, \mathcal{F})$. On pose

$$N(x, A) = \int_A f(x, y) \mu(dy), \quad x \in E, \quad A \in \mathcal{F}.$$

N est un noyau. Si de plus $\int_F f(x, y) \mu(dy) = 1$, pour tout $x \in E$, N est une probabilité de transition.

Définition 2 Soient X et Y deux v.a. à valeurs dans $(E, \mathcal{E})$, respectivement $(F, \mathcal{F})$. On appelle loi conditionnelle de Y sachant X , une probabilité de transition N , de E vers F telle que :

$$\mathbb{E}(g(Y)|X) = \int g(y)N(X, dy) = N(X, g), \quad (4.30)$$

pour toute fonction borélienne positive g .

Remarques :

1) Si ν est une mesure positive sur $(E, \mathcal{E})$, on note indifféremment :

$$\int_E f d\nu = \int_E f(x)\nu(dx) = \int_E f(x)d\nu(x) = \nu(f). \quad (4.31)$$

2) On a vu que $x \rightarrow \int g(y)N(x, dy)$ est mesurable, donc $\int g(y)N(X, dy)$ est une v.a. $\sigma(X)$ -mesurable. Ainsi pour démontrer que N est la loi conditionnelle de Y sachant X , il est nécessaire et suffisant de montrer :

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}\left[f(X) \int_F g(y)N(x, dy)\right] = \int_E f(x) \left(\int_F g(y)N(x, dy)\right) \mu(dx)$$

où μ désigne la loi de X . C'est exactement la démarche que nous avons adoptée dans l'exemple 6.

3) La loi conditionnelle de Y sachant X n'est pas unique. Soient N_1 et N_2 deux lois conditionnelles de Y sachant X . On a seulement :

$$N_1(x, A) = N_2(x, A) \quad \forall A \in \mathcal{F},$$

pour μ -presque tout x , μ désignant la loi de X .

Dans le cas discret de l'exemple 5, N_1 et N_2 sont déterminés uniquement sur $\mathbb{N}$.

4) Formellement $N(x, dy)$ est la loi de Y sachant $X = x$. Si X suit une loi discrète, et x appartient au support de la loi de X , alors,

$$N(x, A) = \mathbb{P}(Y \in A | X = x) = \frac{\mathbb{P}(Y \in A, X = x)}{\mathbb{P}(X = x)}.$$

5) On montre que si E et F sont deux espaces métriques séparables et complets, munis de leur tribu borélienne, il existe alors une loi conditionnelle de Y sachant X . Rentrent dans ce cadre les espaces $\mathbb{R}^n$.

Soient X et Y deux v.a. à valeurs dans E et F respectivement. Si on connaît la loi μ de X et si N décrit la loi conditionnelle de Y sachant X , alors la loi du couple (X, Y) est déterminée. En effet

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)\mathbb{E}[g(Y)|X]] = \mathbb{E}\left[f(X) \int_F g(y)N(X, dy)\right] \quad (4.32)$$

pour toute $f : E \rightarrow \mathbb{R}_+$, $g : F \rightarrow \mathbb{R}_+$, mesurables, bornées.

On va s'intéresser au problème réciproque : calculer la loi conditionnelle de Y sachant X à l'aide de la loi de (X, Y) . Nous allons nous restreindre à $E = \mathbb{R}^n$ et $F = \mathbb{R}^m$, $n \geq 1$ et $m \geq 1$. En pratique, dans de nombreux cas, on rencontre des v.a. à valeurs dans $\mathbb{R}^n$, et dont la loi soit admet une densité (par rapport à la mesure de Lebesgue), soit est discrète. Lorsque X est discrète on supposera, pour simplifier que X est à valeurs dans $\mathbb{Z}^n$. Soit λ_d^n la mesure de comptage des points de $\mathbb{Z}^n$:

$$\lambda_d^n = \sum_{k \in \mathbb{Z}^n} \delta_k. \quad (4.33)$$

λ_d^n est une mesure positive de masse infinie ($\lambda_d^n(\mathbb{Z}^n) = +\infty$) mais σ -finie :

$$\lambda_d^n([-a, a] \times \dots \times [-a, a]) < \infty,$$

pour tout $a \in \mathbb{Z}$. Si X est à valeurs dans $\mathbb{Z}^n$, la loi de X admet une densité f par rapport à λ_d^n et

$$f(m) = \mathbb{P}(X = m), \quad m \in \mathbb{Z}^n.$$

On note λ_l^n la mesure de Lebesgue sur $\mathbb{R}^n$.

Proposition 5 Soient X et Y deux v.a. à valeurs dans $\mathbb{R}^n$, respectivement $\mathbb{R}^m$. On suppose que la loi de (X, Y) admet une densité $\varphi(x, y)$ par rapport à la mesure $\nu_1 \otimes \nu_2$, où ν_1 (resp. ν_2) vaut λ_d^n ou λ_l^n . (resp. λ_d^m ou λ_l^m). On note :

$$\alpha(x) = \int_{\mathbb{R}^m} \varphi(x, y) \nu_2(dy) ; \quad x \in \mathbb{R}^n, \quad (4.34)$$

$$N(x, dy) = \frac{\varphi(x, y)}{\alpha(x)} \mathbf{1}_{\{\alpha(x) > 0\}} \nu_2(dy) ; \quad x \in \mathbb{R}^n. \quad (4.35)$$

Alors,

- 1) La loi de X admet α comme densité par rapport à ν_1 .
- 2) La loi conditionnelle de Y sachant X est donnée par la probabilité de transition N .

Preuve : a) Par définition, si $h : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ est mesurable et bornée, on a,

$$\mathbb{E}(h(X, Y)) = \int_{\mathbb{R}^n \times \mathbb{R}^m} h(x, y) \varphi(x, y) \nu_1(dx) \nu_2(dy). \quad (4.36)$$

En particulier si $h(x, y) = \mathbf{1}_{\{\alpha(x)=0\}}$,

$$\mathbb{P}(\alpha(X) = 0) = \int_{\mathbb{R}^n \times \mathbb{R}^m} \mathbf{1}_{\{\alpha(x)=0\}} \varphi(x, y) \nu_1(dx) \nu_2(dy).$$

On applique le théorème de Fubini :

$$\begin{aligned}\mathbb{P}(\alpha(X) = 0) &= \int_{\mathbb{R}^n} \mathbb{I}_{\{\alpha(x)=0\}} \left(\int_{\mathbb{R}^m} \varphi(x, y) \nu_2(dy) \right) \nu_1(dx) \\ &= \int_{\mathbb{R}^n} \mathbb{I}_{\{\alpha(x)=0\}} \alpha(x) \nu_1(dx) = 0\end{aligned}$$

Donc $\mathbb{P}(\alpha(X) = 0) = 0$. Ce qui signifie que presque sûrement, $\alpha(X) \neq 0$. Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^m \rightarrow \mathbb{R}$, deux fonctions boréliennes et bornées. On déduit de (4.36) :

$$\begin{aligned}\mathbb{E}[f(X)g(Y)] &= \mathbb{E}[f(X)g(Y)\mathbb{I}_{\{\alpha(X)>0\}}] \\ &= \int_{\mathbb{R}^n \times \mathbb{R}^m} f(x)g(y)\varphi(x, y)\mathbb{I}_{\{\alpha(x)>0\}}\nu_1(dx)\nu_2(dy).\end{aligned}\tag{4.37}$$

En particulier si $g = 1$,

$$\begin{aligned}\mathbb{E}[f(X)] &= \int_{\mathbb{R}^n} f(x) \left(\int_{\mathbb{R}^m} \varphi(x, y) \nu_2(dy) \right) \mathbb{I}_{\{\alpha(x)>0\}} \nu_1(dx) \\ &= \int_{\mathbb{R}^n} f(x) \alpha(x) \mathbb{I}_{\{\alpha(x)>0\}} \nu_1(dx) = \int_{\mathbb{R}^n} f(x) \alpha(x) \nu_1(dx).\end{aligned}$$

Ce qui signifie que la loi de X est $\alpha(x)\nu_1(dx)$. On revient à (4.37) et on applique à nouveau le théorème de Fubini :

$$\begin{aligned}\mathbb{E}[f(X)g(Y)] &= \int_{\mathbb{R}^n} f(x) \mathbb{I}_{\{\alpha(x)>0\}} \alpha(x) \left(\int_{\mathbb{R}^m} g(y) \frac{\varphi(x, y)}{\alpha(x)} \nu_2(dy) \right) \nu_1(dx) \\ &= \int_{\mathbb{R}^n} f(x) \mathbb{I}_{\{\alpha(x)>0\}} \alpha(x) \left(\int_{\mathbb{R}^m} g(y) N(x, dy) \right) \nu_1(dx).\end{aligned}$$

On a montré :

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E} \left(f(X) \int_{\mathbb{R}^m} g(y) N(X, dy) \right).\tag{4.38}$$

b) D'après l'exemple 7, N est bien un noyau. Il reste à montrer que N est une probabilité de transition :

$$\begin{aligned}N(x, \mathbb{R}^m) &= \int_{\mathbb{R}^m} \frac{\varphi(x, y)}{\alpha(x)} \mathbb{I}_{\{\alpha(x)>0\}} \nu_2(dy) = \frac{1}{\alpha(x)} \mathbb{I}_{\{\alpha(x)>0\}} \int_{\mathbb{R}^m} \varphi(x, y) \nu_2(dy) \\ &= \frac{1}{\alpha(x)} \mathbb{I}_{\{\alpha(x)>0\}} \alpha(x) = \mathbb{I}_{\{\alpha(x)>0\}}.\end{aligned}$$

On remarque qu'il faut modifier légèrement la définition de N , en définissant N par la formule (4.35) lorsque $\alpha(x) > 0$ et par une probabilité quelconque sur $\mathbb{R}^m$, quand $\alpha(x) = 0$, par exemple $N(x, dy) = \gamma(y)\nu_2(dy)$ convient, où $\gamma : \mathbb{R}^m \rightarrow \mathbb{R}_+$ est mesurable et $\int_{\mathbb{R}^m} \gamma(y)\nu_2(dy) = 1$. Alors

$$N(x, dy) = \left\{ \frac{\varphi(x, y)}{\alpha(x)} \mathbb{I}_{\{\alpha(x)>0\}} + \gamma(y) \mathbb{I}_{\{\alpha(x)=0\}} \right\} \nu_2(dy).\tag{4.39}$$

□

Nous allons à présent décrire aussi explicitement que possible le noyau N lorsque X et Y sont

discrètes et/ou à densité. Lorsque l'une des v.a. est discrète on supposera pour simplifier qu'elle est à valeurs dans $\mathbb{N}$.

a) X est discrète

(i) Y est discrète

On note

$$p_{i,j} = \mathbb{P}(X = i, Y = j) ; i \in \mathbb{N}, j \in \mathbb{N}.$$

Alors,

$$\nu_1 = \nu_2 = \lambda_d^1 = \sum_{k \in \mathbb{Z}} \delta_k \text{ et } \varphi(i, j) = p_{i,j} \mathbb{1}_{\{i \geq 0, j \geq 0\}},$$

$$\alpha(i) = \int_{\mathbb{N}} \varphi(i, j) \nu_2(dj) = \sum_{j \geq 0} \varphi(i, j) = \sum_{j \geq 0} p_{i,j} = p_{i, \cdot}.$$

$$N(i, \{j\}) = \frac{p_{i,j}}{p_{i, \cdot}}.$$

On retrouve ainsi le résultat classique :

$$\mathbb{P}(Y = j | X = i) = \frac{\mathbb{P}(Y = j, X = i)}{\mathbb{P}(X = i)} = \frac{p_{i,j}}{p_{i, \cdot}}.$$

(ii) Y admet une densité

On suppose que Y est à valeurs dans $\mathbb{R}^m$. On a :

$$\nu_1 = \lambda_d^1 = \sum_{k \in \mathbb{Z}} \delta_k \text{ et } \nu_2(dy) = \lambda_l^m(dy) = dy_1 dy_2 \dots dy_m.$$

$$\alpha(i) = \int_{\mathbb{R}^m} \varphi(i, y) dy = \mathbb{P}(X = i) ; i \in \mathbb{N},$$

$$N(i, dy) = \frac{\varphi(i, y)}{\alpha(i)} dy = \mathbb{P}(Y \in dy | X = i) ; i \in \mathbb{N}.$$

En d'autres termes la probabilité $N(i, \cdot)$ admet $\frac{\varphi(i, y)}{\alpha(i)}$ comme densité (par rapport à la mesure de Lebesgue sur $\mathbb{R}^m$). De plus

$$N(i, f) = \int_{\mathbb{R}^m} f(y) N(i, dy) = \mathbb{E}[f(Y) | X = i], f \geq 0.$$

Il s'agit d'un conditionnement usuel par l'événement $\{X = i\}$, lorsque celui-ci est de probabilité non nulle.

b) X admet une densité

On suppose que X est à valeurs dans $\mathbb{R}^n$.

(i) Y est discrète

$$\nu_1(dx) = \lambda_l^n(dx) = dx_1 dx_2 \dots dx_n \text{ et } \nu_2 = \lambda_d^1 = \sum_{k \in \mathbb{Z}} \delta_k,$$

$$\alpha(x) = \int_{\mathbb{N}} \varphi(x, j) \nu_2(dj) = \sum_{k \in \mathbb{Z}} \varphi(x, k) ; x \in \mathbb{R}^n,$$

$$N(x, \{j\}) = \frac{\varphi(x, j)}{\alpha(x)}, \quad j \in \mathbb{N}, x \in \mathbb{R}^n.$$

(ii) Y admet une densité

$$\nu_1(dx) = \lambda_l^n(dx) = dx_1 dx_2 \dots dx_n \text{ et } \nu_2(dy) \lambda_l^m(dy) = dy_1 dy_2 \dots dy_m,$$

$$\alpha(x) = \int_{\mathbb{R}^m} \varphi(x, y) dy ; x \in \mathbb{R}^n,$$

$$N(x, dy) = \frac{\varphi(x, y)}{\alpha(x)} dy \quad x \in \mathbb{R}^n. \quad (4.40)$$

$N(x, \cdot)$ admet $\frac{\varphi(x, y)}{\alpha(x)}$ comme densité.

4.3 Le cas gaussien

Soit Z un vecteur gaussien à valeurs dans $\mathbb{R}^{n+m}$. On note X les n premières coordonnées de Z , et Y les m suivantes : $Z = (X, Y)$. On suppose que X admet une densité. Le but de ce paragraphe est de déterminer la loi de Y sachant X .

Rappelons que si $n = m = 1$, nous avons calculé $\mathbb{E}(Y|X)$ (voir exemple 4). Nous allons adapter cette approche au cas multidimensionnel.

Nous commençons par fixer quelques notations.

— K désigne la matrice de covariance de (X, Y) , on écrira :

$$K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix},$$

- K_{11} est la matrice de covariance de X , carrée et d'ordre n ,
- $K_{12} = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))^*)$ est une matrice $n \times m$ (n lignes, m colonnes),
- $K_{21} = K_{12}^*$ est d'ordre $m \times n$,
- K_{22} est la matrice de covariance de Y , carrée et d'ordre m .

On a supposé que X admet une densité, condition équivalente à :

$$K_{11} \text{ est inversible.} \quad (4.41)$$

Proposition 6 1) Il existe une matrice A , d'ordre $m \times n$, et un vecteur gaussien X' , indépendant de X tel que

$$Y = AX + X' \quad (4.42)$$

2) De plus $A = K_{21}K_{11}^{-1}$ et le vecteur gaussien X' est caractérisé par,

$$\mathbb{E}(X') = \mathbb{E}(Y) - A\mathbb{E}(X), \quad K_{X'} = K_{22} - K_{21}K_{11}^{-1}K_{21}^*, \quad (4.43)$$

où $K_{X'}$ désigne la matrice de covariance de X' .

Remarque.

L'affirmation essentielle de la proposition 6 est le 1) et en particulier (4.42), on en déduit alors facilement le 2). La méthode développée est une extension de celle que nous avons utilisée en dimension 1.

Preuve : (Proposition 6). **a)** La première étape consiste à montrer que si Z est un vecteur gaussien, $Z_1 = B_1 Z$, $Z_2 = B_2 Z$ où B_1 et B_2 sont deux matrices, alors les deux v.a. Z_1 et Z_2 sont indépendantes si et seulement si :

$$K_{Z_1, Z_2} = \mathbb{E}[(Z_1 - \mathbb{E}(Z_1))(Z_2 - \mathbb{E}(Z_2))^*] = 0. \quad (4.44)$$

On adopte les conventions du premier chapitre : les v.a. à valeurs multidimensionnelles sont représentées par une matrice unicolonne. Si Z_i est à valeurs dans $\mathbb{R}^{n_i}$, les deux v.a. Z_1 et Z_2 sont indépendantes si et seulement si :

$$\mathbb{E}[\exp\{i(u_1^* Z_1 + u_2^* Z_2)\}] = \mathbb{E}[\exp(iu_1^* Z_1)]\mathbb{E}[\exp(iu_2^* Z_2)], \quad (4.45)$$

pour tout $u_i \in \mathbb{R}^{n_i}$. Mais (4.45) est équivalente à :

$$u_1^* Z_1 \text{ et } u_2^* Z_2 \text{ sont indépendantes.} \quad (4.46)$$

Il est clair que (4.46) entraîne (4.45), en changeant dans (4.45), u_i en $\lambda_i u_i$, on montre que (4.45) entraîne (4.46). Mais

$$\tilde{Z} = \begin{pmatrix} u_1^* Z_1 \\ u_2^* Z_2 \end{pmatrix} = BZ.$$

Donc $\tilde{Z}$ est un vecteur gaussien, à valeurs dans $\mathbb{R}^2$; on sait que les deux composantes de $\tilde{Z}$ sont indépendantes si et seulement si,

$$\rho = \text{Cov}(u_1^* Z_1, u_2^* Z_2) = 0 ; \quad \forall u_1, \quad \forall u_2. \quad (4.47)$$

Mais

$$u_1^* Z_1 - \mathbb{E}(u_1^* Z_1) = u_1^* Z_1 - u_1^* \mathbb{E}(Z_1) = u_1^* (Z_1 - \mathbb{E}(Z_1)),$$

$$Z_2^* u_2 - \mathbb{E}(Z_2^* u_2) = Z_2^* u_2 - \mathbb{E}(Z_2^*) u_2 = (Z_2^* - \mathbb{E}(Z_2^*)) u_2.$$

D'où

$$\rho \mathbb{E}[u_1^* (Z_1 - \mathbb{E}(Z_1)) (Z_2^* - \mathbb{E}(Z_2^*)) u_2] = u_1^* \mathbb{E}[(Z_1 - \mathbb{E}(Z_1)) (Z_2^* - \mathbb{E}(Z_2^*))] u_2.$$

Par conséquent,

$$\rho = u_1^* K_{Z_1, Z_2} u_2.$$

Il est à présent évident que : (4.47) est équivalent à (4.44).

b) Montrons (4.42). On cherche une matrice A telle que $X' = Y - AX$ soit indépendant de Y . Mais $X' = B_1 Z$ et $X = B_2 Z$ (rappelons que Z a pour coordonnées X et Y). On applique le résultat du a) : X' et X sont deux v.a. indépendantes si et seulement si,

$$K_{X', X} = \mathbb{E}[(Y - AX - \mathbb{E}(Y - AX))(X - \mathbb{E}(X))^*] = 0. \quad (4.48)$$

En utilisant la linéarité, il vient,

$$K_{X', X} = \mathbb{E}[(Y - \mathbb{E}(Y))(X - \mathbb{E}(X))^*] - \mathbb{E}[A(X - \mathbb{E}(X))(X - \mathbb{E}(X))^*],$$

$$K_{X', X} = K_{21} - A \mathbb{E}[(X - \mathbb{E}(X))(X - \mathbb{E}(X))^*] = K_{21} - A K_{11}.$$

Il est à présent aisé de montrer que (4.48) est équivalent à $K_{21} - A K_{11} = 0$. Mais K_{11} est inversible, donc A est unique et,

$$A = K_{21} K_{11}^*.$$

c) Puisque $X' = Y - AX$, par linéarité, on a $\mathbb{E}(X') = \mathbb{E}(Y) - A \mathbb{E}(X)$ et

$$K_{X'} = \mathbb{E}[(Y_0 - AX_0)(Y_0 - AX_0)^*] = \mathbb{E}[(Y_0 - AX_0)(Y_0^* - X_0^* A^*)],$$

où $Y_0 = Y - \mathbb{E}(Y)$ et $X_0 = X - \mathbb{E}(X)$. On développe, il vient,

$$K_{X'} = \mathbb{E}[Y_0 Y_0^*] - A \mathbb{E}[X_0 Y_0^*] - \mathbb{E}[Y_0 X_0^*] A^* + A \mathbb{E}[X_0 X_0^*] A^*,$$

$$K_{X'} = K_{22} - A K_{12} - K_{21} A^* + A K_{11} A^*.$$

Mais

$$K_{21} = A K_{11} \implies -K_{21} A^* + A K_{11} A^* = -K_{21} A^* + K_{21} A^* = 0,$$

d'où

$$K_{X'} = K_{22} - A K_{12} = K_{22} - K_{21} K_{11}^{-1} K_{12} = K_{22} - K_{21} K_{11}^{-1} K_{21}^*.$$

□

Proposition 7 Soit N la loi conditionnelle de Y sachant X . Alors $N(x, \cdot)$ est la loi gaussienne $\mathcal{N}_n(Ax + m', K)$ avec $m' = \mathbb{E}(X') = \mathbb{E}(Y) - A \mathbb{E}(X)$, $K = K_{X'} = K_{22} - K_{21} K_{11}^{-1} K_{21}^*$.

Preuve : a) Si U et V sont deux v.a. indépendantes, et f une fonction positive et mesurable, alors $\mathbb{E}[f(U, V)] = \mathbb{E}[F(U)]$, avec $F(u) = \mathbb{E}[f(u, V)]$.

En effet, si ν (resp. μ) désigne la loi de U (resp. V), $\nu \otimes \mu$ est la loi de (U, V) :

$$\mathbb{E}[f(U, V)] = \iint f(u, v) \nu(du) \mu(dv).$$

On applique le théorème de Fubini,

$$\mathbb{E}[f(U, V)] = \int \left(\int f(u, v) \mu(dv) \right) \nu(du) = \int F(u) \nu(du).$$

b) Soient f et g deux fonctions boréliennes, positives et

$$\Delta = \mathbb{E}[f(X)g(Y)].$$

On applique la proposition 6, puis a) :

$$\Delta = \mathbb{E}[f(X)g(AX + X')] = \mathbb{E}[F(X)]$$

avec $F(x) = \mathbb{E}[f(x)g(Ax + X')] = f(x)\mathbb{E}[g(Ax + X')]$. Mais $Ax + X'$ suit une loi $\mathcal{N}_n(Ax + m', K) = N(x, \cdot)$, donc,

$$F(x) = f(x) \int g(y) N(x, dy),$$

$$\Delta = \mathbb{E} \left[f(X) \int g(y) N(X, dy) \right]$$

□

4.4 Introduction aux martingales

Définitions.

- 1) $(\Omega, \mathcal{A}, \mathbb{P})$ désigne un espace de probabilité usuel.
- 2) Une filtration indexée par I ($I = \{1, 2, \dots, n\}$ ou $\mathbb{N}$) est une famille $\mathcal{F} = (\mathcal{F}_i ; i \in I)$ croissante de sous-tribus de $\mathcal{A}$, ce qui signifie que $\mathcal{F}_i$ est pour tout i une sous-tribu de $\mathcal{A}$ et :

$$\mathcal{F}_i \subset \mathcal{F}_{i+1}, \tag{4.49}$$

pour tout $i \in I$, tel que $i + 1 \in I$.

- 3) Une famille de v.a. $\{X_i ; i \in I\}$ est dite une $\mathcal{F}$ -martingale si

$$X_i \text{ est intégrable pour tout } i \in I, \tag{4.50}$$

$$X_i \text{ est } \mathcal{F}_i\text{-mesurable, } \forall i \in I, \tag{4.51}$$

$$X_i = \mathbb{E}(X_{i+1} | \mathcal{F}_i) \quad \forall i \in I, (i + 1) \in I. \tag{4.52}$$

- 4) On dit que $\{X_i ; i \in I\}$ est une martingale si ce processus est une $\mathcal{F}$ -martingale avec

$$\mathcal{F}_i = \sigma\{X_j ; j \leq i, j \in I\}, \quad \mathcal{F} = (\mathcal{F}_i ; i \in I). \tag{4.53}$$

Exemples.

Exemple 8) Soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. On considère la filtration la plus simple, la filtration constante : $\mathcal{F}_i = \mathcal{B}, \forall i \in I$. On note $\mathcal{F} = (\mathcal{F}_i ; i \in I)$.

Soit $(X_i ; i \in I)$ une $\mathcal{F}$ -martingale. D'après (4.51), X_i est $\mathcal{B}$ -mesurable, donc $X_i = \mathbb{E}(X_{i+1}|\mathcal{B}) = X_{i+1}$. On en déduit que $(X_i ; i \in I)$ est une $\mathcal{F}$ -martingale si et seulement si $i \rightarrow X_i$ est constante, égale à X , et de plus X est intégrable et $\mathcal{B}$ -mesurable.

Exemple 9) Donnons à présent un exemple important de martingales. Soit $(\varepsilon_n ; n \geq 1)$ une suite de v.a. indépendantes. On suppose de plus $\mathbb{E}(|\varepsilon_n|) < +\infty$ et $\mathbb{E}(\varepsilon_n) = 0$, pour tout $n \geq 1$. On pose

$$X_n = \sum_{k=1}^n \varepsilon_k, \quad (4.54)$$

et $\mathcal{F}_n = \sigma(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$. Puisque $X_n - X_{n-1} = \varepsilon_n$, on a $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$. Il est clair que X_n est intégrable et $\mathcal{F}_n$ -mesurable. Vérifions la propriété (4.52) :

$$\mathbb{E}(X_n \mathbb{1}_A) = \mathbb{E}(X_{n+1} \mathbb{1}_A), \quad \forall A \in \mathcal{F}_n. \quad (4.55)$$

Mais $X_{n+1} = X_n + \varepsilon_{n+1}$, par conséquent

$$\mathbb{E}(X_{n+1} \mathbb{1}_A) = \mathbb{E}(X_n \mathbb{1}_A) + \mathbb{E}(\varepsilon_{n+1} \mathbb{1}_A).$$

ε_{n+1} est une v.a. indépendante de $\mathcal{F}_n$, donc

$$\mathbb{E}(\varepsilon_{n+1} \mathbb{1}_A) = \mathbb{E}(\varepsilon_{n+1})\mathbb{P}(A) = 0.$$

L'égalité (4.55) en résulte immédiatement.

Cet exemple constitue en quelque sorte le prototype de martingale. Supposons que ε_n représente le gain (algébrique) d'un joueur, à la n -ième partie. X_n représente la fortune du joueur après n parties. On a fait l'hypothèse que les résultats des différentes parties sont indépendants et qu'en moyenne le jeu est équilibré : à chaque fois le joueur a autant de chance de gagner que de perdre, ce qui se traduit par $\mathbb{E}(\varepsilon_n) = 0, \forall n \geq 1$. Nous verrons (voir théorème 2) qu'il n'existe pas de stratégie optimale permettant d'avoir une espérance de gain positive.

Exemple 10) Soient $\mathcal{F} = (\mathcal{F}_n ; n \geq 0)$ une filtration, U une v.a. intégrable, $\mathcal{A}$ -mesurable. On pose :

$$X_n = \mathbb{E}(U|\mathcal{F}_n). \quad (4.56)$$

Alors $(X_n ; n \geq 0)$ est une $\mathcal{F}$ -martingale. Il est clair que (4.50) et (4.51) sont vérifiées. Par ailleurs (4.49) implique :

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = \mathbb{E}[\mathbb{E}(U|\mathcal{F}_{n+1})|\mathcal{F}_n] = \mathbb{E}(U|\mathcal{F}_n) = X_n.$$

Donc (4.52) est satisfaite.

Pour simplifier nous supposons dans la suite : $I = \mathbb{N}$. Les résultats et définitions s'adaptent aisément au cas où $I = \{1, 2, \dots, n\}$. Sauf mention contraire, $\mathcal{F} = (\mathcal{F}_n ; n \geq 0)$ désigne une filtration donnée et les martingales considérées dans la suite seront des $\mathcal{F}$ -martingales.

Proposition 8 Soient a et b deux réels, $(X_n)_{n \geq 0}$ et $(Y_n)_{n \geq 0}$ deux martingales. Alors

1) $(aX_n + bY_n)_{n \geq 0}$ est une martingale.

2) On a

$$\mathbb{E}(X_m | \mathcal{F}_n) = X_n ; \forall m \geq n \geq 0. \quad (4.57)$$

$$\mathbb{E}(X_n) = \mathbb{E}(X_0), \quad \forall n \geq 0. \quad (4.58)$$

Preuve : 1) Posons $Z_n = aX_n + bY_n$. Il est évident que $(Z_n)_{n \geq 0}$ vérifie les deux propriétés (4.50) et (4.51). De plus

$$\begin{aligned} \mathbb{E}(Z_{n+1} | \mathcal{F}_n) &= \mathbb{E}[aX_{n+1} + bY_{n+1} | \mathcal{F}_n] = a\mathbb{E}(X_{n+1} | \mathcal{F}_n) + b\mathbb{E}(Y_{n+1} | \mathcal{F}_n) \\ &= aX_n + bY_n = Z_n. \end{aligned}$$

2) On suppose n fixé et on raisonne par récurrence sur $m \geq n$. Il est clair que la propriété est vraie pour $m = n$ et $m = n + 1$. On la suppose réalisée pour le rang $m \geq n$, montrons qu'elle a encore lieu pour $m + 1$:

$$\mathbb{E}(X_{m+1} | \mathcal{F}_n) = \mathbb{E}(\mathbb{E}(X_{m+1} | \mathcal{F}_m) | \mathcal{F}_n) = \mathbb{E}(X_m | \mathcal{F}_n) = X_n.$$

3) On prend l'espérance de part et d'autre de (4.57) : $\mathbb{E}(X_n) = \mathbb{E}(X_m)$, pour tout $m \geq n$, ce qui signifie que $n \rightarrow \mathbb{E}(X_n)$ est constante. $\square$

Définition 3 Une application $T : \Omega \rightarrow \mathbb{N} \cup \{+\infty\}$ est un **temps d'arrêt** si :

$$\{T \leq n\} \in \mathcal{F}_n, \quad \text{pour tout } n \geq 0. \quad (4.59)$$

On notera qu'un temps d'arrêt peut prendre la valeur $+\infty$.

Proposition 9 1) Les temps constants sont des temps d'arrêt.

2) Une application $T : \Omega \rightarrow \mathbb{N} \cup \{+\infty\}$ est un temps d'arrêt si et seulement si :

$$\{T = n\} \in \mathcal{F}_n, \quad \text{pour tout } n \geq 0. \quad (4.60)$$

En particulier les temps d'arrêt sont des v.a. $\mathcal{A}$ -mesurables.

3) Si T_1 et T_2 sont deux temps d'arrêt, alors $\sup(T_1, T_2)$, $\inf\{T_1, T_2\}$ et $T_1 + T_2$ sont des temps d'arrêt.

4) Soit $(X_n; n \geq 0)$ une suite de v.a. à valeurs dans $(U, \mathcal{U})$, telle que $X_n : (\Omega, \mathcal{F}_n) \rightarrow (U, \mathcal{U})$ soit mesurable pour tout $n \geq 0$. Pour tout F de $\mathcal{U}$, on note

$$D_F = \inf\{n \geq 0, X_n \in F\} \quad (4.61)$$

avec la convention $\inf \emptyset = +\infty$. Alors D_F est un temps d'arrêt.

Preuve : a) L'assertion 1) est évidente.

b) Soit $n \geq 1$, on a

$$\{T = n\} = \{T \leq n\} \cap \{T > n - 1\} = \{T \leq n\} \cap \{T \leq n - 1\}^c$$

et $\{T \leq n\} = \bigcup_{k=0}^n \{T = k\}$. On en déduit l'équivalence entre (4.59) et (4.60). De plus $\{T = n\} \in \mathcal{F}_n \subset \mathcal{A}$, T est une v.a. $\mathcal{A}$ -mesurable.

c) L'assertion 3 résulte des égalités :

$$\{\sup(T_1, T_2) \leq n\} = \{T_1 \leq n\} \cap \{T_2 \leq n\},$$

$$\{\inf(T_1, T_2) \leq n\} = \{T_1 \leq n\} \cup \{T_2 \leq n\},$$

$$\{T_1 + T_2 = n\} = \bigcup_{k=0}^n \{T_1 = k\} \cap \{T_2 = n - k\}.$$

d) On a,

$$\begin{aligned} \{D_F = n\} &= \{X_0 \notin F, \dots, X_{n-1} \notin F, X_n \in F\} \\ &= \{X_0 \in F^c\} \cap \dots \cap \{X_{n-1} \in F^c\} \cap \{X_n \in F\}. \end{aligned}$$

Donc D_F vérifie (4.60), D_F est bien un temps d'arrêt. Remarquons que :

$$\{D_F = +\infty\} = \{X_n \notin F, \forall n \geq 0\}. \quad (4.62)$$

□

Plaçons-nous dans le cadre de l'exemple . Le joueur peut décider de s'arrêter à un instant fixé à l'avance (il s'agit d'un temps d'arrêt constant). Il peut aussi se retirer du jeu dès que sa fortune est en dessous d'un seuil s , qu'il s'est fixé à l'avance ; il s'arrête en T :

$$T = \inf\{n \geq 0, X_n \leq s\}. \quad (4.63)$$

On a $T = D_F$ avec $F =]-\infty, s]$. T est un temps d'arrêt : le choix de s'arrêter en $T = n$ ne dépend que de $X_0, X_1, \dots, X_n$.

Proposition 10 Soit T un temps d'arrêt fini et $(X_n)_{n \geq 0}$ un processus adapté :

$$X_n \text{ est } \mathcal{F}_n\text{-mesurable pour tout } n \geq 0. \quad (4.64)$$

On pose $X_T(\omega) = X_{T(\omega)}(\omega)$. Alors X_T est une v.a.

Preuve : On suppose que $X_n : (\Omega, \mathcal{F}_n) \rightarrow (U, \mathcal{U})$ est mesurable pour tout $n \geq 0$. Soit $A \in \mathcal{U}$, on a

$$\{X_T \in A\} = \bigcup_{n \geq 0} \{T = n, X_T \in A\} = \bigcup_{n \geq 0} \{T = n, X_n \in A\}.$$

Mais $\{T = n\}$ et $\{X_n \in A\}$ sont deux événements de $\mathcal{F}_n$ donc $\{T = n, X_n \in A\} \in \mathcal{A}$. Par conséquent $\{X_T \in A\} \in \mathcal{A}$. $\square$

Théorème 2 (théorème d'arrêt) *Soit T un temps fini et $(X_n)_{n \geq 0}$ une martingale. On suppose que*

$$T \text{ est un temps d'arrêt borné} \tag{4.65}$$

ou

$$(X_{T \wedge n}; n \geq 0) \text{ est un processus borné,} \tag{4.66}$$

Alors

$$\mathbb{E}(X_T) = \mathbb{E}(X_0). \tag{4.67}$$

Preuve : a) On commence par étudier le cas où T est borné. Il existe un entier k tel que

$$T \leq k. \tag{4.68}$$

D'après (4.57) et (4.60), on a :

$$\mathbb{E}(X_k | \mathcal{F}_n) = X_n \text{ } n \leq k; \{T = n\} \in \mathcal{F}_n.$$

On en déduit :

$$\begin{aligned} \mathbb{E}(X_T) &= \sum_{n=0}^k \mathbb{E}(X_n \mathbb{I}_{\{T=n\}}) = \sum_{n=0}^k \mathbb{E}[\mathbb{E}(X_k | \mathcal{F}_n) \mathbb{I}_{\{T=n\}}] \\ &= \sum_{n=0}^k \mathbb{E}(X_k \mathbb{I}_{\{T=n\}}) = \mathbb{E} \left[X_k \left(\sum_{n=0}^k \mathbb{I}_{\{T=n\}} \right) \right] = \mathbb{E}(X_k). \end{aligned}$$

Mais d'après (4.58), $\mathbb{E}(X_k) = \mathbb{E}(X_0)$.

b) On suppose que (4.66) a lieu : il existe K tel que

$$|X_{T \wedge n}| \leq K, \forall n \geq 0. \tag{4.69}$$

Nous allons utiliser le résultat de l'étape précédente. Posons

$$T_n = T \wedge n = \inf\{n, T\}.$$

T_n est un temps d'arrêt borné, donc $\mathbb{E}(X_{T_n}) = \mathbb{E}(X_0)$.

Mais T est un temps d'arrêt fini, la suite $n \rightarrow T_n$ est croissante et vaut T à partir d'un certain rang, donc X_{T_n} converge p.s. vers X_T . De plus d'après (4.69) la suite de v.a. X_{T_n} est bornée, une application du théorème de convergence dominée, conduit à : $\mathbb{E}(X_T) = \mathbb{E}(X_0)$. $\square$

Remarques.

- 1) En reprenant le contexte de l'exemple , le théorème 2 affirme que l'espérance du gain reste constante, pour toute règle d'arrêt T .
- 2) On peut montrer plus généralement, lorsque (4.66) a lieu, que $(X_{n \wedge T}; n \geq 0)$ est une martingale.
- 3) La propriété d'arrêt (4.67) est vraie avec des hypothèses plus faibles que (4.65) ou (4.66). Toutefois dans les exemples ces hypothèses sont souvent suffisantes.

Les martingales qui ne sont pas "trop grandes" convergent à l'infini. Il s'agit en règle générale de résultats difficiles à établir. Toutefois dans le cadre L^2 , il est facile d'obtenir un résultat de convergence.

Proposition 11 Soit $(X_n)_{n \geq 0}$ une martingale, de carré intégrable :

$$\sup_{n \geq 0} \mathbb{E}(X_n^2) < \infty \quad (4.70)$$

Alors X_n converge dans $L^2(\Omega)$ vers une v.a. X_∞ , de plus

$$X_n = \mathbb{E}(X_\infty | \mathcal{F}_n), \quad \forall n \geq 0. \quad (4.71)$$

Preuve : 1) Soit $n \geq m$. On a :

$$\mathbb{E}((X_n - X_m)^2) = \mathbb{E}(X_n^2) + \mathbb{E}(X_m^2) - 2\mathbb{E}(X_n X_m).$$

Mais $\mathbb{E}(X_n X_m) = \mathbb{E}(\mathbb{E}(X_n X_m | \mathcal{F}_m)) = \mathbb{E}(X_m \mathbb{E}(X_n | \mathcal{F}_m)) = \mathbb{E}(X_m^2)$.

Par conséquent

$$\mathbb{E}((X_n - X_m)^2) = \mathbb{E}(X_n^2) - \mathbb{E}(X_m^2) = a_n - a_m; \quad n \geq m, \quad (4.72)$$

où l'on a posé : $a_n = \mathbb{E}(X_n^2)$. Si $n = m + 1$, on a $a_{m+1} - a_m = \mathbb{E}((X_{m+1} - X_m)^2) \geq 0$.

Ce qui signifie que $n \rightarrow a_n$ est croissante. La propriété (4.70) signifie que (a_n) est majorée donc (a_n) est une suite convergente. Puisque la suite (a_n) est de Cauchy, la relation (4.72) implique que $(X_n)_{n \geq 0}$ est une suite de Cauchy dans $L^2(\Omega)$. Cette suite converge dans $L^2(\Omega)$ vers une v.a. X_∞ , de carré intégrable.

2) Montrons (4.71). On utilise à nouveau (4.57) :

$$\begin{aligned} |\mathbb{E}(X_\infty | \mathcal{F}_n) - X_n| &= |\mathbb{E}(X_\infty | \mathcal{F}_n) - \mathbb{E}(X_{n+k} | \mathcal{F}_n)| = |\mathbb{E}[(X_\infty - X_{n+k}) | \mathcal{F}_n]| \\ &\leq \mathbb{E}(|X_\infty - X_{n+k}| | \mathcal{F}_n), \end{aligned}$$

avec $n \geq 0$ et $k \geq 0$.

On prend l'espérance de part et d'autre de l'inégalité précédente :

$$\mathbb{E}[|\mathbb{E}(X_\infty|\mathcal{F}_n) - X_n|] \leq \mathbb{E}[|X_\infty - X_{n+k}|]$$

Par ailleurs

$$\mathbb{E}[|X_\infty - X_{n+k}|] \leq \{\mathbb{E}[(X_\infty - X_{n+k})^2]\}^{1/2}.$$

Pour tout $n > 0$, la suite $(X_{n+k})_{k \geq 0}$ converge dans $L^2(\Omega)$ vers X_∞ . On fait tendre k vers $+\infty$, on obtient :

$$\mathbb{E}[|\mathbb{E}(X_\infty|\mathcal{F}_n) - X_n|] = 0.$$

D'où $X_n = \mathbb{E}(X_\infty|\mathcal{F}_n)$. □

Remarque.

Réciproquement soit X_∞ une v.a. appartenant à $L^2(\Omega)$, on définit une suite de v.a. $(X_n)_{n \geq 0}$ par (4.71). On a vu dans l'exemple que $(X_n)_{n \geq 0}$ est une martingale. De plus

$$\mathbb{E}[(X_\infty - X_n)^2] = \mathbb{E}(X_\infty^2) - 2\mathbb{E}(X_n X_\infty) + \mathbb{E}(X_n^2),$$

$$\mathbb{E}(X_n X_\infty) = \mathbb{E}(X_n \mathbb{E}(X_\infty|\mathcal{F}_n)) = \mathbb{E}(X_n X_n) = \mathbb{E}(X_n^2).$$

Donc

$$\mathbb{E}((X_\infty - X_n)^2) = \mathbb{E}(X_\infty^2) - \mathbb{E}(X_n^2) \geq 0.$$

En particulier $\mathbb{E}(X_n^2) \leq \mathbb{E}(X_\infty^2)$. Ce qui signifie que $(X_n)_{n \geq 0}$ est une martingale de carré intégrable. D'après la proposition 11, $(X_n)_{n \geq 0}$ converge dans $L^2(\Omega)$ vers $\tilde{X}$. X_∞ et $\tilde{X}$ sont reliées par la relation

$$\mathbb{E}(\tilde{X}|\mathcal{F}_n) = \mathbb{E}(X_\infty|\mathcal{F}_n), \quad \forall n \geq 0.$$

Si l'on suppose de plus que $\mathcal{A} = \bigvee_{n \geq 0} \mathcal{F}_n$, en utilisant le théorème de classe monotone on peut

montrer que $\tilde{X} = X_\infty$.

Les martingales de carré intégrable $(X_n)_{n \geq 0}$, s'identifient aux v.a. X_∞ de $L^2(\Omega)$ à travers (4.71), de plus X_∞ est la limite dans $L^2(\Omega)$ de $(X_n)_{n \geq 0}$.