

Probabilités 2

Y. Aktar

yar@eisti.eu

8 mars 2016

1 Convergence de variables aléatoires

2 Vecteurs aléatoires

3 Vecteurs Gaussiens

La convergence de v.a. est une notion essentielle en probabilités et mène à certains résultats fondamentaux (LGN, TCL, ...). Plusieurs modes de convergence sont possibles en probabilités, ils sont introduits dans ce chapitre, dans lequel les suites de v.a. sont supposées construites sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

Ici nous allons voir 4 types de convergences, à savoir convergence

- presque sûre
- en probabilité
- en norme p
- en loi

La convergence de v.a. correspond à la convergence des fonctions. Pour les fonctions, la convergence la plus faible est la convergence simple qui s'énonce pour des v.a. de la façon suivante :

$$\forall \omega \in \Omega, X_n(\omega) \rightarrow X(\omega).$$

En probabilité (et plus généralement en théorie de la mesure), il est trop restrictif de demander la convergence pour **tous** les $\omega \in \Omega$.

À la place, on considère la convergence presque sûre :

Définition 1

Une suite de v.a. $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement vers X si $\mathbb{P}(X_n \rightarrow X) = 1$. On note cette convergence par $X_n \xrightarrow{p.s.} X$.

Remarque 1 : Ceci est équivalent à dire

$$\exists N \in \mathcal{F}, \mathbb{P}(N) = 0, \forall \omega \in N^c, X_n(\omega) \rightarrow X(\omega).$$

Remarque 2 : Si la limite presque sûre de $(X_n)_n$ existe, alors elle est *essentiellement unique*, i.e.

$$(X_n \xrightarrow{p.s.} X \text{ et } X_n \xrightarrow{p.s.} X') \Rightarrow X = X', \mathbb{P} - p.s.$$

En effet, soit $N, N' \in \mathcal{F}$, t.q. $\mathbb{P}(N) = 0 = \mathbb{P}(N')$ venant de la **Remarque 1**. En prenant $\tilde{N} = N \cup N'$, on obtient grâce à l'inégalité $\mathbb{P}(\tilde{N}) \leq \mathbb{P}(N) + \mathbb{P}(N')$ que pour tout $\omega \in \tilde{N}^c$, $X(\omega) = X'(\omega)$, i.e. $\mathbb{P}(X = X') = 1, \mathbb{P} - p.s.$

Exemple 2

Voici un exemple de suite de v.a. $(X_n)_{n \in \mathbb{N}^*}$, défini sur $(\Omega, \mathcal{F}, \mathbb{P})$ l'espace de probabilité (ou l'espace probabilisé) ($\mathbb{P}(\Omega) = 1$) où $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}([0, 1])$ la tribu borélienne sur l'univers Ω (l'ensemble de toutes les issues (résultats) qui peuvent être obtenues au cours d'une expérience aléatoire) et $\mathbb{P} = \lambda$ mesure de Lebesgue, qui converge presque sûrement mais qui ne converge pas partout :

$$X_n(\omega) = n\chi_{[0, 1/n]}(\omega), \forall n \in \mathbb{N}^*, \forall \omega \in \Omega.$$

On a $X_n \rightarrow 0, \mathbb{P} - p.s.$, en effet si $N = \{0\}$, alors $\omega \in N^c \Rightarrow X_n(\omega) \rightarrow 0$;
et il n'est pas vrai pour cet exemple que $X_n(\omega) \rightarrow 0$ pour tout $\omega \in \Omega$, car $X_n(0) = n \rightarrow \infty$.

Rappel : La tribu borélienne de $[0, 1]$, $\mathcal{B}(\Omega)$ est la plus petite σ -algèbre sur Ω contenant tous les ensembles ouverts de Ω .

Les lemmes de Borel-Cantelli concernent les \liminf et \limsup d'évènements. C'est la clé pour beaucoup de problèmes concernant la convergence presque sûre, comme nous le verrons plus tard.

Théorème 3 (Lemme de Borel-Cantelli)

Soit $(A_n)_{n \in \mathbb{N}}$ une suite (infinie) d'évènements d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On a

i) $\sum_{n \geq 1} P(A_n) < +\infty \Rightarrow \mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) = 0.$

ii) Si les A_n sont indépendants, $\sum_{n \geq 1} P(A_n) = +\infty \Rightarrow \mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) = 1.$

Quelques rappels concernant \limsup et \liminf :

$$\limsup_{n \rightarrow +\infty} A_n \stackrel{\text{déf}}{=} \bigcap_{n \geq 1} \bigcup_{k > n} A_k =: A^* \quad \text{et} \quad \liminf_{n \rightarrow +\infty} A_n \stackrel{\text{déf}}{=} \bigcup_{n \geq 1} \bigcap_{k > n} A_k =: A_*.$$

Noter que $A_* \subset A^*$. En effet $\bigcap_{k > n} A_k \subset A_q$ pour tout $q > n$. On a donc $\bigcap_{k > n} A_k \subset \bigcup_{q > p} A_q$ pour tout $p \in \mathbb{N}^*$. On a alors pour tout n :

$$\bigcap_{k > n} A_k \subset \bigcap_{p \geq 1} \bigcup_{q > p} A_q = A^*, \quad \text{finalement} \quad \bigcup_{n \geq 1} \bigcap_{k > n} A_k \subset A^*.$$

C'est à dire $A_* \subset A^*$.

Puis, les règles élémentaires sur \cup , \cap et la complémentarité, notée par c , donnent sans difficulté

$$(A^*)^c = (A^c)_*.$$

De plus, A^* et A_* ont les interprétations suivantes :

Propriété 1

Soit $(A_j)_{j \in \mathbb{N}}$ une collection infinie d'ensembles. Alors

i) "À partir d'un certain rang, ω est dans tous les A_j " s'écrit

$$\omega \in \bigcup_{i \geq 0} \bigcap_{j > i} A_j \quad (= A_*).$$

ii) " ω est dans une infinité de A_j " s'écrit

$$\omega \in \bigcap_{i \geq 0} \bigcup_{j > i} A_j \quad (= A^*).$$

On écrit parfois $\{A_n \text{ i.s.}\}$ où *i.s.* signifie "infiniment souvent".

Démonstration :

- i) Soit ω qui, à partir d'un certain rang, est dans tous les A_j . On traduit cela de la façon suivante : il existe un rang i tel que pour tout rang $j > i$, ω est dans A_j . D'après la signification des symboles \forall , \exists , \cap et \cup , cela revient à écrire

$$\omega \in \underbrace{\bigcup_{i \geq 0}}_{\text{il existe } i \geq 0} \underbrace{\bigcap_{j > i}}_{\text{pour tout } j > i} \underbrace{A_j}_{\omega \text{ est dans } A_j} .$$

- ii) Dire que ω est dans une infinité de A_j est équivalent à dire que

" pour tout p , il existe $q > p$ avec ω dans A_q ".

En effet, si tel est le cas, ω est bien dans une infinité de A_j car, d'après cette propriété,

- avec $p = 0$, il existe $p_1 > p$ tel que ω est dans A_{p_1} ,
- avec $p = p_1$, il existe $p_2 > p_1$ tel que ω est dans A_{p_2} ,
- avec $p = p_2$, il existe $p_3 > p_2$ tel que ω est dans A_{p_3} ,
- ... ,
- avec $p = p_n$, il existe $p_{n+1} > p_n$ tel que ω est dans $A_{p_{n+1}}$,
- ... ,

et finalement, ω est dans chaque A_{p_n} , $n \in \mathbb{N}^*$, c'est-à-dire dans une infinité de A_j .

Réciproquement, s'il est dans une infinité de A_i , alors pour tout p on trouve $q > p$ tel que $\omega \in A_q$, sinon, ce serait qu'il existe p tel que pour $q > p$, ω n'est pas dans A_q . Ou encore : ω ne peut appartenir qu'aux A_i d'indices $i \leq p$, c'est-à-dire seulement un nombre fini d'entre eux, ce qui est faux.

Donc, pour ce deuxième point, pour tout p , on trouve $q > p$, tel que $\omega \in A_q$, en langage \forall, \exists , cela s'écrit

$$\omega \in \underbrace{\bigcap_{p \geq 0}}_{\text{pour tout } p \geq 0} \underbrace{\bigcup_{q > p}}_{\text{il existe } q > p} \underbrace{A_q}_{\omega \text{ est dans } A_q} .$$

□

Lien entre liminf, limsup d'ensembles et de fonctions

Rappel : Pour une suite réelle $u = (u_n)_{n \in \mathbb{N}}$, on définit ses limites supérieur et inférieure par

$$u_* := \liminf u_n = \sup_{n \geq 0} \inf_{k \geq n} u_k$$

$$u^* := \limsup u_n = \inf_{n \geq 0} \sup_{k \geq n} u_k$$

Ce sont les plus petite et plus grande valeurs d'adhérence de la suite u . On a toujours

$$u_* \leq u^*$$

et il y a égalité ssi la suite u converge ; de plus si tel est le cas $\lim_{n \rightarrow +\infty} u_n = u_* = u^*$. En plus, en changeant le signe, les limites inférieure et supérieure s'échangent :

$$(-u)^* = -u_*$$

$$(-u)_* = -u^*.$$

Pour une suite de fonctions $(f_n)_{n \geq 0}$, on définit des fonctions limites inférieure et supérieure de la façon suivante :

$$f_*(x) = \liminf_{n \rightarrow +\infty} f_n(x), \quad f^*(x) = \limsup_{n \rightarrow +\infty} f_n(x).$$

Propriété 2

Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'ensemble mesurables, on a

$$(\mathbf{1}_A)_* = \mathbf{1}_{A_*}, \quad (\mathbf{1}_A)^* = \mathbf{1}_{A^*}.$$

Démonstration : Il suffit de le faire pour $(\mathbf{1}_A)^*$. Or $\omega \in (\mathbf{1}_A)^*$ si $\forall k, \exists n \geq k$ t.q. $\omega \in A_n$. On a donc $\forall k, \exists n \geq k$ avec $\mathbf{1}_{A_n(\omega)} = 1$ ou encore $\inf_{k \geq 0} \sup_{n \geq k} \mathbf{1}_{A_n(\omega)} = 1$, c'est-à-dire $(\mathbf{1}_{A(\omega)})^* = 1$. Manifestement, il s'agit d'une équivalence. □

Définition 4 (Convergence d'ensembles)

On dit qu'une suite d'ensemble $(A_n)_{n \in \mathbb{N}}$ converge vers A et on note $A_n \rightarrow A, n \rightarrow +\infty$, si

$$A^* = A_* = A.$$

On montre alors aussi :

Propriété 3

$$\mathbb{P}(A_*) \leq (\mathbb{P}(A))_* \leq (\mathbb{P}(A))^* \leq \mathbb{P}(A^*).$$

On en déduit que

$$A_n \rightarrow A \Rightarrow \mathbb{P}(A_n) \rightarrow \mathbb{P}(A), n \rightarrow +\infty.$$

Démonstration : Soient $B_n = \bigcup_{k > n} A_k$ et $C_n = \bigcap_{k > n} A_k$. Alors $(B_n)_{n \in \mathbb{N}}$ est une suite décroissante d'ensembles, de limite A^* et $(C_n)_{n \in \mathbb{N}}$ est une suite croissante d'ensembles, de limite A_* . D'après la propriété de convergence monotone des mesures (de probabilités), on a

$$\mathbb{P}(A_n) \leq \mathbb{P}(B_n) \rightarrow \mathbb{P}(A^*)$$

et

$$\mathbb{P}(A_n) \geq \mathbb{P}(C_n) \rightarrow \mathbb{P}(A_*)$$

On en déduit $\mathbb{P}(A_*) \leq (\mathbb{P}(A))_*$ et $(\mathbb{P}(A))^* \leq \mathbb{P}(A^*)$. Le second point est clair car les limites supérieure et inférieure coïncident lorsque $A_n \rightarrow A, n \rightarrow +\infty$.

Démonstration :

- i) Par définition, $\limsup_{n \rightarrow +\infty} A_n \subset \bigcup_{m \geq n} A_m$ pour tout $n \geq 0$. On a donc

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) \leq \mathbb{P}\left(\bigcup_{m \geq n} A_m\right) \leq \sum_{m \geq n} \mathbb{P}(A_m). \quad (1)$$

Or $\sum_{m \geq n} \mathbb{P}(A_m)$ est le reste d'ordre n de la série convergente $\sum_{n \geq 1} \mathbb{P}(A_n)$. Comme le reste d'une série convergente tend vers 0, un passage à la limite $n \rightarrow +\infty$ dans (1) conclut.

- ii) Soit $M < N < +\infty$. L'indépendance et l'inégalité $1 - x \leq e^{-x}$ impliquent

$$\begin{aligned} \mathbb{P}\left(\bigcap_{n=M}^N (A_n)^c\right) &= \prod_{n=M}^N (1 - \mathbb{P}(A_n)) \leq \prod_{n=M}^N e^{-\mathbb{P}(A_n)} \\ &= e^{-\sum_{n=M}^N \mathbb{P}(A_n)} \rightarrow 0 \text{ quand } N \rightarrow +\infty. \end{aligned}$$

Ainsi $\mathbb{P}(\bigcup_{n \geq M} A_n) = 1$ pour tout M et puisque $\bigcup_{n \geq M} A_n \downarrow A^*$, il s'en suit que $\mathbb{P}(A^*) = 1$. □

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. pas nécessairement indépendantes, ni identiquement distribuées, ainsi qu'une v.a.

Définition 5

Nous disons que $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X si

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

Ceci est noté : $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$

Interprétation : Cette notion de convergence peut se comprendre de la manière suivante. Pour tout écart ε fixé, lorsque n devient très grand, il est de moins en moins probable d'observer un écart, supérieur à l'écart donné, entre X_n et X .

Exemple 6

Soit $a > 0$ un nombre fixé. Soient la v.a. $X = 0$ et la suite de v.a. $(X_n)_{n \in \mathbb{N}}$ t.q. :

$$\forall n \in \mathbb{N}, \mathbb{P}(X_n = n^a) = \frac{1}{n} \text{ et } \mathbb{P}(X_n = 0) = 1 - \frac{1}{n}.$$

Il est facile alors de constater que, pour tout $\varepsilon > 0$ fixé et pour tout $n \geq 1$, nous avons :

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n| > \varepsilon) = \mathbb{P}(X_n = n^a) = \frac{1}{n} \rightarrow 0.$$

Ainsi nous en déduisons $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$.

On rappelle que pour $p \geq 1$, on définit les espaces L^p , pour un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, par :

$$L^p(\Omega, \mathcal{F}, \mathbb{P}) = \{X : \Omega \rightarrow \mathbb{R} \text{ t.q. } \|X\|_p = (\mathbb{E}[|X|^p])^{1/p} < +\infty\}.$$

Ce sont des espaces de Banach pour la norme $\|\cdot\|_p$. Remarquez que pour $p = 2$, $L^2(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de Hilbert car la norme 2 dérive du produit scalaire suivant

$$\langle X, Y \rangle = \mathbb{E}[XY].$$

Définition 7

On dit qu'une suite de v.a. $(X_n)_{n \in \mathbb{N}}$ converge en norme p vers X si $\|X_n - X\|_p \rightarrow 0$ quand $n \rightarrow +\infty$, c'est-à-dire

$$\lim_{n \rightarrow +\infty} \mathbb{E}[\|X_n - X\|^p] = 0.$$

On note la convergence en norme p , par $X_n \xrightarrow[n \rightarrow +\infty]{L^p} X$.

Exemple 8

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a. telle que pour tout $n \in \mathbb{N}^*$, $X_n \sim \mathcal{N}(0, n^{-2})$. Alors Comme $\mathbb{E}[|\sqrt{n}X_n|^2] = n\mathbb{E}[X_n^2] = n \times n^{-2} = \frac{1}{n} \rightarrow 0$, d'où $X_n \xrightarrow[n \rightarrow +\infty]{L^2} 0$.

Définition 9

On dit qu'une suite de v.a. $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si en tout point de continuité $t \in \mathbb{R}$ de $F_X(t)$, nous avons

$$\lim_{n \rightarrow +\infty} F_{X_n}(t) = F_X(t).$$

Ceci est noté $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$.

Exemple 10

Soit $(Y_n)_{n \in \mathbb{N}}$ une suite de v.a. i.i.d. de loi Exponentielle (i.e. de fonction densité $f_{Y_n}(t) = \beta e^{-\beta t} \mathbf{1}_{t>0}$). Nous posons $X_n = \min_{1 \leq i \leq n} Y_i$. On trouve que

$$F_{X_n}(t) = (1 - e^{-n\beta t}) \mathbf{1}_{t>0} \xrightarrow[n \rightarrow +\infty]{} \mathbf{1}_{t>0}.$$

La fonction de répartition de la v.a. définie par $\mathbb{P}(X=0) = 1$ est donnée par $F_X(t) = \mathbf{1}_{t \geq 0}$. Nous constatons donc $\lim_{n \rightarrow +\infty} F_{X_n}(t) = F_X(t)$ en tout point de continuité de F_X , c'est-à-dire pour tout nombre réel $t \neq 0$. Donc on a bien $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$.

Théorème 11

$$X_n \xrightarrow[n \rightarrow +\infty]{p.s.} X \Leftrightarrow \forall \varepsilon > 0, \mathbb{P}\left(\sup_{m \geq n} |X_m - X| > \varepsilon\right) \xrightarrow[n \rightarrow +\infty]{} 0.$$

Démonstration : Comme $X_n \xrightarrow[n \rightarrow +\infty]{p.s.} X \Leftrightarrow X_n - X \xrightarrow[n \rightarrow +\infty]{p.s.} 0$, il suffit de montrer le théorème dans le cas où $X = 0$, i.e.

$$X_n \xrightarrow[n \rightarrow +\infty]{p.s.} 0 \Leftrightarrow \forall \varepsilon > 0, \mathbb{P}\left(\sup_{m \geq n} |X_m| > \varepsilon\right) \xrightarrow[n \rightarrow +\infty]{} 0.$$

Cas \Rightarrow : Soit $\varepsilon > 0$. Posons $A_n(\varepsilon) = \{\sup_{m \geq n} |X_m| > \varepsilon\}$ et $C = \{\lim_{n \rightarrow +\infty} X_n = 0\}$ et $B_n(\varepsilon) = C \cap A_n(\varepsilon)$ pour tout $n \in \mathbb{N}^*$. On a $B_{n+1}(\varepsilon) \subset B_n(\varepsilon)$, et $\bigcap_{n \geq 1} B_n(\varepsilon) = \emptyset$. D'où

$$\lim_{n \rightarrow +\infty} \mathbb{P}(B_n(\varepsilon)) = \mathbb{P}\left(\bigcap_{n \geq 1} B_n(\varepsilon)\right) = 0.$$

Comme $\mathbb{P}(C) = 1, \mathbb{P}(C^c) = 0$, on obtient donc

$$\begin{aligned} \mathbb{P}(B_n(\varepsilon)) &= \mathbb{P}(C \cap A_n(\varepsilon)) = 1 - \mathbb{P}(C^c \cup (A_n(\varepsilon))^c) \\ &= 1 - \mathbb{P}(C^c) - \mathbb{P}((A_n(\varepsilon))^c) + \mathbb{P}(C^c \cap (A_n(\varepsilon))^c) \\ &= \mathbb{P}(A_n(\varepsilon)) + \mathbb{P}(C^c \cap (A_n(\varepsilon))^c) \\ &= \mathbb{P}(A_n(\varepsilon)). \quad \square \end{aligned}$$

Cas \Leftarrow : Supposons $\lim_{n \rightarrow +\infty} \mathbb{P}(A_n(\varepsilon)) = 0$ et posons $D(\varepsilon) = \{|X| \vphantom{X} \}^* > \varepsilon\}$. On a $D(\varepsilon) \subset A_n(\varepsilon)$ pour tout $n \in \mathbb{N}^*$, alors $\mathbb{P}(D(\varepsilon)) = 0$. On a

$$C^c = \left\{ \lim_{n \rightarrow +\infty} X_n \neq 0 \right\} \subset \bigcup_{k \geq 1} \left\{ (|X| \vphantom{X} \}^* > \frac{1}{k} \right\},$$

donc

$$1 - \mathbb{P}(C) \leq \sum_{k \neq 1} \mathbb{P}(D(1/k)) = 0,$$

qui implique que $\mathbb{P}(C) = 1$, i.e. $X_n \xrightarrow[n \rightarrow +\infty]{p.s.} 0$.

Montrons que

$$\forall \varepsilon > 0, \sum_{n \geq 1} \mathbb{P}(|X_n - X| > \varepsilon) < +\infty \Rightarrow X_n \rightarrow X, p.s.$$

Démonstration : Notons $A_n := \{|X_n - X| > \varepsilon\}$ pour tout $n \in \mathbb{N}^*$. Par hypothèse, $\sum_{n \geq 1} \mathbb{P}(A_n) < +\infty$.

D'après le lemme de Borel-Cantelli, on a $\mathbb{P}(A^*) = 0$, i.e. $\mathbb{P}((A^*)^c) = 1$, ou encore $\mathbb{P}((A^c)_*) = 1$. D'où

$$\mathbb{P}(\{\text{\`a partir d'un certain rang } \omega \text{ est dans tous les } (A_i)^c\}) = 1.$$

C'est-à-dire qu'il existe $B \in \mathcal{F}$ tel que $\mathbb{P}(B) = 1$, pour tout $\omega \in B$, il existe $n_0(\omega) \in \mathbb{N}^*$ tel que pour tout $n \geq n_0(\omega)$, $\omega \in (A_n)^c$ (ou encore $\omega \notin A_n$), i.e. $|X_n(\omega) - X(\omega)| \geq \varepsilon$. D'où

$$X_n \rightarrow X \text{ p.-s.}$$



Des variables aléatoires uni-dimensionnelles ont été introduites pendant le cours de Probabilités 1. De façon analogue nous définissons des variables aléatoires multi-dimensionnelles (ou des vecteurs aléatoires).

Quelques rappels :

La Tribu Borélienne de \mathbb{R}^n , $\mathcal{B}(\mathbb{R}^n)$, est définie comme étant la tribu engendrée par les ouverts de \mathbb{R}^n , dont les éléments sont appelés ensembles boréliens. La tribu borélienne \mathbb{R}^n est engendrée par les pavés (produit d'intervalles).

$$\mathcal{B}(\mathbb{R}^n) = \sigma(\{\prod_{i=1}^n]a_i, b_i[, a_i, b_i \in \mathbb{R}, a_i < b_i, \forall i\})$$

La mesure de Lebesgue λ_n sur \mathbb{R}^n est définie par $\lambda_n(\prod_{i=1}^n]a_i, b_i[, a_i, b_i) = \prod_{i=1}^n (b_i - a_i)$ (revoir le cours de Mesure-Intégration)

Définition 12 (vecteur aléatoire n -dimensionnel)

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. On dit que X est vecteur aléatoire de dimension $n \geq 1$ $X : \Omega \rightarrow \mathbb{R}^n$ est mesurable, i.e. lorsque

$$\{\omega \in \Omega \text{ t.q. } X(\omega) \in B\} \in \mathcal{F}, \forall B \in \mathcal{B}(\mathbb{R}^n). \quad (2)$$

$\{\omega \in \Omega \text{ t.q. } X(\omega) \in B\} := X^{-1}(B)$ et la propriété (2) est notée par $X^{-1}(\mathcal{B}(\mathbb{R}^n)) \subset \mathcal{F}$.

Propriété 4

Soit \mathcal{C} un ensemble d'éléments de \mathbb{R}^n t.q. $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R}^n)$. Alors

$$X \text{ est mesurable} \Leftrightarrow X^{-1}(\mathcal{C}) \subset \mathcal{F}.$$

Démonstration : Le sens \Leftarrow étant claire. Montrons le sens \Rightarrow : Supposons que $X^{-1}(C) \in \mathcal{F}$ pour tout $C \in \mathcal{C}$. On doit montrer que $X^{-1}(\Lambda) \in \mathcal{F}$ pour tout $\Lambda \in \mathcal{B}(\mathbb{R}^n)$. Notons que $X^{-1}(\cup_n \Lambda_n) = \cup_n X^{-1}(\Lambda_n)$, $X^{-1}(\cap_n \Lambda_n) = \cap_n X^{-1}(\Lambda_n)$, et $X^{-1}(\Lambda^c) = (X^{-1}(\Lambda))^c$. Soit $U = \{A \in \mathcal{B}(\mathbb{R}^n), \text{ t.q. } X^{-1}(A) \in \mathcal{F}\}$. Alors $\mathcal{C} \subset U$, et comme X^{-1} commutes avec des intersections et unions dénombrables, et compléments, il vient alors U est une tribu. Alors $\sigma(\mathcal{C}) \subset U$, et $U \subset \mathcal{B}(\mathbb{R}^n)$, et comme $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{C})$ on conclut alors $\mathcal{B}(\mathbb{R}^n) = U$, et donc $X^{-1}(\mathcal{B}(\mathbb{R}^n)) \subset \sigma(X^{-1}(\mathcal{C})) \subset \mathcal{F}$. □

Corollaire 1

$X = (X_1, \dots, X_n)$ est un vecteur aléatoire si et seulement si

$$\{\omega \in \Omega \text{ t.q. } X_1(\omega) \leq t_1, \dots, X_n(\omega) \leq t_n\} \in \mathcal{F}, \forall t_1, \dots, t_n \in \mathbb{R}.$$

C'est-à-dire lorsque $\left(\bigcap_{i=1}^n \{X_i \leq t_i\} \right) \in \mathcal{F}, \forall t_1, \dots, t_n \in \mathbb{R}$.

Démonstration : Prendre $\mathcal{C} = \left\{ \prod_{i=1}^n]-\infty, t_i] \right\}, t_i \in \mathbb{R}, \forall i \in \{1, \dots, n\}$. □

Il existe deux classes importantes de vecteurs aléatoires :

Définition 13 (Vecteur aléatoire absolument continue)

Soit X un vecteur aléatoire, on l'appelle vecteur aléatoire absolument continu si

$$\mathbb{P}(\{X \in B\}) = 0, \forall B \in \mathcal{B}(\mathbb{R}^n) \text{ t.q. } \lambda_n(B) = 0.$$

C'est-à-dire que μ_X est absolument continue par rapport à la mesure λ_n où μ_X est la mesure de probabilité définie par

$$\mu_X(B) = \mathbb{P}(\{X \in B\}), \forall B \in \mathcal{B}(\mathbb{R}^n).$$

Rappelons que une mesure μ est absolument continue par rapport à la mesure λ_n , si pour tout $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\lambda_n(A) = 0 \Rightarrow \mu(A) = 0.$$

On dit aussi que μ est dominée par λ_n , que l'on note, dans ce cours, par $\mu \ll \lambda_n$.

Dans ce cas, d'après le Théorème de Radon-Nikodym, il existe une fonction mesurable $f_X : \mathbb{R}^n \rightarrow \mathbb{R}$ positive t.q.

$$\mu_X(B) = \int_B f_X d\lambda_n, \forall B \in \mathcal{B}(\mathbb{R}^n).$$

i.e. f_X est la densité de μ_X par rapport à λ_n . f_X est unique λ_n -presque partout (i.e. $\lambda_n(\{f_X \neq \tilde{f}_X\}) = 0$).

Lorsque X est un vecteur aléatoire absolument continu, on appelle f_X la fonction de densité de X ou la fonction de densité conjointe de X_1, \dots, X_n .

Remarquons que $f_X \geq 0$ et $\int f_X d\lambda_n = 1$. On note $\int g d\lambda_n$ par $\int g(x) dx$.

Définition 14 (Vecteur aléatoire discret)

Soit X un vecteur aléatoire, on l'appelle vecteur aléatoire discret si X prend ses valeurs dans un sous-espace **dénombrable** de \mathbb{R}^n . C'est-à-dire

$$\exists B \in \mathcal{B}(\mathbb{R}^n) \text{ dénombrable t.q. } \mu_X(B) = 1.$$

Dans ce cas, on appelle fonction de masse de X ou distribution de probabilité discrète multivariée des variables aléatoires discrètes X_1, \dots, X_n et on note m_X ou $m_{(X_1, \dots, X_n)}$ la fonction définie par

$$m_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \mathbb{P}\left(\bigcap_{i=1}^n \{X_i = x_i\}\right) = m_X(x) = \mathbb{P}(X = x), x \in B, \quad (3)$$

où B est le domaine (c'est-à-dire l'ensemble de toutes les valeurs permises) du vecteur aléatoire discret X , dénombrable.

m_X a les propriétés suivantes (montrez les) :

- (i) $0 \leq m_X(x) \leq 1$ pour tout $x \in B$;
- (ii) $\sum_{x \in B} m_X(x) = 1$;
- (iii) Si $B_1 \subset B$, alors

$$\mathbb{P}(X \in B_1) = \sum_{x \in B_1} m_X(x).$$

Définition 15

La fonction $F_{(X_1, \dots, X_n)} : \mathbb{R}^n \rightarrow [0, 1]$ définie par $F_{(X_1, \dots, X_n)}(x) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$ est appelée la fonction de répartition de X , notée dans ce cours par, $F_X(x)$.

Donc pour un vecteur aléatoire discret X , on a

$$F_X(x) = \sum_{y \leq x} m_X(y).$$

C'est-à-dire

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \sum_{y_1 \leq x_1, y_1 \in B_1} \cdots \sum_{y_n \leq x_n, y_n \in B_n} \mathbb{P}(X_1 = y_1, \dots, X_n = y_n).$$

Et pour un vecteur aléatoire X absolument continu, on a

$$F_X(x) = \mu_X \left(\prod_{i=1}^n]-\infty, x_i] \right) = \int_{\prod_{i=1}^n]-\infty, x_i]} f_X(y) dy = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_X(y_1, \dots, y_n) dy_1 \dots dy_n.$$

Soit X un vecteur aléatoire discret et soit I un sous-ensemble d'indices de $\{1, \dots, n\}$ de cardinal $k \leq n$. C'est-à-dire qu'il existe $i_1, \dots, i_k \in \{1, \dots, n\}$ t.q. $i_1 < i_2 < \dots < i_k$.

Que vaut $m_{(X_{i_1}, \dots, X_{i_k})}(x_{i_1}, \dots, x_{i_k})$? appelée la fonction de densité marginale de $(X_{i_1}, \dots, X_{i_k})$.

Prenons $n = 3$ et $k = 2$ avec $i_1 = 1$ et $i_2 = 3$ pour simplifier les calculs et donc

$$m_{(X_1, X_3)}(x_1, x_3) = \mathbb{P}(X_1 = x_1, X_3 = x_3) = \mathbb{P}\left(\bigcup_{y \in B_2} \{X_1 = x_1, X_3 = x_3, X_2 = y\}\right).$$

D'où par l'union d'événements disjoints et l'additivité de \mathbb{P} , on obtient :

$$m_{(X_1, X_3)}(x_1, x_3) = \sum_{y \in B_2} \mathbb{P}(\{X_1 = x_1, X_2 = y, X_3 = x_3\}) = \sum_{y \in B_2} m_{(X_1, X_2, X_3)}(x_1, y, x_3).$$

Exercice : Généraliser pour n et k quelconques.

De façon analogue, on obtient les fonctions $F_{(X_{i_1}, \dots, X_{i_k})}$ appelées fonctions de répartitions marginales :

Exemple avec $n = 2$ et $k = 1$ où on prend $i_1 = 1$:

$$F_{X_1}(x) = \sum_{x' \leq x} m_{X_1}(x') = \sum_{x' \leq x} \sum_y m_{(X_1, X_2)}(x', y).$$

Pour la simplicité prenons $n = 2$. C'est-à-dire que nous nous occupons des vecteurs aléatoires bidimensionnels. Le théorème suivant peut se généraliser au cas où $n = 2, 3, 4, \dots$

Théorème 16

Soit $X = (Y, Z)$ un vecteur aléatoire bidimensionnel ayant une densité f sur \mathbb{R}^2 . Alors

a) Y et Z ont pour fonctions de densités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ données par

$$f_Y(t) = \int_{-\infty}^{\infty} f_X(y, z) dz; \quad f_Z(z) = \int_{-\infty}^{\infty} f_X(y, z) dy.$$

b) Y et Z sont indépendantes si et seulement si

$$f_X(y, z) = f_Y(y)f_Z(z) \quad \lambda_2 - \text{presque partout.}$$

c) La formule précédente permet de définir une autre fonction de densité sur \mathbb{R} à tout point $y \in \mathbb{R}$ t.q. $f_Y(t) \neq 0$:

$$f_{Y=y}(z) := \frac{f_X(y, z)}{f_Y(y)}.$$

Démonstration :

a) Pour tout $A \in \mathcal{B}(\mathbb{R})$, on a

$$\begin{aligned}\mathbb{P}(Y \in A) &= \mathbb{P}(X \in A \times \mathbb{R}) = \int \int_{A \times \mathbb{R}} f_X(y, z) dy dz \\ &= \int_A \left(\int_{-\infty}^{\infty} f_X(y, z) dz \right) dy = \int_A f_Y(y) dy.\end{aligned}$$

et comme l'égalité est vraie pour tout $A \in \mathcal{B}(\mathbb{R})$ on obtient donc $\mu_Y(A) = \int_A f_Y(y) dy$ avec $f_Y(y) \geq 0$ et $\mu_Y(\mathcal{B}(\mathbb{R})) = 1$. D'où f_Y est une fonction de densité de Y .

La preuve pour Z est la même.

b) Supposons que $f_X(y, z) = f_Y(y)f_Z(z)$. Alors

$$\begin{aligned}\mathbb{P}(Y \in A, Z \in B) &= \int \int \mathbf{1}_{A \times B}(y, z) f_X(y, z) dy dz \\ &= \dots = \mathbb{P}(Y \in A) \mathbb{P}(Z \in B),\end{aligned}$$

et puisque A et B sont des boréliens arbitraires, Y et Z sont des v.a. indépendantes.

Le sens inverse est laissé en exercice (penser à appliquer le théorème de Tonelli-Fubini).

c) On a

$$\begin{aligned}\int f_{Y=y}(z) dz &= \int_{-\infty}^{\infty} \frac{f_X(y,z)}{f_Y(z)} dz \\ &= \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f_X(y,z) dz = \frac{1}{f_Y(y)} f_Y(y) = 1.\end{aligned}$$

Comme $f_{Y=y}(z)$ est positive, mesurable, d'intégrale égale à 1, alors c'est une fonction de densité. □

Tournons notre attention vers des fonctions des vecteurs aléatoires, ici de dimension n . Le problème est le suivant :

Soit $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ une fonction borélienne, c'est-à-dire que pour tout $B \in \mathcal{B}(\mathbb{R}^n)$, $g^{-1}(B) \in \mathcal{B}(\mathbb{R}^n)$. Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire admettant une densité notée f_X , quelle est la densité de $Y = g(X)$ en termes de f_X ?

Rappelons-nous que si g est une fonction dérivable de G , un ouvert de \mathbb{R}^n , dans \mathbb{R}^n , sa matrice Jacobienne $J_g(x)$ au point G est $J_g(x) = \frac{\partial g}{\partial x}(x)$ (c'est-à-dire $(J_g(x))_{ij} = \frac{\partial g_i}{\partial x_j}(x)$, où $g = (g_1, g_2, \dots, g_n)$). Le Jacobien de g au point x est le déterminant de la matrice Jacobienne $J_g(x)$. Si le Jacobien n'est pas égal à 0, alors g est inversible sur un voisinage de x , et le Jacobien de la fonction inverse g^{-1} au point $y = g(x)$ est l'inverse du Jacobien de g au point x , c'est-à-dire $\det(J_{g^{-1}}(g(x))) = \frac{1}{\det(J_g(x))}$.

Théorème 17 (Formule de Transformation Jacobienne (changement de variable))

Soit G un ouvert de \mathbb{R}^n et soit $g : G \rightarrow \mathbb{R}^n$ une fonction C^1 . Supposons que g est injective sur G et son Jacobien n'est pas nul. Alors pour f mesurable et telle que $f \mathbf{1}_{g(G)}$ est positive ou intégrable par rapport à la mesure de Lebesgue, on a

$$\int_{g(G)} f(y) dy = \int_G f(g(x)) |\det(J_g(x))| dx$$

où $g(G) = \{y \in \mathbb{R}^n : \text{il existe } x \in G \text{ avec } g(x) = y\}$.

Théorème 18

Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire admettant une densité f . Soit $g: \mathbb{R}^n \rightarrow \mathbb{R}^n$ une fonction C^1 et injective, avec son Jacobien qui n'est pas nul. Alors le vecteur aléatoire $Y = g(x)$ admet une densité, appelée f_Y définie par

$$f_Y(y) = f_X(g^{-1}(y)) |\det(J_{g^{-1}}(y))| \mathbf{1}_{g(\mathbb{R}^n)}(y),$$

pour tout $y \in \mathbb{R}^n$ où $g(\mathbb{R}^n) = \{y \in \mathbb{R}^n : \text{il existe } x \in \mathbb{R}^n \text{ avec } g(x) = y\}$.

Démonstration : Les propriétés de g impliquent que G est un ouvert et que la fonction inverse g^{-1} est bien définie sur G et est C^1 avec son Jacobien non nul. Soit $B \in \mathcal{B}(\mathbb{R}^n)$, et $A = g^{-1}(B)$. On a

$$\mathbb{P}(X \in A) = \int_A f_X(x) dx = \int_{g^{-1}(B)} f_X(x) dx = \int_B f_X(g^{-1}(x)) |\det(J_{g^{-1}}(x))| dx$$

grâce la formule transformation Jacobienne avec g^{-1} . Mais on a aussi $\mathbb{P}(Y \in B) = \mathbb{P}(X \in A)$, puisque $\mathbb{P}(X \in A) = \int_B f_Y(y) dy$. Comme $B \in \mathcal{B}(\mathbb{R}^n)$ est arbitraire, on conclut que

$$f_X(g^{-1}(x)) |\det(J_{g^{-1}}(x))| = f_Y(x).$$

On peut donner un corollaire où on peut traiter le cas où g n'est pas injective mais néanmoins C^1 :

Propriété 5

Soit $S \in \mathcal{B}(\mathbb{R}^n)$ tel que S est la réunion disjointe de sous-ensembles S_0, S_1, \dots, S_m tels que $\cup_{i=0}^m S_i = S$, et que $\lambda_n(S_0) = 0$ et que pour tout $i \in \{1, \dots, m\}$, $g : S_i \rightarrow \mathbb{R}^n$ est injective et C^1 de Jacobien non nul. Soit $Y = g(X)$, où X est un vecteur aléatoire n -dimensionnel à valeurs dans S et admettant une densité f_X . Alors Y admet une densité donnée par

$$f_Y(y) = \sum_{i=1}^m f_X(g_i^{-1}(y)) |\det(J_{g_i^{-1}}(y))| \mathbf{1}_{g(S_i)}(y),$$

pour tout $y \in \mathbb{R}^n$ où g_i^{-1} est la fonction inverse $g_i^{-1} : g(S_i) \rightarrow S_i$ et $J_{g_i^{-1}}$ est sa matrice Jacobienne correspondante et $g(S_i) = \{y \in \mathbb{R}^n : \exists x \in S_i, g(x) = y\}$.

Rappelons quelques propriétés connues :

Définition 19 (Covariance de deux v.a. réelles)

Soient X et Y deux v.a. réelles t.q. $\text{Var}(X), \text{Var}(Y) < +\infty$. La covariance de X, Y est définie par

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Notons que $\mathbb{E}[XY]$ existe, en effet X et Y ont des variances finies, donc elles appartiennent à $L^2(\Omega)$. Comme $|XY| \leq \frac{1}{2}X^2 + \frac{1}{2}Y^2$ d'où $XY \in L^1(\Omega)$.

Remarque : $\text{Cov}(X, X) = \text{Var}(X)$.

Théorème 20

Soient X et Y deux v.a. réelles. Si X et Y sont indépendantes, alors $\text{Cov}(X, Y) = 0$.

Démonstration : X et Y indépendants impliquent que $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, d'où le résultat □.

Le sens opposé n'est pas vrai toujours ; en effet soient X et Y , deux v.a. indépendantes, chacune de moments d'ordres un et deux finis t.q. $\mathbb{E}[Y] = 0$. Alors on a $\text{Cov}(X, XY) = \mathbb{E}[X^2 Y] = 0$ qui est nul, alors que X et XY sont dépendants.

Définition 21

Soient X et Y deux v.a. chacune admettant une variance finie. Le coefficient de corrélation de X et Y est le nombre

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)},$$

où $\sigma(X) = \sqrt{\text{Var}(X)}$ et $\sigma(Y) = \sqrt{\text{Var}(Y)}$.

Remarque : Par l'inégalité de Cauchy-Schwarz (i.e. $|E[XY]| \leq (E[X^2]E[Y^2])^{1/2}$) on a toujours $-1 \leq \rho \leq 1$ et X et Y indépendants impliquent que $\rho = 0$ par le Théorème 20.

Définition 22 (espérance d'un vecteur aléatoire)

Soit X un vecteur aléatoire de dimension n , on définit l'espérance de X par

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \mathbb{E}[X_2], \dots, \mathbb{E}[X_n]) \in \mathbb{R}^n.$$

Définition 23 (la matrice de covariance)

Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire de dimension n t.q. pour tout i , $X_i \in L^2(\Omega)$. La matrice de covariance de X , notée par $\mathbb{V}(X)$ est définie comme étant une matrice de taille $n \times n$ de terme général

$$\mathbb{V}(X)_{ij} = \text{Cov}(X_i, X_j).$$

Théorème 24

La matrice des variances-covariances (ou la matrice de covariance) est symétrique et semi-définie positive, i.e. $\mathbb{V}(X)_{ij} = \mathbb{V}(X)_{ji}$ pour tous i, j et $(a, \mathbb{V}(X)a)_{\mathbb{R}^n} \geq 0$ pour tout $a = (a_1, \dots, a_n) \in \mathbb{R}^n$.

Démonstration : Comme pour tous i, j , $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$ d'où $\mathbb{V}(X)$ est symétrique. Soit $a \in \mathbb{R}^n$. On a

$$\begin{aligned}(a, \mathbb{V}(X)a)_{\mathbb{R}^n} &= \sum_{1 \leq i, j \leq n} a_i a_j \text{Cov}(X_i, X_j) \\ &= \sum_{1 \leq i, j \leq n} a_i a_j \mathbb{E}[X_i X_j] - \sum_{1 \leq i, j \leq n} a_i a_j \mathbb{E}[X_i] \mathbb{E}[X_j] \\ &= \mathbb{E} \left[\left(\sum_{i=1}^n a_i X_i \right)^2 \right] - \left(\sum_{i=1}^n a_i \mathbb{E}[X_i] \right)^2 \\ &= \mathbb{E}[(a, X)_{\mathbb{R}^n}^2] - \mathbb{E}[(a, X)_{\mathbb{R}^n}]^2 = \text{Var}((a, X)_{\mathbb{R}^n}) \geq 0 \quad \square.\end{aligned}$$

Théorème 25

$X = (X_1, \dots, X_n)$ un vecteur aléatoire si et seulement si pour tout $1 \leq i \leq n$, X_i est une variable aléatoire.

Théorème 1

Soit X un vecteur aléatoire de dimension n de matrice de covariance $\mathbb{V}(X)$. Soit A une matrice de taille $m \times n$ et posons $Y = AX$. Alors Y est un vecteur aléatoire et sa matrice de covariance est $\mathbb{V}(Y) = ACA'$ où A' est désigne la matrice transposée de A .

Démonstration : Pour montrer que Y est un vecteur aléatoire on utilise le Théorème 25 en remarquant que $Y = (g_1(X), \dots, g_m(X))$ où $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ avec $g_i(x) = \sum_{j=1}^n A_{ij}x_j$ qui est une fonction continue pour tout $i \in \{1, \dots, m\}$ et comme X est un vecteur aléatoire d'où $g_i(X)$ est une variable aléatoire.

Pour tous $i, j \in \{1, \dots, m\}$, $\mathbb{V}(Y)_{ij} = \mathbb{E}[g_i(X)g_j(X)] - \mathbb{E}[g_i(X)]\mathbb{E}[g_j(X)]$ d'où en développant et simplifiant on trouve $\mathbb{V}(Y)_{ij} = \sum_{1 \leq p, q \leq n} A_{ip}A_{iq} \text{Cov}(X_p, X_q) = (A\mathbb{V}(X)A')_{ij}$, i.e.

$$\mathbb{V}(AX) = A\mathbb{V}(X)A'$$

□

Définition 26

Soit X un vecteur aléatoire de dimension n admettant une fonction de densité f_X et $g : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction positive alors on définit l'espérance $\mathbb{E}[g(X)]$ de la v.a. $g(X)$ par la formule

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^n} g(x) f_X(x) dx$$

qui est toujours une quantité bien définie même si elle peut prendre la valeur $+\infty$. Si g est de signe quelconque et $\mathbb{E}[|g(X)|] < \infty$ alors on dit que $g(X)$ est intégrable et on peut définir l'espérance de $g(X)$ par la formule ci-dessus. Si $g(X)$ n'est pas intégrable l'intégrale dans la formule ci-dessus n'est pas bien définie.

Définition 27

Soit X un vecteur aléatoire discret de dimension n , notons m_X sa fonction de masse et B_X l'ensemble dénombrable de $\mathcal{B}_{\mathbb{R}^n}$ t.q. $\mu_X(X \in B_X) = 1$. Soit g une fonction t.q. $\sum_{x \in B_X} |g(x)| < +\infty$, alors on définit $\mathbb{E}[g(X)]$ par

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

Lorsque vous avez une fonction $g: \mathbb{R}^n \rightarrow \mathbb{R}$ on peut utiliser la formule de transformation Jacobienne (Théorème 17), on arrive à déterminer f_Y grâce à la fonction de répartition de Y , F_Y où $Y = g(X)$ avec X admettant une fonction de densité :

On a

$$F_Y(y) = \mathbb{P}(Y \leq y) = \int_{g(x_1, \dots, x_n) \leq y} f_X(x) dx.$$

Supposons qu'il existe une fonction $h: \mathbb{R}^n \rightarrow \mathbb{R}$ t.q.

$$g(x_1, x_2, \dots, x_n) \leq y \Rightarrow x_1 \leq h(y, x_2, \dots, x_n).$$

Alors

$$\begin{aligned} F_Y(y) &= \int_{\{x \in \mathbb{R}^n : g(x) \leq y\}} f_X(x) dx \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\int_{-\infty}^{h(y, x_2, \dots, x_n)} f_X(x_1, x_2, \dots, x_n) \right) dx_2 \dots dx_n. \end{aligned}$$

En dérivant par rapport à y et en assumant que h est C^1 en y et que f_X est continue, alors on obtient

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{\partial h(y, x_2, \dots, x_n)}{\partial y} f_X(h(y, x_2, \dots, x_n), x_2, \dots, x_n) dx_2 \dots dx_n. \end{aligned}$$

Théorème 28

Soit X un vecteur aléatoire de dimension n admettant une densité f_X . Alors on a la formule suivante

$$f_X(t_1, \dots, t_n) = \frac{\partial^n}{\partial t_1 \dots \partial t_n} F_X(t_1, \dots, t_n)$$

pour tout point de continuité (t_1, \dots, t_n) de $\frac{\partial^n}{\partial x_1 \dots \partial x_n} F_X$.

Définition 29

Soit X un vecteur aléatoire de dimension n , on dit que X_1, X_2, \dots, X_n sont **mutuellement indépendants** si pour tout $p \in \mathbb{N}$ t.q. $2 \leq p \leq n$ et pour toute collection de p boréliens $A_{i_1}, \dots, A_{i_p} \in \mathcal{B}_{\mathbb{R}^n}$, on a

$$\mathbb{P}(\cap_{k=1}^p \{X_{i_k} \in A_{i_k}\}) = \prod_{k=1}^p \mathbb{P}(\{X_{i_k} \in A_{i_k}\}).$$

En s'appuyant sur les fonctions caractéristiques on obtient que celle-ci est équivalente à

$$\forall t_1, t_2, \dots, t_n \in \mathbb{R}, F_X(t_1, \dots, t_n) = \prod_{i=1}^n F_{X_i}(t_i).$$

En dérivant par rapport à chacune des variables, on obtient $f_X(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$ où les f_{X_i} , $i \in \{1, \dots, n\}$ sont les densités marginales des v.a. X_i .

Définition 30

Soit X un vecteur aléatoire de dimension n . Sa fonction caractéristique $\phi_X : \mathbb{R}^n \rightarrow \mathbb{R}$ est définie par

$$\phi_X(u) = \mathbb{E}[e^{i(u, X)_{\mathbb{R}^n}}], \forall u \in \mathbb{R}^n.$$

Théorème 2

Soit X un vecteur aléatoire de dimension n . Supposons qu'il existe un entier $m \in \mathbb{N}^*$ t.q. $X^m \in L^1(\Omega)$. Alors la fonction caractéristique ϕ_X de X admet des dérivées partielles d'ordre jusqu'à m , et

$$\frac{\partial^m}{\partial x_{j_1} \cdots \partial x_{j_m}} \phi_X(u) = i^m \mathbb{E} \left[e^{i(u, X)_{\mathbb{R}^n}} \prod_{k=1}^m X_{j_k} \right].$$

Théorème 3

Soit X un vecteur aléatoire de dimension n et $a \in \mathbb{R}^m$. Soit A une matrice de taille $m \times n$. Alors

$$\phi_{a+AX}(u) = e^{i(u, a)_{\mathbb{R}^m}} \phi_X(A'u), \forall u \in \mathbb{R}^m.$$

Voici une application immédiate du Théorème 2 qui permet de calculer les moments d'une variable aléatoire ne utilisant sa fonction caractéristique.

Rappel : k -ième moment d'une variable aléatoire X est $\mathbb{E}[X^k]$.

Pour les deux premiers moments, on obtient :

$$\begin{aligned}\mathbb{E}[X] &= -i\phi'_X(0) \text{ si } \mathbb{E}[|X|] < +\infty \\ \mathbb{E}[X^2] &= -\phi''_X(0) \text{ si } \mathbb{E}[X^2] < +\infty\end{aligned}$$

Exemples :

- Soit X une variable aléatoire de type Bernoulli de paramètre p , notée par $X \sim \mathcal{B}(p)$, alors

$$\phi_X(u) = \mathbb{E}[e^{iuX}] = e^{iu0}(1-p) + e^{iu}p = pe^{iu} + 1-p.$$

- Pour $X \sim \mathcal{B}(n, p)$, on a

$$\phi_X(u) = \mathbb{E}[e^{iuX}] = \sum_{j=0}^n \binom{n}{j} e^{iuj} p^j (1-p)^{n-j} = (pe^{iu} + 1-p)^n.$$

On peut aussi passer par le fait que X s'écrit comme une somme de n variables aléatoires indépendantes $Y_j, j = 1, \dots, n$, de loi identique $\mathcal{B}(p)$, d'où

$$X = \sum_{j=1}^n Y_j,$$

d'où

$$\phi_X(u) = \mathbb{E}[e^{iuX}] = \mathbb{E}[e^{iu\sum_{j=1}^n Y_j}] = \mathbb{E}\left[\prod_{j=1}^n e^{iuY_j}\right] = \prod_{j=1}^n \mathbb{E}[e^{iuY_j}]$$

par indépendance des Y_j ;

$$\phi_X(u) = \prod_{j=1}^n \phi_{Y_j}(u) = (pe^{iu} + 1 - p)^n.$$

• Soit X une v.a. de loi de Poisson de paramètre λ , notée par $X \sim \lambda$, d'où

$$\begin{aligned}\phi_X(u) &= \mathbb{E}[e^{iuX}] = \sum_{k=0}^{+\infty} e^{iuk} \mathbb{P}(X = k) \\ &= \sum_{k=0}^{+\infty} e^{iuk} \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^{+\infty} \frac{(\lambda e^{iu})^k}{k!} e^{-\lambda} \\ &= e^{-\lambda} e^{\lambda e^{iu}} = e^{\lambda(e^{iu} - 1)}.\end{aligned}$$

- Soit X une v.a. de loi uniforme sur $] -a, a[$, notée par $X \sim \mathcal{U}(] -a, a[)$, d'où

$$\phi_X(u) = \mathbb{E}[e^{iuX}] = \frac{1}{2a} \int_{-a}^a e^{iux} dx = \frac{e^{iua} - e^{-iua}}{2ai u};$$

en utilisant le fait que pour $z \in \mathbb{R}$, $e^{iz} = \cos z + i \sin z$, et $\cos(a) = \cos(-a)$, on obtient

$$\phi_X(u) = \frac{2i \sin(au)}{2ai u} = \frac{\sin(au)}{au}.$$

Exercice en classe : calculer $\phi_X(u)$ pour $X \sim \mathcal{N}(0,1)$:

1. Écrire $\phi_X(u)$ en version $a + ib$.
2. Montrer que $\phi_X(u) = \Re(\phi_X(u))$.
3. Montrer que $\phi_X'(u) = -u\phi_X(u)$.
4. Résoudre l'équation différentielle ci-dessus.

Théorème 4

Si deux fonctions de densité μ_1 et μ_2 deux mesures de probabilités sur \mathbb{R}^n ont les mêmes transformées de Fourier, i.e. $\int_{\mathbb{R}^n} e^{i(u,x)_{\mathbb{R}^n}} \mu_1(dx) = \int_{\mathbb{R}^n} e^{i(u,x)_{\mathbb{R}^n}} \mu_2(dx)$ alors elles sont égales presque partout.

Corollaire 2

Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire de dimension n . Alors les X_1, X_2, \dots, X_n sont indépendantes si et seulement si

$$\phi_X(u_1, \dots, u_n) = \prod_{j=1}^n \phi_{X_j}(u_j), \quad \forall u_1, \dots, u_n \in \mathbb{R}.$$

Exercices en classe :

- montrer que $\phi_X(u, 0, \dots, 0) = \phi_{X_1}(u)$ pour tout $u \in \mathbb{R}$.
- montrer que $\phi_X(u, u, \dots, u) = \phi_{X_1 + \dots + X_n}(u)$ pour tout $u \in \mathbb{R}$.
- montrer que si X et Y sont des v.a.r. i.i.d. alors $Z = X - Y$ a une fonction caractéristique symétrique.
- montrer que $\overline{\phi_X(u)} = \phi_X(-u)$ pour tout $u \in \mathbb{R}$ où $\bar{z} = a - ib$ pour $z = a + ib$ avec $a, b \in \mathbb{R}$.
- montrer que pour X_1, \dots, X_n v.a. indépendantes, d'espérances égales à 0, et de moments 3 finies, alors $\mathbb{E}[(X_1 + \dots + X_n)^3] = \mathbb{E}[X_1^3] + \dots + \mathbb{E}[X_n^3]$.
- montrer que si $X \sim \mathcal{N}(0, 1)$ alors pour tout $n \in \mathbb{N}^*$, $\mathbb{E}[X^{2n}] = \frac{(2n)!}{2^n n!} = \prod_{k=1}^n (2k-1)$, que vaut $\mathbb{E}[X^{2n+1}]$ pour tout $n \in \mathbb{N}$?

Définition 31

Un vecteur aléatoire $X = (X_1, \dots, X_n)$ de dimension n est dit Gaussien si toute combinaison linéaire $\sum_{j=1}^n a_j X_j$ avec $a_j \in \mathbb{R}$ pour tout $j \in \{1, \dots, n\}$, suit une loi Gaussienne.

Les fonctions caractéristiques aident beaucoup lors de manipulations de variables aléatoires Gaussiennes.

Théorème 32

Soit X un vecteur aléatoire de dimension n .

X est Gaussien si et seulement si sa fonction caractéristique a la forme

$$\phi_X(u) = e^{i\langle u, \mu \rangle - \frac{1}{2} \langle u, Qu \rangle}$$

où $\mu \in \mathbb{R}^n$ et Q est une matrice symétrique semi-définie positive (i.e. $Q^T = Q$ et $\langle x, Qx \rangle \geq 0$ pour tout $x \in \mathbb{R}^n$). Q est alors la matrice de covariance de X et μ est l'espérance de X , i.e. $\mu_j = \mathbb{E}[X_j]$ pour tout $j \in \{1, \dots, n\}$.

Démonstration : Le sens \Leftarrow : Soit $Y = \sum_{j=1}^n a_j X_j = \langle a, X \rangle$ une combinaison linéaire des composants de X . On a besoin de montrer que Y suit une loi normale unidimensionnelle. On a donc pour tout $v \in \mathbb{R}$:

$$\phi_Y(v) = \phi_X(va) = e^{iv\langle a, \mu \rangle - \frac{v^2}{2} \langle a, Qa \rangle}$$

et donc ϕ_Y est la fonction caractéristique d'une variable aléatoire suivant une loi normale $\mathcal{N}(\langle a, \mu \rangle, \langle a, Qa \rangle)$; et donc par le Théorème 4 (mêmes transformées de Fourier), Y suit une loi normale.

Le sens \Rightarrow : Supposons que X soit un vecteur aléatoire Gaussien, et posons

$$Y = \sum_{j=1}^n a_j X_j = \langle a, X \rangle$$

comme étant une combinaison linéaire des composants de X . Soit $Q = \mathbb{V}(X)$ la matrice de variances-covariances de X . Alors

$$\mathbb{E}[Y] = \langle a, \mu \rangle$$

où $\mu = (\mu_1, \dots, \mu_n)$ et $\mathbb{E}[X_i] = \mu_i$ pour tout $i \in \{1, \dots, n\}$; et donc

$$\sigma^2(Y) = \langle a, Qa \rangle,$$

par le Théorème 24.

Comme Y suit une loi normale par hypothèse, d'où avec l'espérance et la variance de Y , on obtient

$$\phi_Y(v) = e^{iv\langle a, \mu \rangle - \frac{v^2}{2} \langle a, Qa \rangle}.$$

Alors

$$\phi_Y(\mathbf{1}) = \phi_{\langle a, X \rangle}(\mathbf{1}) = \mathbb{E}[e^{i\langle a, X \rangle}] = \phi_X(a),$$

et nous obtenons

$$\phi_X(u) = e^{i\langle u, \mu \rangle - \frac{1}{2} \langle u, Qu \rangle}.$$

□

Notation : Lorsque X est comme dans le théorème précédent, on note par $\mathcal{N}(\mu, Q)$ sa loi.

Exemple : Soit X_1, \dots, X_n des variables aléatoires indépendantes de loi respectives $\mathcal{N}(\mu_1, \sigma_1^2), \dots, \mathcal{N}(\mu_n, \sigma_n^2)$. Alors $X = (X_1, \dots, X_n)$ est Gaussien. En effet

$$\phi_X(u_1, \dots, u_n) = \prod_{j=1}^n \phi_{X_j}(u_j)$$

par le Corollaire 2 ; et donc

$$\begin{aligned} \phi_X(u) &= \prod_{j=1}^n e^{iu_j \mu_j - \frac{1}{2} u_j^2 \sigma_j^2} \\ &= e^{\sum_{j=1}^n iu_j \mu_j - \frac{1}{2} \sum_{j=1}^n u_j^2 \sigma_j^2} \\ &= e^{i\langle u, \mu \rangle - \frac{1}{2} \langle u, Qu \rangle} \end{aligned}$$

où $\mu = (\mu_1, \dots, \mu_n)$ et Q est la matrice diagonale

$$\begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \dots & \\ 0 & & & \sigma_n^2 \end{pmatrix}$$

Comme $\phi_X(u)$ est de la forme $e^{i\langle u, \mu \rangle - \frac{1}{2} \langle u, Qu \rangle}$, nous savons que X suit une loi normale $\mathcal{N}(\mu, Q)$.

Corollaire 3

Soit X un vecteur aléatoire Gaussien de dimension n . Les composants X_j de X sont indépendants si et seulement si la matrice de covariance Q de X est diagonale.

Théorème 33

Soit X un vecteur aléatoire Gaussien. Les composants X_j et X_k de X sont indépendants si et seulement si X_k et X_j sont non-corrélés.

Démonstration : Voir l'exercice ci-dessus pour la condition nécessaire. Quant à la condition de suffisance :

Supposons que Q est diagonale, i.e. $Q = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$. Alors par le théorème précédent on obtient $\phi_X(u) = \prod_{j=1}^n \phi_{X_j}(u_j)$, où $\phi_{X_j}(u_j) = e^{iu_j\mu_j - \frac{1}{2}u_j^2\sigma_j^2}$. Le corollaire précédent donne alors : les X_j sont indépendants où chacun des X_j suit la loi $\mathcal{N}(\mu_j, \sigma_j^2)$.

Théorème 34

Soit X un vecteur aléatoire Gaussien de dimension n , d'espérance μ et de matrice de covariance Q . Alors il existe n variables aléatoires Y_1, \dots, Y_n telles que

$$Y_j \sim \mathcal{N}(0, \lambda_j), \lambda_j \geq 0, \forall j \in \{1, \dots, n\},$$

et une matrice orthogonale A telle que $X = \mu + AY$.

Démonstration : Comme Q est une matrice de covariance, donc elle est symétrique et semi-définie positive, par conséquent il existe toujours une matrice orthogonale A telle que $Q = A\Lambda A^*$ où Λ est une matrice diagonale d'éléments positifs, avec A^* la matrice transposée de A . Comme A est orthogonale, d'où $A^* = A^{-1}$.

On pose $Y = A^*(X - \mu)$ où $\mu_j = \mathbb{E}[X_j]$ pour tout j . Comme X est un vecteur Gaussien par hypothèse, alors Y l'est également (toute combinaison linéaire de Y_j est une combinaison linéaire de X_j , donc suit une loi gaussienne uni-dimensionnelle). De plus, la matrice de covariance de Y est $A^*QA = \Lambda$, comme $X = \mu + AY$ le théorème est donc prouvé.

Corollaire 4

Un vecteur aléatoire Gaussien X a une densité sur \mathbb{R}^n si et seulement si la matrice de covariance Q de X est non-dégénérée (i.e. il n'existe pas de vecteur $a \in \mathbb{R}^n - \{0\}$ t.q. $Qa = 0$, ou de façon équivalent $\det(Q) \neq 0$).

Démonstration : Par le théorème précédent, on sait qu'il existe n variables indépendantes, toutes suivant respectivement une loi normale $\mathcal{N}(0, \lambda_j)$, ($1 \leq j \leq n$), avec $Q = A\Lambda A^*$, pour A orthogonale. Si $\det(Q) \neq 0$, alors $\lambda_j > 0$, pour tout $j \in \{1, \dots, n\}$, car $\det(Q) = \det(\Lambda) = \prod_{j=1}^n \lambda_j$. Comme $\lambda_j > 0$ et $Y_j \sim \mathcal{N}(0, \lambda_j)$, on sait donc que Y a une fonction de densité, donnée par

$$f_Y(y) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi\lambda_j}} e^{-\frac{1}{2\lambda_j}y_j^2},$$

et comme $X = \mu + AY$, on déduit grâce au théorème de changement de variables avec la transformation Jacobienne, que X admet une fonction de densité, donnée par

$$f_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det(Q)}} e^{-\frac{1}{2}\langle x-\mu, Q^{-1}(x-\mu) \rangle}$$

car H est un hyperplan, donc l'avant dernière égalité ci-dessus n'a pas lieu ; d'où X n'admet pas de fonction de densité.

Supposons que Q est dégénérée, i.e. $\det(Q) = 0$. Alors il existe $a \in \mathbb{R}^n$, $a \neq 0$ t.q. $Qa = 0$. La variable aléatoire $Z = \langle a, X \rangle$ a une variance égale à $\langle a, Qa \rangle = 0$, donc Z est presque-sûrement égale à l'espérance $\langle a, \mu \rangle$. Par conséquent $\mathbb{P}(X \in H) = 1$, où H est un hyperplan orthogonal au vecteur a et contenant le vecteur μ , i.e.

$H = \{x \in \mathbb{R}^n : \langle x - \mu, a \rangle = 0\}$. Comme la dimension de H est $n - 1$, la mesure de Lebesgue sur \mathbb{R}^n de H vaut 0. Si X avait une fonction de densité, nous aurions la propriété suivante satisfaite :

$$1 = \mathbb{P}(X \in H) = \int_H f_X(x) dx = \int_H f_X(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Alors que

$$\int \mathbf{1}_H(x_1, \dots, x_n) dx_n = 0.$$

Soit $X \sim \mathcal{N}(0,1)$ et $a > 0$. On pose $Y^a := X\mathbf{1}_{\{|X| < a\}} - X\mathbf{1}_{\{|X| \geq a\}}$.

1. Montrer que Y^a est une variable aléatoire gaussienne et qu'il existe un $b > 0$ tel que

$$\int_0^b x^2 e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} = \frac{1}{4}.$$

2. Calculer la covariance de X et Y^b . Le couple (X, Y^b) forme-t-il un vecteur gaussien ?

1. Soit $u \in \mathbb{R}$, alors $\mathbb{E}[e^{uY^a}] = \int_{\mathbb{R}} e^{u(x\mathbf{1}_{\{|x|<a\}} - x\mathbf{1}_{\{|x|\geq a\}})} f(x) dx$ où $f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$.

Donc $\mathbb{E}[e^{uY^a}] = \int_{-\infty}^{-a} e^{-ux} f(x) dx + \int_{-a}^a e^{ux} f(x) dx + \int_a^{+\infty} e^{-ux} f(x) dx$, en simplifiant il vient

$$\mathbb{E}[e^{uY^a}] = \int_{-\infty}^{-a} e^{-ux} f(x) dx + \int_{-\infty}^a e^{ux} f(x) dx = e^{u^2/2} (\phi_X(u-a) + \phi(-u+a)).$$

Comme $\phi(x) + \phi(-x) = 1$, on a donc $\mathbb{E}[e^{uY^a}] = e^{u^2/2}$, i.e. $Y^a \sim \mathcal{N}(0,1)$.

2. En posant $F(b) = \int_0^b x^2 e^{-x^2/2} \frac{dx}{\sqrt{2\pi}}$, F est continue car $x \mapsto x^2 e^{-x^2/2}$ est continue sur $[0, b]$.

D'où avec $F(+\infty) - \frac{1}{4} = \frac{1}{2} - \frac{1}{4} = \frac{1}{4} > 0$ et $F(0) - \frac{1}{4} = -\frac{1}{4} < 0$, on obtient l'existence d'un $b > 0$ tel que $F(b) = \frac{1}{4}$ par le théorème des valeurs intermédiaires.

$$\text{Cov}(X, Y^b) = \mathbb{E}[XY^b] = \mathbb{E}[X^2(\mathbf{1}_{\{|X|<b\}} - \mathbf{1}_{\{|X|\geq b\}})]] = 2 \int_{-b}^b x^2 e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} - 1 = 4F(b) - 1 = 0.$$

Comme $|X| = |Y^b|$ d'où X et Y^b ne sont pas indépendantes, ce qui implique que (X, Y^b) n'est pas un vecteur gaussien.