

Cours de Probabilités : Modèles et Applications.

Anne PHILIPPE¹ & Marie-Claude VIANO²

Niveau Master 1

Université de Nantes

Année 2009-2010

1. Anne.Philippe@univ-nantes.fr
2. Marie-Claude.Viano@univ-lille1.fr

Table des matières

| | |
|-------------------------------------------------------------------------|----|
| Chapitre 1. Conditionnement | 4 |
| 1. Conditionnement par un événement | 4 |
| 2. Conditionnement par une partition au plus dénombrable d'événements | 6 |
| 3. Conditionnement par une tribu | 10 |
| Annexe : Chaînes de Markov à espace d'états au plus dénombrable | 16 |
| Annexe : Démonstration de la Proposition 1.10 | 19 |
| Chapitre 2. Vecteurs Gaussiens | 21 |
| 1. Rappels : vecteurs aléatoires, covariance, fonction caractéristique. | 21 |
| 2. Lois gaussiennes dans \mathbb{R} | 22 |
| 3. Vecteurs Gaussiens | 23 |
| 4. Vecteur gaussien et indépendance | 24 |
| 5. Existence | 26 |
| 6. Densité d'un vecteur gaussien | 26 |
| 7. Vecteur gaussien et conditionnement | 27 |
| 8. Projection d'un vecteur gaussien | 29 |
| Chapitre 3. Convergences | 32 |
| 1. Définition des modes de convergence | 32 |
| 2. Quelques propriétés | 32 |
| 3. Compléments sur la convergence presque sûre | 33 |
| 4. Théorème de Borel-Cantelli. | 34 |
| Chapitre 4. Sommes de variables aléatoires | 36 |
| 1. Inégalité de Kolmogorov | 36 |
| 2. Loi Forte des Grands Nombres | 37 |
| 3. Théorème de Glivenko-Cantelli | 40 |
| Chapitre 5. Convergence en loi | 44 |
| 1. Définitions | 44 |
| 2. Convergence en loi et convergence en probabilité | 45 |
| 3. Propriétés de la convergence en loi | 46 |
| 4. Convergence en loi et fonctions caractéristiques | 47 |
| 5. Théorème central limite | 48 |
| 6. Théorème de Cramér | 50 |
| 7. Théorème de Lindberg-Feller | 53 |

Conditionnement

Dans ce chapitre, $(\Omega, \mathcal{A}, \mathbf{P})$ désigne un espace probabilisé.

1. Conditionnement par un événement

On se donne B un événement (*i.e.* $B \in \mathcal{A}$) tel que $\mathbf{P}(B) > 0$. Pour tout $A \in \mathcal{A}$, on pose

$$Q_B(A) = \frac{\mathbf{P}(B \cap A)}{\mathbf{P}(B)}.$$

Exercice 1. Montrer que Q_B définit une probabilité sur (Ω, \mathcal{A})

△△

Définition 1.1. Q_B est la mesure de probabilité sachant l'événement B . On la note souvent $\mathbf{P}(\cdot|B)$

Proposition 1.1. Soit X une variable aléatoire.

Si X est \mathbf{P} -intégrable *i.e.* $\mathbb{E}|X| := \int |X| d\mathbf{P} < \infty$ alors X est aussi $\mathbf{P}(\cdot|B)$ -intégrable. De plus, on a

$$\mathbb{E}(X|B) \stackrel{\text{def}}{=} \int X d\mathbf{P}(\cdot|B) = \frac{1}{\mathbf{P}(B)} \int_B X d\mathbf{P} = \frac{1}{\mathbf{P}(B)} \mathbb{E}(X \mathbb{I}_B). \quad (1.1)$$

$\mathbb{E}(X|B)$ est l'espérance de X sachant (ou conditionnelle à) l'événement B .

DÉMONSTRATION. Ce résultat se démontre par étapes, en commençant par une fonction indicatrice, ce qui permet de passer aux fonctions étagées puis aux fonctions mesurables positives, pour conclure sur les fonctions intégrables.

Étape 1 : $X = \mathbb{I}_A$:

$$\begin{aligned} \int \mathbb{I}_A d\mathbf{P}(\cdot|B) &= \mathbf{P}(A|B) = \frac{\mathbf{P}(B \cap A)}{\mathbf{P}(B)} \\ &= \frac{1}{\mathbf{P}(B)} \int \mathbb{I}_{B \cap A} d\mathbf{P} = \frac{1}{\mathbf{P}(B)} \int_B \mathbb{I}_A d\mathbf{P} \end{aligned}$$

Étape 2 : $X = \sum_{j \in J} \lambda_j \mathbb{I}_{A_j}$ où J est un ensemble fini et les (A_j) sont des événements disjoints.

On obtient la relation (1.1) grâce à la linéarité de l'intégrale et l'étape précédente.

Étape 3 : X est mesurable positive. Il existe une suite croissante de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ étagées positives qui converge vers X . D'après l'étape 2., on a

$$\forall n \in \mathbb{N} \quad \int X_n d\mathbf{P}(\cdot|B) = \frac{1}{\mathbf{P}(B)} \int_B X_n d\mathbf{P}.$$

On applique ensuite le théorème de Beppo Levi aux deux intégrales.

Étape 4 : X intégrable. La variable aléatoire X se décompose sous la forme $X = X^+ - X^-$ avec

$$\begin{cases} X^+ = \max(X, 0) \\ X^- = \max(-X, 0) \end{cases}$$

Les variables aléatoires X^+ et X^- sont positives et \mathbf{P} -intégrable. L'étape 3 assure que

$$\int X^+ d\mathbf{P}(\cdot|B) = \frac{1}{\mathbf{P}(B)} \int_B X^+ d\mathbf{P} < +\infty$$

et

$$\int X^- d\mathbf{P}(\cdot|B) = \frac{1}{\mathbf{P}(B)} \int_B X^- d\mathbf{P} < +\infty$$

donc $|X| = X^+ + X^-$ est $\mathbf{P}(\cdot|B)$ -intégrable et on obtient (1.1). □

Pour toute variable aléatoire X (admettons qu'elle est à valeurs dans \mathbb{R} , mais elle pourrait aussi être à valeurs dans un autre espace) ce qui précède définit aussi la probabilité, sachant B , que X appartienne à n'importe quel borélien de \mathbb{R} . Autrement dit on construit facilement la loi conditionnelle de X et, par le théorème de transfert, l'espérance conditionnelle se calcule aussi en utilisant cette loi :

Définition 1.2. La loi de X sachant (ou conditionnelle à) l'événement B est la transportée de $\mathbf{P}(\cdot|B)$ par X . Soit, pour tout $E \in \mathcal{B}(\mathbb{R})$

$$\mathbf{P}_X(E|B) = \mathbf{P}(X^{-1}(E)|B) = \mathbf{P}(X \in E|B)$$

et on a

$$\mathbb{E}(X|B) = \int_{\mathbb{R}} x d\mathbf{P}_X(x|B).$$

Exemple 1.1. Soit (X_1, X_2) un couple de variables aléatoires indépendantes et de même loi définie par

$$\mathbf{P}(X_1 = k) = \alpha^k(1 - \alpha) \quad \forall k \in \mathbb{N}$$

où α est fixé dans $]0, 1[$. (On note $\tilde{G}(\alpha)$ cette loi)

On pose

$$N = \begin{cases} 1 & \text{si } X_1 \leq X_2 \\ 2 & \text{sinon} \end{cases}$$

On va déterminer la loi de X_1 sachant $N = 1$.

Commençons par calculer la loi de la variable aléatoire N . La loi du couple (X_1, X_2) étant symétrique, on a

$$\mathbf{P}(X_1 < X_2) = 1/2(1 - \mathbf{P}(X_1 = X_2)).$$

On en déduit

$$\begin{aligned} \mathbf{P}(N = 1) &= \mathbf{P}(X_1 = X_2) + \mathbf{P}(X_1 < X_2) = \frac{1}{2}(1 + \mathbf{P}(X_1 = X_2)) \\ &= \frac{1}{2} \left(1 + \sum_{k=0}^{\infty} \mathbf{P}(X_1 = k, X_2 = k) \right) = \frac{1}{2} \left(1 + \sum_{k=0}^{\infty} \mathbf{P}(X_1 = k) \mathbf{P}(X_2 = k) \right) \end{aligned}$$

car les variables aléatoires sont indépendantes,

$$\begin{aligned} \mathbf{P}(N = 1) &= \frac{1}{2} \left(1 + (1 - \alpha)^2 \sum_{k=0}^{\infty} \alpha^{2k} \right) \\ &= \frac{1}{2} \left(1 + \frac{(1 - \alpha)^2}{1 - \alpha^2} \right) = \frac{1}{1 + \alpha} \end{aligned}$$

La loi conditionnelle de X_1 sachant l'événement $[N = 1]$ est donné par

$$\begin{aligned} \mathbf{P}(X_1 = k|N = 1) &= \frac{\mathbf{P}(N = 1, X_1 = k)}{\mathbf{P}(N = 1)} = \frac{\mathbf{P}(X_2 \geq k, X_1 = k)}{\mathbf{P}(N = 1)} \\ &= (1 + \alpha) \mathbf{P}(X_1 = k) \sum_{\ell \geq k} \mathbf{P}(X_2 = \ell) \\ &= (1 + \alpha) \alpha^k (1 - \alpha) \sum_{\ell \geq k} (1 - \alpha) \alpha^\ell \\ &= (1 - \alpha^2) \alpha^{2k}, \quad \forall k \in \mathbb{N}. \end{aligned}$$

La loi conditionnelle de X_1 sachant l'événement $[N = 1]$ est une loi $\tilde{G}(\alpha^2)$. On en déduit que l'espérance de X_1 sachant l'événement $[N = 1]$ est égale à

$$\mathbb{E}(X_1|N = 1) = \sum_{k=0}^{\infty} k \mathbf{P}(X_1 = k|N = 1) = \frac{\alpha^2}{1 - \alpha^2}$$

De la même façon, on obtient la loi de X_1 sachant $[N = 2]$

$$\mathbf{P}(X_1 = k|N = 2) = (1 - \alpha^2)\alpha^{k-1}(1 - \alpha^k) \quad \forall k \geq 1$$

et l'espérance conditionnelle de X_1 sachant $N = 2$ est égale à $\frac{1 + \alpha + \alpha^2}{1 - \alpha^2}$.

Remarque :

L'espérance de X_1 s'exprime à partir des espérances conditionnelles $\mathbb{E}(X_1|N = 1)$ et $\mathbb{E}(X_1|N = 2)$. En effet, on peut décomposer X_1 en

$$X_1 = X_1 \mathbb{I}_{N=1} + X_1 \mathbb{I}_{N=2}$$

et grâce à (1.1), on obtient

$$\begin{aligned} \mathbb{E}X_1 &= \mathbb{E}(X_1 \mathbb{I}_{N=1}) + \mathbb{E}(X_1 \mathbb{I}_{N=2}) \\ &= \mathbf{P}(N = 1)\mathbb{E}(X_1|N = 1) + \mathbf{P}(N = 2)\mathbb{E}(X_1|N = 2) \\ &= \frac{\alpha}{1 - \alpha} \end{aligned}$$

||

2. Conditionnement par une partition au plus dénombrable d'événements

Dans cette partie, $(B_n)_{n \in \mathbb{N}}$ désigne une partition de Ω telle que $\mathbf{P}(B_n) \neq 0$ pour tout n . On note \mathcal{B} la tribu engendrée par cette partition $\mathcal{B} = \sigma(B_n, n \in \mathbb{N})$.

2.1. Définition et propriétés de la probabilité conditionnelle. Soit A un événement. On définit la probabilité conditionnelle de A sachant \mathcal{B} comme étant la variable aléatoire qui prend la valeur $\mathbf{P}(A|B_i)$ pour tout $\omega \in B_i$. On la note $\mathbf{P}(A|\mathcal{B})$. On a donc pour tout $\omega \in \Omega$

$$\mathbf{P}(A|\mathcal{B})(\omega) = \sum_{n \in \mathbb{N}} \mathbf{P}(A|B_n) \mathbb{I}_{B_n}(\omega) \quad (1.2)$$

Proposition 1.2. Soit A fixé, la variable aléatoire $\mathbf{P}(A|\mathcal{B})$ satisfait les propriétés suivantes :

$$[\pi_1] : \begin{cases} \mathbf{P}(A|\mathcal{B}) \text{ est } \mathcal{B}\text{-mesurable} \\ \text{pour tout } B \in \mathcal{B}, \text{ on a } \int_B \mathbf{P}(A|\mathcal{B}) \, d\mathbf{P} = \mathbf{P}(A \cap B). \end{cases}$$

DÉMONSTRATION. En tant que fonction constante sur chaque B_i , la variable aléatoire $\mathbf{P}(A|\mathcal{B})$ est mesurable. Soit $B \in \mathcal{B}$. Il s'écrit de la forme $B = \cup_{i \in I_0} B_i$ et pour tout fonction \mathbf{P} -intégrable f , on a

$$\int_B f \, d\mathbf{P} = \sum_{i \in I_0} \int_{B_i} f \, d\mathbf{P}.$$

On obtient

$$\begin{aligned} \int_B \mathbf{P}(A|\mathcal{B}) \, d\mathbf{P} &= \sum_{i \in I_0} \int_{B_i} \mathbf{P}(A|\mathcal{B}) \, d\mathbf{P} \\ &= \sum_{i \in I_0} \int_{B_i} \sum_{j \in \mathbb{N}} \mathbf{P}(A|B_j) \mathbb{I}_{B_j} \, d\mathbf{P} = \sum_{i \in I_0} \sum_{j \in \mathbb{N}} \mathbf{P}(A|B_j) \int_{B_i \cap B_j} \, d\mathbf{P} \\ &= \sum_{i \in I_0} \mathbf{P}(A|B_i) \mathbf{P}(B_i) \quad \text{car } B_i \cap B_j = \emptyset \text{ si } j \neq i \text{ et } B_i \text{ sinon} \\ &= \mathbf{P}(A \cap \cup_{i \in I_0} B_i) = \mathbf{P}(A \cap B) \end{aligned}$$

□

Proposition 1.3. À ω fixé, $\mathbf{P}(\cdot|\mathcal{B})(\omega)$ est une mesure de probabilité.

DÉMONSTRATION. En effet, si on fixe $\omega \in \Omega$, il existe un unique élément de la partition, noté, B^ω tel que $\omega \in B^\omega$. On a alors

$$\mathbf{P}(\cdot|\mathcal{B})(\omega) = \sum_{j \in \mathbb{N}} \mathbf{P}(\cdot|B_j) \mathbb{I}_{B_j}(\omega) = \mathbf{P}(\cdot|B^\omega).$$

C'est bien une probabilité. □

Remarque 1.1. Supposons qu'il existe des éléments de la partition ayant une probabilité nulle. On note $I = \{i : \mathbf{P}(B_i) \neq 0\}$. La probabilité conditionnelle sachant \mathcal{B} est alors définie par

$$\mathbf{P}(A|\mathcal{B})(\omega) = \sum_{n \in I} \mathbf{P}(A|B_n) \mathbb{I}_{B_n}(\omega) \quad (1.3)$$

$\mathbf{P}(A|\mathcal{B})$ est nulle sur l'ensemble $\bigcup_{i \in I^c} B_i$. |

2.2. Définition et propriétés de l'espérance conditionnelle.

Définition 1.3. Soit X une variable aléatoire dans $L^1(\Omega, \mathcal{A}, \mathbf{P})$, l'espérance de X conditionnellement à \mathcal{B} est la variable aléatoire définie par

$$\mathbb{E}(X|\mathcal{B})(\omega) \stackrel{\text{not}}{=} \int X \, d\mathbf{P}(\cdot|\mathcal{B})(\omega) = \sum_{i \in \mathbb{N}} \mathbb{E}(X|B_i) \mathbb{I}_{B_i}(\omega) \quad (1.4)$$

pour tout $\omega \in \Omega$.

Remarque 1.2. Pour tout ω , c'est tout simplement

$$\mathbb{E}(X|\mathcal{B})(\omega) = \int X \, d\mathbf{P}(\cdot|B^\omega) = \mathbb{E}(X|B^\omega)$$

où B^ω est l'unique élément de la partition contenant ω . |

Proposition 1.4. Soit (X, Y) un couple de variables aléatoires intégrables par rapport à \mathbf{P} et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires \mathbf{P} -intégrables.

À ω fixé, l'espérance conditionnelle vérifie les propriétés suivantes

- (1) pour tout couple $(\lambda, \mu) \in \mathbb{R}^2$, $\mathbb{E}(\lambda X + \mu Y|\mathcal{B})(\omega) = \lambda \mathbb{E}(X|\mathcal{B})(\omega) + \mu \mathbb{E}(Y|\mathcal{B})(\omega)$
- (2) si $X \leq Y$ p.s. alors $\mathbb{E}(X|\mathcal{B})(\omega) \leq \mathbb{E}(Y|\mathcal{B})(\omega)$.
- (3) si $(X_n)_n$ est une suite croissante de variables aléatoires positives intégrables telle que $\lim_{n \rightarrow \infty} \uparrow X_n = X$ alors $\lim_{n \rightarrow \infty} \uparrow \mathbb{E}(X_n|\mathcal{B})(\omega) = \mathbb{E}(X|\mathcal{B})(\omega)$

DÉMONSTRATION. Ces propriétés se déduisent immédiatement des propriétés de l'espérance. En effet, à ω fixé $\mathbb{E}(\cdot|\mathcal{B})(\omega)$ est égale à l'espérance par rapport à la mesure de probabilité $\mathbf{P}(\cdot|B^\omega)$ où B^ω est l'unique élément de la partition contenant ω . □

Proposition 1.5. Soit X une variable aléatoire \mathbf{P} -intégrable.

- (1) L'application $\omega \mapsto \mathbb{E}(X|\mathcal{B})(\omega)$ satisfait

$$[\pi_2] : \begin{cases} \mathbb{E}(X|\mathcal{B}) \text{ est } \mathcal{B}\text{-mesurable} \\ \text{pour tout } B \in \mathcal{B}, \text{ on a } \int_B \mathbb{E}(X|\mathcal{B}) \, d\mathbf{P} = \int_B X \, d\mathbf{P}. \end{cases}$$

En particulier en prenant $B = \Omega$, on obtient

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

i.e. les variables aléatoires X et $\mathbb{E}(X|\mathcal{B})$ ont la même valeur moyenne.

- (2) Si Y et XY sont \mathbf{P} -intégrables et si X est \mathcal{B} -mesurable alors

$$\mathbb{E}(XY|\mathcal{B}) = X \mathbb{E}(Y|\mathcal{B})$$

- (3) Si X est indépendante de la tribu \mathcal{B} alors $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}(X)$

DÉMONSTRATION.

Preuve de (1). $\mathbb{E}(X|\mathcal{B})$ est \mathcal{B} -mesurable car elle est constante sur chaque B_i .

Pour tout $B \in \mathcal{B}$, on démontre l'égalité $\int_B E(X|\mathcal{B}) d\mathbf{P} = \int_B X d\mathbf{P}$ par la méthode standard.

Étape 1 : $X = \mathbb{I}_A$ avec $A \in \mathcal{A}$ le résultat est immédiat car on retrouve la relation $[\pi_1]$. Il suffit d'écrire la définition de $\mathbb{E}(X|\mathcal{B})$ i.e. $E(\mathbb{I}_A|\mathcal{B}) = \mathbf{P}(A|\mathcal{B})$.

Étape 2 : $X = \sum_{j \in J} \lambda_j \mathbb{I}_{A_j}$ où J est un sous ensemble fini. Le résultat découle de l'étape précédente en utilisant la linéarité de l'intégrale et de la linéarité de l'espérance conditionnelle.

Étape 3 : X est mesurable positive. Il existe une suite de variables aléatoires étagées positives telle que $\lim \uparrow X_n = X$. L'étape précédente assure que pour tout n , on a

$$\int_B E(X_n|\mathcal{B}) d\mathbf{P} = \int_B X_n d\mathbf{P}.$$

D'après (1)- (2) de la proposition 1.4, $E(X_n|\mathcal{B})$ est presque sûrement croissante positive. On peut donc appliquer le théorème de Beppo Levi, puis la propriété (3) de la proposition 1.4 pour obtenir le résultat.

Étape 4 : X est intégrable. On décompose $X = X^+ - X^-$ avec $\mathbb{E}X^\pm < \infty$. L'étape précédente assure que

$$\int_B E(X^\pm|\mathcal{B}) d\mathbf{P} = \int_B X^\pm d\mathbf{P}$$

pour tout $B \in \mathcal{B}$. On peut ensuite conclure en utilisant les propriétés de linéarité.

Preuve de (2). On a

$$\begin{aligned} \mathbb{E}(XY|\mathcal{B}) &= \sum_{i \in I} \mathbb{E}(XY|B_i) \mathbb{I}_{B_i} = \sum_{i \in I} \int XY d\mathbf{P}(\cdot|B_i) \mathbb{I}_{B_i} \\ &= \sum_{i \in I} \frac{1}{\mathbf{P}(B_i)} \int_{B_i} XY d\mathbf{P} \mathbb{I}_{B_i} \end{aligned}$$

Comme X est \mathcal{B} -mesurable donc X s'écrit de la forme $\sum_{i \in I} x_i \mathbb{I}_{B_i}$ avec $I \subset \mathbb{N}$. On obtient

$$\mathbb{E}(XY|\mathcal{B}) = \sum_{i \in \mathbb{N}} \frac{1}{\mathbf{P}(B_i)} \int_{B_i} XY d\mathbf{P} \mathbb{I}_{B_i} = \sum_{i \in \mathbb{N}} \frac{1}{\mathbf{P}(B_i)} \sum_{j \in \mathbb{N}} x_j \int_{B_i} \mathbb{I}_{B_j} Y d\mathbf{P} \mathbb{I}_{B_i}$$

or

$$\mathbb{I}_{B_j \cap B_i} = \begin{cases} \mathbb{I}_{B_j} & \text{si } i = j \\ 0 & \text{sinon.} \end{cases}$$

d'où

$$\mathbb{E}(XY|\mathcal{B}) = \sum_{i \in I} x_i \frac{1}{\mathbf{P}(B_i)} \int_{B_i} Y d\mathbf{P} \mathbb{I}_{B_i} = \sum_{i \in I} x_i \mathbb{E}(Y|B_i) \mathbb{I}_{B_i}.$$

D'autre part

$$X \mathbb{E}(Y|\mathcal{B}) = \sum_{i \in I} x_i \mathbb{I}_{B_i} \sum_{j \in I} \mathbb{E}(Y|B_j) \mathbb{I}_{B_j} = \sum_{i,j \in I} x_i \mathbb{E}(Y|B_j) \mathbb{I}_{B_j \cap B_i} = \sum_{i \in I} x_i \mathbb{E}(Y|B_i) \mathbb{I}_{B_i}$$

D'où l'égalité annoncée.

Preuve de (3). X étant indépendante de \mathcal{B} , elle est aussi indépendante des variables aléatoires \mathbb{I}_{B_j} . On a donc

$$\mathbb{E}(X|\mathcal{B}) = \sum_{i \in I} \frac{1}{\mathbf{P}(B_i)} \mathbb{E}(X \mathbb{I}_{B_i}) \mathbb{I}_{B_i} = \sum_{i \in I} \frac{1}{\mathbf{P}(B_i)} \mathbb{E}(X) \mathbb{E}(\mathbb{I}_{B_i}) \mathbb{I}_{B_i} = \mathbb{E}(X) \sum_{i \in I} \mathbb{I}_{B_i} = \mathbb{E}(X)$$

□

2.3. Conditionnement par une variable aléatoire discrète. Soit Y une variable aléatoire discrète, $Y(\Omega) = \{y_i, i \in I \subset \mathbb{N}\}$. On peut appliquer les résultats précédents à la tribu engendrée par Y , *i.e.* la tribu engendrée par la partition constituée des événements $B_i = Y^{-1}(\{y_i\})$, $i \in I$. On note $\sigma(Y)$ cette tribu. On définit l'espérance de X sachant Y par

$$\mathbb{E}(X|\sigma(Y)) := \mathbb{E}(X|Y)$$

et de même la loi de X sachant Y

$$\mathbf{P}_X(\cdot|\sigma(Y)) := \mathbf{P}_X(\cdot|Y).$$

En particulier, on a

$$\mathbb{E}(X|Y) = \sum_{i \in \mathbb{N}} \mathbb{E}(X|Y = y_i) \mathbb{I}_{[Y=y_i]} = g(Y).$$

D'après $[\pi_2]$,

$$\int_B g(Y) \, d\mathbf{P} = \int_B X \, d\mathbf{P}$$

pour tout $B \in \sigma(Y)$. Par conséquent pour toute partie C de $\{y_i, i \in I\}$, on a $Y^{-1}(C) \in \sigma(Y)$ et la relation précédente s'écrit

$$\sum_{y \in C} g(y) \mathbf{P}(Y = y) = \int_{Y^{-1}(C)} X \, d\mathbf{P}$$

Exemple 1.2. Reprenons les variables aléatoires définies dans l'exemple 1.1. L'espérance de N conditionnellement à X_1 est donné par

$$\sum_{k=0}^{\infty} E(N|X_1 = k) \mathbb{I}_{X_1=k}.$$

De plus, on a

$$\begin{aligned} E(N|X_1 = k) &= \mathbf{P}(N = 1|X_1 = k) + 2\mathbf{P}(N = 2|X_1 = k) \\ &= \frac{\mathbf{P}(X_2 \geq k | X_1 = k) + 2\mathbf{P}(X_2 < k, X_1 = k)}{\mathbf{P}(X_1 = k)} \\ &= \mathbf{P}(X_2 \geq k) + 2\mathbf{P}(X_2 < k) = 1 + \mathbf{P}(X_2 < k) \\ &= \begin{cases} 1 + \sum_{j=0}^{k-1} (1 - \alpha)\alpha^j = 2 - \alpha^k & \text{si } k \geq 1 \\ 1 & \text{si } k = 0 \end{cases} \\ &= 2 - \alpha^k \end{aligned}$$

Ainsi, on obtient

$$\mathbb{E}(N|X_1) = \sum_{k=0}^{\infty} (2 - \alpha^k) \mathbb{I}_{X_1=k} = 2 - \alpha^{X_1}$$

||

2.4. Quelques remarques. On sait désormais calculer l'espérance conditionnelle par rapport à une tribu engendrée par une partition dénombrable ou par une variable aléatoire discrète.

Il est important de remarquer que les propriétés $[\pi_2]$ définissent l'espérance conditionnelle presque sûrement. En effet,

Proposition 1.6. *Si deux variables aléatoires Y_1 et Y_2 vérifient les propriétés $[\pi_2]$ alors elles sont égales \mathbf{P} -presque sûrement.*

DÉMONSTRATION. On a, pour tout $B \in \mathcal{B}$,

$$\int_B (Y_1 - Y_2) \, d\mathbf{P} = 0 \tag{1.5}$$

Posons $B_1 = \{Y_1 - Y_2 > 0\}$ et $B_2 = \{Y_1 - Y_2 < 0\}$. Ces deux événements sont dans la tribu \mathcal{B} car $Y_1 - Y_2$ est \mathcal{B} -mesurable. En appliquant (1.5) aux événements B_1 et B_2 , on obtient qu'ils sont \mathbf{P} -négligeables. Par conséquent, on a bien $Y_1 - Y_2 = 0$ \mathbf{P} -presque sûrement. \square

Les conséquences de ce résultat sont les suivantes.

- (1) Si on modifie la variable aléatoire $\sum \mathbb{E}(X|B_i) \mathbb{I}_{B_i}$ sur un ensemble \mathcal{B} -mesurable et de probabilité nulle alors elle vérifie toujours les propriétés $[\pi_2]$.
- (2) Si l'on trouve une variable aléatoire qui vérifie les propriétés $[\pi_2]$ alors elle est presque sûrement égale à $\sum \mathbb{E}(X|B_i) \mathbb{I}_{B_i}$.
Dans ce cas on parlera de **version de l'espérance conditionnelle**.

3. Conditionnement par une tribu

L'exemple suivant met en évidence la difficulté de généraliser au conditionnement par une variable aléatoire non nécessairement discrète.

Exemple 1.3. Si (X, Y) est un couple de variables aléatoires indépendantes et de même loi uniforme sur $[0, 1]$. On pose $Z = X + Y$. On souhaite donner un sens à $\mathbf{P}(X \leq a | Z = \alpha)$ et à $\mathbb{E}(X \leq a | Z = \alpha)$.

Ici, la définition précédente n'a pas de sens, puisque les probabilités $\mathbf{P}(X \leq a, Z = \alpha)$ et $\mathbf{P}(Z = \alpha)$ sont toutes les deux nulles.

Une possibilité pourrait être de calculer $\mathbf{P}(X \leq a | Z \in [\alpha, \alpha + h])$ puis de prendre si elle existe la limite quand h tend vers zéro.

On prend $\alpha < 1$. Soit $h > 0$ tel que $[\alpha, \alpha + h] \subset [0, 1]$, on a

$$\mathbf{P}(Z \in [\alpha, \alpha + h]) = 1/2((\alpha + h)^2 - \alpha^2) = 1/2h^2 + \alpha h$$

et

$$\mathbf{P}(X \leq a, Z \in [\alpha, \alpha + h]) = \begin{cases} ah & \text{si } a < \alpha \\ \mathbf{P}(Z \in [\alpha, \alpha + h]) - 1/2(\alpha + h - a)^2 & \text{si } \alpha \leq a \leq h + \alpha \\ \mathbf{P}(Z \in [\alpha, \alpha + h]) & \text{si } a > h + \alpha \end{cases}$$

d'où

$$\mathbf{P}(X \leq a | Z \in [\alpha, \alpha + h]) = \begin{cases} \frac{2ah}{h^2 + 2h\alpha} & \text{si } a < \alpha \\ 1 - \frac{(\alpha + h - a)^2}{h^2 + 2h\alpha} & \text{si } \alpha \leq a \leq h + \alpha \\ 1 & \text{si } a > h + \alpha \end{cases} \xrightarrow{h \rightarrow 0} \begin{cases} \frac{a}{\alpha} & \text{si } a < \alpha \\ 1 & \text{si } a \geq \alpha \end{cases}$$

On reconnaît la loi uniforme sur $[0, \alpha]$. ||

Le problème principal posé par cette définition est l'existence de la limite quand h tend vers zéro.

Une "meilleure" voie est d'utiliser les propriétés $[\pi_2]$ comme définition de l'espérance conditionnelle et de vérifier que les propriétés obtenues dans la section précédente

Définition 1.4. Soit \mathcal{B} une tribu et X une variable aléatoire. L'espérance conditionnelle de X sachant \mathcal{B} est la classe des variables aléatoires $\mathbb{E}(X|\mathcal{B})$ qui vérifient $[\pi_2]$ i.e.

$$\begin{cases} \mathbb{E}(X|\mathcal{B}) \text{ est } \mathcal{B}\text{-mesurable} \\ \text{pour tout } B \in \mathcal{B}, \text{ on a } \int_B \mathbb{E}(X|\mathcal{B}) \, d\mathbf{P} = \int_B X \, d\mathbf{P}. \end{cases}$$

On appelle version de l'espérance conditionnelle un élément de cette classe

La suite de cette section s'organise de la façon suivante :

- 1) On montre que $\mathbb{E}(X|\mathcal{B})$ vérifie les propriétés obtenues dans la section précédente pour l'espérance conditionnelle (Section 3.1)
- 2) On montre l'existence de $\mathbb{E}(X|\mathcal{B})$ sous certaines hypothèses. (Section 3.2)

3.1. Propriétés de $\mathbb{E}(X|\mathcal{B})$. Soit X une variable aléatoire P -intégrable. On suppose qu'il existe une variable aléatoire $\mathbb{E}(X|\mathcal{B})$ qui satisfait les propriétés $[\pi_2]$. Sous ces conditions, la variable aléatoire $\mathbb{E}(X|\mathcal{B})$ vérifie les propriétés suivantes

Propriété A. $\mathbb{E}(\cdot|\mathcal{B})$ est linéaire.

Propriété B. Si X est positive alors $\mathbb{E}(X|\mathcal{B})$ l'est aussi. Sinon, il existe $B \in \mathcal{B}$ de probabilité non nulle telle que $\int_B \mathbb{E}(X|\mathcal{B}) \, d\mathbf{P} < 0$. D'où $\int_B X \, d\mathbf{P} < 0$ ce qui contredit l'hypothèse X positive.

Propriété C. [*Théorème de Beppo Levi pour l'espérance conditionnelle*] Soit $(X_n)_{n \in \mathbb{N}}$ une suite croissante de variables aléatoires positives qui converge vers X alors $\lim \uparrow \mathbb{E}(X_n|\mathcal{B}) = \mathbb{E}(X|\mathcal{B})$.

Notons d'abord que $(\mathbb{E}(X_n|\mathcal{B}))_n$ est une suite croissante de variables aléatoires positives et \mathcal{B} -mesurable. De plus elle est majorée par $\mathbb{E}(X|\mathcal{B})$ donc elle converge et sa limite (notée Y) est \mathcal{B} -mesurable. De plus, en utilisant Beppo Levi, on a

$$\int_B Y \, d\mathbf{P} = \lim \int_B \mathbb{E}(X_n|\mathcal{B}) \, d\mathbf{P} = \lim \int_B X_n \, d\mathbf{P} = \int_B X \, d\mathbf{P}$$

Donc Y vérifie $[\pi_2]$.

Propriété D. [*Lemme de Fatou pour l'espérance conditionnelle*] Si $(X_n)_{n \in \mathbb{N}}$ est une suite positive, alors

$$\mathbb{E}(\lim X_n|\mathcal{B}) \leq \lim \mathbb{E}(X_n|\mathcal{B})$$

Propriété E. [*Théorème de la convergence dominée pour l'espérance conditionnelle*] Si $(X_n)_{n \in \mathbb{N}}$ est une suite qui converge vers X et s'il existe une variable aléatoire Y intégrable telle que pour tout n $X_n \leq Y$ alors

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|\mathcal{B}) = 0.$$

Proposition 1.7. Soit (X, Y) un couple de variables aléatoires. Si X est \mathcal{B} -mesurable alors $\mathbb{E}(XY|\mathcal{B}) = X\mathbb{E}(Y|\mathcal{B})$ presque sûrement.

DÉMONSTRATION. $X\mathbb{E}(Y|\mathcal{B})$ est le produit de variables aléatoires \mathcal{B} -mesurables donc elle l'est aussi. On montre maintenant par la méthode standard l'égalité : pour tout $B \in \mathcal{B}$

$$\int_B X\mathbb{E}(Y|\mathcal{B}) \, d\mathbf{P} = \int_B XY \, d\mathbf{P}$$

Fixons $B \in \mathcal{B}$.

Étape 1 : $X = \mathbb{I}_A$ avec $A \in \mathcal{B}$. On a

$$\begin{aligned} \int_B \mathbb{I}_A \mathbb{E}(Y|\mathcal{B}) \, d\mathbf{P} &= \int_{A \cap B} \mathbb{E}(Y|\mathcal{B}) \, d\mathbf{P} = \int_{A \cap B} Y \, d\mathbf{P} \quad \text{car } A \cap B \in \mathcal{B} \\ &= \int_B \mathbb{I}_A Y \, d\mathbf{P} \end{aligned}$$

Étape 2 : $X = \sum_{i=1}^p \lambda_i \mathbb{I}_{A_i}$. Le résultat découle directement de la linéarité de l'intégrale.

Étape 3 : X est mesurable positive. Il existe une suite croissante de variables aléatoires étagées positives $(X_n)_{n \in \mathbb{N}}$ telle que $\lim \uparrow X_n = X$. L'étape précédente assure que pour tout n , on a

$$\int_B X_n \mathbb{E}(Y|\mathcal{B}) \, d\mathbf{P} = \int_B X_n Y \, d\mathbf{P}. \quad (1.6)$$

◇ La variable aléatoire Y est positive. La suite $(YX_n)_n$ est donc croissante positive et converge vers XY . On peut donc appliquer le Théorème de Beppo Levi aux deux membres de (1.6) et on obtient

$$\int_B X \mathbb{E}(Y|\mathcal{B}) \, d\mathbf{P} = \int_B XY \, d\mathbf{P}$$

◇ Y quelconque. On applique le résultat précédent à Y^+ et Y^- i.e.

$$\int_B X E(Y^\pm | \mathcal{B}) \, d\mathbf{P} = \int_B XY^\pm \, d\mathbf{P}.$$

Ensuite, le résultat découle de la linéarité de l'intégrale et de l'espérance conditionnelle.

Étape 4 : X est intégrable. On applique l'étape précédente à X^+ et X^- et on utilise à nouveau les propriétés de linéarité. □

Proposition 1.8. *Si X est indépendante de \mathcal{B} alors $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}(X)$ presque sûrement.*

DÉMONSTRATION. La variable aléatoire constante et égale à $\mathbb{E}(X)$ est \mathcal{B} -mesurable. De plus, pour tout $B \in \mathcal{B}$, on a

$$\begin{aligned} \int_B \mathbb{E}(X) \, d\mathbf{P} &= \mathbb{E}(X) \int_B d\mathbf{P} = \int X \, d\mathbf{P} \int \mathbb{I}_B \, d\mathbf{P} \\ &= \int X \mathbb{I}_B \, d\mathbf{P} \quad \text{car} \quad X \perp \mathbb{I}_B \\ &= \int_B X \, d\mathbf{P} \end{aligned}$$

□

Proposition 1.9. *Soient \mathcal{B} et \mathcal{B}' deux tribus telles que $\mathcal{B} \subset \mathcal{B}'$, on a*

$$\mathbb{E}(X|\mathcal{B}) = \mathbb{E}\left(\mathbb{E}(X|\mathcal{B}')|\mathcal{B}\right) = \mathbb{E}\left(\mathbb{E}(X|\mathcal{B})|\mathcal{B}'\right) \quad (1.7)$$

DÉMONSTRATION. Rappelons que la condition $\mathcal{B} \subset \mathcal{B}'$ implique que si une variable aléatoire est \mathcal{B} -mesurable alors elle est aussi \mathcal{B}' -mesurable. Ce résultat implique que $\mathbb{E}(X|\mathcal{B})$ est \mathcal{B}' -mesurable et donc d'après la proposition 1.7, on a $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}\left(\mathbb{E}(X|\mathcal{B})|\mathcal{B}'\right)$.

Montrons maintenant que $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}\left(\mathbb{E}(X|\mathcal{B}')|\mathcal{B}\right)$. La \mathcal{B} -mesurabilité de $\mathbb{E}(X|\mathcal{B})$ est acquise, de plus pour tout $B \in \mathcal{B}$

$$\begin{aligned} \int_B \mathbb{E}(X|\mathcal{B}') \, d\mathbf{P} &= \int_B X \, d\mathbf{P} \quad \text{car} \quad B \in \mathcal{B}' \supset \mathcal{B} \\ &= \int_B \mathbb{E}(X|\mathcal{B}) \, d\mathbf{P} \quad \text{car} \quad B \in \mathcal{B} \end{aligned}$$

□

3.2. Existence de $\mathbb{E}(X|\mathcal{B})$ et interprétation géométrique. Dans cette partie, nous montrons que si X est de carré intégrable (c'est à dire $X \in L^2(\Omega, \mathcal{A}, \mathbf{P})$), la meilleure approximation de X par un élément de $L^2(\Omega, \mathcal{B}, \mathbf{P})$ est une version de l'espérance conditionnelle.

Théorème 1.10. *Soit H un espace de Hilbert (i.e. un espace normé complet dont la norme provient d'un produit scalaire). Soit E un sous espace complet de H . Pour tout $x \in H$, il existe un unique élément $y \in E$ tel que*

$$\|y - x\| = \inf_{w \in E} \|w - x\| \quad (1.8)$$

$$y - x \perp E \quad (1.9)$$

On note $y = P_E(x)$. C'est la projection orthogonale sur E .

Rappelons que $L^2(\Omega, \mathcal{A}, \mathbf{P})$ est un espace de Hilbert. Soit \mathcal{B} une sous tribu \mathcal{A} et $L^2(\mathcal{B}) = L^2(\Omega, \mathcal{B}, \mathbf{P})$ le sous espace $L^2(\Omega, \mathcal{A}, \mathbf{P})$ constitué des classes d'équivalence d'applications \mathcal{B} -mesurable. C'est un sous espace complet de $L^2(\Omega, \mathcal{A}, \mathbf{P})$. Les hypothèses du théorème 1.10 sont donc satisfaites. Posons $P_{L^2(\mathcal{B})}$ la projection orthogonale sur $L^2(\mathcal{B})$.

Théorème 1.11. *Dans le contexte décrit ci-dessus, si X une variable aléatoire dans $L^2(\Omega, \mathcal{A}, \mathbf{P})$ alors la variable aléatoire $Y = P_{L^2(\mathcal{B})}(X)$ vérifie la condition $[\pi_2]$ (c'est une version de l'espérance conditionnelle de X sachant \mathcal{B} .)*

DÉMONSTRATION. Montrons que la variable aléatoire Y vérifie la condition $[\pi_2]$. Par construction, la variable aléatoire $Y = P_{L^2(\mathcal{B})}(X)$ est \mathcal{B} -mesurable. Comme la variable $Y - X$ est orthogonale à \mathcal{B} , on a pour tout $B \in \mathcal{B}$, les variables aléatoires \mathbb{I}_B et $Y - X$ sont orthogonales c'est à dire

$$\mathbb{E}((X - Y) \mathbb{I}_B) = 0.$$

Ceci implique que

$$\mathbb{E}(Y \mathbb{I}_B) = \mathbb{E}(X \mathbb{I}_B)$$

ou encore

$$\int_B Y \, d\mathbf{P} = \int_B X \, d\mathbf{P}.$$

Par conséquent, la variable aléatoire Y vérifie $[\pi_2]$. \square

Remarque 1.3. L'existence reste vraie dans L^1 . On admet ici ce résultat. Il est important de remarquer que l'interprétation en terme de projection n'est plus valide. |

3.3. Conditionnement par une variable aléatoire. Soit Y une variable aléatoire à valeur dans l'espace E (muni de la tribu \mathcal{E}). On note $\sigma(Y)$ la tribu engendrée par Y . Pour simplifier les notations, on écrit $\mathbb{E}(X|Y)$ au lieu de $\mathbb{E}(X|\sigma(Y))$.

Lemme 1.12 (Lemme de Doob). *Soit U une variable aléatoire sur Ω . U est $\sigma(Y)$ -mesurable si et seulement si il existe une application mesurable g de E dans \mathbb{R} telle $U = g(Y)$*

L'espérance conditionnelle $\mathbb{E}(X|Y)$ est donc de la forme $g(Y)$ avec g mesurable et

$$\int_B g(Y) \, d\mathbf{P} = \int_B X \, d\mathbf{P} \quad \forall B \in \sigma(Y).$$

L'événement B peut aussi s'exprimer de la forme $Y^{-1}(C)$ avec $C \in \mathcal{E}$. En utilisant le théorème du transfert, on obtient

$$[\pi_3] : \int_C g(y) \, d\mathbf{P}_Y(y) = \int_{Y^{-1}(C)} X \, d\mathbf{P}$$

Remarque 1.4. Pas abus de langage, on utilise aussi la notation $g(y) = \mathbb{E}(X|Y = y)$ qui est différente de $\mathbb{E}(X|\{Y = y\})$ |

Remarque 1.5. *Rappel : une π -classe est une collection de parties de E contenant E et stable par intersection fini.*

Par exemple, $\{]-\infty, a], a \in \mathbb{R}\}$ est une π -classe qui engendre la tribu borélienne sur \mathbb{R}

La propriété $[\pi_1]$ (resp. $[\pi_2]$) est vraie si et seulement si elle est vérifiée pour tout $C \in \mathcal{C}$ où \mathcal{C} est une π -classe qui engendre \mathcal{B} . On admet ce résultat.

Rappel : L'image réciproque d'une π -classe est aussi une π -classe.

Par conséquent $[\pi_3]$ est vraie si et seulement si elle est vérifiée sur une π -classe qui engendre \mathcal{E} . |

Dans ce contexte, les propriétés de l'espérance conditionnelle s'écrivent

- 1) Si X et Y sont indépendantes alors $\mathbb{E}(X|Y) = \mathbb{E}(X)$.
- 2) Pour tout application h de E dans \mathbb{R} mesurable, $\mathbb{E}(h(Y)Z|Y) = h(Y)\mathbb{E}(Z|Y)$.
- 3) Si $Z = h(Y)$ ($\implies \sigma(Z) \subset \sigma(Y)$) alors

$$\mathbb{E}(X|Z) = \mathbb{E}(\mathbb{E}(X|Y)|Z) = \mathbb{E}(\mathbb{E}(X|Z)|Y).$$

3.4. Vecteur à densité. Soit (X, Y) un couple de variables aléatoires qui admet une densité f par rapport à la mesure produit $\mu_1 \otimes \mu_2$. On a en particulier

$$\mathbf{P}(X \in B, Y \in C) = \int_{B \times C} f(x, y) \, d\mu_1 \otimes \mu_2(x, y)$$

Proposition 1.13. Si $\mathbb{E}|X| < \infty$ i.e. $\int |x|f(x, y) \, d\mu_1 \otimes \mu_2(x, y) < \infty$ alors

$$\mathbb{E}(X|Y) = g(Y) = \frac{\int x f(x, Y) \, d\mu_1(x)}{\int f(x, Y) \, d\mu_1(x)} \quad (1.10)$$

DÉMONSTRATION. Montrons d'abord que g est bien définie. Le dénominateur est nul sur $\{\omega, \int f(x, Y(\omega)) \, d\mu_1(x) = 0\}$. On a

$$\begin{aligned} \mathbf{P}\left\{\omega, \int f(x, Y(\omega)) \, d\mu_1(x) = 0\right\} &= \int_{\{y: \int f(x, y) \, d\mu_1(x) = 0\}} d\mathbf{P}_Y(y) \\ &= \int_{\{y: f_Y(y) = 0\}} f_Y(y) \, d\mu_2(y) = 0 \end{aligned}$$

$g(Y(\omega))$ est donc définie \mathbf{P} -p.s.

$g(Y)$ est $\sigma(Y)$ -mesurable (voir un cours sur les mesures produits).

On montre maintenant l'égalité [π₃]. Soit $C \in \mathcal{E}$

$$\begin{aligned} \int_C g(y) \, d\mathbf{P}_Y(y) &= \int_C g(y) f_Y(y) \, d\mu_2(y) \\ &= \int_C \frac{\int x f(x, y) \, d\mu_1(x)}{\int f(x, y) \, d\mu_1(x)} f_Y(y) \, d\mu_2(y) \\ &= \int_C \int x f(x, y) \, d\mu_1(x) \, d\mu_2(y) \\ &= \int_{Y^{-1}(C)} X \, d\mathbf{P} \quad [\text{via le théorème du transfert}] \end{aligned}$$

□

Cette proposition (1.10) nous donne une méthode de calcul de l'espérance conditionnelle dans les situations suivantes :

Situation 1. (X, Y) est un couple de variables aléatoires discrètes

$$\mathbb{E}(X|Y = k) = \frac{1}{\mathbf{P}(Y = k)} \sum_j x_j \mathbf{P}(X = x_j, Y = k)$$

c'est le contexte de l'exemple 1.2.

Situation 2. (X, Y) est un couple de variables aléatoires admettant f pour densité :

$$\mathbb{E}(X|Y = y) = \frac{\int x f(x, y) \, dx}{\int f(x, y) \, dx}$$

3.5. Distribution conditionnelle. À partir de la définition de l'espérance conditionnelle $\mathbb{E}(X|\mathcal{B})$, on obtient l'existence de la probabilité conditionnelle $\mathbf{P}(A|\mathcal{B})$ en prenant $X = \mathbb{I}_A$. Dans le cas général, on a donc

$$\int_B \mathbf{P}(A|\mathcal{B}) \, d\mathbf{P} = \int_B \mathbb{I}_A \, d\mathbf{P} = \mathbf{P}(A \cap B)$$

Si la tribu \mathcal{B} est engendrée par une variable aléatoire Y à valeurs dans \mathbb{R}^p alors $\mathbf{P}(A|Y) = \phi(Y)$ et pour tout $C \in \mathcal{B}(\mathbb{R}^p)$

$$\int_C \phi(y) \, d\mathbf{P}_Y(y) = \mathbf{P}(Y^{-1}(C) \cap A)$$

Comme pour l'espérance conditionnelle, on notera par abus de langage $P(A|Y = y) = \phi(y)$.

Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements disjoints. On peut sélectionner, pour chaque n , une version de $\mathbf{P}(A_n|\mathcal{B})$. On vérifie facilement que

$$\mathbf{P}(\cup_{n \in \mathbb{N}} A_n | \mathcal{B}) = \sum_{n \in \mathbb{N}} \mathbf{P}(A_n | \mathcal{B}) \quad \text{p.s.} \quad (1.11)$$

$\mathbf{P}(\cdot | \mathcal{B})(\omega)$ semble se comporter comme une probabilité mais il y a un problème. En effet l'ensemble de mesure 0 où l'égalité (1.11) n'est pas vérifiée dépend de la suite $(A_n)_{n \in \mathbb{N}}$. Donc $\mathbf{P}(\cdot | \mathcal{B})(\omega)$ ne définit pas, a priori, une probabilité.

Définition 1.5. On appelle version régulière de la probabilité conditionnelle par rapport à \mathcal{B} , une application π de $\Omega \times \mathcal{A}$ dans $[0, 1]$ telle que

- 1) à $A \in \mathcal{A}$ fixé, $\pi(\cdot, A)$ est \mathcal{B} -mesurable et $\pi(\cdot, A) = \mathbb{E}(\mathbb{I}_A | \mathcal{B})(\cdot)$ p.s.
- 2) à ω fixé, $A \mapsto \pi(\omega, A)$ est une mesure de probabilité.

L'existence d'une version régulière n'est pas garantie en général. Le théorème suivant nous donne un exemple où l'existence de la version régulière est assurée.

Théorème 1.14. Soit X une variable aléatoire sur \mathbb{R}^q . Pour tout $D \in \mathcal{B}(\mathbb{R}^q)$, on peut choisir une version de $\mathbf{P}_X(D|\mathcal{B}) = \mathbf{P}(X^{-1}(D)|\mathcal{B})$ telle que

- (1) à $D \in \mathcal{B}(\mathbb{R}^q)$ fixé, $\mathbf{P}_X(D|\mathcal{B})(\cdot)$ est une version de $\mathbb{E}(\mathbb{I}_{X^{-1}(D)}|\mathcal{B})(\cdot)$
- (2) à ω fixé, $\mathbf{P}_X(\cdot|\mathcal{B})(\omega)$ est une mesure de probabilité.

On parle alors de version régulière.

Définition 1.6. $\mathbf{P}_X(\cdot|\mathcal{B})$ est appelée loi conditionnelle de X sachant \mathcal{B} . Si $\mathcal{B} = \sigma(Y)$ avec Y une variable aléatoire à valeurs dans \mathbb{R}^p , alors $\mathbf{P}_X(\cdot|Y)$ est la loi de X sachant Y .

Pour tout $D \in \mathcal{B}(\mathbb{R}^q)$ et tout $B \in \mathcal{B}$, on a

$$\int_B \mathbf{P}_X(D|\mathcal{B}) \, d\mathbf{P} = \mathbf{P}(X^{-1}(D) \cap B)$$

et pour tout $\psi(X)$ intégrable

$$\mathbb{E}(\psi(X)|\mathcal{B})(\omega) = \int_{\mathbb{R}} \psi(x) \, d\mathbf{P}_X(x|\mathcal{B})(\omega)$$

Si on prend \mathcal{B} la tribu engendrée par la variable aléatoire Y , on a pour tout $D \in \mathcal{B}(\mathbb{R}^q)$ et $C \in \mathcal{B}(\mathbb{R}^p)$

$$\int_C \mathbf{P}_X(D|Y = y) \, d\mathbf{P}_Y(y) = \mathbf{P}(X \in D, Y \in C)$$

et pour tout $\psi(X)$ intégrable, on a

$$\mathbb{E}(\psi(X)|Y)(\omega) = \int \psi(x) \, d\mathbf{P}_X(x|Y)(\omega)$$

pour tout ω .

Proposition 1.15. Soit (X, Y) un couple de variables aléatoires qui admet pour densité f par rapport à la mesure $\mu_1 \otimes \mu_2$. La loi conditionnelle de X sachant Y admet pour densité

$$f_X(x|Y = y) = \frac{f(x, y)}{\int f(x, y) \, d\mu_1(x)} \mathbb{I}_{\{y: \int f(x, y) \, d\mu_1(x) \neq 0\}} \quad (1.12)$$

DÉMONSTRATION. Posons $C_Y = \{y : \int f(x, y) \, d\mu_1(x) \neq 0\}$. Il faut montrer que

$$\int_A f_X(x|Y = y) \, d\mu_1(x)$$

est une version de $\mathbf{P}(X \in A|Y)$. Pour tout $B \in \mathcal{B}(\mathbb{R}^p)$, on a

$$\begin{aligned} \int_B \int_A f_X(x|Y=y) \, d\mu_1(x) \, d\mathbf{P}_Y(y) &= \int_B \left(\int_A f_X(x|Y=y) \, d\mu_1(x) \right) f_Y(y) \, d\mu_2(y) \\ &= \int_B \int_A f(x,y) \, d\mu_1(x) \, d\mu_2(y) \quad (\text{Th Fubini}) \\ &= \mathbf{P}(X \in A, Y \in B) \end{aligned}$$

□

Annexe : Chaînes de Markov à espace d'états au plus dénombrable

Soit E un ensemble dénombrable et (X_n) une suite de variables aléatoires à valeurs dans E .

Définition 1.7. $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov si et seulement si pour tout $n \in \mathbb{N}$ et tout $(j_0, \dots, j_n) \in E^{n+1}$

$$\mathbf{P}(X_n = j_0 | X_{n-1} = j_1, \dots, X_0 = j_n) = \mathbf{P}(X_n = j_0 | X_{n-1} = j_1) \stackrel{\text{def}}{=} p_n(j_1, j_0).$$

Si p_n ne dépend pas de n la chaîne de Markov est dite homogène.

Remarque 1.6. L'ensemble E est souvent appelé espace des états de la chaîne. |

Proposition 1.16. Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov homogène.

- (1) La loi de X_n est déterminée par la donnée des $p(i, j) = \mathbf{P}(X_1 = j | X_0 = i)$ et par la loi de X_0 .
- (2) Pour tout entier $h > 0$ et pour tout $(j_0, \dots, j_h) \in E^{h+1}$, on a

$$\mathbf{P}(X_n = j_0 | X_{n-1} = j_1, \dots, X_{n-h} = j_h) = \mathbf{P}(X_n = j_0 | X_{n-1} = j_1).$$

- (3) Pour tout entier $h > 0$ et pour tout $(j_0, \dots, j_{n+1}) \in E^{n+2}$, on a

$$\mathbf{P}(X_{n+h} = j_0 | X_n = j_1, \dots, X_0 = j_{n+1}) = \mathbf{P}(X_{n+h} = j_0 | X_n = j_1).$$

DÉMONSTRATION.

- (1) Rappelons que pour toute suite finie d'événements A_1, \dots, A_n , on a la relation suivante

$$\mathbf{P}(A_1 \cap \dots \cap A_n) = \mathbf{P}(A_n | A_{n-1} \cap \dots \cap A_1) \mathbf{P}(A_{n-1} | A_{n-2} \cap \dots \cap A_1) \cdots \mathbf{P}(A_1). \quad (1.13)$$

Pour tout $j \in E$, on a

$$\begin{aligned} \mathbf{P}(X_n = j) &= \mathbf{P} \left(\bigcup_{i_1, \dots, i_n} [X_n = j, X_{n-1} = i_1, \dots, X_0 = i_n] \right) \\ &= \sum_{i_1, \dots, i_n} \mathbf{P}(X_n = j, X_{n-1} = i_1, \dots, X_0 = i_n) \quad (\text{car les événements sont disjoints}) \\ &= \sum_{i_1, \dots, i_n} \mathbf{P}(X_n = j | X_{n-1} = i_1, \dots, X_0 = i_n) \mathbf{P}(X_{n-1} = i_1 | X_{n-2} = i_2, \dots, X_0 = i_n) \\ &\quad \cdots \mathbf{P}(X_1 = i_{n-1} | X_0 = i_n) \mathbf{P}(X_0 = i_n) \quad (\text{d'après (1.13)}) \\ &= \sum_{i_1, \dots, i_n} p(i_1, j) p(i_2, i_1) \cdots p(i_n, i_{n-1}) \pi_0(i_n) \end{aligned} \quad (1.14)$$

puisque $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov.

- (2) Par définition on a

$$\begin{aligned}
\mathbf{P}(X_n = j_0 | X_{n-1} = j_1, \dots, X_{n-h} = j_h) &= \frac{\mathbf{P}(X_n = j_0, X_{n-1} = j_1, \dots, X_{n-h} = j_h)}{\mathbf{P}(X_{n-1} = j_1, \dots, X_{n-h} = j_h)} \\
&= \sum_{j_{h+1}, \dots, j_n} \frac{\mathbf{P}(X_n = j_0, X_{n-1} = j_1, \dots, X_0 = j_n)}{\mathbf{P}(X_{n-1} = j_1, \dots, X_{n-h} = j_h)} \\
&= \sum_{j_{h+1}, \dots, j_n} \mathbf{P}(X_n = j_0 | X_{n-1} = j_1, \dots, X_0 = j_n) \frac{\mathbf{P}(X_{n-1} = j_1, \dots, X_0 = j_n)}{\mathbf{P}(X_{n-1} = j_1, \dots, X_{n-h} = j_h)} \\
&= \mathbf{P}(X_n = j_0 | X_{n-1} = j_1)
\end{aligned}$$

car le terme $\mathbf{P}(X_n = j_0 | X_{n-1} = j_1, \dots, X_0 = j_n)$ ne dépend pas de $(j_2, \dots, j_{h+1}, \dots, j_n)$ et vaut $\mathbf{P}(X_n = j_0 | X_{n-1} = j_1)$. De plus, on a

$$\begin{aligned}
\sum_{j_{h+1}, \dots, j_n} \frac{\mathbf{P}(X_{n-1} = j_1, \dots, X_0 = j_n)}{\mathbf{P}(X_{n-1} = j_1, \dots, X_{n-h} = j_h)} \\
= \sum_{j_{h+1}, \dots, j_n} \mathbf{P}(X_{n-h-1} = j_{h+1}, \dots, X_0 = j_n | X_{n-1} = j_1, \dots, X_{n-h} = j_h) = 1
\end{aligned}$$

(3) Le terme de gauche s'exprime de la façon suivante

$$\begin{aligned}
\mathbf{P}(X_{n+h} = j_0 | X_n = j_1, \dots, X_0 = j_{n+1}) &= \frac{\mathbf{P}(X_{n+h} = j_0, X_n = j_1, \dots, X_0 = j_{n+1})}{\mathbf{P}(X_n = j_1, \dots, X_0 = j_{n+1})} \\
&= \sum_{i_1, \dots, i_{h-1}} \frac{\mathbf{P}(X_{n+h} = j_0, X_{n+h-1} = i_1, \dots, X_{n+1} = i_{h-1}, X_n = j_1, \dots, X_0 = j_{n+1})}{\mathbf{P}(X_n = j_1, \dots, X_0 = j_{n+1})} \\
&= \sum_{i_1, \dots, i_{h-1}} p(i_1, j_0) \cdots p(j_1, i_{h-1})
\end{aligned}$$

On traite le terme de droite de la même façon *i.e.*

$$\begin{aligned}
\mathbf{P}(X_{n+h} = j_0 | X_n = j_1) &= \frac{\mathbf{P}(X_{n+h} = j_0, X_n = j_1)}{\mathbf{P}(X_n = j_1)} \\
&= \sum_{i_1, \dots, i_{h-1}} \frac{\mathbf{P}(X_{n+h} = j_0, X_{n+h-1} = i_1, \dots, X_{n+1} = i_{h-1}, X_n = j_1)}{\mathbf{P}(X_n = j_1)} \\
&= \sum_{i_1, \dots, i_{h-1}} p(i_1, j_0) \cdots p(j_1, i_{h-1}).
\end{aligned}$$

La dernière égalité est obtenue grâce à la propriété 2. Ceci conclut la preuve. \square

Le lemme suivant donne une méthode pour construire des chaînes de Markov. Il est souvent très utile pour vérifier qu'une suite de variables aléatoires est une chaîne de Markov.

Lemme 1.17. Soit $(Y_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires *i.i.d.* à valeurs dans un espace dénombrable. Si $(X_n)_{n \in \mathbb{N}}$ est définie par la donnée de X_0 indépendante de la suite $(Y_n)_{n \in \mathbb{N}}$ et la relation

$$X_{n+1} = f_n(X_n, Y_n) \quad (1.15)$$

où f_n est une fonction à valeurs dans E , alors $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov.

De plus, si la fonction f_n ne dépend pas de n alors la chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ est homogène.

La preuve de ce résultat repose sur le lemme suivant

Lemme 1.18. Soient U, V et W trois variables aléatoires discrètes à valeur dans trois espaces E, F et G . Si pour tout $(u, v, w) \in E \times F \times G$ la probabilité $\mathbf{P}(U = u | V = v, W = w)$ ne dépend pas de w alors c'est aussi la probabilité $\mathbf{P}(U = u | V = v)$

DÉMONSTRATION. Pour tout w , on a

$$\begin{aligned}\mathbf{P}(U = u, V = v, W = w) &= \mathbf{P}(U = u|V = v, W = w)\mathbf{P}(V = v, W = w) \\ &= F(u, v)\mathbf{P}(V = v, W = w)\end{aligned}$$

où $F(u, v) = \mathbf{P}(U = u|V = v, W = w)$. On obtient

$$\begin{aligned}\mathbf{P}(U = u, V = v) &= \sum_{w \in E} \mathbf{P}(U = u, V = v, W = w) \\ &= F(u, v) \sum_{w \in E} \mathbf{P}(V = v, W = w) \\ &= F(u, v)\mathbf{P}(V = v)\end{aligned}$$

c'est à dire

$$\mathbf{P}(U = u|V = v) = F(u, v)$$

□

DÉMONSTRATION. (*Retour au Lemme 1.17*)

Remarquons que X_n s'exprime comme une fonction des variables aléatoires $(X_0, Y_0, \dots, Y_{n-1})$. Comme le vecteur $(X_0, Y_0, \dots, Y_{n-1})$ est indépendant de la suite (Y_n, Y_{n+1}, \dots) , il en est de même pour X_n . Soit $(j_0, \dots, j_n) \in E^{n+1}$,

$$\begin{aligned}\mathbf{P}(X_n = j_0, \dots, X_0 = j_n) &= \sum_{\{l: f_{n-1}(j_1, l) = j_0\}} \mathbf{P}(Y_{n-1} = l, X_{n-1} = j_1, \dots, X_0 = j_n) \\ &= \sum_{\{l: f_{n-1}(j_1, l) = j_0\}} \mathbf{P}(Y_{n-1} = l)\mathbf{P}(X_{n-1} = j_1, \dots, X_0 = j_n)\end{aligned}$$

d'où

$$\mathbf{P}(X_n = j_0|X_{n-1} = j_1, \dots, X_0 = j_n) = \sum_{\{l: f_{n-1}(j_1, l) = j_0\}} \mathbf{P}(Y_{n-1} = l)$$

Le terme de droite ne dépend pas de (j_2, \dots, j_n) , c'est donc $\mathbf{P}(X_n = j_0|X_{n-1} = j_1)$ d'après le lemme 1.18. La suite $(X_n)_{n \in \mathbb{N}}$ est donc une chaîne de Markov avec

$$p_n(i, j) = \sum_{\{l: f_{n-1}(i, l) = j\}} \mathbf{P}(Y_{n-1} = l) = \sum_{\{l: f_{n-1}(i, l) = j\}} \mathbf{P}(Y_0 = l).$$

De plus, si la fonction f_n ne dépend pas de n alors $p_n(i, j)$ ne dépend pas de n et donc la chaîne de Markov est homogène. □

Exemple 1.4. (*Marche aléatoire*) On pose $X_{n+1} = X_n + Y_n = f(X_n, Y_n)$ où f ne dépend pas de n . D'après le lemme 1.17, (X_n) est une chaîne de Markov homogène et

$$p(i, j) = \sum_{\{l: i+l=j\}} \mathbf{P}(Y_0 = l) = \mathbf{P}(Y_0 = j - i).$$

On prend par exemple $X_0 = 0$ p.s. et la suite (Y_i) i.i.d. suivant la loi de Bernoulli de paramètre p . Il est facile de vérifier que

$$X_n = \sum_{k=1}^n Y_k \quad \forall k \in \mathbb{N}$$

et donc X_n suit une loi binomiale $\sim \mathcal{B}(n, p)$. □

Exemple 1.5. On pose

$$X_{n+1} = \begin{cases} X_n + Y_n & \text{si } X_n \in D \\ X_n & \text{si } X_n \notin D \end{cases}$$

on a donc $f(X_n, Y_n) = X_n + Y_n \mathbb{1}_{X_n \in D}$ et

$$p(i, j) = \sum_{\{l | f(i, l) = j\}} \mathbf{P}(Y_0 = l) = \begin{cases} \mathbf{P}(Y = j - i) & \text{si } i \in D \\ 1 & \text{si } i \notin D \text{ et } i = j \\ 0 & \text{si } i \notin D \text{ et } i \neq j \end{cases}$$

Pour conclure cette section, on considère des chaînes de Markov à espace d'états fini. Fixons $E = \{1, \dots, q\}$. Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov homogène sur E , on définit la matrice

$$Q = \begin{pmatrix} p(1, 1) & p(1, 2) & \cdots & p(1, q) \\ \vdots & \vdots & \ddots & \vdots \\ p(q, 1) & p(q, 2) & \cdots & p(q, q) \end{pmatrix}. \quad (1.16)$$

Définition 1.8. La matrice Q définie en (1.16) est appelée matrice de transition.

La somme des termes d'une ligne vaut 1,

$$\mathbf{P}(X_2 = 1 | X_1 = k) + \cdots + \mathbf{P}(X_2 = q | X_1 = k) = 1 \quad \forall k \in E.$$

L'équation (1.14) s'écrit

$$(\pi_1(1), \dots, \pi_1(q)) = (\pi_0(1), \dots, \pi_0(q))Q \quad \pi_1 = \pi_0 Q$$

D'où $\pi_n = \pi_0 Q^n$ si bien que Q^n peut s'interpréter comme la matrice de transition

$$\mathbf{P}(X_n = j | X_0 = i) = Q^n(i, j).$$

Annexe : Démonstration de la Proposition 1.10

Montrons d'abord que les propriétés (1.8) et (1.9) sont équivalentes.

(1.8) **implique** (1.9) :

Soit w arbitraire dans E , Pour tout $t \in \mathbb{R}$, on a $tw + y \in E$ et

$$\|y - x\|^2 \leq \|tw + y - x\|^2 = t^2 \|w\|^2 + 2t(w|y - x) + \|y - x\|^2.$$

D'où $t^2 \|w\|^2 + 2t(w|y - x) \geq 0$ pour tout $t \in \mathbb{R}$ ceci est possible si et seulement si $(w|y - x) = 0$. On a donc prouvé que $y - x \perp E$.

(1.9) **implique** (1.8) :

Soit $y \in E$ tel que $y - x \perp E$, pour tout $w \in E$, $\|x - w\|^2 = \|x - y\|^2 + \|y - w\|^2$ car $y - w \in E$ est orthogonal à $x - y$. On a donc $\|x - w\|^2 \geq \|x - y\|^2$ et par conséquent $\|y - x\| = \inf_{w \in E} \|w - x\|$.

Existence

On pose $d = \inf_{w \in E} \|w - x\|$. Si $d = 0$ alors $x \in E$ convient. On suppose désormais que $d \neq 0$. Par définition de la borne inférieure il existe une suite $(y_n)_n$ dans E telle que $\|y_n - x\| \rightarrow d$ quand n tend vers l'infini. Soit $\varepsilon > 0$ fixé, il existe un entier N tel que pour tout $n > N$ $\|y_n - x\|^2 < \varepsilon^2 + d^2$. Prenons $n, m > N$, par l'identité du parallélogramme, on a

$$\begin{aligned} \|y_n - x\|^2 + \|y_m - x\|^2 &= \frac{1}{2} \|y_n + y_m - 2x\|^2 + \frac{1}{2} \|y_m - y_n\|^2 \\ &= 2 \left\| \frac{y_n + y_m}{2} - x \right\|^2 + \frac{1}{2} \|y_m - y_n\|^2 \end{aligned}$$

Comme $\frac{y_n + y_m}{2} \in E$, on a $\left\| \frac{y_n + y_m}{2} - x \right\|^2 \geq d^2$. On en déduit que $\|y_m - y_n\|^2 \leq 4\varepsilon^2$. La suite $(y_n)_n$ est donc de Cauchy dans E . Cet espace étant complet, la suite converge dans E , notons y la limite. De plus,

$$d^2 \leq \|y - x\|^2 \leq \|y_n - x\|^2 + \|y_n - y\|^2 \xrightarrow[n \rightarrow \infty]{} d^2 + 0$$

ce qui implique que $d = \|y - x\|$. On a donc prouvé l'existence de $y \in E$ vérifiant (1.8)

Unicité

Soit $(y, z) \in E^2$ tel que $d = \|x - y\| = \|x - z\|$. On a

$$\|x - y\|^2 = \|x - z\|^2 = \|x - y\|^2 + \|y - z\|^2 + 2(x - y|y - z) \quad (1.17)$$

or $(x - y|y - z) = 0$ car $y - z \in E$ et $x - y \perp E$, et donc (1.17) implique que $\|y - z\| = 0$ d'où $y = z$.

Vecteurs Gaussiens

1. Rappels : vecteurs aléatoires, covariance, fonction caractéristique.

Nous commençons ce chapitre par quelques rappels sur les vecteurs aléatoires. Soit $X = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$

un vecteur aléatoire *i.e.* une application mesurable sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Toutes les coordonnées du vecteur X sont des variables aléatoires réelles.

Définition 2.1. Soit X un vecteur aléatoire tel que, pour tout $j \in \{1, \dots, k\}$, la variable aléatoire X_j est \mathbf{P} -intégrable. L'espérance de X est le vecteur $\mathbb{E}(X) = {}^t(\mathbb{E}(X_1) \cdots \mathbb{E}(X_k))$. On dit que le vecteur aléatoire X est centré si $\mathbb{E}X$ est le vecteur nul dans \mathbb{R}^k .

Soit X un vecteur aléatoire dont toutes les coordonnées sont des variables aléatoires réelles de L^2 . La notion de variance pour les variables aléatoires réelles se généralise en considérant la matrice (notée $\text{Var}(X)$ ou Σ_X) constituée des éléments

$$\text{Cov}(X_i, X_j) = \mathbb{E}((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))) \quad \forall 1 \leq i, j \leq k.$$

Un simple calcul permet d'établir l'égalité suivante

$$\text{Cov}(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j).$$

La matrice Σ_X est appelée matrice de covariance de X . Elle peut aussi s'exprimer de la façon suivante :

$$\Sigma_X = \mathbb{E}(X {}^tX) - \mathbb{E}(X) {}^t\mathbb{E}(X)$$

où tX représente la transposée de X .

Soient X et Y deux vecteurs aléatoires, L^2 . On suppose que X (resp. Y) est à valeur dans \mathbb{R}^k (resp. \mathbb{R}^p). La covariance entre les vecteurs X et Y est la matrice constituée des éléments $\text{Cov}(X_i, Y_j)$ pour $i = 1, \dots, k$ et $j = 1, \dots, p$. On la note $\text{Cov}(X, Y)$. Remarquons que si les vecteurs sont indépendants alors la matrice $\text{Cov}(X, Y)$ est la matrice nulle.

Proposition 2.1. Transformation affine : Soit $B \in \mathbb{R}^p$ et A une matrice $(p \times k)$. On pose $Y = AX + B$. Le vecteur aléatoire Y vérifie

- (1) $EY = AEX + B$
- (2) $\text{Var}(Y) = A\text{Var}(X) {}^tA$

Proposition 2.2. Soient X et Y deux vecteurs aléatoires.

- (i) Si X et Y sont indépendants alors $\text{Cov}(X, Y) = 0$
mais attention la réciproque est fausse
- (ii) Posons $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$. Sa matrice de covariance est donnée par

$$\begin{pmatrix} \text{Var}(X) & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & \text{Var}(Y) \end{pmatrix}$$

- (iii) Si X et Y sont des vecteurs de même dimension alors

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

Définition 2.2. Soit X un vecteur aléatoire. Sa fonction caractéristique est définie par

$$\Phi_X(u) = \mathbb{E}(e^{i \text{ }^t u X}) = \mathbb{E}\left(e^{i \sum_{j=1}^k u_j X_j}\right) \quad \forall u \in \mathbb{R}^k$$

Le résultat suivant assure que la loi d'une variable aléatoire est déterminée par sa fonction caractéristique.

Théorème 2.3 (unicité). Soient X et Y deux variables aléatoires. On note ϕ_X et ϕ_Y leurs fonctions caractéristiques. Si $\phi_X = \phi_Y$ alors les variables aléatoires X et Y admettent la même loi.

On vérifie facilement que si Y est une transformation affine du vecteur X c'est à dire $Y = AX + B$ avec $B \in \mathbb{R}^p$ et A une matrice ($p \times k$), les fonctions caractéristiques de X et Y sont liées par la relation suivante

$$\Phi_Y(v) = e^{i \text{ }^t v B} \Phi_X(\text{ }^t A v)$$

pour tout $v \in \mathbb{R}^p$.

La propriété suivante donne une condition nécessaire et suffisante pour obtenir l'indépendance des coordonnées d'un vecteur.

Proposition 2.4. Soit $X = (X_1, \dots, X_k)$ un vecteur aléatoire. Les variables aléatoires (X_1, \dots, X_k) sont indépendantes si et seulement si

$$\Phi_X(u) = \prod_{i=1}^k \Phi_{X_i}(u_i) \quad \forall u = (u_1, \dots, u_k)$$

DÉMONSTRATION.

[\implies] Supposons que les coordonnées de X soient indépendantes. Pour tout $u \in \mathbb{R}^k$, on a

$$\Phi_X(u) := \mathbb{E}\left(e^{i \sum_{j=1}^k u_j X_j}\right) = \prod_{j=1}^k \mathbb{E}\left(e^{i u_j X_j}\right) = \prod_{j=1}^k \Phi_{X_j}(u_j)$$

[\impliedby] Réciproquement, pour tout $u \in \mathbb{R}^k$, on a

$$\Phi_X(u) = \prod_{i=1}^k \Phi_{X_i}(u_i).$$

Alors, on a, pour tout $u \in \mathbb{R}^k$

$$\begin{aligned} \int e^{i \sum_{j=1}^k u_j x_j} d\mathbf{P}_{X_1, \dots, X_k}(x_1, \dots, x_k) &= \prod_{j=1}^k \int e^{i u_j x_j} d\mathbf{P}_{X_j}(x_j) \\ &= \int e^{i \sum_{j=1}^k u_j x_j} \prod_{j=1}^k d\mathbf{P}_{X_j}(x_j) \end{aligned}$$

L'unicité de la fonction caractéristique (voir Théorème 2.3) implique que la loi du vecteur (X_1, \dots, X_k) est la mesure produit. D'où l'indépendance des coordonnées du vecteur X \square

2. Loïs gaussiennes dans \mathbb{R}

Définition 2.3. Soit X une variable aléatoire réelle. On dit que X est une variable aléatoire gaussienne de paramètres (μ, σ^2) avec $\mu \in \mathbb{R}$ et $\sigma \in \mathbb{R}^+$ (on note $X \sim \mathcal{N}(\mu, \sigma^2)$) si et seulement si X vérifie une des deux conditions suivantes

[I] $\sigma > 0$ et X admet pour densité

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) \quad (2.1)$$

[II] $\sigma = 0$ et X est presque sûrement égale à μ .

Remarque 2.1. Dans la situation [II], on parle de lois gaussiennes dégénérées. Dans ce cas, la variable aléatoire n'admet pas de densité par rapport à la mesure de Lebesgue. |

Rappelons quelques propriétés des lois gaussiennes. Soit $X \sim \mathcal{N}(\mu, \sigma^2)$, on a

- l'espérance de X vaut μ .
- la variance est notée $\text{Var}(X)$ et vaut σ^2 .
- la fonction caractéristique de X est égale à

$$\phi_X(t) = \mathbb{E}(e^{itX}) = e^{it\mu} e^{-\frac{t^2\sigma^2}{2}} \quad \forall t \in \mathbb{R}.$$

- si $\sigma > 0$ alors $\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$. Cette loi gaussienne est dite standard
- plus généralement, on a

$$aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

pour tout $(a, b) \in \mathbb{R}^2$.

Dans la suite du chapitre, on utilisera les notations suivantes

- ◊ tA pour la transposée de A
- ◊ I_n pour la matrice identité de dimension $n \times n$
- ◊ $0_{n,m}$ pour la matrice nulle de dimension $n \times m$
- ◊ $\text{diag}(a_1, \dots, a_p)$ désigne la matrice diagonale dont la diagonale est égale à (a_1, \dots, a_p) , c'est à dire

$$\text{diag}(a_1, \dots, a_p) = \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_p \end{pmatrix} = \begin{pmatrix} a_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & a_p \end{pmatrix}$$

3. Vecteurs Gaussiens

Définition 2.4. Un vecteur aléatoire X à valeurs dans \mathbb{R}^k est un vecteur gaussien si et seulement si toute combinaison linéaire de ses coordonnées est une variable aléatoire gaussienne i.e. $\forall \lambda, \sum_{j=1}^k \lambda_j X_j$ suit une loi gaussienne (éventuellement dégénérée).

Remarque 2.2. Si X est un vecteur gaussien alors pour toute partie $\{i_1, \dots, i_p\}$ de $\{1, \dots, k\}$, le vecteur $(X_{i_1}, \dots, X_{i_p})$ est gaussien. |

Proposition 2.5. Si ϕ est une application linéaire de \mathbb{R}^k dans \mathbb{R}^m et si X est un vecteur gaussien de dimension k alors $\phi(X)$ est aussi un vecteur gaussien de dimension m .

DÉMONSTRATION. Il suffit de remarquer que toute combinaison linéaire des coordonnées de $\phi(X)$ est aussi une combinaison linéaire des coordonnées de X . \square

On peut aussi définir un vecteur gaussien à partir de sa fonction de caractéristique. On a l'équivalence suivante :

Théorème 2.6. Un vecteur aléatoire X à valeurs dans \mathbb{R}^k est un vecteur gaussien si et seulement si $X \in L^2$ et il admet pour fonction caractéristique

$$\Phi_X(u) = e^{i {}^t u \mu - \frac{1}{2} {}^t u \Sigma u} \quad \forall u \in \mathbb{R}^k \quad (2.2)$$

avec $\mu = \mathbb{E}(X)$ et $\Sigma = \text{Var}(X)$.

DÉMONSTRATION.

(\implies) Par définition, on a

$$\Phi_X(u) = \mathbb{E}(e^{i {}^t u X}) = \mathbb{E}(e^{i Y}) = \Phi_Y(1)$$

où $Y = \sum_{i=1}^k u_i X_i$. La variable aléatoire Y est une variable aléatoire gaussienne puisque c'est une combinaison linéaire des coordonnées du vecteur gaussien X .

On a donc

$$\Phi_Y(1) = e^{i\mathbb{E}(Y)} e^{-\frac{\text{Var}(Y)}{2}}$$

avec

$$\mathbb{E}(Y) = \mathbb{E}\left(\sum_{i=1}^k u_i X_i\right) = \sum_{i=1}^k u_i \mathbb{E}(X_i) = {}^t u \mu$$

et

$$\begin{aligned} \text{Var}(Y) &= \text{Var}\left(\sum_{i=1}^k u_i X_i\right) = \sum_{i=1}^k \sum_{j=1}^k \text{Cov}(u_i X_i, u_j X_j) \\ &= \sum_{i=1}^k \sum_{j=1}^k u_i u_j \text{Cov}(X_i, X_j) = {}^t u \Sigma u. \end{aligned}$$

Ceci prouve que la fonction caractéristique du vecteur X est bien de la forme annoncée.

(\Leftarrow) Soit ${}^t \Lambda = (\lambda_1, \dots, \lambda_k)$. On pose $Y = \sum_{i=1}^k \lambda_i X_i = {}^t \Lambda X$. Montrons que Y est une variable aléatoire gaussienne. Pour tout $u \in \mathbb{R}$, la fonction caractéristique de Y est donnée par

$$\begin{aligned} \Phi_Y(u) &= \mathbb{E}(e^{iu} {}^t \Lambda X) = \Phi_X(u \Lambda) \\ &= e^{i {}^t (u \Lambda) \mu - \frac{1}{2} {}^t (u \Lambda) \Sigma u \Lambda} \\ &= e^{iu {}^t \Lambda \mu - \frac{1}{2} u^2 {}^t \Lambda \Sigma \Lambda} \end{aligned}$$

On reconnaît la fonction caractéristique de la loi gaussienne $\mathcal{N}({}^t \Lambda \mu; {}^t \Lambda \Sigma \Lambda)$. Donc Y est bien une variable aléatoire gaussienne. \square

Remarque 2.3. Cette propriété nous montre que la loi d'un vecteur gaussien est entièrement déterminée par son espérance et sa matrice de covariance.

Dans la suite, on utilisera la notation $X \sim \mathcal{N}_k(\mu, \Sigma)$ pour désigner un vecteur gaussien de dimension k , d'espérance μ et de matrice de covariance Σ . $|$

Exemple 2.1. Soient X_1, \dots, X_p des variables aléatoires gaussiennes **indépendantes**. On suppose que $X_i \sim \mathcal{N}(m_i, \sigma_i^2)$ pour tout $i \in 1, \dots, p$. A partir du Théorème 2.6, on montre facilement que $X = {}^t (X_1, \dots, X_p)$ est un vecteur gaussien d'espérance ${}^t (\mathbb{E}X_1, \dots, \mathbb{E}X_p)$ de sa matrice de covariance est diagonale et égale à $\text{diag}(\sigma_1^2, \dots, \sigma_p^2)$. Attention : l'exemple 2.2 nous fournit un exemple de vecteur donc les coordonnées sont gaussiennes mais le vecteur n'est pas gaussien. \parallel

Remarque 2.4. Soit A une matrice de dimension $p \times k$ et X un vecteur gaussien, $X \sim \mathcal{N}_k(\mu, \Sigma)$. Le vecteur AX est un vecteur gaussien de dimension p d'après la Proposition 2.5. De plus, on a $X \sim \mathcal{N}_p(A\mu, A\Sigma {}^t A)$. $|$

4. Vecteur gaussien et indépendance

Lorsque ${}^t X = {}^t (X_1, \dots, X_k)$ est un vecteur aléatoire, L^2 , nous avons vu que si les coordonnées (X_1, \dots, X_k) sont indépendantes alors la matrice de covariance est diagonale car $\text{cov}(X_i, X_j) = 0$ pour $i \neq j$. La réciproque est **fausse** en général. Pour les vecteurs gaussiens, nous avons le résultat suivant

Proposition 2.7. *Soit X un vecteur gaussien de dimension k , de moyenne μ et de covariance Σ . Les variables aléatoires (X_1, \dots, X_k) sont indépendantes si et seulement si la matrice Σ est diagonale.*

DÉMONSTRATION. Notons $\sigma_{i,j}$ les éléments de la matrice Σ .

D'après la proposition 2.4, les variables aléatoires (X_1, \dots, X_k) sont indépendantes si seulement si

$$\Phi_X(u) = \prod_{i=1}^k \Phi_{X_i}(u_i) \quad \forall u \in \mathbb{R}^k. \quad (2.3)$$

Or la fonction caractéristique de X s'écrit

$$\Phi_X(u) = e^{i {}^t u \mu - \frac{1}{2} {}^t u \Sigma u}$$

et le terme de droite dans (2.3) s'exprime comme

$$\prod_{i=1}^k \Phi_{X_i}(u_i) = \prod_{j=1}^k e^{i \mu_j u_j - \frac{1}{2} u_j^2 \sigma_{j,j}} = e^{i {}^t u \mu} e^{-\frac{1}{2} \sum_{j=1}^k u_j^2 \sigma_{j,j}}.$$

De plus, on a $\sum_{j=1}^k u_j^2 \sigma_{j,j} = {}^t u D u$ en prenant D la matrice diagonale construite à partir de la diagonale de Σ *i.e.*

$$D = \begin{pmatrix} \sigma_{1,1} & & \\ & \ddots & \\ & & \sigma_{k,k} \end{pmatrix}.$$

L'égalité (2.3) est donc équivalente à $\Sigma = D$. D'où l'indépendance des coordonnées de X si et seulement si $\Sigma = D$. □

Exemple 2.2. Soit X une variable aléatoire gaussienne standard et soit ε une variable aléatoire de loi

$$\mathbf{P}(\varepsilon = 1) = \mathbf{P}(\varepsilon = -1) = 1/2.$$

Supposons que X et ε soient des variables aléatoires indépendantes. On pose $Y = \varepsilon X$.

- Y est une variable aléatoire gaussienne.

En effet la fonction caractéristique de Y est donnée par

$$\begin{aligned} \mathbb{E}(e^{itY}) &= \mathbb{E}(e^{it\varepsilon X}) \\ &= \int \left(\sum_{k \in \{-1,1\}} e^{itkx} \mathbf{P}(\varepsilon = k) \right) d\mathbf{P}_X(x) \quad (\text{indépendance et th. de Fubini}) \\ &= \frac{1}{2} \mathbb{E}e^{itX} + \frac{1}{2} \mathbb{E}e^{-itX} \\ &= \frac{1}{2} (\phi_X(t) + \phi_X(-t)) = e^{-\frac{1}{2}t^2} \quad (\text{car } \phi_X(t) = e^{-\frac{1}{2}t^2}). \end{aligned}$$

- $\text{Cov}(X, Y) = 0$

En effet, la covariance est égale à

$$\begin{aligned} \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) &= \mathbb{E}(XY) && (\text{car } \mathbb{E}(X) = 0) \\ &= \mathbb{E}(X^2\varepsilon) = \mathbb{E}(X^2)\mathbb{E}(\varepsilon) && (\text{via l'indépendance}) \\ &= 0 && (\text{car } \mathbb{E}(\varepsilon) = 0) \end{aligned}$$

- $|X| = |\varepsilon||Y| = |Y|$ donc les variables aléatoires X et Y ne sont pas indépendantes.

Ici, les variables aléatoires X et Y sont gaussiennes mais le vecteur (X, Y) lui n'est pas un vecteur gaussien. ||

Proposition 2.8. Soit X un vecteur gaussien écrit de la forme (Y, Z) avec $Y \in \mathbb{R}^p$ et $Z \in \mathbb{R}^q$. Les vecteurs Y et Z sont indépendants si et seulement si la matrice de covariance de X est diagonale par blocs c'est à dire

$$\begin{pmatrix} A & 0_{p,q} \\ 0_{q,p} & B \end{pmatrix}$$

avec A une matrice de dimension $p \times p$ et B une matrice de dimension $q \times q$.

DÉMONSTRATION. à faire en exercice □

5. Existence

Proposition 2.9. *Soient Σ une matrice symétrique semi définie positive et μ un vecteur de \mathbb{R}^k . Il existe une (unique) loi de probabilité \mathcal{N} telle que*

$$\int e^{i \text{ }^t u x} d\mathcal{N}(x) = e^{i \text{ }^t \mu u - \frac{1}{2} \text{ }^t u \Sigma u}$$

pour tout $u \in \mathbb{R}^k$.

DÉMONSTRATION. La matrice Σ étant symétrique est diagonalisable dans une base orthonormale. Par conséquent, il existe une matrice P telle que

$$P \text{ }^t P = I_k \quad \text{et} \quad \text{ }^t P \Sigma P = \Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_k \end{pmatrix}$$

Les λ_i sont positifs ($\lambda_j \geq 0$) car la matrice Σ est semi définie positive.

On définit le vecteur Y dont les coordonnées Y_j sont indépendantes et tel que, pour tout j , Y_j est distribuée suivant la loi gaussienne $\mathcal{N}(0, \lambda_j)$.

La fonction caractéristique de Y s'écrit alors

$$\Phi_Y(t) = \prod_{j=1}^k e^{-\frac{1}{2} \lambda_j t_j^2}$$

C'est donc un vecteur gaussien centré et de matrice de covariance Λ .

On pose $X = PY + \mu$. Le vecteur X est gaussien. En effet toute combinaison linéaire des composantes de X s'écrit comme la somme d'une constante et d'une combinaison linéaire des composantes de Y . Comme Y est un vecteur gaussien c'est aussi la somme d'une constante et d'une variable gaussienne qui reste une variable gaussienne.

De plus, sa moyenne vaut $P\mathbb{E}Y + \mu = \mu$ et sa matrice de covariance égale à

$$\text{Var}(PY + \mu) = \text{Var}(PY) = P\text{Var}(Y) \text{ }^t P = P\Lambda \text{ }^t P = \Sigma.$$

□

6. Densité d'un vecteur gaussien

On a montré dans la section 5 que la loi normale $\mathcal{N}_k(\mu, \Sigma)$ est la mesure image de la loi $\otimes_{j=1}^k \mathcal{N}(0, \lambda_j)$ par la transformation $y \mapsto Py + \mu$.

Situation 1 : Si pour tout j , $\lambda_j \neq 0$, la mesure $\otimes_{j=1}^k \mathcal{N}(0, \lambda_j)$ admet pour densité

$$\prod_{j=1}^k \frac{1}{\sqrt{2\pi\lambda_j}} \exp\left(-\frac{1}{2} \frac{y_j^2}{\lambda_j}\right) = \frac{1}{(2\pi)^{k/2} \prod_{i=1}^k \lambda_i} \exp\left(-\frac{1}{2} \sum_{i=1}^k \frac{y_i^2}{\lambda_i}\right).$$

Par conséquent, la mesure image par la transformation $x = Py + \mu$ admet aussi une densité égale à

$$\frac{1}{(2\pi)^{k/2} \det(\Sigma)^{1/2}} e^{(-\frac{1}{2} \text{ }^t (x-\mu) \Sigma^{-1} (x-\mu))}. \quad (2.4)$$

en effet

- 1) Le jacobien de la transformation $y \mapsto Py + \mu$ est égal à $|\det(P)| = 1$.
- 2) Le terme $\prod_{i=1}^k \lambda_i$ est égal à $\det(\Sigma)$.

3) Finalement, on a

$$\begin{aligned} \sum_{i=1}^k \frac{y_i^2}{\lambda_i} &= {}^t y \Lambda^{-1} y \\ &= {}^t ({}^t P(x - \mu)) \Lambda^{-1} ({}^t P(x - \mu)) \\ &= {}^t (x - \mu) P \Lambda^{-1} {}^t P(x - \mu) = {}^t (x - \mu) \Sigma^{-1} (x - \mu) \end{aligned}$$

Situation 2 : On note $\lambda_1, \dots, \lambda_q$ les valeurs propres non nulles ($q < k$). La loi de Y s'écrit comme $Q_1 \otimes Q_2$ où Q_1 admet pour densité

$$\prod_{j=1}^q \frac{1}{\sqrt{2\pi\lambda_j}} \exp\left(-\frac{1}{2\lambda_j} y_j^2\right)$$

et Q_2 est la mesure de Dirac au point $(0, \dots, 0) \in \mathbb{R}^{k-p}$. La loi image de $Q_1 \otimes Q_2$ par la transformation $y \mapsto Py + \mu$ n'admet pas de densité par rapport à la mesure de Lebesgue sur \mathbb{R}^k . En effet, le support de la loi est l'image par $y \mapsto Py + \mu$ du sous espace vectoriel $\mathbb{R}^q \times 0^{k-q}$ de dimension q . La loi est concentrée sur un sous espace affine de dimension q dans \mathbb{R}^k qui est de mesure de Lebesgue nulle.

Dans ce cas, on a

$$\begin{aligned} \mathbb{E}[h(X)] &= \int h(PY + \mu) dQ_1 \otimes Q_2 \\ &= \int h(P {}^t(y_1, \dots, y_q, 0, \dots, 0) + \mu) \prod_{j=1}^q \frac{1}{\sqrt{2\pi\lambda_j}} \exp\left(-\frac{1}{2\lambda_j} y_j^2\right) dy_1 \cdots dy_q \end{aligned}$$

7. Vecteur gaussien et conditionnement

Soit X un vecteur gaussien de dimension $k \geq 2$. On pose H_1^{k-1} le sous espace des fonctions affines engendré par X_1, \dots, X_{k-1} . On note $P_{H_1^{k-1}}(X_k)$ la projection orthogonale (dans L^2) de X_k sur H_1^{k-1} .

Théorème 2.10. *Soit X un vecteur gaussien d'espérance μ et de matrice de covariance Σ . On a les propriétés suivantes :*

Propriété 1. La variable aléatoire $X_k - P_{H_1^{k-1}}(X_k)$ est indépendante de (X_1, \dots, X_{k-1}) , elle suit une loi gaussienne centrée et de variance

$$\sigma^2 = \mathbb{E}(X_k - P_{H_1^{k-1}}(X_k))^2 = \|X_k - P_{H_1^{k-1}}(X_k)\|_2^2$$

Propriété 2. $P_{H_1^{k-1}}(X_k) = \mathbb{E}(X_k | X_1, \dots, X_{k-1})$

Propriété 3. La loi conditionnelle de X_k sachant (X_1, \dots, X_{k-1}) est gaussienne d'espérance $P_{H_1^{k-1}}(X_k)$ et de variance $\|X_k - P_{H_1^{k-1}}(X_k)\|_2^2$.

La variance ne dépend donc pas de (X_1, \dots, X_{k-1}) .

DÉMONSTRATION.

• Preuve de la propriété 1.

La variable aléatoire $X_k - P_{H_1^{k-1}}(X_k)$ est orthogonale à H_1^{k-1} . En particulier, elle est orthogonale à $\mathbb{I}_{\mathbb{R}}$ c'est à dire

$$\mathbb{E}((X_k - P_{H_1^{k-1}}(X_k)) \mathbb{I}_{\mathbb{R}}) = \mathbb{E}((X_k - P_{H_1^{k-1}}(X_k))) = 0.$$

La variable aléatoire $X_k - P_{H_1^{k-1}}(X_k)$ est donc centrée.

Posons $Z = (X_1, \dots, X_{k-1}, X_k - P_{H_1^{k-1}}(X_k))$. Par construction, la variable aléatoire $P_{H_1^{k-1}}(X_k) \in H_1^{k-1}$ donc elle s'écrit sous la forme $\alpha_0 + \alpha_1 X_1 + \dots + \alpha_{k-1} X_{k-1}$. On peut donc exprimer Z comme

une transformation affine de X c'est à dire

$$Z = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -\alpha_0 \end{pmatrix} + \begin{pmatrix} & & & \\ & I_p & & \\ & \cdots & & \\ & & -\alpha_{k-1} & \\ & & & 0_{p,1} \\ & & & & 1 \end{pmatrix} X$$

Le vecteur Z est donc un vecteur gaussien. Sa matrice de covariance de Z est de la forme

$$\tilde{\Sigma} = \begin{pmatrix} \Sigma_{k-1} & 0_{k-1,1} \\ 0_{1,k-1} & \sigma^2 \end{pmatrix} \text{ avec } \Sigma_{k-1} = \text{Var}(X_1, \dots, X_{k-1})$$

et $\sigma^2 = \text{Var}(X_k - P_{H_1^{k-1}}(X_k)) = \|X_k - P_{H_1^{k-1}}(X_k)\|_2^2$ car la variable aléatoire $X_k - P_{H_1^{k-1}}(X_k)$ est centrée. La matrice $\tilde{\Sigma}$ étant diagonale par blocs, on peut conclure que $X_k - P_{H_1^{k-1}}(X_k)$ est indépendante de (X_1, \dots, X_{k-1}) . De plus, elle suit une loi gaussienne d'espérance 0 et de variance σ^2 .

• **Preuve de la propriété 2.**

La variable aléatoire $X_k - P_{H_1^{k-1}}(X_k)$ est indépendante des variables aléatoires X_1, \dots, X_{k-1} et donc de toute fonction $\phi(X_1, \dots, X_{k-1})$. Par conséquent, $X_k - P_{H_1^{k-1}}(X_k)$ est orthogonal à $L^2(\Omega, \sigma(X_1, \dots, X_{k-1}), \mathbf{P})$ et $P_{H_1^{k-1}}(X_k)$ est aussi la projection orthogonale de X_k sur l'espace $L^2(\Omega, \sigma(X_1, \dots, X_{k-1}), \mathbf{P})$. D'où le résultat.

• **Preuve de la propriété 3.** La preuve repose sur le lemme suivant

Lemme 2.11. *Si $X = g(Y) + Z$ avec Z et Y indépendantes alors la loi de X conditionnellement à Y est la loi de Z (notée \mathbf{P}_Z) translatée par le vecteur $g(Y)$ i.e. $\mathbf{P}(X \in A | Y = y) = \mathbf{P}_Z(A - g(y))$*

DÉMONSTRATION DU LEMME. D'abord la mesurabilité s'obtient en remarquant que $\mathbf{P}_Z(A - g(Y)) = \int \mathbb{I}_A(Z + g(Y)) d\mathbf{P}_Z$ est mesurable. Puis, pour tout B , on a

$$\begin{aligned} \int_B \mathbf{P}_Z(A - g(y)) d\mathbf{P}_Y(y) &= \int \mathbb{I}_B(y) \mathbb{I}_{A-g(y)}(z) d\mathbf{P}_Y(y) d\mathbf{P}_Z(z) \\ &= \mathbf{P}(Y \in B, Z - g(Y) \in A) = \mathbf{P}(Y \in B, X \in A) \end{aligned}$$

□

• **suite de la preuve de la propriété 3.**

On décompose la variable aléatoire X_k de la façon suivante :

$$X_k = P_{H_1^{k-1}}(X_k) + X_k - P_{H_1^{k-1}}(X_k)$$

avec $P_{H_1^{k-1}}(X_k) = g(X_1, \dots, X_{k-1})$ et $X_k - P_{H_1^{k-1}}(X_k)$ indépendantes. D'après le Lemme 2.11 la loi conditionnelle de X_k sachant (X_1, \dots, X_{k-1}) est la loi de $X_k - P_{H_1^{k-1}}(X_k)$ translatée par le vecteur $P_{H_1^{k-1}}(X_k)$. C'est donc la loi gaussienne $\mathcal{N}(0, \sigma^2)$ translatée par le vecteur $P_{H_1^{k-1}}(X_k)$ ce qui donne la loi gaussienne d'espérance $P_{H_1^{k-1}}(X_k)$ et de variance σ^2 .

□

Exemple 2.3. Soit $X = (X_1, X_2, X_3)$ un vecteur gaussien centré et de variance

$$\begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}$$

La loi conditionnelle de X_1 sachant (X_2, X_3) est une variable aléatoire gaussienne de moyenne $\mathbb{E}(X_1 | X_2, X_3)$ de la forme $\alpha + \beta X_2 + \gamma X_3$ telle que

$$X_1 - (\alpha + \beta X_2 + \gamma X_3) \perp \begin{cases} 1 \\ X_2 \\ X_3 \end{cases}$$

On obtient

$$\begin{cases} \alpha = 0 \\ \beta + \rho\gamma = \rho \\ \beta\rho + \gamma = \rho^2 \end{cases} \implies \begin{cases} \alpha = 0 \\ \gamma = 0 \\ \beta = \rho \end{cases}$$

i.e. $\mathbb{E}(X_1|X_2, X_3) = \rho X_2$ De plus la variance de la loi conditionnelle est égale à

$$\sigma^2 = \|X_1 - \rho X_2\|_2^2 = 1 - \rho^2.$$

||

8. Projection d'un vecteur gaussien

Définition 2.5. Fixons $k \in \mathbb{N}^*$. La loi du χ^2 à k degrés de liberté (notée $\chi^2(k)$) est la loi de probabilité qui admet pour densité

$$f_k(x) = e^{-\frac{x}{2}} x^{\frac{k}{2}-1} \frac{1}{\Gamma(k/2)\sqrt{2}^k} \mathbb{I}_{\mathbb{R}_+^*}(x).$$

Rappelons que la fonction Γ est définie par

$$\Gamma(z) = \int_{\mathbb{R}_+} e^{-x} x^{z-1} dx \quad \forall z > 0.$$

Elle vérifie les propriétés suivantes

- (1) pour tout $z > 0$, $\Gamma(z+1) = z\Gamma(z)$.
- (2) pour tout $n \in \mathbb{N}$, $\Gamma(n+1) = n!$.
- (3) $\Gamma(1/2) = \sqrt{\pi}$

Théorème 2.12. Soit (X_1, \dots, X_k) une suite de variables aléatoires indépendantes et de même loi gaussienne standard i.e. pour tout i , $X_i \sim \mathcal{N}(0, 1)$.

(1) Les variables aléatoires (X_1^2, \dots, X_n^2) sont indépendantes et identiquement distribuées suivant la loi du χ^2 à 1 degré de liberté : $\chi^2(1)$.

(2) $\|X\|^2 = \sum_{i=1}^k X_i^2$ suit une loi du χ^2 à k ddl : $\chi^2(k)$.

DÉMONSTRATION. A faire en exercice. □

Théorème 2.13 (Théorème de Cochran). Soit E_1 et E_2 deux sous espaces orthogonaux supplémentaires de \mathbb{R}^k . $X = (X_1, \dots, X_k)$ un vecteur gaussien centré de covariance I_k . On note Z_1 (resp. Z_2) la projection orthogonale de X sur E_1 (resp. E_2).

Propriété 1 Les variables aléatoires Z_1 et Z_2 sont indépendantes

Propriété 2 $\|Z_1\|^2$ (resp. $\|Z_2\|^2$) suit une loi du χ^2 à p (resp. q) degrés de liberté avec $p+q=k$.

DÉMONSTRATION.

• Preuve de la propriété 1

On note p (resp. q) la dimension du sous espace E_1 (resp. E_2) avec $p+q=k$.

La matrice de projection orthogonale sur E_1 (resp. E_2) s'écrit

$$S_1 = P \begin{pmatrix} I_p & 0_{p,q} \\ 0_{q,p} & 0_{q,q} \end{pmatrix} {}^tP \quad (\text{resp. } S_2 = P \begin{pmatrix} 0_{p,p} & 0_{p,q} \\ 0_{q,p} & I_q \end{pmatrix} {}^tP.)$$

où P est une matrice orthogonale (i.e. $P {}^tP = {}^tPP = I_k$). Les matrices S_1 et S_2 vérifient les propriétés suivantes :

- (1) elles sont symétriques
- (2) $S_i^2 = S_i$ pour $i = 1, 2$
- (3) $S_1 S_2 = S_2 S_1 = 0_{k,k}$.

Posons

$$Z := \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} S_1 X \\ S_2 X \end{pmatrix} = SX \quad \text{avec } S = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}$$

Comme transformation linéaire du vecteur gaussien X , le vecteur aléatoire Z est un vecteur gaussien. Son espérance est égale à

$$\mathbb{E}Z = S\mathbb{E}X = 0$$

et sa matrice de covariance est égale à

$$\begin{aligned} \text{Var}(Z) &= S\text{Var}(X) {}^tS = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix} \begin{pmatrix} {}^tS_1 & {}^tS_2 \end{pmatrix} \\ &= \begin{pmatrix} S_1 {}^tS_1 & 0 \\ 0 & S_2 {}^tS_2 \end{pmatrix} = \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \end{aligned}$$

La matrice de covariance étant diagonale par blocs, les variables aléatoires Z_1 et Z_2 sont indépendantes (voir la proposition 2.7). De plus, le vecteur Z_1 (resp. Z_2) est un vecteur gaussien centré et ayant pour matrice de covariance S_1 (resp. S_2).

• **Preuve de la propriété 2** Considérons maintenant la projection sur E_1 . On définit le vecteur

$$W = {}^tPZ_1 (= {}^tPS_1X).$$

Le vecteur aléatoire W est un vecteur gaussien centré et sa matrice de covariance est

$$\text{Var}(W) = {}^tP\text{Var}(Z_1)P = {}^tPS_1P = \begin{pmatrix} I_p & 0_{p,q} \\ 0_{q,p} & 0_{q,q} \end{pmatrix}$$

Les variables aléatoires W_{p+1}, \dots, W_k sont nulles presque sûrement. Le carré de la norme de W s'écrit donc

$$\|W\|^2 = \sum_{i=1}^p W_i^2.$$

avec W_1, \dots, W_p indépendantes d'après la proposition 2.7. La loi de $\|W\|^2$ est donc donnée par le théorème 2.12, c'est une loi du χ^2 à p degré(s) de liberté. D'autre part, $\|W\|^2$ s'écrit aussi

$$\|W\|^2 = {}^tWW = {}^t(P S_1 X)(P S_1 X) = {}^tX {}^tS_1 {}^tP P S_1 X = {}^tX {}^tS_1 S_1 X = \|S_1 X\|^2 = \|Z_1\|^2.$$

Ceci termine la preuve. □

Dans la proposition suivante nous donnons une application du théorème de Cochran.

Proposition 2.14. *Soit X_1, \dots, X_n une suite de variables aléatoires gaussiennes indépendantes et de même loi $\mathcal{N}(0, 1)$. Les variables aléatoires $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ (appelée moyenne empirique) et $\sum_{i=1}^n (X_i - \bar{X}_n)^2$ sont indépendantes. De plus, la variable aléatoire $\sum_{i=1}^n (X_i - \bar{X}_n)^2$ suit une loi du chi-deux à $n - 1$.*

DÉMONSTRATION. Soit E_1 le sous espace de \mathbb{R}^n engendré par $\delta = \frac{1}{\sqrt{n}} {}^t(1, \dots, 1)$. La projection orthogonale de (X_1, \dots, X_n) sur E_1 est de la forme $\lambda\delta$ avec λ tel que $(X_1 - \frac{\lambda}{\sqrt{n}}, \dots, X_n - \frac{\lambda}{\sqrt{n}})$ soit orthogonal à δ , c'est-à-dire, $\lambda = \sqrt{n}\bar{X}_n$.

Calculons maintenant la projection orthogonale du vecteur ${}^tX = (X_1, \dots, X_n)$ sur le sous espace E_2 qui est orthogonal à E_1 et tel que E_1 et E_2 soient supplémentaires. Elle s'écrit $v = {}^t(v_1, \dots, v_n)$ avec $\sum_{i=1}^n v_i = 0$ (car v et δ sont orthogonaux) et ${}^t(X_1 - v_1, \dots, X_n - v_n) = \alpha\delta$ (car $X - v$ appartient à E_1). D'où $\alpha = -\bar{X}_n$ et $v = {}^t(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)$. Le vecteur ${}^t(X_1, \dots, X_n)$ étant gaussien (voir remarque 2.2), le théorème 2.13 assure l'indépendance des vecteurs aléatoires $\sqrt{n}\bar{X}_n\delta$ et $v = {}^t(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)$. Par conséquent, les variables aléatoires \bar{X}_n et $\sum_{i=1}^n (X_i - \bar{X}_n)^2$ sont indépendantes.

De plus, comme $\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \|v\|^2$, elle suit une loi du chi-deux à $n - 1$ degrés de liberté. □

Remarque 2.5. Lorsque les variables aléatoires X_i sont indépendantes et de même loi $\mathcal{N}(m, \sigma^2)$, $\sigma > 0$, la propriété d'indépendance reste vraie. De plus, si $\sigma^2 \neq 0$ alors

$$\sigma^{-2} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

suit une loi du chi-deux à $n - 1$ degrés de liberté. |

Convergences

1. Définition des modes de convergence

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires et soit X une variable aléatoire. On définit les modes de convergence suivants :

$X_n \xrightarrow{p.s.} X$: $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement vers X si et seulement si

$$\mathbf{P}(\omega : X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X(\omega)) = 1$$

$X_n \xrightarrow{P} X$: $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X si et seulement si

$$\forall \varepsilon > 0 \quad \mathbf{P}(|X_n - X| > \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0$$

$X_n \xrightarrow{L^p} X$: $(X_n)_{n \in \mathbb{N}}$ converge dans L^p ($p \geq 1$) vers X si et seulement si la variable aléatoire X et la suite $(X_n)_{n \in \mathbb{N}}$ sont dans L^p et vérifient

$$\mathbb{E}|X_n - X|^p \xrightarrow[n \rightarrow \infty]{} 0$$

2. Quelques propriétés

- Si $X_n \xrightarrow{P} X$ et $X_n \xrightarrow{P} Y$ alors $X = Y$ presque sûrement.
Les limites (si elles existent) sont unique presque sûrement.
- $X_n \xrightarrow{P} X$ implique $|X_n - X| \xrightarrow{P} 0$
- si $X_n \xrightarrow{P} X$ et $Y_n \xrightarrow{P} Y$ alors pour tout $(a, b) \in \mathbb{R}^2$ on a

$$aX_n + bY_n \xrightarrow{P} aX + bY$$

Ces propriétés sont aussi vérifiées par la convergence presque sûre et la convergence L^p en supposant les variables aléatoires X_n et que la limite X sont dans L^p .

Nous donnons maintenant les implication entre ces modes de convergence.

Théorème 3.1. *Entre ces trois modes de convergence, on a les implications suivantes*

$$(1) X_n \xrightarrow{p.s.} X \implies X_n \xrightarrow{P} X$$

$$(2) X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{P} X$$

$$(3) \text{ Soit } p \geq 1, X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{L^1} X$$

Remarque 3.1.

- 1- L'implication 2 est une conséquence immédiate de l'inégalité de Markov. Si les variables aléatoires X_n et la limite X sont dans L^p alors

$$\mathbf{P}(|X_n - X| > \varepsilon) \leq \frac{1}{\varepsilon^p} \mathbb{E}(|X_n - X|^p) \quad \forall \varepsilon > 0$$

- 2- L'implication 3 découle de l'inégalité de Hölder

$$\mathbb{E}(|X_n - X|) \leq (\mathbb{E}(|X_n - X|^p))^{\frac{1}{p}}$$

3- On peut résumer ce résultat par le graphique suivant pour $p \geq 1$

$$L^p \implies L^1 \implies \text{proba} \\ \uparrow \\ \text{p.s.}$$

Les implications qui ne figurent pas dans ce graphique sont fausses.

Proposition 3.2. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires et soit g une fonction continue.

(1) si $X_n \xrightarrow{p.s.} X$ alors $g(X_n) \xrightarrow{p.s.} g(X)$

(2) si $X_n \xrightarrow{P} X$ alors $g(X_n) \xrightarrow{P} g(X)$

DÉMONSTRATION.

La preuve pour la convergence p.s. est immédiate.

Montrons le résultat pour la convergence en probabilité. Fixons $A > 0$ et $\varepsilon > 0$. La fonction g étant continue, elle est uniformément continue sur tout compact donc en particulier sur $[-A, A]$. Par conséquent, il existe $\eta_{\varepsilon, A}$ tel que

$$|X| < A, \quad |X_n - X| \leq \eta_{\varepsilon, A} \implies |g(X_n) - g(X)| \leq \varepsilon$$

D'où

$$\mathbf{P}(|X| < A, \quad |X_n - X| \leq \eta_{\varepsilon, A}) \leq \mathbf{P}(|g(X_n) - g(X)| \leq \varepsilon)$$

$$\mathbf{P}(|X_n - X| \leq \eta_{\varepsilon, A}) - \mathbf{P}(|X| > A, \quad |X_n - X| \leq \eta_{\varepsilon, A}) \leq \mathbf{P}(|g(X_n) - g(X)| \leq \varepsilon)$$

$\mathbf{P}(|X| > A)$ converge vers zéro quand A tend vers $+\infty$. Fixons $\zeta > 0$, on peut donc choisir A_0 tel que

$$\mathbf{P}(|X| > A_0, \quad |X_n - X| \leq \eta_{\varepsilon, A_0}) \leq \mathbf{P}(|X| > A_0) \leq \zeta.$$

Ensuite, la convergence en probabilité de $(X_n)_{n \in \mathbb{N}}$ vers X assure l'existence de $m > 0$ tel que pour tout $n > m$

$$\mathbf{P}(|X_n - X| \leq \eta_{\varepsilon, A_0}) > 1 - \zeta$$

et donc

$$\mathbf{P}(|g(X_n) - g(X)| \leq \varepsilon) > 1 - 2\zeta$$

D'où le résultat. □

3. Compléments sur la convergence presque sûre

Définition 3.1. Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements, on définit les événements $\overline{\lim} A_n$ et $\underline{\lim} A_n$ de la façon suivante

$$\overline{\lim} A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k \qquad \underline{\lim} A_n = \bigcup_{n \in \mathbb{N}} \bigcap_{k \geq n} A_k$$

ou de façon équivalente,

- $x \in \overline{\lim} A_n$ si et seulement si il appartient à une infinité d'événements de la suite $(A_n)_{n \in \mathbb{N}}$
- $x \in \underline{\lim} A_n$ si et seulement si il appartient à tous les événements la suite $(A_n)_{n \in \mathbb{N}}$ sauf éventuellement un nombre fini.

On a les équivalences suivantes

$$\begin{aligned} X_n \xrightarrow{p.s.} X &\iff \mathbf{P}(\underline{\lim} \{|X_n - X| < \varepsilon\}) = 1 \quad \forall \varepsilon > 0 \\ &\iff \mathbf{P}(\overline{\lim} \{|X_n - X| > \varepsilon\}) = 0 \quad \forall \varepsilon > 0 \\ &\iff \lim_{n \rightarrow \infty} \downarrow \mathbf{P}\left(\bigcup_{m \geq n} \{|X_m - X| > \varepsilon\}\right) = 0 \quad \forall \varepsilon > 0 \\ &\iff \lim_{n \rightarrow \infty} \downarrow \mathbf{P}\left(\sup_{m \geq n} |X_m - X| > \varepsilon\right) = 0 \quad \forall \varepsilon > 0 \end{aligned}$$

Finalement, on obtient le critère suivant de convergence presque sûre :

$$X_n \xrightarrow{p.s.} X \iff \sup_{m \geq n} |X_m - X| \xrightarrow{P} 0$$

Un autre critère de convergence presque sûre est le critère de Cauchy : $(X_n)_{n \in \mathbb{N}}$ a une limite au sens de la convergence presque sûre si et seulement si pour tout ε positif

$$\mathbf{P}\left(\sup_{m, p \geq n} |X_m - X_p| \geq \varepsilon\right) \xrightarrow{n \rightarrow \infty} 0$$

ce qui est équivalent à

$$Z_n := \sup_{m, p \geq n} |X_m - X_p| \xrightarrow{P} 0. \quad (3.1)$$

Cette convergence est aussi équivalente à

$$\sup_{m \geq n} |X_m - X_n| \xrightarrow{P} 0$$

en effet

$$\begin{aligned} \sup_{m \geq n} |X_m - X_n| &\leq \sup_{m, p \geq n} |X_m - X_p| \\ &\leq \sup_{m, p \geq n} (|X_m - X_n| + |X_p - X_n|) \\ &\leq \sup_{m, p \geq n} |X_m - X_n| + \sup_{m, p \geq n} |X_p - X_n| \\ &\leq 2 \sup_{m \geq n} |X_m - X_n| \end{aligned}$$

4. Théorème de Borel-Cantelli.

Le résultat suivant est admis.

Théorème 3.3 (Théorème de Borel-Cantelli). *Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements.*

BC-1 Si $\sum_{k=1}^{\infty} \mathbf{P}(A_k) < \infty$ alors $\mathbf{P}(\overline{\lim} A_n) = 0$

BC-2 Si les événements $(A_n)_{n \in \mathbb{N}}$ sont indépendants et $\sum_{k=1}^{\infty} \mathbf{P}(A_k) = \infty$
alors $\mathbf{P}(\overline{\lim} A_n) = 1$

Les résultats suivants sont des applications “immédiates” de la propriété BC-1 du théorème de Borel Cantelli.

Proposition 3.4. *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires et X une variable aléatoire. Si pour tout $\varepsilon > 0$,*

$$\sum_{n=1}^{\infty} \mathbf{P}(|X_n - X| > \varepsilon) < \infty$$

alors $X_n \xrightarrow{p.s.} X$.

DÉMONSTRATION. On applique BC-1 aux événements $A_n = \{|X_n - X| > \varepsilon\}$. Ceci implique que

$$\mathbf{P}(\overline{\lim}\{|X_n - X| > \varepsilon\}) = 0$$

pour tout $\varepsilon > 0$. D'où la convergence presque sûre. \square

Proposition 3.5. *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires et X une variable aléatoire. Si $X_n \xrightarrow{P} X$ alors on peut extraire une sous suite $(X_{n_k})_{k \in \mathbb{N}}$ qui converge presque sûrement vers X*

DÉMONSTRATION. Comme $X_n \xrightarrow{P} X$, pour tout $\varepsilon > 0$, $P(|X_n - X| > \varepsilon)$ converge vers zéro. On prend $\varepsilon = 1/m$, avec $m \in \mathbb{N}$. Il existe un entier n_m tel que $n_m > n_{m-1}$ et

$$\mathbf{P}(|X_{n_m} - X| > \frac{1}{m}) < 2^{-m}.$$

La série $\sum_{j=1}^{\infty} \mathbf{P}(|X_{n_j} - X| > \frac{1}{j})$ est convergente, donc d'après le lemme de Borel Cantelli BC-1,

$$P(\Lambda) = 0 \text{ avec } \Lambda = \overline{\lim} \left\{ |X_{n_j} - X| > \frac{1}{j} \right\}.$$

Pour tout $\omega \in \Lambda^c$, il existe j_0 tel que pour tout $j > j_0$, $|X_{n_j}(\omega) - X(\omega)| < \frac{1}{j}$ la sous suite $(X_{n_k}(\omega))_{k \in \mathbb{N}}$ converge donc vers $X(\omega)$. Comme $P(\Lambda^c) = 1$, la sous suite $(X_{n_k})_{k \in \mathbb{N}}$ converge presque sûrement vers X . \square

Sommes de variables aléatoires

Dans ce chapitre, on considère $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires et on étudie le comportement asymptotique de la suite S_n définie par

$$S_n = \sum_{i=1}^n X_i$$

ou éventuellement de la suite renormalisée, par exemple $\frac{1}{n}S_n$.

1. Inégalité de Kolmogorov

Commençons par énoncer l'inégalité de Kolmogorov. Ce résultat est admis dans ce cours.

Théorème 4.1. *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes, centrées et de carré intégrable. On a, pour tout $a > 0$*

$$\mathbf{P}\left(\max_{1 \leq k \leq n} |S_k| > a\right) \leq \frac{\text{Var}(S_n)}{a^2} \quad (4.1)$$

La proposition suivante s'obtient grâce à cette inégalité.

Proposition 4.2. *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes, centrées et de carré intégrables.*

Si

$$\sum_{n=1}^{\infty} \text{Var}(X_n) < \infty$$

alors S_n admet une limite au sens de la convergence presque sûre i.e. la série $\sum_{n=1}^{\infty} X_n$ converge presque sûrement.

DÉMONSTRATION. On utilise le critère de Cauchy pour prouver cette convergence. Fixons $\varepsilon > 0$. On montre que

$$\mathbf{P}\left(\sup_{m > n} |S_m - S_n| > \varepsilon\right) \xrightarrow{n \rightarrow \infty} 0.$$

Puisque $S_m - S_n = \sum_{j=n+1}^m X_j$,

$$\sup_{m > n} |S_m - S_n| = \sup_{m > n} \left| \sum_{j=n+1}^m X_j \right| = \lim_{h \rightarrow \infty} \uparrow \max_{n+1 \leq m \leq n+h} \left| \sum_{j=n+1}^m X_j \right|$$

D'après l'inégalité de Kolmogorov, nous avons

$$\mathbf{P}\left(\max_{n+1 \leq m \leq n+h} \left| \sum_{j=n+1}^m X_j \right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} \text{Var}\left(\sum_{j=n+1}^{n+h} X_j\right) \quad (4.2)$$

Comme la suite d'événements $([\max_{n+1 \leq m \leq n+h} |S_m - S_n| > \varepsilon])_{h \geq 1}$ est croissante en h , on a

$$\mathbf{P}\left(\sup_{m > n} |S_m - S_n| > \varepsilon\right) = \mathbf{P}\left(\lim_{h \rightarrow \infty} \left\{ \max_{\substack{n+1 \leq m \\ m \leq n+h}} |S_m - S_n| > \varepsilon \right\}\right) = \lim_{h \rightarrow \infty} \mathbf{P}\left(\max_{\substack{n+1 \leq m \\ m \leq n+h}} |S_m - S_n| > \varepsilon\right)$$

Ensuite, en prenant la limite des deux termes de (4.2), on obtient

$$\begin{aligned} \mathbf{P}(\sup_{m>n} |S_m - S_n| > \varepsilon) &\leq \frac{1}{\varepsilon^2} \lim_{h \rightarrow \infty} \sum_{j=n+1}^{n+h} \text{Var}(X_j) \quad (\text{via l'indépendance}) \\ &\leq \frac{1}{\varepsilon^2} \sum_{j=n+1}^{\infty} \text{Var}(X_j) \xrightarrow{n \rightarrow \infty} 0 \quad (\text{par hypothèse}). \end{aligned}$$

Ceci conclut la preuve. \square

Exemple 4.1. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires L^2 indépendantes et identiquement distribuées et soit $(a_n)_n$ une suite déterministe. Si $\sum_{n \in \mathbb{N}} a_n^2 < \infty$ alors la série $\sum_{n \in \mathbb{N}} a_n X_n$ converge presque sûrement. \parallel

2. Loi Forte des Grands Nombres

Dans cette section, on étudie la convergence de la suite S_n/n . Rappelons que pour des variables aléatoires indépendantes et identiquement distribuées, la loi forte des grands nombres nous donne une condition nécessaire et suffisante pour l'existence de la limite au sens de la convergence presque sûre. Ce résultat ne sera pas démontré dans ce cours.

Théorème 4.3 (Loi Forte des Grands Nombres). *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et de même loi. $\frac{1}{n} \sum_{j=1}^n X_j$ a une limite p.s. si et seulement si $X_1 \in L^1$. Dans ce cas la limite est égale presque sûrement à $\mathbb{E}X_1$*

Nous regardons maintenant ce que devient ce résultat lorsque l'hypothèse d'indépendance est omise.

Définition 4.1. *Soit $(X_n)_{n \in \mathbb{Z}}$ une suite de variables aléatoires. La suite est dite stationnaire d'ordre 2 si elle vérifie les propriétés suivantes :*

- (1) pour tout n , $X_n \in L^2$
- (2) $\mathbb{E}(X_n)$ ne dépend pas de $n \in \mathbb{Z}$
- (3) pour tout $n \in \mathbb{Z}$ et tout $h \in \mathbb{Z}$, la covariance $\text{Cov}(X_n, X_{n+h})$ dépend uniquement de $|h|$.
On la note $\sigma(h) = \sigma(-h)$

Exemple 4.2. Une suite de variables aléatoires de L^2 indépendantes et identiquement distribuées est une suite stationnaire d'ordre 2. \parallel

Théorème 4.4. *Soit $(X_n)_{n \in \mathbb{Z}}$ une suite de variables aléatoires stationnaire d'ordre 2 telle que $\sum_{k \in \mathbb{Z}} |\sigma(k)| < \infty$. On a*

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p.s.} \mathbb{E}(X_1)$$

Lemme 4.5. *Sous les hypothèses du théorème, on a*

$$\frac{1}{n} \text{Var}\left(\sum_{i=1}^n X_i\right) \rightarrow \sigma = \sum_{k \in \mathbb{Z}} \sigma(k)$$

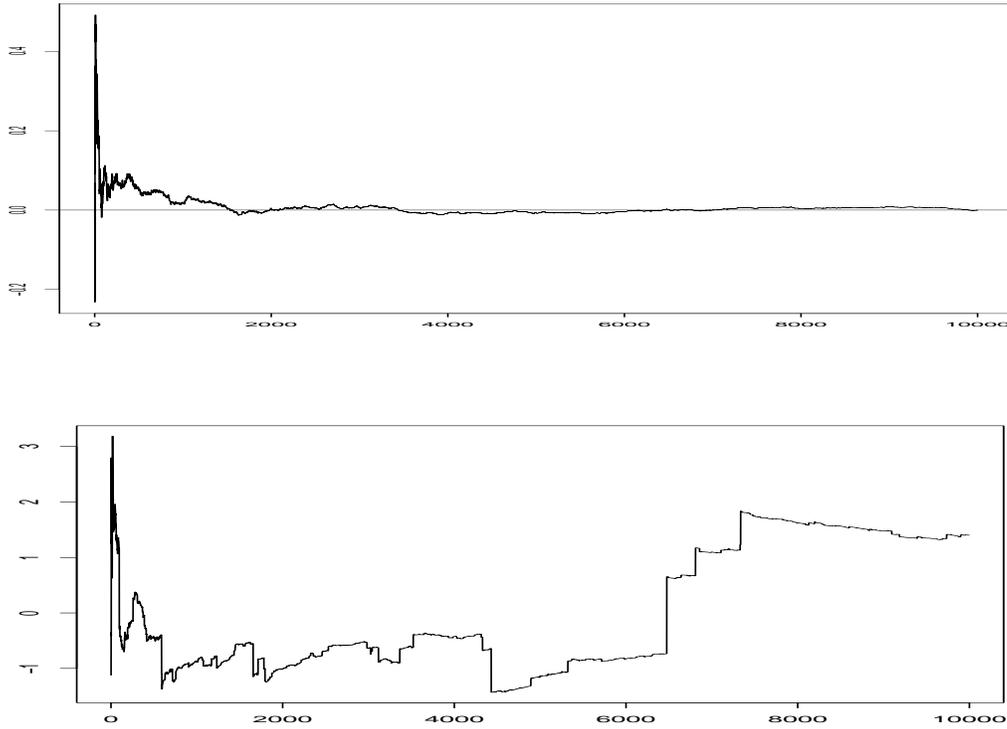


FIGURE 4.1. Représentation de $\frac{S_n}{n}$ en fonction de n : (*haut*) pour des variables aléatoires iid suivant la loi gaussienne standard $\mathcal{N}(0,1)$ (*bas*) pour des variables aléatoires iid suivant la loi de Cauchy c'est-à-dire la loi qui admet pour densité $x \mapsto \frac{1}{\pi} \frac{1}{1+x^2} \mathbb{I}_{\mathbb{R}}(x)$ la loi de Cauchy n'admet pas de moment d'ordre 1

DÉMONSTRATION. On a

$$\begin{aligned}
 \frac{1}{n} \text{Var}\left(\sum_{i=1}^n X_i\right) &= \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n \text{Cov}(X_j, X_i) \\
 &= \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n \sigma(j-i) && \text{(par définition de } \sigma) \\
 &= \frac{1}{n} (n\sigma(0) + 2(n-1)\sigma(1) + \cdots + 2\sigma(n-1)) && \text{(car } \sigma(h) = \sigma(-h)) \\
 &= \frac{1}{n} (s_0 + \cdots + s_{n-1}) && \text{(avec } s_k = \sum_{i=-k}^k \sigma(i))
 \end{aligned}$$

Par hypothèse s_n converge vers σ donc $\frac{1}{n} (s_1 + \cdots + s_n)$ converge vers la même limite. \square

DÉMONSTRATION. (retour à la preuve du théorème) On peut supposer que $\mathbb{E}(X_1) = 0$ (sinon on remplace les X_i par $X_i - \mathbb{E}(X_1)$). On pose

$$S_n = \sum_{i=1}^n X_i \quad \forall n \in \mathbb{N}$$

et

$$U_N = \max_{N^2 \leq k \leq (N+1)^2 - 1} \left| \frac{S_k}{k} \right| \quad \forall N \in \mathbb{N}$$

La convergence presque sûre de $\frac{S_n}{n}$ découle des résultats suivants :

- (1) pour tout $k \in \mathbb{N}$, $\left| \frac{S_k}{k} \right| \leq U_{[\sqrt{k}]}$
- (2) $U_N \xrightarrow{p.s.} 0$

Preuve de 1 Par définition de U_N , on a

$$U_N \geq \left| \frac{S_k}{k} \right| \quad \text{si } k \in \{N^2, \dots, (N+1)^2 - 1\}$$

c'est-à-dire si et seulement si

$$\begin{aligned} N^2 &\leq k \leq (N+1)^2 - 1 \\ &\iff N \leq \sqrt{k} < (N+1) \\ &\iff N = [\sqrt{k}] \end{aligned}$$

d'où le résultat.

Preuve de 2 Pour tout $k \in \mathbb{N}$, il existe $N \in \mathbb{N}$ tel que $N^2 \leq k \leq (N+1)^2 - 1$.

On peut donc décomposer S_k de la façon suivante :

$$\begin{aligned} \frac{S_k}{k} &= \frac{S_{N^2} + X_{N^2+1} + \dots + X_k}{k} = \frac{S_{N^2}}{k} + \frac{1}{k} \sum_{j=N^2+1}^k X_j \\ \left| \frac{S_k}{k} \right| &\leq \frac{|S_{N^2}|}{N^2} + \frac{1}{N^2} \left| \sum_{j=N^2+1}^k X_j \right| \quad (\text{car } k \geq N^2) \end{aligned}$$

D'où

$$\left| \frac{S_k}{k} \right| \leq \left| \frac{S_{N^2}}{N^2} \right| + \frac{1}{N^2} \sum_{j=N^2+1}^k |X_j| \leq \left| \frac{S_{N^2}}{N^2} \right| + \frac{1}{N^2} \sum_{j=N^2+1}^{(N+1)^2} |X_j|$$

Cette inégalité est vraie pour tout $k \in \{N^2, \dots, (N+1)^2 - 1\}$ donc

$$U_N \leq \left| \frac{S_{N^2}}{N^2} \right| + \frac{1}{N^2} \sum_{j=N^2+1}^{(N+1)^2} |X_j|$$

On obtient

$$\mathbb{E}(U_N^2) \leq \frac{2}{N^4} \mathbb{E}(S_{N^2}^2) + \frac{2}{N^4} \mathbb{E} \left(\sum_{j=N^2+1}^{(N+1)^2} |X_j| \right)^2$$

car $(a+b)^2 \leq 2(a^2 + b^2)$

Comme les X_i sont centrés, le lemme 4.5 assure la convergence de $\frac{1}{N^2} \mathbb{E}(S_{N^2}^2)$ vers σ . Pour le second terme, on a

$$\mathbb{E} \left(\sum_{j=N^2+1}^{(N+1)^2} |X_j| \right)^2 \leq \sum_{N^2+1 \leq \ell, j \leq (N+1)^2} \mathbb{E}|X_j X_\ell|$$

Grâce à l'inégalité de Hölder, on obtient

$$\mathbb{E}|X_j X_\ell| \leq \sqrt{\mathbb{E}(X_j^2) \mathbb{E}(X_\ell^2)} = \sigma(0) \quad (\text{car } \mathbb{E}(X_j^2) = \sigma(0)).$$

Finalement

$$\mathbb{E}\left(\sum_{j=N^2+1}^{(N+1)^2} |X_j|\right)^2 \leq \sum_{N^2+1 \leq \ell \leq (N+1)^2} \sum_{N^2+1 \leq j \leq (N+1)^2} \sigma(0) \leq (2N+1)^2 \sigma(0).$$

On a $\mathbb{E}U_N^2$ est un $O(N^{-2})$ donc la série $\sum_N \mathbb{E}(U_N^2)$ converge. Cette condition assure la convergence presque sûre de U_N vers 0 (voir Proposition 3.4). \square

3. Théorème de Glivenko-Cantelli

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées. Soit F la fonction de répartition de la loi commune.

On pose pour tout $x \in \mathbb{R}$ et tout $\omega \in \Omega$

$$F_n(x, \omega) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{]-\infty, x]}(X_i(\omega)) \quad (4.3)$$

On pose $F_n(x) = F_n(x, \cdot)$. La variable aléatoire F_n est appelée fonction de répartition empirique.

A x fixé, la loi forte des grands nombres assure la convergence presque sûre de $F_n(x)$ vers la fonction de répartition $F(x)$. En effet, $(\mathbb{I}_{]-\infty, x]}(X_i))_{i \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes intégrables et identiquement distribuées. De plus

$$\mathbb{E}(\mathbb{I}_{]-\infty, x]}(X_1)) = \mathbf{P}(X_1 \in]-\infty, x]) = F(x).$$

Par conséquent, pour tout $x \in \mathbb{R}$, il existe Ω_x tel que $\mathbf{P}(\Omega_x) = 1$ et pour tout $\omega \in \Omega_x$ $F_n(x, \omega) \rightarrow F(x)$. Le théorème suivant montre que sans hypothèse supplémentaire, on peut construire un ensemble de mesure 1, indépendant de x sur lequel $F_n(x)$ converge vers $F(x)$. Le résultat est le suivant

Théorème 4.6 (Théorème de Glivenko Cantelli). *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires tel que pour tout borélien A*

$$\frac{1}{n} \sum_{i=1}^n \mathbb{I}_A(X_i) \rightarrow \mathbf{P}(X \in A) = \int_A dF \quad (4.4)$$

où F représente la fonction de répartition. Sous ces conditions, on a

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{p.s.} 0$$

i.e. il existe $\tilde{\Omega}$ tel que $\mathbf{P}(\tilde{\Omega}) = 1$ et pour tout $\omega \in \tilde{\Omega}$

$$\sup_{x \in \mathbb{R}} |F_n(x, \omega) - F(x)| \xrightarrow{p.s.} 0$$

Remarque 4.1. Le résultat s'applique en particulier pour des variables aléatoires indépendantes et identiquement distribuées. \square

DÉMONSTRATION. Fixons $x \in \mathbb{R}$. En prenant $A =]-\infty, x]$ puis $]-\infty, x[$ dans l'équation (4.4), on obtient les convergences suivantes

$$F_n(x) \xrightarrow{p.s.} F(x) \quad F_n(x^-) \xrightarrow{p.s.} F(x^-). \quad (4.5)$$

Fixons $N \in \mathbb{N}$, pour tout $j \in \{0, 1, \dots, N\}$, on pose $x_{j,N} = \inf \{x : F(x) \geq \frac{j}{N}\}$ (on prend $x_{0,N} = -\infty$ et éventuellement $x_{N,N} = \infty$).

Il y a deux possibilités

– soit il existe x tel que $F(x) = \frac{j}{N}$. Dans ce cas, on a

$$F(x_{j,N}) = \frac{j}{N} \quad \text{et} \quad x_{j,N} = \inf \left\{ x : F(x) = \frac{j}{N} \right\}.$$

– soit il existe x_0 tel que

$$F(x_0^-) < \frac{j}{N} \quad \text{et} \quad F(x_0^+) > \frac{j}{N}.$$

Dans ce cas, on a

$$x_{j,N} = x_0 \quad \text{et} \quad F(x_{j,N}) > \frac{j}{N}.$$

Dans les deux cas, on a l'inégalité suivante

$$F(x_{j,N}^-) \leq \frac{j}{N} \leq F(x_{j,N}) \quad (4.6)$$

qui implique que

$$F(x_{j,N}) + \frac{1}{N} \geq F(x_{j+1,N}^-) \quad (4.7)$$

Maintenant, on définit les intervalles

- $\Delta_{j,N} = [x_{j,N}, x_{j+1,N}[$ pour $j = 1, \dots, N-1$,
- $\Delta_{0,N} =]-\infty, x_{1,N}[$
- $\Delta_{N,N} = \begin{cases} [x_{N,N}, \infty[& \text{si } x_{N,N} < \infty \\ \emptyset & \text{si } x_{N,N} = \infty. \end{cases}$

Pour tout $x \in \Delta_{j,N}$, on a

$$F_n(x_{j,N}) - F(x_{j+1,N}^-) \leq F_n(x) - F(x) \leq F_n(x_{j+1,N}^-) - F(x_{j,N})$$

car F_n et F sont des fonctions croissantes. Puis, l'inégalité (4.7) implique que

$$\underbrace{F_n(x_{j,N}) - F(x_{j,N})}_{=D_{n,j,N}} - \frac{1}{N} \leq F_n(x) - F(x) \leq \underbrace{F_n(x_{j+1,N}^-) - F(x_{j+1,N}^-)}_{=D'_{n,j,N}} + \frac{1}{N}.$$

Finalement, on a

$$|F_n(x) - F(x)| \leq \frac{1}{N} + \max(D_{n,j,N}, D'_{n,j,N}) \quad \forall x \in \Delta_{j,N}$$

et donc

$$\sup_{x \in \Delta_{j,N}} |F_n(x) - F(x)| \leq \frac{1}{N} + \max(D_{n,j,N}, D'_{n,j,N}) \leq \frac{1}{N} + \underbrace{\max_{0 \leq j \leq N} \{\max(D_{n,j,N}, D'_{n,j,N})\}}_{=D_{n,N}}$$

Comme le terme de droite ne dépend pas de j , on en déduit que

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \frac{1}{N} + D_{n,N}$$

D'après (4.5), il existe $\Omega_{j,N}$ tel que $\mathbf{P}(\Omega_{j,N}) = 1$ et pour tout $\omega \in \Omega_{j,N}$,

$$\max(D_{n,j,N}, D'_{n,j,N}) \xrightarrow{n \rightarrow \infty} 0$$

Par conséquent,

$$D_{n,N} \xrightarrow{n \rightarrow \infty} 0 \quad \forall \omega \in \bigcap_{j=1}^N \Omega_{j,N}$$

puis

$$\overline{\lim}_x |F_n(x, \omega) - F(x)| \leq \frac{1}{N} \quad \forall \omega \in \bigcap_{N \in \mathbb{N}} \bigcap_{j=1}^N \Omega_{j,N}$$

avec

$$\mathbf{P}\left(\bigcap_{N \in \mathbb{N}} \bigcap_{j=1}^N \Omega_{j,N}\right) = 1.$$

Comme ceci est vrai pour tout $N \in \mathbb{N}$, le résultat donc démontré avec

$$\tilde{\Omega} = \bigcap_{N \in \mathbb{N}} \bigcap_{j=1}^N \Omega_{j,N}.$$

□

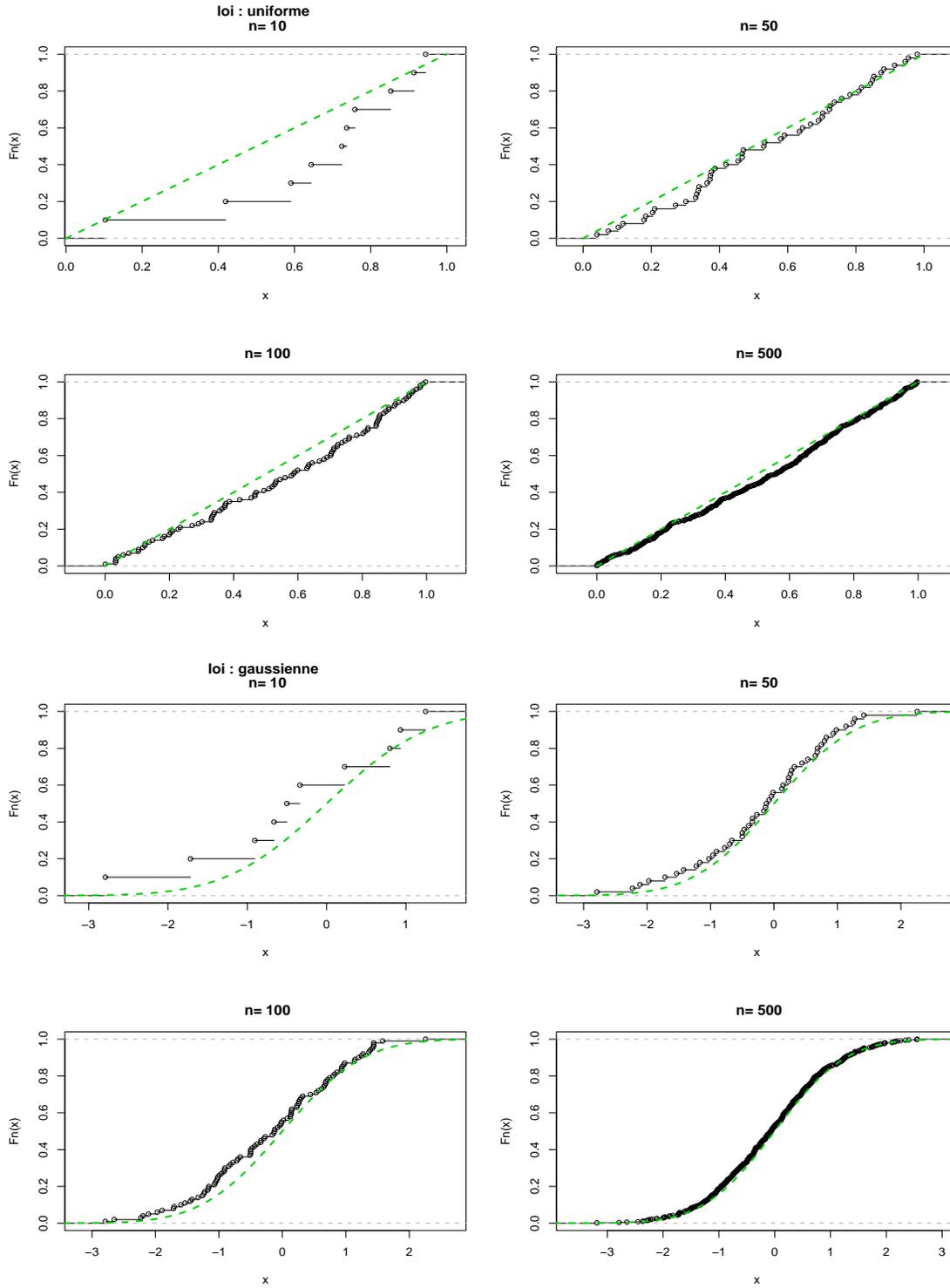


FIGURE 4.2. Illustration de la convergence de la fonction de répartition empirique F_n . Les variables aléatoires sont simulées suivant la loi uniforme sur $[0, 1]$ (haut) et la loi normale $\mathcal{N}(0, 1)$ (bas). On fait varier n : $n = 10, 50, 100, 500$

Convergence en loi

Pour tout vecteur aléatoire $X = (X^{(1)}, \dots, X^{(d)})'$ à valeurs dans \mathbb{R}^d , on note F_X sa fonction de répartition

$$F_X(x) = P(X^{(1)} \leq x_1, \dots, X^{(d)} \leq x_d) \quad \forall x = (x_1, \dots, x_d) \in \mathbb{R}^d$$

et ϕ_X sa fonction caractéristique

$$\phi_X(t) = \mathbb{E}(e^{it'X}) \quad \forall t \in \mathbb{R}^d.$$

Pour toute fonction f , on pose

$$\mathcal{C}_f = \{x : f \text{ est continue en } x\}.$$

1. Définitions

Définition 5.1. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires à valeurs dans \mathbb{R}^d . La suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si et seulement si

$$F_{X_n}(x) \rightarrow F_X(x)$$

en tout point de continuité x de F i.e. pour tout $x \in \mathcal{C}_F$.

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

Exemple 5.1.

– La loi de X_n est la mesure de Dirac au point $\frac{1}{n}$

$$F_{X_n}(x) = \mathbb{I}_{[\frac{1}{n}, \infty[}(x) \rightarrow \mathbb{I}_{]0, \infty[}(x) \quad \forall x \in \mathbb{R}$$

ceci implique

$$F_{X_n}(x) \rightarrow \mathbb{I}_{]0, \infty[}(x) := F(x) \quad \forall x \in \mathcal{C}_F$$

F est la fonction de répartition de la mesure de Dirac en zéro.

– La loi de Y_n est la mesure de Dirac au point $-\frac{1}{n}$

$$F_{Y_n}(x) = \mathbb{I}_{[-\frac{1}{n}, \infty[}(x) \rightarrow \mathbb{I}_{]0, \infty[}(x) \quad \forall x \in \mathbb{R}$$

Les deux suites convergent en loi vers la mesure de Dirac en 0 car leurs fonctions de répartition convergent vers $\mathbb{I}_{]0, \infty[}(x)$ lorsque $x \neq 0$. ||

Remarque 5.1. Il est important de remarquer que contrairement aux modes de convergence définis dans le chapitre 3, la limite est connue uniquement à travers sa loi. En conséquence, X peut être remplacée par n'importe quelle variable aléatoire qui a même loi.

C'est pour cette raison que l'on note souvent $X_n \xrightarrow{\mathcal{L}} \mathbf{P}$ où \mathbf{P} représente la loi de la limite X . Par exemple, on note $X_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$ au lieu de “ $X_n \xrightarrow{\mathcal{L}} X$ avec $X \sim \mathcal{N}(0, 1)$ ” |

De nombreuses propriétés prouvées dans le chapitre 3 pour la convergence en probabilité, presque sûre et dans L^p ne sont pas vérifiées par la convergence en loi. Voici quelques exemples

(1) Il n'y a plus unicité presque sûre de la limite.

Exemple 5.2. Si $X_n = X$ avec $X \sim \mathcal{N}(0, 1)$ alors $X_n \xrightarrow{\mathcal{L}} X$ et $X_n \xrightarrow{\mathcal{L}} -X$ car X et $-X$ ont même loi. Mais, les limites ne sont pas égales (au sens presque sûr) en effet

$$\mathbf{P}(X = -X) = \mathbf{P}(X = 0) = 0.$$

||

(2) Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} Y$ alors $X_n + Y_n$ ne converge pas nécessairement vers $X + Y$.

Exemple 5.3. Soit X une variable gaussienne standard. On prend $X_n = X$ et $Y_n = -X$. Comme la loi de X est symétrique, on a

$$X_n \xrightarrow{\mathcal{L}} X \quad Y_n \xrightarrow{\mathcal{L}} X$$

mais $X_n + Y_n = 0$ et donc ne converge pas vers $2X$. La loi limite de $X_n + Y_n$ est la mesure de Dirac en 0. ||

Les équivalences suivantes donnent différents critères de convergence en loi. Elles ne sont pas démontrées ici.

Proposition 5.1. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires à valeurs dans \mathbb{R}^d . Les affirmations suivantes sont équivalentes :

[L-1] : $F_{X_n}(x) \xrightarrow{n \rightarrow \infty} F(x)$ pour tout $x \in \mathcal{C}_F$

[L-2] : $\mathbb{E}g(X_n) \xrightarrow{n \rightarrow \infty} \mathbb{E}g(X)$ pour toute fonction $g : \mathbb{R}^d \rightarrow \mathbb{R}$ continue bornée

[L-3] : $\mathbb{E}g(X_n) \xrightarrow{n \rightarrow \infty} \mathbb{E}g(X)$ pour toute fonction $g : \mathbb{R}^d \rightarrow \mathbb{R}$ uniformément continue bornée

[L-4] : $\mathbb{E}g(X_n) \xrightarrow{n \rightarrow \infty} \mathbb{E}g(X)$ pour toute fonction $g : \mathbb{R}^d \rightarrow \mathbb{R}$ continue à support compact

[L-5] : $\mathbb{E}g(X_n) \xrightarrow{n \rightarrow \infty} \mathbb{E}g(X)$
pour toute fonction $g : \mathbb{R}^d \rightarrow \mathbb{R}$ bornée mesurable telle que $\mathbf{P}(X \in C_g) = 1$.

2. Convergence en loi et convergence en probabilité

Ce premier résultat complète le théorème 3.1.

Proposition 5.2. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires.

Si $X_n \xrightarrow{P} X$ alors $X_n \xrightarrow{\mathcal{L}} X$.

DÉMONSTRATION. On utilise le critère [L-3]. Soit g une fonction uniformément continue bornée.

Fixons $\varepsilon > 0$. Comme g est uniformément continue, il existe un réel $\eta > 0$ tel que

$$|X_n - X| < \eta \implies |g(X_n) - g(X)| \leq \varepsilon.$$

On décompose $|\mathbb{E}g(X_n) - \mathbb{E}g(X)|$ de la façon suivante :

$$|\mathbb{E}g(X_n) - \mathbb{E}g(X)| \leq \int_{|X_n - X| < \eta} |g(X_n) - g(X)| \, d\mathbf{P} + \int_{|X_n - X| \geq \eta} |g(X_n) - g(X)| \, d\mathbf{P}$$

Le premier terme du membre de droite est majoré par

$$\varepsilon \mathbf{P}(|X_n - X| < \eta) \leq \varepsilon$$

et le second est majoré par

$$2 \sup_x (|g(x)|) \mathbf{P}(|X_n - X| \geq \eta) = 2\|g\|_\infty \mathbf{P}(|X_n - X| \geq \eta)$$

qui converge vers zéro quand $n \rightarrow \infty$ puisque $X_n \xrightarrow{P} X$ et $\|g\|_\infty < \infty$.

On a donc

$$\overline{\lim} |\mathbb{E}g(X_n) - \mathbb{E}g(X)| \leq \varepsilon \quad \forall \varepsilon > 0$$

et la convergence en loi est prouvée. □

On peut ainsi compléter le graphique

$$L^p \quad \text{avec } p \geq 1 \implies L^1 \implies \text{proba} \implies \text{Loi}$$

$$\uparrow$$

$$\text{p.s.}$$

Exemple 5.4. Soit $X_n = (-1)^n X$ avec $X \sim \mathcal{N}(0, 1)$. On a $X_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$ mais $(X_n)_{n \in \mathbb{N}}$ ne converge pas en probabilité. La réciproque de la proposition 5.2 est donc fautive. Cependant lorsque la limite est une constante, on a le résultat suivant. ||

Proposition 5.3. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires et soit c une constante. Si $X_n \xrightarrow{\mathcal{L}} c$ alors $X_n \xrightarrow{P} c$.

DÉMONSTRATION. Puisque $X_n \xrightarrow{\mathcal{L}} c$, le critère $[\mathcal{L}-1]$ nous donne

$$F_{X_n}(t) \rightarrow \begin{cases} 0 & \text{si } t < c \\ 1 & \text{si } t > c \end{cases} \quad (5.1)$$

Soit $\varepsilon > 0$ fixé. On peut exprimer la probabilité $\mathbf{P}(|X_n - c| \geq \varepsilon)$ de la façon suivante :

$$\begin{aligned} \mathbf{P}(|X_n - c| \geq \varepsilon) &= 1 - \mathbf{P}(|X_n - c| \leq \varepsilon) \\ &= 1 - \underbrace{F_{X_n}(c + \varepsilon)}_{\substack{\rightarrow 1 \\ \text{car } c + \varepsilon > c}} + \underbrace{F_{X_n}(c - \varepsilon)}_{\substack{\rightarrow 0 \\ \text{car } c - \varepsilon < c}} \xrightarrow{n \rightarrow +\infty} 0 \end{aligned}$$

les différentes limites sont obtenues à partir de (5.1). □

3. Propriétés de la convergence en loi

Dans cette partie $(X_n)_{n \in \mathbb{N}}$ désigne une suite de vecteurs aléatoires à valeurs dans \mathbb{R}^d . Nous avons les propriétés suivantes

Proposition 5.4. Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ une fonction mesurable.

Si $X_n \xrightarrow{\mathcal{L}} X$ et si $\mathbf{P}(X \in C_f) = 1$ alors $f(X_n) \xrightarrow{\mathcal{L}} f(X)$

DÉMONSTRATION. On utilise le critère $[\mathcal{L}-2]$. Soit $g : \mathbb{R}^k \rightarrow \mathbb{R}$ continue bornée. La fonction $g \circ f$ est bornée et elle est continue sur C_f . Par conséquent, on a $\mathbf{P}(X \in C_{g \circ f}) = 1$ et $g \circ f : \mathbb{R}^d \rightarrow \mathbb{R}$ est mesurable bornée donc d'après le critère $[\mathcal{L}-5]$,

$$\mathbb{E}g(f(X_n)) \rightarrow \mathbb{E}g(f(X)).$$

Ceci assure la convergence $f(X_n) \xrightarrow{\mathcal{L}} f(X)$. □

Exemple 5.5. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires.

- Si $X_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$ alors $X_n^2 \xrightarrow{\mathcal{L}} \chi^2(1)$
- Si $X_n \xrightarrow{\mathcal{L}} X \sim \mathcal{N}(0, 1)$ alors $\frac{1}{X_n} \xrightarrow{\mathcal{L}} \frac{1}{X}$ car

$$\mathbf{P}(X \neq 0) = \mathbf{P}(X \in \{x : x \mapsto x^{-1} \text{ est continue en } x\}) = 1$$

Exercice 2. Calculer la loi de la variable aléatoire $\frac{1}{X}$ △△ ||

Proposition 5.5. Si $X_n \xrightarrow{\mathcal{L}} X$ et si $X_n - Y_n \xrightarrow{P} 0$ alors $Y_n \xrightarrow{\mathcal{L}} X$

DÉMONSTRATION. On utilise le critère $[\mathcal{L}-3]$. Soit $g : \mathbb{R}^d \rightarrow \mathbb{R}$ uniformément continue bornée. A $\varepsilon > 0$ fixé, il existe $\eta > 0$ tel que $|X_n - Y_n| < \eta$ implique que $|g(X_n) - g(Y_n)| < \varepsilon$.

On décompose $|\mathbb{E}g(Y_n) - \mathbb{E}g(X)|$ de la façon suivante :

$$\begin{aligned} |\mathbb{E}g(Y_n) - \mathbb{E}g(X)| &\leq |\mathbb{E}g(Y_n) - \mathbb{E}g(X_n)| + |\mathbb{E}g(X_n) - \mathbb{E}g(X)| \\ &\leq \int_{|X_n - Y_n| < \eta} |g(Y_n) - g(X_n)| \, d\mathbf{P} + \int_{|X_n - Y_n| \geq \eta} |g(X_n) - g(Y_n)| \, d\mathbf{P} + |\mathbb{E}g(X_n) - \mathbb{E}g(X)| \quad (5.2) \end{aligned}$$

Comme $X_n \xrightarrow{\mathcal{L}} X$ le troisième terme de (5.2) converge vers zéro. D'autre part, grâce à l'uniforme continuité de g , on a

$$\int_{|X_n - Y_n| < \eta} |g(Y_n) - g(X_n)| \, d\mathbf{P} \leq \varepsilon \mathbf{P}(|X_n - Y_n| < \eta) \leq \varepsilon$$

et

$$\int_{|X_n - Y_n| \geq \eta} |g(Y_n) - g(X_n)| \, d\mathbf{P} \leq 2 \sup_{x \in \mathbb{R}} (|g(x)|) \mathbf{P}(|X_n - Y_n| \geq \eta) \xrightarrow{n \rightarrow \infty} 0$$

en utilisant la convergence $X_n - Y_n \xrightarrow{P} 0$ et le fait que g est bornée.

En conclusion, pour tout $\varepsilon > 0$, on a

$$\overline{\lim} |\mathbb{E}g(Y_n) - \mathbb{E}g(X)| \leq \varepsilon$$

d'où le résultat. \square

Proposition 5.6. *Si $X_n \xrightarrow{\mathcal{L}} X$ et si $Y_n \xrightarrow{\mathcal{L}} c$ où c est une constante alors $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} X \\ c \end{pmatrix}$*

DÉMONSTRATION. Notons d'abord que

$$\mathbf{P} \left(\left| \begin{pmatrix} X_n \\ Y_n \end{pmatrix} - \begin{pmatrix} X_n \\ c \end{pmatrix} \right| \geq \varepsilon \right) = \mathbf{P}(|Y_n - c| \geq \varepsilon) \xrightarrow{n \rightarrow +\infty} 0$$

car $Y_n \xrightarrow{P} c$. Comme $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} - \begin{pmatrix} X_n \\ c \end{pmatrix} \xrightarrow{P} 0$, d'après la proposition 5.5, il reste maintenant à prouver que $\begin{pmatrix} X_n \\ c \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} X \\ c \end{pmatrix}$. Pour ce faire, on utilise le critère [L-2]. Soit g continue bornée de $\mathbb{R}^d \times \mathbb{R}^k \rightarrow \mathbb{R}$. Il est facile de remarquer que la fonction $h : x \mapsto g(x, c)$ est continue bornée de $\mathbb{R}^d \rightarrow \mathbb{R}$. Par conséquent, en utilisant le critère [L-2] et la convergence $X_n \xrightarrow{\mathcal{L}} X$, on obtient le résultat

$$\mathbb{E}h(X_n) \rightarrow \mathbb{E}h(X)$$

soit encore

$$\mathbb{E}g(X_n, c) \rightarrow \mathbb{E}g(X, c)$$

d'où la convergence en loi de $\begin{pmatrix} X_n \\ c \end{pmatrix}$ vers $\begin{pmatrix} X \\ c \end{pmatrix}$. Ceci conclut la preuve. \square

Proposition 5.7. *Soit $f : \mathbb{R}^d \times \mathbb{R}^k \rightarrow \mathbb{R}$ mesurable.*

Si $X_n \xrightarrow{\mathcal{L}} X$, $Y_n \xrightarrow{\mathcal{L}} c$ avec c une constante et si $\mathbf{P}((X, c) \in \mathcal{C}_f) = 1$ alors $f(X_n, Y_n) \xrightarrow{\mathcal{L}} f(X, c)$

DÉMONSTRATION. C'est une conséquence des propositions 5.4 et 5.6. \square

Corollaire 5.8. *Soit c une constante.*

- *Si $X_n \xrightarrow{\mathcal{L}} X$, $Y_n \xrightarrow{\mathcal{L}} c$ alors $X_n + Y_n \xrightarrow{\mathcal{L}} X + c$*
- *Si $X_n \xrightarrow{\mathcal{L}} X$, $Y_n \xrightarrow{\mathcal{L}} c$ alors $X'_n Y_n \xrightarrow{\mathcal{L}} X'c$*
- *Si $X_n \xrightarrow{\mathcal{L}} X$, et si Y_n est une suite de variables aléatoires réelles telle que $Y_n \xrightarrow{\mathcal{L}} c \neq 0$ alors $\frac{X_n}{Y_n} \xrightarrow{\mathcal{L}} \frac{X}{c}$*

4. Convergence en loi et fonctions caractéristiques

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires à valeurs dans \mathbb{R}^d . Pour tout $n \in \mathbb{N}$, on note ϕ_{X_n} la fonction caractéristique de X_n .

Théorème 5.9. *La suite de vecteurs aléatoires $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si et seulement si pour tout $t \in \mathbb{R}^d$,*

$$\phi_{X_n}(t) \rightarrow \phi_X(t).$$

DÉMONSTRATION. Ce résultat est admis. \square

Le résultat suivant montre que l'étude de la convergence en loi pour des suites de vecteurs aléatoires peut être réduit à un problème de convergence en loi unidimensionnel.

Proposition 5.10. *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires à valeurs dans \mathbb{R}^d avec $d > 1$. $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si et seulement si pour tout $\lambda \in \mathbb{R}^d$,*

$$\lambda' X_n = \sum_{i=1}^d \lambda_i X_n^{(i)} \xrightarrow{\mathcal{L}} \sum_{i=1}^d \lambda_i X^{(i)} = \lambda' X$$

DÉMONSTRATION.

[Preuve de \implies] Supposons que $X_n \xrightarrow{\mathcal{L}} X$. Soit $\lambda \in \mathbb{R}^d$ fixé. Pour tout $t \in \mathbb{R}$, on a

$$\phi_{\lambda' X_n}(t) = \mathbb{E}(e^{it\lambda' X_n}) = \mathbb{E}(e^{i(t\lambda)' X_n}) = \phi_{X_n}(t\lambda') \xrightarrow[n \rightarrow \infty]{} \phi_X(t\lambda') = \mathbb{E}(e^{i(t\lambda)' X}) = \phi_{\lambda' X}(t)$$

[Preuve de \impliedby] Supposons que $\lambda' X_n \xrightarrow{\mathcal{L}} \lambda' X$ pour tout $\lambda \in \mathbb{R}^d$. Pour tout $t \in \mathbb{R}^d$,

$$\phi_{X_n}(t) = \mathbb{E}(e^{it' X_n}) = \phi_{\sum_{i=1}^d t_i X_n^{(i)}}(1) \xrightarrow[n \rightarrow \infty]{} \phi_{\sum_{i=1}^d t_i X^{(i)}}(1) = \phi_X(t)$$

□

5. Théorème central limite

Théorème 5.11 (Théorème central limite). *Soit $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires de L^2 indépendantes et identiquement distribuées.*

On suppose que $\sigma^2 := \text{Var}(X_1) \neq 0$. Sous ces hypothèses, on a

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j - \mathbb{E}(X_j)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2) \quad (5.3)$$

ce qui est équivalent à

$$\frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n (X_j - \mathbb{E}(X_j)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Remarque 5.2. Les variables aléatoires

$$Z_j = \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n (X_j - \mathbb{E}(X_j))$$

sont d'espérance nulle et de variance 1. La loi limite vérifie aussi ces propriétés. |

Remarque 5.3. Si les X_j sont iid suivant la loi $\mathcal{N}(0, \sigma^2)$, la convergence est triviale puisque les Z_j sont iid suivant la loi $\mathcal{N}(0, 1)$ |

DÉMONSTRATION. La preuve est essentiellement basée sur la caractérisation de la convergence en loi par la convergence de la fonction caractéristique (voir 5.9).

Lemme 5.12. *Soit X une variable aléatoire de L^2 , sa fonction caractéristique s'écrit*

$$\phi_X(t) = 1 + it\mathbb{E}(X) - \frac{1}{2}t^2\mathbb{E}(X^2) + o(t^2) \quad (5.4)$$

DÉMONSTRATION. A faire en exercice □

Retour à la preuve du théorème central limite

Posons $Y_j = \frac{X_j - \mathbb{E}(X_j)}{\sigma}$

$$\begin{aligned} \phi_{\frac{\sum_{j=1}^n Y_j}{\sqrt{n}}}(t) &= \phi_{\sum_{j=1}^n Y_j}\left(\frac{t}{\sqrt{n}}\right) \\ &= \prod_{j=1}^n \phi_{Y_j}\left(\frac{t}{\sqrt{n}}\right) \quad (\text{car les v.a. sont indépendantes}) \\ &= \left(\phi_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n \quad (\text{car les v.a. sont de même loi}) \end{aligned}$$

D'après le lemme 5.12, on a

$$\phi_Y(t) = 1 + it\mathbb{E}(Y) - \frac{1}{2}t^2\mathbb{E}(Y^2) + o(t^2) = 1 - \frac{t^2}{2} + o(t^2)$$

car $\mathbb{E}(Y) = 1$ et $\mathbb{E}(Y^2) = \text{Var}(X)/\sigma^2 = 1$. Par conséquent, on a

$$\phi_{\frac{\sum_{j=1}^n Y_j}{\sqrt{n}}}(t) = \left(1 - \frac{1}{2} \frac{t^2}{n} + o\left(\frac{t^2}{n}\right)\right)^n$$

On utilise l'inégalité

$$|a^n - b^n| \leq n|a - b| \quad \text{si } |a| < 1 \text{ et } |b| < 1.$$

A t fixé, pour n assez grand, on a

$$\left| \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n - \left(1 - \frac{t^2}{2n}\right)^n \right| \leq o\left(\frac{t^2}{n}\right)n \xrightarrow[n \rightarrow \infty]{t \text{ fixé}} 0$$

Comme

$$\left(1 - \frac{t^2}{2n}\right)^n \xrightarrow[n \rightarrow \infty]{t \text{ fixé}} e^{-t^2/2},$$

on peut conclure. \square

En fait on peut démontrer sans hypothèse supplémentaire, la convergence uniforme de la fonction de répartition de X_n vers celle de la loi gaussienne standard.

Proposition 5.13. *Si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires de L^2 iid telle que $\sigma^2 := \text{Var}(X_1) \neq 0$ alors on a uniformément en $x \in \mathbb{R}$*

$$\mathbf{P}\left(\frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n (X_j - \mathbb{E}(X_j)) \leq x\right) \rightarrow \Phi(x) \quad (5.5)$$

où Φ est la fonction de répartition de la loi $\mathcal{N}(0, 1)$.

ceci est équivalent à

$$\sup_{x \in \mathbb{R}} \left| \mathbf{P}\left(\frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n (X_j - \mathbb{E}(X_j)) \leq x\right) - \Phi(x) \right| \rightarrow 0 \quad (5.6)$$

DÉMONSTRATION. Ce résultat est admis \square

En fait si l'on suppose que $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires de L^3 indépendantes et identiquement distribuées, on peut contrôler la vitesse de convergence de l'erreur vers zéro.

Théorème 5.14 (Théorème de Berry Esseen). *Si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires de L^3 indépendantes et identiquement distribuées. On suppose que*

$$\mathbb{E}(X_1) = 0, \quad \mathbb{E}(X_1^2) = \sigma^2 \neq 0, \quad \mathbb{E}(|X_1|^3) = \rho.$$

Sous ces hypothèses

$$\sup_{x \in \mathbb{R}} \left| \mathbf{P}\left(\frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n (X_j - \mathbb{E}(X_j)) \leq x\right) - \Phi(x) \right| \leq \frac{1}{\sqrt{n}} \frac{2\rho}{\sigma^3} \quad (5.7)$$

Théorème 5.15. *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires de L^2 indépendants et identiquement distribués. On pose $\Sigma = \text{Var}(X_1)$. On a*

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n (X_j - \mathbb{E}X_j) \xrightarrow{\mathcal{L}} \mathcal{N}_d(0, \Sigma)$$

DÉMONSTRATION. Ce résultat découle de la version unidimensionnelle du théorème central limite (Théorème 5.11) et de la proposition 5.10. Fixons $\lambda \in \mathbb{R}^d$. Les variables aléatoires réelles $(\lambda'X_i)_{i \in \mathbb{N}}$ sont indépendantes et identiquement distribuées. On peut donc appliquer le Théorème Central Limite.

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n (\lambda'X_k - \mathbb{E}(\lambda'X_k)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \text{Var}(\lambda'X_1))$$

soit encore

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n (\lambda' X_k - \lambda' \mathbb{E}(X_k)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \lambda' \text{Var}(X_1) \lambda).$$

Finalement, on a

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n (\lambda' X_k - \lambda' \mathbb{E}(X_k)) = \lambda' \left(\frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - \mathbb{E}(X_k)) \right) \xrightarrow{\mathcal{L}} \lambda' X \quad \text{avec } \mathcal{N}(0, \text{Var}(X_1)).$$

On peut ensuite conclure grâce à la proposition 5.10. \square

6. Théorème de Cramér

Théorème 5.16 (Théorème de Cramér). *Soit g de \mathbb{R}^d dans \mathbb{R}^k une fonction de classe \mathcal{C}^1 au voisinage de μ . Soit $(X_n)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires dans \mathbb{R}^d telle que*

$$\sqrt{n}(U_n - \mu) \xrightarrow{\mathcal{L}} U \quad (5.8)$$

alors

$$\sqrt{n}(g(U_n) - g(\mu)) \xrightarrow{\mathcal{L}} {}^t(\nabla g(\mu)) U \quad (5.9)$$

où $\nabla g(\mu) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_k}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial g_1}{\partial x_d} & \cdots & \frac{\partial g_k}{\partial x_d} \end{pmatrix}.$

Corollaire 5.17. *Soit g une fonction \mathcal{C}^1 au voisinage de μ . Si*

$$\sqrt{n}(U_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma)$$

alors

$$\sqrt{n}(g(U_n) - g(\mu)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, {}^t \nabla g(\mu) \Sigma \nabla g(\mu))$$

Exemple 5.6. Si $(X_n)_{n \in \mathbb{N}}$ est une suite iid de L^2 . Posons $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$. D'après le TCL,

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

d'où, pour toute fonction g de classe \mathcal{C}^1 au voisinage de μ , on a

$$\sqrt{n}(g(\bar{X}_n) - g(\mu)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, (\nabla g(\mu))^2 \sigma^2)$$

En prenant $g(x) = x^2$, on obtient

$$\sqrt{n}(\bar{X}_n^2 - \mu^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\mu^2 \sigma^2)$$

La limite est une loi gaussienne dégénérée si $\mu = 0$. \parallel

DÉMONSTRATION. D'après le corollaire 5.8, on a $U_n - \mu \xrightarrow{\mathcal{L}} 0$ car

$$U_n - \mu = \sqrt{n}(U_n - \mu) \frac{1}{\sqrt{n}}$$

Ensuite, comme g est de classe \mathcal{C}^1 au voisinage de μ , on peut écrire le développement de Taylor de cette fonction à l'ordre 1 au voisinage de μ . Par conséquent, Il existe $\eta > 0$ tel que si $|x - \mu| \leq \eta$ alors on a

$$g(x) = g(\mu) + \left(\int_0^1 \nabla g(\mu + \theta(x - \mu))^T d\theta \right) (x - \mu).$$

Sur l'ensemble $\{|U_n - \mu| \leq \eta\}$, on a donc

$$\sqrt{n}(g(U_n) - g(\mu)) = \left(\int_0^1 \nabla g(\mu + \theta(U_n - \mu))^T d\theta \right) \sqrt{n}(U_n - \mu)$$

D'où la décomposition

$$\sqrt{n}(g(U_n) - g(\mu)) = \left(\int_0^1 \nabla g(\mu + \theta(U_n - \mu))^T d\theta \right) \sqrt{n}(U_n - \mu) + R_n \mathbb{I}_{|U_n - \mu| \geq \eta} \quad (5.10)$$

avec

$$R_n = \sqrt{n}(g(U_n) - g(\mu)) - \left(\int_0^1 \nabla g(\mu + \theta(U_n - \mu))^T d\theta \right) \sqrt{n}(U_n - \mu)$$

(1) La fonction

$$y \mapsto \int_0^1 \nabla g(\mu + \theta y)^T d\theta$$

est continue en zéro et $U_n - \mu \xrightarrow{P} 0$ donc

$$\int_0^1 \nabla g(\mu + \theta(U_n - \mu))^T d\theta \xrightarrow{P} \nabla g(\mu)^T$$

(2) Comme $\sqrt{n}(U_n - \mu) \xrightarrow{\mathcal{L}} U$, la proposition 5.7, avec $f(x, y) = x^T y$, donne la convergence en loi du premier terme de (5.10) vers $\nabla g(\mu)^T U$.

(3) Pour le second terme de (5.10), on a

$$\mathbf{P}(|R_n| \mathbb{I}_{|U_n - \mu| \geq \eta} \geq \varepsilon) \leq \mathbf{P}(|U_n - \mu| \geq \eta) \rightarrow 0$$

On conclut en appliquant la proposition 5.7 avec $f(x, y) = x + y$. □

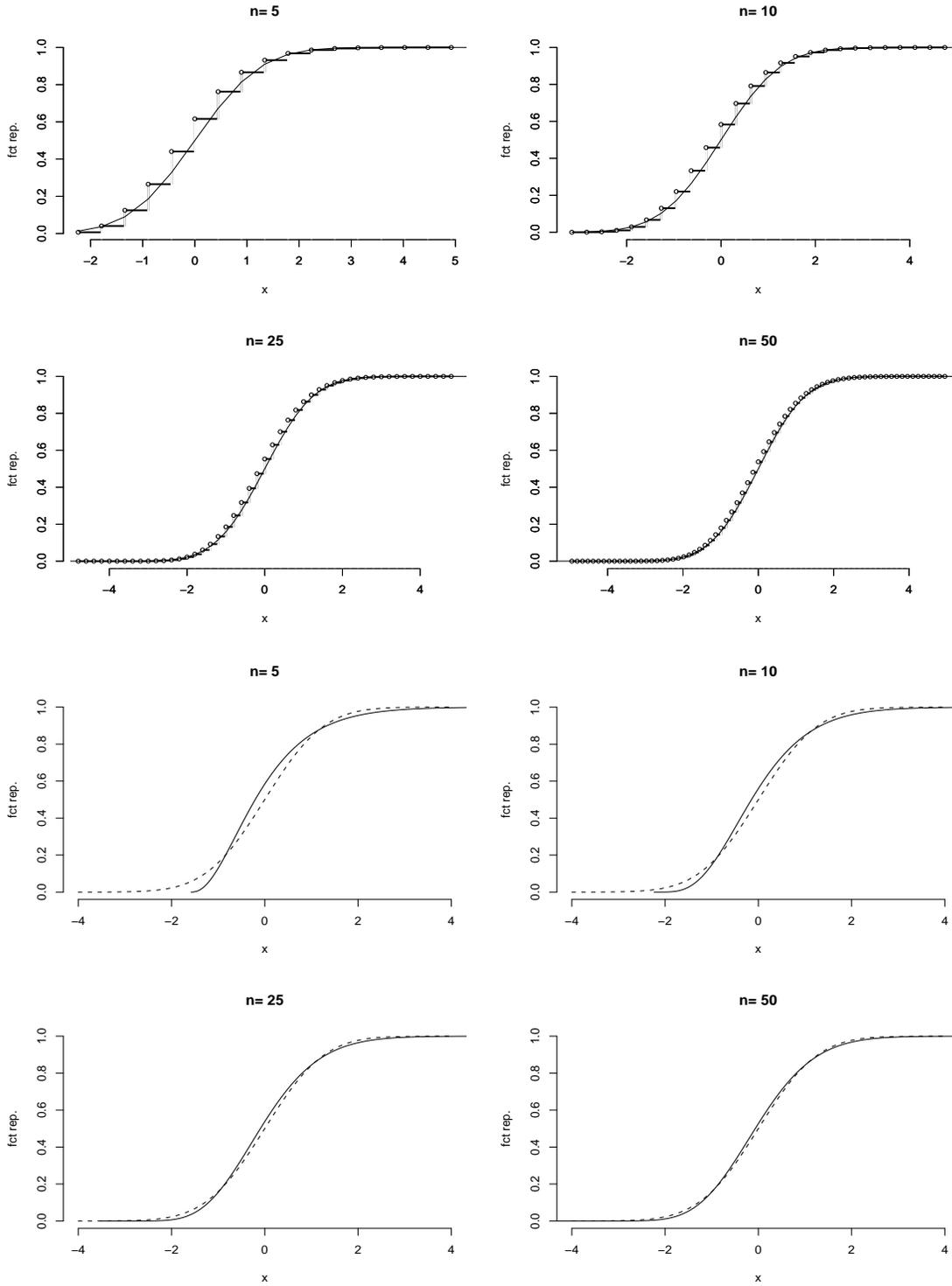


FIGURE 5.1. Convergence de la fonction de répartition F_{Y_n} avec $Y_n = \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n (X_j - \mathbb{E}(X_j))$ vers $\Phi(x)$ la fonction de répartition de la loi gaussienne standard. (haut) $(X_n)_{n \in \mathbb{N}}$ iid suivant la loi de Poisson de paramètre 1 (bas) $(X_n)_{n \in \mathbb{N}}$ iid suivant la loi gamma de paramètre $(1/2, 1/2)$ *i.e.* de densité $x \mapsto \frac{1}{\Gamma(1/2)\sqrt{2}} \sqrt{x} e^{-x/2}$

7. Théorème de Lindberg-Feller

Dans cette section, nous donnons une version du théorème central limite pour des variables aléatoires non équidistribuées.

On se donne un tableau triangulaire de variables aléatoires

$$X_{n,1}, \dots, X_{n,k_n} \quad n = 1, \dots, \quad k_n \in \mathbb{N}$$

qui vérifie les conditions suivantes (notées $[\mathcal{L}]$) :

- à n fixé, les variables $X_{n,1}, \dots, X_{n,k_n}$ sont indépendantes
- pour tout $n \in \mathbb{N}$ et pour tout $j \in \{1, \dots, k_n\}$, $X_{n,j}$ est dans L^2 .
- pour tout $t > 0$

$$\frac{1}{D_n^2} \sum_{j=1}^{k_n} \mathbb{E} \left[(X_{n,j} - \mathbb{E}(X_{n,j}))^2 \mathbb{I}_{|X_{n,j} - \mathbb{E}(X_{n,j})| > t D_n} \right] \rightarrow 0 \quad n \rightarrow \infty \quad (5.11)$$

où

$$D_n^2 = \sum_{j=1}^{k_n} \mathbb{E} (X_{n,j} - \mathbb{E}(X_{n,j}))^2 \quad (5.12)$$

Remarque 5.4. Si l'on se place dans le contexte du théorème central limite classique : $k_n = n$ et $X_{n,j} = Y_j$ où (Y_j) est une suite de variables aléatoires L^2 et i.i.d..

On a $D_n^2 = n\sigma^2$, où $\sigma^2 = \text{Var}(Y_1)$ et la condition (5.11) qui devient

$$\mathbb{E} \left[(Y_1 - \mathbb{E}(Y_1))^2 \mathbb{I}_{|Y_1 - \mathbb{E}(Y_1)| > t\sigma\sqrt{n}} \right] \rightarrow 0 \quad n \rightarrow \infty$$

est vérifiée. |

Lemme 5.18. *La condition (5.11) implique la convergence uniforme suivante*

$$\sup_{1 \leq j \leq k_n} \frac{1}{D_n^2} \mathbb{E} (X_{n,j} - \mathbb{E}(X_{n,j}))^2 \rightarrow 0 \quad n \rightarrow \infty \quad (5.13)$$

DÉMONSTRATION. Fixons $\varepsilon \in [0, 1]$.

L'espérance $\mathbb{E} (X_{n,j} - \mathbb{E}(X_{n,j}))^2$ se décompose sous la forme

$$\mathbb{E} \left[(X_{n,j} - \mathbb{E}(X_{n,j}))^2 \mathbb{I}_{|X_{n,j} - \mathbb{E}(X_{n,j})| \leq \varepsilon D_n} \right] + \mathbb{E} \left[(X_{n,j} - \mathbb{E}(X_{n,j}))^2 \mathbb{I}_{|X_{n,j} - \mathbb{E}(X_{n,j})| > \varepsilon D_n} \right]$$

Uniformement en j , le premier terme est borné par $\varepsilon^2 D_n^2$ et le second par

$$\sum_{j=1}^{k_n} \mathbb{E} \left[(X_{n,j} - \mathbb{E}(X_{n,j}))^2 \mathbb{I}_{|X_{n,j} - \mathbb{E}(X_{n,j})| > \varepsilon D_n} \right].$$

Pour tout $j \leq k_n$, on a donc

$$\sup_{1 \leq j \leq k_n} \frac{1}{D_n^2} \mathbb{E} (X_{n,j} - \mathbb{E}(X_{n,j}))^2 \leq \varepsilon^2 + \frac{1}{D_n^2} \sum_{j=1}^{k_n} \mathbb{E} \left[(X_{n,j} - \mathbb{E}(X_{n,j}))^2 \mathbb{I}_{|X_{n,j} - \mathbb{E}(X_{n,j})| > \varepsilon D_n} \right].$$

D'où

$$\limsup_{n \rightarrow \infty} \sup_{1 \leq j \leq k_n} \frac{1}{D_n^2} \mathbb{E} (X_{n,j} - \mathbb{E}(X_{n,j}))^2 < \varepsilon^2$$

d'après (5.11). Comme ε est fixé arbitrairement, on a bien, quand $n \rightarrow \infty$,

$$\sup_{1 \leq j \leq k_n} \frac{1}{D_n^2} \mathbb{E} (X_{n,j} - \mathbb{E}(X_{n,j}))^2 \rightarrow 0$$

□

Théorème 5.19. *Sous les conditions $[\mathcal{L}]$, on a*

$$\frac{1}{D_n} \sum_{i=1}^{k_n} (X_{n,i} - \mathbb{E}(X_{n,i})) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (5.14)$$

DÉMONSTRATION. On suppose sans perte de généralité que les variables aléatoires $X_{n,j}$ sont centrées et on va prouver la convergence des fonctions caractéristiques. Posons $\phi_{n,j}(t) = \mathbb{E}e^{itX_{n,j}}$ et $\sigma_{n,j}^2 = \text{Var}(X_{n,j})$.

Rappelons que

$$|R(u)| \leq \min(u^2, \frac{|u|^3}{6})$$

avec $R(u) = e^{iu} - 1 - iu + u^2/2$.

Fixons $t \in \mathbb{R}$.

$$\phi_{n,j}(t) = 1 - \frac{1}{2}\sigma_{n,j}^2 t^2 + R_{n,j}(t)$$

avec

$$|R_{n,j}(t)| \leq \mathbb{E}|R(tX_{n,j})| \mathbb{E}X_{n,j}^2$$

Soit $\varepsilon > 0$ fixé.

$$\begin{aligned} |R_{n,j}(t)| &\leq \mathbb{E}(|R(tX_{n,j})| \mathbb{I}_{|X_{n,j}| > \varepsilon D_n}) + \mathbb{E}(|R(tX_{n,j})| \mathbb{I}_{|X_{n,j}| \leq \varepsilon D_n}) \\ &\leq t^2 \mathbb{E}(X_{n,j}^2 \mathbb{I}_{|X_{n,j}| > \varepsilon D_n}) + \frac{|t|^3}{6} \mathbb{E}(|X_{n,j}|^3 \mathbb{I}_{|X_{n,j}| \leq \varepsilon D_n}) \\ &\leq t^2 \mathbb{E}(X_{n,j}^2 \mathbb{I}_{|X_{n,j}| > \varepsilon D_n}) + \frac{|t|^3}{6} \varepsilon D_n \mathbb{E}(X_{n,j}^2) \\ &\leq t^2 \mathbb{E}(X_{n,j}^2 \mathbb{I}_{|X_{n,j}| > \varepsilon D_n}) + \frac{|t|^3}{6} \varepsilon D_n \sigma_{n,j}^2 \end{aligned}$$

Calculons maintenant la fonction caractéristique ϕ_n de $\frac{1}{D_n} \sum_{j=1}^{k_n} X_{n,j}$. On a, en utilisant l'indépendance des variables $X_{n,1}, \dots, X_{n,k_n}$,

$$\begin{aligned} \phi_n(t) &= \prod_{j=1}^{k_n} \phi_{n,j}(t/D_n) \\ &= \prod_{j=1}^{k_n} \left(1 - \sigma_{n,j}^2 \frac{t^2}{2D_n^2} + R_{n,j}(t/D_n) \right) \end{aligned}$$

On utilise l'inégalité suivante

$$\left| \prod_{j=1}^{k_n} a_j - \prod_{j=1}^{k_n} b_j \right| \leq \sum_{j=1}^{k_n} |a_j - b_j|$$

si $|a_j| \leq 1$ et $|b_j| \leq 1$ pour tout j . On applique cette inégalité à

$$\begin{aligned} a_j &= 1 - \sigma_{n,j}^2 \frac{t^2}{2D_n^2} \\ b_j &= \phi_{n,j}(t/D_n). \end{aligned}$$

On a bien $|b_j| \leq 1$ pour tout couple (j, n) et (5.13) assure que $|a_j| \leq 1$ pour $n > n_0$ et tout $j = 1, \dots, k_n$.

On obtient

$$\begin{aligned} \left| \phi_n(t) - \prod_{j=1}^{k_n} \left(1 - \sigma_{n,j}^2 \frac{t^2}{2D_n^2} \right) \right| &\leq \frac{t^2}{D_n^2} \sum_{j=1}^{k_n} \mathbb{E}(X_{n,j}^2 \mathbb{I}_{|X_{n,j}| > \varepsilon D_n}) + \frac{|t|^3}{6} \varepsilon \sum_{j=1}^{k_n} \frac{\sigma_{n,j}^2}{D_n^2} \\ &\leq \frac{t^2}{D_n^2} \sum_{j=1}^{k_n} \mathbb{E}(X_{n,j}^2 \mathbb{I}_{|X_{n,j}| > \varepsilon D_n}) + \frac{|t|^3}{6} \varepsilon \end{aligned}$$

D'après (5.11),

$$\limsup_{n \rightarrow \infty} \left| \phi_n(t) - \prod_{j=1}^{k_n} \left(1 - \sigma_{n,j}^2 \frac{t^2}{2D_n^2} \right) \right| \leq \frac{|t|^3}{6} \varepsilon$$

Comme ε est fixé arbitrairement, on a donc,

$$\lim_{n \rightarrow \infty} \left| \phi_n(t) - \prod_{j=1}^{k_n} \left(1 - \sigma_{n,j}^2 \frac{t^2}{2D_n^2} \right) \right| = 0.$$

Il reste à prouver la convergence du produit $\prod_{j=1}^{k_n} \left(1 - \sigma_{n,j}^2 \frac{t^2}{2D_n^2} \right)$. On a

$$\log \prod_{j=1}^{k_n} \left(1 - \sigma_{n,j}^2 \frac{t^2}{2D_n^2} \right) = \sum_{j=1}^{k_n} \log \left(1 - \sigma_{n,j}^2 \frac{t^2}{2D_n^2} \right) = -t^2 \sum_{j=1}^{k_n} \frac{\sigma_{n,j}^2}{2D_n^2} + r_n(t)$$

où

$$|r_n(t)| \leq \sum_{j=1}^{k_n} \frac{t^4 \sigma_{n,j}^4}{D_n^4} \leq t^4 \sum_{j=1}^{k_n} \frac{\sigma_{n,j}^2}{D_n^2} \sup_{1 \leq j \leq k_n} \frac{\sigma_{n,j}^2}{D_n^2} = t^4 \sup_{1 \leq j \leq k_n} \frac{\sigma_{n,j}^2}{D_n^2}.$$

D'après (5.13), le terme de droite converge vers zéro. On en déduit finalement que

$$\lim_{n \rightarrow \infty} \phi_n(t) = \lim_{n \rightarrow \infty} \prod_{j=1}^{k_n} \left(1 - \sigma_{n,j}^2 \frac{t^2}{2D_n^2} \right) = e^{-t^2/2}.$$

□

Sous des conditions plus fortes que $[\mathcal{L}]$ mais généralement plus simple à vérifier, on a le théorème central limite suivant.

Théorème 5.20 (Théorème de Lyapounov). *Soit un tableau triangulaire de variables aléatoires*

$$X_{n,1}, \dots, X_{n,k_n} \quad n = 1, \dots, \quad k_n \in \mathbb{N}$$

qui vérifie les conditions suivantes

- à n fixé, les variables $X_{n,1}, \dots, X_{n,k_n}$ sont indépendantes
- il existe $\delta > 0$ tel que pour tout $n \in \mathbb{N}$ et pour tout $j \in \{1, \dots, k_n\}$,

$$\alpha_{n,j} := \mathbb{E}|X_{n,j}|^{2+\delta} < \infty$$

et

$$\frac{1}{D_n^{2+\delta}} \sum_{j=1}^{k_n} \alpha_{n,j} \rightarrow 0 \quad n \rightarrow \infty \quad (5.15)$$

où D_n^2 est définie en (5.12).

Sous ces hypothèses, la condition $[\mathcal{L}]$ est vérifiée et

$$\frac{1}{D_n} \sum_{i=1}^{k_n} (X_{n,i} - \mathbb{E}(X_{n,i})) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (5.16)$$

DÉMONSTRATION. Il suffit de prouver la convergence (5.11). Soit $\varepsilon > 0$ fixé. On a

$$\begin{aligned} \mathbb{E} \left[(X_{n,j} - \mathbb{E}(X_{n,j}))^2 \mathbb{I}_{|X_{n,j} - \mathbb{E}(X_{n,j})| > \varepsilon D_n} \right] &\leq \\ &\left(\mathbb{E}|X_{n,j} - \mathbb{E}(X_{n,j})|^{2+\delta} \right)^{2/(2+\delta)} \left(\mathbb{E} \mathbb{I}_{|X_{n,j} - \mathbb{E}(X_{n,j})| > \varepsilon D_n} \right)^{\delta/(2+\delta)} \end{aligned}$$

par l'inégalité de Holder avec $p = (2 + \delta)/2$ et $q = (2 + \delta)/\delta$. Ensuite, l'inégalité de Markov, nous donne

$$\mathbb{E} \mathbb{I}_{|X_{n,j} - \mathbb{E}(X_{n,j})| > \varepsilon D_n} \leq \frac{1}{(\varepsilon D_n)^{2+\delta}} \mathbb{E}|X_{n,j} - \mathbb{E}(X_{n,j})|^{2+\delta}$$

D'où

$$\begin{aligned} \mathbb{E} \left[(X_{n,j} - \mathbb{E}(X_{n,j}))^2 \mathbb{I}_{|X_{n,j} - \mathbb{E}(X_{n,j})| > \varepsilon D_n} \right] &\leq \frac{1}{(\varepsilon D_n)^\delta} \mathbb{E}|X_{n,j} - \mathbb{E}(X_{n,j})|^{2+\delta} \\ &\leq \frac{2^{3+\delta}}{(\varepsilon D_n)^\delta} \mathbb{E}|X_{n,j}|^{2+\delta} \end{aligned}$$

En effet, on a

$$|X_{n,j} - \mathbb{E}(X_{n,j})|^{2+\delta} \leq (|X_{n,j}| + |\mathbb{E}(X_{n,j})|)^{2+\delta} \leq 2^{1+\delta} \left(|X_{n,j}|^{2+\delta} + |\mathbb{E}(X_{n,j})|^{2+\delta} \right)$$

par convexité de la fonction $h(x) = x^{2+\delta}$. Puis l'inégalité de Jensen pour la même fonction convexe, nous donne $|\mathbb{E}(X_{n,j})|^{2+\delta} \leq \mathbb{E}|X_{n,j}|^{2+\delta}$. On obtient

$$\mathbb{E}|X_{n,j} - \mathbb{E}(X_{n,j})|^{2+\delta} \leq 2^{2+\delta} \mathbb{E}|X_{n,j}|^{2+\delta}$$

Finalement, d'après (5.15), quand $n \rightarrow \infty$

$$\frac{1}{D_n^2} \sum_{j=1}^{k_n} E \left[(X_{n,j} - \mathbb{E}(X_{n,j}))^2 \mathbb{I}_{|X_{n,j} - \mathbb{E}(X_{n,j})| > \varepsilon D_n} \right] \leq \frac{C}{D_n^{2+\delta}} \sum_{j=1}^{k_n} \mathbb{E}|X_{n,j}|^{2+\delta} \rightarrow 0.$$

D'où la conclusion. □