

COURS/TD – PROBABILITÉS ET STATISTIQUE

Ce cours a pour objectif de rappeler, ou de présenter pour la première fois, les différentes notions de probabilités qui seront utiles dans la formation. Toutes ces notions sont fondamentales pour pouvoir comprendre les socles théoriques des modules de statistiques : il est donc important de le travailler sérieusement et régulièrement. Nous tenons à remercier particulièrement Alexandre Genadot et Adrien Richou pour une bonne partie du contenu.

Nous souhaitons que ce cours nécessite un minimum de bagage de théorie de la mesure. Quelques théorèmes importants seront rappelés mais nous n'en ferons pas les démonstrations.

La seconde partie du cours fera l'objet d'un second polycopié : elle sera largement consacrée à une première introduction de la statistique inférentielle (estimation paramétrique ponctuelle et par intervalle de confiance).

Évaluations :

- ♣ des interrogations aléatoires (30 minutes environ),
- ♣ un DNS (devoir non surveillé),
- ♣ un DS,
- ♣ un examen.

Table des matières

1	Espaces de probabilité	3
1.1	Axiomatique des probabilités : l'espace probabilisé	3
1.2	Propriétés élémentaires d'une probabilité	6
1.3	Conditionnement par rapport à un événement	8
1.4	Indépendance d'événements	10
1.5	Exercices additionnels	10
2	Variables et vecteurs aléatoires	13
2.1	Généralités	13
2.1.1	Définition	13
2.1.2	Loi et fonction de répartition	13
2.1.3	Variables aléatoires discrètes	16
2.1.4	Variables aléatoires continues	18
2.2	Espérance	21
2.2.1	Définition et transfert	21
2.2.2	Moments supérieurs	26
2.3	Loi et espérance conditionnelle sachant un événement	28
2.4	Fonctions caractéristique et génératrice	29
2.5	Indépendance et corrélation de variables aléatoires	32
2.5.1	Définition et caractérisation de l'indépendance	33
2.5.2	Covariance et corrélation de deux variables aléatoires réelles	36
2.6	Exercices	38
3	Convergences stochastiques et théorèmes limites	42
3.1	Convergence en probabilité, presque-sûre et L^p	42
3.1.1	Convergence en probabilité	42
3.1.2	Convergence dans L^p	44
3.1.3	Convergence presque-sûre	45
3.2	Convergence en loi	51
3.2.1	Définition et premières propriétés	51
3.2.2	Critère de convergence en loi des suites de variables discrètes	53
3.2.3	Continuité, opérations, projections	53
3.2.4	Liens entre la convergence en loi et les autres convergences aléatoires	55
3.3	Bilan des liens entre les différentes convergences stochastiques	57
3.4	Théorèmes Limites	58
3.4.1	Loi faible des grands nombres	58

3.4.2	Loi forte des grands nombres	58
3.4.3	Le Théorème Central Limite	60
3.4.4	Généralisation des Théorèmes Limites dans le cas de suites de vecteurs aléatoires	61
4	Variables et Vecteurs Aléatoires Gaussiens	63
4.1	Variables aléatoires réelles gaussiennes (ou normales)	63
4.1.1	La loi normale centrée réduite	63
4.1.2	Lois (réelles) normales : Définition et propriétés générales	64
4.2	Vecteurs aléatoires gaussiens : Définition	65
4.2.1	Notations, Rappels et Compléments sur la matrice de (variance-)covariance	65
4.2.2	Définition de la notion de vecteur gaussien	68
4.3	Vecteurs aléatoires gaussiens : Quelques propriétés	69
4.3.1	Fonction caractéristique d'un vecteur gaussien	69
4.3.2	Transformations affines d'un vecteur gaussien	70
4.3.3	Vecteurs gaussiens et Indépendance	70
4.3.4	Densité : Condition d'existence et Formule explicite	72
4.3.5	Retour sur le T.C.L. vectoriel : Compléments et Démonstration	72
4.4	Lois du Khi-deux et de Student	73
4.5	Exercices supplémentaires	75
5	Appendice : rappels sur l'intégrale de Lebesgue.	77
5.1	Pour les fonctions mesurables positives.	77
5.2	Pour des fonctions à valeurs réelles.	78
5.3	Pour des fonctions à valeurs complexes.	79
5.4	Pour des fonctions vectorielles.	80
5.5	Espaces fonctionnels L^p	80

1 ESPACES DE PROBABILITÉ

§ 1.1. Axiomatique des probabilités : l'espace probabilisé.

DEFINITION – 1.1 On appelle *expérience aléatoire* une expérience renouvelable, et qui, renouvelée dans des conditions identiques – pour autant que l'observateur puisse s'en assurer – ne donne pas forcément le même résultat à chaque renouvellement.

Nous allons donner ici le formalisme classique des probabilités. La notion de *tribu* définie ci-dessous joue un rôle fondamental : elle correspond à l'ensemble des résultats d'une expérience aléatoire, sur laquelle on pourra poser une application probabilité, et uniquement sur celle-ci. On ne pourra pas calculer la probabilité de n'importe quoi.

Prenons un ensemble d'évènements, au sens plus heuristique connu au lycée, c'est-à-dire des *résultats* ou *issues possibles* d'une expérience aléatoire. Vous calculez jusque là par exemple ce type de probabilité :

- la probabilité vue comme une fréquence qu'un évènement (formé d'issues possibles) ait lieu,
- la probabilité que deux évènements soient réalisés en même temps (intersection),
- la probabilité que l'un ou l'autre des deux évènements soit réalisé (union),
- la probabilité d'un évènement sachant un autre, etc ...

Une tribu \mathcal{A} va contenir un ensemble de parties d'un ensemble appelé *univers* et noté Ω donné, qui vérifient certaines propriétés de stabilité. L'univers contiendra l'ensemble des résultats d'une expérience aléatoire. Pour donner un sens aux probabilités ci-dessus, il faudrait donc que l'ensemble des évènements soit, au moins, stable par réunion, et passage au complémentaire.

Dans cette partie on cherche donc à donner une définition rigoureuse de la notion d'*univers*, d'*évènement aléatoire* et de *probabilité*.

DEFINITION – 1.2 Soit Ω un ensemble appelé *univers*. On appelle *tribu* (ou σ -*algèbre*) sur Ω un sous-ensemble \mathcal{A} des parties de Ω vérifiant :

- (i) $\emptyset \in \mathcal{A}$,
- (ii) (Stabilité par passage au complémentaire) si $A \in \mathcal{A}$ alors $A^c \in \mathcal{A}$,
- (iii) (Stabilité par réunion dénombrable) si $(A_n)_{n \in \mathbb{N}}$ est une suite d'éléments de \mathcal{A} alors $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$.

Le couple (Ω, \mathcal{A}) est alors appelé *espace mesurable*.

Notez que l'on aurait pu sous (iii) remplacer l'axiome (i) par $\Omega \in \mathcal{A}$. Les éléments de Ω sont appelés les *résultats* ou *issues*, notés généralement ω . Les éléments de \mathcal{A} sont appelés les *évènements*. \emptyset est appelé *évènement impossible* et Ω *évènement quasi-certain*.

À partir des trois axiomes principaux, on peut en déduire d'autres, qui auraient pu être pris d'ailleurs comme définition.

PROPOSITION – 1.1 Soit Ω un ensemble et \mathcal{A} une tribu sur Ω . Alors

- (iv) $\Omega \in \mathcal{A}$,
- (v) \mathcal{A} est stable par réunion finie,
- (vi) \mathcal{A} est stable par intersection dénombrable. En particulier, par intersection finie.

□ PREUVE.

(iv) Comme $\emptyset \in \Omega$, vu que \mathcal{A} est stable par passage au complémentaire, $\emptyset^c = \Omega \in \mathcal{A}$.

(v) On sait que \mathcal{A} est stable par réunion dénombrable. L'idée est qu'une réunion finie est en particulier dénombrable en la prolongeant par le vide. Considérons $A_1, \dots, A_N \in \mathcal{A}$. Posant $A_k = \emptyset$ pour tout $k \geq N + 1$, on a

$$\bigcup_{k=1}^{\infty} A_k = \left(\bigcup_{k=1}^N A_k \right) \cup \emptyset = \bigcup_{k=1}^N A_k \in \mathcal{A}.$$

(vi) Soit $(A_n)_{n \in \mathbf{N}}$ une suite d'éléments de \mathcal{A} . Alors $(A_n^c)_{n \in \mathbf{N}}$ est aussi une suite d'éléments de \mathcal{A} . D'après (iii), on obtient :

$$\left(\bigcup_{n \in \mathbf{N}} A_n^c \right)^c = \bigcap_{n \in \mathbf{N}} (A_n^c)^c = \bigcap_{n \in \mathbf{N}} A_n \in \mathcal{A}.$$

┘

Donnons sans plus tarder deux exemples triviaux de tribus.

EXEMPLE – 1.1 Sur un ensemble quelconque Ω , on peut y mettre deux tribus évidentes :

- $\{\emptyset, \Omega\}$ appelée *tribu grossière*,
- l'ensemble $\mathcal{P}(\Omega)$ des parties de Ω , appelée *tribu discrète*.

Par ailleurs, à partir d'une famille de parties quelconque, on peut construire une tribu qui la contient.

PROPOSITION – 1.2 Soit Ω un ensemble et $\mathcal{C} \subset \mathcal{P}(\Omega)$. Il existe une unique tribu, notée $\sigma(\mathcal{C})$, et appelée *tribu engendrée* par \mathcal{C} , qui contient \mathcal{C} et qui est minimale pour cette propriété : si \mathcal{A}' est une autre tribu qui contient \mathcal{C} alors $\sigma(\mathcal{C}) \subset \mathcal{A}'$. Autrement dit, $\sigma(\mathcal{C})$ est la plus petite tribu sur Ω qui contient \mathcal{C} .

On dit aussi que \mathcal{C} engendre la tribu $\sigma(\mathcal{C})$.

REMARQUE – 1.1 Ce genre de construction est habituelle en Mathématiques : quitte à ajouter des éléments on peut toujours enrichir une structure pour qu'elle vérifie certaines propriétés (ici celle de tribu). On peut par exemple la rapprocher de celle d'espace vectoriel engendré : si E est un espace vectoriel et $(f_1, \dots, f_m) \in E^m$, $\text{Vect}\{f_1, \dots, f_m\}$ est le plus petit sous-espace vectoriel de E contenant $\{f_1, \dots, f_m\}$. On peut montrer qu'il s'agit en fait des combinaisons linéaires finies des f_i .

EXEMPLE – 1.2 Soit Ω un ensemble.

- si $A \in \mathcal{P}(\Omega)$, alors $\{\emptyset, A, A^c, \Omega\}$ est la tribu engendrée par A . Ou encore avec les notations précédentes :

$$\sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\} = \sigma(\{A^c\}).$$

- **TRIBU BORÉLIENNE UNIDIMENSIONNELLE.** On note $\mathcal{B}(\mathbf{R})$ la tribu engendrée par les ouverts de \mathbf{R} . Pour rappel, on dit qu'un sous-ensemble $O \subset \mathbf{R}$ est *ouvert* si

$$\forall x \in O, \exists a, b \in O, x \in]a, b[\subset O.$$

$]1, 2[$ est ouvert, par contre $[1, 2[$ ne l'est pas. Est-il possible de trouver un intervalle ouvert, contenant 1, qui soit dans $]1, 2[$? Notons \mathcal{O} l'ensemble des ouverts de \mathbf{R} . Avec les notations précédentes, on définit la *tribu borélienne* $\mathcal{B}(\mathbf{R})$ sur \mathbf{R} comme la tribu engendrée par ses ouverts :

$$\mathcal{B}(\mathbf{R}) = \sigma(\mathcal{O}).$$

On peut démontrer par ailleurs que

$$\mathcal{B}(\mathbf{R}) = \sigma(\{]-\infty, a[\}_{a \in \mathbf{R}}) = \sigma(\{]a, \infty[\}_{a \in \mathbf{R}}) = \sigma(\{]a, b[\}_{a < b}) = \sigma(\{]a, b] \}_{a < b}).$$

- **TRIBU BORÉLIENNE.** On peut très facilement généraliser à \mathbf{R}^d , $d \geq 1$. Pour rappel, on dit qu'un sous-ensemble $O \subset \mathbf{R}^d$ est *ouvert* si

$$\forall (x_1, \dots, x_d) \in O, \forall i \in \{1, \dots, d\}, \exists a_i, b_i \in O, x_i \in]a_i, b_i[, \prod_{i=1}^d]a_i, b_i[\subset O.$$

Notons \mathcal{O} l'ensemble des ouverts de \mathbf{R}^d . On définit la tribu borélienne sur \mathbf{R}^d par

$$\mathcal{B}(\mathbf{R}^d) = \sigma(\mathcal{O}).$$

On peut démontrer que les pavés l'engendrent

$$\mathcal{B}(\mathbf{R}^d) = \sigma(\{]-\infty, a_1] \times \dots \times]-\infty, a_d[\}_{(a_1, \dots, a_d) \in \mathbf{R}^d}) = \sigma(\{]-\infty, a_1[\times \dots \times]-\infty, a_d[\}_{(a_1, \dots, a_d) \in \mathbf{R}^d}) = \dots$$

DEFINITION – 1.3 (MESURE ET MESURE DE PROBABILITÉ) Étant donné un ensemble Ω et une tribu \mathcal{A} sur Ω , on appelle *mesure de probabilité* (ou plus simplement *probabilité*) sur (Ω, \mathcal{A}) une application $\mathbf{P} : \mathcal{A} \rightarrow [0, \infty]$ telle que

(i) (Additivité dénombrable ou σ -additivité) pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} **deux à deux disjoints** (i.e. $A_n \cap A_m = \emptyset$ pour tout $n \neq m$)

$$\mathbf{P} \left(\bigcup_{n \in \mathbb{N}} A_n \right) = \sum_{n \in \mathbb{N}} \mathbf{P}(A_n),$$

(ii) $\mathbf{P}(\Omega) = 1$.

On appelle *espace mesuré* un triplet $(\Omega, \mathcal{A}, \mathbf{P})$ où \mathcal{A} est une tribu sur Ω , et \mathbf{P} une probabilité sur (Ω, \mathcal{A}) . Un événement A vérifiant $\mathbf{P}(A) = 0$ est dit *négligeable* (ou *quasi-impossible* parfois). Un événement A tel que $\mathbf{P}(A) = 1$ est dit *quasi-certain*, on dit aussi que A a lieu *presque sûrement*, p.s. en abrégé.

Les axiomes (i) et $\mathbf{P}(\Omega) = 1$ sont ceux d'une *mesure* sur $(\Omega, \mathcal{A}, \mathbf{P})$: une probabilité est donc une mesure vérifiant (iii). Remarquez que le membre de gauche de l'égalité du (i) reste inchangé si l'on permute les A_n (l'union ne dépend pas de l'ordre).

REMARQUE – 1.2 Pour que la définition ait un sens, il faut donc que le membre de droite soit aussi invariant par permutation des A_n .

Or, dans toute série absolument convergente, on peut permuter les termes et la somme reste inchangée. Ici c'est bien le cas comme

$$\sum_{n \in \mathbb{N}} |\mathbf{P}(A_n)| = \sum_{n \in \mathbb{N}} \mathbf{P}(A_n) = 1 < \infty.$$

⚠ Attention. Le point important est qu'une probabilité mesure uniquement les éléments de \mathcal{A} , et c'est tout. Avec les objets \mathbf{P} et \mathcal{A} définis plus haut, on ne peut en particulier pas calculer la probabilité d'une union non dénombrable.

⚠ Attention. Dans les exercices, prenez de suite le réflexe de bannir les « probabilités » qui ne seraient pas dans $[0, 1]$.

Citons une propriété d'extension basée sur la notion de tribu engendrée. On l'admettra ici, elle découle d'un lemme plus général appelé *lemme des classes monotones*.

PROPOSITION – 1.3 (MESURE ET TRIBU ENGENDRÉE) Soit (Ω, \mathcal{A}) un espace mesuré par deux mesures finies \mathbf{P} et \mathbf{Q} . Supposons que :

(i) il existe une partie \mathcal{C} , **stable par intersection finie**, telle que $\sigma(\mathcal{C}) = \mathcal{A}$,

(ii) pour tout $A \in \mathcal{C}$, $\mathbf{P}(A) = \mathbf{Q}(A)$,

alors $\mathbf{P} = \mathbf{Q}$ partout.

REMARQUE – 1.3 (MESURE DE LEBESGUE) On note λ la mesure de Lebesgue sur $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ avec $d \in \mathbf{N}^*$. Pour rappel, il s'agit d'une application $\lambda : \mathcal{B}(\mathbf{R}^d) \rightarrow [0, \infty[$ qui unifie le concept de longueur d'intervalle lorsque $d = 1$, d'aire lorsque $d = 2$ et de volume lorsque $d = 3$. Rappelons également, mais ce ne sera pas fondamental dans la suite, comment on la construit :

— pour définir une mesure sur une tribu, il suffit de la définir (cf. théorème d'extension de HAHN) sur une classe d'éléments non vide, stable par différence et union finie qui l'engendre en tant que tribu.

— Ici, $\mathcal{B}(\mathbf{R}^d) = \sigma \left(\{]-\infty, a_1[\times \dots \times]-\infty, a_d[\}_{a_1 < b_1, \dots, a_d < b_d} \right) = \sigma \left(\{]a_1, b_1[\times \dots \times]a_d, b_d[\}_{a_1 < b_1, \dots, a_d < b_d} \right)$, on définit λ par

$$\lambda(]a_1, b_1[\times \dots \times]a_d, b_d[) = \prod_{i=1}^d (b_i - a_i).$$

— On démontre qu'il existe une unique mesure, en utilisant la Proposition 1.3, sur $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ vérifiant cette propriété. On l'appelle *mesure de Lebesgue*.

On constate immédiatement que : si $d = 1$, $\lambda(]a, b[) = b - a$ et si $d = 2$, $\lambda(]a, b[\times]c, d[) = (b - a)(d - c)$ avec $a < b$ et $c < d$. Dans le premier cas on mesure bien la longueur de l'intervalle, dans le second l'aire du pavé.

Comme pour les tribus plus haut, donnons maintenant quelques exemples d'espaces mesurés.

EXEMPLE – 1.3 (MASSE DE DIRAC) Soit Ω un ensemble et $a \in \Omega$. On définit une mesure de probabilité sur $(\Omega, \mathcal{P}(\Omega))$ en posant pour tout $A \in \mathcal{P}(\Omega)$:

$$\delta_a(A) = \begin{cases} 1 & \text{si } a \in A, \\ 0 & \text{sinon.} \end{cases}$$

On vérifie facilement que pour tout $a \in \Omega$, δ_a est une mesure de probabilité sur $(\Omega, \mathcal{P}(\Omega))$, elle est appelée *masse de Dirac* en a .

EXERCICE 1 (PROBABILITÉ UNIFORME DISCRÈTE)

Soit Ω un ensemble fini non vide. Pour tout $A \in \mathcal{P}(\Omega)$, posons

$$\mu_{\Omega}^{\#}(A) = \frac{\#A}{\#\Omega},$$

où $\#A$ désigne le cardinal de A . Alors montrer que $\mu_{\Omega}^{\#}$ définit une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ appelée *probabilité uniforme* sur Ω .

EXERCICE 2 (PROBABILITÉ UNIFORME CONTINUE)

On note λ la mesure de Lebesgue sur $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. Étant donné $a < b$ on définit une application sur $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ en posant pour tout $B \in \mathcal{B}(\mathbf{R})$,

$$\mu^{[a,b]}(B) = \frac{\lambda(B \cap [a,b])}{b-a}.$$

Montrer qu'il s'agit d'une mesure de probabilité. Elle est appelée *probabilité uniforme sur l'intervalle* $[a,b]$. Notez que $\lambda(B \cap [a,b])$ a bien un sens si B est un Borelien.

La notion d'espace mesuré va nous permettre de donner une formulation mathématique aux expériences aléatoires, donc d'associer à chaque expérience un triplet $(\Omega, \mathcal{A}, \mathbf{P})$. En voici quelques exemples :

EXERCICE 3 (UN JEU PILE OU FACE NON TRUQUÉ)

Le pile ou face en une étape est modélisé par l'espace $(\Omega, \mathcal{A}, \mathbf{P})$ avec $\Omega = \{P, F\} =_{def} \{P\} \cup \{F\}$,

$$\mathcal{A} = \mathcal{P}(\Omega) = \{\emptyset, \{P\}, \{F\}, \{P, F\}\},$$

et \mathbf{P} est la probabilité uniforme donnée par $\mathbf{P}(\{P\}) = \mathbf{P}(\{F\}) = \frac{1}{2}$. Vérifier que l'on obtient un espace mesuré.

EXERCICE 4

Une main au poker est modélisée par l'espace $(\Omega, \mathcal{A}, \mathbf{P})$ avec $\Omega = \mathcal{P}_5(\{\text{cartes}\})$ l'ensemble des parties à cinq éléments de l'ensemble des cartes, $\mathcal{A} = \mathcal{P}(\Omega)$, le nombre de cartes en main dépendant de la manche, et \mathbf{P} est la probabilité uniforme (sur Ω) donnée par

$$\forall A \in \mathcal{A}, \quad \mathbf{P}(A) = \frac{\#A}{\#\Omega} = \frac{\#A}{\binom{52}{5}}.$$

Vérifier que l'on obtient un espace mesuré.

EXERCICE 5 (COMMENT MODÉLISER UN JEU INFINI DE PILE OU FACE ?)

Il est naturel d'introduire comme espace des états l'ensemble Ω des suites $(\omega_1, \omega_2, \dots)$ où les ω_i valent 0 ou 1 (on note 1 lorsqu'on fait pile et 0 lorsqu'on fait face). On a ainsi

$$\Omega = \{0, 1\}^{\mathbf{N}^*},$$

l'ensemble des suites à valeurs dans $\{0, 1\}$. c'est-à-dire $\Omega = \{P, F\}^{\mathbf{N}^*}$.

On définit une tribu \mathcal{A} sur Ω de la façon suivante : la tribu \mathcal{A} est la tribu engendrée par tous les événements $C_{i,\varepsilon}$

$$C_{i,\varepsilon} = \{\omega \in \Omega, \quad \omega_i = \varepsilon\},$$

où i décrit \mathbf{N}^* et $\varepsilon \in \{0, 1\}$. On l'appelle en général *tribu cylindrique*. Par exemple, l'événement $C_{10,1}$ correspond à faire pile au dixième lancer. Si on joue avec une pièce qui donne pile avec probabilité p , il est alors naturel d'associer à tout événement $C_{i,\varepsilon}$ la probabilité p si $\varepsilon = 1$ et 0 sinon. On peut alors montrer qu'il existe une unique mesure de probabilité \mathbf{P} définie sur (Ω, \mathcal{A}) telle que pour tous $n \in \mathbf{N}^*$, $i_1, \dots, i_n \in \mathbf{N}^*$, $\varepsilon_1, \dots, \varepsilon_n \in \{0, 1\}$ on ait

$$\mathbf{P}(C_{i_1, \varepsilon_1} \cap \dots \cap C_{i_n, \varepsilon_n}) = p^{\varepsilon_1 + \dots + \varepsilon_n} (1-p)^{n - (\varepsilon_1 + \dots + \varepsilon_n)}.$$

Vérifier que l'on obtient un espace mesuré.

§ 1.2. Propriétés élémentaires d'une probabilité.

PROPOSITION – 1.4 Un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ vérifie les propriétés suivantes, pour deux événements $A, B \in \mathcal{A}$,

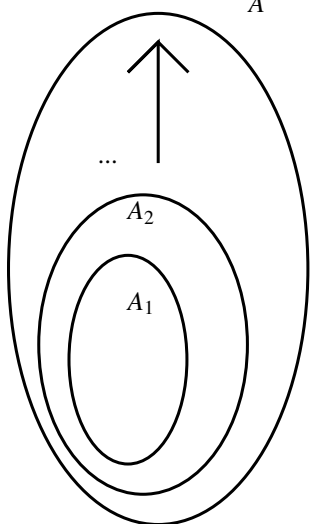
- (Différence) $\mathbf{P}(B \setminus A) = \mathbf{P}(B) - \mathbf{P}(A \cap B)$. En particulier si $A \subset B$ alors $\mathbf{P}(B \setminus A) = \mathbf{P}(B) - \mathbf{P}(A)$.
- (Formule d'inclusion/exclusion) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$,

- (Additivité) si $A \cap B = \emptyset$ alors $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$. En particulier, $\mathbf{P}(\emptyset) = 0$ et $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$,
- (Monotonie pour \subset) si $A \subset B$ alors $\mathbf{P}(A) \leq \mathbf{P}(B)$.

EXERCICE 6

Faire la démonstration de la proposition précédente.

Dans toute la suite du cours, étant donné une suite réelle $(u_n)_{n \in \mathbf{N}}$ et $l \in \mathbf{R} \cup \{\infty\}$, on écrit $u_n \nearrow l$ lorsque la suite $(u_n)_{n \in \mathbf{N}}$ est croissante et tend vers l . De même, si $l \in \mathbf{R} \cup \{-\infty\}$, on écrit $u_n \searrow l$ lorsque la suite $(u_n)_{n \in \mathbf{N}}$ est décroissante et tend vers l . Étant donnée une suite $(A_n)_{n \in \mathbf{N}}$ d'ensembles, la notation $A_n \nearrow A$ signifie que la suite est croissante (pour l'inclusion, i.e. $A_n \subset A_{n+1}$ pour tout $n \in \mathbf{N}$) et que $\bigcup_{n \in \mathbf{N}} A_n = A$. On dit aussi que $(A_n)_{n \in \mathbf{N}}$ est une *suite croissante exhaustive de parties* de A .



Enfin $A_n \searrow A$ signifie que la suite est décroissante pour l'inclusion et que $\bigcap_{n \in \mathbf{N}} A_n = A$.

PROPOSITION – 1.5 (CROISSANCE MONOTONE) Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité, $(A_n)_{n \in \mathbf{N}}$ une suite de parties de \mathcal{A} et $A \in \mathcal{A}$. Alors

$$A_n \nearrow A \implies \mathbf{P}(A_n) \nearrow \mathbf{P}(A).$$

REMARQUE – 1.4 L'hypothèse de monotonie est fondamentale. Considérons par exemple \mathbf{P} la probabilité uniforme sur $\{0, 1\}$ et posons $A_0 = \{0\}$ et $A_n = \{1\}$ pour $n \in \mathbf{N}^*$. Alors $\lim_{n \rightarrow \infty} \mathbf{P}(A_n) = \frac{1}{2}$ mais

$$\mathbf{P}\left(\bigcup_{n \in \mathbf{N}} A_n\right) = \mathbf{P}(\{0, 1\}) = 1.$$

▮ **PREUVE.** On pose $C_0 = A_0$ et $C_n = A_n \setminus A_{n-1}$ pour tout $n \in \mathbf{N}^*$. Les propriétés suivantes sont laissées en exercice :

- les C_n sont deux à deux disjoints,
- $A_n = \bigcup_{k=0}^n C_k$ pour tout $n \in \mathbf{N}$,
- $\bigcup_{k=0}^{\infty} C_k = \bigcup_{k=0}^{\infty} A_k$.

Par σ -additivité, on obtient :

$$\mathbf{P}(A_n) = \sum_{k=0}^n \mathbf{P}(C_k) \nearrow \sum_{k=0}^{\infty} \mathbf{P}(C_k) = \mathbf{P}\left(\bigcup_{k=0}^{\infty} C_k\right) = \mathbf{P}\left(\bigcup_{k=0}^{\infty} A_k\right).$$

▮

Citons immédiatement la propriété analogue pour l'intersection.

PROPOSITION – 1.6 Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité et soient $(B_n)_{n \in \mathbf{N}}$, B des éléments de \mathcal{A} . Alors :

$$B_n \searrow B, \implies \mathbf{P}(B_n) \searrow \mathbf{P}(B).$$

Si la suite $(A_n)_{n \in \mathbf{N}}$ n'est pas monotone, on a en revanche qu'une inégalité.

PROPOSITION – 1.7 Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité et soit $(A_n)_{n \in \mathbf{N}}$ une suite d'événements de \mathcal{A} . On a alors la propriété de *sous-additivité dénombrable* :

$$\mathbf{P}\left(\bigcup_{n \in \mathbf{N}} A_n\right) \leq \sum_{n=0}^{\infty} \mathbf{P}(A_n).$$

EXERCICE 7

Faire la preuve de la proposition.

§ 1.3. Conditionnement par rapport à un évènement.

Un énoncé de probabilité conditionnelle est un énoncé du type : « si B se produit alors la probabilité que A se produise est p ». On peut penser par exemple que B est l'évènement « il neige » et A « le bus est en retard ». Si je sais qu'il neige, la probabilité que le bus soit en retard est augmentée (dans le cas de la TBM, on peut remplacer l'évènement B par « je prends le TRAM » ou tout autre évènement trivial ...).

Autre exemple, si on joue au jeu suivant : il y a trois portes, derrière l'une des trois se cache un trésor. J'ai le droit de regarder successivement derrière chacune des portes et d'en choisir une à chaque tour. Sachant que j'ai déjà regardé une porte, j'ai plus de chance au tour suivant de tomber sur la bonne.

Il faut donc voir le conditionnement comme de l'information disponible. Mathématiquement cela donne la définition suivante.

DEFINITION – 1.4 Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace probabilisé. Soit B un évènement vérifiant $\mathbf{P}(B) > 0$ (B est donc supposé **non-négligeable**). Pour tout évènement A on pose

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

Cette quantité est appelée *probabilité de A sachant B* ou *probabilité conditionnellement à B* .

⚠ Attention. Nous n'avons pas défini l'évènement $\{A | B\}$, seulement la probabilité conditionnelle sachant B . À ne pas confondre avec l'évènement $A \setminus B := A \cap B^c$.

REMARQUE – 1.5 Un évènement B non négligeable étant fixé, la probabilité conditionnelle sachant B est simplement :

$$\mathbf{P}(A|B) = \mathbf{P}_B(A),$$

où \mathbf{P}_B est la probabilité sur (Ω, \mathcal{A}) définie par :

$$\forall C \in \mathcal{A}, \quad \mathbf{P}_B(C) = \frac{\mathbf{P}(C \cap B)}{\mathbf{P}(B)}.$$

Finalement, une probabilité conditionnelle est **quasiment** la probabilité d'intersection, il faut cependant rajouter un facteur $\frac{1}{\mathbf{P}(B)}$ pour faire de \mathbf{P}_B une mesure de probabilité.

EXERCICE 8

Une famille a deux enfants. Quelle est la probabilité que les deux soient des garçons conditionnellement au fait qu'au moins l'un des deux est un garçon ?

► **CORRECTION.**

Avec des notations évidentes, l'espace de probabilité est $\Omega = \{GG, GF, FG, FF\}$ muni de la probabilité uniforme. La probabilité cherchée est

$$\mathbf{P}(GG|\{GG, GF, FG\}) = \frac{\mathbf{P}(\{GG\})}{\mathbf{P}(\{GG, GF, FG\})} = \frac{1/4}{3/4} = \frac{1}{3}.$$

On en déduit directement de la définition une formule de permutation de conditionnement appelée *formule de BAYES*.

PROPOSITION – 1.8 (FORMULE DE BAYES) Soit A et B des événements non négligeables, alors

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

La formule des probabilités totales, rappelée ci-dessous, se reformule avec des conditionnements.

DEFINITION – 1.5 Soient Ω un ensemble et I une partie dénombrable. Considérons $(A_i)_{i \in I}$ une famille de parties de Ω . On dit que $(A_i)_{i \in I}$ est une *partition* de Ω si les deux propriétés suivantes sont vérifiées :

- (i) $\bigcup_{i \in I} A_i = \Omega$,
- (ii) les A_i sont deux à deux disjoints.

Autrement dit, la famille $(A_i)_{i \in I}$ est une partition de Ω si pour tout $\omega \in \Omega$, il existe un et un seul $i \in I$ tel que $\omega \in A_i$.

EXEMPLE – 1.4 Pour tout $A \subset \Omega$, l'ensemble $\{A, A^c\}$ est une partition de Ω .

PROPOSITION – 1.9 (FORMULE DES PROBABILITÉS TOTALES) Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace probabilisé, I une partie dénombrable et $(B_i)_{i \in I}$ une partition de Ω en événements. Alors, pour tout événement A ,

$$\begin{aligned} \mathbf{P}(A) &= \sum_{i \in I} \mathbf{P}(A \cap B_i) \\ &= \sum_{i \in I} \mathbf{P}(A|B_i) \mathbf{P}(B_i), \quad \text{si } \mathbf{P}(B_i) > 0 \text{ pour tout } i. \end{aligned}$$

▮ **PREUVE.** Comme les B_i recouvrent Ω on a

$$A = A \cap \Omega = A \cap \left(\bigcup_{i \in I} B_i \right) = \bigcup_{i \in I} (A \cap B_i).$$

Et comme les B_i sont deux à deux disjoints, les $A \cap B_i$ le sont aussi. De plus, I est fini ou dénombrable donc

$$\mathbf{P}(A) = \sum_{i \in I} \mathbf{P}(A \cap B_i).$$

▮

REMARQUE – 1.6 Un cas particulier qui revient souvent est la formule

$$\mathbf{P}(A) = \mathbf{P}(A|B)\mathbf{P}(B) + \mathbf{P}(A|B^c)\mathbf{P}(B^c)$$

valable dès que $0 < \mathbf{P}(B) < 1$. C'est le cas associé à la partition $A \cup A^c = \Omega$.

EXERCICE 9 (FAUX POSITIFS)

Une maladie affecte une personne sur 1000. Le test de dépistage n'est pas parfait : le résultat est toujours positif pour une personne malade et pour une personne saine il est positif (donc erroné) 2 fois sur 100. Quelle est la probabilité qu'une personne ayant un résultat positif au test soit effectivement malade ?

►CORRECTION.

Soit T l'événement « le test est positif » et M l'événement « la personne est malade ». On cherche $\mathbf{P}(M|T)$. On écrit

$$\mathbf{P}(M|T) = \mathbf{P}(T|M) \frac{\mathbf{P}(M)}{\mathbf{P}(T)}.$$

D'après les données du problème $\mathbf{P}(T|M) = 1$ et $\mathbf{P}(M) = 0,001$. De plus

$$\mathbf{P}(T) = \mathbf{P}(T|M)\mathbf{P}(M) + \mathbf{P}(T|M^c)\mathbf{P}(M^c) = 1 \times 0,001 + 0,02 \times 0,999.$$

En regroupant tout, on trouve que $\mathbf{P}(M|T)$ est de l'ordre de 5%. Le test est probablement erroné.

§ 1.4. Indépendance d'événements.

De façon intuitive, on dit que A est indépendant de B si savoir B ne change pas la probabilité de A . C'est-à-dire si

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

Pour que cette formule ait un sens on est obligé de supposer que $\mathbf{P}(B) > 0$, ce qui n'est pas le cas dans la définition suivante. Mais si c'est le cas l'égalité précédente signifie simplement

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

DEFINITION – 1.6 Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace probabilisé. Deux événements A et B sont dits *indépendants* si

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

Plus généralement, soit $(A_i)_{i \in I}$ une famille d'événements avec I dénombrable. On dit que les A_i sont *mutuellement indépendants* (on dit parfois seulement *indépendants*) si

$$\mathbf{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbf{P}(A_j)$$

pour tout J fini inclus dans I . On dit que les A_i sont *indépendants deux à deux* si pour tous $i, j \in I$, A_i et A_j sont indépendants.

EXEMPLE – 1.5 On tire une carte dans un paquet de 52. L'événement A : « tirer un roi » est indépendant de l'événement B : « tirer un pique ». En effet $\mathbf{P}(A \cap B)$ est la probabilité de tirer le roi de pique, soit $1/52$, qui est bien égal à $\mathbf{P}(A)\mathbf{P}(B) = (4/52) \times (13/52) = 1/52$.

EXERCICE 10

Montrer qu'un événement A est indépendant de lui-même si et seulement si $\mathbf{P}(A) \in \{0, 1\}$.

►CORRECTION.

Écrire la définition, et constater que $\mathbf{P}(A)$ vérifie l'équation $x^2 = x$.

EXEMPLE – 1.6 (POUR TROIS ÉVÈNEMENTS) Montrer que A , B et C sont indépendants si les quatre propriétés suivantes sont vérifiées

$$\begin{aligned}\mathbf{P}(A \cap B) &= \mathbf{P}(A)\mathbf{P}(B), \\ \mathbf{P}(A \cap C) &= \mathbf{P}(A)\mathbf{P}(C), \\ \mathbf{P}(B \cap C) &= \mathbf{P}(B)\mathbf{P}(C), \\ \mathbf{P}(A \cap B \cap C) &= \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C).\end{aligned}$$

Les trois premières propriétés signifient que les événements A , B et C sont deux à deux indépendants.

EXERCICE 11

On jette deux pièces. Montrer que les événements suivants

- la première pièce tombe sur pile,
- la deuxième pièce tombe sur pile,
- les deux pièces donnent le même résultat,

sont deux à deux indépendants, mais pas mutuellement indépendants.

§ 1.5. Exercices additionnels.

EXERCICE 12

On tire simultanément 5 cartes au hasard dans un jeu de 32 cartes. Quelle est la probabilité :

- 1) p_2 de tirer exactement deux as ?
- 2) p_3 de tirer au moins deux as ?
- 3) p_4 de tirer une double paire ?
- 4) p_5 de tirer un full (trois cartes d'une valeur et deux d'une seconde valeur) ?

On pensera, pour une fois, à modéliser l'expérience aléatoire, c'est-à-dire expliciter l'espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ sous-jacent.

EXERCICE 13

On jette deux dés non pipés, un dé noir et un dé blanc. Soit A l'évènement « le chiffre du dé noir est pair », B l'évènement « le chiffre du dé blanc est impair » et C l'évènement « les deux chiffres ont même parité ». Montrer que ces trois évènements sont indépendants deux à deux mais ne sont pas mutuellement indépendants.

EXERCICE 14 (ROULETTE RUSSE)

Un revolver à 6 coups contient une seule balle mais on ne sait pas à quel endroit du barillet. Le premier joueur place le revolver sur sa tempe et presse la gâchette. S'il survit le deuxième joueur fait de même. Vaut-il mieux jouer en premier ou en second ? La taille du barillet importe-t-elle ?

EXERCICE 15

Une urne contient six boules dont 4 blanches et 2 noires. On extrait une boule de l'urne, on note sa couleur, puis on la remet dans l'urne. On effectue ensuite des tirages sans remise jusqu'à l'obtention d'une boule de la même couleur que précédemment. Soit $k \in \mathbf{N}^*$. Déterminer la probabilité que le nombre de tirage après remise de la boule initialement tirée soit k .

EXERCICE 16

Une urne U contient a boules blanches et b boules rouges tandis qu'une urne V contient b boules blanches et a boules rouges. On effectue une suite de tirages successifs d'une boule dans U ou dans V selon les règles suivantes. On commence par tirer une boule dans U . Si on a obtenu une boule blanche, le tirage suivant s'effectue dans U , et si on a obtenu une boule rouge, le tirage suivant s'effectue dans V . Après chaque tirage, la boule est remise dans l'urne dont elle provient. Pour $n \geq 1$, soit p_n la probabilité d'obtenir une boule blanche au n ème tirage.

- 1) Montrer, pour tout $n \geq 1$, la relation $p_{n+1} = cp_n + d$ avec c et d à déterminer.
- 2) En déduire la valeur de p_n et sa limite quand n tend vers l'infini.

EXERCICE 17

Un lot de montres identiques est reçu par un détaillant parisien. Celui-ci provient de façon équiprobable soit de Hong-Kong, soit de Singapour. L'usine de Hong-Kong produit un article défectueux sur 1000 en moyenne, celle de Singapour un sur 200. Le détaillant inspecte une première montre : elle marche. Sachant ceci, quelle est la probabilité que la deuxième montre inspectée marche elle aussi ?

EXERCICE 18

Dans un élevage de poulets un animal est atteint par une maladie M avec une probabilité de 0,2. Un test de dépistage est utilisé. La probabilité qu'un poulet non atteint par M soit négatif au test est de 0,9. La probabilité qu'un poulet atteint par M soit positif au test est de 0,8. Quelle est la probabilité qu'un poulet pris au hasard et ayant une réaction positive soit atteint par M ?

EXERCICE 19

Un jeu télévisé oppose un présentateur à un candidat. Ce candidat est placé devant trois portes fermées numérotées de 1 à 3 : une voiture a été placée de manière aléatoire derrière l'une d'elles tandis qu'il n'y a rien derrière les deux autres. Le candidat commence par désigner la porte 1. Puis le présentateur, sachant quelle est la bonne porte dès le début, aide le candidat en ouvrant une porte vide parmi les portes restantes : il choisit d'ouvrir la porte 2. Il propose ensuite au candidat de choisir entre ouvrir la porte 1 qu'il a choisie initialement, ou bien ouvrir la porte 3. Déterminer si le candidat augmente ses chances de gagner la voiture en changeant son choix initial.

EXERCICE 20

On choisit au hasard un des n premiers entiers $1, 2, \dots, n$. Soit $1 \leq p \leq n$ et soit A_p l'évènement correspondant à ce que le nombre choisi soit divisible par p .

- 1) Calculer $\mathbf{P}(A_p)$ lorsque p divise n .
- 2) Si p_1, p_2, \dots, p_k sont des diviseurs premiers de n distincts, montrer que les évènements $A_{p_1}, A_{p_2}, \dots, A_{p_k}$ sont indépendants.

- 3) On appelle *fonction indicatrice d'Euler* la fonction ϕ définie sur les entiers naturels dont la valeur $\phi(n)$ est égale au nombre d'entiers inférieurs à n et premiers avec n . Montrer que

$$\phi(n) = n \prod_{i=1}^k \left(1 - \frac{1}{p_i}\right).$$

EXERCICE 21

On considère n « menteurs » I_1, I_2, \dots, I_n . Le premier menteur I_1 reçoit une information sous la forme de « oui » ou « non ». Il transmet l'information à I_2 , ainsi de suite jusqu'à I_n qui l'annonce au monde. Chacun des menteurs transmet ce qu'il a entendu avec la probabilité p et le contraire avec la probabilité $1 - p$ où $0 < p < 1$. De plus, les réponses des n individus sont indépendantes.

- 1) Soit p_n la probabilité que l'information soit fidèlement transmise. Déterminer une relation liant p_n et p_{n+1} .
- 2) En déduire la valeur de p_n et sa limite lorsque n tend vers l'infini.

EXERCICE 22

Un livre contient 4 erreurs. À chaque relecture une faute non corrigée est corrigée avec une probabilité $1/3$. Les relectures sont indépendantes les unes des autres.

- 1) Combien faut-il faire de relectures pour que la probabilité qu'il ne subsiste aucune erreur soit supérieure à 0,90 ?
- 2) Traiter la même question en supposant que le nombre x d'erreurs est réparti uniformément sur $\{0, 1, 2, 3, 4\}$.

2 VARIABLES ET VECTEURS ALÉATOIRES

§ 2.1. Généralités

2.1.1 Définition

On ne sera pas toujours intéressé par le résultat complet d'une expérience aléatoire mais plutôt par une conséquence de ce résultat, c'est-à-dire une fonction de ce dernier. Une telle fonction, ayant une propriété dite de *mesurabilité*, est appelée *variable aléatoire* si la quantité d'intérêt est à valeurs dans \mathbf{R} ou *vecteur aléatoire* si la quantité est à valeurs dans \mathbf{R}^d , $d \geq 2$. Voici donc un tout premier exemple très simple.

EXERCICE 23

On lance deux pièces. On appelle X le nombre de « pile » obtenus et on pose $Y = 1$ si on obtient au moins une face et 0 sinon. Les quantités X et Y sont des variables aléatoires et le couple (X, Y) forme un vecteur aléatoire. De manière formelle, on a $\Omega = \{PP, PF, FP, FF\}$, muni de la probabilité uniforme si l'expérience se déroule avec des pièces équilibrées, et X et Y sont définies par

$$\begin{aligned} X(PP) &= 2, & Y(PP) &= 0, & X(PF) &= 1, & Y(PF) &= 1, \\ X(FP) &= 1, & Y(FP) &= 1, & X(FF) &= 0, & Y(FF) &= 1. \end{aligned}$$

Le vecteur (X, Y) est lui défini par

$$(X, Y)(PP) = (X(PP), Y(PP)) = (2, 0), \quad (X, Y)(PF) = (1, 1), \quad (X, Y)(FP) = (1, 1), \quad (X, Y)(FF) = (0, 1).$$

Il sera fait référence à cet exemple dans la suite. Pour toute la suite du chapitre, on se donne un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ et un espace mesurable (E, \mathcal{E}) égal quasiment toujours à $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ pour un certain $d \in \mathbf{N}^*$.

DEFINITION – 2.1 Une application $X : \Omega \rightarrow E$ est dite *mesurable* si

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

Notez bien que la notion de mesurabilité dépend des tribus sous-jacentes associées aux espaces de départ et d'arrivée.

Notations. Pour simplifier, nous noterons parfois

- **v.a.r.** (ou parfois simplement **v.a.**) pour **variable aléatoire réelle**.
- **i.e.** pour abrégé *c'est-à-dire* (puisque *i.e.* est l'abréviation de l'expression latine *id est* qui signifie *c'est-à-dire*).

DEFINITION – 2.2 Une application mesurable $X : \Omega \rightarrow E$ est

- une *variable aléatoire réelle* si $E = \mathbf{R}$,
- une *variable aléatoire* ou un *vecteur aléatoire* si $E = \mathbf{R}^d$ avec $d \geq 2$. Dans ce cas, $X = (X_1, \dots, X_d)$ où les X_i sont des variables aléatoires appelées *composantes* ou *marginales* de X ,
- une *variable aléatoire complexe* s'il existe X_1 et X_2 deux variables aléatoires réelles telles que $X = X_1 + iX_2$.

La troisième définition est équivalente à la mesurabilité par rapport à la tribu borélienne complexe (les ouverts étant définis par rapport au module).

On dira simplement variable aléatoire si il n'y a pas lieu de distinguer le cas réel du cas vectoriel. La notion de mesurabilité permet donc de mesurer toutes les parties du type $X^{-1}(A)$ avec $A \in \mathcal{E}$. En considérant ce type d'application, on pourra donc calculer des probabilités que X soit dans un certain ensemble de \mathcal{E} .

Dans l'Exercice 23, on connaissait les valeurs de X et Y associées à chaque évènement. En pratique, on connaît seulement la *loi* de X et Y , c'est-à-dire seulement les valeurs que prennent X et Y avec une certaine probabilité.

Formalisons cela dans la suite.

2.1.2 Loi et fonction de répartition

DEFINITION – 2.3 Soit $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$ une variable aléatoire. On appelle *loi de X* la mesure de probabilité \mathbf{P}_X sur (E, \mathcal{E}) donnée, pour $A \in \mathcal{E}$, par

$$\mathbf{P}_X(A) = \mathbf{P}(X^{-1}(A)).$$

On notera même souvent pour $\mathbf{P}(X^{-1}(A))$, $\mathbf{P}(X \in A)$.

REMARQUE – 2.1 En théorie de la mesure, \mathbf{P}_X est la mesure image de la probabilité \mathbf{P} par la variable aléatoire X . Plus généralement, n'importe quelle application mesurable permet de transporter une mesure de probabilité sur l'espace mesuré de départ, vers l'espace mesurable d'arrivée.

DEFINITION – 2.4 On appelle *fonction de répartition*

— de la variable aléatoire réelle X , la fonction $F_X : \mathbf{R} \rightarrow [0, 1]$ donnée, pour $x \in \mathbf{R}$, par

$$\begin{aligned} F_X(x) &= \mathbf{P}_X([-\infty, x]) \\ &= \mathbf{P}(X \leq x). \end{aligned}$$

— ou plus généralement du vecteur aléatoire $X = (X_1, \dots, X_d)$, pour $d \geq 2$, l'application $F_{(X_1, \dots, X_d)} : \mathbf{R}^d \rightarrow [0, 1]$ donnée, pour $(x_1, \dots, x_d) \in \mathbf{R}^d$, par

$$\begin{aligned} F_{(X_1, \dots, X_d)}(x_1, \dots, x_d) &= \mathbf{P}_{(X_1, \dots, X_d)}([-\infty, x_1] \times \dots \times [-\infty, x_d]) \\ &= \mathbf{P}(X_1 \leq x_1, \dots, X_d \leq x_d). \end{aligned}$$

On distinguera souvent par souci de clarté le cas unidimensionnel du cas vectoriel.

EXERCICE 24

On considère une variable aléatoire X à valeurs dans $\{1, \dots, 6\}$ donnant le résultat du lancé d'un dé à 6 faces. Donner la loi de X et calculer sa fonction de répartition.

EXERCICE 25

Si on considère le vecteur aléatoire (X, Y) de l'Exemple 23, sa fonction de répartition est une fonction de deux variables donnée, pour $(x, y) \in \mathbf{R}^2$, par

$$\begin{aligned} F_{(X, Y)}(x, y) &= \frac{1}{4} \mathbb{1}_{[0, 1] \times [1, +\infty[}(x, y) + \frac{1}{4} \mathbb{1}_{[2, +\infty[\times [0, 1]}(x, y) + \frac{3}{4} \mathbb{1}_{[1, 2] \times [1, +\infty[}(x, y) + \mathbb{1}_{[2, +\infty[\times [1, +\infty[}(x, y). \end{aligned}$$

Remarquons que

$$\lim_{y \rightarrow \infty} F_{(X, Y)}(x, y) = \frac{1}{4} \mathbb{1}_{[0, 1]}(x) + \frac{3}{4} \mathbb{1}_{[1, 2]}(x) + \mathbb{1}_{[2, +\infty[}(x) = F_X(x).$$

Ce qui, comme nous le verrons, est un fait général. Le passage à la limite permet d'obtenir la fonction de répartition de chacune des marginales. Passons maintenant à quelques propriétés systématiques des fonctions de répartition.

PROPOSITION – 2.1 (PROPRIÉTÉS ANALYTIQUES) Soit $X : \Omega \rightarrow \mathbf{R}$ une variable aléatoire et F_X sa fonction de répartition. Alors

- (i) F_X est croissante,
- (ii) F_X est continue à droite, et donc possède une limite à gauche,
- (iii) $\lim_{x \rightarrow -\infty} F_X(x) = 0$ et $\lim_{x \rightarrow \infty} F_X(x) = 1$.

On écrit parfois aussi que F_X est càdlag pour la propriété (ii).

La continuité à droite pour F_X en a s'écrit : $F_X \rightarrow_{a+} F(a)$. Ou séquentiellement :

$$\text{pour toute suite } (x_n)_{n \in \mathbf{N}} \text{ telle que } x_n \rightarrow_{n \rightarrow \infty} a^+, \quad F(x_n) \rightarrow_{n \rightarrow \infty} F(a).$$

▮ **PREUVE.** Soit $x \leq y$, alors $\{X \leq x\} \subset \{X \leq y\}$ donc $\mathbf{P}(X \leq x) \leq \mathbf{P}(X \leq y)$, ce qui montre que F_X est croissante. Soit $x \in \mathbf{R}$, on sait que F_X est croissante bornée par 0 et 1, donc d'après le théorème de la limite monotone, F_X admet une limite à droite et à gauche en chacun de ses points. En conséquence

$$\lim_{y \rightarrow x, y > x} F_X(y) = \lim_{n \rightarrow \infty} F_X\left(x + \frac{1}{n}\right).$$

Remarquons ensuite que $\left\{X \leq x + \frac{1}{n}\right\} \searrow \{X \leq x\}$. Donc, par convergence monotone établie dans le premier chapitre,

$$F_X\left(x + \frac{1}{n}\right) = \mathbf{P}\left(X \leq x + \frac{1}{n}\right) \searrow \mathbf{P}(X \leq x) = F_X(x).$$

Ainsi F_X est continue à droite en x . Or $\{X \leq -n\} \searrow \{X \leq -\infty\} = \emptyset$ donc

$$F_X(-n) = \mathbf{P}(X \leq -n) \searrow 0.$$

Enfin, $\{X \leq n\} \nearrow \{X \leq +\infty\} = \Omega$ donc $F_X(n) = \mathbf{P}(X \leq n) \nearrow 1$, d'où (iii). ┘

On notera la *limite à gauche* par $F_X(x-)$:

$$F_X(x-) = \lim_{y \rightarrow x, y < x} F_X(y).$$

REMARQUE – 2.2 Ces propriétés caractérisent la notion de fonction de répartition au sens suivant : si F_X est une fonction vérifiant les trois propriétés précédentes alors on peut trouver un espace de probabilité $(\Omega, \mathcal{A}, \mathbf{P})$ et une variable aléatoire $X : \Omega \rightarrow \mathbf{R}$ telle que $\mathbf{P}(X \leq x) = F_X(x)$ pour tout $x \in \mathbf{R}$.

PROPOSITION – 2.2 (PROPRIÉTÉS PROBABILISTES) Soit $X : \Omega \rightarrow \mathbf{R}$ une variable aléatoire et F_X sa fonction de répartition. Soient $x \leq y$ des réels, on a

(i) $\mathbf{P}(X > x) = 1 - F_X(x)$,

(ii) $\mathbf{P}(x < X \leq y) = F_X(y) - F_X(x)$,

(iii) $\mathbf{P}(X < x) = F_X(x-)$,

(iv) $\mathbf{P}(X = x) = F_X(x) - F_X(x-)$.

En particulier, F_X est continue en x si et seulement si $\mathbf{P}(X = x) = 0$.

Lorsque pour tout $x \in \mathbf{R}$, $\mathbf{P}(X = x) = 0$ on dit que X est *sans atome*. F_X est alors continue en chaque point : la continuité à droite est systématique, celle à gauche découle de (iv).

┐ **PREUVE.** Les propriétés (i) et (ii) découlent directement de la définition de F_X , en écrivant que pour tous $x, y \in \mathbf{R}$ tels que $x \leq y$,

$$\{x < X \leq y\} = \{X \leq y\} \setminus \{X \leq x\},$$

on passe ensuite à la probabilité. Pour (iii), on remarque que $\left\{X \leq x - \frac{1}{n}\right\} \nearrow \{X < x\}$, donc par convergence monotone,

$$F_X\left(x - \frac{1}{n}\right) = \mathbf{P}\left(X \leq x - \frac{1}{n}\right) \nearrow \mathbf{P}(X < x).$$

Comme on a aussi $\lim_{n \rightarrow \infty} F_X\left(x - \frac{1}{n}\right) = F_X(x-)$ puisque F_X admet une limite à gauche en tout point, (iii) est démontrée. On obtient alors (iv) en écrivant $\mathbf{P}(X = x) = \mathbf{P}(X \leq x) - \mathbf{P}(X < x)$. ┘

EXERCICE 26

On reprend la variable aléatoire X de l'exemple 23. On a $F_X(1-) = \frac{1}{4}$ et $F_X(1) = \frac{3}{4}$ ce qui montre que $\mathbf{P}(X = 1) = F_X(1) - F_X(1-) = \frac{1}{2}$.

Comme remarqué dans l'exemple, explicitons maintenant le lien entre la fonction de répartition d'un vecteur aléatoire et la fonction de répartition de ses marginales : c'est *in fine* un simple passage à la limite.

PROPOSITION – 2.3 Pour $d \in \mathbf{N}^*$, soit (X_1, \dots, X_d) un vecteur aléatoire et $F_{(X_1, \dots, X_d)}$ sa fonction de répartition. Alors la fonction de répartition de la composante X_i pour $i \in \{1, \dots, d\}$ est donnée, pour $x \in \mathbf{R}$, par

$$F_{X_i}(x) = \lim_{(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d) \rightarrow (\infty, \dots, \infty)} F_{(X_1, \dots, X_d)}(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_d).$$

▮ PREUVE. Soient $x_1^{(k)}, \dots, x_{i-1}^{(k)}, x_{i+1}^{(k)}, \dots, x_d^{(k)}$, $d - 1$ suites croissantes tendant toutes vers l'infini. Alors

$$\{X_1 \leq x_1^{(k)}, \dots, X_{i-1} \leq x_{i-1}^{(k)}, X_i \leq x, X_{i+1} \leq x_{i+1}^{(k)}, \dots, X_d \leq x_d^{(k)}\} \nearrow \{X_i \leq x\}.$$

Donc, par convergence monotone,

$$\begin{aligned} F_{(X_1, \dots, X_n)}(x_1^{(k)}, \dots, x_{i-1}^{(k)}, x, x_{i+1}^{(k)}, \dots, x_d^{(k)}) &= \mathbf{P}(X_1 \leq x_1^{(k)}, \dots, X_{i-1} \leq x_{i-1}^{(k)}, X_i \leq x, X_{i+1} \leq x_{i+1}^{(k)}, \dots, X_n \leq x_d^{(k)}) \\ &\nearrow \mathbf{P}(X_i \leq x) \\ &= F_{X_i}(x). \end{aligned}$$

▮

Le théorème suivant est fondamental et permet de répondre au problème initial, qui est que la fonction de répartition caractérise la loi.

THÉORÈME – 2.1 La fonction de répartition *caractérise la loi*. Plus précisément, si X et Y sont deux variables aléatoires définies sur un même espace mesuré $(\Omega, \mathcal{A}, \mathbf{P})$ ayant même fonction de répartition, alors :

$$\mathbf{P}_X = \mathbf{P}_Y.$$

Autrement dit, si deux variables aléatoires réelles vérifient $\mathbf{P}(X \leq x) = \mathbf{P}(Y \leq x)$ pour tout $x \in \mathbf{R}$, alors

$$\mathbf{P}(X \in B) = \mathbf{P}(Y \in B) \quad \text{pour tout borélien } B.$$

L'énoncé du théorème est aussi vrai dans le cas vectoriel.

▮ PREUVE. [Cas réel] Dans le cas unidimensionnel, si $\mathbf{P}(X \leq x) = \mathbf{P}(Y \leq x)$ pour tout $x \in \mathbf{R}$, \mathbf{P}_X et \mathbf{P}_Y coïncident sur les ensembles du type $] -\infty, x]$ avec $x \in \mathbf{R}$. D'après la Proposition 1.3, elles sont égales partout. ▮

⚠ Attention. Le fait que X et Y aient la même loi ne dit rien sur $\mathbf{P}(X = Y)$ ou sur l'égalité presque-sûre de X et Y . En fait, X et Y peuvent avoir la même loi tout en étant définies sur des espaces probabilisés différents, auquel cas la quantité $\mathbf{P}(X = Y)$ n'a même aucun sens. En revanche, si X, Y sont définies sur un même espace probabilisé Ω alors

$$\{X = Y\} = \{\omega \in \Omega, \quad X(\omega) = Y(\omega)\},$$

et $X = Y$ p.s. (c'est-à-dire $\mathbf{P}(X = Y) = 1$) est beaucoup plus fort que $\mathbf{P}_X = \mathbf{P}_Y$.

2.1.3 Variables aléatoires discrètes

DEFINITION – 2.5 Soit $D \subset \mathbf{R}^d$ avec $d \geq 1$. L'ensemble D est dit *dénombrable* s'il existe un sous-ensemble N de \mathbf{N}^d (éventuellement fini) et une bijection $\Phi : D \rightarrow N$.

Autrement dit on peut énumérer les éléments de D , donc en particulier calculer des sommes.

⚠ Attention. Une autre convention de vocabulaire est obtenue avec $N = \mathbf{N}$, donc dénombrable dans ce cas ne contient pas les ensembles finis. Ici on adopte la première.

DEFINITION – 2.6 Une variable aléatoire X est dite *discrète* s'il existe $D \subset \mathbf{R}$ (ou $D \subset \mathbf{R}^d$, avec $d \in \mathbf{N}^*$ dans le cas vectoriel) tel que D soit dénombrable et tel que

$$\mathbf{P}(X \in D) = 1.$$

Pour que cette définition ait un sens, il faut que $\{X \in D\}$ soit un événement. C'est bien le cas puisqu'on peut l'écrire comme une réunion dénombrable d'événements :

$$\{X \in D\} = \bigcup_{x \in D} \{X = x\}.$$

D'après les axiomes des tribus, la dénombrabilité est donc suffisante (car pour tout x , $X^{-1}(x) = \{X = x\}$). Une variable aléatoire discrète est donc presque-sûrement à valeurs dans l'ensemble D . On commence avec deux lois discrètes classiques, une liste de telles lois sera proposée ensuite dans un tableau.

EXEMPLE – 2.1 Une variable aléatoire X suit la loi de Bernoulli de paramètre $p \in [0, 1]$ et on note $X \sim \mathcal{B}(p)$ si

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

La variable aléatoire est alors discrète car $\mathbf{P}(X \in \{0, 1\}) = \mathbf{P}(X = 0) + \mathbf{P}(X = 1) = 1$ et $\{0, 1\}$ est un ensemble fini.

EXEMPLE – 2.2 Une variable aléatoire $X = (X_1, \dots, X_d)$ à valeurs dans \mathbf{N}^d , suit une loi multinomiale de paramètres $n \in \mathbf{N}$, $p_1, \dots, p_d \in [0, 1]$, $p_1 + \dots + p_d = 1$, notée $\mathcal{M}(n, p_1, \dots, p_d)$ si

$$\mathbf{P}(X = (n_1, \dots, n_d)) = \frac{n!}{n_1! \dots n_d!} p_1^{n_1} \dots p_d^{n_d}$$

avec $n_1 + \dots + n_d = n$ et $n_1, \dots, n_d \in \mathbf{N}$.

Cette loi s'interprète ainsi : si on jette n boules aléatoirement une par une dans d boîtes différentes, chaque boule ayant une probabilité p_i d'être jetée dans la i -ème boîte, les nombres (X_1, \dots, X_d) de boules dans les boîtes $1, \dots, d$ suivent une loi $\mathcal{M}(n, p_1, \dots, p_d)$.

DEFINITION – 2.7 Soit X une variable discrète, on appelle fonction de masse ou fonction densité de X la fonction

$$f_X : x \in \mathbf{R}^d \mapsto \mathbf{P}(X = x)$$

avec $d = 1$ dans le cas unidimensionnel et $d \geq 2$ dans le cas vectoriel. On appelle support de X l'ensemble $\text{supp}(X) = \{x \in \mathbf{R}^d : f_X(x) \neq 0\}$.

REMARQUE – 2.3 Dans le cas d'une variable aléatoire discrète, on parlera plutôt de fonction de masse. Elle vérifie donc les propriétés suivantes :

- l'ensemble $\text{supp}(X) = \{x \in \mathbf{R}^d : f_X(x) \neq 0\}$ est fini ou dénombrable, c'est le support de X , c'est aussi le support de f_X en tant que support de fonction,
- on a $\sum_{x \in \text{supp}(X)} f_X(x) = 1$. Cette égalité est une réécriture de la notion de loi.

EXEMPLE – 2.3 La fonction de masse d'une variable de Bernoulli de paramètre p est ainsi donnée par

$$\forall x \in \mathbf{R}, \quad f(x) = p\mathbb{1}_1(x) + (1 - p)\mathbb{1}_0(x).$$

REMARQUE – 2.4 (LIEN ENTRE FONCTION DE MASSE ET FONCTION DE RÉPARTITION.) La fonction de répartition F_X d'une variable aléatoire discrète X est une fonction en escalier avec éventuellement une infinité de « marches ». La connaissance de F_X détermine f_X et réciproquement.

Par exemple, dans le cas unidimensionnel, les points $x \in \text{supp}(f_X)$ où f_X est non nulle sont les points où F_X fait un saut et $f_X(x)$ est la taille du saut que fait F_X en x . Tout ceci se résume en une équation :

$$\forall t \in \mathbf{R}, \quad F_X(t) = \sum_{x \in \text{supp}(X)} f_X(x) \mathbb{1}_{[x, +\infty[}(t).$$

EXERCICE 27

Justifier proprement cette dernière formule.

Le tableau suivant rassemble quelques lois discrètes usuelles.

Nom	Paramètre(s)	Notation	Support	$\mathbf{P}(X = k)$
Bernoulli	$p \in [0, 1]$	$\mathcal{B}(p)$	$\{0, 1\}$	$p\mathbb{1}_1(k) + (1 - p)\mathbb{1}_0(k)$
Binomiale	$(n, p) \in \mathbf{N}^* \times [0, 1]$	$\mathcal{B}(n, p)$	$\{0, \dots, n\}$	$\binom{n}{k} p^k (1 - p)^{n-k}$
Binomiale négative	$(n, p) \in \mathbf{N}^* \times [0, 1]$	$\mathcal{B}^-(n, p)$	\mathbf{N}	$\binom{n+k-1}{n-1} p^n (1 - p)^k$
Géométrique	$p \in [0, 1]$	$\mathcal{G}(p)$	\mathbf{N}^*	$p(1 - p)^{k-1}$
Hyper-Géométrique	$p \in [0, 1], N \in \mathbf{N}, n \in \{0, \dots, N\}, Np \in \mathbf{N}^*$	$\mathcal{H}(N, n, p)$	$\{a, \dots, b\}$ $a = (n - N(1 - p))^+$ $b = \min(n, Np)$	$\frac{\binom{Np}{k} \binom{N(1-p)}{n-k}}{\binom{N}{n}}$
Poisson	$\lambda \in]0, \infty[$	$\mathcal{P}(\lambda)$	\mathbf{N}	$e^{-\lambda} \frac{\lambda^k}{k!}$

Ces différentes lois modélisent toutes des expériences aléatoires concrètes. Mais nous avons besoin encore de la notion d'indépendance pour en parler.

REMARQUE – 2.5 (EXPRESSION *littérale* D'UNE LOI DISCRÈTE.) Soit X une variable discrète. Remarquons que

$$\mathbf{P}(X \notin \text{supp}(X)) = 1 - \mathbf{P}(X \in \text{supp}(X)) = 1 - 1 = 0.$$

Donc pour tout borélien B ,

$$\mathbf{P}_X(B) = \mathbf{P}(X^{-1}(B)) = \mathbf{P}(X^{-1}(B \cap \text{supp}(X))) = \mathbf{P}\left(\bigcup_{x \in B \cap \text{supp}(X)} X^{-1}(\{x\})\right) = \sum_{x \in B \cap \text{supp}(X)} \mathbf{P}(X = x),$$

par un argument d'union dénombrable disjointe. On a donc

$$\mathbf{P}_X(B) = \sum_{x \in \text{supp}(X)} \mathbf{P}(X = x) \delta_x(B),$$

ce qui veut dire que la loi de X s'écrit

$$\boxed{\mathbf{P}_X = \sum_{x \in \text{supp}(X)} \mathbf{P}(X = x) \delta_x} \iff \forall A \in E, \quad \mathbf{P}_X(A) = \sum_{x \in \text{supp}(X)} \mathbf{P}(X = x) \delta_x(A).$$

On aurait donc très bien pu définir l'ensemble des variables aléatoires discrètes comme l'ensemble des variables aléatoires de loi une probabilité de la forme précédente. Il n'est pas compliqué de constater que ce type de loi est à support dénombrable.

EXEMPLE – 2.4 Dans le cas de la loi binomiale de paramètres n et p , on a

$$\mathbf{P}_X = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \delta_k.$$

PROPOSITION – 2.4 Soit X une variable aléatoire à valeurs dans \mathbf{R} ou \mathbf{R}^d avec $d \geq 2$. Alors X est discrète si et seulement s'il existe un ensemble dénombrable I et deux familles $(p_i)_{i \in I} \in ([0, 1])^I$ et $(x_i)_{i \in I} \in (\mathbf{R}^d)^I$ telles que

$$\mathbf{P}_X = \sum_{i \in I} p_i \delta_{x_i}, \quad \text{avec} \quad \sum_{i \in I} p_i = 1.$$

C'est-à-dire si \mathbf{P}_X est *combinaison convexe dénombrable* de masses de Dirac. On continue avec une proposition qui fait le lien entre loi des marginales d'un vecteur aléatoire et la loi dudit vecteur.

PROPOSITION – 2.5 (LIEN ENTRE LA LOI D'UN VECTEUR ET CELLE DE SES MARGINALES, VERSION DISCRÈTE) On considère un entier $N \geq 2$. Soit (X_1, \dots, X_N) un vecteur aléatoire à valeurs dans \mathbf{R} , discret de supports $\text{supp}X_1, \dots, \text{supp}X_N$. Alors les variables aléatoires X_1, \dots, X_N sont des variables aléatoires réelles discrètes, de lois (ou fonction de masse) respectives f_{X_1}, \dots, f_{X_N} données par :

$$\forall x \in \mathbf{R}, i \in \{1, \dots, N\}, \quad f_{X_i}(x) = \mathbf{P}(X_i = x) \\ = \sum_{\substack{(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N) \\ \in \text{supp}X_1 \times \dots \times \text{supp}X_{i-1} \times \text{supp}X_{i+1} \times \dots \times \text{supp}X_N}} \mathbf{P}(X_1 = x_1, \dots, X_{i-1} = x_{i-1}, \dots, X_{i+1} = x_{i+1}, \dots, X_N = x_N).$$

EXERCICE 28

Écrire la proposition lorsque $N = 2$, la démontrer ensuite dans ce cas particulier.

2.1.4 Variables aléatoires continues

DEFINITION – 2.8 Soit $X : \Omega \rightarrow \mathbf{R}$ une variable aléatoire et F_X sa fonction de répartition. On dira que la variable X est *continue* s'il existe une fonction mesurable positive $f : \mathbf{R} \rightarrow [0, +\infty[$ telle que pour tout $x \in \mathbf{R}$,

$$F_X(x) = \int_{-\infty}^x f(t) dt.$$

Soit $X : \Omega \rightarrow \mathbf{R}^d$ une variable aléatoire, on dit que X est continue s'il existe une fonction mesurable positive $f : \mathbf{R}^d \rightarrow [0, +\infty[$ telle que pour tout $(x_1, \dots, x_d) \in \mathbf{R}^d$,

$$F_{X_1, \dots, X_d}(x_1, \dots, x_d) = \int_{]-\infty, x_1] \times \dots \times]-\infty, x_d]} f(t_1, \dots, t_d) dt_1 \dots dt_d.$$

La fonction f est appelée *densité* (ou *fonction de masse*) de la variable aléatoire X . Elle sera généralement notée f_X dans la suite.

REMARQUE – 2.6 (DÉFINITION ALTERNATIVE) Soit $X : \Omega \rightarrow \mathbf{R}$ une variable aléatoire. On dit que X est *continue* s'il existe une fonction mesurable positive $f : \mathbf{R} \rightarrow [0, +\infty[$ telle que pour tout borélien B de \mathbf{R} :

$$\mathbf{P}_X(B) = \int_B f(t) dt.$$

Même chose pour le cas vectoriel en considérant un borélien B de \mathbf{R}^d . Ou encore, d'après la Proposition 1.3, s'il existe une fonction mesurable positive $f : \mathbf{R} \rightarrow [0, +\infty[$ telle que pour tous réels $a < b$,

$$\mathbf{P}_X(]a, b]) = \mathbf{P}(a < X \leq b) = \int_a^b f(t) dt.$$

Dans le langage de la théorie de la mesure on dit que \mathbf{P}_X , la loi de X , est *absolument continue* par rapport à la mesure de Lebesgue. On écrit

$$\boxed{d\mathbf{P}_X = f \cdot dx},$$

ou parfois aussi $\mathbf{P}_X(dx) = f(x)dx$. Avec notre définition, les définitions alternatives en sont des conséquences. Elles sont résumées dans la proposition suivante :

PROPOSITION – 2.6 (CAS RÉEL) Soit $X : \Omega \rightarrow \mathbf{R}$ une variable continue de densité f_X . Alors

(i) $\int_{-\infty}^{\infty} f_X(t) dt = 1,$

(ii) $\mathbf{P}(X = x) = 0$ pour tout réel x , c'est-à-dire X n'a pas d'atome,

(iii) $\mathbf{P}(a \leq X \leq b) = \int_a^b f_X(t) dt$ pour tous réels $a \leq b$,

(iv) pour tout borélien $B \in \mathcal{B}(\mathbf{R})$,

$$\mathbf{P}_X(B) = \int_B f_X(t) dt.$$

▮ **PREUVE.** (i) découle de $\lim_{x \rightarrow \infty} F_X(x) = 1$, comme

$$\int_{-\infty}^{\infty} f_X(t) dt =_{def} \lim_{x \rightarrow \infty} \int_{-\infty}^x f_X(t) dt.$$

Remarquons que la fonction de répartition F_X d'une variable continue est continue, ce qui implique (ii) : pour tout $x \in \mathbf{R}$, $\mathbf{P}(X = x) = F(x) - F(x^-) = 0$. Pour (iii) on a alors

$$\mathbf{P}(a \leq X \leq b) = \mathbf{P}(a < X \leq b) = F(b) - F(a) = \int_a^b f_X(t) dt.$$

(iv) découle d'un argument classique : les ensembles du type $[a, b]$ ou $]a, b]$ engendrent en tant que tribu $\mathcal{B}(\mathbf{R})$ et est une classe stable par intersection finie : (iv) est donc une conséquence de (iii) d'après la proposition 1.3. ▮

Noter que dans (iii) avec (ii), on peut ouvrir ou fermer les bornes de (a, b) cela ne change rien. Dans le cas vectoriel, on a la proposition suivante :

PROPOSITION – 2.7 (CAS VECTORIEL) Soit $X : \Omega \rightarrow \mathbf{R}^d$ un vecteur aléatoire continu de densité f_X . Alors

(i) $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(t_1, \dots, t_d) dt_1 \dots dt_d = 1,$

(ii) $\mathbf{P}(X = (x_1, \dots, x_d)) = 0$ pour tout $(x_1, \dots, x_d) \in \mathbf{R}^d$,

(iii) $\mathbf{P}(X \in \prod_{i=1}^d [a_i, b_i]) = \int_{a_1}^{b_1} \cdots \int_{a_d}^{b_d} f_X(t_1, \dots, t_d) dt_1 \dots dt_d$ pour tous réels $a_i \leq b_i, i \in \{1, \dots, d\}$,

(iv) pour tout borélien $B \in \mathcal{B}(\mathbf{R}^d)$,

$$\mathbf{P}(X \in B) = \int_B f(t_1, \dots, t_d) dt_1 \dots dt_d.$$

⚠ **Attention.** Soit maintenant D un sous-ensemble fini ou dénombrable de \mathbf{R}^d . Alors, si X est continue, on a

$$\mathbf{P}(X \in D) = \sum_{x \in D} \mathbf{P}(X = x) = 0.$$

Cette remarque reflète bien une grande différence entre les deux types de variables : une variable aléatoire discrète n'a que des atomes, une variable aléatoire continue n'en a aucun.

Par définition, dans le cas continu la fonction de répartition s'obtient en primitivant une fonction L^2 que l'on appelle densité. Cette dernière étant seulement L^2 , rien ne dit que la fonction de répartition est dérivable : c'est le cas si la densité est continue. Le résultat suivant donne un élément de réponse.

PROPOSITION – 2.8 Soit X une variable aléatoire réelle. Si F_X est continue sur \mathbf{R} , dérivable sur un ensemble du type $\mathbf{R} \setminus I$ où I est un ensemble fini de points (on dit aussi *dérivable presque-partout pour la mesure de Lebesgue*), alors :

X est une variable à densité f_X , donnée par $f_X = F'_X$ pour tout $x \in \mathbf{R} \setminus I$.

Le tableau suivant rassemble quelques lois continues usuelles.

Nom	Paramètre(s)	Notation	Densité
Uniforme	$a < b$ deux réels	$\mathcal{U}_{[a,b]}$	$\frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$
Paréto	$p \in]1, +\infty[$	Paréto(p)	$\frac{p-1}{x^p} \mathbb{1}_{[1, +\infty[}(x)$
Exponentielle	$\lambda \in]0, +\infty[$	$\mathcal{E}(\lambda)$	$\lambda e^{-\lambda x} \mathbb{1}_{[0, +\infty[}(x)$
Gamma	$(p, \lambda) \in]0, +\infty[\times]0, +\infty[$	$\mathcal{G}(p, \lambda)$	$\frac{\lambda^p}{\Gamma(p)} e^{-\lambda x} x^{p-1} \mathbb{1}_{[0, +\infty[}(x)$
Béta	$(p, q) \in]0, +\infty[\times]0, +\infty[$	$\beta(p, q)$	$\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1} (1-x)^{q-1} \mathbb{1}_{]0,1[}(x)$
Laplace	$\lambda \in]0, \infty[$	$\mathcal{L}(\lambda)$	$\frac{\lambda}{2} e^{-\lambda x }$
Normale 1D	$(m, \sigma^2) \in \mathbf{R} \times]0, +\infty[$	$\mathcal{N}(m, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$
Normale dD	$(M, K) \in \mathbf{R}^d \times \mathcal{G}\mathcal{L}_{d,d}(\mathbf{R})$	$\mathcal{N}_d(M, K)$	$\frac{1}{(2\pi)^{d/2} \sqrt{\det K}} e^{-\frac{1}{2} \mathbf{T}(X-M) K^{-1} (X-M)}$
Cauchy	$c \in]0, \infty[$	$\mathcal{C}(c)$	$\frac{c}{\pi} \frac{x}{x^2 + c^2}$

Rappelons que la fonction Gamma est définie pour tout $x > 0$ par $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$.

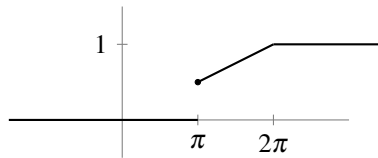
REMARQUE – 2.7 (LOI NORMALE : CAS PARTICULIER DE LA VARIANCE NULLE.) Noter que généralement on parle aussi de loi normale lorsque $\sigma = 0$. En effet si on regarde la densité Gaussienne on se rend compte que lorsque $\sigma \rightarrow 0$, on se rapproche d'une Dirac en m . On note donc parfois :

$$\mathcal{N}(m, 0) = \delta_m.$$

⚠ **Attention.** D'autre part, il existe des variables aléatoires qui ne sont ni discrètes, ni continues. Un exemple est donné maintenant.

EXERCICE 29

Soit Θ un variable aléatoire suivant une loi uniforme sur $[0, 2\pi]$. Posons $X = \max(\Theta, \pi)$. Montrer que la fonction de répartition de X a le graphe suivant :



En déduire que X n'est ni discrète ni continue.

► **CORRECTION.**

La variable X n'est donc pas continue puisque $\mathbf{P}(X = \pi) = \frac{1}{2}$. Elle n'est pas non plus discrète puisque sa fonction de répartition n'est pas en escalier. Plus précisément, on voit sur le graphe que $\mathbf{P}(X = x) = 0$ pour tout $x \neq \pi$. Par conséquent, si D est un ensemble dénombrable, alors $\mathbf{P}(X \in D)$ vaut 0 ou $\frac{1}{2}$ selon que π appartienne à D ou pas. Dans tous les cas $\mathbf{P}(X \in D) < 1$ et X ne peut pas être discrète.

EXERCICE 30

Soit X une variable exponentielle de paramètre $\lambda > 0$. Montrer la propriété d'absence de mémoire :

$$\mathbf{P}(X \geq s + t | X \geq s) = \mathbf{P}(X \geq t),$$

pour tout $s, t > 0$. Expliquer ce qu'elle représente sur un dessin.

PROPOSITION – 2.9 (LIEN ENTRE LA LOI D'UN VECTEUR ET CELLE DE SES MARGINALES, VERSION CONTINUE)

On considère un entier $N \geq 2$. Soit (X_1, \dots, X_N) un vecteur aléatoire à valeurs dans \mathbf{R} , continue, de densité $f_{(X_1, \dots, X_N)}$. Alors les variables aléatoires X_1, \dots, X_N sont des variables aléatoires réelles continues, de densités respectives f_{X_1}, \dots, f_{X_N} données par :

$$\forall x \in \mathbf{R}, i \in \{1, \dots, N\}, \quad f_{X_i}(x) = \int_{\mathbf{R}^{N-1}} f_{(X_1, \dots, X_N)}(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_N) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_N.$$

En particulier : soit (X, Y) un vecteur aléatoire à valeurs dans \mathbf{R}^2 , continue, de densité $f_{(X, Y)}$. Alors les variables aléatoires X et Y sont des variables aléatoires réelles continues, de densités respectives f_X et f_Y données par :

$$\forall x \in \mathbf{R}, \quad f_X(x) = \int_{\mathbf{R}} f_{(X, Y)}(x, y) dy, \quad \text{et} \quad f_Y(y) = \int_{\mathbf{R}} f_{(X, Y)}(x, y) dx.$$

EXERCICE 31

Démontrer la proposition pour un vecteur aléatoire (X, Y) à valeurs dans \mathbf{R}^2 .

§ 2.2. **Espérance**

2.2.1 Définition et transfert

Soit $X : \Omega \rightarrow \mathbf{R}^d$ une variable aléatoire et $h : \mathbf{R}^d \rightarrow \mathbf{R}$ une fonction mesurable. Vous trouverez la définition générale de mesurabilité dans l'appendice, ainsi que des rappels sur l'intégrale de Lebesgue.

Si $h(X)$ est positive ou intégrable pour la mesure \mathbf{P} , l'intégrale

$$\int_{\Omega} h(X(\omega)) d\mathbf{P}(\omega)$$

est bien définie. Elle vaut éventuellement $+\infty$ si $h(X)$ est positive \mathbf{P} -ps.

DEFINITION – 2.9 On appelle cette intégrale *espérance* de $h(X)$ et on la note

$$\mathbf{E}(h(X)) = \int_{\Omega} h(X(\omega)) d\mathbf{P}(\omega).$$

En particulier pour $h = \text{Id}$, on appelle *espérance de X* ou *moment d'ordre 1* la quantité

$$\mathbf{E}(X) = \int_{\Omega} X(\omega) d\mathbf{P}(\omega).$$

Une variable aléatoire X sera dite *centrée* si $\mathbf{E}(X) = 0$.

Si $X : \Omega \rightarrow \mathbf{R}$ est une variable unidimensionnelle alors $\mathbf{E}(X)$ est bien définie dès lors que X est positive ou intégrable ($\mathbf{E}(|X|) < \infty$). On rappelle que (cf. Appendice) $L^1(\Omega)$ désigne l'espace des variables ou vecteurs aléatoires $X = (X_1, \dots, X_d)$ intégrables contre la mesure \mathbf{P} . Dans le cas vectoriel, cela signifie que chaque coordonnée est intégrable, et l'intégrale est le vecteur formé par les intégrales des coordonnées :

$$\mathbf{E}(X) = (\mathbf{E}(X_1), \dots, \mathbf{E}(X_d)).$$

Donnons quelques exemples.

EXEMPLE – 2.5 (VARIABLE ALÉATOIRE CONSTANTE) Si $X = c$ avec $c \in \mathbf{R}$ alors $X = c\mathbb{1}_\Omega$, donc son espérance vaut

$$\mathbf{E}(X) = \int_{\Omega} c\mathbb{1}_\Omega(\omega) d\mathbf{P}(\omega) = c \int_{\Omega} \mathbb{1}_\Omega(\omega) d\mathbf{P}(\omega) = c\mathbf{P}(\Omega) = c.$$

(!) provient de la définition de l'intégrale de Lebesgue pour les fonctions étagées. Dans le même ordre d'idée, pour tout événement A :

$$\mathbf{E}(\mathbb{1}_A) = \mathbf{P}(A).$$

L'espérance est une intégrale, elle en vérifie donc les mêmes propriétés :

PROPOSITION – 2.10 Soient X, Y deux variables aléatoires réelles intégrables.

- (i) (Linéarité) $\mathbf{E}(\alpha X + \beta Y) = \alpha\mathbf{E}(X) + \beta\mathbf{E}(Y)$,
- (ii) (Monotonie) si $X \leq Y$ alors $\mathbf{E}(X) \leq \mathbf{E}(Y)$,
- (iii) si $X = c$ p.s. pour un $c \in \mathbf{R}$, alors $\mathbf{E}(X) = c$.

EXERCICE 32

On effectue n lancers d'une pièce équilibrée et on note $Y_i = 1$ si le i -ème lancer donne pile et 0 sinon. Déterminer l'espérance de Y_i pour tout i . En déduire l'espérance du nombre de piles.

►CORRECTION.
On a alors

$$\mathbf{E}(Y_i) = \mathbf{E}(\mathbb{1}_{\{1\}}(Y_i)) = \mathbf{P}(Y_i = 1) = \frac{1}{2}.$$

Le nombre de piles obtenu est $X = Y_1 + \dots + Y_n$. Son espérance est, par linéarité,

$$\mathbf{E}(X) = \mathbf{E}(Y_1) + \dots + \mathbf{E}(Y_n) = \frac{n}{2}.$$

EXEMPLE – 2.6 Si $Y = (Y_1, \dots, Y_d)$ est un vecteur aléatoire constitué de d variables de Bernoulli de paramètre $p \in [0, 1]$ alors

$$\mathbf{E}(Y) = (p, \dots, p).$$

L'espérance peut servir à contrôler les *quantiles d'une loi* ou *queues d'une loi* grâce à l'inégalité de Markov que l'on cite maintenant.

THÉORÈME – 2.2 (INÉGALITÉ DE MARKOV) Soit X une variable aléatoire réelle positive ou intégrable. Alors pour tout $a > 0$,

$$\mathbf{P}(X > a) \leq \frac{\mathbf{E}(X)}{a}.$$

En particulier, pour tout entier $k \in \mathbf{N}$, si X^k admet une espérance on a

$$\mathbf{P}(X > a) \leq \frac{\mathbf{E}(X^k)}{a^k}.$$

┌ PREUVE. Puisque $X \geq 0$, on a, presque sûrement :

$$X = X\mathbb{1}_{X>a} + X\mathbb{1}_{X\leq a} \geq X\mathbb{1}_{X>a} > a\mathbb{1}_{X>a}.$$

En prenant l'espérance on obtient bien

$$\mathbf{E}(X) \geq a\mathbf{P}(X > a).$$

On peut ensuite constater que

$$\mathbf{P}(X > a) = \mathbf{P}(X^k > a^k).$$

└

En pratique on ne connaît le plus souvent que la loi d'une variable aléatoire X , et pas la probabilité \mathbf{P} , ni-même l'espace Ω . Il est donc inutile d'espérer calculer des espérances en la gardant sous cette forme. La formule de transfert répond précisément au problème : l'espérance que l'on cherche à calculer se transporte (et est égale) à une intégrale contre \mathbf{P}_X . Si X est une variable aléatoire continue dans \mathbf{R} , on tombe sur une intégrale de Riemann habituelle puisque

$$d\mathbf{P}_X = f_X \cdot dx.$$

Si X est une variable aléatoire discrète, la loi s'exprime comme combinaison convexe de masses de Dirac : on tombe sur une série.

THÉORÈME – 2.3 (TRANSFERT) Soit $X : \Omega \rightarrow \mathbf{R}^d$ une variable aléatoire et $h : \mathbf{R}^d \rightarrow \mathbf{R}$ une fonction mesurable. Si $h(X)$ est positive ou intégrable pour la mesure \mathbf{P} , alors

$$\mathbf{E}(h(X)) = \int_{\mathbf{R}^d} h(t) d\mathbf{P}_X(t).$$

En particulier :

— si X est discrète, h est mesurable positive ou vérifie $\sum_{x \in \text{supp}(X)} \mathbf{P}(X = x) |h(x)| < \infty$, alors

$$\mathbf{E}(h(X)) = \sum_{x \in \text{supp}(X)} \mathbf{P}(X = x) h(x),$$

— si X possède une densité notée f , h est mesurable positive ou vérifie $\int_{\mathbf{R}^d} |h(t)| f(t) dt < \infty$, alors

$$\mathbf{E}(h(X)) = \int_{\mathbf{R}^d} h(t) f(t) dt.$$

EXERCICE 33

Calculer l'espérance de la variable aléatoire X définie dans l'Exercice 29.

► **CORRECTION.**

Remarquons que $\max(\Theta, \pi) = \pi \mathbb{1}_{\Theta < \pi} + \Theta \mathbb{1}_{\Theta \geq \pi}$. Par linéarité de l'espérance on a donc

$$\mathbf{E}(X) = \pi \mathbf{E}(\mathbb{1}_{\Theta < \pi}) + \mathbf{E}(\Theta \mathbb{1}_{\Theta \geq \pi}).$$

Comme Θ est une variable uniforme sur $[0, 2\pi]$, on a

$$\begin{aligned} \mathbf{E}(\mathbb{1}_{\Theta < \pi}) &= \frac{1}{2\pi} \int_0^{2\pi} \mathbb{1}_{x < \pi} dx = \frac{1}{2}, \\ \mathbf{E}(\Theta \mathbb{1}_{\Theta \geq \pi}) &= \frac{1}{2\pi} \int_0^{2\pi} x \mathbb{1}_{x \geq \pi} dx = \frac{3\pi}{4}. \end{aligned}$$

Donc

$$\mathbf{E}(X) = \frac{\pi}{2} + \frac{3\pi}{4} = \frac{5\pi}{4}.$$

▮ **PREUVE.** Comme souvent, on revient à la définition de l'intégrale pour montrer une telle propriété : on commence avec des fonctions étagées, puis on passe à la limite.

Prenons d'abord h de la forme $h = \mathbb{1}_B$ pour un certain borélien B . Alors

$$\mathbf{E}(\mathbb{1}_B(X)) = \mathbf{P}_X(B) = \int_{\mathbf{R}^d} \mathbb{1}_B(t) d\mathbf{P}_X(t),$$

donc l'égalité est vraie. Si h est une fonction étagée positive, alors h est combinaison linéaire d'indicatrices, et l'égalité reste vraie par linéarité de l'intégrale et de l'espérance.

Si h est mesurable positive, il existe une suite $(h_n)_{n \geq 0}$ de fonctions étagées positives telle que $h_n \nearrow h$. On a alors

$$\mathbf{E}(h_n(X)) = \int_{\mathbf{R}^d} h_n(t) d\mathbf{P}_X(t)$$

pour tout n . Par convergence monotone, on obtient l'égalité cherchée en passant à la limite. Enfin si h est une fonction quelconque, on obtient le résultat en écrivant $h = h_+ - h_-$ (partie positive moins partie négative).

Soit X une variable discrète. On a vu que la loi de X peut s'écrire sous la forme

$$\mathbf{P}_X = \sum_{x \in \text{supp}(X)} \mathbf{P}(X = x) \delta_x.$$

Donc

$$\mathbf{E}(h(X)) = \int_{\mathbf{R}^d} h(t) d\mathbf{P}_X(t) = \int_{\mathbf{R}^d} h(t) \sum_{x \in \text{supp}(X)} \mathbf{P}(X = x) d\delta_x(t) = \sum_{x \in \text{supp}(X)} \mathbf{P}(X = x) \int_{\mathbf{R}^d} h(t) d\delta_x(t).$$

Or $\int_{\mathbf{R}^d} h(t) d\delta_x(t) = h(x)$ d'où

$$\mathbf{E}(h(X)) = \sum_{x \in \text{supp}(X)} \mathbf{P}(X = x) h(x).$$

L'espérance $\mathbf{E}(h(X))$ est donc bien définie dès que h est mesurable et soit positive soit telle que

$$\sum_{x \in \text{supp}(X)} \mathbf{P}(X = x) |h(x)| < \infty.$$

Le cas continu est laissé au lecteur. ┘

EXERCICE 34

Soit X une variable suivant une loi exponentielle de paramètre $\lambda > 0$. Calculer son espérance.

► CORRECTION.

$$\mathbf{E}(X) = \int_{\mathbf{R}} x \lambda e^{-\lambda x} \mathbb{1}_{]0, +\infty[}(x) dx = \lambda \int_0^{\infty} x e^{-\lambda x} dx = \frac{1}{\lambda}.$$

EXERCICE 35 (UNE AUTRE EXPRESSION DE L'ESPÉRANCE DANS LE CAS DISCRET)

Soit X une variable aléatoire discrète à valeurs dans \mathbf{N} .

1) a. Pour tout $k \in \mathbf{N}^*$, exprimer $\mathbf{P}(X > k - 1)$ en fonction de $\mathbf{P}(X = k)$ et $\mathbf{P}(X > k)$.

$$\text{En déduire que : } \forall n \in \mathbf{N}^*, \sum_{k=1}^n k \mathbf{P}(X = k) = \sum_{k=0}^{n-1} \mathbf{P}(X > k) - n \mathbf{P}(X > n).$$

b. On suppose que la série $\sum_{k \geq 0} \mathbf{P}(X > k)$ converge. Montrer que X admet une espérance.

c. Réciproquement, on suppose que X admet une espérance. Montrer qu'alors la suite $(n \mathbf{P}(X > n))_{n \geq 1}$ converge vers 0, puis que la série $\sum_{k \geq 0} \mathbf{P}(X > k)$ converge. Que vaut sa somme ?

2) Application : on considère deux variables aléatoires indépendantes X et Y qui suivent chacune une loi géométrique de paramètres p_1 et p_2 respectivement. On note $Z = \max\{X, Y\}$.

a. Pour tout $k \in \mathbf{N}$, calculer $\mathbf{P}(X > k)$ puis $\mathbf{P}(Z > k)$.

b. Montrer que Z admet une espérance et la calculer.

REMARQUE – 2.8 (DÉFINITION ALTERNATIVE DE L'ESPÉRANCE) On peut donc affirmer que

$$X \text{ est une variable aléatoire intégrable à valeur dans } \mathbf{N} \iff \sum_{k \geq 0} \mathbf{P}(X > k) < \infty.$$

En plus l'espérance est donnée par :

$$\mathbf{E}(X) = \sum_{k \geq 0} \mathbf{P}(X > k) < \infty.$$

EXERCICE 36

Un mobile se déplace sur un axe de la façon suivante :

- à l'instant 0, il est au point d'abscisse 0 ;
- si, à l'instant n , le mobile est au point d'abscisse k , alors, à l'instant $n + 1$, soit il sera au point d'abscisse $k + 1$ avec une probabilité $p \in]0, 1[$, soit il retournera au point 0 avec une probabilité $1 - p$.

On note X_n la variable aléatoire égale à l'abscisse du mobile à l'instant n .

1) Déterminer la loi de X_1 .

- 2) Déterminer l'ensemble des valeurs prises par X_n .
- 3) Soient $n \in \mathbf{N}^*$ et $k \in \mathbf{N}$ tel que $1 \leq k \leq n$; trouver une relation entre $\mathbf{P}(X_n = k)$ et $\mathbf{P}(X_{n-1} = k - 1)$.
- 4) En déduire une relation de récurrence entre $\mathbf{E}(X_n)$ et $\mathbf{E}(X_{n-1})$ pour $n \geq 1$.
- 5) En déduire une expression de $\mathbf{E}(X_n)$ en fonction de n et p .

EXERCICE 37 (INÉGALITÉ DE JENSEN, VERSION \mathcal{C}^1)

Soit X une variable aléatoire et ϕ une fonction convexe de classe \mathcal{C}^1 . On suppose que X et $\phi(X)$ sont intégrables. On pose $m = \mathbf{E}(X)$. En utilisant le fait que ϕ est au-dessus de sa tangente en m , montrer que

$$\phi(\mathbf{E}(X)) \leq \mathbf{E}(\phi(X)).$$

REMARQUE – 2.9 L'hypothèse \mathcal{C}^1 n'est pas nécessaire : si ϕ est seulement continue convexe, il existe une minorante affine, et on peut refaire le raisonnement précédent.

Jusqu'à maintenant on a vu que la densité était la fonction apparaissant sous l'intégrale de la fonction de répartition d'une part, ou de manière équivalente sous l'intégrale de la loi. Voici une autre caractérisation, qui peut être plus simple à utiliser dans certains contextes.

PROPOSITION – 2.11 (DENSITÉ VIA DES FONCTIONS TESTS) Soit $X : \Omega \rightarrow \mathbf{R}^d$ une variable aléatoire. Supposons qu'il existe une fonction f mesurable telle que pour toute fonction h mesurable bornée, on ait

$$\mathbf{E}(h(X)) = \int_{\mathbf{R}^d} h(x) d\mathbf{P}_X(x) = \int_{\mathbf{R}^d} h(x) f(x) dx.$$

Alors X est une variable aléatoire continue et f est la densité de X .

▮ **PREUVE.** En particulier, pour tout borélien B de \mathbf{R}^d ,

$$\mathbf{P}(X \in B) = \mathbf{E}(\mathbf{1}_B(X)) = \int_B f(x) dx.$$

Le résultat est démontré. ▮

Donnons une application importante de ce principe. Remarquons d'abord qu'une fonction continue d'une variable aléatoire n'est pas forcément une variable aléatoire continue.

En effet, si ϕ est la fonction nulle et X une variable continue, alors $\phi(X)$ est la fonction constante égale à 0 qui n'est pas une variable continue ($\mathbf{P}(\phi(X) = 0) = 1$, elle possède un atome).

DEFINITION – 2.10 Soient U et V des ouverts de \mathbf{R}^d et $\phi : U \rightarrow V$ une fonction. On dit que ϕ est un \mathcal{C}^1 -difféomorphisme si ϕ est de classe \mathcal{C}^1 , bijective, et si sa réciproque est de classe \mathcal{C}^1 .

Pour $x \in U$, on notera $J_\phi(x)$ le jacobien de ϕ en x , c'est-à-dire le déterminant $\det D\phi(x)$ de la matrice jacobienne de ϕ en x notée $D\phi(x)$.

THÉORÈME – 2.4 Soit X une variable aléatoire continue à valeurs dans un ouvert U de \mathbf{R}^d et soit $\phi : U \rightarrow V$ un \mathcal{C}^1 -difféomorphisme. Alors $Y = \phi(X)$ est une variable aléatoire continue de densité

$$g(y) = f \circ \phi^{-1}(y) \left| J_{\phi^{-1}}(y) \right| \mathbf{1}_{\phi(U)}(y).$$

où f est une densité de X .

▮ **PREUVE.** Par hypothèse, la densité f de X est nulle en dehors de U . Soit h une fonction mesurable positive. En appliquant

la formule de changement de variable à ϕ^{-1} on obtient :

$$\begin{aligned} \mathbf{E}(h(Y)) &= \mathbf{E}(h(\phi(X))) = \int_U h \circ \phi(x) f(x) dx \\ &= \int_{\phi(U)} h(y) f \circ \phi^{-1}(y) \left| J_{\phi^{-1}}(y) \right| dy \\ &= \int_{\mathbf{R}^d} h(y) g(y) dy. \end{aligned}$$

En appliquant la remarque précédente, on en déduit que Y est une variable continue de densité g . ┘

REMARQUE – 2.10 La fonction h utilisée dans le raisonnement précédent est souvent appelée *fonction test*, on parle de *méthode de la fonction test*. On peut aussi remplacer *mesurable bornée* par *mesurable positive*, ou d'autres classes de fonctions.

EXERCICE 38 (TRANSLATION D'UNE V.A. CONTINUE)

Soit X une variable aléatoire réelle de densité f et soient a, b des réels, avec $a \neq 0$. Alors $Y = aX + b$ est une variable continue de densité

$$y \mapsto \frac{1}{|a|} f\left(\frac{y-b}{a}\right).$$

EXERCICE 39

On rappelle que le vecteur (X_1, X_2) suit une *loi normale de paramètres* $0_{\mathbf{R}^2}$ et $I_{\mathcal{M}_2(\mathbf{R})}$ si sa densité est, pour $(x, y) \in \mathbf{R}^2$,

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}.$$

Posons $Z = (X_1, X_1 + X_2)$. Déterminer la densité de Z par la méthode de la fonction test.

► **CORRECTION.**

Soit h une fonction mesurable positive de \mathbf{R}^2 dans \mathbf{R} , on a

$$\begin{aligned} \mathbf{E}(h(Z)) &= \mathbf{E}(h(X_1, X_1 + X_2)) \\ &= \int_{\mathbf{R}^2} h(x, x+y) \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dx dy \\ &= \int_{\mathbf{R}^2} h(\phi(x, y)) \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dx dy. \end{aligned}$$

L'application ϕ est définie sur \mathbf{R}^2 par $\phi(x, y) = (x, x+y)$. Elle est bijective de \mathbf{R}^2 dans lui-même et d'application réciproque définie sur \mathbf{R}^2 par $\phi^{-1}(u, v) = (u, v-u)$. Sa matrice jacobienne est donc

$$\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}.$$

Le jacobien vaut donc 1. Ainsi

$$\mathbf{E}(h(Z)) = \int_{\mathbf{R}^2} h(u, v) \frac{1}{2\pi} e^{-\frac{u^2+(v-u)^2}{2}} du dv.$$

La densité de Z est donc donnée sur \mathbf{R}^2 par

$$f_Z(u, v) = \frac{1}{2\pi} e^{-\frac{u^2+v^2-2uv}{2}}.$$

2.2.2 Moments supérieurs

DEFINITION – 2.11 (MOMENTS) Soit X une variable aléatoire réelle. La quantité $\mathbf{E}(X^k)$, si elle existe, est appelée *moment d'ordre k* de X . La quantité $\mathbf{E}(|X|^k)$, si elle existe, est appelée *moment absolu d'ordre k* de X .

Le moment d'ordre k de X est défini si $X \geq 0$ presque sûrement (dans ce cas il peut prendre la valeur $+\infty$) ou si X^k est intégrable (dans ce cas il est forcément fini). Pour rappel, X est de moment absolu d'ordre k fini signifie exactement que X est dans $L^k(\Omega)$.

Rappelons que pour une mesure de probabilité, les espaces L^p sont emboîtés : si $p \leq q$ alors $L^q(\Omega) \subset L^p(\Omega)$.

DEFINITION – 2.12 (VARIANCE ET MATRICE DE COVARIANCE) Soit $X : \Omega \rightarrow \mathbf{R}$ une variable aléatoire de carré intégrable. Sa variance est définie par

$$\mathbf{V}(X) = \mathbf{E}((X - \mathbf{E}(X))^2) = \mathbf{E}(X^2) - \mathbf{E}(X)^2.$$

Une variable aléatoire réelle sera dite *réduite* si $\mathbf{V}(X) = 1$.

Dans le cas vectoriel, si $X = (X_1, \dots, X_d)$, à la variance se substitue la *matrice de variance-covariance* ou *matrice de dispersion* définie par

$$\mathbf{Cov}(X) = (\mathbf{E}((X_i - \mathbf{E}(X_i))(X_j - \mathbf{E}(X_j))))_{1 \leq i, j \leq d} = (\mathbf{E}(X_i X_j) - \mathbf{E}(X_i)\mathbf{E}(X_j))_{1 \leq i, j \leq d}.$$

Pour $1 \leq i, j \leq d$, la quantité

$$\mathbf{E}(X_i X_j) - \mathbf{E}(X_i)\mathbf{E}(X_j)$$

s'appelle la *covariance* entre X_i et X_j et se note $\mathbf{Cov}(X_i, X_j)$.

EXERCICE 40

Soit X le nombre de pile obtenus au bout de n lancers à pile ou face. Rappelons que pour $k \in \{0, \dots, n\}$,

$$\mathbf{P}(X = k) = \binom{n}{k} \frac{1}{2^n}.$$

Calculer $\mathbf{E}(X(X - 1))$.

►CORRECTION.
D'après la formule de transfert

$$\begin{aligned} \mathbf{E}(X(X - 1)) &= \sum_{k=0}^n k(k - 1)\mathbf{P}(X = k) \\ &= \sum_{k=2}^n k(k - 1) \binom{n}{k} \frac{1}{2^n} \\ &= n(n - 1) \frac{1}{2^n} \sum_{k=2}^n \binom{n-2}{k-2} \\ &= \frac{n(n-1)}{4}. \end{aligned}$$

On en déduit

$$\mathbf{V}(X) = \mathbf{E}(X(X - 1)) + \mathbf{E}(X) - \mathbf{E}(X)^2 = \frac{n(n-1)}{4} + \frac{n}{2} - \frac{n^2}{4} = \frac{n}{4}.$$

EXERCICE 41

Soit Y une variable aléatoire telle que $\mathbf{P}(Y = 0) = \mathbf{P}(Y = 1) = \frac{1}{2}$. On pose $Z = (Y, -Y)$. Déterminer l'espérance et la covariance du vecteur Z .

►CORRECTION.

$$\mathbf{E}(Z) = \left(\frac{1}{2}, -\frac{1}{2} \right),$$

et $\mathbf{V}(Y) = \mathbf{V}(-Y) = \frac{1}{4}$. De plus $\mathbf{E}(Y(-Y)) - \mathbf{E}(Y)\mathbf{E}(-Y) = -\mathbf{E}(Y^2) + \mathbf{E}(Y)^2 = -\mathbf{V}(Y) = -\frac{1}{4}$, de sorte que

$$\mathbf{Cov}(Z) = \frac{1}{4} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

PROPOSITION – 2.12 Soient X, Y deux variables aléatoires réelles de carré intégrable, et soient λ et c deux réels. On a

- (i) $\mathbf{V}(c) = 0$,
- (ii) $\mathbf{V}(\lambda X + c) = \lambda^2 \mathbf{V}(X)$,
- (iii) $\mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y) + 2\mathbf{Cov}(X, Y)$.

EXERCICE 42

Démontrer la propriété.

Nous verrons diverses propriétés de la matrice de dispersion d'un vecteur aléatoire dans le chapitre sur les vecteurs Gaussiens.

La variance mesure la déviation de X à sa moyenne. Plus la variance est grande, plus X a de chances d'être loin de sa moyenne. Plus précisément, on a le résultat suivant.

THÉORÈME – 2.5 (INÉGALITÉ DE BIENAYMÉ-TCHEBICHEV) Soit $X : \Omega \rightarrow \mathbf{R}$ une variable dans L^2 et $a > 0$. Alors

$$\mathbf{P}(|X - \mathbf{E}(X)| \geq a) \leq \frac{\mathbf{V}(X)}{a^2}.$$

▮ PREUVE. L'inégalité de Markov appliquée à $Y = |X - \mathbf{E}(X)|$ donne

$$\mathbf{P}(Y \geq a) = \mathbf{P}(Y^2 \geq a^2) \leq \frac{\mathbf{E}(Y^2)}{a^2} = \frac{\mathbf{V}(X)}{a^2}.$$

EXEMPLE – 2.7 Notons X_n le nombre de pile obtenu au bout de n lancers. On s'intéresse à la fréquence des piles $\frac{X_n}{n}$.

On a $\mathbf{E}(X_n/n) = \mathbf{E}(X_n)/n = \frac{1}{2}$ et $\mathbf{V}(X_n/n) = \mathbf{V}(X_n)/n^2 = 1/(4n)$. Étant donné $\varepsilon > 0$, par l'inégalité de Bienaymé-Tchebychev on a

$$\mathbf{P}\left(\left|\frac{X_n}{n} - \frac{1}{2}\right| > \varepsilon\right) \leq \frac{1}{4n\varepsilon^2}.$$

En particulier, la probabilité que X_n/n dévie de sa moyenne d'au moins ε tend vers 0 quand n tend vers l'infini. On dit que X_n/n tend vers $1/2$ en probabilité. Ce type de convergence sera systématisé dans le dernier chapitre.

§ 2.3. Loi et espérance conditionnelle sachant un évènement

On considère toujours un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$. Soit $B \in \mathcal{A}$ un évènement non-négligeable, c'est-à-dire $\mathbf{P}(B) > 0$. On a déjà défini une nouvelle probabilité de conditionnement \mathbf{P}_B sur le même espace $(\Omega, \mathcal{A}, \mathbf{P})$ par

$$\forall A \in \mathcal{A}, \quad \mathbf{P}_B(A) = \mathbf{P}(A|B).$$

On est alors amené à la définition suivante

DEFINITION – 2.13 (LOI ET ESPÉRANCE CONDITIONNELLE SACHANT UN ÉVÈNEMENT) Soit X une variable aléatoire admettant une espérance (donc positive ou dans $L^1(\Omega, \mathcal{A}, \mathbf{P})$).

On appelle *loi conditionnelle de X sachant B* la mesure de probabilité définie par

$$\forall A \in \mathcal{A}, \quad \mathbf{P}_{X|B}(A) = \mathbf{P}_B(X^{-1}(A)).$$

On appelle *espérance conditionnelle de X sachant B* la quantité

$$\mathbf{E}(X|B) = \int_{\Omega} X(\omega) d\mathbf{P}_B(\omega) = \frac{\mathbf{E}(X\mathbb{1}_B)}{\mathbf{P}(B)}.$$

Elle s'interprète comme la valeur moyenne de X lorsque B est réalisé.

REMARQUE – 2.11 D'autres notions plus générales d'espérances conditionnelles existent. On peut généraliser facilement à une variable aléatoire Y et définir $\mathbf{E}(X|Y)$ comme

$$\mathbf{E}(X|Y) = \varphi(Y),$$

où φ est la fonction définie par :

$$\varphi(y) = \begin{cases} \mathbf{E}(X|Y=y) & \text{si } \mathbf{P}(Y=y) > 0 \\ 0 & \text{ailleurs} \end{cases}.$$

L'espérance conditionnelle est alors une variable aléatoire. Cette définition s'applique donc bien aux conditionnements par une variable aléatoire discrète.

On généralise ensuite encore avec un conditionnement sachant une tribu \mathcal{G} : cette notion la plus générale fait intervenir la propriété de projection dans le Hilbert $L^2(\Omega, \mathcal{A}, \mathbf{P})$ sur le sous-espace vectoriel des fonctions L^2 \mathcal{G} -mesurables. Cette

définition générale permet d'aboutir à la définition de *martingale*.

Notez aussi que toutes les théorèmes classiques d'intégration se généralisent avec des espérances conditionnelles : par exemple l'inégalité de Jensen, la convergence dominée, l'inégalité de Cauchy-Schwarz et le théorème de Fubini existent tous dans une version conditionnelle.

Nous n'aurons pas besoin de ces objets dans la suite du cours.

§ 2.4. Fonctions caractéristique et génératrice

Nous avons déjà vu deux méthodes pour caractériser une loi :

- un calcul direct,
- une comparaison des fonctions de répartition,
- une comparaison des fonctions de densité (éventuellement avec la méthode la fonction test) ou de masse dans le cas discret.

Dans cette sous section, on rajoute deux autres outils : les *fonctions génératrices* (ou *transformée de Fourier*) et *fonctions caractéristiques* (ou *transformée de Laplace*). En plus de caractériser les lois, elles permettent d'avoir accès très facilement à beaucoup d'autres informations sur la loi (espérance, variance, moments supérieurs).

Nous désignons par $\langle \cdot, \cdot \rangle$ le produit scalaire Euclidien dans \mathbf{R}^d .

DEFINITION – 2.14 (FONCTION CARACTÉRISTIQUE) Soit $X : \Omega \rightarrow \mathbf{R}^d$ un vecteur aléatoire. On appelle *fonction caractéristique de X*, et on note $\phi_X : \mathbf{R}^d \rightarrow \mathbf{C}$ la fonction définie par :

$$\phi_X : t \in \mathbf{R}^d \mapsto \mathbf{E}(e^{i\langle t, X \rangle}) = \mathbf{E}(\cos(\langle t, X \rangle)) + i\mathbf{E}(\sin(\langle t, X \rangle)).$$

En particulier, pour $d = 1$ on obtient

$$\phi_X : t \in \mathbf{R} \mapsto \mathbf{E}(e^{itX}).$$

Notez que l'espérance est bien définie puisque pour tout $t \in \mathbf{R}^d$

$$|e^{i\langle t, X \rangle}| = 1,$$

et que la fonction constante égale à 1 est intégrable.

REMARQUE – 2.12 (LIEN AVEC LA TRANSFORMÉE DE FOURIER D'UNE FONCTION) Soit $f \in L^1(\mathbf{R})$. Alors pour tout $x \in \mathbf{R}$, on appelle transformée de Fourier de f notée $\mathcal{F}(f) : \mathbf{R} \rightarrow \mathbf{R}$ la fonction définie par

$$\mathcal{F}(f)(t) = \int_{-\infty}^{\infty} f(x)e^{-2i\pi tx} dx.$$

Remarquer que sous l'hypothèse $L^1(\mathbf{R})$, l'intégrale précédente est parfaitement définie. Alors, si X est une variable aléatoire continue et réelle, de densité $f \in L^1(\mathbf{R})$, on a

$$\phi_X(t) = \mathcal{F}(f)\left(-\frac{t}{2\pi}\right) = \mathcal{F}\left(\frac{d\mathbf{P}_X}{dx}\right)\left(-\frac{t}{2\pi}\right).$$

On peut remarquer que \mathcal{F} est un opérateur linéaire :

$$\mathcal{F} : L^1(\mathbf{R}) \longrightarrow L^\infty(\mathbf{R}).$$

Sa restriction à $L^2(\mathbf{R})$ est encore notée \mathcal{F} dans la suite :

$$\mathcal{F} : L^2(\mathbf{R}) \longrightarrow L^2(\mathbf{R}).$$

Si $g \in L^2(\mathbf{R})$, alors on définit $\mathcal{F}^{-1}(g)$ par

$$\mathcal{F}^{-1}(g)(t) = \int_{-\infty}^{\infty} g(x)e^{2i\pi tx} dx.$$

Ce nouvel opérateur $g \mapsto \mathcal{F}^{-1}(g)$ est en fait comme la notation l'indique l'opérateur inverse de \mathcal{F} , il vérifie en particulier

$$\mathcal{F} \circ \mathcal{F}^{-1} = \text{Id}_{L^2(\mathbf{R}), L^2(\mathbf{R})}.$$

EXEMPLE – 2.8 Soit X une variable aléatoire de loi $\mathcal{E}(1)$. Alors, par le théorème de transfert, pour tout $t \in \mathbf{R}$,

$$\phi_X(t) = \mathbf{E}(e^{itX}) = \int_0^\infty e^{itx} e^{-x} dx = \frac{1}{1-it}.$$

EXEMPLE – 2.9 Soit X une variable aléatoire de loi $\mathcal{B}(n, p)$. Alors, pour tout $t \in \mathbf{R}$,

$$\phi_X(t) = \sum_{k=0}^n e^{itk} \mathbf{P}(X = k) = \sum_{k=0}^n e^{itk} \binom{n}{k} p^k (1-p)^{n-k} = (1-p + pe^{it})^n.$$

Comme son nom l'indique, la fonction caractéristique caractérise la loi. On énonce ce fait dans le théorème suivant :

THÉORÈME – 2.6 Si X et Y sont deux vecteurs aléatoires de lois \mathbf{P}_X et \mathbf{P}_Y telles que $\phi_X = \phi_Y$, alors $\mathbf{P}_X = \mathbf{P}_Y$.

▮ **PREUVE.** [Version Fourier inverse dans le cas continu et unidimensionnel] On a déjà vu que pour tout $t \in \mathbf{R}$,

$$\phi_X(t) = \mathcal{F}\left(\frac{d\mathbf{P}_X}{dx}\right)\left(-\frac{t}{2\pi}\right) = \mathcal{F}(f)\left(-\frac{t}{2\pi}\right),$$

en notant f la densité de X . Donc $\mathcal{F}(f)(t) = \phi_X(-2\pi t)$. Or $\psi_X : t \mapsto \phi_X(-2\pi t)$ est une fonction dans $L^2(\mathbf{R})$, donc d'après la formule de transformation (vu en Remarque 2.12) inverse on a $f = \mathcal{F}^{-1}(\psi_X)$. Si $\phi_X = \phi_Y$, on a aussi $\psi_X = \psi_Y$ et donc en notant f et g les densités respectives de X et Y , on a $f = g = \mathcal{F}^{-1}(\psi_X) = \mathcal{F}^{-1}(\psi_Y)$. Donc $\mathbf{P}_X = \mathbf{P}_Y$. ▮

Nous citons rapidement le théorème des classes monotones fonctionnelles utilisé dans la preuve générale.

THÉORÈME – 2.7 (CLASSE MONOTONE FONCTIONNELLE) Soit \mathcal{C} un ensemble de fonctions réelles bornées sur Ω , stable par multiplication, et contenant les constantes. Soit E un espace vectoriel de fonctions contenant les constantes, et stable par convergence monotone bornée (on dit que E est *monotone*). Alors :

E contient les fonctions mesurables par rapport à $\sigma(\mathcal{C})$.

L'idée de ce théorème est la suivante :

on a deux ensembles \mathcal{C} et E de fonctions contenant les constantes, dont l'un (l'ensemble E) est en plus un espace vectoriel. Alors sous l'hypothèse de monotonie, l'espace vectoriel E contient en fait plus : les fonctions mesurables par rapport à $\sigma(\mathcal{C})$.

▮ **PREUVE.** [Version classe monotone fonctionnelle] La démonstration utilise le théorème de classes monotones fonctionnel. On note e_1, \dots, e_d la base canonique de \mathbf{R}^d . Pour tout $t \in \mathbf{R}^d$, l'égalité des parties réelles (resp. imaginaires) de ϕ_X et ϕ_Y donne $\mathbf{E}(\cos(\langle t, X \rangle)) = \mathbf{E}(\cos(\langle t, Y \rangle))$ (resp. $\mathbf{E}(\sin(\langle t, X \rangle)) = \mathbf{E}(\sin(\langle t, Y \rangle))$). Notons \mathcal{C} l'ensemble des combinaisons linéaires finies des fonctions $x \mapsto \cos(\langle t, x \rangle)$ et $x \mapsto \sin(\langle t, x \rangle)$. En particulier, la fonction $x \mapsto n \sin(\langle e_i/n, x \rangle)$ appartient à \mathcal{C} et sa limite simple, la projection sur la i -ème coordonnée, est mesurable par rapport à la tribu $\sigma(\mathcal{C})$ engendrée par \mathcal{C} . Donc $\sigma(\mathcal{C}) = \mathcal{B}(\mathbf{R}^d)$.

Soit maintenant \mathcal{H} l'espace vectoriel des fonctions boréliennes bornées ϕ telles que $\mathbf{E}(\phi(X)) = \mathbf{E}(\phi(Y))$. L'espace \mathcal{H} contient les constantes et est stable par convergence monotone bornée. De plus $\mathcal{C} \subset \mathcal{H}$ et \mathcal{C} est stable par multiplication (linéariser un produit de sinus et de cosinus). Le théorème des classes monotones fonctionnelles montre alors que \mathcal{H} contient toute fonction bornée mesurable par rapport à $\sigma(\mathcal{C}) = \mathcal{B}(\mathbf{R}^d)$, donc toute fonction borélienne. Le résultat s'en suit. ▮

PROPOSITION – 2.13 Soit X une variable aléatoire réelle de fonction caractéristique ϕ_X et $t \in \mathbf{R}$. Alors :

- (i) $\phi_X(-t) = \overline{\phi_X(t)}$, $\phi_X(0) = 1$,
- (ii) si $\mathbf{E}(|X|^n) < \infty$, alors ϕ_X est n -fois dérivable, de dérivée k -ième ($k \leq n$),

$$\phi_X^{(k)}(t) = i^k \mathbf{E}(X^k e^{itX}).$$

En particulier, $\phi_X^{(k)}(0) = i^k \mathbf{E}(X^k)$,

- (iii) si X admet une espérance, alors ϕ_X est uniformément continue.

▮ PREUVE. Nous démontrons une partie du point (ii), le reste de la preuve est laissée au lecteur. Démontrons donc que ϕ_X est dérivable en tout point t de \mathbf{R} lorsque $\mathbf{E}(|X|) < \infty$. Pour tout $h \neq 0$,

$$\frac{\phi_X(t+h) - \phi_X(t)}{h} = \int_{\mathbf{R}} e^{itx} \frac{e^{ihx} - 1}{h} d\mathbf{P}_X(x).$$

Or,

$$\left| e^{itx} \frac{e^{ihx} - 1}{h} \right| \leq |x|$$

qui est intégrable pour \mathbf{P}_X , indépendamment de h . D'après le théorème de convergence dominée,

$$\phi_X'(t) = \lim_{h \rightarrow 0} \int_{\mathbf{R}} e^{itx} \frac{e^{ihx} - 1}{h} d\mathbf{P}_X(x) = \int_{\mathbf{R}} ix e^{itx} d\mathbf{P}_X(x) = i\mathbf{E}(X e^{itX}).$$

On a redémontré ici dans ce cas particulier le théorème de dérivation sous l'intégrale. Mais vous pouvez bien sûr l'utiliser directement et vérifier les différentes hypothèses. ▮

EXERCICE 43

Rappeler la définition d'uniforme continuité, et démontrer le point (iii)

Le second objet caractérisant la loi est ce qu'on appelle la fonction génératrice (des moments). La fonction caractéristique est définie partout, cependant la fonction génératrice ne l'est pas forcément. Cela n'est pas problématique pour la propriété de caractérisation d'une loi.

Notez également que la fonction génératrice des moments, mise à part la propriété d'analyticité qui sera vue après, ne présente aucun avantage par rapport à la fonction caractéristique. Cependant elle existe dans la littérature et il faut en connaître les principales propriétés.

DEFINITION – 2.15 Si $X : \Omega \rightarrow \mathbf{R}^d$ est un vecteur aléatoire, on appelle *fonction génératrice des moments* ou *transformée de Laplace* la fonction

$$M_X : t \in \mathbf{R}^d \mapsto \mathbf{E}(e^{\langle t, X \rangle}),$$

définie pour les valeurs de t où $e^{\langle t, X \rangle}$ est intégrable. En particulier, si $d = 1$

$$M_X : t \in \mathbf{R} \mapsto \mathbf{E}(e^{tX}).$$

La fonction génératrice des moments, si elle est définie dans un voisinage de 0, caractérise la loi comme la fonction caractéristique.

REMARQUE – 2.13 (FONCTION GÉNÉRATRICE) Parfois dans la littérature on rencontre également un dernier type de fonction. Si X est une variable aléatoire réelle, on appelle *fonction génératrice de X* la fonction définie par $G_X(s) = \mathbf{E}(s^X)$ pour les s tels que s^X soit intégrable.

Supposons que X est une variable aléatoire discrète, alors

$$G_X(s) = \sum_{x \in \text{supp} X} \mathbf{P}(X = x) s^x.$$

On reconnaît là à nouveau une série entière en s , qui par majoration est au moins définie sur $] -1, 1[$.

EXERCICE 44 (BINOMIALE NÉGATIVE)

Soient $m \in \mathbf{N}$, $m \geq 2$ et $p \in]0, 1]$. La pièce est lancée une infinité de fois.

Soit X le nombre de lancers nécessaires jusqu'à ce qu'on obtienne m fois pile.

Soit A l'évènement : « le nombre de piles reste toujours $< m$ ».

1) Quelles sont les valeurs prises par X ? Avec quelle probabilité? On dit que X suit une $\mathcal{B}^-(n, p)$.

2) Montrer que $\mathbf{P}(A) = 0$. On établira au préalable le développement en série entière

$$|x| < 1, \quad (1+x)^{-m} = \sum_{k=m}^{\infty} \binom{k-1}{m-1} (-x)^{k-m}.$$

3) Déterminer la fonction génératrice de X notée G_X et en déduire $\mathbf{E}(X)$ et $\mathbf{V}(X)$.

4) Montrer que si $p \neq 1$, $\mathbf{E}\left(\frac{m-1}{X-1}\right) = p$ et $\mathbf{E}\left(\frac{m}{X}\right) > p$.

PROPOSITION – 2.14 Soit $X : \Omega \rightarrow \mathbf{R}$ une variable aléatoire réelle telle que e^{tX} est intégrable pour t dans un intervalle ouvert contenant 0.

Alors la fonction génératrice des moments est définie sur un intervalle ouvert contenant 0 et c'est une série entière en 0 :

$$M_X(t) = \sum_{n \in \mathbf{N}} \frac{t^n}{n!} \mathbf{E}(X^n)$$

pour tout t dans ce voisinage.

En particulier, pour tout $n \in \mathbf{N}$,

$$M_X^{(n)}(0) = \mathbf{E}(X^n).$$

▮ PREUVE. Supposons M_X définie sur $] -\varepsilon, \varepsilon[$, pour un $\varepsilon > 0$. Puisque

$$\sum_{n \in \mathbf{N}} \frac{|tx|^n}{n!} = e^{|tx|} \leq e^{tx} + e^{-tx},$$

le théorème de convergence dominée montre que pour tout $|t| < \varepsilon$,

$$M_X(t) = \mathbf{E}(e^{tX}) = \sum_{n \in \mathbf{N}} \frac{1}{n!} \mathbf{E}((tX)^n) = \sum_{n \in \mathbf{N}} \frac{t^n}{n!} \mathbf{E}(X^n),$$

ce qui démontre que M_X est une série entière en 0. La propriété de dérivation des séries entières permet d'obtenir la formule des moments. ▮

Le tableau suivant rassemble l'espérance, la variance et la fonction caractéristique de certaines lois usuelles.

Nom	Paramètre(s)	Espérance	Variance	$\phi_X(t)$
Uniforme	$a < b$ deux réels	$\frac{a+b}{2}$	$\frac{(a-b)^2}{12}$	$\frac{e^{itb} - e^{ita}}{it(b-a)}$
Exponentielle	$\lambda \in]0, +\infty[$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{1}{1 - \frac{it}{\lambda}}$
Gamma	$(p, \lambda) \in]0, +\infty[\times]0, +\infty[$	$\frac{p}{\lambda}$	$\frac{p}{\lambda^2}$	$\left(\frac{1}{1 - \frac{it}{\lambda}}\right)^p$
Laplace		0	2	$\frac{1}{1+t^2}$
Normale	$(m, \sigma^2) \in \mathbf{R} \times]0, +\infty[$	m	σ^2	$e^{itm - \frac{\sigma^2 t^2}{2}}$
Bernoulli	$p \in [0, 1]$	p	$p(1-p)$	$1 - p + pe^{it}$
Binomiale	$(n, p) \in \mathbf{N}^* \times [0, 1]$	np	$np(1-p)$	$(1 - p + pe^{it})^n$
Géométrique	$p \in [0, 1]$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^{it}}{1 - (1-p)e^{it}}$
Poisson	$\lambda \in]0, \infty[$	λ	λ	$e^{\lambda(e^{it}-1)}$

Pour achever ce chapitre, voici une autre notion fondamentale en probabilités : l'indépendance de variables aléatoires. Nous avons déjà rencontré cette notion pour des familles d'évènements, les deux sont liées et nous le verrons.

EXERCICE 45

Calculer la fonction caractéristique d'une uniforme, d'une exponentielle et d'une loi normale.

§ 2.5. Indépendance et corrélation de variables aléatoires

On a déjà regardé la notion d'indépendance pour les familles d'évènements. On s'intéresse ici à la notion d'indépendance pour les variables aléatoires.

2.5.1 Définition et caractérisation de l'indépendance

DEFINITION – 2.16 Une famille quelconque de variables aléatoires $(X_i)_{i \in I}$ sur Ω , et à valeurs dans \mathbf{R}^d est dite *mutuellement indépendante* (ou *indépendante*) si pour tout sous-ensemble $J \subset I$ fini et tous les $B_j \in \mathcal{B}(\mathbf{R}^d)$ avec $j \in J$,

$$\mathbf{P}\left(\bigcap_{j \in J} \{X_j \in B_j\}\right) = \prod_{j \in J} \mathbf{P}(X_j \in B_j).$$

Si on a seulement

$$\mathbf{P}(X_i \in A, X_j \in B) = \mathbf{P}(X_i \in A)\mathbf{P}(X_j \in B)$$

pour tous les $i \neq j$ et $A, B \in \mathcal{B}(\mathbf{R}^d)$, on dit que les variables $X_i, i \in I$ sont *deux à deux indépendantes*.

Des variables aléatoires indépendantes suivant toutes la même loi sont dite *i.i.d.* (pour « indépendantes identiquement distribuées »).

REMARQUE – 2.14 (LIEN AVEC L'INDÉPENDANCE D'ÉVÈNEMENTS) La famille $(X_i)_{i \in I}$ est donc composée de variables aléatoires indépendantes (ou mutuellement indépendantes) (resp. indépendantes deux à deux) si et seulement si la famille $(\mathbb{1}_{\{X_j \in B_j\}})_{j \in J}$ est indépendante (resp. indépendante deux à deux) en tant qu'évènements pour tout ensemble $J \subset I$ fini et famille B_j de boréliens.

EXERCICE 46

Soient A et B deux évènements d'un espace $(\Omega, \mathcal{A}, \mathbf{P})$. Montrer que A et B sont indépendants si et seulement si $\mathbb{1}_A$ et $\mathbb{1}_B$ le sont.

Comme d'habitude, on peut réécrire facilement la définition d'une autre manière. Dans le cas d'une famille finie de variables aléatoires, on peut réécrire la notion d'indépendance en terme de loi du vecteur aléatoire formé par la famille.

PROPOSITION – 2.15 (VERSION LOI/FONCTION DE RÉPARTITION) Soit $N \geq 2$ un entier. Un vecteur aléatoire $X = (X_1, \dots, X_N)$ est à *marginales indépendantes* (ou dit autrement les variables aléatoires X_i pour $1 \leq i \leq N$ sont *mutuellement indépendantes*) si et seulement si

$$\mathbf{P}_X = \mathbf{P}_{X_1} \otimes \dots \otimes \mathbf{P}_{X_N},$$

ou encore si et seulement si pour tous $(x_1, \dots, x_N) \in \mathbf{R}^N$,

$$\mathbf{P}(X_1 \leq x_1, \dots, X_N \leq x_N) = \prod_{i=1}^N \mathbf{P}(X_i \leq x_i).$$

La loi $\mathbf{P}_{X_1} \otimes \dots \otimes \mathbf{P}_{X_N}$ est appelée *loi produit* et est définie par l'égalité suivante :

$$(\mathbf{P}_{X_1} \otimes \dots \otimes \mathbf{P}_{X_N})(B_1 \times \dots \times B_N) = \prod_{i=1}^N \mathbf{P}_{X_i}(B_i),$$

pour toute famille B_1, \dots, B_N de boréliens.

▮ PREUVE. La première partie de la proposition est donc simplement une réécriture de la définition. Pour la deuxième partie, on constate simplement qu'une implication est triviale, et que l'autre découle de la Proposition 1.3. ▮

PROPOSITION – 2.16 (VIA LES FONCTIONS-TESTS) Un vecteur aléatoire $X = (X_1, \dots, X_N)$ est à marginales indépendantes (ou dit autrement les variables aléatoires X_i pour $1 \leq i \leq N$ sont mutuellement indépendantes) si et seulement si pour toutes fonctions boréliennes bornées $h_1, \dots, h_N : \mathbf{R} \rightarrow \mathbf{R}$

$$\mathbf{E}\left(\prod_{j=1}^N h_j(X_j)\right) = \prod_{j=1}^N \mathbf{E}(h_j(X_j)).$$

▮ PREUVE. Démonstration classique et déjà faite. On commence par les fonctions étagées, puis on écrit toute fonction mesurable positive comme limite simple de fonctions étagées. Inversement, on choisit pour les h_j des indicatrices de boréliens et on retrouve la définition d'indépendance. ▮

PROPOSITION – 2.17 (CAS DES VARIABLES DISCRÈTES) Soient X et Y des variables discrètes, X et Y sont indépendantes si et seulement si pour tous $x, y \in \mathbf{R}$,

$$\mathbf{P}(X = x, Y = y) = \mathbf{P}(X = x)\mathbf{P}(Y = y).$$

EXERCICE 47

Généraliser la proposition précédente à un vecteur aléatoire à marginales continues quelconques (X_1, \dots, X_N) avec $N \geq 2$ un entier.

⚠ **Attention.** Ceci n'est vrai que pour les variables discrètes. Par exemple, si X et Y sont continues, on a toujours

$$\mathbf{P}(X = x, Y = y) = 0 = \mathbf{P}(X = x)\mathbf{P}(Y = y),$$

bien que toutes les variables aléatoires continues ne soient pas indépendantes au sens précédent.

▮ **PREUVE.** On suppose que l'égalité a lieu. Soient A et B des boréliens de \mathbf{R} , notons $A' = A \cap \text{supp}X$ et $B' = B \cap \text{supp}Y$. On a

$$\begin{aligned} \mathbf{P}(X \in A, Y \in B) &= \mathbf{P}(X \in A', Y \in B') = \sum_{i \in A', j \in B'} \mathbf{P}(X = i, Y = j) \\ &= \sum_{i \in A', j \in B'} \mathbf{P}(X = i)\mathbf{P}(Y = j) \\ &= \sum_{i \in A'} \mathbf{P}(X = i) \sum_{j \in B'} \mathbf{P}(Y = j) \\ &= \mathbf{P}(X \in A')\mathbf{P}(Y \in B') = \mathbf{P}(X \in A)\mathbf{P}(Y \in B), \end{aligned}$$

ce qui montre que X et Y sont indépendantes. La réciproque est évidente. ▮

EXERCICE 48

On lance deux dés, on appelle X et Y les résultats respectifs du premier et du deuxième dé. Montrer qu'ils sont indépendants.

► **CORRECTION.**
On a

$$\mathbf{P}(X = 2, Y = 3) = \frac{1}{36} = \mathbf{P}(X = 2)\mathbf{P}(Y = 3).$$

Donc $\{X = 2\}$ est indépendant de $\{Y = 3\}$ et de même pour les autres valeurs de X et Y . Ainsi X est indépendant de Y .

PROPOSITION – 2.18 (CAS DES VARIABLES CONTINUES) Si le vecteur aléatoire (X, Y) possède une densité $f : \mathbf{R}^2 \rightarrow \mathbf{R}$ qui s'écrit sous forme produit :

$$\forall (x, y) \in \mathbf{R}^2, \quad f(x, y) = g(x)h(y),$$

où les fonctions g et h sont boréliennes positives, alors X et Y sont indépendantes de densités respectives

$$\frac{g}{\int_{\mathbf{R}} g(x) dx} \quad \text{et} \quad \frac{h}{\int_{\mathbf{R}} h(y) dy}.$$

Réciproquement, si les variables réelles X et Y sont indépendantes de densités respectives f_X et f_Y , alors le vecteur aléatoire (X, Y) admet pour densité

$$f_{(X, Y)} = f_X f_Y.$$

▮ **PREUVE.** La densité du couple (X, Y) étant sous forme produit, on a

$$\int_{\mathbf{R}} g(x) dx \int_{\mathbf{R}} h(y) dy = \int_{\mathbf{R}^2} g(x)h(y) dx dy = \int_{\mathbf{R}^2} f(x, y) dx dy = 1.$$

On a déjà vu d'après la Proposition 2.9 que la densité de X est

$$x \mapsto g(x) \times \int_{\mathbf{R}} h(y) dy = \frac{g(x)}{\int_{\mathbf{R}} g(t) dt},$$

celle de Y

$$y \mapsto \int_{\mathbf{R}} g(x) dx \times h(y) = \frac{h(y)}{\int_{\mathbf{R}} h(t) dt}.$$

Par suite, pour toutes fonctions ϕ et ψ telles que $\phi(X)$ et $\psi(Y)$ sont intégrables, on a

$$\begin{aligned} \mathbf{E}(\phi(X)\psi(Y)) &= \int_{\mathbf{R}^2} \phi(x)\psi(y)f(x,y) dx dy \\ &= \int_{\mathbf{R}^2} \phi(x)\psi(y)f_X(x)f_Y(y) dx dy \\ &= \int_{\mathbf{R}} \phi(x)f_X(x) dx \int_{\mathbf{R}} \psi(y)f_Y(y) dy \\ &= \mathbf{E}(\phi(X))\mathbf{E}(\psi(Y)). \end{aligned}$$

D'où l'indépendance de X et Y d'après la caractérisation par les fonctions tests.

Réciproquement, soit Q la mesure sur \mathbf{R}^2 de densité $f_X f_Y$. On veut montrer que pour tout borélien A de \mathbf{R}^2 , on a

$$P_{(X,Y)}(A) = Q(A).$$

Soit A_1 et A_2 deux boréliens de \mathbf{R} , par indépendance, on a,

$$\begin{aligned} \mathbf{P}((X,Y) \in A_1 \times A_2) &= \mathbf{P}(X \in A_1, Y \in A_2) \\ &= \mathbf{P}(X \in A_1)\mathbf{P}(Y \in A_2) \\ &= \int_{A_1} f_X(x) dx \int_{A_2} f_Y(y) dy \\ &= \int_{A_1 \times A_2} f_X(x)f_Y(y) dx dy \\ &= Q(A_1 \times A_2). \end{aligned}$$

Les mesures $P_{(X,Y)}$ et Q coïncident donc sur la classe des pavés, qui forment une classe monotone qui engendre les boréliens de \mathbf{R}^2 . On en déduit d'après la Proposition 1.3 que $P_{(X,Y)} = Q$ sur $\mathcal{B}(\mathbf{R}^2)$ tout entier. \square

Dans la dernière étape de la preuve, on aurait pu raisonner aussi avec les fonctions de répartition.

EXERCICE 49

Généraliser la proposition précédente à un vecteur aléatoire à marginales continues quelconques (X_1, \dots, X_N) avec $N \geq 2$ un entier.

EXERCICE 50

Si X et Y sont deux variables aléatoires indépendantes de lois respectives $\mathcal{U}_{[a,b]}$ et $\mathcal{U}_{[c,d]}$, alors montrer que le vecteur (X, Y) est un vecteur aléatoire de densité

$$\forall (x,y) \in \mathbf{R}^2, \quad f_{(X,Y)}(x,y) = \frac{1}{(b-a)(d-c)} \mathbb{1}_{[a,b] \times [c,d]}(x,y).$$

PROPOSITION – 2.19 (INDÉPENDANCE ET FONCTION CARACTÉRISTIQUE) Un vecteur aléatoire $X = (X_1, \dots, X_N)$ est à marginales indépendantes (ou dit autrement les variables aléatoires X_i pour $1 \leq i \leq N$ sont mutuellement indépendantes) si et seulement si

$$\phi_{X_1 + \dots + X_N} = \phi_{X_1} \dots \phi_{X_N}.$$

EXERCICE 51

Démontrer cette caractérisation dans le cas $N = 2$.

EXEMPLE – 2.10 Soit X et Y des variables aléatoires indépendantes suivant des lois de Poisson de paramètres λ et μ respectivement. On a, pour $t \in \mathbf{R}$:

$$\begin{aligned} \phi_{X+Y}(t) &= \phi_X(t)\phi_Y(t) \\ &= e^{\lambda(e^{it}-1)} e^{\mu(e^{it}-1)} \\ &= e^{(\lambda+\mu)(e^{it}-1)}. \end{aligned}$$

On reconnaît la fonction caractéristique d'une loi de Poisson de paramètre $\lambda + \mu$. Comme la fonction caractéristique caractérise la loi, la variable $X + Y$ suit la loi de Poisson de paramètre $\lambda + \mu$.

2.5.2 Covariance et corrélation de deux variables aléatoires réelles

Rappelons la définition de la covariance.

DEFINITION – 2.17 On appelle *covariance* de deux variables de carré intégrable la quantité

$$\mathbf{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y) = \mathbf{E}((X - \mathbf{E}(X))(Y - \mathbf{E}(Y))).$$

On dit que deux variables de carré intégrable sont *non corrélées* si $\mathbf{Cov}(X, Y) = 0$. Dans le cas contraire, elles sont dites *corrélées*.

Comme $2|XY| \leq X^2 + Y^2$, la covariance de variables aléatoires réelles X et Y est bien définie dès que X et Y sont de carré intégrable.

La covariance apparaît naturellement dans la formule de variance d'une somme que l'on constatera plus tard :

$$\mathbf{V}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbf{V}(X_i) + 2 \sum_{1 \leq i < j \leq n} \mathbf{Cov}(X_i, X_j)$$

pour toute famille X_1, \dots, X_n de variables de carré intégrable.

PROPOSITION – 2.20 Si X, Y sont indépendantes et de carré intégrable alors $\mathbf{Cov}(X, Y) = 0$.

▮ PREUVE. En effet, dans ce cas, on a $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$. ▮

Attention, la réciproque n'est pas vraie, on peut avoir $\mathbf{Cov}(X, Y) = 0$ sans que X et Y soient indépendantes.

PROPOSITION – 2.21 (PROPRIÉTÉ DE LA COVARIANCE) $\mathbf{Cov} : L^2(\Omega) \times L^2(\Omega) \rightarrow \mathbf{R}$ est un produit scalaire sur $L^2(\Omega)$: elle est bilinéaire, symétrique, définie positive.

Par ailleurs, soit X une variable aléatoire réelle de carré intégrable. Alors

- (i) $\mathbf{Cov}(X, X) = \mathbf{V}(X)$,
- (ii) $\mathbf{Cov}(X, c) = 0$ pour toute constante $c \in \mathbf{R}$.

EXERCICE 52

Faire la preuve de ces deux résultats.

PROPOSITION – 2.22 (VARIANCE D'UNE SOMME) Soit (X_1, \dots, X_n) une famille de variables aléatoires. Alors :

$$\mathbf{V}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbf{V}(X_i) + 2 \sum_{1 \leq i < j \leq n} \mathbf{Cov}(X_i, X_j).$$

En particulier, si X_1, \dots, X_n est une suite de variables indépendantes on a

$$\mathbf{V}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbf{V}(X_i).$$

EXERCICE 53

Démontrer la dernière proposition.

EXERCICE 54 (MOYENNE & VARIANCE EMPIRIQUES)

Soit $n \in \mathbf{N}^*$, $\mu \in \mathbf{R}$, et $\sigma > 0$. Soient aussi X_1, \dots, X_n des variables aléatoires réelles admettant une espérance commune μ et une variance commune σ^2 .

1) Quelle est l'espérance de \overline{X}_n où $\overline{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$?

2) On suppose dans cette question les v.a. i.i.d.

a. Soit $l \in \{1, 2, \dots, n\}$. Que vaut $\mathbf{E}(X_l^2)$? Calculer $\mathbf{E}(\overline{X}_n \cdot X_l)$ et $\mathbf{E}(\overline{X}_n^2)$, puis $\mathbf{V}(\overline{X}_n)$ et $\mathbf{Cov}(\overline{X}_n, X_l - \overline{X}_n)$.

b. Calculer l'espérance de $S_n^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \overline{X}_n)^2$.

3) On suppose dans cette question que les v.a. sont i.i.d. On note $Y_i = X_i - \mu$ et on suppose que toutes les variables aléatoires admettent un moment d'ordre 4 noté μ_4 . Vérifier que

$$nS_n^2 = \left(\sum_{j=1}^n Y_j^2 \right) - n\overline{Y}_n^2.$$

Exprimer en fonction de n , σ^2 et μ_4 l'espérance des variables aléatoires suivantes :

$$Y = \left(\sum_{j=1}^n Y_j^2 \right)^2, \quad V = \overline{Y}_n^2 \sum_{j=1}^n Y_j^2, \quad W = \overline{Y}_n^4.$$

En déduire $\mathbf{E}(S_n^4)$ puis $\mathbf{V}(S_n^2)$.

Rappelons pour finir les deux tableaux de lois cités plus haut.

Nom	Paramètre(s)	Notation	Support	$\mathbf{P}(X = k)$
Bernoulli	$p \in [0, 1]$	$\mathcal{B}(p)$	$\{0, 1\}$	$p\mathbb{1}_1(k) + (1-p)\mathbb{1}_0(k)$
Binomiale	$(n, p) \in \mathbf{N}^* \times [0, 1]$	$\mathcal{B}(n, p)$	$\{0, \dots, n\}$	$\binom{n}{k} p^k (1-p)^{n-k}$
Binomiale négative	$(n, p) \in \mathbf{N}^* \times [0, 1]$	$\mathcal{B}^-(n, p)$	\mathbf{N}	$\binom{n+k-1}{n-1} p^n (1-p)^k$
Géométrique	$p \in [0, 1]$	$\mathcal{G}(p)$	\mathbf{N}^*	$p(1-p)^{k-1}$
Hyper - Géométrique	$p \in [0, 1], N \in \mathbf{N}, n \in \{0, \dots, N\}, Np \in \mathbf{N}^*$	$\mathcal{H}(N, n, p)$	$\{a, \dots, b\}$ $a = (n - N(1-p))^+$ $b = \min(n, Np)$	$\frac{\binom{Np}{k} \binom{N(1-p)}{n-k}}{\binom{N}{n}}$
Poisson	$\lambda \in]0, \infty[$	$\mathcal{P}(\lambda)$	\mathbf{N}	$e^{-\lambda} \frac{\lambda^k}{k!}$

Nom	Paramètre(s)	Notation	Densité
Uniforme	$a < b$ deux réels	$\mathcal{U}_{[a,b]}$	$\frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$
Paréto	$p \in]1, +\infty[$	Paréto(p)	$\frac{p-1}{x^p} \mathbb{1}_{[1, +\infty[}(x)$
Exponentielle	$\lambda \in]0, +\infty[$	$\mathcal{E}(\lambda)$	$\lambda e^{-\lambda x} \mathbb{1}_{[0, +\infty[}(x)$
Gamma	$(p, \lambda) \in]0, +\infty[\times]0, +\infty[$	$\gamma(p, \lambda)$	$\frac{\lambda^p}{\Gamma(p)} e^{-\lambda x} x^{p-1} \mathbb{1}_{[0, +\infty[}(x)$
Béta	$(p, q) \in]0, +\infty[\times]0, +\infty[$	$\beta(p, q)$	$\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1} (1-x)^{q-1} \mathbb{1}_{]0,1[}(x)$
Laplace	$\lambda \in]0, \infty[$	$\mathcal{L}(\lambda)$	$\frac{\lambda}{2} e^{-\lambda x }$
Normale	$(m, \sigma^2) \in \mathbf{R} \times]0, +\infty[$	$\mathcal{N}(m, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$
Cauchy	$c \in]0, \infty[$	$\mathcal{C}(c)$	$\frac{c}{\pi} \frac{x}{x^2 + c^2}$

REMARQUE – 2.15 (MODÉLISATION ET LOIS USUELLES) Donnons quelques interprétation.

- BERNOULLI $\mathcal{B}(p)$ avec $p \in [0, 1]$ — expérience à deux issues dont l'une est de probabilité p (exemple : lancer d'une pièce truquée ou non).
- BINOMIALE $\mathcal{B}(n, p)$ avec $p \in [0, 1]$ et $n \in \mathbf{N}^*$ — c'est une somme de n Bernouillis indépendantes de paramètre p . La simulation d'une Binomiale correspond donc au nombre de succès (si 1 représente un succès) dans n jeux de type Bernouilli.
- BINOMIALE NÉGATIVE $\mathcal{B}^-(n, p)$ avec $p \in [0, 1]$ et $n \in \mathbf{N}^*$ — on est toujours dans le cadre de tirages indépendants selon un schéma de Bernouilli, mais la Binomiale négative modélise le nombre d'échecs obtenus avant l'obtention de n succès.

§ 2.6. Exercices

EXERCICE 55 (SOMME DE LOIS USUELLES)

- 1) a. Soit X_1 une v.a. de loi binomiale de paramètres n et p , $n \geq 2$ entier et $p \in [0, 1]$. Calculer la fonction génératrice G_1 de X_1 . En déduire l'espérance et la variance de X_1 .
- b. Soient X_1 et Y_1 deux v.a. indépendantes de lois binomiales respectives $\mathcal{B}(n, p)$ et $\mathcal{B}(m, p)$. Donner la loi de probabilité de $X_1 + Y_1$.
- c. Soient X_1, \dots, X_N , N variables aléatoires indépendantes de loi $\mathcal{B}(p)$. Déterminer la loi de $X_1 + \dots + X_N$. Retrouver alors l'espérance et la variance d'une Binomiale $\mathcal{B}(n, p)$.
- 2) a. Soit X_2 une v.a. de loi de Poisson de paramètre $\lambda > 0$. Calculer la fonction génératrice G_2 de X_2 . En déduire l'espérance et la variance de X_2 .
- b. Soient X_2 et Y_2 deux v.a. indépendantes de lois de Poisson respectives $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$. Donner la loi de probabilité de $X_2 + Y_2$.

►CORRECTION.

Soient $t \in \mathbf{R}$ et $X \sim \mathcal{B}(n, p)$, alors d'après la formule du binôme on a

$$\phi_X(t) = \sum_{k=0}^n e^{itk} \binom{n}{k} p^k (1-p)^{n-k} = (1-p + pe^{it})^n.$$

Soit $|s| < 1$. Alors on peut calculer aussi la fonction génératrice

$$G_X(s) = \mathbf{E}(s^X) = \sum_{k=0}^n s^k \binom{n}{k} p^k (1-p)^{n-k} = (ps + 1 - p)^n.$$

En dérivant et en évaluant en 1 on trouve $\mathbf{E}(X_1) = np$, et $\mathbf{V}(X) = np(1-p)$. Soient X_1 et Y_1 deux v.a. indépendantes de lois binomiales respectives $\mathcal{B}(n, p)$ et $\mathcal{B}(m, p)$. Alors par caractérisation de l'indépendance avec les fonctions tests on a

$$G_{X_1+X_2}(s) = G_{X_1}(s)G_{X_2}(s) = (ps + 1 - p)^n (ps + 1 - p)^m = (ps + 1 - p)^{n+m}.$$

La dernière fonction est celle d'une génératrice associée à une $\mathcal{B}(n+m, p)$, or la fonction génératrice caractérise la loi puisqu'elle est définie sur un voisinage ouvert de 0 ici, donc $X_1 + X_2 \sim \mathcal{B}(n+m, p)$.

Pour $X_1 + \dots + X_N$ c'est pareil :

$$G_{X_1+X_2+\dots+X_N}(s) = \prod_{k=1}^N (ps + 1 - p) = (ps + 1 - p)^N.$$

La dernière fonction est une génératrice de loi $\mathcal{B}(N, p)$.

EXERCICE 56 (SOMME DE LOIS UNIFORMES)

Soient X et Y deux variables aléatoires indépendantes de même loi uniforme sur $\{1, \dots, n\}$.

- 1) Calculer $\mathbf{P}(X = Y)$.
- 2) Trouver la loi de $X + Y$.
- 3) On pose $U = \min\{X, Y\}$ et $V = \max\{X, Y\}$. Trouver les lois de (U, V) , U et V . U et V sont-elles indépendantes ?

►CORRECTION.

1) Par indépendance, on a $\mathbf{P}(X = Y) = \sum_{k=1}^n \mathbf{P}(X = k)\mathbf{P}(Y = k) = \frac{1}{n}$.

2) Clairement $X + Y$ est à valeurs dans $\{2, \dots, n\}$. Donc soit $2 \leq k \leq n$, on a encore par indépendance

$$\mathbf{P}(X + Y = k) = \sum_{l=1}^n \mathbf{P}(X = k-l)\mathbf{P}(Y = l) = \frac{1}{n} \sum_{l=1}^n \mathbf{P}(X = k-l) = \frac{1}{n} \sum_{l=1}^n \mathbf{P}(X = k-l) = \frac{1}{n^2} \sum_{l=k-n}^{k-1} \mathbb{1}_{\{1, \dots, n\}}(l).$$

La toute dernière somme se réécrit

$$\mathbf{P}(X + Y = k) = \begin{cases} \frac{1}{n} & \text{si } k \geq n, \\ \frac{k}{n^2} & \text{sinon.} \end{cases}$$

3) U et V prennent ses valeurs dans $\{1, \dots, n\}$. Soit $(k, l) \in \{1, \dots, n\}^2$, alors

$$\mathbf{P}(U = k, V = l) = \mathbf{P}(X = k, Y = l) + \mathbf{P}(X = l, Y = k) = \mathbf{P}(X = k)\mathbf{P}(Y = l) + \mathbf{P}(X = l)\mathbf{P}(Y = k) = \frac{2}{n^2}.$$

En terme de fonction de masse, on a

$$f_{(U,V)}(k, l) = \frac{2}{n^2} \mathbb{1}_{\{1, \dots, n\}^2}(k, l).$$

On en déduit en sommant, les lois des marginales : si $k \in \{1, \dots, n\}$,

$$\mathbf{P}(U = k) = \sum_{l=1}^n \mathbf{P}(U = k, V = l) = \frac{2}{n},$$

de-même pour V . La fonction de masse du vecteur n'est pas le produit des deux fonctions de masses, les variables aléatoires ne sont donc pas indépendantes.

EXERCICE 57

Un homme rentre chez lui le soir et dispose d'un trousseau de k clés ($k \geq 2$). Lorsqu'il n'a pas bu, il essaie une clé au hasard puis, si le résultat est infructueux, il la met de côté et essaie une clé au hasard parmi celles restantes. À chaque insuccès, il répète ce procédé. Lorsque cet homme est ivre, il essaie une clé au hasard puis, à chaque insuccès il remet la clé avec les autres et choisit à nouveau une clé au hasard. On définit les événements suivants : $A = \ll \text{l'homme n'a pas bu} \gg$ et $B = \ll \text{l'homme est ivre} \gg$.

- 1) Dans toute cette question on suppose que l'homme est ivre. On note X_B la variable aléatoire représentant le nombre d'essais nécessaires à l'homme pour parvenir à ouvrir sa porte quand il est ivre. Quelle est la loi de X_B ?
- 2) Dans toute cette question on suppose que l'homme n'a pas bu. On note X_A la variable aléatoire représentant le nombre d'essais nécessaires à l'homme pour parvenir à ouvrir sa porte quand il n'a pas bu. Calculer $\mathbf{P}(X_A = 1)$, $\mathbf{P}(X_A = 2)$ et $\mathbf{P}(X_A = 3)$. Quelle est la loi de X_A ?
- 3) On suppose maintenant que $\mathbf{P}(A) = \mathbf{P}(B) = \frac{1}{2}$. Calculer la probabilité p_n que l'homme soit ivre, sachant qu'il est parvenu à ouvrir sa porte au terme d'exactly n essais ($n \in \mathbf{N}^*$). ♣ INDICATION – On pensera à distinguer les cas où $n \leq k$ et $n > k$.

►CORRECTION.

- 1) On est exactement dans le cadre d'une expérience aléatoire de loi géométrique de paramètre $p = \frac{1}{k}$ (il y a remise si l'homme est ivre). Autrement dit X_B est à support $\{1, \dots, \infty\}$ et pour tout $k \geq 1$,

$$\mathbf{P}(X_B = k) = p(1-p)^{k-1}.$$

- 2) $\mathbf{P}(X_A = 1) = \frac{1}{k} = p$, $\mathbf{P}(X_A = 2) = \frac{k-1}{k} \frac{1}{k-1} = p$, et $\mathbf{P}(X_A = 3) = \frac{k-1}{k} \frac{k-2}{k-1} \frac{1}{k-2} = \frac{1}{k} = p$. Pour les autres valeurs, on peut recommencer, et on obtient une $\mathcal{U}\{1, \dots, k\}$.

- 3) L'homme est maintenant soit ivre, soit sobre de manière équiprobable. Soit X la variable aléatoire donnant le nombre d'essais nécessaires pour ouvrir la porte. Alors X est définie en deux morceaux :

$$X(\omega) = \begin{cases} X_A(\omega) & \text{si } \omega \in A, \\ X_B(\omega) & \text{si } \omega \in B. \end{cases}$$

$$\text{On souhaite calculer } p_n = \mathbf{P}(B | X = n) = \frac{\mathbf{P}(B \cap \{X = n\})}{\mathbf{P}(X = n)} = \frac{\mathbf{P}(\{X_B = n\} | B)\mathbf{P}(B)}{\mathbf{P}(X = n)} = \frac{\mathbf{P}(\{X_B = n\} | B)\mathbf{P}(B)}{\mathbf{P}(\{X_A = n\} | A)\mathbf{P}(A) + \mathbf{P}(\{X_B = n\} | B)\mathbf{P}(B)}.$$

Si $n > k$, alors on obtient

$$p_n = \frac{\mathbf{P}(\{X_B = n\} | B)\mathbf{P}(B)}{\mathbf{P}(\{X_B = n\} | B)\mathbf{P}(B)} = 1,$$

ce qui est attendu : si on a essayé plus de fois que le nombre de clés, c'est qu'on est ivre.

Maintenant passons au second : si $k \leq n$, on obtient

$$p_n = \frac{p(1-p)^{n-1}}{1/k + p(1-p)^{n-1}}.$$

EXERCICE 58 (SIMULATION 1)

Il est facile de générer des variables aléatoires à partir de la loi uniforme sur l'intervalle $[0, 1]$. Soit U une variable aléatoire de loi uniforme sur $[0, 1]$.

- 1) Si $X = -\ln(U)/\lambda$ avec $\lambda > 0$, montrer que X suit la loi exponentielle $\mathcal{E}(\lambda)$.
- 2) Soit $Y = \lfloor 1 + nU \rfloor$ avec $n \in \mathbf{N}^*$ et où $\lfloor \cdot \rfloor$ désigne la partie entière. Donner la loi de Y .

►CORRECTION.

- 1) Soit $x \in \mathbf{R}$, alors

$$\mathbf{P}(X \leq x) = \mathbf{P}(U \geq \exp(-\lambda x)) = \int_{\exp(-\lambda x)}^1 \mathbb{1}_{[0,1]}(x) dx = \int_{-\lambda x}^0 \mathbb{1}_{[0,1]}(e^y) e^y dy = \int_{-\infty}^x \mathbb{1}_{[0,1]}(e^{-\lambda t}) e^{-\lambda t} \lambda dt = \int_{-\infty}^x \mathbb{1}_{[0,+\infty[}(t) e^{-\lambda t} \lambda dt.$$

On retrouve une densité exponentielle.

- 2) Y est clairement à valeurs dans $\{1, \dots, n+1\}$. Soit $k \in \{1, \dots, n+1\}$, alors par définition de la partie entière :

$$\mathbf{P}(Y = k) = \mathbf{P}(k-1 \leq nU < k) = \mathbf{P}((k-1)/n \leq U < k/n) = \int_{(k-1)/n}^{k/n} \mathbb{1}_{[0,1]}(t) dt = \frac{1}{n}.$$

Ainsi Y suit une uniforme sur $\{1, \dots, n+1\}$.

EXERCICE 59 (SIMULATION 2)

Soit X une variable aléatoire réelle de fonction de répartition F . On suppose que F est inversible, d'inverse F^{-1} .

- 1) Si U est une variable aléatoire de loi uniforme sur $[0, 1]$, montrer que la variable aléatoire $F^{-1}(U)$ a même loi que X .
- 2) Montrer que $F(X)$ suit une loi uniforme sur $[0, 1]$.

REMARQUE – 2.16 En pratique, ce sont ces résultats que les logiciels de calcul formel utilisent pour simuler n'importe quelle loi à partir d'une uniforme.

La question 2) de l'Exercice 58 ne rentre pas dans le cadre de l'exercice précédent, la fonction $x \mapsto \lfloor 1 + nx \rfloor$ n'étant pas inversible. C'est en fait une fonction *inverse généralisé* : si F est une fonction de répartition continue, alors $F^{-1}(t) = \inf\{y : F(y) > t\}$ est un inverse à gauche pour F . Si F est en particulier inversible, cet inverse à gauche coïncide avec le vrai inverse F^{-1} .

EXERCICE 60

Soit ε une variable aléatoire discrète à valeurs dans $\{-1, 1\}$ et soit X une variable aléatoire à densité f . On suppose que ε et X sont indépendantes.

- 1) Montrer que $Y = \varepsilon X$ possède une densité et la calculer.
- 2) Donner une condition nécessaire et suffisante pour que Y et X aient la même loi.

EXERCICE 61 (CENTRAGE ET RÉDUCTION)

Si $X \sim \mathcal{N}(0, 1)$, déterminer la loi de $\sigma X + m$ avec $\sigma > 0$ et $m \in \mathbf{R}$.

EXERCICE 62 (LOI LOG-NORMALE)

- 1) On considère la variable aléatoire $Y = \exp(X)$ où $X \sim \mathcal{N}(0, 1)$. On note respectivement F et G les fonctions de répartition de Y et X . Exprimer G à l'aide de F , montrer alors que Y admet une densité g que l'on déterminera.
- 2) On dit que Y suit une loi *log-normale*. Est-ce que Y admet une espérance ? Si oui, calculer $\mathbf{E}(Y)$.
- 3) Est-ce que Y^2 admet une espérance ? Si oui, calculer $\mathbf{E}(Y^2)$ puis $\mathbf{V}(Y)$.

EXERCICE 63 (GAMMA ET KHI-DEUX)

On dit qu'une variable aléatoire X suit la loi Gamma $\mathcal{G}(a, \lambda)$ avec $a, \lambda > 0$, si sa densité de probabilité est donnée par

$$f_X(x) = \frac{1}{\Gamma(a)} \lambda^a x^{a-1} \exp(-\lambda x) \mathbb{1}_{x>0}.$$

- 1) Calculer l'espérance et la variance de X .
- 2) Soient X et Y deux variables aléatoires indépendantes et de lois $\mathcal{G}(a, \lambda)$ et $\mathcal{G}(b, \lambda)$ avec $a, b, \lambda > 0$. Si

$$U = X + Y \quad \text{et} \quad V = \frac{X}{X + Y},$$

montrer que U et V sont indépendantes et trouver leurs lois marginales.

- 3) En déduire que

$$B(a, b) := \int_0^1 x^{a-1} (1-x)^{b-1} dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

- 4) Si X_1, \dots, X_n sont des variables aléatoires indépendantes et de même loi $\mathcal{E}(\lambda)$ avec $\lambda > 0$, donner la loi de $S_n = X_1 + \dots + X_n$ et trouver son espérance et sa variance.
- 5) Si Y_1, \dots, Y_n sont des variables aléatoires indépendantes et de même loi $\mathcal{N}(0, 1)$, donner la loi de $T_n = Y_1^2 + \dots + Y_n^2$ et trouver son espérance et sa variance.

EXERCICE 64

Soient X et Y deux variables aléatoires indépendantes de loi uniforme sur $[-a, a]$ avec $a > 0$. On pose

$$U = \frac{X + Y}{2} \quad \text{et} \quad V = \frac{X - Y}{2}.$$

- 1) Soient \mathcal{D} le carré $[-a, a]^2$ et Δ le losange de base $[-a, a]$ et de hauteur $[-a, a]$. Soit h l'application de \mathcal{D} dans Δ définie, pour tout $(x, y) \in \mathcal{D}$, par

$$h(x, y) = \left(\frac{x+y}{2}, \frac{x-y}{2} \right).$$

Montrer que h est un difféomorphisme dont le jacobien ne s'annule pas sur \mathcal{D} .

- 2) Calculer la densité de probabilité du couple (U, V) .
- 3) Montrer que U et V suivent la loi triangulaire symétrique dont la densité est donnée par

$$f(w) = \frac{1}{a^2} (a - |w|) \mathbb{1}_{\{|w| \leq a\}}.$$

- 4) Les variables aléatoires U et V sont-elles indépendantes ?

►CORRECTION.

1) h est bien une fonction de Jacobien non nul, puisqu'il vaut

$$\det \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix} = -1/2.$$

De plus rappelons que le losange de côté a est la boule de rayon a pour la norme 1 :

$$\Delta = \{(x_1, x_2) \in \mathbf{R}^2 : |x_1| + |x_2| \leq a\}.$$

De-même \mathcal{D} est la boule de rayon a pour la norme infinie :

$$\mathcal{D} = \{(x_1, x_2) \in \mathbf{R}^2 : \max(|x_1|, |x_2|) \leq a\}.$$

Soit donc $(u, v) \in \mathcal{D}$ alors on majore

$$\left| \frac{u+v}{2} \right| + \left| \frac{u-v}{2} \right|$$

si $(u, v) \in \mathcal{D}$. On regarde les différents cas, par exemple si $u \geq v \geq 0$, $\left| \frac{u+v}{2} \right| + \left| \frac{u-v}{2} \right| = u \leq |u| \leq a$. De-même pour les autres.

On constate facilement que l'inverse est donné par

$$h^{-1}(u, v) = (u+v, u-v),$$

cet inverse est bien dans le bon espace. Par ailleurs l'application et son inverse sont bien \mathcal{C}^1 car chaque coordonnée est un polynôme en (x, y) donc \mathcal{C}^1 .

2) On utilise ici la méthode des fonctions tests (l'énoncé y incite clairement). D'après la question précédente, on peut réaliser le changement de variable donné par $(u', v') = h(u, v)$.

Soit $\phi : \mathbf{R}^2 \rightarrow \mathbf{R}$ une fonction mesurable bornée. Alors par indépendance la densité du vecteur (X, Y) est le produit des densités et on obtient par transfert

$$\begin{aligned} \mathbf{E}(\phi(U, V)) &= \frac{1}{4a^2} \int_{\mathbf{R}^2} \phi\left(\frac{u+v}{2}, \frac{u-v}{2}\right) \mathbb{1}_{[-a, a]}(u) \mathbb{1}_{[-a, a]}(v) du dv = \frac{1}{4a^2} \int_{\mathcal{D}} \phi\left(\frac{u+v}{2}, \frac{u-v}{2}\right) du dv \\ &= \frac{1}{4a^2} \int_{\Delta} \phi(u', v') |J_{h^{-1}}|(u', v') du' dv' \\ &= \frac{1}{2a^2} \int_{\Delta} \phi(u', v') du' dv' \end{aligned}$$

La densité du couple (U, V) est donc

$$(u, v) \mapsto \frac{1}{2a^2} \mathbb{1}_{\Delta}(u, v).$$

3) On en déduit les densités marginales

$$f_U(u) = \frac{1}{2a^2} \int_{\mathbf{R}} \mathbb{1}_{\Delta}(u, v) dv = \frac{1}{2a^2} \mathbb{1}_{\{|u| \leq a\}} \int_{\mathbf{R}} \mathbb{1}_{\{|v| \leq a-|u|\}}(u, v) dv = \frac{1}{2a^2} \mathbb{1}_{\{|u| \leq a\}} \int_{|u|-a}^{a-|u|} dv = \frac{1}{a^2} (a - |u|) \mathbb{1}_{\{|u| \leq a\}}.$$

4) La densité du couple n'est pas le produit des densités, il n'y a pas indépendance.

3 CONVERGENCES STOCHASTIQUES ET THÉORÈMES LIMITES

Dans tous le chapitre, les variables aléatoires sont définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ et sont, sauf mention contraire, à valeurs dans \mathbf{R}^d . On notera $|\cdot|$ la norme Euclidienne sur \mathbf{R}^d .

Nous allons voir que les convergences stochastiques sont à manipuler avec beaucoup de précautions. Beaucoup de propriétés classiques sur les convergences des suites numériques ne sont plus vraies pour certains modes de convergence probabiliste.

Un exemple typique : si $(x_n)_{n \in \mathbf{N}}$ est une suite numérique telle que $x_n \xrightarrow{n \rightarrow \infty} x$, on aura pas toujours

$$x_n \xrightarrow{n \rightarrow \infty} x \iff x_n - x \xrightarrow{n \rightarrow \infty} 0.$$

§ 3.1. Convergence en probabilité, presque-sûre et L^p

3.1.1 Convergence en probabilité

Revenons sur une situation déjà rencontrée au chapitre précédent. Soit $(X_i)_{i \geq 1}$ une suite de variables indépendantes et de même loi de Bernoulli de paramètre $p \in]0, 1[$. En utilisant l'inégalité de Tchebychev, on a, pour tout $\varepsilon > 0$,

$$\mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - p \right| \geq \varepsilon \right) \leq \frac{\mathbf{V} \left(\frac{1}{n} \sum_{i=1}^n X_i \right)}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2}.$$

Ceci montre que,

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - p \right| \geq \varepsilon \right) = 0.$$

On dit que la suite de variables aléatoires $\left(\frac{1}{n} \sum_{i=1}^n X_i \right)_{n \geq 1}$ converge vers p en probabilité.

DEFINITION – 3.1 On dit qu'une suite $(Z_n)_{n \geq 0}$ de variables aléatoires ou vecteurs aléatoires converge vers Z en probabilité si

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbf{P}(|Z_n - Z| \geq \varepsilon) = 0.$$

On note alors : $Z_n \xrightarrow{n \rightarrow \infty}^{\mathbf{P}} Z$.

Une limite en probabilité jouit également de la propriété d'unicité presque-sûre.

PROPOSITION – 3.1 Une limite en probabilité est presque sûrement unique. Autrement dit : si Z et Z' sont deux limites en probabilité, alors $\mathbf{P}(Z = Z') = 1$.

▮ **PREUVE.** En effet, considérons Z et Z' deux variables aléatoires étant limites en probabilité d'une suite $(Z_n)_{n \geq 0}$. Pour tout $\varepsilon > 0$, on a pour tout entier n les inclusions successives suivantes :

$$\{|Z - Z'| \geq \varepsilon\} \subset \{|Z - Z_n| + |Z' - Z_n| \geq \varepsilon\} \subset \{|Z - Z_n| \geq \varepsilon/2\} \cup \{|Z' - Z_n| \geq \varepsilon/2\}$$

d'où par croissance de la probabilité \mathbf{P} (pour l'inclusion) :

$$\mathbf{P}(|Z - Z'| \geq \varepsilon) \leq \mathbf{P}(|Z - Z_n| + |Z' - Z_n| \geq \varepsilon) \leq \mathbf{P}(|Z - Z_n| \geq \varepsilon/2) + \mathbf{P}(|Z' - Z_n| \geq \varepsilon/2) \xrightarrow{n \rightarrow \infty} 0.$$

Donc $\mathbf{P}(|Z - Z'| \geq \varepsilon) = 0$ pour tout $\varepsilon > 0$ et

$$\mathbf{P}(|Z - Z'| > 0) \leq \mathbf{P} \left(\bigcup_{k=1}^{\infty} \left\{ |Z - Z'| \geq \frac{1}{k} \right\} \right) \leq \sum_{k=1}^{\infty} \mathbf{P} \left(|Z - Z'| \geq \frac{1}{k} \right) = \sum_{k=1}^{\infty} 0 = 0.$$

Ainsi $\mathbf{P}(|Z - Z'| = 0) = 1$ ce qui équivaut à $\mathbf{P}(Z = Z') = 1$ et donc $Z = Z'$ \mathbf{P} -p.s. ▮

PROPOSITION – 3.2 (CONTINUITÉ, "OPÉRATIONS" ET CONVERGENCE EN PROBABILITÉ.)

Soit $(Z_n)_{n \geq 0}$ une suite de variables ou vecteurs aléatoires réelles à valeurs dans \mathbf{R}^d telle que $Z_n \xrightarrow{\mathbf{P}} Z$.

Alors pour tout entier d' et toute application $g : \mathbf{R}^d \rightarrow \mathbf{R}^{d'}$ continue, $g(Z_n) \xrightarrow{\mathbf{P}} g(Z)$.

Par ailleurs, si (Y_n) une autre suite de vecteurs aléatoires tels que $Y_n \xrightarrow{\mathbf{P}} Y$,

(i) $Y_n + Z_n \xrightarrow{\mathbf{P}} Y + Z$,

(ii) si $d = 1$, $Y_n Z_n \xrightarrow{\mathbf{P}} YZ$, si $d \geq 2$, on a $\langle Y_n, Z_n \rangle \xrightarrow{\mathbf{P}} \langle Y, Z \rangle$,

(iii) si $d = 1$ et $Z > 0$ ps, $Y_n/Z_n \xrightarrow{\mathbf{P}} Y/Z$.

EXERCICE 65

1) Justifier que pour tout $n \in \mathbf{N}$,

$$\forall \eta > 0, \forall \varepsilon > 0, \{|g(Z_n) - g(Z)| \geq \eta\} = \{|g(Z_n) - g(Z)| \mathbb{1}_{|Z_n - Z| \geq \varepsilon} \geq \eta\} \cup \{|g(Z_n) - g(Z)| \mathbb{1}_{|Z_n - Z| < \varepsilon} \geq \eta\},$$

puis l'inclusion

$$\forall \eta > 0, \forall \varepsilon > 0, \{|g(Z_n) - g(Z)| \geq \eta\} \subset \{|Z_n - Z| \geq \varepsilon\} \cup \{|g(Z_n) - g(Z)| \geq \eta\} \cap \{|Z_n - Z| < \varepsilon\}.$$

2) Démontrer la première propriété en utilisant la continuité de g .

3) Démontrer le point (i). ♣ INDICATION – on pourra remarquer en le justifiant que l'on a, pour tout $n \in \mathbf{N}$, les inclusions successives suivantes :

$$\forall \varepsilon > 0, \{|Y_n + Z_n - (Y + Z)| \geq \varepsilon\} \subset \{|Y_n - Y| + |Z_n - Z| \geq \varepsilon\} \subset \{|Y_n - Y| \geq \varepsilon/2\} \cup \{|Z_n - Z| \geq \varepsilon/2\}.$$

Voici une autre propriété qui relie la convergence en probabilité d'une suite de vecteurs et de leurs coordonnées, propriété que l'on connaît évidemment déjà pour les suites numériques usuelles de vecteurs.

PROPOSITION – 3.3 (CONVERGENCE EN PROBABILITÉ DES SUITES DE VECTEURS ET DE LEURS COORDONNÉES.)

Soit $(Z_n)_{n \geq 0}$ une suite de variables aléatoires et à valeurs dans \mathbf{R}^d avec $d \geq 2$ entier. Pour tout $n \in \mathbf{N}$, on note $Z_n = (Z_{1,n}, \dots, Z_{d,n})$ ou encore $Z_n = (Z_{j,n} : 1 \leq j \leq d)$

Soit $Z = (Z_1, \dots, Z_d) = (Z_j : 1 \leq j \leq d)$ un vecteur aléatoire de \mathbf{R}^d . Alors on a :

$(Z_n) \xrightarrow{\mathbf{P}_{n \rightarrow \infty}} Z$ ssi pour tout $1 \leq j \leq d$, la suite de variables aléatoires réelles $(Z_{j,n})$ converge en probabilité vers Z_j .

EXERCICE 66

Démontrer l'équivalence énoncée dans la Proposition 3.3.

On peut également définir la convergence en probabilité comme la convergence associée à une métrique $d^{\mathbf{P}}$ définie sur l'ensemble des variables aléatoires sur $(\Omega, \mathcal{A}, \mathbf{P})$.

Plus précisément, considérons deux variables aléatoires X, Y , alors

$$d^{\mathbf{P}}(X, Y) = \mathbf{E}(\min(|X - Y|, 1))$$

est une distance sur l'espace des variables aléatoires.

PROPOSITION – 3.4 Soit $(Z_n)_{n \geq 0}$ une suite de variables aléatoires, et Z une variable aléatoire à valeurs dans \mathbf{R}^d . Alors

$$Z_n \xrightarrow{\mathbf{P}_{n \rightarrow \infty}} Z \iff d^{\mathbf{P}}(Z_n, Z) \xrightarrow{n \rightarrow \infty} 0.$$

Autrement dit, la convergence en probabilité est métrisable.

▮ PREUVE. Soit $\varepsilon \in]0, 1]$. Alors

$$\mathbf{P}(|Z_n - Z| > \varepsilon) \leq \mathbf{P}(\min(|Z_n - Z|, 1) > \varepsilon) \leq \frac{d(Z_n, Z)}{\varepsilon}.$$

Par ailleurs, on a pour tout $\varepsilon > 0$

$$d^{\mathbf{P}}(Z_n, Z) = \mathbf{E}(\min(|Z_n - Z|, 1)) = \mathbf{E}(\min(|Z_n - Z|, 1)\mathbb{1}_{|Z_n - Z| \geq \varepsilon}) + \mathbf{E}(\min(|Z_n - Z|, 1)\mathbb{1}_{|Z_n - Z| < \varepsilon}),$$

d'où $d^{\mathbf{P}}(Z_n, Z) \leq \varepsilon + \mathbf{P}(|Z_n - Z| \geq \varepsilon)$.

Ainsi si Z_n converge en probabilité, $\lim_{n \rightarrow \infty} d^{\mathbf{P}}(Z_n, Z) \leq \varepsilon$, en faisant tendre ε vers 0, on a la convergence au sens de $d^{\mathbf{P}}$.

Inversement si $d^{\mathbf{P}}(Z_n, Z) \rightarrow_{n \rightarrow \infty} 0$, alors $\mathbf{P}(|Z_n - Z| > \varepsilon) \rightarrow_{n \rightarrow \infty} 0$ pour tout $\varepsilon \in]0, 1]$, si $\varepsilon > 1$, on écrit simplement que

$$\mathbf{P}(|Z_n - Z| > \varepsilon) \leq \mathbf{P}(|Z_n - Z| > 1),$$

et le cas précédent s'applique. ┘

3.1.2 Convergence dans L^p

Remarquons aussi que pour l'exemple considéré, la convergence en probabilité de la suite de variables aléatoires $\left(\frac{1}{n} \sum_{i=1}^n X_i\right)_{n \geq 1}$

vers p était une conséquence de la convergence de $\mathbf{V}\left(\frac{1}{n} \sum_{i=1}^n X_i\right)$ vers 0.

C'est-à-dire de : $\mathbf{E}\left(\left(\frac{1}{n} \sum_{i=1}^n X_i - p\right)^2\right) \rightarrow_{n \rightarrow \infty} 0$.

On dit que la suite de variables aléatoires $\left(\frac{1}{n} \sum_{i=1}^n X_i\right)_{n \geq 1}$ converge vers p dans $L^2(\Omega, \mathcal{A}, \mathbf{P})$.

Consultez les rappels de l'appendice pour plus de détails sur ces espaces.

DEFINITION – 3.2 Soit $p \geq 1$ un réel fixé. On dit qu'une suite $(Z_n)_{n \geq 0}$ de variables ou vecteurs aléatoires de $L^p(\Omega, \mathcal{A}, \mathbf{P})$ converge vers Z dans $L^p(\Omega, \mathcal{A}, \mathbf{P})$ et on note $Z_n \xrightarrow[n \rightarrow \infty]{L^p} Z$ si

$$\lim_{n \rightarrow \infty} \mathbf{E}(|Z_n - Z|^p) = 0 \iff \lim_{n \rightarrow \infty} \|Z_n - Z\|_{L^p} = 0.$$

Comme tous les $L^p(\Omega, \mathcal{A}, \mathbf{P})$ sont des espaces de Banach, ce sont en particulier des espaces fermés. La limite Z est alors forcément dans L^p .

La convergence L^p est par définition métrisable pour la distance associée à la norme $\|\cdot\|_{L^p}$.

PROPOSITION – 3.5 Soit $p \geq 1$ un réel fixé. Soit $(Z_n)_{n \in \mathbb{N}}$ une suite de variables ou vecteurs aléatoires. Alors on a :

- (i) $Z_n \xrightarrow[n \rightarrow \infty]{L^p} Z \implies Z_n \xrightarrow[n \rightarrow \infty]{\mathbf{P}} Z$.
- (ii) La limite L^p est unique presque-sûrement.
- (iii) La convergence L^p implique la convergence L^q pour tout $1 \leq q < p$.

┐ **PREUVE.** (i) L'inégalité de Markov montre que pour tout $\varepsilon > 0$ et tout entier $n \geq 0$,

$$\mathbf{P}(|Z_n - Z| \geq \varepsilon) = \mathbf{P}(|Z_n - Z|^p \geq \varepsilon^p) \leq \frac{1}{\varepsilon^p} \mathbf{E}(|Z_n - Z|^p).$$

La convergence dans L^p implique donc la convergence en probabilité.

La propriété (ii) est donc une conséquence de l'unicité \mathbf{P} -p.s. de la limite en probabilité énoncée dans la Proposition 3.1.

Enfin, le point (iii) résulte directement de l'inégalité qui suit la Proposition 5.4 : on a $\|X_n\|_q \leq \|X_n\|_p$ comme $\mathbf{P}(\Omega) = 1$. ┘

EXERCICE 67

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles indépendantes et de même loi, d'espérance m et de variance $\sigma^2 > 0$. Pour tout $n \geq 1$, on pose

$$S_n = \sum_{k=1}^n \frac{X_k}{k}.$$

♣ INDICATION – Pour cet exercice, on pourra utiliser (sans le démontrer) le développement asymptotique

$$\sum_{k=1}^n \frac{1}{k} = \log n + \gamma + \frac{1}{2n} + o\left(\frac{1}{n}\right) \quad \text{où } \gamma \text{ est la constante d'Euler.}$$

1) Calculer l'espérance et la variance de S_n .

2) Montrer les deux égalités suivantes : $\mathbf{E} \left[\left(\frac{S_n}{\log n} - m \right)^2 \right] = \frac{\text{Var}(S_n)}{(\log n)^2} + \left(\frac{\mathbf{E}[S_n]}{\log n} - m \right)^2 = O\left(\frac{1}{(\log n)^2}\right)$.

3) En déduire que la suite $\frac{S_n}{\log n}$ converge en probabilité vers m lorsque $n \rightarrow +\infty$.

3.1.3 Convergence presque-sûre

Dans l'exemple considéré, a-t-on **P**-p.s. la convergence ponctuelle de la suite $\left(\frac{1}{n} \sum_{i=1}^n X_i \right)_{n \geq 1}$? C'est-à-dire, peut-on montrer que l'ensemble des $\omega \in \Omega$ tels que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i(\omega) = p$$

est de probabilité 1 ?

DEFINITION – 3.3 On dit qu'une suite $(Z_n)_{n \geq 0}$ de variables ou vecteurs aléatoires converge **P**-presque sûrement vers Z (ou simplement converge presque sûrement vers Z) et on note $Z_n \xrightarrow{\mathbf{P}\text{-p.s.}} Z$ ou simplement $Z_n \xrightarrow{p.s.} Z$ si

$$\mathbf{P} \left(\lim_{n \rightarrow \infty} Z_n = Z \right) = 1.$$

Ou, de manière équivalente, s'il existe un événement **P**-négligeable N de \mathcal{A} (c'est-à-dire que $\mathbf{P}(N) = 0$) tel que

$$\forall \omega \in \Omega \setminus N, \quad Z_n(\omega) \xrightarrow{n \rightarrow \infty} Z(\omega),$$

en tant que suite de \mathbf{R}^d .

Ce mode de convergence est donc fortement lié à la convergence des suites usuelles à valeurs dans \mathbf{R}^d en Analyse : on fixe un ω en dehors d'un ensemble négligeable, et on regarde la convergence de $(Z_n(\omega))_{n \in \mathbf{N}}$ en tant que suite de \mathbf{R}^d . Autrement dit, ceci correspond à la convergence ponctuelle (simple) des suites de fonctions $Z_n : \Omega \rightarrow \mathbf{R}^d$ sur tout Ω hormis un ensemble négligeable.

Les Propositions 3.2 et 3.3 (sur resp., la composition de la convergence en probabilité avec une application continue ou une combinaison d'opérations sur des suites convergant en probabilité; resp. sur les liens avec la convergence des suites des coordonnées) s'étendent à la convergence presque sûre. Ceci résulte de la proposition qui suit.

Vue la définition de la convergence presque-sûre, il n'est pas surprenant qu'elle hérite des propriétés habituelles des limites de suites numériques.

PROPOSITION – 3.6 (CONTINUITÉ, PROJECTIONS ET CONVERGENCE PRESQUE SÛRE.)

Soit $(Z_n)_{n \in \mathbf{N}}$ une suite de variables ou vecteurs aléatoires à valeurs dans \mathbf{R}^d (avec $d \geq 1$ entier) qui converge en probabilité vers Z .

(i) Pour tout entier $d' \in \mathbf{N}^*$ et toute application $g : \mathbf{R}^d \rightarrow \mathbf{R}^{d'}$ continue, $g(Z_n) \xrightarrow{p.s.} g(Z)$,

(ii) si $d \geq 2$, on note pour tout $1 \leq j \leq d$, $Z_{j,n}$ (resp. Z_j) la variable aléatoire égale à la $j^{\text{ème}}$ coordonnée de Z_n (resp. Z). Alors on a l'équivalence suivante :

$$Z_n \xrightarrow{p.s.} Z \iff \forall 1 \leq j \leq d, \quad Z_{j,n} \xrightarrow{p.s.} Z_j.$$

L'idée maîtresse de cette preuve est la suivante :
utiliser les résultats connus pour les suites numériques sur des ensembles non négligeables.

▮ PREUVE. Pour (i), supposons que N désigne un événement de \mathcal{A} tel que :

$$\mathbf{P}(N) = 0 \quad \text{et} \quad \text{pour tout } \omega \in \Omega \setminus N, Z_n(\omega) \xrightarrow{n \rightarrow \infty} Z(\omega).$$

La fonction $g : \mathbf{R}^d \rightarrow \mathbf{R}^{d'}$ étant continue sur \mathbf{R}^d , on sait que : pour toute suite (z_n) de \mathbf{R}^d convergeant vers un élément $z \in \mathbf{R}^d$, la suite des images converge avec $(g(z_n)) \xrightarrow{n \rightarrow \infty} g(z)$ (dans $\mathbf{R}^{d'}$). On en déduit que :

$$\text{pour tout } \omega \in \Omega \setminus N, \quad g(Z_n(\omega)) \xrightarrow{n \rightarrow \infty} g(Z(\omega)).$$

Donc l'évènement $\Omega' = \{\omega \in \Omega : \lim_{n \rightarrow \infty} g(Z_n(\omega)) = g(Z(\omega))\}$ est quasi-certain puisque l'on a $N^c \subset \Omega'$.

Pour (ii) : la condition nécessaire résulte du fait que les projections $p_j : x = (x_1, \dots, x_d) \in \mathbf{R}^d \mapsto x_j \in \mathbf{R}$ sont évidemment continues.

Pour la réciproque : on sait par hypothèse qu'il existe d événements notés N_1, \dots, N_d de \mathcal{A} tous \mathbf{P} -négligeables et tels que :

$$\text{pour tout } 1 \leq j \leq d : \quad \forall \omega \in \Omega \setminus N_j, Z_{j,n}(\omega) \xrightarrow{n \rightarrow \infty} Z_j(\omega). \quad (3.1)$$

On considère alors l'évènement $N \in \mathcal{A}$ défini par : $N = \bigcup_{j=1}^d N_j$. On a la convergence (3.1) pour tout $\omega \in N^c$. De plus, N étant une réunion finie d'évènements \mathbf{P} -négligeables, il l'est aussi.

▮

Afin d'établir des propriétés importantes de la convergence presque sûre, nous allons la reformuler. Notamment des limites supérieures d'ensembles apparaîtront de manière naturelle, nous commençons donc par donner un sens à ces notions.

REMARQUE – 3.1 (LIMITES SUPÉRIEURE ET INFÉRIEURE.)

Pour des suites : soit $(u_n)_{n \in \mathbf{N}}$ une suite numérique.

Les suites $\left(\sup_{k \geq n} u_k \right)_{n \in \mathbf{N}}$ et $\left(\inf_{k \geq n} u_k \right)_{n \in \mathbf{N}}$ sont monotones (en n , la première est décroissante, la seconde croissante). Elles convergent donc, vers une limite éventuellement infinie, que l'on note respectivement

$$\limsup_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} \sup_{k \geq n} u_k \quad \text{et} \quad \liminf_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} \inf_{k \geq n} u_k$$

qui sont deux éléments de $\overline{\mathbf{R}}$ appelées *Limite supérieure* (resp. *inférieure*) de la suite $(u_n)_{n \in \mathbf{N}}$. L'avantage de ces deux objets est qu'ils existent toujours, *a contrario* ce n'est pas le cas pour les limites.

Sur la même idée (remplacer la borne supérieure par une réunion, inférieure par une intersection), considérons une suite $(A_n)_{n \in \mathbf{N}}$ d'évènements de \mathcal{A} . On définit

$$\boxed{\limsup_n A_n = \bigcap_{n \in \mathbf{N}} \bigcup_{k \geq n} A_k \quad \text{et} \quad \liminf_n A_n = \bigcup_{n \in \mathbf{N}} \bigcap_{k \geq n} A_k.}$$

Ces deux ensembles sont bien des événements de \mathcal{A} appelés *limite supérieure* (resp. *inférieure*) de la suite $(A_n)_{n \in \mathbf{N}}$.

On peut alors reformuler la définition de la convergence presque-sûre de nombreuses façons. Remarquons qu'en écrivant la définition de la limite, on a

$$\left\{ \lim_{n \rightarrow \infty} Z_n = Z \right\} = \bigcap_{\varepsilon > 0} A_\varepsilon \quad \text{où} \quad A_\varepsilon = \bigcup_{N \in \mathbf{N}} \bigcap_{n \geq N} \{|Z_n - Z| < \varepsilon\}.$$

En effet : $Z_n \xrightarrow[n \rightarrow \infty]{p.s.} Z$ si et seulement si il existe un évènement négligeable N tel que

$$\forall \omega \in \Omega \setminus N, \quad \forall \varepsilon > 0, \exists N, \quad \forall n \geq N, \quad |Z_n(\omega) - Z(\omega)| \leq \varepsilon.$$

L'ensemble des $\omega \in \Omega$ vérifiant

$$\forall \varepsilon > 0, \exists N, \quad \forall n \geq N, \quad |Z_n(\omega) - Z(\omega)| \leq \varepsilon,$$

est exactement

$$\bigcap_{\varepsilon > 0} \bigcup_{N \in \mathbf{N}} \bigcap_{n \geq N} \{|Z_n - Z| < \varepsilon\} = \bigcap_{\varepsilon > 0} A_\varepsilon.$$

Les ensembles A_ε étant croissants en ε , on montre facilement (par convergence monotone des probabilités, Voir Chap. 2) que

$$\bigcap_{\varepsilon > 0} A_\varepsilon = \bigcap_{k \geq 1} A_{\frac{1}{k}},$$

ce qui a l'avantage de nous ramener à une intersection dénombrable. Et on a de manière évidente en raisonnant par l'absurde que :

$$\mathbf{P}\left(\lim_{n \rightarrow \infty} Z_n = Z\right) = 1 \iff \forall k \geq 1, \quad \mathbf{P}\left(A_{\frac{1}{k}}\right) = 1,$$

soit en regardant le complémentaire pour faire intervenir la limite supérieure :

$$\begin{aligned} \mathbf{P}\left(\lim_{n \rightarrow \infty} Z_n = Z\right) = 1 &\iff \forall k \geq 1, \quad \mathbf{P}\left(A_{1/k}^c\right) = \mathbf{P}\left(\bigcap_{N \in \mathbf{N}} \bigcup_{n \geq N} \{|Z_n - Z| \geq 1/k\}\right) = 0 \\ &\iff \forall k \geq 1, \quad \mathbf{P}\left(\limsup_{n \rightarrow \infty} \{|Z_n - Z| \geq 1/k\}\right) = 0. \end{aligned}$$

Par ailleurs, dire que $\omega \in \limsup_{n \rightarrow \infty} \{|Z_n - Z| \geq 1/k\}$ équivaut à dire que $|Z_n(\omega) - Z(\omega)| \geq 1/k$ pour une infinité de n .

Remarquons que les évènements $\bigcup_{n \geq N} \{|Z_n - Z| \geq 1/k\}$ sont décroissants en N de tel sorte que

$$\mathbf{P}\left(\limsup_{n \rightarrow \infty} \{|Z_n - Z| \geq 1/k\}\right) = \mathbf{P}\left(\bigcap_{N \in \mathbf{N}} \bigcup_{n \geq N} \{|Z_n - Z| \geq 1/k\}\right) = \lim_{N \rightarrow \infty} \mathbf{P}\left(\bigcup_{n \geq N} \{|Z_n - Z| \geq 1/k\}\right).$$

Donc

$$\mathbf{P}\left(\lim_{n \rightarrow \infty} Z_n = Z\right) = 1 \iff \forall k \geq 1, \quad \lim_{N \rightarrow \infty} \mathbf{P}\left(\bigcup_{n \geq N} \{|Z_n - Z| \geq 1/k\}\right) = 0.$$

Énonçons ce résultat sous forme de proposition.

PROPOSITION – 3.7 Une suite $(Z_n)_{n \geq 0}$ de variables aléatoires converge presque-sûrement vers Z si et seulement si

$$\forall \varepsilon > 0, \quad \mathbf{P}\left(\limsup_{n \rightarrow \infty} \{|Z_n - Z| \geq \varepsilon\}\right) = 0$$

ou de manière équivalente :

$$\forall \varepsilon > 0, \quad \lim_{N \rightarrow \infty} \mathbf{P}\left(\bigcup_{n \geq N} \{|Z_n - Z| \geq \varepsilon\}\right) = 0.$$

On en déduit alors très facilement que la convergence presque-sûre implique la convergence en probabilité puisque pour tout $\varepsilon > 0$ et $N \in \mathbf{N}$,

$$\mathbf{P}(\{|Z_N - Z| \geq \varepsilon\}) \leq \mathbf{P}\left(\bigcup_{n \geq N} \{|Z_n - Z| \geq \varepsilon\}\right) \xrightarrow{N \rightarrow \infty} 0.$$

Par conséquent, l'unicité de la limite en probabilité implique l'unicité pour la convergence presque-sûre. On a alors la propriété suivante.

PROPOSITION – 3.8 (i) La convergence presque-sûre implique la convergence en probabilité : si $Z_n \xrightarrow[n \rightarrow \infty]{p.s.} Z$ alors $Z_n \xrightarrow[n \rightarrow \infty]{\mathbf{P}} Z$.

(ii) La limite presque-sûre d'une suite de variables aléatoires est presque-sûrement unique.

Retournons à l'exemple initial : comment montre-t-on la convergence presque-sûre de la suite $\left(\frac{1}{n} \sum_{i=1}^n X_i\right)_{n \geq 1}$ vers p ?

Le plus souvent on pourra faire appel à des théorèmes limites. Sinon, un outil très utile est le lemme de Borel-Cantelli dont le but est de caractériser la négligeabilité d'une limite supérieure d'ensembles.

PROPOSITION – 3.9 (LEMME DE BOREL-CANTELLI) Soit $(A_n)_{n \geq 0}$ une suite d'événements de \mathcal{A} .

(i) Si $\sum_{n \geq 1} \mathbf{P}(A_n) < \infty$, alors $\mathbf{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 0$.

(ii) Si les événements A_n sont indépendants et si $\sum_{n \geq 1} \mathbf{P}(A_n) = \infty$ alors $\mathbf{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 1$.

On obtient un critère de convergence presque-sûre qui se démontre en combinant la Proposition 3.7 et le Lemme de Borel-Cantelli (Proposition 3.9 (a)).

PROPOSITION – 3.10 (CRITÈRE DE CONVERGENCE P.S.) Soient $(Z_n)_{n \geq 0}$ une suite de variables aléatoires et Z une autre variable aléatoire à valeurs dans \mathbf{R}^d avec $d \in \mathbf{N}^*$.

$$\forall \varepsilon > 0, \quad \sum_{n \geq 1} \mathbf{P}(|Z_n - Z| > \varepsilon) < \infty \quad \implies \quad Z_n \xrightarrow[n \rightarrow \infty]{p.s.} Z.$$

Notez l'analogie suivante : la limite en probabilité est à la convergence d'un terme général de série vers zéro, ce que la limite presque-sûre est à la convergence de la série associée.

▮ **PREUVE.** (i) Si $\sum_{n \geq 1} \mathbf{P}(A_n) < \infty$, on a déjà remarqué que

$$\mathbf{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = \mathbf{P}\left(\bigcap_{N \in \mathbf{N}} \bigcup_{n \geq N} A_n\right) = \lim_{N \rightarrow \infty} \mathbf{P}\left(\bigcup_{n \geq N} A_n\right) \leq \lim_{N \rightarrow \infty} \sum_{n \geq N} \mathbf{P}(A_n) = 0,$$

qui est un reste de série convergente.

(ii) Si les événements A_n sont indépendants et si $\sum_{n \geq 1} \mathbf{P}(A_n) = \infty$, on a, en utilisant en premier qu'une intersection d'événements quasi-certains est quasi-certaine, puis l'indépendance des A_n^c :

$$\begin{aligned} \mathbf{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 1 &\iff \forall N \geq 1, \quad \mathbf{P}\left(\bigcup_{n \geq N} A_n\right) = 1 \iff \forall N \geq 1, \quad \mathbf{P}\left(\bigcap_{n \geq N} A_n^c\right) = 0 \\ &\iff \forall N \geq 1, \quad \lim_{m \rightarrow \infty} \mathbf{P}\left(\bigcap_{n=N}^m A_n^c\right) = 0 \\ &\iff \forall N \geq 1, \quad \lim_{m \rightarrow \infty} \prod_{n=N}^m \mathbf{P}(A_n^c) = 0 \\ &\iff \forall N \geq 1, \quad \lim_{m \rightarrow \infty} \prod_{n=N}^m (1 - \mathbf{P}(A_n)) = 0. \end{aligned}$$

Or $\prod_{n=N}^m (1 - \mathbf{P}(A_n)) \leq \prod_{n=N}^m e^{-\mathbf{P}(A_n)} = e^{-\sum_{n=N}^m \mathbf{P}(A_n)}$ et puisque la série diverge :

$$\lim_{m \rightarrow \infty} \prod_{n=N}^m (1 - \mathbf{P}(A_n)) \leq \lim_{m \rightarrow \infty} e^{-\sum_{n=N}^m \mathbf{P}(A_n)} = 0.$$

D'où le résultat.

▮

Il en découle le résultat suivant connu sous le nom de *Loi du 0-1*, *Loi de Blumenthal* ou encore *Loi du "tout ou rien"*.

COROLLAIRE – 3.1 (LOI DU 0-1) Soit $(A_n)_{n \geq 1}$ une suite d'événements de \mathcal{A} .

Si les événements A_n sont indépendants, alors $\mathbf{P} \left(\limsup_{n \rightarrow \infty} A_n \right) \in \{0, 1\}$.

Dans l'exemple, en utilisant l'inégalité de Markov, pour tout $\varepsilon > 0$, on a

$$\sum_{n=1}^{\infty} \mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - p \right| \geq \varepsilon \right) = \sum_{n=1}^{\infty} \mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - p \right|^4 \geq \varepsilon^4 \right) \leq \sum_{n=1}^{\infty} \frac{1}{\varepsilon^4} \mathbf{E} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - p \right|^4 \right).$$

Montrons qu'il existe une constante $C > 0$ telle que

$$\mathbf{E} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - p \right|^4 \right) \leq \frac{C}{n^2}.$$

Auquel cas, la convergence la série est assurée et la convergence presque-sûre établie d'après le lemme de Borel Cantelli.

On note pour tout entier $i \geq 1$, $Y_i := X_i - p$. Alors les Y_i sont iid, centrées et il faut développer la somme :

$$\left| \frac{1}{n} \sum_{i=1}^n X_i - p \right|^4 = \frac{1}{n^4} \left(\sum_{i=1}^n Y_i \right)^4 = \frac{1}{n^4} \left(\sum_{1 \leq i \leq n} Y_i^4 + 4 \sum_{1 \leq i \neq j \leq n} Y_i^3 Y_j + 6 \sum_{1 \leq i \neq j \leq n} Y_i^2 Y_j^2 + 12 \sum_{1 \leq i \neq j \neq k \leq n} Y_i Y_j Y_k^2 + 24 \sum_{1 \leq i \neq j \neq k \neq l \leq n} Y_i Y_j Y_k Y_l \right).$$

En utilisant la linéarité de l'espérance et le fait que les variables aléatoires $(Y_i)_{i \geq 1}$ sont indépendantes et centrées, on déduit :

$$\mathbf{E} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - p \right|^4 \right) = \frac{1}{n^4} \left(\sum_{1 \leq i \leq n} \mathbf{E}(Y_i^4) + 6 \sum_{1 \leq i \neq j \leq n} \mathbf{E}(Y_i^2) \mathbf{E}(Y_j^2) \right).$$

Comme les variables aléatoires Y_i sont identiquement distribuées, on a de plus :

$$\mathbf{E} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - p \right|^4 \right) = \frac{1}{n^4} \left(\sum_{1 \leq i \leq n} \mathbf{E}(Y_1^4) + 6 \sum_{1 \leq i \neq j \leq n} \mathbf{E}(Y_1^2)^2 \right) = \frac{1}{n^4} (n \mathbf{E}(Y_1^4) + 3n(n-1) \mathbf{E}(Y_1^2)^2).$$

Or comme X_1 est dans $L^4(\Omega)$, et même dans $L^\infty(\Omega)$ ici, $Y_1 = X_1 - p$ aussi. Donc Y_1 admet un moment d'ordre quatre et *a fortiori* d'ordre deux. Donc la quantité $\tilde{C}_p := \max(\mathbf{E}(Y_1^4), \mathbf{E}(Y_1^2)^2) \in]0, \infty[$, puisque

$$\mathbf{V}(Y_1^2) = \mathbf{V}((X_1 - p)^2) = \mathbf{V}(X_1^2) - 4p \mathbf{Cov}(X_1, X_1^2) + 4p^2 \mathbf{V}(X_1) = p - p^2 \neq 0.$$

D'où :

$$\mathbf{E} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - p \right|^4 \right) = \frac{1}{n^4} (n \mathbf{E}(Y_1^4) + 3n(n-1) \mathbf{E}(Y_1^2)^2) \leq \frac{(3n^2 - 2n) \tilde{C}_p}{n^4} \leq 3 \tilde{C}_p / n^2.$$

Ainsi, par comparaison, la série $\sum_{n=1}^{\infty} \mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - p \right| \geq \varepsilon \right)$ converge. D'après le lemme de Borel-Cantelli, on en déduit que pour tout $\varepsilon > 0$

$$\mathbf{P} \left(\limsup_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{i=1}^n X_i - p \right| \geq \varepsilon \right) = 0,$$

c'est-à-dire que pour tout $\varepsilon > 0$, l'événement $\left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - p \right| \geq \varepsilon \right\}$ a lieu pour seulement un nombre fini de n et donc :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = p \quad \mathbf{P} - \text{p.s.}$$

En fait cette convergence rentre dans le cadre des lois des grands nombres : elles seront vues plus tard.

EXERCICE 68 (LA CONVERGENCE EN PROBABILITÉ N'IMPLIQUE PAS LA CONVERGENCE P.S.)

Soit $(Z_n)_{n \geq 1}$ une suite de variables réelles indépendantes telles que, pour $n \geq 1$,

$$\mathbf{P}(Z_n = 0) = 1 - \frac{1}{n}, \quad \mathbf{P}(Z_n = 1) = \frac{1}{n}.$$

- 1) Montrer que la suite $(Z_n)_{n \geq 1}$ tend vers 0 en probabilité mais pas presque-sûrement.
- 2) La suite $(Z_n)_{n \geq 1}$ converge-t-elle dans L^1 ? Et dans L^p pour $p \geq 1$ quelconque? Justifiez.

►CORRECTION.

1) Pour tout $\varepsilon > 0$, on a

$$\mathbf{P}(|Z_n| \geq \varepsilon) = \mathbf{P}(Z_n = 1) = \frac{1}{n} \xrightarrow{n \rightarrow \infty} 0.$$

Donc la suite $(Z_n)_{n \geq 1}$ tend vers 0 en probabilité. Mais comme

$$\sum_{n=1}^{\infty} \mathbf{P}(Z_n = 1) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty,$$

d'après le lemme de Borel-Cantelli (cas où l'on a indépendance), l'événement $\{Z_n = 1\}$ a lieu pour une infinité de n , P-p.s. Donc la suite $(Z_n)_{n \geq 1}$ ne tend pas presque-sûrement vers 0.

2) Puisque $Z_n \xrightarrow{\mathbf{P}}_{n \rightarrow \infty} 0$, si la suite $(Z_n)_{n \geq 1}$ converge dans L^1 c'est nécessairement vers 0. Or ceci est vrai car :

$$\mathbf{E}(|Z_n - 0|) \stackrel{\text{ici}}{=} \mathbf{E}(Z_n) = 0 \times \mathbf{P}(Z_n = 0) + 1 \times \mathbf{P}(Z_n = 1) = 1/n \xrightarrow{n \rightarrow \infty} 0.$$

D'autre part, on peut remarquer que l'on a : pour tout réel $p \geq 1$ et tout $n \in \mathbf{N}$, $Z_n^p = Z_n$ puisque Z_n prend les valeurs 0 et 1 (avec $0^p = 0$ et $1^p = 1$). Par conséquent, la suite (Z_n) converge vers 0 dans tous les espaces L^p (avec $p \geq 1$ quelconque).

EXERCICE 69 (LA CONVERGENCE EN PROBABILITÉ N'IMPLIQUE PAS LA CONVERGENCE L^p)

Soit, pour $n \geq 1$, Z_n suivant une loi de Cauchy de paramètre $\frac{1}{n}$, c'est-à-dire ayant pour densité

$$\forall x \in \mathbf{R}, \quad f_n(x) = \frac{1/n}{\pi((1/n)^2 + x^2)}.$$

- 1) Pour tout $n \in \mathbf{N}$, Z_n appartient-elle à $L^p(\Omega)$ pour un certain $p \geq 1$?
- 2) Étudier la convergence L^p et la convergence en probabilité de $(Z_n)_{n \in \mathbf{N}}$.

►CORRECTION.

Remarquons que Z_n n'appartient à aucun des espaces L^p pour $p \geq 1$ car :

$$\mathbf{E}(|Z_n|) = \int_{\mathbf{R}} |x| f_n(x) dx = \frac{2}{\pi n} \int_0^{+\infty} \frac{x}{(1/n)^2 + x^2} dx$$

qui vaut $+\infty$ en utilisant le fait que cette intégrale diverge puisque qu'on a par exemple :

$$\int_1^{+\infty} \frac{x}{\pi((1/n)^2 + x^2)} dx \geq \int_1^{+\infty} \frac{1}{2x} dx$$

et la dernière intégrale est divergente et vaut $+\infty$ par le Critère de Riemann.

Donc la suite $(Z_n)_{n \geq 1}$ ne peut pas converger dans un espace L^p avec $p \geq 1$. Néanmoins, pour tout $\varepsilon > 0$,

$$\mathbf{P}(|Z_n| \geq \varepsilon) = 1 - \int_{-\varepsilon}^{\varepsilon} f_n(x) dx = 1 - \frac{2}{\pi} \arctan(n\varepsilon) \xrightarrow{n \rightarrow \infty} 0.$$

D'où la convergence en probabilité de la suite $(Z_n)_{n \geq 1}$ vers 0.

Cependant, nous avons la propriété suivante

PROPOSITION – 3.11 (CONVERGENCE EN PROBA ET PS À UNE SOUS-SUITE PRÊT) Si une suite $(Z_n)_{n \geq 0}$ de variables aléatoires converge en probabilité vers Z , alors il existe une sous-suite $(n_k)_{k \in \mathbf{N}}$ telle que $Z_{n_k} \xrightarrow[p.s.]{k \rightarrow \infty} Z$.

▮ PREUVE. D'après la définition de la convergence en probabilité, il existe pour tout $p \in \mathbf{N}$, un entier n_p tel que

$$\mathbf{P}\left(|Z_{n_p} - Z| > \frac{1}{p}\right) \leq \frac{1}{2^p}.$$

On peut supposer que $(n_p)_{p \in \mathbf{N}}$ est croissante. Soit $\varepsilon > 0$, on choisit p assez grand tel que $1/p \leq \varepsilon$, et

$$\mathbf{P}(|Z_{n_p} - Z| > \varepsilon) \leq \mathbf{P}\left(|Z_{n_p} - Z| > \frac{1}{p}\right) \leq \frac{1}{2^p}.$$

D'après le critère de la Proposition 3.10, on a convergence presque-sûre. ▮

EXERCICE 70

- 1) Soit X une variable aléatoire admettant un moment d'ordre 2.
Soit $a > 0$, montrer que

$$\mathbf{E} \left(X^2 \mathbb{1}_{|X| \geq \sqrt{na}} \right) \rightarrow 0 \text{ quand } n \rightarrow +\infty.$$

- 2) Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles indépendantes et de même loi. On suppose que la loi commune possède un moment d'ordre 2 fini. Pour tout $n \geq 1$, on pose

$$M_n = \max_{1 \leq k \leq n} |X_k|.$$

Montrer que la suite $(M_n/\sqrt{n})_{n \geq 1}$ converge en probabilité vers 0.

§ 3.2. Convergence en loi

3.2.1 Définition et premières propriétés

La convergence la plus faible que nous ayons vu jusqu' alors est la convergence en probabilité. Cette convergence implique une certaine convergence des fonctions de répartition.

Commençons par une heuristique pour des variables aléatoires **réelles**. Soit $(Z_n)_{n \geq 0}$ une suite de variables aléatoires réelles convergeant en probabilité vers Z . Soit $x \in \mathbf{R}$, on note F la fonction de répartition de Z et F_n la fonction de répartition de Z_n . On va essayer d'établir, sous l'hypothèse de la convergence en probabilité, une propriété de convergence pour la suite des fonctions de répartitions.

On a pour tout $\varepsilon > 0$,

$$\begin{aligned} F_n(x) = \mathbf{P}(Z_n \leq x) &= \mathbf{P}(Z_n \leq x, Z \leq x + \varepsilon) + \mathbf{P}(Z_n \leq x, Z > x + \varepsilon) \\ &\leq \mathbf{P}(Z \leq x + \varepsilon) + \mathbf{P}(Z - Z_n \geq \varepsilon) \\ &\leq F(x + \varepsilon) + \mathbf{P}(|Z - Z_n| \geq \varepsilon). \end{aligned}$$

Donc $F_n(x) - F(x) \leq F(x + \varepsilon) - F(x) + \mathbf{P}(|Z_n - Z| \geq \varepsilon)$. D'autre part,

$$\begin{aligned} F(x - \varepsilon) = \mathbf{P}(Z \leq x - \varepsilon) &= \mathbf{P}(Z_n \leq x, Z \leq x - \varepsilon) + \mathbf{P}(Z_n > x, Z \leq x - \varepsilon) \\ &\leq \mathbf{P}(Z_n \leq x) + \mathbf{P}(Z_n - Z \geq \varepsilon) \\ &\leq F_n(x) + \mathbf{P}(|Z_n - Z| \geq \varepsilon). \end{aligned}$$

Donc $F(x) - F_n(x) \leq F(x) - F(x - \varepsilon) + \mathbf{P}(|Z_n - Z| \geq \varepsilon)$. Finalement on obtient l'encadrement suivant en combinant les deux précédents :

$$\boxed{F(x - \varepsilon) - F(x) - \mathbf{P}(|Z_n - Z| \geq \varepsilon) \leq F_n(x) - F(x) \leq F(x + \varepsilon) - F(x) + \mathbf{P}(|Z_n - Z| \geq \varepsilon)}.$$

Soit $\eta > 0$, et supposons que x est un point de continuité de F . Alors on peut choisir ε tel que

$$-\frac{\eta}{2} \leq F(x - \varepsilon) - F(x) \leq \frac{\eta}{2}, \quad \text{et} \quad -\frac{\eta}{2} \leq F(x + \varepsilon) - F(x) \leq \frac{\eta}{2}.$$

On a alors

$$-\frac{\eta}{2} - \mathbf{P}(|Z_n - Z| \geq \varepsilon) \leq F_n(x) - F(x) \leq \frac{\eta}{2} + \mathbf{P}(|Z_n - Z| \geq \varepsilon).$$

Pour ce ε , par convergence en probabilité, on peut trouver n_0 tel que pour tout $n \geq n_0$,

$$-\frac{\eta}{2} \leq \mathbf{P}(|Z_n - Z| \geq \varepsilon) \leq \frac{\eta}{2}$$

et donc

$$-\eta \leq F_n(x) - F(x) \leq \eta.$$

Donc

$$\boxed{\lim_{n \rightarrow \infty} F_n(x) = F(x) \text{ pour tout point de continuité } x \text{ de } F.}$$

Ceci nous conduit à la notion de convergence en loi.

DEFINITION – 3.4 Soient $d \in \mathbf{N}^*$ et Z une variable aléatoire à valeurs dans \mathbf{R}^d . On dit qu'une suite $(Z_n)_{n \geq 0}$ de variables

aléatoires de \mathbf{R}^d converge en loi vers Z et on note $Z_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Z$ si

$$\text{pour tout point de continuité } x \text{ de } F_Z \text{ on a : } \lim_{n \rightarrow \infty} F_{Z_n}(x) = F_Z(x).$$

Notez que, pour cette convergence, il n'est pas nécessaire que les variables aléatoires soient définies sur un même espace probabilisé. La définition garde son sens si on considère une suite $(\mathbf{P}_n)_{n \in \mathbf{N}}$ où \mathbf{P}_n est associée à chaque Z_n .

Par contre – cela est nécessaire pour les convergences en probabilité, presque sûre et dans L^p .

REMARQUE – 3.2 (EXIGENCE EN LES POINTS DE CONTINUITÉ, POURQUOI ?) Déjà en les points où F_Z n'est pas continue le raisonnement précédent ne fonctionne pas. Il ne serait alors plus clair que la convergence en probabilité implique la convergence en loi.

Par ailleurs, si on demandait la convergence en tout point de \mathbf{R} , la définition serait trop restrictive. En effet, soit $Z_n \sim \mathcal{U}([0, 1/n])$ pour $n \geq 1$. Alors \mathbf{P} -p.s.

$$0 \leq Z_n \leq \frac{1}{n},$$

donc $(Z_n)_{n \geq 1}$ converge presque-sûrement vers 0 et donc en loi vers 0 (ce sera vu plus tard).

Par contre, la fonction de répartition de la variable aléatoire Z constante égale à 0 est

$$F_Z = \mathbb{1}_{[0, \infty[}$$

qui n'est pas continue en 0 et telle que $F_Z(0) = 1$. On a $F_{Z_n}(0) = 0$ pour tout n et donc $\lim_{n \rightarrow \infty} F_{Z_n}(0) \neq F_Z(0)$.

REMARQUE – 3.3 (Attention : NON UNICITÉ DE LA LIMITE POUR LA CONVERGENCE EN LOI.) Il n'y a pas unicité presque sûre de la limite en loi que l'on soit dans le cas de suites réelles ou vectorielles. Ceci vient du fait que deux variables aléatoires peuvent avoir même loi tout en étant très différentes. En général, si $(Z_n)_{n \in \mathbf{N}}$ converge en loi à la fois vers Z et vers Z' alors

$$Z \text{ et } Z' \text{ sont presque sûrement distinctes } \textit{autrement dit } \mathbf{P}(Z = Z') < 1.$$

Par exemple, supposons que $Z_n = X \sim \mathcal{N}(0, 1)$ pour tout $n \geq 0$.

Alors, comme X et $-X$ ont même loi (par exemple en raison de la parité de la densité de la loi $\mathcal{N}(0, 1)$), la suite $(Z_n)_{n \geq 0}$ converge en loi à la fois vers X et vers $-X$. Or il n'y a pas égalité presque sûre de ces deux limites ici puisqu'en fait $\mathbf{P}(X = -X) = 0$. En effet :

$$\mathbf{P}(X = -X) = \mathbf{P}(2X = 0) = \mathbf{P}(X = 0) = 0 \quad \text{car la loi } \mathcal{N}(0, 1) \text{ est continue.}$$

On dispose de deux autres formulations équivalentes de la notion de convergence en loi donnée en Définition 3.4, qui sont parfois prises en tant que définition.

PROPOSITION – 3.12 (DÉFINITIONS ALTERNATIVES.) Soient $d \in \mathbf{N}^*$ et Z une variable aléatoire à valeurs dans \mathbf{R}^d . Une suite $(Z_n)_{n \geq 0}$ de variables aléatoires de \mathbf{R}^d converge en loi vers Z si et seulement si l'une des deux conditions équivalentes suivantes est vérifiée :

(i) (Fonction test, convergence étroite) pour toute fonction $f : \mathbf{R}^d \rightarrow \mathbf{R}$ continue bornée, on a

$$\lim_{n \rightarrow \infty} \mathbf{E}(f(Z_n)) = \mathbf{E}(f(Z)).$$

(ii) (Convergence simple des fonctions caractéristiques) pour tout $t \in \mathbf{R}^d$,

$$\lim_{n \rightarrow \infty} \phi_{Z_n}(t) = \phi_Z(t).$$

Nous admettrons les équivalences de cette proposition.

REMARQUE – 3.4 Avec (i), on peut définir d'autres modes de convergences, en remplaçant la classe des fonctions continues bornées par par exemple les fonctions \mathcal{C}^∞ à support compact (*convergence vague*), ou encore les fonctions continues qui tendent vers 0 en $+\infty$ (*convergence faible*). On a alors immédiatement :

$$(Z_n)_{n \in \mathbf{N}} \text{ converge étroitement} \implies (Z_n)_{n \in \mathbf{N}} \text{ converge faiblement} \implies (Z_n)_{n \in \mathbf{N}} \text{ converge vaguement.}$$

Notez cependant l'implication "(i) ⇒ (ii)" qui est évidente.

En effet, en utilisant la linéarité de l'espérance, il est clair que si l'assertion (i) est vraie, elle l'est aussi pour toutes les fonctions continues et bornées $f : \mathbf{R}^d \rightarrow \mathbf{C}$ en regardant les parties réelles et imaginaires.

Elle est donc vraie pour notamment pour la famille des fonctions complexes $\{f_t, t \in \mathbf{R}^d\}$ données par

$$\forall t \in \mathbf{R}^d, f_t : z \in \mathbf{R}^d \mapsto e^{i\langle t, z \rangle}.$$

L'assertion (ii) s'ensuit en remarquant que : pour toute variable aléatoire X de \mathbf{R}^d , sa fonction caractéristique ϕ_X est donnée par :

$$\forall t \in \mathbf{R}^d, \phi_X(t) = \mathbf{E}(f_t(X)).$$

D'autre part, l'équivalence entre la Définition 3.4 et l'assertion (ii) semble très naturelle puisqu'on a vu (cf. Chapitre 2) que la loi d'une variable aléatoire réelle est caractérisée par sa fonction de répartition ainsi que par sa fonction caractéristique.

3.2.2 Critère de convergence en loi des suites de variables discrètes

Dans le cas particulier des suites de vecteurs (ou variables) aléatoires **discrètes**, la convergence en loi est aussi caractérisée comme suit via les fonctions de masse.

PROPOSITION – 3.13 Soient $(Z_n)_{n \geq 1}$ et Z des variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ et à valeurs dans \mathbf{R}^d avec $d \geq 1$ entier quelconque.

On suppose que les Z_n et Z sont des variables discrètes possédant un support commun dénombrable D , i.e. pour tout $n \in \mathbf{N}$, $\mathbf{P}(Z_n \in D) = 1$.

Alors :

$$Z_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Z \iff \forall z \in D, \mathbf{P}(Z_n = z) \xrightarrow[n \rightarrow \infty]{} \mathbf{P}(Z = z).$$

L'exercice suivant démontre cette propriété dans le cas particulier où $d = 1$ et $D = \mathbf{N}$.

Notez que cette proposition ne permet pas forcément d'identifier la loi limite. On a besoin de savoir que Z est discrète !

Ce n'est pas surprenant : si par exemple $\mathbf{P}(Z_n = z) \xrightarrow[n \rightarrow \infty]{} 0$ pour tout z dans son support, alors $(Z_n)_{n \in \mathbf{N}}$ convergerait vers toutes les lois continues ?

EXERCICE 71

Soient $(X_n)_{n \geq 0}$ et X des variables aléatoires discrètes à valeurs dans \mathbf{Z} . Montrer que $(X_n)_{n \geq 0}$ converge en loi vers X si et seulement si pour tout $k \in \mathbf{Z}$, $\mathbf{P}(X_n = k)$ tend vers $\mathbf{P}(X = k)$.

3.2.3 Continuité, opérations, projections

PROPOSITION – 3.14 (CONTINUITÉ ET CONVERGENCE EN LOI)

Soit $(Z_n)_{n \geq 0}$ une suite de variables aléatoires toutes à valeurs dans \mathbf{R}^d avec $d \in \mathbf{N}^*$.

Si $Z_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Z$ alors pour toute application $g : \mathbf{R}^d \rightarrow \mathbf{R}^{d'}$ continue, avec $d' \in \mathbf{N}^*$ quelconque, $g(Z_n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} g(Z)$.

EXERCICE 72

Démontrer la proposition.

Dans la suite de cette section, nous énonçons certaines propriétés de la convergence en loi et nous mettons en évidence certaines de ses spécificités.

REMARQUE – 3.5 (CONVERGENCE EN LOI ET OPÉRATIONS.)

La convergence en loi est un mode de convergence stochastique très particulier : elle ne vérifie pas de bonnes propriétés arithmétiques.

D'abord, notons qu'en général $Z_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Z$ **n'implique pas** $Z_n - Z \xrightarrow[n \rightarrow \infty]{\mathcal{L}} 0$.

D'autre part : si (Y_n) et (Z_n) sont deux suites de variables réelles qui convergent en loi vers respectivement Y et Z , on ne peut en général **rien dire** sur la convergence des suites $(Y_n + Z_n)$ et $(Y_n Z_n)$.

L'exercice suivant propose une illustration de cette remarque.

EXERCICE 73

Soit X une variable aléatoire réelle qui suit la loi $\mathcal{U}([-1, 1])$. On définit la suite $(Z_n)_{n \geq 0}$ en posant : $Z_n = X$ pour tout entier n .

- 1) Justifier que la suite $(Z_n)_n$ converge en loi vers la variable $Z = -X$.
- 2) Démontrer que $(Z_n - Z)_n$ ne converge pas en loi vers 0.

Cependant, dans certains cas, on a des résultats positifs. Voici un exemple usuel que nous énonçons pour $d = 2$ connu sous le nom de *Théorème de Slutsky* : c'est le cas particulier où l'une des deux suites converge en loi vers une constante.

PROPOSITION – 3.15 (THÉORÈME DE SLUTSKY— OPÉRATIONS DE CONVERGENCE EN LOI SI L'UNE EST CSTE)
 Soient $(Y_n)_n$ et $(Z_n)_n$ des suites de variables aléatoires réelles telles qu'il existe une constante $c \in \mathbf{R}$ et une variables aléatoires réelle (quelconque) Z telles que

$$Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} c \quad \text{et} \quad Z_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Z \quad \text{lorsque } n \rightarrow \infty.$$

Alors :

- (i) La suite vectorielle (Y_n, Z_n) converge aussi en loi vers le vecteur (c, Z) .
- (ii) ainsi, on a en particulier :
 - (a) $Y_n + Z_n \xrightarrow{\mathcal{L}} c + Z$,
 - (b) $Y_n Z_n \xrightarrow{\mathcal{L}} cZ$.

EXERCICE 74 (DÉMONSTRATION DE LA PROPOSITION 3.15)

1) Pour simplifier, on pose $X = (c, Z)$ et $X_n = (Y_n, Z_n)$ pour tout entier $n \geq 0$.

Soit $t = (u, v)$ un élément quelconque de \mathbf{R}^2 . Le but est de prouver que la suite $\phi_{X_n}(t)$ converge dans \mathbf{C} vers $\phi_X(t)$.

- a. Sur quel domaine de \mathbf{R} , la fonction $\psi : x \in \mathbf{R} \mapsto e^{ix}$ est-elle continue ? Uniformément continue ?
- b. Justifier que l'on a pour tout entier $n \in \mathbf{N}$: pour tout $t = (u, v) \in \mathbf{R}^2$,

$$|\phi_{X_n}(t) - \phi_X(t)| \leq \left| \mathbf{E} \left(e^{ivZ_n} (e^{iuY_n} - e^{iuc}) \right) \right| + |\phi_{Z_n}(v) - \phi_Z(v)|, \tag{3.2}$$

et que pour tout $\varepsilon > 0$,

$$\left| \mathbf{E} \left(e^{ivZ_n} (e^{iuY_n} - e^{iuc}) \right) \right| \leq \mathbf{E} \left(\left| e^{iuY_n} - e^{iuc} \right| \mathbb{1}_{|Y_n - c| < \varepsilon} \right) + 2\mathbf{P}(|Y_n - c| \geq \varepsilon) \tag{3.3}$$

Puis utiliser la question (a) pour prouver que $(Y_n, Z_n) \xrightarrow{\mathbf{P}} (c, Z)$.

2) En déduire les convergences du point (ii) de la Proposition 3.15.

PROPOSITION – 3.16 (LIENS ENTRE CONVERGENCE EN LOI D'UNE SUITE DE VECTEURS ET DE SES COORDONNÉES.)

Soit $(Z_n)_{n \in \mathbf{N}}$ une suite de variables aléatoires à valeurs dans \mathbf{R}^d avec $d \geq 2$ entier.
 Pour tout entier $n \geq 0$, on note $Z_n = (Z_{1,n}, \dots, Z_{d,n})$ ou encore $Z_n = (Z_{j,n} : 1 \leq j \leq d)$.
 On considère aussi $Z = (Z_1, \dots, Z_d) = (Z_j : 1 \leq j \leq d)$ un vecteur aléatoire de \mathbf{R}^d .

- (i) Si la suite (Z_n) converge en loi vers le vecteur Z alors pour tout $1 \leq j \leq d$, la suite de variables aléatoires réelles $(Z_{j,n})$ converge en loi vers Z_j .
- (ii) **Attention** : Par contre, la réciproque est fautive en général.

La propriété (i) ne se généralise pas à n'importe quel vecteur, c'est ce qui est précisé dans la prochaine proposition.

EXERCICE 75

Démontrer l'assertion (i) de la Proposition 3.16.

Nous allons établir le point (ii) de cette proposition via un contre-exemple dans l'exercice suivant. Des commentaires et compléments sur la Proposition 3.16 seront développés ci-dessous dans la Remarque 3.6 et la Proposition 3.15.

EXERCICE 76

Considérons X une variable aléatoire réelle qui suit la loi dite de Rademacher de paramètre $1/2$ autrement dit : X est une variable réelle discrète qui prend les valeurs -1 et 1 avec $\mathbf{P}(X = -1) = \mathbf{P}(X = 1) = 1/2$. On définit les suites (Y_n) et (Z_n) en posant :

$$\forall n \in \mathbf{N}, \quad Y_n = X \quad \text{et} \quad Z_n = -X.$$

- 1) Justifier rapidement que la suite $((Y_n, Z_n))_n$ converge en loi vers le couple $(X, -X)$.
- 2) Montrer que les suites $(Y_n)_n$ et $(Z_n)_n$ convergent toutes les deux en loi vers la variable X .
- 3) Le but de cette question est de montrer que la suite $((Y_n, Z_n))_n$ ne converge pas en loi vers (X, X) .
 - a. Calculer la fonction caractéristique ϕ_X de X puis celle du couple $(X, -X)$.
 - b. Déterminer rapidement la fonction caractéristique de couple (X, X) .
 - c. Conclure en utilisant tout ce qui précède.

►CORRECTION.

Evident pour la première puisque les variables Y_n sont toutes égales à X donc la suite (Y_n) converge dans tous les modes vers X : en particulier $Y_n \xrightarrow{\mathcal{L}} X$.

De même : $Z_n \xrightarrow{\mathcal{L}} -X$. Mais il est très facile de vérifier que $-X$ a même loi que X (autrement dit : la loi de Rademacher de paramètre $1/2$ est symétrique). En effet comme X prend seulement les valeurs ± 1 , il en est de même pour $-X$. De plus :

$$\mathbf{P}(-X = -1) = \mathbf{P}(X = 1) = 1/2 \quad \text{donc} \quad \mathbf{P}(-X = -1) = \mathbf{P}(X = -1).$$

Donc, par définition de la convergence en loi, on a aussi : $Z_n \xrightarrow{\mathcal{L}} X$.

Pour tout entier $n \in \mathbf{N}$, on a : $(Y_n, Z_n) = (X, -X)$. Donc cette suite converge en loi vers le couple $(X, -X)$.

Montrons que la suite (Y_n, Z_n) ne converge pas en loi vers le couple (X, X) . Pour cela, il suffit par exemple de trouver un vecteur $t \in \mathbf{R}^2$ tel que la suite $(\phi_{(Y_n, Z_n)}(t))$ ne converge pas vers $\phi_{(X, X)}(t)$:

- Soit $t = (u, v)$ un élément quelconque de \mathbf{R}^2 . On a pour tout entier $n \in \mathbf{N}$,

$$\phi_{(Y_n, Z_n)}(t) = \phi_{(X, -X)}(t) = \mathbf{E} \left(e^{iuX + iv(-X)} \right) = \mathbf{E} \left(e^{i(u-v)X} \right) = \frac{1}{2} \left(e^{i(u-v)} + e^{-i(u-v)} \right).$$

On a donc pour tout entier $n \in \mathbf{N}$,

$$\forall t = (u, v) \in \mathbf{R}^2, \quad \phi_{(Y_n, Z_n)}(t) = \phi_{(X, -X)}(t) = \cos(u - v).$$

- De même pour tout $t = (u, v) \in \mathbf{R}^2$,

$$\phi_{(X, X)}(t) = \mathbf{E} \left(e^{iuX + ivX} \right) = \mathbf{E} \left(e^{i(u+v)X} \right) = \frac{1}{2} \left(e^{i(u+v)} + e^{-i(u+v)} \right) = \cos(u + v).$$

On peut alors remarquer que $\phi_{(X, -X)}(\pi/2, \pi/2) = \cos(0) = 1$ alors que $\phi_{(X, X)}(\pi/2, \pi/2) = \cos(\pi) = -1$.

Donc les fonctions caractéristiques des vecteurs $(X, -X)$ et (X, X) sont distinctes et que l'on a vu que la suite (Y_n, Z_n) converge vers le couple $(X, -X)$, elle ne peut pas converger vers le couple (X, X) !

REMARQUE – 3.6 (COMPLÉMENTS.) La Proposition 3.16 est en fait très naturelle si l'on se souvient (Voir Chapitre 2) du fait que la connaissance de la loi d'un vecteur aléatoire nous donne les lois de ses coordonnées (lois marginales). Mais, en général, la connaissance des lois marginales ne suffit pas pour obtenir complètement la loi du vecteur associé à moins qu'elles soient indépendantes.

3.2.4 Liens entre la convergence en loi et les autres convergences aléatoires

Voici les liens de la convergence en loi (dans le cadre réel et vectoriel) avec les autres modes de convergences aléatoires : c'est le plus faible des convergences stochastiques.

PROPOSITION – 3.17 Soient $(Z_n)_{n \geq 0}$ et Z des variables aléatoires définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ et à valeurs dans \mathbf{R}^d avec $d \geq 1$ entier quelconque.

- (i) $Z_n \xrightarrow{\mathbf{P}}_{n \rightarrow \infty} Z \implies Z_n \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} Z,$
- (ii) soit $p \geq 1$, $Z_n \xrightarrow{L^p}_{n \rightarrow \infty} Z \implies Z_n \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} Z,$
- (iii) $Z_n \xrightarrow{p.s.}_{n \rightarrow \infty} Z \implies Z_n \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} Z.$

▮ **PREUVE.** Le point (i) a été démontré pour $d = 1$ dans la Section 3.2.1 (juste avant la Définition 3.4). Le cas vectoriel où $d \geq 2$ est laissé en exercice.

Les points (ii) et (iii) s'en déduisent alors directement puisqu'on a vu que les convergences presque-sûre et dans L^p impliquent la convergence en probabilité. ▮

⚠ **Attention.** En général, la réciproque de la Proposition 3.17 est fautive !

Contre-Exemple : On considère la suite (Z_n) de variables aléatoires réelles telle que pour tout entier n , on ait

$$Z_n = (-1)^n X,$$

où X est une variable aléatoire réelle qui suit la loi normale centrée réduite $\mathcal{N}(0, 1)$.

On peut remarquer que, comme la loi $\mathcal{N}(0, 1)$ est symétrique - car par exemple la densité de la loi $\mathcal{N}(0, 1)$ est paire - (i.e. que $-X$ et X sont de même loi), on a pour tout entier $n \geq 0$, $Z_n \sim \mathcal{N}(0, 1)$. Donc forcément la suite Z_n converge en loi vers la variable X ainsi que $-X$.

S'il y avait convergence en probabilité, alors la suite (Z_n) convergerait en probabilité à la fois vers X et vers $-X$. Par unicité de la limite en probabilité, ceci impliquerait $\mathbf{P}(X = -X) = 1$. Absurde ici car cette probabilité vaut ici 0 :

$$\mathbf{P}(X = -X) = \mathbf{P}(2X = 0) = \mathbf{P}(X = 0) = 0 \quad \text{puisque la loi } \mathcal{N}(0, 1) \text{ est continue.}$$

La convergence en probabilité est donc en général plus forte que la convergence en loi.

Cependant, ces 2 modes de convergences sont équivalents dans le cas particulier où la limite est constante presque-sûrement. En effet, on a :

PROPOSITION - 3.18 Soient $d \in \mathbf{N}^*$ et $(Z_n)_n$ une suite de variables définies sur $(\Omega, \mathcal{A}, \mathbf{P})$, à valeurs dans \mathbf{R}^d . Soit $a \in \mathbf{R}^d$. Alors on a lorsque $n \rightarrow \infty$:

$$Z_n \xrightarrow{\mathbf{P}}_{n \rightarrow \infty} a \iff Z_n \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} a.$$

▮ **PREUVE.** La condition nécessaire est toujours vraie comme déjà signalé (cf. Proposition 3.17).

Démontrons maintenant la réciproque dans le cas $d = 1$ via l'Exercice suivant.

EXERCICE 77

On considère $(Z_n)_{n \in \mathbf{N}}$ une suite de variables aléatoires réelles qui converge en loi vers une constante $a \in \mathbf{R}$. Le but de cet exercice est de démontrer la convergence en probabilité.

- 1) Calculer la fonction de répartition notée F_a de la constante a et donner son ensemble de continuité.
- 2) Soit $\varepsilon > 0$. Démontrer que l'on a :

$$0 \leq \mathbf{P}[|Z_n - a| \geq \varepsilon] \leq F_{Z_n}(a - \varepsilon) + 1 - F_{Z_n}(a + \varepsilon/2).$$

Puis conclure.

► CORRECTION.

- 1) On suppose que $Z_n \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} a$ ce qui se traduit par exemple par :

$$F_{Z_n}(t) \rightarrow F_a(t) \quad \text{en tout point } t \text{ de continuité de la fonction de répartition } F_a.$$

Calculons F_a . On a par définition $F_a(t) := \mathbf{P}(a \leq t)$ pour tout $t \in \mathbf{R}$.

Or pour tout réel t , comme a est un réel fixé, on a :

$$\{a \leq t\} = \begin{cases} \emptyset & \text{si } t < a, \\ \Omega & \text{si } t \geq a. \end{cases}$$

Par conséquent $F_a(t) = \mathbb{1}_{[a, +\infty[}(t)$ qui est une fonction continue sur $\mathbf{R} \setminus \{a\}$. Ainsi montrer " $Z_n \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} a$ " équivaut à montrer : lorsque $n \rightarrow +\infty$,

$$F_{Z_n}(t) := \mathbf{P}(Z_n \leq t) \rightarrow \begin{cases} 0 & \text{si } t < a \\ 1 & \text{si } t > a. \end{cases} \quad (3.4)$$

- 2) Le but est de démontrer que $Z_n \xrightarrow{\mathbf{P}} a$ i.e. que :

$$\forall \varepsilon > 0, \quad \mathbf{P}[|Z_n - a| \geq \varepsilon] \rightarrow 0 \quad \text{lorsque } n \rightarrow +\infty. \quad (3.5)$$

Fixons $\varepsilon > 0$. On a : $\{|Z_n - a| \geq \varepsilon\} = \{Z_n \leq a - \varepsilon\} \cup \{Z_n \geq a + \varepsilon\}$ qui est une réunion disjointe. Par suite :

$$\begin{aligned} \mathbf{P}[|Z_n - a| \geq \varepsilon] &= \mathbf{P}(Z_n \leq a - \varepsilon) + \mathbf{P}(Z_n \geq a + \varepsilon) \\ &\leq \mathbf{P}(Z_n \leq a - \varepsilon) + \mathbf{P}(Z_n > a + \varepsilon/2). \end{aligned}$$

On a donc pour tout $n \in \mathbf{N}$,

$$0 \leq \mathbf{P}[|Z_n - a| \geq \varepsilon] \leq F_{Z_n}(a - \varepsilon) + 1 - F_{Z_n}(a + \varepsilon/2). \quad (3.6)$$

Puisque $a - \varepsilon < a$ et que $a + \varepsilon/2 > a$, on a par (3.4) : $F_{Z_n}(a - \varepsilon) \rightarrow 0$ et $F_{Z_n}(a + \varepsilon/2) \rightarrow 1$ lorsque $n \rightarrow +\infty$.
 Par conséquent, $F_{Z_n}(a - \varepsilon) + 1 - F_{Z_n}(a + \varepsilon/2) \rightarrow 0$ lorsque $n \rightarrow +\infty$.
 Ceci implique par (3.6) que : $\mathbf{P}[|Z_n - a| \geq \varepsilon] \rightarrow 0$ lorsque $n \rightarrow +\infty$.

EXERCICE 78

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et de même loi uniforme sur $[0, a]$ avec $a > 0$. On pose, pour $n \geq 1$,

$$U_n = \max(X_1, \dots, X_n) \quad \text{et} \quad V_n = n(U_n - a).$$

- 1) En calculant, pour $\varepsilon > 0$ et $n \geq 1$, $\mathbf{P}(|U_n - a| > \varepsilon)$, montrer que $U_n \rightarrow a$ **P**-p.s..
- 2) Déterminer la fonction de répartition de V_n . En déduire que $(V_n)_{n \geq 1}$ converge en loi vers une variable aléatoire V dont on précisera la loi.

EXERCICE 79

On considère ici $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles et on pose pour tout entier $n \geq 1$, $Z_n = X_n/n$.

- 1) On suppose que pour tout $n \geq 1$, X_n suit la loi uniforme discrète sur $\{1, \dots, n\}$. Montrer que la suite $(Z_n)_{n \geq 1}$ converge en loi vers la loi $\mathcal{U}([0, 1])$.
- 2) Soit $\lambda > 0$ un réel fixé. Ici, on suppose que pour tout $n \geq 1$, X_n suit la loi géométrique $\mathcal{G}(\lambda/n)$. Montrer que la suite $(Z_n)_{n \geq 1}$ converge en loi vers la loi $\mathcal{E}(\lambda)$. La suite aléatoire $(X_n)_{n \in \mathbf{N}}$ converge-t-elle en loi ?

EXERCICE 80

Soit (X_n) une suite de variables aléatoires indépendantes de loi exponentielle $\mathcal{E}(1/n)$ et soit $Y_n = X_n - \lfloor X_n \rfloor$ où $\lfloor X_n \rfloor$ désigne la partie entière de X_n .

- 1) Quel est l'ensemble des valeurs prises par la variable aléatoire Y_n ?
- 2) Montrer que (Y_n) converge en loi vers Y dont on précisera la loi.

EXERCICE 81

- 1) Soit (X_n) une suite de variables aléatoires réelles indépendantes de loi binomiale $X_n \sim \mathcal{B}\left(n, \frac{\lambda}{n}\right)$ avec $0 < \lambda < 1$.
 Montrer que X_n converge en loi vers une variable aléatoire dont on déterminera la loi.

- 2) On tire 100 fois avec remise une boule dans une urne contenant 99 boules blanches et une boule rouge. En utilisant la question 1), déterminer une valeur approchée de la probabilité d'obtenir au moins 2 boules rouges sur ces 100 tirages.

EXERCICE 82

Soit $(\varepsilon_n)_{n \in \mathbf{N}}$ une suite iid de variables aléatoires suivant une loi de Rademacher de paramètre $p = 1/2$, $\mathcal{R}(1/2)$.

Soit $(a_n)_{n \in \mathbf{N}}$ une suite de nombres réels. Pour tout entier $n \in \mathbf{N}^*$, on définit $S_n = \sum_{k=1}^n a_k \varepsilon_k$.

- 1) Soient $n \in \mathbf{N}^*$ et $t > 0$ fixés. Montrer que pour tout $h > 0$,

$$\mathbf{P}(S_n > t) = \mathbf{P}(\exp(hS_n) > \exp(ht)) \leq e^{-ht} \mathbf{E}(\exp(hS_n)).$$

- 2) En déduire que

$$\mathbf{P}(|S_n| > t) \leq 2 \exp\left(-\frac{t^2}{2 \sum_{k=1}^n a_k^2}\right).$$

- 3) Dans cette question on suppose que pour tout $k \in \mathbf{N}^*$, on a $a_k = \frac{1}{2^k}$.

Étudier la convergence en loi de la suite $(S_n)_{n \in \mathbf{N}}$.

- 4) Dans cette question on suppose que pour tout $k \in \mathbf{N}^*$, on a $a_k = \frac{\sqrt{k}}{\ln(k+1)}$.

Étudier la convergence presque-sûre de la suite $(S_n/n)_{n \in \mathbf{N}}$.

§ 3.3. Bilan des liens entre les différentes convergences stochastiques

On considère une suite $(Z_n)_{n \in \mathbf{N}}$ de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ et à valeurs dans \mathbf{R}^d avec $d \in \mathbf{N}^*$. Soit $Z : \Omega \rightarrow \mathbf{R}^d$ une variable aléatoire.

D'après les sections précédentes, on a les liens suivants entre les différents types de convergence stochastique : pour le diagramme bilan, **voir cours**.

§ 3.4. Théorèmes Limites

Il existe trois résultats fondamentaux sur la convergence de sommes (ou de moyennes) de variables aléatoires indépendantes et de même loi (sous certaines conditions d'intégrabilité) : à savoir *les Loïs dite faible et forte des grands nombres* (cf. les Sections 3.4.1 et 3.4.2) et *le Théorème dit Limite Central* (cf. Section 3.4.3).

Nous énonçons et démontrons ces théorèmes dans le cas de suites de variables aléatoires **réelles**. Puis nous évoquerons le cas des suites de vecteurs aléatoires (à valeurs dans un certain \mathbf{R}^d avec $d \geq 2$ entier) dans la Section 3.4.4.

Voici une définition usuelle de la moyenne de variables qui sera très utilisée dans le cadre des applications et notamment en Statistique (cf. la seconde partie du cours sur l'introduction à la estimation paramétrique ponctuelle).

DEFINITION – 3.5 Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles. Pour tout entier $n \geq 1$, on pose $S_n = \sum_{i=1}^n X_i$. On appelle *moyenne empirique des n premiers X_i* ou simplement *moyenne empirique des X_i* , la variable notée \bar{X}_n et définie par :

$$\bar{X}_n := \frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n}.$$

3.4.1 Loi faible des grands nombres

On rappelle que "dire que des variables aléatoires sont **i.i.d.**" signifie que ces variables sont indépendantes (sous-entendu (mutuellement)) et **identiquement distribuées** autrement dit qu'elles sont indépendantes et de même loi de probabilité.

THÉORÈME – 3.1 (LOI FAIBLE DES GRANDS NOMBRES.) Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles i.i.d. et définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$. On suppose qu'elles sont intégrables i.e. que : $\mathbf{E}(|X_1|) < \infty$.

Posons, pour tout $n \in \mathbf{N}^*$: $S_n = \sum_{i=1}^n X_i$.

Alors la suite $\bar{X}_n = \frac{S_n}{n}$ converge en probabilité vers l'espérance $\mathbf{E}(X_1)$ lorsque $n \rightarrow \infty$.

┌ PREUVE. Voir cours pour la démonstration détaillée. ─

REMARQUE – 3.7 Ce résultat a été établi au début de ce Chapitre (cf. Section 3.1.1) dans le cas où la loi commune des X_i est une loi de Bernoulli de paramètre p . Notez que dans ce cas particulier, nous avons proposé une autre démonstration basée sur le fait que ces X_i sont en fait aussi de carré intégrable i.e. dans L^2 .

3.4.2 Loi forte des grands nombres

En fait on a une convergence plus forte - à savoir presque sûre - sous les mêmes hypothèses.

THÉORÈME – 3.2 (LOI FORTE DES GRANDS NOMBRES.) Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$. On suppose que ces variables sont i.i.d. et intégrables (i.e. $\mathbf{E}(|X_1|) < \infty$).

Alors la suite $\bar{X}_n = \frac{S_n}{n}$ converge presque sûrement vers $\mathbf{E}(X_1)$ lorsque $n \rightarrow \infty$.

REMARQUE – 3.8 Notez que dans le cas particulier où la loi commune des X_i est une loi de Bernoulli de paramètre p , ce résultat a été établi (juste après le Corollaire 3.1) en utilisant le fait que ces X_i sont en fait aussi dans L^4 .

REMARQUE – 3.9 (LIEN ENTRE LOIS FAIBLE ET FORTE DES GRANDS NOMBRES) La loi forte des grands nombres implique la Loi faible puisque la convergence presque sûre implique la convergence en probabilité !

REMARQUE – 3.10 (COMPLÉMENTS) Pour être tout à fait complet, notez que l'on peut en fait démontrer que la condition "X₁ intégrable" est une condition nécessaire mais aussi suffisante à la "convergence presque sûre de la suite $\bar{X}_n = S_n/n$ ".

La démonstration du Théorème 3.2 est plus complexe que celle du Théorème 3.1. Il faut procéder en 2 étapes : l'idée est d'abord d'établir le résultat dans le cas où les variables sont de plus dans L^4 . Ce point est détaillé ci-dessous. Le cas général est admis : sa démonstration consiste à utiliser le fait qu'une variable de L^1 peut être approximée par une variable de L^4 .

▮ PREUVE. Preuve du Théorème 3.2 dans le cas L^4 .

• Comme pour la loi faible des grands nombres, on voit facilement qu'il suffit de faire la démonstration dans le cas de variables centrées i.e. que $\mu := \mathbf{E}(X_1) = 0$ (en effet : dans le cas général où $\mu \neq 0$, on centre les variables aléatoires X_i i.e. que l'on considère les variables $Y_i := X_i - \mu$ qui sont centrées (par linéarité de l'espérance). Donc la suite $S'_n := \sum_{i=1}^n Y_i$ sera telle que : S'_n/n converge presque sûrement vers $\mathbf{E}(Y_1) = 0$ (si l'on admet le cas $\mu = 0$). Or comme on a trivialement : $S_n/n = S'_n/n + \mu$, ceci implique que S_n/n converge presque sûrement vers $0 + \mu = \mu$!

• On suppose donc ici que les X_i sont centrées et appartiennent de plus à l'espace L^4 . Soit $\varepsilon > 0$. Par l'inégalité de Markov, on a :

$$\text{pour tout } n \geq 1, \quad \mathbf{P}(|S_n| \geq \varepsilon n) = \mathbf{P}(|S_n|^4 \geq \varepsilon^4 n^4) \leq \mathbf{E}(S_n^4) / (\varepsilon^4 n^4). \quad (3.7)$$

Or on a :

$$S_n^4 = \sum_{1 \leq i \leq n} X_i^4 + 4 \sum_{1 \leq i \neq j \leq n} X_i^3 X_j + 6 \sum_{1 \leq i \neq j \leq n} X_i^2 X_j^2 + 12 \sum_{1 \leq i \neq j \neq k \leq n} X_i X_j X_k^2 + 24 \sum_{1 \leq i \neq j \neq k \neq l \leq n} X_i X_j X_k X_l.$$

En utilisant la linéarité de l'espérance et le fait que les variables aléatoires $(X_i)_{i \geq 1}$ sont indépendantes et centrées, on déduit :

$$\mathbf{E}(S_n^4) = \sum_{1 \leq i \leq n} \mathbf{E}(X_i^4) + 6 \sum_{1 \leq i \neq j \leq n} \mathbf{E}(X_i^2) \mathbf{E}(X_j^2).$$

Comme les variables X_i sont toutes de même loi, elles ont en particulier les mêmes moments et donc :

$$\mathbf{E}(S_n^4) = \sum_{1 \leq i \leq n} \mathbf{E}(X_1^4) + 6 \sum_{1 \leq i \neq j \leq n} \mathbf{E}(X_1^2)^2 = n \mathbf{E}(X_1^4) + 3n(n-1) \mathbf{E}(X_1^2)^2.$$

Or comme X_1 est dans L^4 , elle admet un moment d'ordre 4 et a fortiori d'ordre 2. Donc la quantité C défini par $C := \max(\mathbf{E}(X_1^4), \mathbf{E}(X_1^2)^2)$ est dans $]0, \infty[$. D'où :

$$\mathbf{E}(S_n^4) = n \mathbf{E}(X_1^4) + 3n(n-1) \mathbf{E}(X_1^2)^2 \leq (3n^2 - 2n)C \leq 3n^2 C.$$

Donc l'inégalité (3.7) donne :

$$0 \leq \sum_{n \geq 1} \mathbf{P}(|S_n| \geq \varepsilon n) \leq \frac{3C}{\varepsilon^4} \leq \sum_{n \geq 1} \frac{1}{n^2}. \quad (3.8)$$

Ainsi, par comparaison, on déduit que pour tout $\varepsilon > 0$, la série $\sum_{n \geq 1} \mathbf{P}(|S_n/n| \geq \varepsilon)$ converge. ceci implique par le Lemme de Borel-Cantelli que la suite S_n/n converge presque sûrement vers 0. ▮

EXERCICE 83

1) Montrer que pour toute suite de variable aléatoire $(X_n)_{n \in \mathbf{N}^*}$, on a

$$\frac{S_n}{n} \text{ converge presque-sûrement} \implies \frac{X_n}{n} \xrightarrow[n \rightarrow \infty]{p.s.} 0.$$

2) Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de variables aléatoires indépendantes de loi

$$\mathbf{P}(X_n = n) = \mathbf{P}(X_n = -n) = \frac{1}{2(n+1)(\ln(n+1))},$$

$$\mathbf{P}(X_n = 0) = 1 - \frac{1}{(n+1)(\ln(n+1))},$$

a) Justifier que l'on ne peut pas ici appliquer les lois faible et Forte des grands nombres à cette suite.

b) Montrer que les variables X_n ($n \in \mathbf{N}^*$) sont toutes centrées.

c) Démontrer que la suite $(\frac{X_n}{n})$ ne converge pas presque sûrement vers 0 lorsque $n \rightarrow \infty$.

Expliquer pourquoi ceci implique que la suite S_n/n ne converge pas presque sûrement, ni en probabilité vers 0.

3.4.3 Le Théorème Central Limite

Une autre façon d'énoncer la Loi forte des grands nombres est de dire que si $\mathbf{E}(|X_1|) < \infty$, alors $S_n/n = \mathbf{E}(X_1) + o(1)$ **P**-p.s. lorsque $n \rightarrow \infty$. En un certain sens, le Théorème Limite Central donne un terme de plus dans le développement asymptotique de S_n/n , précisant la vitesse de convergence de S_n/n vers $\mathbf{E}(X_1)$ i.e. le comportement limite en loi du terme d'erreur $o(1)$ (**modulo** une hypothèse supplémentaire sur la loi des X_i !). Il permet d'approximer la loi de S_n/n lorsque n est grand.

THÉORÈME – 3.3 (THÉORÈME CENTRAL LIMITE OU DE LA LIMITE CENTRALE (T.C.L. OU T.L.C.))

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles qui sont i.i.d. On suppose de plus que les variables X_n sont de carrés intégrables i.e. que : $\mathbf{E}[X_1^2] < \infty$.

On note $\mu = \mathbf{E}[X_1]$ leur espérance et $\sigma^2 = \mathbf{V}(X_1)$ leur variance. On suppose que $\sigma > 0$.

On pose pour tout entier $n \in \mathbf{N}^*$, $\bar{X}_n := \frac{1}{n} \sum_{k=1}^n X_k$.

Alors on a :

$$\sqrt{n} \cdot \frac{\bar{X}_n - \mu}{\sigma} = \frac{\sum_{k=1}^n X_k - n\mu}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Z \quad \text{avec } Z \sim \mathcal{N}(0, 1). \quad (3.9)$$

REMARQUE – 3.11 (COMMENTAIRES ET COMPLÉMENTS.)

1) Si X_1 est de carré intégrable avec $\mathbf{V}(X_1) > 0$, on pose $Z_n := \sqrt{n}(\bar{X}_n - \mathbf{E}(X_1)) / \sqrt{\mathbf{V}(X_1)}$ pour tout entier $n \geq 1$.

On vérifie facilement que les variables Z_n sont toutes centrées et réduites.

Le fait remarquable sur le résultat du TLC est que la loi limite de la suite (Z_n) ne dépend pas de loi particulière (commune) des X_i . En outre, ceci illustre qu'une certaine régularité apparaît dans les phénomènes aléatoires : on dit que *la loi normale a un caractère universel* dans le sens où le résultat précédent ne dépend pas de la loi de départ des X_i mais juste de leurs espérance et variance (sous réserve de leurs existences!).

2) Notez d'autre part que la convergence (3.9) équivaut à dire que l'on a :

$$\text{pour tous } a < b \text{ dans } \bar{\mathbf{R}}, \quad \mathbf{P}\left(\sqrt{n} \cdot \frac{\bar{X}_n - \mu}{\sigma} \in]a, b]\right) \rightarrow \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx. \quad (3.10)$$

Cette reformulation est utile pour les applications concrètes et notamment en Statistique : elle sert par exemple à construire ce que l'on appelle *des intervalles de confiance* (ou fourchettes) (ce point sera abordé - si le temps le permet - dans la partie "Introduction à l'estimation paramétrique").

EXERCICE 84

Justifier l'équivalence entre les assertions (3.9) et (3.10).

REMARQUE – 3.12 (REFORMULATION DU TCL. COMMENTAIRES SUR LE CAS $\sigma = 0$.)

La convergence (3.9) est équivalente à la convergence en loi suivante :

$$\sqrt{n} \cdot (\bar{X}_n - \mu) = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Y \quad \text{avec } Y \sim \mathcal{N}(0, \sigma^2). \quad (3.11)$$

Cette équivalence résulte des propriétés des lois normales : dans le Chapitre 4 sur "les Variables et Vecteurs aléatoires gaussiens", nous verrons que si $Z \sim \mathcal{N}(0, 1)$ alors la variable $Y = \sigma Z$ suit la loi $\mathcal{N}(0, \sigma^2)$ qui admet pour fonction caractéristique $\phi_Y(t) = e^{-t^2 \sigma^2 / 2}$ pour tout $t \in \mathbf{R}$.

En outre, on peut noter que la convergence (3.11) est en fait vraie aussi dans le cas particulier où $\sigma = 0$ (donc pour tout $\sigma \geq 0$).

REMARQUE – 3.13 L'indépendance des variables est **primordiale dans le résultat du T.C.L.**

On peut s'en convaincre en considérant par exemple la suite $(X_n)_{n \in \mathbf{N}}$ constante telle que pour tout n , $X_n = X$ où X est une variable aléatoire réelle qui suit une loi de Rademacher de paramètre 1/2. Autrement dit : X prend les valeurs -1 et 1 avec même probabilité 1/2.

En effet : dans ce cas, on a $\mathbf{E}(X) = 0$, $\mathbf{V}(X) = 1$ et

$$\text{pour tout entier } n \geq 0, \quad Z_n := \sqrt{n} \times \frac{\bar{X}_n - \mathbf{E}(X)}{\sqrt{\mathbf{V}(X)}} = \sqrt{n}X.$$

Pour tout entier $n \geq 0$, la variable $Z_n = \sqrt{n}X$ est clairement discrète :

$$Z_n \text{ est à support dans } \{-\sqrt{n}, \sqrt{n}\} \quad \text{avec } \mathbf{P}(Z_n = \pm\sqrt{n}) = 1/2.$$

Donc on a pour tout entier $n \geq 2$, $\mathbf{P}(Z_n \in]-\sqrt{2}, \sqrt{2}[) = 0$. Ceci implique par exemple

$$\forall n \geq 2, \mathbf{P}(Z_n \in]-1, 1]) = 0 \quad \text{donc évidemment} \quad \mathbf{P}(Z_n \in]-1, 1]) \xrightarrow{n \rightarrow \infty} 0.$$

Or si $Z \sim \mathcal{N}(0, 1)$,

$$\mathbf{P}(Z \in]-1, 1]) = \frac{1}{\sqrt{2\pi}} \int_{-1}^1 e^{-x^2/2} dx > 0.$$

On ne peut donc pas avoir convergence de la suite (Z_n) vers une loi $\mathcal{N}(0, 1)$ car, sinon, on aurait en particulier :

$$\lim_{n \rightarrow \infty} \mathbf{P}(Z_n \in]-1, 1]) = \mathbf{P}(Z \in]-1, 1])$$

mais c'est impossible puisqu'on vient de montrer que la limite ci-dessus est nulle tandis que $\mathbf{P}(Z \in]-1, 1]) > 0$!

▮ PREUVE. **Preuve du TCL : VOIR COURS.** ▮

3.4.4 Généralisation des Théorèmes Limites dans le cas de suites de vecteurs aléatoires

Les Loix faible et forte des grands nombres données dans les Théorèmes 3.1 et 3.2 se généralisent sans difficulté au cas vectoriel.

THÉORÈME – 3.4 (LOIS FAIBLE ET FORTE DES GRANDS NOMBRES : CAS RÉEL ET VECTORIEL.) Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires i.i.d., définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ et à valeurs dans \mathbf{R}^d avec $d \in \mathbf{N}^*$. Si $X_1 = (X_{j,1} : 1 \leq j \leq d)$ est intégrable i.e. si :

$$\mathbf{E}(|X_1|) < \infty \quad \text{où } |X_1| := \sqrt{\sum_{j=1}^d X_{j,1}^2}$$

alors lorsque $n \rightarrow \infty$, la suite $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ converge en probabilité et \mathbf{P} -presque sûrement vers le vecteur $\mathbf{E}(X_1) = (\mathbf{E}(X_{1,1}), \dots, \mathbf{E}(X_{d,1})) \in \mathbf{R}^d$.

Le cas vectoriel $d \geq 2$ est une conséquence directe du cas réel ($d = 1$) puisqu'on a vu que la convergence en probabilité (resp. presque sûre) d'une suite de vecteurs de \mathbf{R}^d équivaut à celle des suites des d coordonnées (Cf. les Propositions 3.3 et 3.6).

Il existe aussi une "version vectorielle" du T.C.L. qui s'énonce comme suit.

THÉORÈME – 3.5 (THÉORÈME LIMITE CENTRAL : CAS RÉEL ET VECTORIEL.) Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires i.i.d., définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ et à valeurs dans \mathbf{R}^d avec $d \in \mathbf{N}^*$.

On suppose que $X_1 = (X_{j,1} : 1 \leq j \leq d)$ est dans L^2 autrement dit de carré intégrable i.e. que $\mathbf{E}(|X_1|^2) < \infty$.

On note :

- $m = \mathbf{E}(X_1)$ l'espérance de X_1 (c'est le vecteur de \mathbf{R}^d donné par $m = (\mathbf{E}(X_{i,1}) : 1 \leq i \leq d)$),
- $\Gamma = \mathbf{V}(X_1)$ la matrice de variance-covariance de X_1 (c'est la matrice carrée symétrique d'ordre d donnée par $\Gamma = (\text{cov}(X_{i,1}, X_{j,1}))_{1 \leq i, j \leq d}$).

Alors la suite de variables aléatoires

$$\sqrt{n} \left(\frac{1}{n} \sum_{k=1}^n X_k - m \right)$$

converge en loi vers un vecteur aléatoire Y à valeurs dans \mathbf{R}^d ayant pour fonction caractéristique

$$\forall t \in \mathbf{R}^d, \quad \phi_Y(t) = e^{-\frac{1}{2}t' \Gamma t}.$$

Notez que l'on dit que ce vecteur Y est un vecteur gaussien de \mathbf{R}^d (ou que Y suit la loi normale sur \mathbf{R}^d) d'espérance $0_{\mathbf{R}^d}$ et de matrice de covariance Γ . Cette loi est usuellement notée $\mathcal{N}_d(0_{\mathbf{R}^d}, \Gamma)$ (Voir Chapitre 4).

La démonstration du Théorème 3.5 sera détaillée dans le Chapitre 4 sur "les Variables et Vecteurs aléatoires Gaussiens" (cf. le Théorème 4.3 et sa preuve).

Notez que cette version vectorielle du TCL n'est pas une conséquence directe de la version "réelle". En effet, comme nous l'avons signalé notamment dans la Proposition 3.16, la convergence en loi des suites de chacune des coordonnées (qui est une conséquence du TLC uni-dimensionnel (Théorème 3.3), dès lors que ces suites ont été correctement renormalisées) ne suffit pas, en général, pour assurer celle de la suite de vecteurs.

EXERCICE 85

Soit (X_n) une suite de variables aléatoires réelles indépendantes de loi de Poisson $X_n \sim \mathcal{P}(a_n)$ avec $a_n > 0$. On note $s_n = a_1 + \dots + a_n$ et $S_n = X_1 + \dots + X_n$. On suppose que $\lim_{n \rightarrow +\infty} s_n = +\infty$. Montrer que, lorsque $n \rightarrow +\infty$,

$$\frac{S_n - s_n}{\sqrt{s_n}} \rightarrow \mathcal{N}(0, 1) \quad \text{en loi.}$$

EXERCICE 86

Soit $p \in]0, 1[$. On considère $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles indépendantes et de même loi de Rademacher de paramètre p , notée $\mathcal{R}(p)$.

Rappel : On dit qu'une variable aléatoire réelle X suit la loi de Rademacher de paramètre p et on note $X \sim \mathcal{R}(p)$ si :

$$X \text{ est discrète et prend les valeurs } -1 \text{ et } 1 \text{ avec } \mathbf{P}(X = 1) = 1 - \mathbf{P}(X = -1) = p.$$

Pour tout entier $n \geq 1$, on pose : $Y_n = X_n X_{n+1}$ et $S_n = Y_1 + \dots + Y_n$.

1) Soit $n \in \mathbb{N}^*$. Montrer que Y_n suit une loi de Rademacher dont on déterminera le paramètre π .

2) Montrer la convergence presque sûre suivante : $\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{\mathbf{P}\text{-p.s.}} (2p - 1)^2$.

3) Donner le théorème limite centrale pour chacune des suites (P_n) et (Q_n) définies par

$$\text{pour tout entier } n \geq 1, \quad P_n = \sum_{k=1}^n Y_{2k} \quad \text{et} \quad Q_n = \sum_{k=1}^n Y_{2k-1}.$$

4) Est-il possible d'en déduire un théorème limite centrale associé à la suite (S_n) ? Expliquer votre réponse.

4 VARIABLES ET VECTEURS ALÉATOIRES GAUSSIENS

Le TCL justifie l'importance des lois normales (et en particulier de la loi normale centrée réduite $\mathcal{N}(0, 1)$). En effet, on approxime (en loi) - sous de bonnes conditions - de nombreuses variables (resp. vecteurs) aléatoires par des variables aléatoires réelles normales (resp. vecteurs gaussiens). Il est donc important d'étudier ces variables et vecteurs en détails.

Dans tout ce chapitre, les variables et vecteurs aléatoires considérés sont tous définis sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$.

Notations. Pour simplifier, nous noterons

- **v.a.r.** (ou parfois simplement **v.a.**) pour **variable aléatoire réelle**.
- **i.e.** pour abrégé *c'est-à-dire* (puisque *i.e.* est l'abréviation de l'expression latine *id est* qui signifie *c'est-à-dire*).

§ 4.1. Variables aléatoires réelles gaussiennes (ou normales)

4.1.1 La loi normale centrée réduite

DEFINITION – 4.1 On dit qu'une v.a.r. X suit **la loi normale (ou gaussienne) centrée réduite** et on note $X \sim \mathcal{N}(0, 1)$ si X admet pour densité

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{pour tout } x \in \mathbf{R}.$$

La fonction f_X définie ci-dessus est bien une densité de probabilité sur \mathbf{R} . En effet, elle est clairement positive. De plus, comme elle est continue sur \mathbf{R} , elle est mesurable. Il reste à prouver que son intégrale sur \mathbf{R} vaut 1 : Voir l'exercice suivant.

EXERCICE 87

1. On souhaite ici calculer l'intégrale $I = \int_{\mathbf{R}} e^{-x^2/2} dx$. Expliquer pourquoi le calcul direct de I n'est pas possible.
2. Démontrer que l'on a : $\int_{\mathbf{R}^2} e^{-(x^2+y^2)/2} dx dy = 2\pi$.
3. Conclure en utilisant le fait admis suivant (qui repose sur le Théorème de Fubini-Tonelli) que

$$\int_{\mathbf{R}^2} e^{-(x^2+y^2)/2} dx dy = \int_{\mathbf{R}} e^{-x^2/2} dx \times \int_{\mathbf{R}} e^{-y^2/2} dy.$$

Voici quelques propriétés fondamentales de la loi $\mathcal{N}(0, 1)$.

PROPOSITION – 4.1 Si $X \sim \mathcal{N}(0, 1)$ alors

- (a) la variable $-X$ suit aussi la loi $\mathcal{N}(0, 1)$ (on dit que la loi $\mathcal{N}(0, 1)$ est *symétrique*),
- (b) $\mathbf{E}(X) = 0$ et $\text{Var}(X) = 1$ (autrement dit : X est *une variable centrée et réduite*),
- (c) *Fonction caractéristique de X* : $\phi_X(t) = e^{-t^2/2}$ pour tout $t \in \mathbf{R}$.

▮ PREUVE. Pour (a) et (b) : Voir COURS. Le point (c) est établi dans l'exercice suivant.

Dans le Chapitre 2, le calcul de la fonction caractéristique de $X \sim \mathcal{N}(0, 1)$ a déjà été fait : vous avez calculé $\mathbf{E}(e^{zX})$ pour tout $z \in \mathbf{R}$.

Puis grâce au Principe de Prolongement Analytique, le résultat se prolonge à tout $z \in \mathbf{C}$. Voici une autre méthode essentiellement basée sur la résolution d'une équation différentielle :

EXERCICE 88 (FONCTION CARACTÉRISTIQUE DE LA LOI $\mathcal{N}(0, 1)$)

Soit X une v.a.r. qui suit la loi $\mathcal{N}(0, 1)$. On cherche à déterminer sa fonction caractéristique ϕ_X .

On peut montrer (ADMIS) que ϕ_X est dérivable sur \mathbf{R} et que l'on peut "permuter dérivée et intégrale" (i.e. que sa dérivée s'obtient en dérivant le terme sous l'intégrale) i.e. que $\phi_X'(t) = \mathbf{E}(iX e^{itX})$.

1. Démontrer que la fonction ϕ_X est solution de l'équation différentielle : (E) $y'(t) = -ty(t)$ pour tout $t \in \mathbf{R}$.

2. En déduire l'expression de ϕ_X en fonction de t .

REMARQUE – 4.1 (COMMENTAIRES SUR LA FONCTION DE RÉPARTITION DE LA LOI $\mathcal{N}(0, 1)$.)

Si X suit la loi $\mathcal{N}(0, 1)$, sa fonction de répartition de X est définie par : $F_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-x^2/2} dx$ pour tout $t \in \mathbf{R}$.

1) Cas $t = 0$: Dans le cas particulier où $t = 0$, on a :

LEMME – 4.1 Si $X \sim \mathcal{N}(0, 1)$ alors $F_X(0) = 1/2$ et même $\mathbf{P}(X \leq 0) = \mathbf{P}(X \geq 0) = 1/2$.

▮ PREUVE. Voir COURS pour les détails. ▮

2) Cas $t \neq 0$: Par contre pour tout réel $t \neq 0$, la quantité $F_X(t)$ ne peut pas être calculée explicitement en fonction de t .

Pourquoi ? C'est parce que la fonction $x \mapsto e^{-x^2/2}$ n'admet pas de primitive sur \mathbf{R} .

Cependant, on dispose d'approximations de $F_X(t)$ pour un grand nombre de réels fixés $t > 0$: ces approximations sont données dans un tableau appelé *table statistique*.

4.1.2 Lois (réelles) normales : Définition et propriétés générales

DEFINITION – 4.2 Soient m et σ deux réels tels que $\sigma \geq 0$.

On dit qu'une v.a.r. X *suit la loi normale (ou gaussienne) de paramètres m et σ^2* et on note $X \sim \mathcal{N}(m, \sigma^2)$ si :

a) soit $\sigma = 0$: alors X est \mathbf{P} -presque sûrement constante égale à m ce que l'on note souvent " $X = m$ \mathbf{P} -p.s." autrement dit X est une variable discrète telle que $\mathbf{P}(X = m) = 1$,

b) soit $\sigma > 0$: alors X admet une densité sur \mathbf{R} donnée par : $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$, $\forall x \in \mathbf{R}$.

⚠ **Attention.** Lorsque l'on demande de montrer qu'une variable aléatoire suit une loi normale, il est sous-entendu (et impératif) de préciser les paramètres m et σ^2 !

REMARQUE – 4.2 Signalons que dans certains ouvrages, le cas $\sigma = 0$ n'est pas être inclus dans la définition des lois normales. Autrement dit : les lois $\mathcal{N}(m, \sigma^2)$ sont définies uniquement pour les couples $(m, \sigma) \in \mathbf{R} \times]0, +\infty[$: ce sont alors toutes des lois continues (cf. Cas **b**) ci-dessus). Nous insistons bien sur le fait que dans ce cours, nous considérons le cas $\sigma \in \mathbf{R}^+$: nous incluons donc les masses de Dirac dans la famille des lois normales (en effet : pour tout $m \in \mathbf{R}$, comme $X \sim \mathcal{N}(m, 0)$ signifie (cf. Cas **a**)) que $\mathbf{P}(X = m) = 1$, cela équivaut à dire que $\mathbf{P}_X = \delta_m$ autrement dit que $\mathcal{N}(m, 0) = \delta_m$). Nous motiverons ce choix dans la Remarque 4.4 de la Section 4.2.

Voici une propriété très utilisée qui relie toute loi normale à la loi normale centrée réduite.

PROPOSITION – 4.2 (CENTRAGE ET RÉDUCTION DES LOIS GAUSSIENNES.) Soient $m \in \mathbf{R}$ et $\sigma \in \mathbf{R}^+$.

a) Si $X \sim \mathcal{N}(m, \sigma^2)$ alors on peut l'écrire $X = \sigma Z + m$ avec $Z \sim \mathcal{N}(0, 1)$.

b) Dans le cas $\sigma > 0$, on a l'équivalence : $X \sim \mathcal{N}(m, \sigma^2) \iff \frac{X-m}{\sigma} \sim \mathcal{N}(0, 1)$.

▮ PREUVE. Voir COURS. ▮

Ainsi tout calcul de probabilité sur une loi gaussienne quelconque se ramène à un calcul sur la loi normale centrée réduite. Nous allons donc pouvoir établir l'essentiel des propriétés sur les lois normales de paramètres (m, σ^2) quelconques à partir des propriétés sur la loi $\mathcal{N}(0, 1)$.

EXERCICE 89

Soient $m \in \mathbf{R}$ et $\sigma \in \mathbf{R}^+$. Considérons X une v.a.r. de loi $\mathcal{N}(m, \sigma^2)$. On note F_X sa fonction de répartition.

1. Supposons que $\sigma > 0$. Identifier la ou les valeur(s) de t pour laquelle (ou lesquelles) on sait expliciter $F_X(t)$.
2. Même question dans le cas $\sigma = 0$.

PROPOSITION – 4.3 Soient m et σ deux réels tels que $\sigma \geq 0$. Si $X \sim \mathcal{N}(m, \sigma^2)$ alors :

- a) $\mathbf{E}(X) = m$ et $\mathbf{V}(X) = \sigma^2$,
- b) pour tout $t \in \mathbf{R}$, $\phi_X(t) = \exp\left(imt - \frac{\sigma^2 t^2}{2}\right)$.

┌ PREUVE. Voir COURS. ─

EXERCICE 90 (CARACTÉRISATION DES V.A.R. P-PRESQUE SÛREMENT CONSTANTES.)

Soient $m \in \mathbf{R}$ et X une v.a.r. Démontrer l'équivalence suivante :

$$(X = m \text{ P-p.s.}) \iff (X \in L^2 \text{ avec } \mathbf{E}(X) = m \text{ et } \mathbf{V}(X) = 0).$$

► **CORRECTION.**

On rappelle que $\mathbf{V}(X) = \mathbf{E}[(X - \mathbf{E}(X))^2]$.

L'implication \implies est alors évidente puisque "X = m P-p.s." assure que $\mathbf{E}(X) = \mathbf{E}(m) = m$.

Pour la réciproque \impliedby , on utilise le fait que : si Y est une v.a.r. dans L^2 alors $\mathbf{E}(Y^2) = 0$ ssi " $Y = 0$ P-p.s.". On applique cette propriété en prenant $Y = X - \mathbf{E}(X) = X - m$ et en remarquant qu'alors $\mathbf{V}(X) = \mathbf{E}(Y^2)$.

PROPOSITION – 4.4 (LOIS GAUSSIENNES ET "QUELQUES OPÉRATIONS".)

- a) Toute transformation affine d'une v.a.r. gaussienne est encore gaussienne. Plus précisément : si $X \sim \mathcal{N}(m, \sigma^2)$ alors pour tous $(a, b) \in \mathbf{R}^2$: $aX + b \sim \mathcal{N}(am + b, a^2\sigma^2)$.
- b) La somme de 2 v.a.r. indépendantes est encore une variable gaussienne. Ainsi : si $X \sim \mathcal{N}(m, \sigma^2)$, $Y \sim \mathcal{N}(m', \sigma'^2)$ et si X et Y sont **indépendantes** alors : $X + Y \sim \mathcal{N}(m + m', \sigma^2 + \sigma'^2)$.

REMARQUE – 4.3 Il est évident que le point **b)** de la proposition précédente se généralise (par récurrence) au cas d'une somme (finie) quelconque de variables gaussiennes (mutuellement) indépendantes.

⚠ **Attention.** L'hypothèse d'indépendance du point **b)** est **primordiale**. Un contre-exemple sera notamment donné dans l'Exercice 95 (dans la Section 4.2.2).

┌ PREUVE. (Preuve de la Proposition 4.4). Voir COURS. ─

§ 4.2. Vecteurs aléatoires gaussiens : Définition

4.2.1 Notations, Rappels et Compléments sur la matrice de (variance-)covariance

a) Notations et Rappels

► **Notations :** Nous allons considérer essentiellement des vecteurs aléatoires (ou non) à valeurs dans \mathbf{R}^n avec $n \in \mathbf{N}^*$ fixé. Voici les notations et conventions, parfois différentes de celles des Chapitres 2 et 3, utilisés dans la suite de ce cours. Fixer ceci est indispensable car nous allons manipuler des vecteurs, des matrices et faire des opérations (d'Algèbre linéaire usuel) dessus.

(N1) Les vecteurs seront ici notés - c'est très fréquent - en colonne : ainsi tout vecteur u de \mathbf{R}^n s'écrit de deux manières différentes via la transposition (que l'on notera ".^T"): :

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \quad \text{ou encore} \quad u = (u_1, \dots, u_n)^T.$$

N.B. : Parfois par abus (et par habitude) quand on introduira un vecteur u de \mathbf{R}^n on le notera en ligne sans le "signe transposé" : ceci ne posait pas de problème dans les chapitres précédents car on ne faisait aucune opération sur les vecteurs considérés !

(N2) Le produit scalaire et la norme euclidiens sur \mathbf{R}^n sont notés respectivement $\langle \cdot, \cdot \rangle$ et $\|\cdot\|$. On rappelle que pour tous vecteurs u et v de \mathbf{R}^n on a :

$$\langle u, v \rangle = u^T v = \sum_{j=1}^n u_j v_j \quad \text{et} \quad \|u\| = \sqrt{\langle u, u \rangle} = \sqrt{\sum_{j=1}^n u_j^2}.$$

(N3) S'il n'y a pas d'ambiguïté le vecteur nul de \mathbf{R}^n i.e. le vecteur $(0, \dots, 0)^T$ sera noté $0_{\mathbf{R}^n}$ ou simplement 0 .

(N4) I_n désigne la matrice identité d'ordre n .

Nous allons, afin de manipuler ses notations et de bien fixer les choses, rappeler des points importants donnés sur les vecteurs aléatoires (quelconques) dans le Chapitre 2 et qui nous seront utiles pour la suite.

► **Rappels** : Soit $n \in \mathbf{N}^*$ et soit $X = (X_1, \dots, X_n)^T$ un vecteur aléatoire de \mathbf{R}^n .

(R1) X est dit intégrable si les n composantes de X sont intégrables autrement dit si X est intégrable ($X \in L^1$), i.e. si

$$"\forall 1 \leq j \leq n, \mathbf{E}(|X_j|) < +\infty" \quad \text{ou de manière équivalente} \quad \mathbf{E}(\|X\|) < +\infty."$$

Dans ce cas, on dit aussi que X est admet un moment d'ordre 1 appelé *espérance* : c'est le vecteur \mathbf{R}^n donné par

$$\mathbf{E}[X] = (\mathbf{E}(X_1), \dots, \mathbf{E}(X_n))^T.$$

(R2) Si les n composantes de X sont toutes de carrés intégrables autrement dit si X est de carré intégrable ($X \in L^2$), i.e. si

$$"\forall 1 \leq j \leq n, \mathbf{E}(X_j^2) < +\infty" \quad \text{ou de manière équivalente} \quad \mathbf{E}(\|X\|^2) < +\infty"$$

alors on appelle *matrice de variance-covariance* de X ou simplement *matrice de covariance* de X la matrice carrée d'ordre n notée $\Gamma_X = (\Gamma_{ij})_{1 \leq i, j \leq n}$ définie par :

$$\forall 1 \leq i, j \leq n, \quad \Gamma_{ij} = \mathbf{cov}(X_i, X_j) := \mathbf{E}\left[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])\right] = \mathbf{E}[X_i X_j] - \mathbf{E}[X_i]\mathbf{E}[X_j].$$

De plus, Γ_X est une matrice symétrique (i.e. $\forall 1 \leq i, j \leq n, \Gamma_{ji} = \Gamma_{ij}$) puisque la covariance est une forme symétrique. Enfin, on a (via la définition de la variance) : $\forall 1 \leq i \leq n, \Gamma_{ii} = \mathbf{V}(X_i)$.

EXERCICE 91

Soient X_1, X_2, X_3 trois variables aléatoires indépendantes de même loi exponentielle $\mathcal{E}(1)$. On pose :

$$Y_1 = 2X_1 + X_2 - 5 \quad ; \quad Y_2 = 3X_2 - 4X_3 \quad ; \quad Y_3 = X_3.$$

Déterminer l'espérance et la matrice de covariance du vecteur $Y = (Y_1, Y_2, Y_3)^T$.

EXERCICE 92

Soit $X = (X_1, \dots, X_n)^T$ un vecteur aléatoire à valeurs dans \mathbf{R}^n . On suppose que X est de carré intégrable, on note m son espérance et Γ sa matrice de covariance.

Fixons $a = (a_1, \dots, a_n)^T$ un vecteur quelconque (non aléatoire) de \mathbf{R}^n . Montrer que la v.a.r. $\langle a, X \rangle$ est de carré intégrable et que :

$$\mathbf{E}[\langle a, X \rangle] = \langle a, m \rangle \quad \text{et} \quad \mathbf{V}(\langle a, X \rangle) = a^T \Gamma a.$$

► **CORRECTION.**

On a : $\langle a, X \rangle = \sum_{j=1}^n a_j X_j$. Comme les X_j sont toutes dans L^2 qui est un \mathbf{R} -espace vectoriel, $\langle a, X \rangle$ est aussi dans L^2 . Par les propriétés de linéarité de l'espérance, on a :

$$\mathbf{E}[\langle a, X \rangle] = \sum_{j=1}^n \mathbf{E}[a_j X_j] = \sum_{j=1}^n a_j \mathbf{E}[X_j] = \langle a, \mathbf{E}[X] \rangle = \langle a, m \rangle.$$

Pour la variance, en se ramenant à la covariance puis en utilisant de bilinéarité de la covariance, on a :

$$\mathbf{V}(\langle a, X \rangle) = \mathbf{cov}(\langle a, X \rangle; \langle a, X \rangle) = \mathbf{cov}\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^n a_j X_j\right) = \sum_{i,j=1}^n a_i a_j \mathbf{cov}(X_i, X_j).$$

On conclut $\mathbf{V}(\langle a, X \rangle) = a^T \Gamma a$ puisque $\Gamma = (\Gamma_{ij})_{1 \leq i, j \leq n}$ avec $\Gamma_{ij} := \mathbf{cov}(X_i, X_j)$.

b) Compléments sur la matrice de (variance-)covariance

On rappelle que dans le cas où X est une variable aléatoire réelle de carré intégrable, sa variance $\mathbf{V}(X)$ est bien définie. On la note souvent σ^2 puisque c'est un réel toujours positif ou nul. Ainsi, elle admet une racine carrée : $\sigma = \sqrt{\mathbf{V}(X)}$ qui est l'*écart-type* de X vérifie bien $\mathbf{V}(X) = \sigma^2$.

La notion de matrice de covariance pour les vecteurs aléatoires (de L^2) généralise celle de variance pour les v.a.r. (de L^2) (notez qu'en prenant $n = 1$ dans le rappel (R2), on a $\Gamma_X = \mathbf{V}(X)$!). Le but de cette section est de voir que ceci "se généralise (en un certain sens)" à la matrice de covariance d'un vecteur de carré intégrable.

PROPOSITION – 4.5 Soient $n \in \mathbf{N}^*$ et X un vecteur aléatoire à valeurs dans \mathbf{R}^n que l'on suppose dans L^2 . Alors Γ_X sa matrice de covariance est définie et elle s'écrit : $\Gamma_X = \Lambda_X \Lambda_X^T$ où Λ_X est une matrice réelle de dimension $n \times n$. On dit parfois que Λ_X est (une) "racine carrée de Γ_X " et on la note $\Lambda_X = \sqrt{\Gamma_X}$.

Ceci repose sur des arguments d'Algèbre Linéaire et vient du fait que Γ_X est une matrice symétrique et semi-définie positive : nous allons justifier ci-après.

c) Preuve de la Proposition 4.5

Rappelons tout d'abord les notions de matrice (semi-)définie positive.

✕ **RAPPEL –** Soit $n \in \mathbf{N}^*$. Soit $A = (a_{ij})_{1 \leq i, j \leq n}$ une matrice réelle carrée d'ordre n . On suppose que A est **symétrique** (autrement dit $A^T = A$). On dit que :

- a) A est une *matrice semi-définie positive* si : $u^T A u \geq 0$ pour tout $u \in \mathbf{R}^n$.
- b) A est *définie positive* si : $u^T A u > 0$ pour tout $u \neq 0$ de \mathbf{R}^n .

✕

LEMME – 4.2 (PROPRIÉTÉS DES MATRICES DE COVARIANCE.)

Soient $n \in \mathbf{N}^*$ et X un vecteur aléatoire à valeurs dans \mathbf{R}^n . On suppose que X est dans L^2 . La matrice de covariance Γ_X de X est une matrice symétrique et semi-définie positive.

EXERCICE 93

Démontrer le Lemme 4.2.

► **CORRECTION.**

La définition de Γ_X et la symétrie de la covariance (i.e. $\mathbf{cov}(U, V) = \mathbf{cov}(V, U)$ pour tout couple (U, V) de v.a.r. de L^2) implique aisément la symétrie de la matrice Γ_X .

De plus, pour tout $u \in \mathbf{R}^n$, on a déjà vu (cf. Exercice 92) que : $u^T \Gamma_X u = \mathbf{V}(\langle u, X \rangle)$ qui est évidemment ≥ 0 (comme toute variance !).

La justification de la Proposition 4.5 repose sur le Lemme 4.2 combiné au résultat suivant.

LEMME – 4.3 (RACINES CARRÉES MATRICIELLES DES MATRICES SYMÉTRIQUES ET SEMI-DÉFINIES POSITIVES.)

Étant donné $n \in \mathbf{N}^*$, on note $\mathcal{M}_n(\mathbf{R})$ l'ensemble des matrices réelles de dimension $n \times n$.

Si $A \in \mathcal{M}_n(\mathbf{R})$ est une matrice symétrique semi-définie positive, alors il existe une matrice $B \in \mathcal{M}_n(\mathbf{R})$ telle que : $A = B B^T$. Cette matrice B est appelée *racine carrée* de A (on la note parfois \sqrt{A}). Attention : en général B n'est pas unique !

Pour être complet, nous donnons la démonstration de ce lemme qui repose sur des arguments assez simples d'Algèbre Linéaire.

▮ **PREUVE.** D'abord, comme A est une matrice symétrique, elle est diagonalisable. Ainsi, si l'on note λ_i avec $1 \leq i \leq n$ les n valeurs propres de A (certaines peuvent évidemment être égales), A peut s'écrire :

$$A = P D P^T \quad \text{avec } D = \text{diag}(\lambda_1, \dots, \lambda_n) \quad \text{et } P \text{ matrice orthogonale (i.e. telle que } P P^T = P^T P = I_n)$$

où l'on rappelle que les vecteurs colonnes de P sont vecteurs propres de A .

Ensuite, la semi-définie positivité de A implique que ses valeurs propres $\lambda_1, \dots, \lambda_n$ sont toutes ≥ 0 . En effet : pour tout $1 \leq i \leq n$, si on désigne par f_i la $i^{\text{ème}}$ colonne de P qui est par construction un vecteur propre normalisé associé à λ_i , on a alors :

$$\forall 1 \leq i \leq n, \quad A f_i = \lambda_i f_i \quad \text{et} \quad \langle f_i, f_i \rangle = 1. \quad (4.1)$$

Fixons un entier $1 \leq i \leq n$. Comme A est semi-définie positive, on sait que $f_i^T A f_i \geq 0$. Par ailleurs, en utilisant (4.1) on a les égalités successives suivantes :

$$f_i^T A f_i = f_i^T (\lambda_i f_i) = \lambda_i \langle f_i, f_i \rangle = \lambda_i.$$

On a prouvé que $\lambda_i \geq 0$.

On introduit alors la matrice \tilde{D} diagonale donnée par : $\tilde{D} = \text{diag}(\sqrt{\lambda_i}, 1 \leq i \leq n)$.

On voit aisément que $\tilde{D}^2 = D$ et évidemment que $\tilde{D}^T = \tilde{D}$. Donc si l'on pose $B = \tilde{D}P$, la relation $A = PDP^T$ nous donne bien $A = BB^T$.

┘

4.2.2 Définition de la notion de vecteur gaussien

DEFINITION – 4.3 Soient $n \in \mathbf{N}^*$ et $X = (X_1, \dots, X_n)^T$ un vecteur aléatoire de \mathbf{R}^n .

On dit que X est un **vecteur gaussien** si toute combinaison linéaire de ses composantes est une v.a.r gaussienne i.e. :

$$\forall a = (a_1, \dots, a_n)^T \in \mathbf{R}^n, \quad \langle a, X \rangle = \sum_{j=1}^n a_j X_j \text{ est une v.a.r gaussienne.}$$

REMARQUE – 4.4 (COMMENTAIRES ET COMPLÉMENTS À CETTE DÉFINITION.)

1) On peut noter que cette définition inclut le cas $a = \mathbf{0}_{\mathbf{R}^n}$ auquel correspond la variable aléatoire réelle (déterministe) constante égale à 0 : il faut donc que 0 soit une v.a.r. gaussienne. Par conséquent inclure les v.a. constantes (ou \mathbf{P} -presque sûrement constantes) dans la Définition 4.2 des variables gaussiennes trouve sa motivation !

2) Nous verrons qu'un vecteur gaussien est dans L^2 (c'est par exemple une conséquence de la Proposition 4.6 ci-dessous : en effet, les composantes d'un vecteur gaussien sont nécessairement des v.a.r. gaussiennes qui sont dans L^2). Ainsi, on peut reformuler la définition précédente (en utilisant notamment aussi l'Exercice 92) :

si X est un vecteur aléatoire de \mathbf{R}^n de carré intégrable, d'espérance $m = \mathbf{E}[X]$ et de matrice de covariance $\Gamma = \Gamma_X$, alors :

$$X \text{ est un vecteur gaussien } \text{ssi } \langle a, X \rangle \sim \mathcal{N}(\langle a, m \rangle, a^T \Gamma a) \text{ pour tout } a \in \mathbf{R}^n.$$

Nous renvoyons aussi à la Définition 4.4 pour une autre définition (équivalente) des vecteurs gaussiens faisant intervenir l'espérance et la matrice de covariance.

EXERCICE 94

Soient $(m_1, m_2, \sigma_1, \sigma_2) \in \mathbf{R}^4$ avec $\sigma_1 \geq 0$ et $\sigma_2 \geq 0$.

On considère X_1 et X_2 deux v.a.r. indépendantes telles que $X_j \sim \mathcal{N}(m_j, \sigma_j^2)$ pour $j = 1$ et 2.

Justifier que le vecteur $V := (3X_1 + 2, X_1 + X_2)^T$ est un vecteur aléatoire gaussien.

EXERCICE 95

Soit X une variable aléatoire gaussienne centrée réduite.

Soit ε une variable aléatoire indépendante de X et qui suit une loi de Rademacher $\mathcal{R}(1/2)$.

1. On pose $Y = \varepsilon X$. Déterminer la loi de Y .
2. Calculer $\mathbf{P}(X + Y = 0)$. La loi de $X + Y$ est-elle gaussienne ? Justifiez votre réponse.
3. Le vecteur $(X, Y)^T$ est-il gaussien ? Justifiez.

PROPOSITION – 4.6 (LOI DES COORDONNÉES D'UN VECTEUR GAUSSIEN.)

Soit $n \geq 2$ un entier. Soit $X = (X_1, \dots, X_n)^T$ un vecteur aléatoire de \mathbf{R}^n que l'on suppose dans L^2 .

On note $m = (m_1, \dots, m_n)^T$ son vecteur espérance et $\Gamma = (\Gamma_{i,j})_{1 \leq i, j \leq n}$ sa matrice de covariance.

Si X est un vecteur gaussien alors toutes ses composantes sont des v.a.r. gaussiennes avec :

$$\text{pour tout } 1 \leq j \leq n, \quad X_j \sim \mathcal{N}(m_j, \Gamma_{j,j}).$$

┘ PREUVE. Voir COURS. ┘

⚠ **Attention.** La réciproque de la Proposition 4.5 est fautive ! En général : pour qu'un vecteur soit gaussien il faut **mais** il ne suffit pas que toutes ses composantes soient des gaussiennes ! Un contre-exemple est donné en particulier fourni par l'Exercice 95 ci-dessus.

Nous verrons dans la Section 4.3.3 une condition assurant qu'il y a équivalence.

§ 4.3. Vecteurs aléatoires gaussiens : Quelques propriétés

4.3.1 Fonction caractéristique d'un vecteur gaussien

D'après la Proposition 4.6, un vecteur gaussien est nécessairement de carré intégrable : ainsi son espérance et sa matrice de covariance sont bien définies. En fait, ces deux quantités déterminent complètement la fonction caractéristique.

THÉORÈME – 4.1 Soient $n \in \mathbf{N}^*$ et $X = (X_1, \dots, X_n)^T$ un vecteur gaussien de \mathbf{R}^n admettant m pour espérance m et Γ pour matrice de covariance. Alors on a :

$$\text{pour tout } t = (t_1, \dots, t_n)^T \in \mathbf{R}^n, \quad \phi_X(t) := \mathbf{E}[e^{i\langle t, X \rangle}] = \exp\left(i\langle t, m \rangle - \frac{1}{2}t^T \Gamma t\right). \quad (4.2)$$

▮ PREUVE. Voir COURS. ▮

Ce résultat montre donc (puisque la fonction caractéristique d'un vecteur aléatoire caractérise sa loi) que la loi d'un vecteur gaussien est caractérisée par son vecteur espérance et sa matrice de covariance. Ceci induit une autre définition des lois gaussiennes multi-dimensionnelles.

DEFINITION – 4.4 (DÉFINITION ALTERNATIVE DES LOIS GAUSSIENNES MULTI-DIMENSIONNELLES.)

Soient $n \in \mathbf{N}^*$ et $X = (X_1, \dots, X_n)^T$ un vecteur aléatoire de \mathbf{R}^n de carré intégrable. On note $m = \mathbf{E}(X)$ son espérance et Γ sa matrice de covariance.

- a) On dit que X est un **vecteur gaussien de moyenne (ou espérance) m et de matrice de covariance Γ** si sa fonction caractéristique est donnée sur \mathbf{R}^n par la formule (4.2). On note alors $X \sim \mathcal{N}_n(m, \Gamma)$.
- b) De plus, on dit que X est un **vecteur gaussien centré** si $m = 0 = 0_{\mathbf{R}^n}$ (le vecteur nul de \mathbf{R}^n). Et on parle de **vecteur gaussien réduit** si $\Gamma = I_n$ (la matrice identité d'ordre n).
Enfin, si $X \sim \mathcal{N}_n(0, I_n)$: on dit que X est un **vecteur gaussien centré et réduit de \mathbf{R}^n** .

REMARQUE – 4.5 Dans le cas réel $n = 1$, on retrouve les lois réelles gaussiennes : dans ce cas l'espérance m est un réel, la matrice de covariance est simplement la variance notée usuellement σ^2 . De plus, au lieu d'écrire $\mathcal{N}_1(m, \sigma^2)$ on notera simplement $\mathcal{N}(m, \sigma^2)$ comme dans la Section 4.1.

REMARQUE – 4.6 (COMPLÉMENTS AU LEMME 4.2.) Pour compléter la Section 4.2.1c), voici la **caractérisation des matrices de covariance** : Soient $n \in \mathbf{N}^*$ et A une matrice réelle de dimension $n \times n$. Alors :

A est la matrice de covariance d'un vecteur aléatoire de \mathbf{R}^n ssi A est une matrice symétrique et semi-définie positive.

EXERCICE 96 (RECTIFICATIF de l'énoncé du Mercredi 28/09/16.)

(a) Montrer que $A = \begin{pmatrix} 2 & -2 \\ -2 & 3 \end{pmatrix}$ est une matrice de covariance d'un vecteur de \mathbf{R}^2 (Indication : Utiliser la Remarque 4.6).

(b) Soit $(X, Y)^T$ un couple de fonction caractéristique : $\phi_{(X, Y)}(t) = \exp(it_1 - 4it_2 - t_1^2 + 2t_1t_2 - \frac{3}{2}t_2^2) \quad \forall t = (t_1, t_2)^T \in \mathbf{R}^2$.

Donner la loi du vecteur $(X, Y)^T$ ainsi que ses lois marginales.

(c) ATTENTION : La fonction ψ donnée pour tout $t = (t_1, t_2)^T \in \mathbf{R}^2$ par $\psi(t) = \exp(it_1 - 4it_2 - t_1^2 + 3t_1t_2 - \frac{1}{2}t_2^2)$ n'est pas la fonction caractéristique d'un vecteur gaussien de \mathbf{R}^2 .

Pourquoi ? (Indication : Utiliser la formule (4.2) et la Remarque 4.6).

4.3.2 Transformations affines d'un vecteur gaussien

PROPOSITION – 4.7 Toute transformation affine d'un vecteur gaussien est encore gaussien. Ainsi :

- a) Soient $n \geq 1$ et $p \geq 1$ deux entiers. Soient b un vecteur de \mathbf{R}^p et A une matrice réelle de taille $p \times n$. Si $X \sim \mathcal{N}_n(m, \Gamma)$ alors $AX + b \sim \mathcal{N}_p(Am + b, A\Gamma A^T)$.
- b) En particulier, tout sous-vecteur de X est encore un vecteur gaussien (un *sous-vecteur de X* est un vecteur construit à partir de X en extrayant une partie de ses composantes donc de la forme $(X_i)_{i \in I}$ où I est une partie de $\{1, \dots, n\}$).

▮ PREUVE. Voir COURS. ▮

EXERCICE 97

On considère $X = (X_1, X_2)^T$ un couple aléatoire gaussien qui est centré et de matrice de covariance $\Gamma_X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.

1. Déterminer les lois des variables aléatoires X_1 et X_2 .
2. Déterminer la loi du vecteur $Y = (X_1 + X_2, -2X_1 + X_2 - 5)^T$.

Dans le cas réel $n = 1$ (Section 4.1), nous avons vu (Proposition 4.2) que toute loi réelle gaussienne peut-être reliée - par centrage et éventuellement réduction (lorsque c'est possible) - à la loi normale centrée réduite $\mathcal{N}(0, 1)$.

Question naturelle : Qu'en est-il pour les vecteurs gaussiens ? Nous allons voir qu'il existe un analogue à cette propriété à savoir que l'on peut "ramener" toute loi gaussienne sur \mathbf{R}^n (avec $n \in \mathbf{N}^*$) à la loi $\mathcal{N}_n(0, I_n)$.

PROPOSITION – 4.8 (CENTRAGE ET RÉDUCTION DES VECTEURS GAUSSIENS.) Soient $n \in \mathbf{N}^*$ et X un vecteur aléatoire à valeurs dans \mathbf{R}^n . On suppose que X est dans L^2 .

On note $m = \mathbf{E}(X)$ son espérance. Si Γ désigne sa matrice de covariance : il existe (Voir Proposition 4.5) une matrice notée Λ de dimension $n \times n$ telle que $\Gamma = \Lambda \Lambda^T$.

- (a) Si $X \sim \mathcal{N}_n(m, \Gamma)$ alors $X = \Lambda Z + m$ avec $Z \sim \mathcal{N}_n(0, I_n)$.
- (b) De plus si Γ est **inversible**, alors on a : $X \sim \mathcal{N}_n(m, \Gamma) \iff \Lambda^{-1}(X - m) \sim \mathcal{N}_n(0, I_n)$.

▮ PREUVE. La démonstration est laissée en exercice au lecteur : elle s'obtient aisément à partir de Proposition 4.7. ▮

4.3.3 Vecteurs gaussiens et Indépendance

On a déjà vu que (cf. Proposition 4.6) si X est un vecteur gaussien alors toutes ses composantes sont des v.a.r. gaussiennes. **Mais** la réciproque est fautive en général : un contre-exemple est donné dans l'Exercice 95 ci-dessus.

Voici une condition suffisante sur X assurant que la réciproque est vraie.

PROPOSITION – 4.9 Soit $X = (X_1, \dots, X_n)^T$ un vecteur aléatoire de \mathbf{R}^n .

Si X_1, \dots, X_n sont des v.a.r. gaussiennes et **indépendantes** alors $X = (X_1, \dots, X_n)^T$ est un vecteur gaussien.

▮ PREUVE. On propose ici une preuve basée sur la Définition 4.3 de vecteurs gaussiens (Notez que l'on peut aussi prouver cela via les fonction caractéristique : cet argument est très proche de certains points de la démonstration du Théorème 4.2 ci-dessous).

Soit $a = (a_1, \dots, a_n)^T \in \mathbf{R}^n$. Alors $\langle a, X \rangle = \sum_{j=1}^n Y_j$ en posant $Y_j = a_j X_j$ pour tout $1 \leq j \leq n$.

On utilise alors les 2 propriétés de la Proposition 4.4 :

- d'une part, comme toute transformation affine (donc en particulier linéaire) d'une v.a.r. gaussienne est encore gaussienne : on en déduit que Y_j est une gaussienne pour tout j ,
- d'autre part, les $(X_j)_{j=1, \dots, n}$ sont indépendantes donc les $(Y_j)_{j=1, \dots, n}$ aussi. La somme de v.a.r. gaussiennes indépendantes étant encore une variable gaussienne, on en conclut que $\langle a, X \rangle$ est une v.a.r. gaussienne.

▮

On a déjà vu (cf. Chapitre 2) que : l'indépendance de deux v.a.r. implique leur non-corrélation (évidemment à condition que ces v.a.r. soient dans L^2); **mais** la réciproque est fautive en général SAUF dans le cas d'un vecteur gaussien :

THÉORÈME – 4.2 (CARACTÉRISATION DE L'INDÉPENDANCE DES COMPOSANTES D'UN VECTEUR GAUSSIEN.)

Soient $n \geq 2$ un entier et $X = (X_1, \dots, X_n)^T$ un vecteur **gaussien** de \mathbf{R}^n .

Alors les quatre assertions suivantes sont équivalentes :

- (a) les v.a. X_1, \dots, X_n sont (mutuellement) indépendantes,
- (b) les v.a. X_1, \dots, X_n sont 2 à 2 indépendantes,
- (c) les v.a. X_1, \dots, X_n sont 2 à 2 non corrélées,
- (d) la matrice de covariance Γ_X de X est diagonale.

▮ PREUVE. Voir COURS. ▮

► **CORRECTION.**

Il est clair que (a) \Rightarrow (b) \Rightarrow (c) (N.B : Ceci est général i.e. vrai pour tout vecteur aléatoire X de \mathbf{R}^n qui est dans L^2). De plus, comme les éléments hors-diagonaux de Γ_X sont les corrélations entre 2 composantes distinctes X_i et X_j (avec $i \neq j$) de X , on a aisément : (c) \Rightarrow (d).

Preuve de (d) \Rightarrow (a) : L'assertion (d) signifie que $\Gamma_X = \text{diag}(\mathbf{V}(X_j) : 1 \leq j \leq n)$. Par conséquent, on a :

$$\forall t = (t_1, \dots, t_n)^T \in \mathbf{R}^n, \quad t^T \Gamma_X t = \sum_{j=1}^n \mathbf{V}(X_j) t_j^2$$

et donc, si on note $m = \mathbf{E}(X)$ l'espérance de X ,

$$\phi_X(t) = \exp\left(i \langle m, t \rangle - \frac{1}{2} \sum_{j=1}^n \mathbf{V}(X_j) t_j^2\right) = \exp\left(i \sum_{j=1}^n \mathbf{E}(X_j) t_j - \frac{1}{2} \sum_{j=1}^n \mathbf{V}(X_j) t_j^2\right) = \prod_{j=1}^n \exp\left(i \mathbf{E}(X_j) t_j - \frac{1}{2} \mathbf{V}(X_j) t_j^2\right)$$

ce qui se traduit par : $\forall t \in \mathbf{R}^n, \phi_X(t) = \prod_{j=1}^n \phi_{X_j}(t_j)$ puisque l'on sait que comme X est gaussien de loi $\mathcal{N}_n(m, \Gamma_X)$, on a en particulier $X_j \sim \mathcal{N}(\mathbf{E}(X_j), \mathbf{V}(X_j))$ pour tout $1 \leq j \leq n$. Ceci implique et même équivaut, d'après les propriétés des fonctions caractéristiques (cf. Chapitre 2), à l'assertion (a).

En combinant les Propositions 4.6 et 4.9 au Théorème 4.2, on déduit aisément l'équivalence suivante.

PROPOSITION – 4.10 Soit $n \geq 2$ un entier. Si $X = (X_1, \dots, X_n)^T$ est un vecteur aléatoire à valeurs dans \mathbf{R}^n . Alors les deux assertions suivantes sont équivalentes :

- a) X est un vecteur gaussien et sa matrice de covariance Γ_X est diagonale,
- b) X_1, \dots, X_n sont des v.a.r. gaussiennes et indépendantes.

Ceci conduit à une autre définition de la loi d'un vecteur gaussien centré réduit.

COROLLAIRE – 4.1 (VECTEUR GAUSSIEN CENTRÉ RÉDUIT : DÉFINITION ALTERNATIVE.)

Soient $n \geq 2$ un entier et $X = (X_1, \dots, X_n)^T$ un vecteur aléatoire de \mathbf{R}^n . Alors :

$$X \sim \mathcal{N}_n(0, I_n) \quad \text{ssi} \quad \text{les composantes } X_1, \dots, X_n \text{ de } X \text{ sont indépendantes et toutes de loi } \mathcal{N}(0, 1).$$

Nous ne détaillerons pas la démonstration de ce résultat qui se déduit aisément de la Proposition 4.10 et des définitions du vecteur espérance et de la matrice de covariance.

EXERCICE 98

Soient X et ε deux v.a.r. indépendantes telles que $X \sim \mathcal{N}(0, 1)$ et $\varepsilon \sim \mathcal{R}(1/2)$. Soit $Y = \varepsilon X$.

Dans l'Exercice 95, on a montré que la variable Y suit une loi $\mathcal{N}(0, 1)$ et que $(X, Y)^T$ n'est pas un vecteur gaussien.

1. Les variables X et Y sont-elles corrélées ? Justifier.
2. Les variables X et Y sont-elles indépendantes ? Justifier.
3. Questions complémentaires :
 - (a) Déterminer la fonction caractéristique de $V = (X, Y)^T$ puis de $S = X + Y$.
 - (b) Comparer les valeurs de : $\mathbf{P}[(X > 1) \cap (Y > 1)]$ et $\mathbf{P}(X > 1) \mathbf{P}(Y > 1)$. Commenter.

4.3.4 Densité : Condition d'existence et Formule explicite

Comme pour les v.a.r. gaussiennes, nous allons voir que certains vecteurs gaussiens sont continus et d'autres pas. Plus précisément, nous allons donner la condition nécessaire et suffisante assurant qu'un vecteur gaussien est continu puis nous expliciterons (lorsqu'elle existe) la densité.

PROPOSITION – 4.11 Soit $n \geq 2$ un entier et soit $X = (X_1, \dots, X_n)^T$ un vecteur aléatoire à valeurs dans \mathbf{R}^n .

1) Vecteur gaussien centré réduit : $X \sim \mathcal{N}_n(0, I_n)$ ssi X admet pour densité :

$$f_X(x) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_{j=1}^n x_j^2} = \frac{1}{(2\pi)^{n/2}} e^{-\|x\|^2/2} \quad \text{pour tout } x = (x_1, \dots, x_n)^T \in \mathbf{R}^n.$$

2) Plus généralement : si $X \sim \mathcal{N}_n(m, \Gamma)$ alors

(a) X admet une densité (par rapport à la mesure de Lebesgue) sur \mathbf{R}^n ssi Γ est inversible.

(b) Dans ce cas, si on note $\det(\Gamma)$ le déterminant de la matrice Γ , alors cette densité est donnée par :

$$f_X(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(\Gamma)}} \exp\left(-\frac{1}{2}(x-m)^T \Gamma^{-1} (x-m)\right), \quad \forall x \in \mathbf{R}^n.$$

▮ PREUVE. Pour 1) : Voir COURS.

Le point 2)(a) est ADMIS.

Ici, nous allons détailler la preuve du 2)(b) : celui-ci se déduit de 1) par la Méthode de la Fonction Test (ou Fonction Muette) et un changement de variable en posant $X = \varphi(Z)$ avec $\varphi(z) = \Lambda z + m$ pour tout $z \in \mathbf{R}^n$.

• Il est clair que φ un \mathcal{C}^1 -difféomorphisme de \mathbf{R}^n sur \mathbf{R}^n ssi Γ est inversible.

On vérifie aisément que dans ce cas, la réciproque ψ de φ est donnée par : $\psi(x) = \Lambda^{-1}(x - m)$.

• On suppose (cf. la preuve du Lemme 4.3 du c) de la Section 4.2.1) ici que la matrice Λ est donnée par : $\Lambda = P\tilde{D}$ où

- i) $\tilde{D} = \text{diag}(\sqrt{\lambda_i}, 1 \leq i \leq n)$, les λ_i étant les valeurs propres de Γ . Ces λ_i sont toutes > 0 puisque : elles sont toutes ≥ 0 car Γ est semi-définie positive; d'autre part, elles sont non nulles puisque Γ est inversible (Γ est ici définie positive),
- ii) et P est une matrice orthogonale dont les colonnes sont des vecteurs propres orthonormés de Γ .

• De plus, la matrice Jacobienne \mathcal{J}_ψ de ψ vaut : $\mathcal{J}_\psi(x) = \Lambda^{-1} = (P\tilde{D})^{-1}$ pour tout $x \in \mathbf{R}^n$. Le jacobien de ψ satisfait

$$\det(\mathcal{J}_\psi(x)) = \det(\Lambda^{-1}) = \det((P\tilde{D})^{-1}) = \det(\tilde{D})^{-1} = 1/\det(\tilde{D}) \quad \text{car } \det(P) = 1 \text{ puisque } P \text{ est orthogonale.}$$

On conclut puisque $\det(\tilde{D}) = \sqrt{\det(\Gamma)}$. En effet : $\det(\Gamma) = \det(PDP^T) = \det(D) = (\det(\tilde{D}))^2$ qui est > 0 car $\det(\tilde{D}) = \prod_{i=1}^n \lambda_i$ et que toutes les λ_i sont > 0 . ▮

4.3.5 Retour sur le T.C.L. vectoriel : Compléments et Démonstration

Dans la Section 3.4.3 du Chapitre 3, nous avons énoncé et démontré le T.C.L. pour des v.a.r. (Théorème 3.3).

Nous nous intéressons ici à la version vectorielle du T.C.L. que nous rappelons :

THÉORÈME – 4.3 (THÉORÈME LIMITE CENTRAL : CAS RÉEL ET VECTORIEL.)

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires i.i.d., définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ et à valeurs dans \mathbf{R}^d avec $d \in \mathbf{N}^*$.

On suppose que $X_1 = (X_{j,1} : 1 \leq j \leq d)$ est dans L^2 autrement dit de carré intégrable i.e. que $\mathbf{E}(|X_1|^2) < \infty$.

On note :

- $m = \mathbf{E}(X_1)$ l'espérance de X_1 (c'est le vecteur de \mathbf{R}^d donné par $m = (\mathbf{E}(X_{i,1}) : 1 \leq i \leq d)$),
- $\Gamma = \mathbf{V}(X_1)$ la matrice de variance-covariance de X_1 (c'est la matrice carrée symétrique d'ordre d donnée par $\Gamma = (\text{cov}(X_{i,1}; X_{j,1}))_{1 \leq i, j \leq d}$).

On pose pour tout entier $n \in \mathbf{N}^*$, $\bar{X}_n := \frac{1}{n} \sum_{k=1}^n X_k$: c'est un vecteur de \mathbf{R}^d que l'on note $\bar{X}_n = (\bar{X}_{j,n} : 1 \leq j \leq d)$.

(a) Alors on a :

$$Y_n := \sqrt{n}(\bar{X}_n - m) \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} Y \quad \text{avec } Y \sim \mathcal{N}_d(0_{\mathbf{R}^d}, \Gamma). \quad (4.3)$$

(b) Dans le cas où Γ est inversible et s'écrit $\Gamma = \Lambda \Lambda^T$ alors la convergence (4.3) équivaut à :

$$Z_n := \sqrt{n} \Lambda^{-1} (\bar{X}_n - m) \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} Z \quad \text{avec } Z \sim \mathcal{N}_d(0_{\mathbf{R}^d}, I_d). \quad (4.4)$$

▮ **PREUVE. Preuve du T.C.L. vectoriel (Théorème 4.3).**

D'après ce qui a été rappelé, le cas réel $d = 1$ est déjà établi.

Pour le cas $d \geq 2$: nous détaillons ci-dessous la preuve. Nous allons voir qu'elle utilise justement la version du T.C.L dans le cas uni-dimensionnel.

Pour tout entier $n \geq 1$, on pose $Y_n = \sqrt{n}(\bar{X}_n - m)$: c'est un vecteur de \mathbf{R}^d .

On souhaite montrer la convergence suivante :

$$Y_n \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} Y \quad \text{avec } Y \sim \mathcal{N}_d(0_{\mathbf{R}^d}, \Gamma).$$

Cela revient à démontrer que :

$$\forall t \in \mathbf{R}^d, \quad \phi_{Y_n}(t) \xrightarrow{n \rightarrow +\infty} \phi_Y(t) = \exp\left(-\frac{1}{2} t^T \Gamma t\right).$$

Nous allons voir que cette démonstration utilise le TCL uni-dimensionnel (Théorème 3.3).

• Fixons $t \in \mathbf{R}^d$. Le but est de montrer que : $\phi_{Y_n}(t) \xrightarrow{n \rightarrow +\infty} \phi_Y(t)$.

Or, on a :

$$\phi_{Y_n}(t) = \mathbf{E}[e^{i \langle t, Y_n \rangle}] = \mathbf{E}\left[e^{i(S_n - n \langle t, m \rangle) / \sqrt{n}}\right] \quad (\star)$$

en posant $S_n = \sum_{k=1}^n Z_k$ avec $Z_k = \langle t, X_k \rangle$ pour tout entier $1 \leq k \leq n$.

• Comme $(X_k)_{k \geq 1}$ est une suite de vecteurs (de \mathbf{R}^d) qui sont i.i.d., dans L^2 , d'espérance m et de matrice de covariance Γ , on en déduit que $(Z_k)_{k \geq 1}$ est une suite de v.a.r. qui sont i.i.d et toutes dans L^2 avec (Cf. Exercice 93 page 61) :

$$\mathbf{E}(Z_1) = \langle t, m \rangle \quad \text{et} \quad \mathbf{V}(Z_1) = t^T \Gamma t.$$

Donc, le TCL uni-dimensionnel appliqué à cette suite $(Z_k)_k$ nous assure que :

$$W_n := \frac{S_n - n \langle t, m \rangle}{\sqrt{n}} \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} W \quad \text{avec } W \sim \mathcal{N}(0, t^T \Gamma t).$$

Ceci se traduit par :

$$\forall u \in \mathbf{R}, \quad \phi_{W_n}(u) \xrightarrow{n \rightarrow +\infty} \phi_W(u) = \exp\left(-\frac{u^2}{2} t^T \Gamma t\right). \quad (\star\star)$$

Or d'après (\star) , on remarque que l'on a : $\phi_{Y_n}(t) = \phi_{W_n}(1)$ et $\phi_Y(t) = \phi_W(1)$. On conclut alors grâce à la convergence $(\star\star)$ appliquée à $u = 1$. \square

▮

§ 4.4. Lois du Khi-deux et de Student

Les lois du Khi-deux et de Student sont des lois définies à partir des lois normales.

DEFINITION – 4.5 a) Si X est une v.a.r. de loi $\mathcal{N}(0, 1)$, on dit que la variable $Y = X^2$ suit la **loi du khi-deux à 1 degré de liberté** que l'on note $\chi^2(1)$.

b) Plus généralement : soient $d \geq 2$ un entier et X_1, \dots, X_d des v.a.r. indépendantes et de même loi $\mathcal{N}(0, 1)$.

La v.a.r. $\sum_{k=1}^d X_k^2$ suit la loi appelée **loi du khi-deux à d degrés de liberté**, que l'on note $\chi^2(d)$.

Voici des propriétés importantes des lois du Khi-Deux.

PROPOSITION – 4.12 (PROPRIÉTÉS DES LOIS DU KHI-DEUX.)

- (a) Etant donné $(d, d') \in (\mathbf{N}^*)^2$, si X et Y sont deux v.a.r. **indépendantes** avec $X \sim \chi^2(d)$ et $Y \sim \chi^2(d')$ alors leur somme suit encore une loi du Khi-Deux : $X + Y \sim \chi^2(d + d')$.
- (b) Si $X \sim \chi^2(1)$ alors sa fonction caractéristique vaut : $\phi_X(t) = \frac{1}{(1 - 2it)^{1/2}}$ pour tout $t \in \mathbf{R}$.
- (c) *Définition alternative* : pour tout entier $d \geq 1$, la loi du Khi-deux $\chi^2(d)$ est en fait la loi Gamma de paramètres $(d/2, 1/2)$ i.e. $\mathcal{G}(d/2, 1/2)$ aussi notée $\Gamma(d/2, 1/2)$.
- (d) Si $d \in \mathbf{N}^*$, la loi $\chi^2(d)$ a pour fonction caractéristique : $\phi(t) = \frac{1}{(1 - 2it)^{d/2}}$ pour tout $t \in \mathbf{R}$.

▮ PREUVE. Pour (a) : Voir COURS. Pour (b) à (d) : Cf. la correction de Exercice 65 (Question 5)) p. 37 du Chapitre 2. ▮

REMARQUE – 4.7 (DÉFINITION ALTERNATIVE ET DENSITÉ.) Grâce au fait que pour $d \in \mathbf{N}^*$, la loi du Khi-deux $\chi^2(d)$ est la loi $\mathcal{G}(d/2, 1/2)$ on peut noter que c'est une loi continue de densité :

$$f_X(x) = \frac{1}{2^{d/2}\Gamma(d/2)} x^{d/2-1} e^{-x/2} \mathbf{1}_{x>0}.$$

DEFINITION – 4.6 Soient $d \in \mathbf{N}^*$, X et Y deux v.a.r. indépendantes avec $X \sim \mathcal{N}(0, 1)$ et $Y \sim \chi^2(d)$.

La variable $T_d = \sqrt{d} \frac{X}{\sqrt{Y}}$ suit la loi dite **loi de Student à d degrés de liberté** que l'on note usuellement $\mathcal{T}(d)$.

REMARQUE – 4.8 Pour être complet, signalons que la loi $\mathcal{T}(d)$ est une loi continue sur \mathbf{R} qui admet une densité explicite qui dépend de d . Celle-ci est donnée par :

$$\forall t \in \mathbf{R}, \quad f_T(t) = \frac{1}{\sqrt{d\pi}} \frac{\Gamma((d+1)/2)}{\Gamma(d/2)} \left(1 + \frac{t^2}{d}\right)^{-\frac{d+1}{2}}.$$

Pour conclure ce chapitre, voici un résultat, connu sous le nom de *Théorème de FISHER*, que nous allons établir car celui-ci sera souvent utilisé en Statistique.

THÉORÈME – 4.4 (THÉORÈME DE FISHER.)

Soient $n \in \mathbf{N}$, X_1, \dots, X_n des v.a.r indépendantes et de même loi (i.i.d.) normale $\mathcal{N}(m, \sigma^2)$ avec $(m, \sigma) \in \mathbf{R} \times \mathbf{R}^{+*}$.

On définit les variables \bar{X}_n et S_n^2 par :

- $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ qui est appelée *moyenne empirique*,
- $S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$ dite *variance empirique corrigée*.

Alors on a les propriétés suivantes :

- a) Les variables \bar{X}_n et S_n^2 sont indépendantes.
- b) $\bar{X}_n \sim \mathcal{N}(m, \sigma^2/n)$ et $(n-1) \frac{S_n^2}{\sigma^2} \sim \chi^2(n-1)$.
- c) Ainsi la variable $T_n := \sqrt{n} \left(\frac{\bar{X}_n - m}{S_n} \right)$ suit la loi de Student $\mathcal{T}(n-1)$.

▮ PREUVE. Il existe au moins 2 démonstrations possibles (qui sont néanmoins liées). Nous allons en présenter une au travers de l'Exercice 104 ci-dessous. La seconde - plus classique (mais plus "boîte noire") repose sur un résultat général lié aux vecteurs gaussiens, le *Théorème de Cochran* : pour plus de détails, on pourra par exemple consulter le livre de Jean-Yves Ouvrard intitulé *Probabilités* (Tome 2), Cassini (2008). ▮

§ 4.5. Exercices supplémentaires

EXERCICE 99

Soit X une v.a.r. telle que $X \sim \mathcal{N}(0, 1)$. On considère Y la variable (égale au signe de X i.e.) donnée par : $Y = X/|X|$.

- Justifier que la variable Y est **P**-presque-sûrement bien définie.
- Déterminer la loi de Y .

►CORRECTION.

1) La variable $Y = X/|X|$ est **P**-presque-sûrement bien définie ssi la v.a.r. $|X|$ est **P**-presque-sûrement non nulle i.e. que $\mathbf{P}(|X| \neq 0) = 1$.

Ceci équivaut à vérifier (par passage au complémentaire) que : $\mathbf{P}(|X| = 0) = 0$.

Or, comme pour tout $x \in \mathbf{R}$, on a : $|x| = 0 \iff x = 0$, on a les égalités suivantes :

$$\{|X| = 0\} = \{X = 0\} \quad \text{et donc} \quad \mathbf{P}(|X| = 0) = \mathbf{P}(X = 0).$$

Or $\mathbf{P}(X = 0) = 0$ puisque X (qui suit la loi $\mathcal{N}(0, 1)$) est une v.a.r. continue!

$$2) \text{ On a : } x = \begin{cases} -x & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}. \text{ Donc pour tout } x \in \mathbf{R}^* : \frac{x}{|x|} = \begin{cases} -1 & \text{si } x < 0 \\ 1 & \text{si } x > 0. \end{cases}$$

On déduit du 1), que la v.a. Y prend **P**-presque-sûrement les valeurs -1 et 1 avec $\mathbf{P}(Y = 1) = \mathbf{P}(X > 0) = \mathbf{P}(X < 0) = \frac{1}{2}$ puisque la loi $\mathcal{N}(0, 1)$ est une loi continue et symétrique. Conclusion : $Y = X/|X|$ suit la loi de Rademacher de paramètre $1/2$.

EXERCICE 100 (MOMENTS DES LOIS NORMALES : EXISTENCE ET QUELQUES CALCULS)

- Soit X une v.a.r. qui suit la loi $\mathcal{N}(0, 1)$. Démontrer que X a des moments de tous ordres, puis que l'on a : pour tout $k \in \mathbf{N}$, $\mathbf{E}(X^{2k+1}) = 0$ et $\mathbf{E}(X^{2k}) = (2k)!/(2^k k!)$.
- On considère ici X une v.a.r. qui suit une loi normale $\mathcal{N}(m, \sigma^2)$ avec $(m, \sigma) \in \mathbf{R} \times \mathbf{R}^+$ quelconque. Justifier que X admet des moments de tous ordres (**N.B.** : on ne demande pas de les calculer!).

►CORRECTION.

Voir correction manuscrite.

EXERCICE 101 (LOI LOG-NORMALE)

La loi Log-normale est utilisée en linguistique pour modéliser le nombre de mots dans une phrase. On dit que X suit une loi Log-normale de paramètres m et $\sigma^2 > 0$ si $X = \exp(Y)$ avec Y de loi normale $\mathcal{N}(m, \sigma^2)$.

- En se ramenant à la loi $\mathcal{N}(0, 1)$, déterminer la densité de probabilité de X .
- Calculer de façon astucieuse l'espérance et la variance de X .
- Si $Z = X^a$ avec $a > 0$, montrer que Z suit une loi Log-normale.
- En déduire les moments entiers positifs de tous les ordres associés à X .

►CORRECTION.

Voir correction manuscrite.

EXERCICE 102

Soit $(\sigma_n^2)_{n \in \mathbf{N}}$ une suite de réels positifs ou nuls. On considère $(X_n)_{n \in \mathbf{N}}$ une suite de v.a.r. gaussiennes telles que pour tout entier $n \in \mathbf{N}$, $X_n \sim \mathcal{N}(0, \sigma_n^2)$. Démontrer l'équivalence suivante :

$$\text{la suite } (X_n)_n \text{ converge en loi} \iff \text{la suite } (\sigma_n^2) \text{ converge.}$$

►CORRECTION.

Voir correction manuscrite.

EXERCICE 103

On considère 3 v.a.r. U, V et W indépendantes avec $U \sim \mathcal{N}(1, 2)$ et $V \sim \mathcal{N}(0, 2)$ et $W \sim \mathcal{N}(-1, 3)$. On définit les variables X, Y et Z en posant :

$$X = U + V \quad \text{et} \quad Y = U - V + 1 \quad \text{et} \quad Z = -2U + W.$$

- Montrer que $(X, Y, Z)^T$ est un vecteur gaussien. Calculer sa moyenne m et sa matrice de covariance Γ .
- Déterminer la loi du vecteur $(X, Z)^T$.
- Déterminer (en justifiant votre réponse) la loi de la variable $S := \frac{(X-1)^2 + (Y-2)^2}{4}$.

►CORRECTION.

Voir cours

EXERCICE 104 (PREUVE DU THÉORÈME DE FISHER.)

On suppose que $(X_n)_{n \geq 1}$ une suite de v.a.r. indépendantes et de même loi $\mathcal{N}(m, \sigma^2)$ avec $m \in \mathbf{R}$ et $\sigma > 0$. On pose

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j, \quad \Sigma_n = \frac{1}{n} \sum_{j=1}^n (X_j - m)^2 \quad \text{et} \quad S_n^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2.$$

1. Déterminer les lois des variables \bar{X}_n , $U_n = \sqrt{n}(\bar{X}_n - m)/\sigma$ et $V_n = n\Sigma_n/\sigma^2$.
2. Le but de cette question est de montrer que les variables \bar{X}_n et S_n^2 sont indépendantes.
 - a) Montrer que le vecteur $(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n, \bar{X}_n)^T$ est un vecteur gaussien.
 - b) Montrer que pour tout $1 \leq j \leq n$, $\text{cov}(X_j - \bar{X}_n, \bar{X}_n) = 0$.
 - c) En déduire que la variable \bar{X}_n est indépendante du vecteur $(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)^T$. Puis conclure.
3. Montrer que : $(n-1)S_n^2 + n(\bar{X}_n - m)^2 = \sum_{j=1}^n (X_j - m)^2$.
4. En déduire l'expression de la fonction caractéristique de la variable $(n-1)S_n^2/\sigma^2$. Qu'en conclut-on ?
5. Puis, justifier que la variable $T_n := \sqrt{n} \left(\frac{\bar{X}_n - m}{S_n} \right)$ suit la loi de Student $\mathcal{T}(n-1)$.

►CORRECTION.

La correction de la question 2) est détaillée dans la correction manuscrite. Les autres questions ont été traitées en cours

EXERCICE 105

L'algorithme de Box-Muller permet la génération de variables aléatoires gaussiennes à partir de variables aléatoires de loi uniforme sur l'intervalle $[0, 1]$. Soient U et V deux variables aléatoires indépendantes et de loi uniforme sur $[0, 1]$. On pose

$$X = \sqrt{-2 \log U} \cos(2\pi V) \quad \text{et} \quad Y = \sqrt{-2 \log U} \sin(2\pi V).$$

Montrer que X et Y sont deux variables aléatoires indépendantes et de même loi $\mathcal{N}(0, 1)$.

►CORRECTION.

Voir correction manuscrite.

EXERCICE 106

Soit $(X, Y, Z)^T$ un vecteur aléatoire gaussien $\mathcal{N}(m, \Gamma)$ avec

$$m = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \quad \text{et} \quad \Gamma = \begin{pmatrix} 6 & 0 & 1 \\ 0 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix}.$$

- 1) Déterminer les lois des variables aléatoires $X, Y, Z, X + 3Y$ et $X + Y + Z$.
- 2) Déterminer les lois des vecteurs aléatoires $U = (X, Y)^T$ et $V = (2X - 5, 4Z + 6)^T$.

►CORRECTION.

Voir correction en cours

5 APPENDICE : RAPPELS SUR L'INTÉGRALE DE LEBESGUE.

On considère dans tout ce paragraphe un espace mesuré $(\Omega, \mathcal{A}, \mu)$. Nous rappelons la définition de mesure positive :

DEFINITION – 5.1 (MESURE POSITIVE) Une *mesure positive* μ sur l'espace mesurable (Ω, \mathcal{A}) est une application $\mu : \mathcal{A} \rightarrow [0, \infty]$ vérifiant :

- (i) $\mu(\emptyset) = 0$,
- (ii) (Additivité dénombrable ou σ -additivité) pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} **deux à deux disjoints** (i.e. $A_n \cap A_m = \emptyset$ pour tout $n \neq m$),

$$\mu \left(\bigcup_{n \in \mathbb{N}} A_n \right) = \sum_{n \in \mathbb{N}} \mu(A_n).$$

DEFINITION – 5.2 (FONCTIONS MESURABLES) Une application $f : \Omega \rightarrow (E, \mathcal{E})$ est dite *mesurable* si

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

On considèrera des fonctions mesurables à valeurs dans \mathbf{R} , \mathbf{C} ou \mathbf{R}^d , \mathbf{R} , \mathbf{C} et \mathbf{R}^d étant munis de leur tribu borélienne respective.

L'intégrale de Riemann, que vous connaissez depuis longtemps, est définie de la façon suivante : on commence par les *fonctions étagées* (ou *fonctions simples*) sur un segment, qui s'écrivent comme combinaison linéaire d'indicatrices. On voit ensuite que toute fonction continue sur ce même segment peut s'approcher par de telles fonctions.

On généralise ensuite à des intervalles quelconques en passant à la limite dans des intégrales du type $\int_a^x f(t) dt$ avec f continue.

Considérons à présent une fonction $f : \Omega \rightarrow \mathbf{R}$. Le même type de construction peut se faire pour ce type de fonction, en supposant une hypothèse de *mesurabilité*. L'objet ainsi construit sera appelé *intégrale de Lebesgue*. Pour $\mu = \lambda$, on retrouve notre brave intégrale de Riemann, ceci explique pourquoi la mesure de Lebesgue est parfois notée dx .

L'avantage de l'intégrale de Lebesgue est que l'on construit directement une intégration sur tout ensemble muni d'une tribu et pas seulement sur des segments ou une classe plus petite. La notion de mesurabilité étant plus faible que la continuité ceci n'est pas surprenant. Il n'y a donc pas d'*intégrales généralisées* à proprement parlé. Un autre avantage est que toute série peut s'écrire sous forme d'une intégrale de Lebesgue, on unifie donc les deux concepts (mais on pouvait déjà en voir le lien avec des théorèmes du type *comparaison série / intégrale de Riemann*).

§ 5.1. Pour les fonctions mesurables positives.

Supposons d'abord f de la forme $f = \mathbb{1}_A$ avec $A \in \mathcal{A}$, c'est une fonction mesurable par rapport à \mathcal{A} et $\mathcal{B}(\mathbf{R})$.

On appelle *intégrale de f contre μ* , notée $\int_{\Omega} f d\mu$, la quantité

$$\int_{\Omega} f d\mu = \int_{\Omega} \mathbb{1}_A(\omega) d\mu(\omega) := \mu(A),$$

et pour tout $B \in \mathcal{A}$

$$\int_B f d\mu := \mu(A \cap B).$$

Si μ est une probabilité sur Ω ,

$$\int_{\Omega} d\mu = \mu(\Omega) = 1.$$

En revanche dans le cas général, l'intégrale d'une telle fonction peut être éventuellement infinie : rappelons que les mesures sont à valeurs dans $[0, \infty]$ et non $[0, \infty[$. Et en fait ce sera le cas pour toutes les intégrales de fonctions mesurables positives.

EXEMPLE – 5.1 Si $\mathcal{A} = \mathcal{B}(\mathbf{R})$, on a pour $A =]a, b]$, $a < b$,

$$\int \mathbb{1}_A d\lambda = b - a = \int \mathbb{1}_A(x) dx.$$

Afin d'obtenir une propriété de linéarité pour notre intégrale, on l'étend de manière linéaire aux fonctions dites *étagées*. Si f a la forme $f(\omega) = \sum_{1 \leq i \leq n} a_i \mathbb{1}_{A_i}(\omega)$ avec A_i des ensembles mesurables disjoints, et $a_i \in \mathbf{R}$, on pose

$$\int_B f \, d\mu = \sum_{1 \leq i \leq n} a_i \mu(A_i \cap B) = \sum_{1 \leq i \leq n} a_i \int_B \mathbb{1}_{A_i} \, d\mu.$$

Pour que cette définition ait un sens il faudrait justifier que la décomposition choisie de f n'a aucune incidence sur son intégrale.

DEFINITION – 5.3 Soit f une fonction mesurable positive sur $(\Omega, \mathcal{A}, \mu)$. On appelle pour toute partie mesurable $B \in \mathcal{A}$ *intégrale de Lebesgue de f contre μ sur B* , la quantité

$$\int_B f \, d\mu = \sup \left\{ \int_B g \, d\mu : g \text{ étagée positive, } g \leq f \right\}.$$

REMARQUE – 5.1 Une définition plus concrète est la suivante : on écrit toute fonction mesurable positive comme une limite de fonctions étagées. Plus précisément, il existe une suite croissante de fonctions étagées convergeant simplement vers f . Elle est donnée par

$$f_n = \varphi_n \circ f,$$

où $\varphi_n : \mathbf{R}^+ \rightarrow \mathbf{R}^+$ est définie par

$$\varphi_n(t) = \begin{cases} \frac{1}{2^n} [2^n t] & \text{si } 0 \leq t < n \\ n & \text{si } t \geq n \end{cases}.$$

On pose alors

$$\int_B f \, d\mu = \lim_{n \rightarrow \infty} \int_B f_n \, d\mu.$$

En fait cela est vrai pour toute famille croissante de fonctions.

THÉORÈME – 5.1 (CONVERGENCE MONOTONE) Soit $(f_n)_{n \in \mathbf{N}}$ une suite croissante de fonctions mesurables positives sur $(\Omega, \mathcal{A}, \mathbf{P})$, convergeant simplement vers une fonction f . Alors f est mesurable positive elle aussi et

$$\int f_n \, d\mu \xrightarrow{n \rightarrow \infty} \int f \, d\mu.$$

Et comme corollaire classique, la version où f_n est une somme partielle de série de fonction.

COROLLAIRE – 5.1 (INTÉGRATION TERME À TERME MONOTONE) Soit $(f_n)_{n \in \mathbf{N}}$ une suite de fonctions mesurables positives sur $(\Omega, \mathcal{A}, \mathbf{P})$, et soit $f = \sum_{n \in \mathbf{N}} f_n$. Alors f est mesurable positive elle aussi et

$$\int f \, d\mu = \sum_{n \in \mathbf{N}} \int f_n \, d\mu.$$

§ 5.2. Pour des fonctions à valeurs réelles.

Comme pour l'intégrale de Riemann, on décompose f en sa partie positive et négative.

On rappelle que si $x \in \mathbf{R}$, $x^+ = \max(x, 0)$ et $x^- = -\min(x, 0)$. On reconstitue avec ces quantités l'identité et la valeur absolue :

$$x = x^+ - x^-, \quad |x| = x^+ + x^-.$$

DEFINITION – 5.4 Soit $f = f^+ - f^-$ une fonction mesurable décomposée en sa partie positive et sa partie négative. On dit que f est μ -intégrable sur $B \in \mathcal{A}$ si

$$\int_B |f| \, d\mu < \infty.$$

Dans ce cas on définit l'intégrale de f sur B par

$$\int_B f \, d\mu = \int_B f^+ \, d\mu - \int_B f^- \, d\mu. \quad (5.1)$$

Alors que l'intégrale d'une fonction positive est tout le temps définie, pour une fonction quelconque on impose la condition d'intégrabilité mentionnée au-dessus. Elle implique notamment que $\int_B f^+ \, d\mu < \infty$ et $\int_B f^- \, d\mu < \infty$: on exclut alors le cas "+ ∞ - ∞ " dans l'égalité (5.2).

⚠ Attention. Attention aussi au vocabulaire : pour une fonction mesurable positive, même si l'intégrale éventuellement infinie existe toujours, elle est dite *intégrable* si elle l'est au sens de la définition précédente, c'est-à-dire si elle est d'intégrale finie.

Il y a donc une nuance entre « être intégrable », et « l'intégrale existe »

Voici une liste de propriétés non exhaustive :

PROPOSITION – 5.1 soient f, g deux fonctions mesurables sur (Ω, \mathcal{A}) , et μ, ν deux mesures positives. Soient aussi $(\alpha, \beta) \in \mathbf{R}^2$ et $A, B \in \mathcal{A}$ deux parties mesurables. Alors :

- (i) (Croissance) si $f \leq g$, $\int_B f \, d\mu \leq \int_B g \, d\mu$,
- (ii) (Monotonie) si $A \subset B$, et $f \geq 0$, alors $\int_A f \, d\mu \leq \int_B f \, d\mu$,
- (iii) (Linéarité) $\int_B (\alpha f + \beta g) \, d\mu = \alpha \int_B f \, d\mu + \beta \int_B g \, d\mu$,
- (iv) (Nullité) si $f = 0$ p.s. alors $\int_B f \, d\mu = 0$,
- (v) si B est μ -négligeable, alors $\int_B f \, d\mu = 0$,
- (vi) si $f \geq 0$, alors $\int_B f \, d\mu = \int f \mathbf{1}_B \, d\mu$,
- (vii) $\alpha, \beta \geq 0$, $\int_B f \, d(\alpha\mu + \beta\nu) = \alpha \int_B f \, d\mu + \beta \int_B f \, d\nu$.

REMARQUE – 5.2 (LIEN SÉRIE/INTÉGRALE) Soit f une fonction mesurable sur $(\Omega, \mathcal{A}, \mu)$ à valeurs dans $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. Alors f est donc intégrable par rapport à $\mu = \delta_\omega$ (masse de DIRAC), et

$$\int f \, d\delta_\omega = f(\omega).$$

Pour justifier complètement cette formule, commencer avec des fonctions étagées puis étendre par linéarité/passage à la limite à toute fonction mesurable. Plus généralement par linéarité d'après (vii), si $\mu = \sum_{1 \leq i \leq n} a_i \delta_{\omega_i}$ avec $a_i \geq 0$ et $\omega_i \in \Omega$, alors on a

$$\int f \, d\mu = \sum_{i=1}^n a_i f(\omega_i).$$

§ 5.3. Pour des fonctions à valeurs complexes.

Dans ce paragraphe, $|\cdot|$ désigne le module.

DEFINITION – 5.5 Soit $f = \text{Ré}(f) + i\text{Im}(f)$ une fonction mesurable à valeurs complexes, c'est-à-dire telle que $\text{Ré}(f), \text{Im}(f)$ soient mesurables. On dit que f est μ -intégrable sur $B \in \mathcal{A}$ si

$$\int_B |f| \, d\mu < \infty.$$

Dans ce cas on définit l'intégrale de f sur B par

$$\int_B f \, d\mu = \int_B \text{Ré} f \, d\mu + i \int_B \text{Im} f \, d\mu. \quad (5.2)$$

Citons pour terminer un autre théorème de convergence très important. Le résultat est le même que celui de convergence monotone, mais le jeu d'hypothèses différent.

THÉORÈME – 5.2 (CONVERGENCE DOMINÉE) Soit $(f_n)_{n \in \mathbf{N}}$ une suite de fonctions mesurables convergeant simplement vers une fonction f . On suppose en plus une *hypothèse de domination* : il existe une fonction positive g intégrable telle que $|f_n| \leq g$ μ -ps.

Alors :

$$\int f_n d\mu \xrightarrow{n \rightarrow \infty} \int f d\mu.$$

On note $f_n \xrightarrow{L^1} f$, et on dit que $(f_n)_{n \in \mathbf{N}}$ converge vers f dans L^1 .

THÉORÈME – 5.3 (CONVERGENCE DOMINÉE, VERSION ESPÉRANCE) Soit $(X_n)_{n \in \mathbf{N}}$ une suite de variable aléatoires convergeant presque-sûrement vers une variable aléatoire X . On suppose en plus une *hypothèse de domination* : il existe une variable aléatoire intégrable Y , telle que $|X_n| \leq Y$ \mathbf{P} -ps.

Alors :

$$\mathbf{E}(X_n) \xrightarrow{n \rightarrow \infty} \mathbf{E}(X).$$

En particulier, si X_n est bornée uniformément en n , i.e. s'il existe une constante $M \in \mathbf{R}^+$ telle que $|X_n| \leq M$ \mathbf{P} -ps, alors l'hypothèse de domination est vérifiée. Attention, ceci n'est pas vrai sur un espace mesuré général : cela vient du fait qu'une constante est intégrable dans le cas d'un espace probabilisé

$$\mathbf{E}(M) = M\mathbf{E}(1) = M\mathbf{P}(\Omega) = M,$$

pour tout $M \in \mathbf{R}^+$.

§ 5.4. Pour des fonctions vectorielles.

On considère un entier $d \in \mathbf{N}^*$, et $|\cdot|$ la norme Euclidienne sur \mathbf{R}^d .

DEFINITION – 5.6 Soit $f = (f_1, \dots, f_d)$ une fonction mesurable (i.e. chaque coordonnée est mesurable) à valeurs dans \mathbf{R}^d . On dit que f est μ -intégrable sur $B \in \mathcal{A}$ si

$$\int_B |f| d\mu < \infty.$$

Dans ce cas on définit l'intégrale de f sur B comme le vecteur

$$\int_B f d\mu = \left(\int_B f_1 d\mu, \dots, \int_B f_d d\mu \right) \quad (5.3)$$

Par théorème d'équivalence des normes en dimension finie, la condition d'intégrabilité est équivalente à l'intégrabilité de chacune des coordonnées au sens des fonctions à valeurs réelles.

§ 5.5. Espaces fonctionnels L^p .

Pour terminer ces rappels regardons les espaces de fonctions intégrables. On rappelle sans démonstration quelques propriétés importantes.

On considère un espace mesuré $(\Omega, \mathcal{A}, \mathbf{P})$ et des fonctions à valeurs complexes ou dans un \mathbf{R}^d avec $d \geq 1$.

Soit $p \geq 1$.

DEFINITION – 5.7 On appelle $\mathcal{L}^p(\Omega, \mu)$ l'espace défini par

$$\mathcal{L}^p(\Omega, \mu) = \left\{ f : \Omega \rightarrow \mathbf{C} \text{ ou } \mathbf{R}^d, \quad \|f\|_p < \infty \right\},$$

avec

$$\|f\|_p = \left(\int_{\Omega} |f|^p d\mu \right)^{1/p}.$$

Par ailleurs, on appelle $\mathcal{L}^\infty(\Omega, \mu)$ l'espace défini par

$$\mathcal{L}^\infty(\Omega, \mu) = \left\{ f : \Omega \rightarrow \mathbf{C} \text{ ou } \mathbf{R}^d, \quad \|f\|_\infty < \infty \right\},$$

avec

$$\|f\|_\infty = \mu - \text{esssup } f := \inf \{ M \geq 0 : |f(x)| \leq M \quad \mu - \text{ps} \}.$$

Ici $|\cdot|$ désigne soit le module soit la norme euclidienne de \mathbf{R}^d suivant la classe de fonctions considérée. Parfois si le contexte est clair on note simplement \mathcal{L}^p au lieu de $\mathcal{L}^p(\Omega, \mu)$.

On considère généralement que des $p \geq 1$ pour la raison suivante : les espaces précédents sont bien définis même si $p < 1$, mais ils ne sont pas des espaces vectoriels. Il est facile de vérifier que si f est dans \mathcal{L}^p alors pour tout $\alpha \in \mathbf{C}$ ou $\alpha \in \mathbf{R}$, αf est mesurable et que $\|\alpha f\|_p = |\alpha| \|f\|_p$. La seule propriété non triviale est la stabilité par somme :

$$f, g \in \mathcal{L}^p(\Omega, \mathcal{A}) \implies f + g \in \mathcal{L}^p(\Omega, \mathcal{A}),$$

par définition cela revient à vérifier l'inégalité triangulaire pour $\|\cdot\|_p$.

L'inégalité suivante d'Hölder répond au problème de l'inégalité triangulaire, valable pour $p \geq 1$ **uniquement**.

PROPOSITION – 5.2 (INÉGALITÉ D'HÖLDER) Pour f, g mesurables, on a

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

pour tout couple (p, q) d'exposants conjugués, i.e. tel que $\frac{1}{p} + \frac{1}{q} = 1$. En particulier, si $p = \infty$ (et $q = 1$), on a

$$\|fg\|_1 \leq \|f\|_1 \|g\|_\infty.$$

Notez que pour $p = 2$ et $q = 2$ on retrouve l'inégalité de Cauchy-Schwarz.

COROLLAIRE – 5.2 (INÉGALITÉ DE MINKOWSKI) Pour f, g mesurables

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

On se rapproche tout doucement des axiomes d'une norme, mais il en reste un qui n'est pas encore très clair : celui de *séparation*. A-t-on

$$\|f\|_p = 0 \implies f = 0 \quad ?$$

La réponse est « pas tout à fait ». *A priori*, nous avons seulement d'après un résultat classique d'intégration :

$$\|f\|_p = 0 \implies f = 0 \quad \mu - \text{ps}.$$

Pour pallier à ce problème, on considère plutôt $\mathcal{L}^p \setminus \sim$ où \sim désigne la relation d'équivalence μ -presque sûre définie par

$$f \sim g \iff f = g \quad \mu - \text{ps}.$$

DEFINITION – 5.8 On appelle $L^p(\Omega, \mu)$ l'espace défini par

$$L^p(\Omega, \mu) = \mathcal{L}^p(\Omega, \mu) \setminus \sim.$$

PROPOSITION – 5.3 (RIESZ-FISCHER) Pour tout $p \geq 1$, $(L^p(\Omega, \mu), \|\cdot\|_p)$ est un espace vectoriel normé complet. C'est donc un Banach.

En pratique on écrit seulement $f \in L^p$ au lieu de considérer les classes de fonctions égales presque-sûrement et de noter $\bar{f} \in L^p$. De-même on note $\|f\|_p$ pour $\|\bar{f}\|_p$. Par ailleurs le choix du représentant d'une classe d'équivalence ne change pas la valeur de l'intégrale.

REMARQUE – 5.3 (VERSION PROBABILISTE) Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace probabilisé. Une variable aléatoire X sur Ω est donc dans L^p si

$$\mathbf{E}(|X|^p)^{1/p} < \infty,$$

en particulier : si X est bornée alors $X \in L^p$ pour tout $p \geq 1$. Attention, ceci n'est pas vrai dans le cas d'un espace mesuré général : cela vient du fait comme déjà dit qu'une constante est intégrable dans le cas d'un espace probabilisé

$$\mathbf{E}(M) = M\mathbf{E}(1) = M\mathbf{P}(\Omega) = M,$$

pour tout $M \in \mathbf{R}^+$.

THÉORÈME – 5.4 (CONVERGENCE DOMINÉE DANS L^p) Soit $(f_n)_{n \in \mathbf{N}}$ une suite de fonctions mesurables dans L^p avec $p < \infty$ convergeant simplement vers une fonction f . On suppose en plus une *hypothèse de domination* : il existe une fonction positive $g \in L^p$ telle que $|f_n| \leq g$ μ -ps.

Alors :

$$\int |f_n - f|^p d\mu \xrightarrow{n \rightarrow \infty} 0,$$

on note $f_n \xrightarrow[n \rightarrow \infty]{L^p} f$, et on dit que $(f_n)_{n \in \mathbf{N}}$ converge vers f dans L^p .

Cette version du théorème de convergence dominée est une conséquence directe de la première. Pour terminer, précisons les inclusions entre ces différents espaces. Elles sont valables sur tout espace $(\Omega, \mathcal{A}, \mu)$ où

$$\mu(\Omega) < \infty,$$

donc en particulier sur les espaces probabilisés.

PROPOSITION – 5.4 (INCLUSIONS) Si $1 \leq p < q \leq \infty$, alors $L^q \subset L^p$. De plus, les normes $\|\cdot\|_p$ et $\|\cdot\|_q$ sont équivalentes sur L^q .

Par ailleurs si f et $(f_n)_{n \in \mathbf{N}}$ sont respectivement une fonction et suite de fonctions mesurables telles que $f_n \xrightarrow[n \rightarrow \infty]{L^q} f$, alors $f_n \xrightarrow[n \rightarrow \infty]{L^p} f$.

Donnons simplement l'inégalité qui permet de le démontrer : par l'inégalité d'Hölder on a pour toute fonction mesurable $f \in L^q$,

$$\int_{\Omega} |f|^p d\mu \leq \left(\int_{\Omega} 1^{\frac{q}{q-p}} d\mu \right)^{\frac{q-p}{q}} \left(\int_{\Omega} |f|^q d\mu \right)^{\frac{p}{q}} = \mu(\Omega)^{\frac{q-p}{q}} \|f\|_q^p.$$

REMARQUE – 5.4 On connaît beaucoup d'autres choses encore sur les espaces L^p : leur dual, des parties denses, la régularité d'opérateurs de translation, ... Ici nous citons les propriétés importantes uniquement pour le cours de probabilités.

Par ailleurs L^2 est souvent étudié plus en détail car sa norme dérive d'un produit scalaire :

$$\|f\|_2^2 = \int_{\Omega} |f|^2 d\mu.$$