Probabilités

Licence de Mathématiques

Table des matières

Chapitre 1

Bla-bla introductif

1. Vocabulaire intuitif

1.1. Expériences aléatoires et évènements.

• *Expérience aléatoire* : n'importe quelle "expérience" (réelle ou fictive) dont on ne peut pas dire avec certitude ce qu'elle va donner.

• *Résultat d'une expérience* : la chose qui nous intéresse dans l'expérience.

Exemple. Vous entrez dans un bar à bières bien achalandé, et vous buvez une pinte de toutes les bières proposées, pour savoir laquelle vous préférez. Le résultat n'est pas votre taux d'alcoolémie à la fin de cette dégustation, mais le nom de la bière que vous avez préféré.

• *Évènement relatif à une expérience* : n'importe quel évènement imaginable dont on peut dire avec certitude s'il se produit ou non dès lors qu'on connait le résultat de l'expérience.

Remarque. Un évènement est toujours décrit par une phrase du langage courant ; et *on identifie l'évènement avec la phrase qui le décrit.*

Exemples. L'évènement "vous allez vomir" *n'est pas* un évènement relatif à l'expérience précédente, car sa réalisation ne dépend pas du résultat de l'expérience (le nom de la bière que vous avez préféré) ; mais l'évènement "la bière que vous avez préférée est belge" en est un, dès lors que vous connaissez la provenance de toutes les bières que vous avez testées.

• *Évènement élémentaire* : tout évènement de la forme "on obtient ω", où ω est un résultat possible de l'expérience. Notation : E_ω.

• *Évènement certain* : un évènement qui se produit toujours, quel que soit le résultat de l'expérience.

Exemple. Si on lance 2 dés, l'évènement "la somme des chiffres est au moins égale à 2" est certain.

• *Évènement impossible* : un évènement qui ne se produit jamais, quel que soit le résultat de l'expérience.

Exemple. Si on lance 2 dés, l'évènement "la somme des chiffres est multiple de 13" est impossible.

• *Conjonction, disjonction* : si $E_1, \ldots, E_n$ sont des évènements, on note $E_1 \wedge \cdots \wedge E_n$ l'évènement "E_1 et E_2 et ... et E_n" (conjonction), et $E_1 \vee \cdots \vee E_n$ l'évènement "E_1 ou ... ou E_n" (disjonction).

• *Évènement contraire* : si E est un évènement, on note $\neg E$ l'évènement "non E". De façon précise, si E est décrit par une certaine phrase Φ, alors $\neg E$ est l'évènement décrit

par la *négation* de la phrase Φ. Ainsi, par définition, $\neg E$ se produit si et seulement si E ne se produit pas. (En particulier, E est un évènement certain si et seulement si $\neg E$ est impossible.)

• *Évènements incompatibles* : deux évènements E et F qui ne peuvent pas se produire en même temps, *i.e.* tels que $E \wedge F$ est impossible. Exemples : si ω et ω' sont deux résultats possibles différents, alors les évènements élémentaires E_ω et $E_{\omega'}$ sont incompatibles.

• *Évènements équivalents* : deux évènements E et E' tels que, quel que soit le résultat de l'expérience, E se produit si et seulement si E' se produit. Autrement dit E est incompatible avec $\neg E'$ et E' est incompatible avec $\neg E$. Notation : $E \equiv E'$.

Exemple. Un supporter de Lens et un supporter du PSG regardent ensemble la finale de la coupe de France, qui oppose Lens au PSG. Chacun pleure si son équipe favorite perd, et se moque de l'autre si elle gagne. Les évènements "le supporter du PSG pleure" et "le supporter de Lens se moque du supporter du PSG" sont équivalents.

Exercice. Montrer que s'il n'y a qu'un nombre fini de résultats possibles, alors tout évènement E est équivalent à une disjonction $E_{\omega_1} \vee \cdots \vee E_{\omega_N}$ d'évènements élémentaires, unique à l'ordre près. (On dit que ces évènements élémentaires "composent" E.)

1.2. Probabilité. Intuitivement, la *probabilité* d'un évènement E relatif à une expérience aléatoire est le "degré de plausibilité" qu'on accorde à E, mesuré sur une échelle allant de 0 à 1 :

$$\mathrm{Prob}(E) = \frac{\%\text{ de chances que } E \text{ se produise}}{100}.$$

Difficultés.

- Ce n'est pas une définition précise.
- La probabilité dépend fortement de l'"observateur".
- Il n'est pas clair qu'un observateur soit en mesure d'attribuer de façon exacte un degré de plausibilité à n'importe quel évènement.

Fait 1. La probabilité de tout évènement E est un nombre compris entre 0 et 1. On a $\mathrm{Prob}(E) = 1$ si E est certain, et $\mathrm{Prob}(E) = 0$ si E est impossible.

Fait 2. Les probabilités s'ajoutent : si E et F sont deux évènements *incompatibles*, alors

$$\mathrm{Prob}(E \vee F) = \mathrm{Prob}(E) + \mathrm{Prob}(F).$$

Fait 3. Deux évènements équivalents ont la même probabilité.

Conséquence. *S'il n'y a qu'un nombre fini de résultats possibles*, alors *la somme des probabilités de* tous *les évènements élémentaires doit être égale à* 1.

Démonstration. Si on note $\omega_1, \ldots, \omega_N$ les résultats possibles (supposés 2 à $2 \neq$), alors l'évènement $E := E_{\omega_1} \vee \cdots \vee E_{\omega_N}$ est certain, donc $\mathrm{Prob}(E) = 1$; mais les E_{ω_i} sont incompatibles, donc on a aussi $\mathrm{Prob}(E) = \mathrm{Prob}(E_{\omega_1}) + \cdots + \mathrm{Prob}(E_{\omega_N})$. □

Remarque. Plus généralement, la probabilité d'un évènement E est la somme des probabilités des évènements élémentaires qui le composent.

2. Un exemple très simple

QUESTION. Quand on jette 2 dés, y a-t-il plus de chances que la somme des chiffres fasse 7 ou bien qu'elle fasse 8 ?

• L'expérience considérée : jeter 2 dés.

• Le résultat : deux chiffres entre 1 et 6.

• Évènement élémentaire typique : “on obtient (i et j)”, où i et j sont deux chiffres entre 1 et 6. Notation : $E_{i,j}$. (Donc $E_{i,j} = E_{j,i}$.)

• Les évènements qui nous intéressent : “la somme des chiffres fait 8” (noté S_8) ; “la somme des chiffres fait 7” (noté S_7). On a

$$S_7 \equiv E_{1,6} \vee E_{2,5} \vee E_{3,4} \qquad \text{et} \qquad S_8 \equiv E_{2,6} \vee E_{3,5} \vee E_{4,4}.$$

• Probabilités.

- Intuitivement, il y a 2 fois plus de chances d'obtenir un résultat mixte (deux chiffres différents) plutôt qu'un “double”.
- Tous les “doubles” ont la même probabilité p, et tous les résultats mixtes ont la même probabilité $q = 2p$.
- Il y a 6 doubles possibles et 15 résultats mixtes possibles (exo).
- On doit donc avoir $6p + 15q = 1$, i. e. $6p + 30p = 1$; d'où $p = \frac{1}{36}$ et $q = \frac{1}{18}$.

• La question qui nous intéresse : on a $\text{Prob}(S_7) = \text{Prob}(E_{1,6}) + \text{Prob}(E_{2,5}) + \text{Prob}(E_{3,4}) = 3 \times \frac{1}{18} = \frac{1}{6}$, et $\text{Prob}(S_8) = \text{Prob}(E_{2,6}) + \text{Prob}(E_{3,5}) + \text{Prob}(E_{4,4}) = 2 \times \frac{1}{18} + \frac{1}{36} = \frac{5}{36}$.
Conclusion : il y plus de chances que la somme fasse 7 plutôt que 8.

3. “Modélisation”

On considère une certaine expérience aléatoire $\boldsymbol{\mathcal{E}}$. On cherche à *modéliser* $\boldsymbol{\mathcal{E}}$, i. e. à trouver un cadre mathématique qui la décrit de façon à peu près satisfaisante. Pour cela, on peut procéder comme suit.

• On choisit un ensemble “mathématique” Ω dont les éléments *s'identifient de manière satisfaisante* aux résultats possibles de l'expérience.

• Pour tout évènement E relatif à $\boldsymbol{\mathcal{E}}$, on pose

$$\widetilde{E} = \{\omega \in \Omega;\ \text{ l'évènement } E \text{ se produit si l'expérience donne le “résultat” } \omega\}.$$

Exercice. Pourquoi y a-t-il des guillemets à “résultat” ?

FAIT. Deux évènements E et E' sont équivalents si et seulement si $\widetilde{E} = \widetilde{E'}$.

Démonstration. Micro-exo. □

• On note $\mathfrak{A}$ la famille de toutes les parties A de Ω de la forme $A = \widetilde{E}$, où E est un évènement relatif à l'expérience $\boldsymbol{\mathcal{E}}$.

• On définit $\mathbb{P} : \mathfrak{A} \to [0, 1]$ par

$$\mathbb{P}(\widetilde{E}) = \text{Prob}(E) \qquad \text{pour tout évènement } E.$$

Il n'y a pas d'ambiguité : si $\widetilde{E} = \widetilde{E'}$, alors les deux évènements E et E' sont équivalents et donc $\text{Prob}(E) = \text{Prob}(E')$.

PROPRIÉTÉS.

- Si E est un évènement impossible, alors $\widetilde{E} = \varnothing$, et si E est certain, alors $\widetilde{E} = \Omega$. Donc : $\mathfrak{A}$ contient $\varnothing$ et Ω.

- Si E et F sont des événements, alors

$$\widetilde{E \vee F} = \widetilde{E} \cup \widetilde{F}\,, \quad \widetilde{E \wedge F} = \widetilde{E} \cap \widetilde{F} \quad \text{et} \quad \widetilde{\neg E} = \widetilde{E}^c.$$

Donc : la famille $\mathfrak{A}$ doit être stable par unions finies, par intersections finies et par complémentation ; autrement dit, si $A, B \in \mathfrak{A}$, alors $A \cup B \in \mathfrak{A}$, $A \cap B \in \mathfrak{A}$ et $A^c \in \mathfrak{A}$. On dit que $\mathfrak{A}$ est une **algèbre** de parties de Ω.

- Si $\omega \in \Omega$, alors $\widetilde{E_\omega} = \{\omega\}$. Donc : la famille $\mathfrak{A}$ contient tous les singletons.

- Un évènement E impossible a pour probabilité 0, et un évènement certain a pour probabilité 1. Donc : $\mathbb{P}(\varnothing) = 0$ et $\mathbb{P}(\Omega) = 1$.

- Deux évènements E et F sont incompatibles si et seulement si $\widetilde{E} \cap \widetilde{F} = \varnothing$. Donc : si $A, B \in \mathfrak{A}$, alors

$$A \cap B = \varnothing \implies \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

On dit que la "fonction d'ensembles" $\mathbb{P}$ est **finiment additive**.

En résumé. Il est raisonnable de chercher à modéliser une expérience aléatoire par un triplet $(\Omega, \mathfrak{A}, \mathbb{P})$, où Ω est un ensemble non vide, $\mathfrak{A}$ est une algèbre de parties de Ω contenant tous les singletons, et $\mathbb{P} : \mathfrak{A} \to [0,1]$ est une fonction d'ensembles finiment additive telle que $\mathbb{P}(\Omega) = 1$.

4. Retour à l'"exemple très simple"

Une modélisation.

- On peut identifier un résultat possible "i et j" avec la paire $(i,j) \in [\![1,6]\!] \times [\![1,6]\!]$. Mais "$i$ et j" et "j et i" sont le même résultat, donc on ne doit pas mettre à la fois (i,j) et (j,i) dans Ω. Par exemple, on peut décider qu'on écrit le plus petit chiffre en premier, *i.e.* qu'on ne prend que les (i,j) tels que $i \leqslant j$. Bref : on peut prendre

$$\Omega = \big\{(i,j) \in [\![1,6]\!] \times [\![1,6]\!];\ i \leqslant j\big\}.$$

- On a $\mathbb{P}(\{(i,i)\}) = \text{Prob}(\text{"double } i\text{"}) = \frac{1}{36}$ pour tout i, et $\mathbb{P}(\{(i,j)\}) = \frac{1}{18}$ pour $i < j$.

- On a $\widetilde{S_7} = \{(1,6),(2,5),(3,4)\}$ et $\widetilde{S_8} = \{(2,6),(3,5),(4,4)\}$. Donc

$$\text{Prob}(S_7) = \mathbb{P}(\widetilde{S_7}) = \mathbb{P}\big(\{(1,6),(2,5),(3,4)\}\big) = 3 \times \frac{1}{18} = \frac{1}{6},$$

et de même $\text{Prob}(S_8) = \mathbb{P}(\widetilde{S_8}) = \mathbb{P}\big(\{(2,6),(3,5),(4,4))\}\big) = 2 \times \frac{1}{18} + \frac{1}{36} = \frac{5}{36}$.

Une autre modélisation.

- On imagine qu'on lance les 2 dés l'un après l'autre (ce qui est une expérience "équivalente"). Intérêt : de cette façon, on *différencie* les 2 dés. Donc, on peut prendre

$$\Omega = [\![1,6]\!] \times [\![1,6]\!],$$

où une paire $(i,j) \in \Omega$ correspond à l'évènement " i pour le 1er dé et j pour le 2ème".

- Avec ce choix de Ω, tous les évènements élémentaires deviennent équiprobables ; donc $\mathbb{P}(\{(i,j)\}) = \frac{1}{36}$ pour tout $(i,j) \in \Omega$.

- On en déduit

$$\mathrm{Prob}(S_7) = \mathbb{P}\big(\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}\big) = 6 \times \frac{1}{36} = \frac{1}{6},$$

et de même $\mathrm{Prob}(S_8) = \mathbb{P}\big(\{(2,6),(3,5),(4,4),(5,3),(6,2)\}\big) = \frac{5}{36}$.

5. Hypothèses supplémentaires

Pour que le modèle $(\Omega, \mathfrak{A}, \mathbb{P})$ soit mathématiquement plus "performant", on va en fait demander plus à $\mathfrak{A}$ et à $\mathbb{P}$: on exigera que $\mathfrak{A}$ soit stable par réunions et intersections *dénombrables*, et que $\mathbb{P}$ soit *dénombrablement additive*. Autrement dit, on demande que $\mathfrak{A}$ soit une *tribu* et que $\mathbb{P}$ soit une *mesure* sur $(\Omega, \mathfrak{A})$.

Exercice. Soit Ω un ensemble fini. Montrer que toute algèbre $\mathfrak{A}$ de parties de Ω est en fait une tribu, et que toute fonction finiment additive $\mu : \mathfrak{A} \to [0, \infty]$ telle que $\mu(\varnothing) = 0$ est en fait une mesure.

Chapitre 2

Espaces de probabilité

1. Généralités

1.1. Lois de probabilité.

DÉFINITION 1.1. *Un* ***univers probabilisable*** *est un ensemble non vide* Ω *muni d'une tribu* $\mathfrak{A}$ *contenant tous les singletons. Les ensembles mesurables (i.e. les éléments de la tribu* $\mathfrak{A}$*) s'appellent des* ***évènements****. Les singletons* $\{\omega\}$*,* $\omega \in \Omega$ *s'appellent les* ***évènements élémentaires****. Une* ***loi de probabilité*** *sur* $(\Omega, \mathfrak{A})$ *est une mesure positive* $\mathbb{P}$ *telle que* $\mathbb{P}(\Omega) = 1$*. On dit que* $(\Omega, \mathfrak{A}, \mathbb{P})$ *est un* ***espace de probabilité****. Si* A *est un évènement, alors* $\mathbb{P}(A)$ *s'appelle la* ***probabilité de*** A*.*

Remarque 0. Imposer que la tribu $\mathfrak{A}$ contienne tous les singletons est une "commodité technique".

Remarque 1. Comme la tribu $\mathfrak{A}$ contient tous les singletons, elle contient tous les ensembles *dénombrables*.

Remarque 2. Si la tribu $\mathfrak{A}$ est clairement spécifiée par le contexte, on parle de loi de probabilité *sur* Ω. En particulier :

- pour un Ω *dénombrable*, on a toujours $\mathfrak{A} = \mathcal{P}(\Omega)$;
- pour $\Omega = \mathbb{R}^d$, on prendra toujours $\mathfrak{A} = \mathfrak{B}(\mathbb{R}^d)$, la tribu borélienne de $\mathbb{R}^d$.

Remarque 3. Si E est un évènement, alors

$$\mathbb{P}(E^c) = 1 - \mathbb{P}(E).$$

Remarque 4. Comme $\mathbb{P}$ est une mesure *finie*, on a

$$\mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \lim_{n\to\infty} \mathbb{P}(A_n) = \sup_n \mathbb{P}(A_n) \quad \text{pour toute suite } \textit{croissante } (A_n) \subseteq \mathfrak{A},$$

et

$$\mathbb{P}\Big(\bigcap_{n\in\mathbb{N}} B_n\Big) = \lim_{n\to\infty} \mathbb{P}(B_n) = \inf_n \mathbb{P}(B_n) \quad \text{pour toute suite } \textit{décroissante } (B_n) \subseteq \mathfrak{A}.$$

1.2. Évènements presque sûrs.

DÉFINITION 1.2. *Soit* $(\Omega, \mathfrak{A}, \mathbb{P})$ *un espace de probabilité. Un évènement* E *est dit* ***presque sûr*** *si* $\mathbb{P}(E) = 1$ *; autrement dit si* $\mathbb{P}(E^c) = 0$ *(puisque* $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$*). Une propriété* (P) *relative aux éléments de* Ω *est dite* ***vraie presque sûrement*** *si l'ensemble des* $\omega \in \Omega$ *vérifiant* (P) *est un évènement presque sûr.*

Remarque. On écrira "ps" au lieu de "presque sûrement".

FAIT ESSENTIEL. Une intersection dénombrable d'évènements presque sûrs est encore un évènement presque sûr.

Démonstration. Si $(E_n)_{n\in\mathbb{N}}$ est une suite d'évènements presque sûrs et si on pose $E = \bigcap_0^\infty E_n$, alors $E^c = \bigcup_0^\infty E_n^c$ et donc $\mathbb{P}(E^c) \leqslant \sum_0^\infty \mathbb{P}(E_n^c) = 0$. □

REFORMULATION. Une *conjonction dénombrable* de propriétés presque sûres est encore une propriété presque sûre. Autrement dit, si $(\mathrm{P}_n)_{n\in\mathbb{N}}$ est une suite de propriétés dépendant de $\omega \in \Omega$ et si on sait que

$$\forall n \in \mathbb{N} \ : \ \big(\mathrm{P}_n(\omega) \text{ est vraie} \quad \text{ps}\big),$$

alors on peut conclure que

$$\big(\forall n \in \mathbb{N} \ : \ \mathrm{P}_n(\omega) \text{ est vraie}\big) \qquad \text{ps}.$$

1.3. Intégration. Comme les lois de probabilité sont des mesures, toute la théorie de l'intégration s'applique.

FAIT IMPORTANT. Si $(\Omega, \mathfrak{A}, \mathbb{P})$ est un espace de probabilité, alors toute fonction mesurable *bornée* $f : \Omega \to \mathbb{R}$ est intégrable par rapport à $\mathbb{P}$.

Démonstration. On a $\int_\Omega |f| \, d\mathbb{P} \leqslant \|f\|_\infty \times \mathbb{P}(\Omega) < \infty$. □

1.4. Classes monotones. Dans ce qui suit, Ω est un ensemble non vide.

DÉFINITION 1.3. *Soit $\mathcal{M}$ une famille de parties de Ω. On dit que $\mathcal{M}$ est une* ***classe monotone*** *si*

- $\Omega \in \mathcal{M}$;
- $(A, B \in \mathcal{M} \text{ et } A \subseteq B) \implies (B \backslash A \in \mathcal{M})$;
- $\bigcup_{n=0}^\infty A_n \in \mathcal{M}$ *pour toute suite croissante* $(A_n)_{n\in\mathbb{N}}$ *d'éléments de* $\mathcal{M}$.

EXEMPLE. On prend $\Omega = \mathbb{R}$. Si μ_1 et μ_2 sont deux mesures boréliennes finies sur $\mathbb{R}$ telles que $\mu_1(\mathbb{R}) = \mu_2(\mathbb{R})$, alors $\mathcal{M} = \{A \in \mathfrak{B}(\mathbb{R});\ \mu_1(A) = \mu_2(A)\}$ est une classe monotone.

Démonstration. Exo. □

DÉFINITION 1.4. *Soit $\mathcal{C}$ une famille de parties de Ω. On dit que $\mathcal{C}$ est* ***π-système*** *si $\mathcal{C}$ est* stable par intersections finies.

EXEMPLE. On prend $\Omega = \mathbb{R}$. La famille $\mathcal{C} = \big\{]-\infty, t];\ t \in \mathbb{R}\big\}$ est un π-système.

Démonstration. C'est évident. □

THÉORÈME DES CLASSES MONOTONES. Si $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ est un π-système et si $\mathcal{M} \subseteq \mathcal{P}(\Omega)$ est une classe monotone contenant $\mathcal{C}$, alors $\mathcal{M}$ contient la tribu $\sigma(\mathcal{C})$ engendrée par $\mathcal{C}$.

Démonstration. On donne juste les grandes lignes ; les détails sont laissés en exo. L'idée est de considérer la classe monotone $\mathcal{M}_0$ *engendrée par* $\mathcal{C}$ (*i.e.* la plus petite classe monotone contenant $\mathcal{C}$), et de montrer que $\mathcal{M}_0$ est une tribu.

(i) On commence par montrer que $A \cap B \in \mathcal{M}_0$ pour tout $A \in \mathcal{C}$ et pour tout $B \in \mathcal{M}_0$. Pour cela, on fixe $A \in \mathcal{C}$ et on vérifie que la famille $\mathcal{M}_A := \{B \subseteq \Omega;\ A \cap B \in \mathcal{M}_0\}$ est une classe monotone qui contient $\mathcal{C}$.

(ii) Ensuite, on montre que $A \cap B \in \mathcal{M}_0$ pour tous $A, B \in \mathcal{M}_0$, en fixant $B \in \mathcal{M}_0$ et en observant que la classe monotone $\mathcal{M}_B := \{A \subseteq \Omega;\ A \cap B \in \mathcal{M}_0\}$ contient $\mathcal{C}$ par (i). À ce stade, on sait donc que $\mathcal{M}_0$ est stable par intersections finies.

(iii) Par (ii) et comme la classe monotone $\mathcal{M}_0$ est stable par complémentation, on voit que $\mathcal{M}_0$ est stable par réunions finies. On en déduit alors que $\mathcal{M}_0$ est stable par réunion dénombrables, en utilisant la stabilité par réunions dénombrables *croissantes*. □

COROLLAIRE 1.5. *Si μ_1 et μ_2 sont deux mesures boréliennes finies sur $\mathbb{R}$ telles que $\mu_1\big(]-\infty, t]\big) = \mu_2\big(]-\infty, t]\big)$ pour tout $t \in \mathbb{R}$, alors $\mu_1 = \mu_2$.*

Démonstration. Les mesures μ_1 et μ_2 sont finies, et $\mu_1(\mathbb{R}) = \lim_{n\to\infty} \mu_1(]-\infty, n]) = \lim_{n\to\infty} \mu_2(]-\infty, n]) = \mu_2(\mathbb{R})$. Donc la famille

$$\mathcal{M} := \{A \in \mathfrak{B}(\mathbb{R});\ \mu_1(A) = \mu_2(A)\}$$

est une classe monotone. Par le Théorème des classes monotones appliqué avec $\mathcal{M}$ et le π-système $\mathcal{C} = \big\{]-\infty, t];\ t \in \mathbb{R}\big\}$, on en déduit que $\mathcal{M}$ contient $\sigma(\mathcal{C}) = \mathfrak{B}(\mathbb{R})$, et donc que $\mu_1(A) = \mu_2(A)$ pour tout borélien $A \subseteq \mathbb{R}$. □

2. Lois discrètes et lois à densité

2.1. Lois discrètes.

DÉFINITION 2.1. *Soit $(\Omega, \mathfrak{A})$ un univers probabilisable. On dit qu'une loi de probabilité $\mathbb{P}$ sur $(\Omega, \mathfrak{A})$ est **discrète** si $\mathbb{P}$ est* portée par un ensemble dénombrable*, autrement dit s'il existe un ensemble dénombrable $D \subseteq \Omega$ tel que $\mathbb{P}(\Omega\backslash D) = 0$.*

Exemple. Si $a \in \Omega$, alors $\mathbb{P} = \delta_a$ est une loi discrète car elle est portée par $\{a\}$.

Remarque. Par définition, si l'univers Ω est dénombrable, alors toute loi de probabilité sur Ω est discrète.

DÉFINITION 2.2. *Soit Ω un ensemble non vide. Une **densité discrète** sur Ω est une fonction $\rho : \Omega \to \mathbb{R}$ telle que $\rho \geqslant 0$ et $\sum\limits_{\omega\in\Omega} \rho(\omega) = 1$. Autrement dit, ρ est une fonction $\geqslant 0$ telle que $\int_\Omega \rho\, d\mu_c = 1$, où μ_c est la* mesure de comptage *sur Ω.*

PROPOSITION 2.3. *Soit $(\Omega, \mathfrak{A})$ un univers probabilisable.*

1. *Si $\mathbb{P}$ est une loi de probabilité discrète sur $(\Omega, \mathfrak{A})$, alors la fonction ρ définie par $\rho(\omega) = \mathbb{P}(\{\omega\})$ est une densité discrète sur Ω.*
2. *Inversement, si $\rho : \Omega \to \mathbb{R}$ est une densité discrète sur Ω, il existe une unique loi de probabilité $\mathbb{P}$ sur $(\Omega, \mathfrak{A})$ telle que $\mathbb{P}(\{\omega\}) = \rho(\omega)$ pour tout $\omega \in \Omega$, et cette loi $\mathbb{P}$ est discrète.*
3. *Si $\mathbb{P}$ est une loi de probabilité discrète sur $(\Omega, \mathfrak{A})$, de densité associée ρ, alors on a pour tout $A \in \mathfrak{A}$:*

$$\mathbb{P}(A) = \sum_{\omega\in A} \rho(\omega) = \sum_{\omega\in A} \mathbb{P}(\{\omega\}).$$

Démonstration. (1) Soit $\mathbb{P}$ une loi discrète sur $(\Omega, \mathfrak{A})$, et soit $D \subseteq \Omega$ un ensemble dénombrable tel que $\mathbb{P}(\Omega\backslash D) = 0$. Alors $\mathbb{P}(D) = 1 - \mathbb{P}(\Omega\backslash D) = 1$. Mais comme D est dénombrable, on a aussi $\mathbb{P}(D) = \mathbb{P}\big(\bigcup_{\omega\in D}\{\omega\}\big) = \sum_{\omega\in D} \mathbb{P}(\{\omega\})$; donc $\sum_{\omega\in D} \mathbb{P}(\{\omega\}) = 1$. Enfin, $\mathbb{P}(\{\omega\}) = 0$ si $\omega \notin D$ car $\mathbb{P}(\{\omega\}) \leqslant \mathbb{P}(\Omega\backslash D) = 0$; donc $\sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = \sum_{\omega\in D} \mathbb{P}(\{\omega\}) = 1$. Donc $\rho(\omega) = \mathbb{P}(\{\omega\})$ est une densité discrète sur Ω.

(2) **Existence.** Pour $A \in \mathfrak{A}$, on pose

$$\mathbb{P}(A) := \sum_{\omega \in A} \rho(\omega).$$

(i) $\mathbb{P}$ *est une loi de probabilité.* Pour tout $A \in \mathcal{P}(\Omega)$, on a

$$\mathbb{P}(A) = \sum_{\omega \in \Omega} \mathbf{1}_A(\omega)\rho(\omega) = \int_\Omega \mathbf{1}_A \rho \, d\mu_c = \int_A \rho \, d\mu_c,$$

où μ_c est la mesure de comptage sur Ω. Donc $\mathbb{P}$ est une mesure sur $(\Omega, \mathfrak{A})$ (*cf* cours d'intégration). De plus $\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \rho(\omega) = 1$, donc $\mathbb{P}$ est une loi de probabilité ; et par définition $\mathbb{P}(\{\omega\}) = \rho(\omega)$ pour tout $\omega \in \Omega$.

(ii) $\mathbb{P}$ *est une loi discrète.*

Fait 2.4. *Si μ est une mesure* finie *sur $(\Omega, \mathfrak{A})$, alors $D := \{\omega \in \Omega;\ \mu(\{\omega\}) \neq 0\}$ est un ensemble dénombrable.*

Preuve du Fait 2.4. On a $D = \bigcup_{k \in \mathbb{N}^*} D_k$, où $D_k = \{\omega \in \Omega;\ \mu(\{\omega\}) \geqslant 1/k\}$; donc il suffit de montrer que chaque ensemble D_k est *fini*.

Supposons que D_k soit infini pour un certain k. Soit $N \in \mathbb{N}^*$ quelconque. Comme D_k est infini, il contient au moins N points distincts $\omega_1, \ldots, \omega_N$. On a alors

$$\mu(D_k) \geqslant \mu(\{\omega_1, \ldots, \omega_N\}) = \sum_{i=1}^{N} \mu(\{\omega_i\}) \geqslant N \times \frac{1}{k}.$$

Donc $\mu(\Omega) \geqslant \frac{N}{k}$ pour tout $N \in \mathbb{N}^*$, absurde puisque $\mu(\Omega) < \infty$.

□

Par le Fait 2.4, l'ensemble $D = \{\omega \in \Omega;\ \rho(\omega) \neq 0\} = \{\omega \in \Omega;\ \mathbb{P}(\{\omega\}) \neq 0\}$ est dénombrable. Par définition de $\mathbb{P}$ et D, on a $\mathbb{P}(D) = \sum_{\omega \in D} \rho(\omega) = \sum_{\omega \in \Omega} \rho(\omega) = 1$, donc $\mathbb{P}(\Omega \backslash D) = 0$ puisque $\mathbb{P}(\Omega) = 1$. Donc $\mathbb{P}$ est portée par D, et donc $\mathbb{P}$ est une loi discrète.

Unicité. Soit $\mathbb{P}'$ autre loi vérifiant $\mathbb{P}'(\{\omega\}) = \rho(\omega)$ pour tout $\omega \in \Omega$. Si on pose à nouveau $D = \{\omega \in \Omega;\ \rho(\omega) \neq 0\}$, alors ($D$ est toujours dénombrable et) on montre comme plus haut que $\mathbb{P}'(\Omega \backslash D) = 0$. Pour tout $A \in \mathfrak{A}$, on a donc

$$\mathbb{P}'(A) = \mathbb{P}'(A \cap D) = \sum_{\omega \in A \cap D} \mathbb{P}'(\{\omega\}) \quad \text{car } A \cap D \text{ est dénombrable;}$$

autrement dit

$$\mathbb{P}'(A) = \sum_{\omega \in A \cap D} \rho(\omega) = \sum_{\omega \in A} \rho(\omega) = \mathbb{P}(A).$$

(3) Découle la preuve de (2). □

Corollaire 2.5. *Une loi de probabilité $\mathbb{P}$ sur $(\Omega, \mathfrak{A})$ est discrète si et seulement si*

$$\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 1.$$

Démonstration. Exo. □

REMARQUE. Si $\mathbb{P}$ est une loi discrète sur $(\Omega, \mathfrak{A})$ de densité associée ρ, on a pour tout $A \in \mathfrak{A}$:

$$\mathbb{P}(A) = \int_A \rho \, d\mu_c.$$

Cette formule permet de prolonger $\mathbb{P}$ en une loi de probabilité *définie sur* $\mathcal{P}(\Omega)$ *tout entier.*

EXEMPLE 1. Soit Ω un ensemble fini. Il existe une unique loi de probabilité sur Ω telle que

$$\mathbb{P}(\{\omega\}) = \frac{1}{\#\Omega} \qquad \text{pour tout } \omega \in \Omega.$$

On dit que $\mathbb{P}$ est la **loi uniforme sur** Ω. Par définition, on a

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} \qquad \text{pour tout } A \subseteq \Omega.$$

Démonstration. Si on pose $\rho(\omega) = \frac{1}{\#\Omega}$ pour tout $\omega \in \Omega$, alors $\rho \geqslant 0$ et $\sum_{\omega\in\Omega} \rho(\omega) = 1$, donc ρ est une densité discrète. □

EXEMPLE 2. Soit $p \in [0, 1]$. Il existe une unique loi de probabilité sur $\Omega = \{0, 1\}$ telle que

$$\mathbb{P}(\{1\}) = p \qquad \text{et} \qquad \mathbb{P}(\{0\}) = 1 - p.$$

On dit que $\mathbb{P}$ est la **loi de Bernoulli de paramètre** p, et on la note $\mathcal{B}(p)$.

Démonstration. Exo. □

EXEMPLE 3. Soit $n \in \mathbb{N}$, et soit $p \in [0, 1]$. Il existe une unique loi de probabilité $\mathbb{P}$ sur $\Omega = [\![0, n]\!]$ telle que

$$\mathbb{P}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{pour tout } k \in [\![0, n]\!].$$

On dit que $\mathbb{P}$ est la **loi binomiale** $\mathcal{B}(n, p)$.

Démonstration. Si on pose $\rho(k) = \binom{n}{k} p^k (1-p)^{n-k}$ pour tout $k \in [\![0, n]\!]$, alors $\rho \geqslant 0$ et

$$\sum_{k\in[\![0,n]\!]} \rho(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \big(p + (1-p)\big)^n = 1.$$

□

EXEMPLE 4. Soit $\lambda > 0$. Il existe une unique loi de probabilité sur $\Omega = \mathbb{N}$ telle que

$$\mathbb{P}(\{k\}) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad \text{pour tout } k \in \mathbb{N}.$$

On dit que $\mathbb{P}$ est la **loi de Poisson de paramètre** λ, et on la note $\mathcal{P}(\lambda)$.

Démonstration. Si on pose $\rho(k) = \frac{\lambda^k}{k!} e^{-\lambda}$ pour tout $k \in \mathbb{N}$, alors

$$\sum_{k\in\mathbb{N}} \rho(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \times e^{\lambda} = 1.$$

□

LEMME 2.6. (intégration)
Soit $\mathbb{P}$ *une loi de probabilité discrète sur un univers* Ω*, de densité associée* $\rho : \Omega \to \mathbb{R}^+$.

(i) *Pour toute fonction (mesurable)* positive f *sur* Ω*, on a*

$$(*) \qquad \int_\Omega f\, d\mathbb{P} = \sum_{\omega\in\Omega} f(\omega)\rho(\omega) = \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\})f(\omega).$$

(ii) *Une fonction (mesurable)* $f : \Omega \to \mathbb{R}$ *est intégrable par rapport à* $\mathbb{P}$ *si et seulement si* $\sum_{\omega\in\Omega} |f(\omega)|\rho(\omega) < \infty$*, et alors* $(*)$ *est vraie.*

Démonstration. (i) Il s'agit de montrer que pour toute fonction (mesurable) $f : \Omega \to \mathbb{R}^+$, on a

$$\int_\Omega f\, d\mathbb{P} = \int_\Omega f\rho\, d\mu_c,$$

où μ_c est la mesure de comptage sur Ω. Si f est une fonction indicatrice, $f = \mathbf{1}_A$, c'est OK d'après la Proposition 2.3 :

$$\int_\Omega \mathbf{1}_A\, d\mathbb{P} = \mathbb{P}(A) = \sum_{\omega\in A} \rho(\omega) = \sum_{\omega\in\Omega} \mathbf{1}_A(\omega)\rho(\omega) = \int_\Omega \mathbf{1}_A\rho\, d\mu_c.$$

Par linéarité, c'est encore OK pour une fonction f *étagée* positive ; et on conclut par convergence monotone.

(ii) Exo. □

Lemme 2.7. (évènements presque sûrs)
Soit $\mathbb{P}$ *une loi discrète sur* Ω*, de densité associée* $\rho : \Omega \to \mathbb{R}^+$*. Un évènement* $E \subseteq \Omega$ *est presque sûr si et seulement si* $\rho(\omega) = 0$ *pour tout* $\omega \notin E$*.*

Démonstration. Évident puisque $\mathbb{P}(E^c) = \sum_{\omega\notin E} \rho(\omega)$. □

Corollaire 2.8. *Si* $\rho(\omega) > 0$ *pour tout* $\omega \in \Omega$*, alors le seul évènement presque sûr est* $E = \Omega$ *; autrement dit : une propriété (dépendant de* $\omega \in \Omega$*) est vraie presque sûrement si et seulement si elle est vraie pour tout* $\omega \in \Omega$*.*

2.2. Lois à densité. Dans ce qui suit, on note λ_d la mesure de Lebesgue sur $\mathbb{R}^d$, et pour une fonction (borélienne) f positive ou intégrable sur $\mathbb{R}^d$, on pose

$$\int_{\mathbb{R}^d} f(x)\, dx := \int_{\mathbb{R}^d} f\, d\lambda_d$$

Définition 2.9. *Soit* $d \in \mathbb{N}^*$*. Une* ***densité lebesguienne*** *sur* $\mathbb{R}^d$ *est une fonction borélienne* $\rho : \mathbb{R}^d \to \mathbb{R}$ *telle que* $\rho \geqslant 0$ *presque partout et* $\int_{\mathbb{R}^d} \rho(x)dx = 1$*.*

Remarque. On peut étendre la définition en autorisant ρ à n'être définie que λ_d-presque partout.

Lemme 2.10. *Si* $\rho : \mathbb{R}^d \to \mathbb{R}$ *est une densité lebesguienne, on définit une loi de probabilité sur* $\mathbb{R}^d$ *en posant pour tout borélien* $A \subseteq \mathbb{R}^d$ *:*

$$\mathbb{P}(A) = \int_A \rho(x)\, dx.$$

Démonstration. On sait (*cf* cours d'intégration) que cette formule définit une mesure borélienne sur $\mathbb{R}^d$; et on a $\mathbb{P}(\mathbb{R}^d) = \int_{\mathbb{R}^d} \rho(x)\, dx = 1$. □

Définition 2.11. *Une dit qu'une loi de probabilité* $\mathbb{P}$ *sur* $\mathbb{R}^d$ *est une* ***loi à densité*** *si elle provient d'une densité lebesguienne, i.e. s'il existe une densité lebesguienne* ρ *telle que* $\mathbb{P}(A) = \int_A \rho(x)dx$ *pour tout borélien* $A \subseteq \mathbb{R}^d$*. On écrit alors* $\mathbb{P} = \rho dx$*, ou* $\mathbb{P} = \rho(x)dx$*, ou encore* $d\mathbb{P}(x) = \rho(x)dx$*.*

REMARQUE 1. La loi détermine la densité : si $\rho_1 dx = \rho_2 dx$, alors $\rho_1(x) = \rho_2(x)$ presque partout (relativement à λ_d).

Démonstration. Supposons que ρ_1 et ρ_2 ne soient pas presque partout égales, autrement dit que $\lambda_d(\{\rho_1 \neq \rho_2\}) > 0$. Alors on a par exemple $\lambda_d(\{\rho_1 > \rho_2\}) > 0$. Si on pose $E = \{\rho_1 > \rho_2\}$, alors $\rho_1 - \rho_2 > 0$ sur E et $\lambda_d(E) > 0$, donc $\int_E (\rho_1 - \rho_2)\, d\lambda_d > 0$. Donc $\int_E \rho_1(x)\, dx > \int_E \rho_2(x)\, dx$, et donc $\rho_1 dx \neq \rho_2 dx$. □

REMARQUE 2. Si $\mathbb{P}$ est une loi à densité sur $\mathbb{R}^d$, alors $\mathbb{P}$ est une mesure **diffuse** : on a $\mathbb{P}(\{x\}) = 0$ pour tout $x \in \mathbb{R}^d$.

Démonstration. On sait que $\lambda_d(\{x\}) = 0$ pour tout $x \in \mathbb{R}^d$; donc $\mathbb{P}(\{x\}) = \int_{\{x\}} \rho\, d\lambda_d = 0$. □

EXEMPLE 1. Si K est un borélien de $\mathbb{R}^d$ tel que $0 < \lambda_d(K) < \infty$, la fonction

$$\rho := \frac{1}{\lambda_d(K)}\, \mathbf{1}_K$$

est une densité lebesguienne sur $\mathbb{R}^d$. La loi $\mathbb{P} = \frac{1}{\lambda_d(K)} \mathbf{1}_K\, dx$ s'appelle la **loi uniforme sur** K. Par définition, on a

$$\mathbb{P}(A) = \frac{\lambda_d(A \cap K)}{\lambda_d(K)} \qquad \text{pour tout borélien } A \subseteq \mathbb{R}^d,$$

et donc

$$\mathbb{P}(A) = \frac{\lambda_d(A)}{\lambda_d(K)} \qquad \text{pour tout borélien } A \subseteq K.$$

En particulier, la *loi uniforme sur* $[0,1]$ est $\mathbb{P} = \mathbf{1}_{[0,1]} dx$.

Démonstration. On a $\int_{\mathbb{R}^d} \mathbf{1}_K(x)\, dx = \int_{\mathbb{R}^d} \mathbf{1}_K\, d\lambda_d = \lambda_d(K)$, donc $\int_{\mathbb{R}^d} \rho(x)\, dx = 1$. □

EXEMPLE 2. La fonction $\rho : \mathbb{R} \to \mathbb{R}$ définie par

$$\rho(x) := \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$$

est une densité lebesguienne sur $\mathbb{R}$. Plus généralement, si $m \in \mathbb{R}$ et $\sigma > 0$ sont donnés, la fonction

$$\rho_{m,\sigma}(x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

est une densité lebesguienne. La loi $\mathbb{P} = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2} dx$ s'appelle la **loi normale** $\mathcal{N}(m, \sigma^2)$**.**

Démonstration. On a

$$\begin{aligned}
\int_{\mathbb{R}} \rho_{m,\sigma}(x)\, dx &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} e^{-\frac{(x-m)^2}{2\sigma^2}}\, dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} e^{-\frac{u^2}{2}}\, \sigma du \qquad \text{en posant } u = \tfrac{(x-m)}{\sigma} \\
&= 1 \qquad \text{car} \quad \textstyle\int_{\mathbb{R}} e^{-\frac{u^2}{2}}\, du = \sqrt{2\pi}.
\end{aligned}$$

□

Remarque. Il faut savoir redémontrer que $\int_{\mathbb{R}} e^{-u^2/2} du = \sqrt{2\pi}$.

EXEMPLE 3. La fonction $\rho : \mathbb{R} \to \mathbb{R}$ définie par

$$\rho(x) := \frac{1}{\pi} \frac{1}{1+x^2}$$

est une densité lebesguienne sur $\mathbb{R}$. La loi $\mathbb{P} = \frac{1}{\pi} \frac{dx}{1+x^2}$ s'appelle la **loi de Cauchy**.

Démonstration. On a $\int_{\mathbb{R}} \rho(x)\, dx = 1$ car $\int_{\mathbb{R}} \frac{dx}{1+x^2} = \big[\arctan(x)\big]_{-\infty}^{\infty} = \pi$. □

EXEMPLE 4. Soit $\lambda > 0$. La fonction $\rho : \mathbb{R} \to \mathbb{R}$ définie par

$$\rho(x) := \lambda\, e^{-\lambda x} \mathbf{1}_{]0,\infty[}(x)$$

est une densité lebesguienne sur $\mathbb{R}$. La loi $\mathbb{P} = \lambda\, e^{-\lambda x} \mathbf{1}_{]0,\infty[}(x) dx$ s'appelle la **loi exponentielle de paramètre** λ, et se note $\mathcal{E}(\lambda)$.

Démonstration. On a $\int_{\mathbb{R}} \rho(x)\, dx = \lambda \int_0^{\infty} e^{-\lambda x} dx = \lambda \times \big[-\frac{1}{\lambda}\, e^{-\lambda x}\big]_0^{\infty} = 1$. □

LEMME 2.12. (intégration)
Soit $\mathbb{P}$ une loi à densité sur $\mathbb{R}^d$, de densité associée $\rho : \mathbb{R}^d \to \mathbb{R}^+$.

(i) *Pour toute fonction borélienne positive f sur $\mathbb{R}^d$, on a*

$$(*) \qquad \int_{\mathbb{R}^d} f\, d\mathbb{P} = \int_{\mathbb{R}^d} f(x)\rho(x)\, dx.$$

(ii) *Une fonction borélienne $f : \mathbb{R}^d \to \mathbb{R}$ est intégrable par rapport à $\mathbb{P}$ si et seulement si $\int_{\mathbb{R}^d} |f(x)|\rho(x)dx < \infty$, et alors $(*)$ est vraie.*

Démonstration. La preuve est exactement la même que celle du Lemme 2.6. □

LEMME 2.13. (évènements presque sûrs)
Soit $\mathbb{P}$ une loi à densité sur $\mathbb{R}^d$, de densité $\rho : \mathbb{R}^d \to \mathbb{R}$. Un évènement $E \subseteq \mathbb{R}^d$ est presque sûr si et seulement si $\rho(x) = 0$ pour presque tout $x \notin E$ relativement à la mesure de Lebesgue λ_d, autrement dit si $\lambda_d(E^c \cap \{\rho \neq 0\}) = 0$.

Démonstration. C'est évident puisque $\mathbb{P}(E^c) = \int_{E^c} \rho\, d\lambda_d$. □

3. Probabilités conditionnelles

On se donne un espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$, et on suppose que $(\Omega, \mathfrak{A}, \mathbb{P})$ modèlise une "expérience réelle". Donc on identifie complètement les "évènements mathématiques" (*i.e.* les $E \in \mathfrak{A}$) avec des "évènements réels"

DÉFINITION INTUITIVE. Soit E un évènement. Si A est un autre évènement, la **probabilité de A sachant E**, notée $\mathbb{P}(A \,|\, E)$, est le degré de plausibilité qu'on accorderait à A si on était certain que l'évènement E se produit.

Exemple 1. Si on joue 2 fois à pile ou face, il est intuitivement clair que le résultat du 1er lancer n'influe pas sur les chances d'obtenir "pile" au 2ème jet. Donc on a par exemple $\mathbb{P}(\text{"pile" au 2ème jet} \,|\, \text{"face" au 1er jet}) = 1/2$.

Exemple 2. Si on tire 2 cartes l'une après l'autre dans un jeu de 52 cartes, il est intuitivement clair que $\mathbb{P}(\text{la 2ème carte est un "pique"} \,|\, \text{la 1ère est un "pique"}) = \frac{12}{51}$ et que $\mathbb{P}(\text{la 2ème carte est un "pique"} \,|\, \text{la 1ère est un "coeur"}) = \frac{13}{51}$.

PROPRIÉTÉS INTUITIVES.

(o) Si E est un évènement "presque impossible", *i.e.* $\mathbb{P}(E) = 0$, alors $\mathbb{P}(A \,|\, E)$ n'a pas de sens. (Si l'impossible se produit, on ne peut plus rien dire de sensé...).

(i) L'application $A \mapsto \mathbb{P}(A \,|\, E)$ est une loi de probabilité.

(ii) $\mathbb{P}(E \,|\, E) = 1$. (Si on est certain que E se produit... on est certain que E se produit.)

(iii) Pour les évènements A dont la réalisation implique celle de E, *i.e.* les $A \subseteq E$, $\mathbb{P}(A \,|\, E)$ est proportionnelle à $\mathbb{P}(A)$. (Le fait de savoir que E est réalisé ne change "esentiellement pas" le degré de plausibilité qu'on accorde à A si de toutes façons A entraine E.)

FAIT 3.1. *Si A et E sont deux évènements et si $\mathbb{P}(E) \neq 0$, alors*

$$(*) \qquad \mathbb{P}(A \,|\, E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)}.$$

Démonstration. Par (i) et (ii) on doit avoir $\mathbb{P}(E^c \,|\, E) = 1 - \mathbb{P}(E \,|\, E) = 0$, donc

$$\mathbb{P}(A \,|\, E) = \mathbb{P}\big((A \cap E^c) \,|\, E\big) + \mathbb{P}\big((A \cap E) \,|\, E\big) = \mathbb{P}\big((A \cap E) \,|\, E\big);$$

et donc par (iii) :

$$\mathbb{P}(A \,|\, E) = c\,\mathbb{P}(A \cap E),$$

où c est une constante indépendante de A. Comme $\mathbb{P}(E \,|\, E) = 1$ et $\mathbb{P}(E) \neq 0$ par (iii) et (o), on a $c = 1/\mathbb{P}(E)$; donc OK. □

INTÉRÊT. La formule $(*)$ donne une définition mathématique précise de la probabilité conditionnelle.

Exercice. Montrer que si Ω est fini et si $\mathbb{P}$ est la loi uniforme sur Ω, alors la restriction de $\mathbb{P}(\,\cdot \,|\, E)$ à $\mathcal{P}(E)$ est la loi uniforme sur E.

NOTATION. Si $A, E_1, \ldots, E_n$ sont des évènements et si $\mathbb{P}(E_1 \cap \cdots \cap E_n) \neq 0$, on pose

$$\mathbb{P}(A \,|\, E_1, \ldots, E_n) = \mathbb{P}(A \,|\, E_1 \cap \cdots \cap E_n).$$

C'est la "probabilité que A soit réalisé sachant que E_1, E_2 ... et E_n sont réalisés".

PROPOSITION 3.2. *Si $E_1, \ldots, E_N$ sont des évènements tels que $\mathbb{P}(E_1 \cap \cdots \cap E_N) \neq 0$, alors on a la* ***formule des probabilités en cascade*** *:*

$$\mathbb{P}(E_1 \cap \cdots \cap E_N) = \mathbb{P}(E_1)\,\mathbb{P}(E_2 \,|\, E_1)\,\mathbb{P}(E_3 \,|\, E_1, E_2) \cdots \mathbb{P}(E_N \,|\, E_1, \ldots, E_{N-1}).$$

Démonstration. Le second membre est égal à

$$\mathbb{P}(E_1) \times \frac{\mathbb{P}(E_2 \cap E_1)}{\mathbb{P}(E_1)} \times \frac{\mathbb{P}\big(E_3 \cap (E_2 \cap E_1)\big)}{\mathbb{P}(E_2 \cap E_1)} \times \cdots \times \frac{\mathbb{P}\big(E_N \cap (E_{N-1} \cap \cdots \cap E_1)\big)}{\mathbb{P}(E_{N-1} \cap \cdots \cap E_1)},$$

qui se simplifie "en diagonale" pour donner $\mathbb{P}(E_N \cap \cdots \cap E_1)$. □

REMARQUE. C'est cette formule qui justifie l'utilisation d'"arbres pondérés" pour calculer des probabilités.

Exemple. On tire 4 cartes au hasard dans un jeu de 52 cartes, et on veut savoir quelle est la probabilité d'obtenir un carré.

Il y a 13 carrés possibles, qui sont évidemment "équiprobables" ; donc la probabilité cherchée est $p = 13q$, où q est la probabilité d'obtenir un carré d'as. On considère que les cartes sont tirées l'une après l'autre. Si on note E_i l'évènement "la i-ème carte tirée est un as" (pour $i = 1, 2, 3, 4$) alors

$$\begin{aligned} q &= \mathbb{P}(E_1)\,\mathbb{P}(E_2 \mid E_1)\,\mathbb{P}(E_3 \mid E_1, E_2)\,\mathbb{P}(E_4 \mid E_1, E_2, E_3) \\ &= \frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} \times \frac{1}{49}, \end{aligned}$$

d'où après simplification : $p = \frac{1}{20825}$.

Proposition 3.3. (formule des probabilités totales)
Si $(E_i)_{i\in I}$ est une partition dénombrable *de Ω en évènements tels que $\mathbb{P}(E_i) \neq 0$, alors on a pour tout évènement A :*

$$\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A \mid E_i)\,\mathbb{P}(E_i).$$

Démonstration. On a $A = \bigcup_{i\in I}(A \cap E_i)$, réunion dénombrable disjointe, donc $\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A \cap E_i) = \sum_{i\in I} \mathbb{P}(A \mid E_i)\mathbb{P}(E_i)$. □

Exemple. Une petite fille aimant les bonbons au citron mais pas les bonbons à la banane a devant elle trois boites opaques, la première contenant 10 bonbons au citron et 30 bonbons à la banane, la 2è contenant 20 bonbons au citron et 20 bonbons à la banane, et la 3è contenant 20 bonbons au citron et 30 bonbons à la banane. Elle choisit une boite au hasard et pioche un bonbon dedans. On veut savoir quelle est la probabilité qu'elle mange le bonbon.

La probabilité cherchée est $p = \mathbb{P}(A)$, où A est l'évènement "la petite fille pioche un bonbon au citron". Si on numérote les boites et qu'on note E_i l'évènement "la petite fille choisit la boite numéro i" (pour $i = 1, 2, 3$), on a $\mathbb{P}(E_i) = \frac{1}{3}$ pour $i = 1, 2, 3$. Donc, par la formule des probabilités totales,

$$\begin{aligned} p &= \frac{1}{3}\big(\mathbb{P}(A \mid E_1) + \mathbb{P}(A \mid E_2) + \mathbb{P}(A \mid E_3)\big) \\ &= \frac{1}{3} \times \big(\frac{10}{40} + \frac{20}{40} + \frac{20}{50}\big); \end{aligned}$$

soit finalement $p = \frac{23}{60}$.

Proposition 3.4. (Théorème de Bayes)

(1) *Si E et E' sont deux évènements tels que $\mathbb{P}(E) \neq 0$ et $\mathbb{P}(E') \neq 0$, alors*

$$\mathbb{P}(E' \mid E) = \frac{\mathbb{P}(E \mid E')\,\mathbb{P}(E')}{\mathbb{P}(E)}.$$

(2) *Soit $(E_i)_{i\in I}$ une partition* dénombrable *de Ω en évènements tels que $\mathbb{P}(E_i) \neq 0$. Pour tout évènement E tel que $\mathbb{P}(E) \neq 0$ et pour tout $i_0 \in I$, on a*

$$\mathbb{P}(E_{i_0} \mid E) = \frac{\mathbb{P}(E \mid E_{i_0})\,\mathbb{P}(E_{i_0})}{\sum_{i\in I} \mathbb{P}(E \mid E_i)\,\mathbb{P}(E_i)}.$$

Démonstration. (1) est évident.

(2) Par (1)

$$\mathbb{P}(E_{i_0} \mid E) = \frac{\mathbb{P}(E \mid E_{i_0})\,\mathbb{P}(E_{i_0})}{\mathbb{P}(E)};$$

d'où le résultat par la formule des probabilités totales. □

Exemple. Dans un certain ensemble d'objets, on estime que 1 objet sur N est défectueux. On met au point un test pour détecter les objets défectueux. Le test est efficace à 100% (un objet défectueux a 100% de chances d'être détecté), mais il y a ε % de chances qu'un objet non défectueux soit déclaré défectueux par le test ($\varepsilon > 0$). On veut savoir quelle est la **fiabilité** du test, autrement dit la probabilité qu'un objet déclaré défectueux par le test soit effectivement défectueux.

L'expérience consiste à tirer un objet au hasard et à le tester. si on note D l'évènement "l'objet est défectueux" et DD l'évènement "l'objet est déclaré défectueux", alors la probabilité cherchée est $p = \mathbb{P}(D \mid DD)$. On applique la théorème de Bayes avec la partition (D, D^c) :

$$\begin{aligned} p &= \frac{\mathbb{P}(DD \mid D)\,\mathbb{P}(D)}{\mathbb{P}(DD \mid D)\,\mathbb{P}(D) + \mathbb{P}(DD \mid D^c)\,\mathbb{P}(D^c)} \\ &= \frac{1 \times \frac{1}{N}}{1 \times \frac{1}{N} + \frac{\varepsilon}{100} \times (1 - \frac{1}{N})}, \end{aligned}$$

autrement dit

$$p = \frac{100}{100 + \varepsilon(N-1)}.$$

Application numérique : avec $N = 1000$ et $\varepsilon = 1$, on obtient $p = \frac{100}{1099} < \frac{1}{10}$.

4. Évènements indépendants

Définition 4.1. *Soit $(\Omega, \mathfrak{A}, \mathbb{P})$ un espace de probabilité.*

(1) *On dit que deux évènements E et F sont **indépendants** si on a $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$.*

(2) *Plus généralement, étant donné une famille $(E_i)_{i\in I}$ d'évènements, on dit que **les E_i sont indépendants**, ou que **la famille $(E_i)_{i\in I}$ est indépendante** si, pour tous $i_1, \dots, i_r \in I$ deux à deux distincts, on a*

$$\mathbb{P}(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_r}) = \mathbb{P}(E_{i_1})\mathbb{P}(E_{i_2}) \cdots \mathbb{P}(E_{i_r}).$$

Remarque 1. Si E et F sont deux évènements tels que $\mathbb{P}(E) > 0$ et $\mathbb{P}(F) > 0$, alors

$$E \text{ et } F \text{ indépendants} \iff \mathbb{P}(E \mid F) = \mathbb{P}(E) \iff \mathbb{P}(F \mid E) = \mathbb{P}(F).$$

Démonstration. Évident puisque $\mathbb{P}(E \cap F) = \mathbb{P}(E \mid F)\mathbb{P}(F) = \mathbb{P}(F \mid E)\mathbb{P}(E)$. □

Remarque 2. Plus généralement, si $(E_i)_{i\in I}$ est une famille quelconque d'évènements avec $\mathbb{P}(E_i) > 0$ pour tout i, alors dire que les E_i sont indépendants signifie que pour tout $i \in I$ et pour tous $j_1, \dots, j_m$ différents de i, on a $\mathbb{P}(E_i \mid E_{j_1}, \dots, E_{j_m}) = \mathbb{P}(E_i)$.

Démonstration. Exo. □

Remarque 3. Par définition, une famille d'évènements $(E_i)_{i\in I}$ est indépendante si et seulement si toute sous-famille *finie* $(E_i)_{i\in F}$ est indépendante. En particulier, si les E_i sont indépendants, alors ils sont **deux à deux indépendants** : pour tous i, j tels que $i \neq j$, les évènements E_i et E_j sont indépendants.

EXEMPLE 1. Une boite opaque contient $N_1 \geqslant 1$ dragibus noirs et $N_2 \geqslant 1$ dragibus rouges. On tire $n \geqslant 2$ dragibus l'un après l'autre dans la boite, et pour $i = 1, \dots, n$, on note E_i l'évènement "le i-ème dragibus tiré est noir". Alors :

(i) si on remet à chaque fois le dragibus tiré dans la boite, les évènements $E_1, \dots, E_n$ sont indépendants ;

(ii) si on mange à chaque fois le dragibus tiré, les évènements $E_1, \dots, E_n$ ne sont pas indépendants.

Démonstration. On note $N = N_1 + N_2$ le nombre total de dragibus.

(i) Comme on remet les dragibus à chaque fois, la couleur du i-ème n'est pas influencé par celle des autres. Donc, il est intuitivement "évident" que si si $i \in [\![1, n]\!]$ et $j_1, \dots, j_m \neq i$, alors $\mathbb{P}(E_i \,|\, E_{j_1}, \dots, E_{j_m}) = N_1/N = \mathbb{P}(E_i)$.

(ii) On va montrer que E_1 et E_2 ne sont déjà pas indépendants.

Il est intuitivement "évident" que $\mathbb{P}(E_1) = \frac{N_1}{N}$; et comme on mange à chaque fois le dragibus tiré, il est aussi "évident" que $\mathbb{P}(E_2 \,|\, E_1) = \frac{N_1-1}{N-1}$. Pour conclure, il suffit donc de montrer que $\mathbb{P}(E_2) = \frac{N_1}{N}$ (ce qui prouvera que $\mathbb{P}(E_2 \,|\, E_1) \neq \mathbb{P}(E_2)$ puisque $N_1 \neq N$). Ce dernier fait parait pas "évident", mais se démontre avec un petit calcul : par la formule des probabilités totales, on a

$$\begin{aligned}
\mathbb{P}(E_2) &= \mathbb{P}(E_2 \,|\, E_1)\,\mathbb{P}(E_1) + \mathbb{P}(E_2 \,|\, E_1^c)\,\mathbb{P}(E_1^c) \\
&= \frac{N_1 - 1}{N - 1} \times \frac{N_1}{N} + \frac{N_1}{N-1} \times \left(1 - \frac{N_1}{N}\right) \\
&= \frac{1}{N(N-1)} \big((N_1 - 1)N_1 + N_1(N - N_1)\big) \\
&= \frac{1}{N(N-1)} \times (-N_1 + N_1 N) = \frac{N_1}{N}.
\end{aligned}$$

□

EXEMPLE 2. On lance deux dés l'un après l'autre. On note E_1 l'évènement "le 1er chiffre obtenu est pair", E_2 l'évènement "le 2è chiffre obtenu est impair", et enfin E l'évènement "la somme des chiffres est paire". Alors E_1, E_2, E sont deux à deux indépendants, mais pas indépendants.

Démonstration. Exo. □

Le fait suivant est intuitivement raisonnable, mais pas "évident".

FAIT 4.2. *Soit $(E_i)_{i\in I}$ une famille indépendante d'évènements. Si $(F_i)_{i\in I}$ est une famille d'évènements telle que $\forall i \in I$: $F_i = E_i$ ou E_i^c, alors les F_i sont indépendants.*

Démonstration. Tout repose sur un

SOUS-FAIT. Le résultat est vrai s'il existe $j_0 \in I$ tel que $F_{j_0} = E_{j_0}^c$ et $F_i = E_i$ pour $i \neq j_0$.

Preuve du sous-fait. Soient $i_1, \dots, i_r \in I$ deux à deux différents. On veut montrer que $\mathbb{P}(F_{i_1} \cap \cdots \cap F_{i_r}) = \mathbb{P}(F_{i_1}) \cdots \mathbb{P}(F_{i_r})$.

Si tous les i_k sont $\neq j_0$, il n'y a rien à faire puisque $F_{i_k} = E_{i_k}$ pour tout k et que les E_i sont indépendants.

Si par exemple $i_1 = j_0$, alors

$$\begin{aligned}
\mathbb{P}(F_{i_1} \cap \cdots \cap F_{i_r}) &= \mathbb{P}(E_{j_0}^c \cap E_{i_2} \cap \cdots \cap E_{i_r}) \\
&= \mathbb{P}\big((E_{i_2} \cap \cdots \cap E_{i_r}) \backslash (E_{j_0} \cap E_{i_2} \cap \cdots \cap E_{i_r})\big) \\
&= \mathbb{P}(E_{i_2} \cap \cdots \cap E_{i_r}) - \mathbb{P}(E_{j_0} \cap E_{i_2} \cap \cdots \cap E_{i_r}) \\
&= \mathbb{P}(E_{i_2}) \cdots \mathbb{P}(E_{i_r}) - \mathbb{P}(E_{j_0})\, \mathbb{P}(E_{i_2}) \cdots \mathbb{P}(E_{i_r}) \\
&= \big(1 - \mathbb{P}(E_{j_0})\big)\, \mathbb{P}(E_{i_2}) \cdots \mathbb{P}(E_{i_r}) \\
&= \mathbb{P}(E_{j_0}^c)\, \mathbb{P}(E_{i_2}) \cdots \mathbb{P}(E_{i_r}) \\
&= \mathbb{P}(F_{i_1}) \cdots \mathbb{P}(F_{i_r}).
\end{aligned}$$

□

Soient $i_1, \dots, i_r \in I$ deux à deux $\neq$. En appliquant r fois le sous-fait, on voit que la famille $(F_{i_1}, E_{i_2}, \dots, E_{i_r})$ est indépendante, puis que la famille $(F_{i_1}, F_{i_2}, E_{i_3}, \dots, E_{i_r})$ est indépendante, etc jusqu'à $(F_{i_1}, \dots, F_{i_r})$. Ainsi, toute sous-famille finie de la famille $(F_i)_{i \in I}$ est indépendante, donc OK. □

Chapitre 3

Variables aléatoires

1. Définition et exemples

DÉFINITION 1.1. *Soit* $(\Lambda, \mathfrak{B})$ *un univers probabilisable. Une* ***variable aléatoire à valeurs dans*** $(\Lambda, \mathfrak{B})$ *est une application mesurable* $X : (\Omega, \mathfrak{A}, \mathbb{P}) \to (\Lambda, \mathfrak{B})$*, où* $(\Omega, \mathfrak{A}, \mathbb{P})$ *est un certain espace de probabilité.*

Remarque 1. S'il n'y a pas de doutes sur qui est la tribu $\mathfrak{B}$, on parle de variable aléatoire *à valeurs dans* Λ.

Remarque 2. On écrira "va" au lieu de "variable aléatoire".

Remarque 3. Une **va réelle** est une va à valeurs dans $\mathbb{R}$.

Remarque 4. On appellera encore variable aléatoire toute "va presque sûrement définie", autrement dit toute application mesurable $X : \Omega_0 \to \Lambda$, où $\Omega_0 \subseteq \Omega$ est mesurable avec $\mathbb{P}(\Omega_0) = 1$.

EXEMPLE 1. Soit $(\Omega, \mathfrak{A}, \mathbb{P})$ un espace de probabilité. Si E est un évènement, alors $\mathbf{1}_E$ est une va définie sur Ω, à valeurs dans $\{0, 1\}$.

EXEMPLE 2. On joue n fois à pile ou face, et on appelle X le nombre de "piles" obtenus. On modélise la situation en prenant $\Omega = \{0, 1\}^n$, où 1 correspond à "pile" et 0 correspond à "face". Tous les résultats sont équiprobables, donc on prend pour $\mathbb{P}$ la loi uniforme. Alors X devient une va définie sur Ω, à valeurs dans $\Lambda = [\![0, n]\!]$. En fait, si $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ où $\omega_i \in \{0, 1\}$, alors $X(\omega) = \sum_{i=1}^n \omega_i$.

EXEMPLE 3. Soit $(\Omega, \mathfrak{A}, \mathbb{P})$ un espace de probabilité quelconque. Si on pose $X(\omega) = \omega$ pour tout $\omega \in \Omega$, alors X est une va à valeurs dans $(\Omega, \mathfrak{A})$.

2. Loi d'une variable aléatoire

NOTATION. Si $X : (\Omega, \mathfrak{A}, \mathbb{P}) \to (\Lambda, \mathfrak{B})$ est une va, on pose pour tout $A \in \mathfrak{B}$:

$$\{X \in A\} := \{\omega \in \Omega;\ X(\omega) \in A\} = X^{-1}(A).$$

LEMME 2.1. *Si* X *est une va définie sur* $(\Omega, \mathfrak{A}, \mathbb{P})$ *et à valeurs dans* $(\Lambda, \mathfrak{B})$*, on définit une loi de probabilité* $\mathbb{P}_X$ *sur* $(\Lambda, \mathfrak{B})$ *en posant pour tout* $A \in \mathfrak{B}$ *:*

$$\mathbb{P}_X(A) = \mathbb{P}\big(\{X \in A\}\big).$$

On dit que $\mathbb{P}_X$ *est la* ***loi de*** X ***sous la probabilité*** $\mathbb{P}$.

Démonstration. La définition a un sens car $\{X \in A\} = X^{-1}(A) \in \mathfrak{A}$ pour tout $A \in B$ et donc on peut bien écrire $\mathbb{P}(\{X \in A\})$. Si $(A_k)_{n\in\mathbb{N}}$ est une suite d'éléments de $\mathfrak{B}$ deux à disjoints, alors $\big\{X \in \bigcup_{k\in\mathbb{N}} A_k\big\} = \bigcup_{k\in\mathbb{N}}\{X \in A_k\}$ où la réunion est disjointe, donc $\mathbb{P}_X\big(\bigcup_{k\in\mathbb{N}} A_k\big) = \mathbb{P}\big(\bigcup_{k\in\mathbb{N}}\{X \in A_k\}\big) = \sum_{k=0}^{\infty}\mathbb{P}(X \in A_k) = \sum_{k=0}^{\infty}\mathbb{P}_X(A_k)$. Donc $\mathbb{P}_X$ est une mesure ; et on a $\mathbb{P}_X(\Lambda) = \mathbb{P}(X \in \Lambda) = \mathbb{P}(\Omega) = 1$. □

REMARQUE. Pour alléger, on écrit $\mathbb{P}(X \in A)$ au lieu de $\mathbb{P}(\{X \in A\})$. Donc

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A).$$

EXEMPLE 1. Soit $(\Omega, \mathfrak{A}, \mathbb{P})$ un espace de probabilité, E un évènement, et $p = \mathbb{P}(E)$. Si $X = \mathbf{1}_E$, alors $\mathbb{P}_X$ est la loi de Bernoulli $\mathcal{B}(p)$.

Démonstration. On a $\mathbb{P}_X(\{1\}) = \mathbb{P}(X = 1) = \mathbb{P}(E) = p$; et de même $\mathbb{P}_X(\{0\}) = \mathbb{P}(E^c) = 1 - p$. □

EXEMPLE 2. On joue n fois à pile ou face et on appelle X le nombre de "piles" obtenus. Alors X suit la loi binomiale $\mathcal{B}(n, \frac{1}{2})$.

Démonstration. On a dit plus haut qu'on modélise en prenant $\Omega = \{0,1\}^n$, où 1 correspond à "pile" et 0 à "face", avec pour $\mathbb{P}$ la loi uniforme. Alors, on a pour tout $k \in [\![0, n]\!]$:

$$\begin{aligned}\mathbb{P}(X = k) &= \frac{\#\Big\{\omega \in \Omega;\ \#\{i \in [\![1,n]\!];\ \omega_i = 1\} = k\Big\}}{\#\Omega} \\ &= \frac{\#\big\{A \subseteq [\![1,n]\!];\ \#A = k\big\}}{2^n} \\ &= \binom{n}{k}\frac{1}{2^n};\end{aligned}$$

autrement dit $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ avec $p = 1/2$. □

EXEMPLE 2'. Soit $(\Omega, \mathfrak{A}, \mathbb{P})$ un espace de probabilité, et soit $p \in [0,1]$. Soient également $E_1, \dots, E_n$ des évènements indépendants tesl que $\mathbb{P}(E_1) = \cdots = \mathbb{P}(E_n) = p$. Pour $\omega \in \Omega$, on pose $X(\omega) := \sharp\big\{i \in [\![1,n]\!];\ \omega \in E_i\big\}$. Alors X suit la loi binomiale $\mathcal{B}(n,p)$.

Démonstration. L'application X est à valeurs dans $[\![0,n]\!]$. Le fait que X soit effectivement une va (*i.e.* soit mesurable) est laissé en exo.

Pour $A \subseteq [\![1,n]\!]$, posons

$$E^A := \big\{\omega \in \Omega;\ \omega \in E_i \text{ si } i \in A \text{ et } \omega \notin E_i \text{ si } i \notin A\big\} = \bigcap_{i \in A} E_i \cap \bigcap_{j \in [\![1,n]\!]\setminus A} E_j^c.$$

Par indépendance, on a

$$\mathbb{P}(E^A) = p^{\#A}(1-p)^{n-\#A} \qquad \text{pour tout } A \subseteq [\![1,n]\!].$$

De plus, pour $k \in [\![0,n]\!]$ donné, l'évènement $\{X = k\}$ est la réunion disjointe des E^A pour les A vérifiant $\#A = k$. Donc

$$\mathbb{P}(X = k) = \sum_{\big\{A \subseteq [\![1,n]\!];\ \#A = k\big\}} p^{\#A}(1-p)^{n-\#A} = \binom{n}{k} p^k (1-p)^{n-k}.$$

□

REMARQUE. Cet exemple s'interprète de la façon suivante. Supposons qu'on répète n fois une expérience "binaire" (le résultat est soit un "succès", soit un "échec") où la probabilité de "succès" est $p \in [0,1]$. Alors, en considérant que les répétitions sont indépendantes, le nombre X de "succès" suit la loi $\mathcal{B}(n,p)$. On est bien dans la situation de l'exemple précédent, en prenant pour E_i l'évènement "succès au i-ème coup".

EXEMPLE 3. Si μ est une loi de probabilité quelconque sur un univers probabilisable $(\Lambda, \mathfrak{B})$, alors il existe une va X à valeurs dans $(\Lambda, \mathfrak{B})$ telle que $\mathbb{P}_X = \mu$.

Démonstration. On prend $(\Omega, \mathfrak{A}, \mathbb{P}) = (\Lambda, \mathfrak{B}, \mu)$ et on pose $X(\omega) = \omega$ pour tout $\omega \in \Omega$ (micro-exo). □

Exercice. On lance 2 dés, et on appelle X la somme des chiffres obtenus. Montrer que pour $k = 2, \dots, 12$, on a $\mathbb{P}_X(\{k\}) = \frac{k-1}{36}$ si $2 \leqslant k \leqslant 6$ et $\mathbb{P}_X(\{k\}) = \frac{13-k}{36}$ si $6 < k \leqslant 12$.

REMARQUE. On dit qu'une va X à valeurs dans Λ est **discrète** si sa loi $\mathbb{P}_X$ est discrète ; autrement dit s'il existe un ensemble *dénombrable* D tel que X est presque sûrement à valeurs dans D. On dit qu'une va X à valeurs dans $\mathbb{R}^d$ est une **va à densité** si sa loi $\mathbb{P}_X$ est à densité. Donc (*cf* le Chapitre 2)

- si X est une va discrète, on a pour tout $A \subseteq \Lambda$:
$$\mathbb{P}_X(A) = \sum_{a \in A} \mathbb{P}_X(\{a\}) = \sum_{a \in A} \mathbb{P}(X = a);$$
- si X est une va à densité à valeurs dans $\mathbb{R}^d$, de densité ρ_X, on a pour tout borélien $A \subseteq \mathbb{R}^d$:
$$\mathbb{P}_X(A) = \int_A \rho_X(x)\, dx.$$

Exercice. Soit X une va réelle telle que $\mathbb{P}_X$ soit une mesure diffuse (par exemple une va à densité), et soit $Z = (X, X)$, qui est une va à valeurs dans $\mathbb{R}^2$. Montrer que $\mathbb{P}_Z$ est une mesure diffuse, mais qu'elle n'est pas à densité.

3. "Théorème de transfert"

THÉORÈME 3.1. *Soit X une va définie sur $(\Omega, \mathfrak{A}, \mathbb{P})$ et à valeurs dans $(\Lambda, \mathfrak{B})$.*

(1a) *Pour toute fonction mesurable positive $f : \Lambda \to \mathbb{R}$, on a*

$$(*) \qquad \int_\Lambda f\, d\mathbb{P}_X = \int_\Omega f \circ X\, d\mathbb{P}.$$

(1b) *Une fonction mesurable $f : \Lambda \to \mathbb{R}$ est intégrable sur Λ par rapport à $\mathbb{P}_X$ si et seulement $f \circ X$ est intégrable sur Ω par rapport à $\mathbb{P}$; et dans ce cas $(*)$ est vraie. En particulier, ceci a lieu pour toute fonction mesurable bornée $f : \Lambda \to \mathbb{R}$.*

(2) *L'identité $(*)$* caractérise $\mathbb{P}_X$ *: si μ est une loi de probabilité sur $(\Lambda, \mathfrak{B})$ telle que $\int_\Omega f \circ X\, d\mathbb{P} = \int f\, d\mu$ pour toute fonction mesurable positive $f : \Lambda \to \mathbb{R}$, alors $\mu = \mathbb{P}_X$.*

Démonstration. (1a) Si f est une fonction indicatrice, $f = \mathbf{1}_E$ où $E \subseteq \Lambda$, on a $f \circ X = \mathbf{1}_E \circ X = \mathbf{1}_{\{X \in E\}}$, donc $\int_\Omega f \circ X\, d\mathbb{P} = \mathbb{P}(X \in E) = \mathbb{P}_X(E) = \int_\Lambda \mathbf{1}_E\, d\mathbb{P}_X = \int_\Lambda f\, d\mathbb{P}_X$. Par linéarité, on en déduit que $(*)$ est vraie pour toute fonction étagée positive $f : \Lambda \to \mathbb{R}$; et par convergence monotone, on obtient $(*)$ pour toute fonction mesurable positive.

(1b) s'obtient en appliquant (1a) d'abord à $|f|$ pour l'intégrabilité, puis à f^+ et f^- pour $(*)$.

(2) D'après (1a), on a $\int_\Lambda f\, d\mu = \int_\Omega f \circ X\, d\mathbb{P} = \int_\Lambda f\, d\mathbb{P}_X$ pour toute fonction mesurable $f : \Lambda \to \mathbb{R}^+$. En particulier, en prenant $f = \mathbf{1}_E$, on obtient $\mu(E) = \mathbb{P}_X(E)$ pour tout ensemble mesurable $E \subseteq \Lambda$, et donc $\mu = \mathbb{P}_X$. □

REMARQUE. Au lieu de $f \circ X$, on écrira souvent $f(X)$. Donc, la formule $(*)$ s'écrit

$$\int_\Omega f(X)\, d\mathbb{P} = \int_\Lambda f(x)\, d\mathbb{P}_X(x).$$

COROLLAIRE 3.2. *Si X est une va discrète, à valeurs dans un ensemble dénombrable I, on a pour toute fonction $f : I \to \mathbb{R}^+$:*

$$\int_\Omega f(X)\, d\mathbb{P} = \sum_{i\in I} f(i)\, \mathbb{P}(X = i).$$

Démonstration. Comme $\mathbb{P}_X$ est une loi discrète, on a

$$\int_I f\, d\mathbb{P}_X = \sum_{i\in I} f(i)\mathbb{P}_X(\{i\}) = \int_I f(i)\mathbb{P}(X = i);$$

donc OK puisque $\int_\Omega f(X)\, d\mathbb{P} = \int_i f\, d\mathbb{P}_X$. □

COROLLAIRE 3.3. *Si X est une va à densité à valeurs dans $\mathbb{R}^d$, de densité associée $\rho_X : \mathbb{R}^d \to \mathbb{R}$, on a pour toute fonction borélienne $f : \mathbb{R}^d \to \mathbb{R}^+$:*

$$\int_\Omega f(X)\, d\mathbb{P} = \int_{\mathbb{R}^d} f(x)\, \rho_X(x)dx.$$

Démonstration. Exo. □

Le théorème de transfert donne une méthode "mécanique" pour déterminer la loi d'une variable aléatoire $X\Omega \to \Lambda$, qu'on appelle la **méthode la fonction muette** : on considère une fonction borélienne $f : \Lambda \to \mathbb{R}^+$ absolument quelconque (la fonction "muette") et on essaye de transformer $\int_\Omega f(X)\, d\mathbb{P}$ pour obtenir $\int_\Lambda f(x)\, d\mu(x)$, où μ est une certaine mesure sur Λ. Si on y arrive, alors on peut conclure que $\mathbb{P}_X = \mu$. Voici un exemple typique.

EXEMPLE. Soit X une va réelle définie sur $(\Omega, \mathfrak{A}, \mathbb{P})$ et uniformément distribuée sur $]0, 1[$, *i.e.* $\mathbb{P}_X = \mathbf{1}_{]0,1[}(x)dx$. Alors la va $Y = -\log(X)$ suit une loi exponentielle de paramètre $\lambda = 1$.

Démonstration. La va Y est bien définie presque sûrement car $X > 0$ ps. Comme X est presque sûrement à valeurs dans $]0, 1[$, on peut supposer qu'en fait $X(\omega) \in]0, 1[$ pour tout $\omega \in \Omega$ (quitte à redéfinir X sur $\{X \notin]0, 1[\}$). Pour toute fonction borélienne $f : \mathbb{R} \to \mathbb{R}^+$, on a

$$\int_\Omega f(Y)\, d\mathbb{P} = \int_\Omega f\big(-\log(X)\big)\, d\mathbb{P} = \int_\Omega g(X)\, d\mathbb{P},$$

où $g : \mathbb{R} \to \mathbb{R}^+$ est la fonction définie par $g(x) = f(-\log(x))$ si $x \in]0,1[$ et par exemple $g(x) = 0$ sinon. Donc, par le Théorème de transfert appliqué à X,

$$\begin{aligned}\int_\Omega f(Y)\, d\mathbb{P} &= \int_{\mathbb{R}} g(x)\, d\mathbb{P}_X(x) \\ &= \int_{\mathbb{R}} g(x)\, \mathbf{1}_{]0,1[}(x)\, dx \\ &= \int_0^1 f\big(-\log(x)\big)\, dx \\ &= \int_0^\infty f(u)e^{-u}du \qquad \text{en posant } u = -\log(x) \\ &= \int_{\mathbb{R}} f(u)\mathbf{1}_{]0,\infty[}(u)e^{-u}du.\end{aligned}$$

Par le Théorème de transfert, on en déduit que $\mathbb{P}_Y = \mathbf{1}_{]0,\infty[}(u)e^{-u}du$, ce qui est le résultat annoncé. □

Exercice. Montrer que si X est une va réelle suivant une loi $\mathcal{N}(m, \sigma^2)$ et si $a, b \in \mathbb{R}$ avec $a \neq 0$, alors $aX + b$ suit la loi $\mathcal{N}(am + b, a^2\sigma^2)$.

4. Compositions

Notation. Si X est une va (définie sur $(\Omega, \mathfrak{A}, \mathbb{P})$) à valeurs dans $(\Lambda, \mathfrak{B})$ et si $\Phi : (\Lambda, \mathfrak{B}) \to (\Lambda', \mathfrak{B}')$ est une application mesurable, on pose

$$\Phi(X) = \Phi \circ X.$$

Ainsi, $\Phi(X)$ est une va (définie sur $(\Omega, \mathfrak{A}, \mathbb{P})$ et) à valeurs dans Λ'.

Proposition 4.1. *Soient I, J des univers dénombrables, et soit $X : (\Omega, \mathfrak{A}, \mathbb{P}) \to I$ une va à valeurs dans I. Soit également $\Phi : I \to J$. Alors*

$$\forall j \in J \ : \ \mathbb{P}(\Phi(X) = j) = \sum_{\{i \in I;\ \Phi(i) = j\}} \mathbb{P}(X = i).$$

Démonstration. Si $\omega \in \Omega$ et si $j \in J$, alors

$$\Phi(X(\omega)) = j \iff \exists i \in I \ : \ \big(X(\omega) = i \quad \text{et} \quad \Phi(i) = j\big).$$

Donc

$$\{\Phi(X) = j\} = \bigcup_{i \in I_j} \{X = i\}, \quad \text{où} \quad I_j = \{i \in I;\ \Phi(i) = j\};$$

et comme il s'agit d'une réunion dénombrable disjointe, on en déduit $\mathbb{P}(\Phi(X) = j) = \sum_{i \in I_j} \mathbb{P}(X = i)$, ce qui est le résultat souhaité. □

Exemple. Soient X_1 et X_2 deux va à valeurs dans $\mathbb{N}$ définies sur Ω. Alors, pour tout $n \in \mathbb{N}$, on a

$$\mathbb{P}(X_1 + X_2 = n) = \sum_{k=0}^{n} \mathbb{P}(X_1 = k, X_2 = n - k).$$

Démonstration. On applique la proposition avec la va $X = (X_1, X_2)$ – à valeurs dans $\mathbb{N} \times \mathbb{N}$ – et la fonction $\Phi : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ définie par $\Phi(u, v) = u + v$. Le résultat est

$$\begin{aligned}\mathbb{P}(X_1 + X_2 = n) &= \sum_{\{(u,v)\in\mathbb{N}\times\mathbb{N};\ u+v=n\}} \mathbb{P}\big((X_1, X_2) = (u, v)\big) \\ &= \sum_{k=0}^{n} \mathbb{P}\big((X_1, X_2) = (k, n-k)\big) \\ &= \sum_{k=0}^{n} \mathbb{P}(X_1 = k, X_2 = n-k).\end{aligned}$$

□

PROPOSITION 4.2. *Soit $X : (\Omega, \mathfrak{A}, \mathbb{P}) \to \mathbb{R}^d$ une va à densité à valeurs dans $\mathbb{R}^d$, de densité $\rho_X : \mathbb{R}^d \to \mathbb{R}^+$. On suppose que X est ps à valeurs dans un certain ouvert $U \subset \mathbb{R}^d$ (autrement dit que $\rho_x(x) = 0$ ps sur $\mathbb{R}^d \backslash U$). Soit également $\Phi : U \to V$ un difféomorphisme de U sur un autre ouvert $V \subset \mathbb{R}^d$. Alors $Y := \Phi(X)$ est une va à densité, et sa densité ρ_Y est donnée par*

$$\rho_Y(y) = \rho_X(\Psi(y))\,|J_\psi(y)|, \quad \text{où} \quad \Psi = \Phi^{-1}.$$

Démonstration. Pour toute fonction borélienne $f : \mathbb{R}^d \to \mathbb{R}^+$, on a

$$\begin{aligned}\int_\Omega f(\Phi(X))\, d\mathbb{P} &= \int_{\mathbb{R}^d} f(\Phi(x))\,\rho_X(x)dx \\ &= \int_U f(\Phi(x))\,\rho_X(x)dx \qquad \text{car } \rho_X(x) = 0 \text{ ps sur } \mathbb{R}^d\backslash U \\ &= \int_V f(y)\,\rho_X(\Psi(y))|J_\Psi(y)|dy \qquad \text{en posant } y = \Phi(x), \textit{ i.e. } x = \Psi(y) \\ &= \int_{\mathbb{R}^d} f(y)\,\rho_X(\Psi(y))|J_\Psi(y)|dy \qquad \text{car ... ;}\end{aligned}$$

donc $Y = \Phi(X)$ a pour loi $\mathbb{P}_Y = \rho_X(\Psi(y))|J_\Psi(y)|dy$. □

Exercice. Utiliser la proposition précédente pour montrer que si X est une va réelle uniformément distribuée sur $]0, 1[$, alors $Y = -\log(X)$ suit la loi exponentielle $\mathcal{E}(1)$.

5. Variables aléatoires indépendantes

DÉFINITION 5.1. *Soit $(\Omega, \mathfrak{A}, \mathbb{P})$ un espace de probabilité, et soit $(X_i)_{i\in I}$ une famille de va définies sur $(\Omega, \mathfrak{A}, \mathbb{P})$, chaque X_i étant à valeurs dans un certain $(\Lambda_i, \mathfrak{B}_i)$. On dit que* ***les X_i sont indépendantes*** *si "elles n'ont aucune influence les unes sur les autres" ; ce qui se traduit mathématiquement comme suit : pour tous $i_1, \dots, i_r \in I$ deux à deux distincts et pour tous $A_{i_1} \in \mathfrak{B}_{i_1}, \dots, A_{i_r} \in \mathfrak{B}_{i_r}$, on a*

$$\mathbb{P}(X_{i_1} \in A_{i_1}, \dots, X_{i_r} \in A_{i_r}) = \mathbb{P}(X_{i_1} \in A_{i_1}) \cdots \mathbb{P}(X_{i_1} \in A_{i_r}).$$

En particulier, deux *va X et X' à valeurs dans $(\Lambda, \mathfrak{B})$ et $(\Lambda', \mathfrak{B}')$ respectivement sont dites indépendantes si on a*

$$\mathbb{P}(X \in A, X' \in A') = \mathbb{P}(X \in A)\,\mathbb{P}(X \in A')$$

pour tous $A \in \mathfrak{B}$ et $A' \in \mathfrak{B}'$.

REMARQUE. Quand il n'y a qu'un nombre fini de va, la définition se simplifie : des va $X_1, \dots, X_d$ à valeurs dans $(\Lambda_1, \mathfrak{B}_1), \dots, (\Lambda_d, \mathfrak{B}_d)$ sont indépendantes si et seulement si

$$(*) \qquad \mathbb{P}(X_1 \in A_1, \dots, X_d \in A_d) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_d \in A_d)$$

pour tous $A_1 \in \mathfrak{B}_1, \dots, A_d \in \mathfrak{B}_d$.

Démonstration. Exo. □

EXEMPLE 1. On joue n fois à pile ou face Pour $i = 1, \dots, n$, on pose $X_i = 0$ si le i-ème jet donne "pile", et $X_i = 1$ si le i-ème jet donne "face". Alors les va $X_1, \dots, X_n$ sont indépendantes.

Démonstration. On modélise par $\Omega := \{0,1\}^n$ muni de la loi uniforme, en identifiant 0 à "pile" et 1 à "face". Alors $X_i : \Omega \to \{0,1\}$ est formellement définie par $X_i(\omega) = \omega_i$ pour $\omega = (\omega_1, \dots, \omega_n) \in \Omega$.

Soient $A_1, \dots, A_n \subseteq \{0,1\}$. Par définition, on a

$$\{X_1 \in A_1, \dots, X_n \in A_n\} = A_1 \times \cdots \times A_n.$$

Donc, par définition de la loi uniforme,

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \frac{\#(A_1 \times \cdots \times A_n)}{\#\Omega} = \frac{\#A_1 \cdots \#A_n}{2^n};$$

autrement dit

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \Big(\frac{\#A_1}{2}\Big) \cdots \Big(\frac{\#A_n}{2}\Big).$$

Par ailleurs, on a aussi $\{X_i \in A_i\} = \{0,1\} \times \cdots \times A_i \times \cdots \times \{0,1\}$, donc

$$\mathbb{P}(X_i \in A_i) = \frac{\#\big(\{0,1\} \times \cdots \times A_i \times \cdots \times \{0,1\}\big)}{\#\Omega} = \frac{2^{n-1}\#A_i}{2^n} = \frac{\#A_i}{2}.$$

Donc $\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n)$. □

GÉNÉRALISATION. Soient $\Omega_1, \dots, \Omega_n$ des ensembles finis, et soit $\Omega = \Omega_1 \times \cdots \times \Omega_n$ muni de la loi uniforme. Pour $i = 1, \dots, n$, soit $X_i : \Omega \to \Omega_i$ la va définie par $X_i(\omega_1, \dots, \omega_n) = \omega_i$. Alors les X_i sont indépendantes.

Démonstration. Exo. □

EXEMPLE 2. Soit $\Omega = [0,1] \times [0,1]$, muni de la mesure de Lebesgue (c'est bien un espace de probabilité puisque $\lambda_2([0,1] \times [0,1]) = 1$). On note $X : \Omega \to [0,1]$ et $Y : \Omega \to [0,1]$ les va définies par $X(x,y) = x$ et $Y(x,y) = y$. Alors X et Y sont indépendantes.

Démonstration. Pour tous boréliens $A, B \subseteq [0,1]$, on a

$$\{X \in A, Y \in B\} = A \times B.$$

Donc

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B) &= \lambda_2(A \times B) \\ &= \lambda_1(A)\, \lambda_1(B) \\ &= \lambda_2(A \times [0,1])\, \lambda_2([0,1] \times B) \\ &= \mathbb{P}(X \in A)\, \mathbb{P}(Y \in B). \end{aligned}$$

□

EXEMPLE 3. Soit $(\Omega, \mathfrak{A}, \mathbb{P})$ un espace de probabilité (quelconque), et soit $(E_i)_{i\in I}$ une famille d'évènements. Alors les va $\mathbf{1}_{E_i}$ sont indépendantes si et seulement si les évènements E_i sont indépendants.

Démonstration. Exo non évident. (Il faut utiliser le Fait 4.2 du Chapitre 2.) □

LEMME 5.2. *Soient X et Y des va* discrètes *définies sur $(\Omega, \mathfrak{A}, \mathbb{P})$ et à valeurs dans des univers dénombrables I et J. Pour pouvoir conclure que X et Y sont indépendantes, il suffit d'avoir vérifié que*

$$(*) \qquad \mathbb{P}(X = i, Y = j) = \mathbb{P}(X = i)\,\mathbb{P}(Y = j) \qquad \textit{pour tout } i \in I \textit{ et pour tout } j \in J.$$

Démonstration. Supposons $(*)$ vérifiée, et soient $A \subseteq I$ et $B \subseteq J$ quelconques. On a

$$\{X \in A, Y \in B\} = \bigcup_{(i,j)\in A\times B} E_{i,j}, \quad \text{où} \quad E_{i,j} = \{X = i, Y = j\}.$$

Comme les $E_{i,j}$ sont deux à deux disjoints et que $A \times B$ est dénombrable, on en déduit

$$\begin{aligned}
\mathbb{P}(X \in A, Y \in B) &= \sum_{(i,j)\in A\times B} \mathbb{P}(E_{i,j}) \\
&= \sum_{(i,j)\in A\times B} \mathbb{P}(X = i, Y = j) \\
&= \sum_{(i,j)\in A\times B} \mathbb{P}(X = i)\,\mathbb{P}(Y = j) \qquad \text{par } (*) \\
&= \left(\sum_{i\in A} \mathbb{P}(X = i)\right)\left(\sum_{j\in B} \mathbb{P}(Y = j)\right) \\
&= \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B).
\end{aligned}$$

Donc X et Y sont indépendantes. □

ATTENTION. Le lemme n'est valable que pour des va discrètes.

6. Fonction de répartition d'une va réelle

DÉFINITION 6.1. *Soit X une va* réelle *définie sur un espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$. La **fonction de répartition** de X est la fonction $F_X : \mathbb{R} \to \mathbb{R}$ définie par*

$$F_X(t) = \mathbb{P}(X \leqslant t) = \mathbb{P}_X\big(]-\infty, t]\big) \qquad \textit{pour tout } t \in \mathbb{R}.$$

EXEMPLE 1. Supposons que X soit une va à densité, de densité associée $\rho_X : \mathbb{R} \to \mathbb{R}^+$. Alors on a

$$F_X(t) = \int_{-\infty}^{t} \rho_X(s)\,ds \qquad \text{pour tout } t \in \mathbb{R}.$$

En particulier, pour une va à densité on peut dire les choses suivantes.

(1) La fonction F_X est continue sur $\mathbb{R}$.

(2) Si la densité $\rho_X : \mathbb{R} \to \mathbb{R}^+$ est *continue*, alors F_X est de classe $\mathcal{C}^1$ et $F_X' = \rho_X$. Plus précisément, F_X est l'unique primitive de ρ_X qui tend vers 0 en $-\infty$, ou encore l'unique primitive de ρ_X qui tend vers 1 en $+\infty$.

Démonstration. Tout est évident sauf (1). Soit $t_0 \in \mathbb{R}$ fixé, et soit $(t_n)_{n\geqslant 1}$ une suite tendant vers t. Pour tout $n \in \mathbb{N}$, on a

$$F_X(t_n) = \int_{-\infty}^{t_n} \rho_X(s)\, ds = \int_{\mathbb{R}} \mathbf{1}_{[s,\infty[}(t_n)\rho_X(s)\, ds.$$

Si $s \in \mathbb{R}$, la fonction $\mathbf{1}_{[s,\infty[}$ est continue en tout point $t \neq s$. Comme $t_n \to t_0$, on en déduit que $\mathbf{1}_{[s,\infty[}(t_n) \to \mathbf{1}_{[s,\infty[}(t_0)$ pour tout $s \neq t_0$, donc pour *presque tout* $s \in \mathbb{R}$. Donc, par convergence dominée (exo),

$$F_X(t_n) = \int_{\mathbb{R}} \mathbf{1}_{[s,\infty[}(t_n)\rho_X(s)\, ds. \to \int_{\mathbb{R}} \mathbf{1}_{[s,\infty[}(t_0)\rho_X(s)\, ds = F_X(t_0).$$

Ainsi, F_X est continue en tout point $t_0 \in \mathbb{R}$. □

Exemple 2. Soit X une va réelle suivant une loi de Bernoulli de paramètre $p \in [0,1]$. On a

$$F_X(t) = \begin{cases} 0 & \text{si} \quad t < 0 \\ 1-p & \text{si} \quad 0 \leqslant t < 1 \\ 1 & \text{si} \quad t \geqslant 1 \end{cases}$$

Démonstration. Micro-exo. □

Exemple 3. Soit X une va réelle uniformément distribuée sur $]0,1[$, *i.e.* $\mathbb{P}_X = \mathbf{1}_{]0,1[}(x)dx$. On a

$$F_X(t) = \begin{cases} 0 & \text{si} \quad t \leqslant 0 \\ t & \text{si} \quad 0 < t < 1 \\ 1 & \text{si} \quad t \geqslant 1 \end{cases}$$

Démonstration. Micro-exo à nouveau. □

Exercice. Déterminer F_X quand X est une va prenant un nombre fini de valeurs $a_1, \ldots, a_N$, avec $\mathbb{P}(X = a_i) = p_i$.

Fait 6.2. *Si X est un va réelle quelconque, sa fonction de répartition possède les propriétés suivantes :*

(i) *F_X est croissante et* continue à droite *en tout point ;*

(ii) *$0 \leqslant F_X \leqslant 1$, $\lim_{t\to-\infty} F_X(t) = 0$ et $\lim_{t\to+\infty} F_X(t) = 1$.*

Démonstration. (i) Il est évident que F_X est croissante (et donc admet une limite à gauche et à droite en tout point). On a d'une part $F_X(t^+) = \lim_{n\to\infty} F_X(t + \frac{1}{n}) = \lim_{n\to\infty} \mathbb{P}(X \leqslant t + \frac{1}{n})$; et d'autre part $\{X \leqslant t\} = \bigcap_{n\in\mathbb{N}^*}\{X \leqslant t + \frac{1}{n}\}$ où l'intersection est décroissante. Comme $\mathbb{P}$ est une mesure finie, on en déduit $F_X(t) = \mathbb{P}(X \leqslant t) = \lim_{n\to\infty} \mathbb{P}(X \leqslant t + \frac{1}{n}) = F_X(t^+)$; donc F_X est continue à droite en tout point $t \in \mathbb{R}$.
La partie (ii) est laissée en exo. □

Fait 6.3. *Si X est une va réelle, on a pour tout $t \in \mathbb{R}$:*

$$\mathbb{P}(X = t) = \mathbb{P}_X(\{t\}) = F_X(t) - F_X(t^-).$$

En particulier, F_X est continue en un point t si et seulement si $\mathbb{P}_X(\{t\}) = 0$; et donc F_X est continue sur $\mathbb{R}$ si et seulement si $\mathbb{P}_X$ est une mesure diffuse.

Démonstration. On a $\mathbb{P}_X(\{t\}) = \mathbb{P}(X = t) = \mathbb{P}(X \leqslant t) - \mathbb{P}(X < t) = F_X(t) - \mathbb{P}(X < t)$. De plus, $\{X < t\} = \bigcup_{n\in\mathbb{N}^*}\{X \leqslant t - \frac{1}{n}\}$ où la réunion est croissante, donc $\mathbb{P}(X < t) = \lim_{n\to\infty} \mathbb{P}(X \leqslant t - \frac{1}{n}) = \lim_{n\to\infty} F_X(t - \frac{1}{n}) = F_X(t^-)$. Ainsi $\mathbb{P}_X(\{t\}) = F_X(t) - F_X(t^-)$. Comme de plus F_X est continue à droite, on en déduit que F_X est continue au point t si et seulement si $\mathbb{P}_X(\{t\}) = 0$. □

THÉORÈME 6.4. *La fonction de répartition* caractérise la loi *: si X_1 et X_2 sont deux va réelles telles que $F_{X_1} = F_{X_2}$, alors $\mathbb{P}_{X_1} = \mathbb{P}_{X_2}$.*

Démonstration. On l'a déjà démontré : c'est le Corollaire 1.5 du Chapitre 2. □

COROLLAIRE 6.5. *Soit X une va réelle. Si F_X est de classe $\mathcal{C}^1$ sur $\mathbb{R}$, alors X est à densité. Plus généralement, si F_X est continue sur $\mathbb{R}$ et de classe $\mathcal{C}^1$ sur $\mathbb{R}\backslash K$ où K est un ensemble fini, alors X est à densité, de densité $\rho_X = F_X'$.*

Démonstration. La fonction $\rho = F_X'$ est définie sur $\mathbb{R}\backslash K$, donc presque partout ; et $\rho \geqslant 0$ car la fonction F_X est croissante.

FAIT. Pour tous $a, b \in \mathbb{R}$ tels que $a < b$, on a

$$(*) \qquad F_X(b) - F_X(a) = \int_a^b \rho(x)\,dx.$$

Preuve du Fait. Soient $t_1 < \cdots < t_{N-1}$ les points de K compris entre a et b, et posons $t_0 := a$ et $t_N = b$. On a

$$F_X(b) - F_X(a) = \sum_{k=0}^{N-1} \big(F_X(t_{k+1}) - F_X(t_k)\big).$$

De plus $F_X(t_{k+1}) - F_X(t_k) = \int_{t_k}^{t_{k+1}} F_X'(x)\,dx$ pour $k = 0, \ldots, N-1$ car F est $\mathcal{C}^1$ sur $]t_k, t_{k+1}[$ et continue sur $[t_k, t_{k+1}]$ (exo). D'où $(*)$ par Chasles. □

En prenant $b = t$ quelconque et en faisant tendre a vers $-\infty$ dans $(*)$ on obtient

$$(**) \qquad F_X(t) = \int_{-\infty}^{t} \rho(x)\,dx \qquad \text{pour tout } t \in \mathbb{R}.$$

En faisant maintenant tendre t vers $+\infty$, on en déduit $\int_{-\infty}^{\infty} \rho(x)\,dx = 1$. Donc ρ est une densité lebesguienne. Si on choisit une va $\widetilde{X}$ de loi $\mathbb{P}_{\widetilde{X}} = \rho(x)dx$, on obtient alors $F_X = F_{\widetilde{X}}$ par $(**)$, et donc $\mathbb{P}_X = \mathbb{P}_{\widetilde{X}} = \rho(x)dx$ puisque la fonction de répartition caractérise la loi. □

PROPOSITION 6.6. *Toute fonction $F : \mathbb{R} \to \mathbb{R}$ croissante, continue à droite en tout point et vérifiant $\lim_{t\to-\infty} F(t) = 0$ et $\lim_{t\to+\infty} F(t) = 1$, est une fonction de répartition.*

Démonstration. On va la faire en supposant que F est *continue* et *strictement croissante*. Dans ce cas F est une bijection de $\mathbb{R}$ sur $]0,1[$ et $F^{-1} :]0,1[\to \mathbb{R}$ est borélienne (en fait, continue). Soit alors U une va réelle uniformément distribuée sur $]0,1[$, *i.e.* $\mathbb{P}_U = \mathbf{1}_{]0,1[}(x)dx$. Si on pose $X = F^{-1}(U) := F^{-1} \circ U$, alors X est une va bien définie puisque F^{-1} est borélienne ; et comme F est strictement croissante, on a $F_X(t) = \mathbb{P}(F^{-1}(U) \leqslant t) = \mathbb{P}(U \leqslant F(t)) = F(t)$ pour tout $t \in \mathbb{R}$, par définition de la loi uniforme ($\mathbb{P}(U \leqslant F(t)) = \lambda_1(]0,1[\cap]-\infty F(t)]) = \lambda_1(]0, F(t)]) = F(t)$ car $0 < F(t) < 1$). Donc $F_X = F$. □

Chapitre 4

Produits

1. Univers produits

DÉFINITION 1.1. *Soit $(\Omega_i, \mathfrak{A}_i)_{i\in I}$ une famille dénombrable d'univers probabilisables, et soit $\Omega = \prod_{i\in I} \Omega_i$. Les éléments de Ω sont donc des "points" de la forme $\omega = (\omega_i)_{i\in I}$, où $\omega_i \in \Omega_i$ pour tout $i \in I$.*

(1) *Un **cylindre de** Ω est un ensemble $C \subseteq \Omega$ de la forme*

$$C = \{\omega \in \Omega;\ \omega_{i_1} \in A_{i_1}, \dots, \omega_{i_r} \in A_{i_r}\}$$

où $i_1, \dots, i_r \in I$ sont deux à deux distincts et $A_{i_k} \in \mathfrak{A}_{i_k}$ pour $k = 1, \dots, r$. Un tel cylindre se notera toujours

$$C = \ll A_{i_1} \times \cdots \times A_{i_r} \gg .$$

(2) *La **tribu produit** sur Ω est la tribu $\mathfrak{A}$ engendrée par les cylindres. On écrit $\mathfrak{A} = \otimes_{i\in I}\mathfrak{A}_i$*

Remarque 1. La tribu $\mathfrak{A}$ contient tous les singletons ; donc $(\Omega, \mathfrak{A})$ est un univers probabilisable, qu'on appelle l'**univers produit** des univers $(\Omega_i, \mathfrak{A}_i)$.

Démonstration. Soit $\alpha = (\alpha_i)_{i\in I}$ un point de Ω fixé. Alors

$$\{\alpha\} = \bigcap_{i\in I}\{\omega \in \Omega;\ \omega_i = \alpha_i\} = \bigcap_{i\in I} \ll \{\alpha_i\} \gg,$$

donc $\{\alpha\} \in \mathfrak{A}$ car I est dénombrable. □

Remarque 2. On peut parfaitement définir la tribu produit dans le cas d'une famille *non dénombrable* d'univers $(\Omega_i, \mathfrak{A}_i)$. Mais dans ce cas, la tribu $\mathcal{A} = \bigotimes_{i\in I} \mathfrak{A}_i$ ne contient pas nécessairement tous les singletons, et donc $(\Omega, \mathcal{A})$ n'est pas forcément un univers probabilisable au sens de la définition qui a été adoptée.

FAIT IMPORTANT. Pour tout $i \in I$, la "projection canonique" $\pi_i : \Omega \to \Omega_i$ est mesurable.

Démonstration. Si $A \subseteq \Omega_i$, alors $\pi_i^{-1}(A) = \{\omega \in \Omega;\ \omega_i \in A\}$. Donc $\pi_i^{-1}(A)$ est un cylindre de Ω pour tout $A \in \mathfrak{A}_i$, et donc π_i est mesurable. □

CAS PARTICULIER. Supposons qu'il n'y ait qu'un nombre fini d'univers, notés $(\Omega_1, \mathfrak{A}_1), \dots, (\Omega_d, \mathfrak{A}_d)$, donc $\Omega = \Omega_1 \times \cdots \times \Omega_d$.

• Un cylindre C de $\Omega = \Omega_1 \times \cdots \times \Omega_d$ peut toujours s'écrire sous la forme

$$C = \{\omega = (\omega_1, \dots, \omega_d);\ \omega_1 \in A_1, \dots, \omega_d \in A_d\} = A_1 \times \cdots \times A_d,$$

où $A_1 \in \mathfrak{A}_1, \dots, A_d \in \mathfrak{A}_d$. En effet : selon la définition générale, C est *a priori* de la forme $C = \{\omega;\ \omega_{i_1} \in A_{i_1}, \dots, \omega_{i_r} \in A_{i_r}\}$, où $i_1 < \cdots < i_r$; et donc $C = A_1 \times \cdots \times A_d$ en posant $A_i := \Omega_i$ pour $i \notin \{i_1, \dots, i_r\}$.

• La tribu produit se note $\otimes_{i=1}^{d}\mathfrak{A}_i$ ou $\mathfrak{A}_1\otimes\cdots\otimes\mathfrak{A}_d$. Si on a seulement deux univers $(\Omega_1,\mathfrak{A}_1)$ et $(\Omega_2,\mathfrak{A}_2)$, on écrit $\mathfrak{A}_1\otimes\mathfrak{A}_2$.

EXEMPLE 1. Si Ω_1 et Ω_2 sont deux univers *dénombrables*, alors $\mathcal{P}(\Omega_1)\otimes\mathcal{P}(\Omega_2)=\mathcal{P}(\Omega_1\times\Omega_2)$.

Démonstration. La tribu $\mathcal{P}(\Omega_1)\otimes\mathcal{P}(\Omega_2)$ contient tous les singletons de $\Omega_1\times\Omega_2$, donc toutes les parties de $\Omega_1\times\Omega_2$ puisque $\Omega_1\times\Omega_2$ est dénombrable. □

EXEMPLE 2. Si $p,q\in\mathbb{N}^*$, alors $\mathfrak{B}(\mathbb{R}^p)\otimes\mathfrak{B}(\mathbb{R}^q)=\mathfrak{B}(\mathbb{R}^{p+q})$. Si $d\in\mathbb{N}^*$, alors $\mathfrak{B}(\mathbb{R})\otimes\cdots\otimes\mathfrak{B}(\mathbb{R})=\mathfrak{B}(\mathbb{R}^d)$.

Démonstration. Vu en principe dans le cours de théorie de l'intégration. □

Exercice. Montrer que si $\Omega_1,\ldots,\Omega_d$ sont des espaces métriques *séparables*, alors $\mathfrak{B}(\Omega_1\times\times\Omega_d)=\mathfrak{B}(\Omega_1)\otimes\cdots\otimes\mathfrak{B}(\Omega_d)$.

2. Lois produits

2.1. Définition. Dans cette section, on se donne une famille dénombrable d'espaces de probabilité $(\Omega_i,\mathfrak{A}_i,\mathbb{P}_i)_{i\in I}$, on pose $\Omega=\prod_{i\in I}\Omega_i$, et on munit Ω de la tribu produit $\mathfrak{A}=\otimes_{i\in I}\mathfrak{A}_i$.

THÉORÈME 2.1. *Il existe une unique loi de probabilité $\mathbb{P}$ sur l'univers produit $(\Omega,\mathfrak{A})$ telle que*

$$(*)\qquad \mathbb{P}(\ll A_{i_1}\times\cdots\times A_{i_r}\gg)=\mathbb{P}_{i_1}(A_{i_1})\cdots\mathbb{P}_{i_r}(A_{i_r})$$

pour tout cylindre $C=\ll A_{i_1}\times\cdots\times A_{i_r}\gg\subseteq\Omega$. On dit que $\mathbb{P}$ est la ***loi produit*** *des lois $\mathbb{P}_i$, $i\in I$; et on écrit $\mathbb{P}=\otimes_{i\in I}\mathbb{P}_i$.*

Preuve de l'unicité. Soient $\mathbb{P}$ et $\mathbb{P}'$ deux lois de probabilité vérifiant $(*)$. Posons $\mathcal{M}=\{A\in\mathfrak{A};\ \mathbb{P}(A)=\mathbb{P}'(A)\}$, et notons $\mathcal{C}$ la famille des cylindres de Ω. Comme $\mathbb{P}$ et $\mathbb{P}'$ sont des lois de probabilité, la famille $\mathcal{M}$ est une classe monotone ; et $\mathcal{M}$ contient $\mathcal{C}$ par hypothèse sur $\mathbb{P}$ et $\mathbb{P}'$. De plus, $\mathcal{C}$ est un π-système (micro-exo). D'après le Théorème des classes monotones, $\mathcal{M}$ contient la tribu $\sigma(\mathcal{C})$, qui est par définition égale à $\mathfrak{A}$. Donc $\mathbb{P}(A)=\mathbb{P}'(A)$ pour tout $A\in\mathfrak{A}$, *i.e.* $\mathbb{P}=\mathbb{P}'$. □

Preuve de l'existence. Faite en principe dans le cours de théorie de l'intégration dans le cas de 2 univers Ω_1, Ω_2, et même avec deux espaces mesurés $(\Omega_1,\mathcal{A}_1,\mu_1)$ et $(\Omega_1,\mathcal{A}_1,\mu_1)$ seulement supposés sigma-finis. Donc OK pour un nombre fini d'univers $\Omega_1,\ldots,\Omega_d$. On *admettra* le cas général. □

REMARQUE. Le Théorème 2.1 est en fait valable pour une famille quelconque (possiblement non dénombrable) d'espaces de probabilité $(\Omega_i,\mathfrak{A}_i,\mathbb{P}_i)$.

EXEMPLE 1. Soient $\Omega_1,\ldots,\Omega_d$ des ensembles finis, et pour $i=1,\ldots,d$, soit $\mathbb{P}_i$ la loi uniforme sur Ω_i. Alors $\mathbb{P}_1\otimes\cdots\otimes\mathbb{P}_d$ est la loi uniforme sur $\Omega=\Omega_1\times\cdots\times\Omega_d$.

Démonstration. Pour tout $\omega = (\omega_1, \dots, \omega_d) \in \Omega$, on a

$$\begin{aligned}\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_d(\{\omega\}) &= \mathbb{P}(\{\omega_1\} \times \cdots \{\omega_d\}) \\ &= \mathbb{P}_1(\{\omega_1\}) \cdots \mathbb{P}_d(\{\omega_d\}) \\ &= \frac{1}{\#\Omega_1} \cdots \frac{1}{\#\Omega_d} \\ &= \frac{1}{\#\Omega};\end{aligned}$$

d'où le résultat. □

Exemple 2. Soient $\mathbb{P}_1 = \rho_1(x)dx$ et $\mathbb{P}_2 = \rho_2(y)dy$ deux lois à densité sur $\mathbb{R}$. Alors $\mathbb{P}_1 \otimes \mathbb{P}_2 = \rho(x,y)dxdy$, où $\rho(x,y) = \rho_1(x)\rho_2(y)$.

Démonstration. On a pour tous boréliens $A, B \subseteq \mathbb{R}$:

$$\begin{aligned}\mathbb{P}_1 \otimes \mathbb{P}_2(A \times B) &= \mathbb{P}_1(A)\,\mathbb{P}_2(B) \\ &= \left(\int_A \rho_1(x)dx\right)\left(\int_B \rho_2(y)dy\right) \\ &= \int_{A\times B} \rho_1(x)\rho_2(y)dxdy \qquad \text{par Fubini.}\end{aligned}$$

Donc la mesure $\rho(x,y)dxdy$ coïncide avec $\mathbb{P}_1 \otimes \mathbb{P}_2$ sur les cylindres de $\mathbb{R}^2$, et donc ces deux mesures sont égales. □

2.2. Fubini.

Théorème 2.2. *Soient $(\Omega_1, \mathfrak{A}_1, \mathbb{P}_1), (\Omega_2, \mathfrak{A}_2, \mathbb{P}_2)$ deux espaces de probabilité. Si $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ est une fonction mesurable positive ou intégrable sur $\Omega_1 \times \Omega_2$ par rapport à $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$, alors (on a le droit d'écrire)*

$$\begin{aligned}\int_{\Omega_1\times\Omega_2} f\, d\mathbb{P}_1 \otimes \mathbb{P}_2 &= \int_{\Omega_1} \left(\int_{\Omega_2} f(x,y)\, d\mathbb{P}_2(y)\right) d\mathbb{P}_1(x) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(x,y)\, d\mathbb{P}_1(x)\right) d\mathbb{P}_2(y).\end{aligned}$$

Démonstration. Faite en principe dans le cours de théorie de l'intégration. □

Remarque 1. Le Théorème de Fubini reste valable dans le cas d'une mesure produit $\mu_1 \otimes \mu_2$ "générale" (*i.e.* avec des mesures μ_2, μ_2 seulement supposées sigma-finies).

Remarque 2. Pour montrer qu'une fonction mesurable $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ est intégrable par rapport à $\mu_1 \otimes \mu_2$, on peut appliquer le "cas positif" du Théorème de Fubini à la fonction $|f|$. Ainsi, f est intégrable sur $\Omega_1 \otimes \Omega_2$ si et seulement si

$$\int_{\Omega_1} \left(\int_{\Omega_2} |f(x,y)|\, d\mu_2(y)\right) d\mu_1(x) < \infty,$$

si et seulement si

$$\int_{\Omega_2} \left(\int_{\Omega_1} |f(x,y)|\, d\mu_1(x)\right) d\mu_2(y) < \infty.$$

3. Produits et indépendance

3.1. Variables aléatoires produits.

LEMME 3.1. *Soit $(\Omega, \mathfrak{A}, \mathbb{P})$ un espace de probabilité, et soit $(\Lambda_i, \mathfrak{A}_i)_{i\in I}$ une famille dénombrable d'univers probabilisables. On munit $\Lambda = \prod_{i\in I} \Lambda_i$ de la tribu produit.*

(1) *Si, pour tout $i \in I$, on se donne une va $X_i : (\Omega, \mathfrak{A}, \mathbb{P}) \to \Lambda_i$, alors $X = (X_i)_{i\in I}$ est une va à valeurs dans $\Lambda = \prod_{i\in I} \Lambda_i$. La loi $\mathbb{P}_X = \mathbb{P}_{(X_i)_{i\in I}}$ s'appelle la **loi jointe** des va X_i.*

(2) *Inversement, si X est une va (définie sur Ω) à valeurs dans $\prod_{i\in I} \Lambda_i$ et si on écrit $X(\omega) = (X_i(\omega))_{i\in I}$, alors les X_i sont des va. Les lois $\mathbb{P}_{X_i}$ s'appellent les **marginales** de la loi $\mathbb{P}_X$.*

Démonstration. (1) Il s'agit de montrer que l'application X est mesurable de $(\Omega, \mathfrak{A})$ dans $(\Lambda, \otimes_{i\in I}\mathfrak{A}_i)$; et comme la tribu $\otimes_{i\in I}\mathfrak{A}_i$ est engendrée par les cylindres, il suffit de vérifier que $X^{-1}(C) \in \mathfrak{A}$ pour tout cylindre $C = \ll A_{i_1} \times \cdots \times A_{i_r} \gg$. Mais ceci est évident : on a

$$X^{-1}(\ll A_{i_1} \times \cdots \times A_{i_r} \gg) = \{X_{i_1} \in A_{i_1}\} \cap \cdots \{X_{i_r} \in A_{i_r}\},$$

et donc $X^{-1}(\ll A_{i_1} \times \cdots \times A_{i_r} \gg) \in \mathfrak{A}$ puisque les X_i sont mesurables.

(2) On a $X_i = \pi_i \circ X$, où $\pi_i = \Lambda \to \Lambda_i$ est la "projection canonique" ; donc X_i est mesurable car X et π_i le sont. □

EXEMPLE 1. Soit $(\Omega_i, \mathfrak{A}_i, \mathbb{P}_i)_{i\in I}$ une famille dénombrable d'espaces de probabilité, et soit $\Omega = \prod_{i\in I} \Omega_i$ muni de tribu produit $\mathfrak{A} = \otimes_{i\in I}\mathfrak{A}_i$ et de la loi produit $\mathbb{P} = \otimes_{i\in I}\mathbb{P}_i$. Notons $X : (\Omega, \mathfrak{A}, \mathbb{P}) \to (\Omega, \mathfrak{A})$ la va définie par $X(\omega) = \omega$. Alors $\mathbb{P}_{X_i} = \mathbb{P}_i$ pour tout $i \in I$.

Démonstration. Fixons $i \in I$. Par définition, $X_i : \Omega \to \Omega_i$ est la "projection canonique" d'indice i. Pour tout $A_i \in \mathfrak{A}_i$, l'ensemble $\{X_i \in A_i\}$ est le cylindre $\ll A_i \gg \subseteq \Omega$; donc $\mathbb{P}(X_i \in A_i) = \mathbb{P}_i(A_i)$ par définition de $\mathbb{P}$. Ainsi, $\mathbb{P}_{X_i} = \mathbb{P}_i$. □

EXEMPLE 2. Si Z est une va à densité à valeurs dans $\mathbb{R}^d$, alors ses lois marginales sont à densité.

Démonstration. On la fait pour $d = 2$. Soit Z une v.a. à densité à valeurs dans $\mathbb{R}^2$ définie sur $(\Omega, \mathfrak{A}, \mathbb{P})$, de densité $\rho : \mathbb{R}^2 \to \mathbb{R}^+$. On écrit $Z = (X, Y)$. Pour toute fonction borélienne $f : \mathbb{R} \to \mathbb{R}^+$, on a $\int_\Omega f(X)\, d\mathbb{P} = \int_\Omega g(Z)\, d\mathbb{P}$, où $g : \mathbb{R}^2 \to \mathbb{R}^+$ est la fonction définie par $g(x, y) := f(x)$. Donc $\int_\Omega f(X)\, d\mathbb{P} = \int_{\mathbb{R}^2} g(x, y)\, d\mathbb{P}_Z(x, y) = \int_{\mathbb{R}^2} f(x)\, \rho(x, y) dx dy = \int_{\mathbb{R}} g(x) \left(\int_{\mathbb{R}} \rho(x, y) dy\right) dx$. Donc X est à densité, avec $\rho_X(x) := \int_{\mathbb{R}} \rho(x, y) dy$. □

3.2. Va produits et indépendance.

LEMME 3.2. *Soit $X = (X_i)_{i\in I}$ une famille dénombrable de va définies sur une même espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$, chaque X_i étant à valeurs dans un univers probabilisable $(\Omega_i, \mathfrak{A}_i)$. On considère X comme une va à valeurs dans $\prod_{i\in I} \Omega_i$. Alors les X_i sont indépendantes si et seulement si $\mathbb{P}_X = \otimes_{i\in I}\mathbb{P}_{X_i}$.*

Démonstration. Par définition, les X_i sont indépendantes si et seulement si on a

$$\mathbb{P}(X_{i_1} \in A_{i_1}, \ldots, X_{i_r} \in A_{i_r}) = \mathbb{P}(X \in A_{i_1}) \cdots \mathbb{P}(X_{i_r} \in A_{i_r})$$

pour tous $i_1, \dots, i_r \in I$ deux à deux distincts et pour tous $A_{i_1} \in \mathfrak{A}_{i_1}, \dots, A_{i_r} \in \mathfrak{A}_{i_r}$. Il revient au même de dire que

$$\mathbb{P}_X(\ll A_{i_1} \times \cdots \times A_{i_r} \gg) = \mathbb{P}_{X_{i_1}}(A_{i_1}) \cdots \mathbb{P}_{X_{i_r}}(A_{i_r})$$

pour tout cylindre $\ll A_{i_1} \times \cdots \times A_{i_r} \gg \subseteq \prod_{i\in I} \Omega_i$; ce qui signifie exactement que $\mathbb{P}_X = \otimes_{i\in I} \mathbb{P}_{X_i}$. □

Corollaire 3.3. *Si $(\mathbb{P}_i)_{i\in I}$ est une famille (dénombrable) quelconque de lois de probabilité, chaque $\mathbb{P}_i$ étant définie sur un univers $(\Omega_i, \mathfrak{A}_i)$, alors il existe une famille $(X_i)_{i\in I}$ de va définies sur un même espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$ telle que les X_i sont indépendantes et $\mathbb{P}_{X_i} = \mathbb{P}_i$ pour tout $i \in I$.*

Démonstration. On prend $\Omega = \prod_{i\in I} \Omega_i$ muni de la tribu produit et de la loi produit $\mathbb{P} = \otimes_{i\in I} \mathbb{P}_i$, et on applique le lemme à la va $X : \Omega \to \Omega$ définie par $X(\omega) = \omega$; autrement dit en prenant pour $X_i : \Omega \to \Omega_i$ la "projection canonique" d'indice i. On a vu que $\mathbb{P}_{X_i} = \mathbb{P}_i$ pour tout $i \in I$, et bien sûr $\mathbb{P}_X = \mathbb{P}$. Donc $\mathbb{P}_X = \otimes_{i\in I} \mathbb{P}_{X_i}$, et donc les X_i sont indépendantes par le lemme. □

Corollaire 3.4. *Soient X_1, X_2 deux va réelles, et soient $\mathbb{P}_1$, $\mathbb{P}_2$ deux lois de probabilité sur $\mathbb{R}$. Alors, les propriétés suivantes sont équivalentes :*

(i) *X_1 et X_2 sont indépendantes et de lois respectives $\mathbb{P}_1$ et $\mathbb{P}_2$;*

(ii) *la va $X = (X_1, X_2)$ est telle que $\mathbb{P}_X = \mathbb{P}_1 \otimes \mathbb{P}_2$.*

Démonstration. L'implication (i) $\Longrightarrow$ (ii) est évidente par le lemme. Inversement, supposons que $\mathbb{P}_X = \mathbb{P}_1 \otimes \mathbb{P}_2$. Pour tout borélien $A_1 \subseteq \mathbb{R}$, on a $\mathbb{P}_{X_1}(A_1) = \mathbb{P}_X(A_1 \times \mathbb{R}) = \mathbb{P}_1(A_1)\,\mathbb{P}_2(\mathbb{R}) = \mathbb{P}_1(A_1)$; donc $\mathbb{P}_{X_1} = \mathbb{P}_1$, et de même $\mathbb{P}_{X_2} = \mathbb{P}_2$. Par conséquent $\mathbb{P}_{(X_1,X_2)} = \mathbb{P}_X = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$, donc X_1 et X_2 sont indépendantes par le lemme. □

Corollaire 3.5. *Soient X_1 et X_2 deux va réelles, et soient ρ_1 et ρ_2 des densités lebesguiennes sur $\mathbb{R}$. Alors X_1 et X_2 sont indépendantes, à densité et de densités respectives ρ_1 et ρ_2 si et seulement si la va $X = (X_1, X_2)$ a pour loi $\mathbb{P}_X = \rho_1(x)\rho_2(y)dxdy$.*

Démonstration. Exo. □

Exercice. Soient X et Y deux va réelles indépendantes suivant la loi $\mathcal{N}(0,1)$. Déterminer la loi de $Z = (X - Y, X + Y)$ et en déduire que les va $X + Y$ et $X - Y$ sont indépendantes.

Chapitre 5

Indépendance (4ème couche)

1. Sommes de va indépendantes

PROPOSITION 1.1. *Soient X_1 et X_2 deux va réelles indépendantes et à densité, de densités ρ_1 et ρ_2. Alors $X = X_1 + X_2$ est une va à densité, de densité $\rho = \rho_1 * \rho_2$ (convolée de ρ_1 et ρ_2), i.e. $\rho(x) = \int_{\mathbb{R}} \rho_1(x-t)\rho_2(t)\,dt$.*

Remarque. Comme ρ_1 et ρ_2 sont dans $L^1(\mathbb{R})$ on sait que $\rho = \rho_1 * \rho_2$ est bien définie presque partout et appartient à $L^1(\mathbb{R})$. De plus on a par Fubini $\int_{\mathbb{R}} \rho = \left(\int_{\mathbb{R}} \rho_1\right)\left(\int_{\mathbb{R}} \rho_2\right) = 1$. Donc ρ est une densité lebesguienne, et donc l'énoncé a un sens.

Preuve de la proposition. Pour toute fonction borélienne $f : \mathbb{R} \to \mathbb{R}^+$, on a

$$\int_\Omega f(X)\,d\mathbb{P} = \int_\Omega f(X_1 + X_2)\,d\mathbb{P} = \int_\Omega g(X_1, X_2)\,d\mathbb{P},$$

où $g : \mathbb{R}^2 \to \mathbb{R}$ est la fonction définie par

$$g(x, y) = f(x + y).$$

D'après le Théorème de transfert et l'indépendance de X_1 et X_2, on en déduit

$$\int_\Omega f(X)\,d\mathbb{P} = \int_{\mathbb{R}^2} f(x+y)\,d\mathbb{P}_{(X_1,X_2)}(x,y) = \int_{\mathbb{R}^2} f(x+y) d\mathbb{P}_{X_1}(x) d\mathbb{P}_{X_2}(y).$$

Ainsi

$$\begin{aligned}
\int_\Omega f(X_1 + X_2)\,d\mathbb{P} &= \int_{\mathbb{R}^2} f(x+y)\rho_1(x)\rho_2(y)dxdy \\
&= \int_{\mathbb{R}} \rho_2(y) \left(\int_{\mathbb{R}} f(u)\rho_1(u-y)du \right) dy \qquad \text{en posant } u = x + y \\
&= \int_{\mathbb{R}} f(u) \left(\int_{\mathbb{R}} \rho_1(u-y)\rho_2(y)\,dy \right) du \\
&= \int_{\mathbb{R}} f(u)\,\rho_1 * \rho_2(u)du.
\end{aligned}$$

Donc $X = X_1 + X_2$ a pour loi $\rho_1 * \rho_2(u)du$, d'après le Théorème de transfert. □

COROLLAIRE 1.2. *Si X et Y sont des va indépendantes telles que X suit une loi normale $\mathcal{N}(m_X, \sigma_X^2)$ et Y suit une loi normale $\mathcal{N}(m_Y, \sigma_Y^2)$, alors $X + Y$ suit la loi $\mathcal{N}(m, \sigma^2)$, où $m = m_X + m_Y$ et $\sigma^2 = \sigma_X^2 + \sigma_Y^2$.*

Démonstration. (i) Supposons d'abord que $m_X = 0 = m_Y$, *i.e.* que X et Y suivent les lois $\mathcal{N}(0, \sigma_X^2)$ et $\mathcal{N}(0, \sigma_Y^2)$.

Notons ρ_X et ρ_Y les densités de X et Y,

$$\rho_X(u) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{u^2}{2\sigma_X^2}} \qquad \text{et} \qquad \rho_Y(u) = \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{u^2}{2\sigma_Y^2}}.$$

D'après la proposition, on sait que $X+Y$ est une va à densité, de densité

$$\begin{aligned}\rho(x) = \rho_X * \rho_Y(x) &= \frac{1}{2\pi\sigma_X\sigma_Y}\int_{\mathbb{R}} e^{-\frac{(x-t)^2}{2\sigma_X^2}}\, e^{-\frac{t^2}{2\sigma_Y^2}}\, dt \\ &= \frac{1}{2\pi\sigma_X\sigma_Y}\int_{\mathbb{R}} e^{-\frac{1}{2\sigma_X^2\sigma_Y^2}\left(\sigma_Y^2(x-t)^2+\sigma_X^2t^2\right)} dt.\end{aligned}$$

Ensuite, en se souvenant que $\sigma_X^2+\sigma_Y^2=\sigma^2$, on écrit

$$\begin{aligned}\sigma_Y^2(x-t)^2+\sigma_X^2t^2 &= \sigma_Y^2x^2 - 2\sigma_Y^2xt + \sigma^2t^2 \\ &= \sigma_Y^2x^2 + \Big(\sigma t - \frac{\sigma_Y^2}{\sigma}x\Big)^2 - \frac{\sigma_Y^4}{\sigma^2}x^2 \\ &= \frac{\sigma_X^2\sigma_Y^2}{\sigma^2}\,x^2 + \Big(\sigma t - \frac{\sigma_Y^2}{\sigma}x\Big)^2 \\ &= \frac{\sigma_X^2\sigma_Y^2}{\sigma^2}\,x^2 + \sigma^2(t-\tau)^2,\end{aligned}$$

où on a posé $\tau = \frac{\sigma_Y^2}{\sigma^2}x$, qui ne dépend pas de t.
En revenant à l'expression de $\rho(x)$, on en déduit

$$\begin{aligned}\rho(x) &= \frac{1}{2\pi\sigma_X\sigma_Y}\, e^{-\frac{x^2}{2\sigma^2}} \int_{\mathbb{R}} e^{-\frac12\left(\frac{\sigma}{\sigma_X\sigma_Y}(t-\tau)\right)^2} dt \\ &= \frac{1}{2\pi\sigma_X\sigma_Y}\, e^{-\frac{x^2}{2\sigma^2}} \int_{\mathbb{R}} e^{-\frac{u^2}{2}}\, \frac{\sigma_X\sigma_Y}{\sigma} du \qquad \text{en posant} \quad u = \tfrac{\sigma}{\sigma_X\sigma_Y}(t-\tau)\cdot\end{aligned}$$

Comme $\int_{\mathbb{R}} e^{-u^2/2}du = \sqrt{2\pi}$, on obtient donc finalement

$$\rho(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{x^2}{2\sigma^2}},$$

ce qui montre que $X+Y$ suit la loi $\mathcal{N}(0,\sigma^2)$.

(ii) Supposons maintenant m_X et m_Y quelconques. On utilise le fait suivant, qui est souvent utile.

Fait. Si Z est une v.a. suivant un loi normale $\mathcal{N}(m,\sigma^2)$ et si $a,b\in\mathbb{R}$ avec $a>0$, alors la v.a. $\widetilde{Z} := aZ+b$ suit la loi $\mathcal{N}(am+b, a^2\sigma^2)$.

Preuve du Fait. Pour toute fonction borélienne $f:\mathbb{R}\to\mathbb{R}^+$, on a $\int_\Omega f(\widetilde{Z})\,d\mathbb{P} = \int_\Omega f(aZ+b)\,d\mathbb{P} = \int_{\mathbb{R}} f(az+b)\,d\mathbb{P}_Z(z) = \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} f(az+b)e^{-(z-m)^2/2\sigma^2}dz$. En posant $x := az+b$, on obtient donc $\int_\Omega f(\widetilde{Z})\,d\mathbb{P} = \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} f(x)e^{-\frac{1}{2\sigma^2}(\frac{x-b}{a}-m)^2}\frac{dx}{a} = \frac{1}{\sqrt{2\pi}a\sigma}\int_{\mathbb{R}} f(x)e^{-\frac{1}{2a^2\sigma^2}(x-am-b)^2}dx$; d'où le résultat. □

Par le Fait, les v.a. $\widetilde{X} := X-m_X$ et $\widetilde{Y} := Y-m_Y$ suivent les lois $\mathcal{N}(0,\sigma_X^2)$ et $\mathcal{N}(0,\sigma_Y^2)$; et elles sont indépendantes. Donc $\widetilde{X}+\widetilde{Y}$ suit la loi $\mathcal{N}(0,\sigma^2)$ d'après ce qu'on vient de voir ; et donc $X+Y=\widetilde{X}+\widetilde{Y}+m$ suit la loi $\mathcal{N}(m,\sigma^2)$. □

Notation. Si $\Lambda_1,\dots,\Lambda_n$ sont des parties de $\mathbb{R}$, on pose

$$\Lambda_1+\cdots+\Lambda_n = \big\{\lambda_1+\cdots+\lambda_n;\ \lambda_1\in\Lambda_1,\dots,\lambda_n\in\Lambda_n\big\}.$$

PROPOSITION 1.3. *Soient $X_1, \dots, X_n$ des va discrètes indépendantes, à valeurs dans des ensembles dénombrables $\Lambda_1, \dots, \Lambda_n \subseteq \mathbb{R}$, et soit $X = X_1 + \cdots + X_n$. Alors X est une va discrète à valeurs dans $\Lambda = \Lambda_1 + \cdots + \Lambda_n$, et pour tout $\lambda \in \Lambda$, on a*

$$\mathbb{P}(X = \lambda) = \sum_{\lambda_1 + \cdots + \lambda_n = \lambda} \mathbb{P}(X_1 = \lambda_1) \cdots \mathbb{P}(X_n = \lambda_n).$$

Démonstration. En appliquant la Proposition 4.1 du Chapitre 3 à l'application $\Phi : \Lambda_1 \times \cdots \times \Lambda_n \to \Lambda$ définie par $\Phi(\lambda_1, \dots, \lambda_n) = \lambda_1 + \cdots + \lambda_n$, on obtient

$$\mathbb{P}(X = \lambda) = \sum_{\lambda_1 + \cdots + \lambda_n = \lambda} \mathbb{P}(X_1 = \lambda_1, \, \dots, X_n = \lambda_n);$$

d'où le résultat par indépendance. □

COROLLAIRE 1.4. *Si $X_1, \dots, X_n$ sont des va indépendantes suivant une loi de Bernoulli de paramètre p, alors $X = X_1 + \cdots + X_n$ suit la loi binomiale $\mathcal{B}(n, p)$.*

Démonstration. On prend $\Lambda_1 = \cdots = \Lambda_n = \{0, 1\}$. Alors $\Lambda_1 + \cdots + \Lambda_n = [\![0, n]\!]$; et pour tout $k \in [\![0, n]\!]$, on a

$$\mathbb{P}(X = k) = \sum_{\bar{\varepsilon} \in E_k} \mathbb{P}(X_1 = \varepsilon_1) \cdots \mathbb{P}(X_n = \varepsilon_n),$$

où

$$E_k = \{\bar{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n) \in \{0, 1\}^n ; \ \varepsilon_1 + \cdots + \varepsilon_n = k\}.$$

Maintenant, un $\bar{\varepsilon} \in \{0, 1\}^n$ s'identifie à $A_{\bar{\varepsilon}} = \{i \in [\![1, n]\!] ; \ \varepsilon_i = 1\} \subseteq [\![1, n]\!]$. On a $\bar{\varepsilon} \in E_k$ si et seulement si $\# A_{\bar{\varepsilon}} = k$, et dans ce cas $\mathbb{P}(X_1 = \varepsilon_1) \cdots \mathbb{P}(X_n = \varepsilon_n) = p^k (1 - p)^{n-k}$, qui ne dépend pas de $\bar{\varepsilon}$. Donc

$$\mathbb{P}(X = k) = p^k (1 - p)^{n-k} \times \#\big\{A \subseteq [\![1, n]\!] ; \ \# A = k\big\} = \binom{n}{k} p^k (1 - p)^{n-k}.$$

□

COROLLAIRE 1.5. *Si X et Y sont des va indépendantes suivant des lois de Poisson $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$, alors $X + Y$ suit la loi $\mathcal{P}(X + Y)$.*

Démonstration. On prend $\Lambda_1 = \Lambda_2 = \mathbb{N}$, et donc $\Lambda_1 + \Lambda_2 = \mathbb{N}$. Pour tout $k \in \mathbb{N}$, on a

$$\begin{aligned}
\mathbb{P}(X = k) &= \sum_{k_1 + k_2 = k} \mathbb{P}(X_1 = k_1) \, \mathbb{P}(X_2 = k_2) \\
&= \sum_{i=0}^{k} \mathbb{P}(X_1 = i) \, \mathbb{P}(X_2 = k - i) \\
&= \sum_{i=0}^{k} \frac{\lambda^i}{i!} e^{-\lambda} \, \frac{\mu^{k-i}}{(k-i)!} e^{-\mu} \\
&= \frac{e^{-(\lambda+\mu)}}{k!} \sum_{i=0}^{k} \binom{k}{i} \lambda^i \mu^{k-i} = \frac{e^{-(\lambda+\mu)}}{k!} \, (\lambda + \mu)^k.
\end{aligned}$$

□

2. Borel-Cantelli

NOTATIONS. Soit Ω un ensemble, et soit $(E_i)_{i\in\mathbb{N}}$ une suite de parties de Ω. On note $\overline{\lim}\, E_i$ l'ensemble des $\omega \in \Omega$ appartenant à *une infinité de E_i*, et $\underline{\lim}\, E_i$ l'ensemble des $\omega \in \Omega$ appartenant à *tous les E_i à partir d'un certain rang*. Ainsi, pour $\omega \in \Omega$, on a les équivalences suivantes :

$$\omega \in \overline{\lim}\, E_i \iff \forall n \in \mathbb{N}\ \exists i \geqslant n \ : \ \omega \in E_i,$$

$$\omega \in \underline{\lim}\, E_i \iff \exists n \in \mathbb{N}\ \forall i \geqslant n \ : \ \omega \in E_i.$$

Autrement dit,

$$\overline{\lim}\, E_i = \bigcap_{n\in\mathbb{N}} \bigcup_{i\geqslant n} E_i \qquad \text{et} \qquad \underline{\lim}\, E_i = \bigcup_{n\in\mathbb{N}} \bigcap_{i\geqslant n} E_i.$$

Remarque. En particulier, on voit que si $(\Omega, \mathfrak{A}, \mathbb{P})$ est un espace de probabilité et si les E_i sont des évènements, alors $\overline{\lim}\, E_i$ et $\underline{\lim}\, E_i$ sont aussi des évènements.

Exercice. Montrer qu'on a $\mathbf{1}_{\overline{\lim} E_i} = \overline{\lim}\ \mathbf{1}_{E_i}$ et $\mathbf{1}_{\underline{\lim} E_i} = \underline{\lim}\ \mathbf{1}_{E_i}$.

THÉORÈME 2.1. ("Lemme de Borel-Cantelli")
Soit $(\Omega, \mathfrak{A}, \mathbb{P})$ un espace de probabilité, et soit $(E_i)_{i\in\mathbb{N}}$ une suite d'évènements.

(1) *Si on a* $\sum_{i=0}^{\infty} \mathbb{P}(E_i) < \infty$, *alors* $\mathbb{P}(\overline{\lim}\, E_i) = 0$; *et donc* $\mathbb{P}(\underline{\lim}\, E_i^c) = 1$.

(2) *Si on a* $\sum_{i=0}^{\infty} \mathbb{P}(E_i) = \infty$ et si les E_i sont indépendants, *alors* $\mathbb{P}(\overline{\lim}\, E_i) = 1$.

Démonstration. (1) Comme $\overline{\lim}\, E_i = \bigcap_{n\in\mathbb{N}} \bigcup_{i\geqslant n} E_i$, on a

$$\mathbb{P}(\overline{\lim}\, E_i) \leqslant \mathbb{P}\Big(\bigcup_{i\geqslant n} E_i\Big) \leqslant \sum_{i\geqslant n} \mathbb{P}(E_i) \qquad \text{pour tout } n \in \mathbb{N}.$$

Donc $\mathbb{P}(\overline{\lim}\, E_i) = 0$ si la série $\sum \mathbb{P}(E_i)$ converge (en faisant $n \to \infty$) ; et par conséquent $\mathbb{P}(\underline{\lim}\, E_i^c) = 1$ car $\underline{\lim}\, E_i^c = (\overline{\lim}\, E_i)^c$ par définition d'une $\overline{\lim}$ et d'une $\underline{\lim}$.

(2) Supposons que les E_i soient indépendants et qu'on ait $\sum_0^\infty \mathbb{P}(E_i) = \infty$. Comme $\overline{\lim}\, E_i = \bigcap_{n\in\mathbb{N}} \bigcup_{i\geqslant n} E_i$ et comme les évènements $B_n = \bigcup_{i\geqslant n} E_i$ forment une suite décroissante, on a $\mathbb{P}(\overline{\lim}\, E_i) = \lim_{n\to\infty} \mathbb{P}\big(\bigcup_{i\geqslant n} E_i\big)$. Donc il suffit (pour montrer que $\mathbb{P}(\overline{\lim}\, E_i) = 1$) de vérifier qu'on a

$$\mathbb{P}\Big(\bigcup_{i\geqslant n} E_i\Big) = 1 \qquad \text{pour tout } n \in \mathbb{N};$$

autrement dit, que

$$\mathbb{P}\Big(\bigcap_{i\geqslant n} E_i^c\Big) = 0.$$

On va le faire en montrant que $\log \mathbb{P}\Big(\bigcap_{i\geqslant n} E_i^c\Big) = -\infty$.
Pour tout $N \geqslant n$, on a par indépendance des E_i^c :

$$\mathbb{P}\Big(\bigcap_{i=n}^{N} E_i^c\Big) = \prod_{i=n}^{N} \mathbb{P}(E_i^c) = \prod_{i=n}^{N} \big(1 - \mathbb{P}(E_i)\big).$$

Comme $\mathbb{P}\Big(\bigcap_{i=n}^{N} E_i^c\Big)$ tend vers $\mathbb{P}\Big(\bigcap_{i\geqslant n} E_i^c\Big)$ quand $N \to \infty$, on en déduit en prenant le logarithme :

$$\log \mathbb{P}\Big(\bigcap_{i\geqslant n} E_i^c\Big) = \sum_{i=n}^{\infty} \log\big(1-\mathbb{P}(E_i)\big);$$

la somme du membre de droite ayant un sens (fini ou $-\infty$) car tous ses termes sont négatifs. Comme de plus $\log\big(1-\mathbb{P}(E_i)\big) \leqslant -\mathbb{P}(E_i)$ et comme la série $\sum \mathbb{P}(E_i)$ diverge, on obtient donc le résultat souhaité :

$$\log \mathbb{P}\Big(\bigcap_{i\geqslant n} E_i^c\Big) \leqslant -\sum_{i=n}^{\infty} \mathbb{P}(E_i) = -\infty.$$

□

Remarque 1. On dit souvent que (1) est la "partie triviale" du Lemme de Borel-Cantelli ; simplement parce que sa preuve est en effet très simple.

Remarque 2. La partie (2) est en fait encore valable si on suppose seulement que les E_i sont *deux à deux indépendants* ; mais la preuve est un peu plus délicate, et moins naturelle.

Exemple 1. Soit $(X_i)_{i\in\mathbb{N}}$ une suite de va réelles. Si on a $\sum_{i=0}^{\infty} \mathbb{P}(|X_i| \geqslant \varepsilon) < \infty$ pour tout $\varepsilon > 0$, alors X_i tend presque sûrement vers 0.

Démonstration. Soit $A := \big\{\omega \in \Omega;\ X_i(\omega) \to 0\big\}$: il s'agit de montrer que $\mathbb{P}(A) = 1$. Pour $\varepsilon > 0$, posons

$$A_\varepsilon := \big\{\omega \in \Omega;\ |X_i(\omega)| < \varepsilon \quad \text{à partir d'un certain rang}\big\}.$$

Alors

$$A = \bigcap_{\varepsilon>0} A_\varepsilon = \bigcap_{k\in\mathbb{N}^*} A_{1/k}.$$

Donc il suffit de montrer que $\mathbb{P}(A_{1/k}) = 1$ pour tout $k \in \mathbb{N}^*$, puisqu'une intersection dénombrable d'évènements presque sûrs est encore un évènement presque sûr.
Fixons $k \in \mathbb{N}^*$. Par définition, on a $A_{1/k} = \underline{\lim}\, F_i$, où $F_i = \{|X_i| < 1/k\}$; autrement dit $A_{1/k} = \underline{\lim}\, E_i^c$, où $E_i = \{|X_i| \geqslant 1/k\}$. Par hypothèse, on sait que $\sum_{i=0}^{\infty} \mathbb{P}(E_i) < \infty$. Donc $\mathbb{P}(\underline{\lim}\, E_i^c) = 1$ par Borel-Cantelli, ce qui termine la démonstration. □

Exemple 2. Si on joue une infinité de fois à pile ou face, il est presque sûr qu'on obtiendra 350000 "piles" consécutifs une infinité de fois.

Démonstration. On modélise la situation en considérant une suite (X_n) de va indépendantes (définies sur un certain $(\Omega, \mathfrak{A}, \mathbb{P})$) suivant une loi de Bernoulli $\mathcal{B}(1/2)$, l'évènement $\{X_n = 1\}$ correspondant à "pile au n-ième jet".
Soit $K = 350000 - 1$. Il s'agit de montrer que pour presque tout $\omega \in \Omega$, il existe une infinité d'entiers n tels que $X_n(\omega) = X_{n+1}(\omega) = \cdots = X_{n+K}(\omega) = 1$.
Soit $(n_i)_{i\in\mathbb{N}}$ une suite d'entiers telle que $n_{i+1} > n_i + K$ pour tout i, et posons

$$E_i := \{X_{n_i} = 1, \ldots, X_{n_i+K} = 1\}.$$

Les évènements E_i sont indépendants car les va X_n sont indépendantes et les ensembles $[\![n_i, n_i + K]\!]$ sont deux à deux disjoints (exo : se fait tout seul). De plus, toujours par indépendance des X_n, on a

$$\mathbb{P}(E_i) = \frac{1}{2^{K+1}} := \varepsilon \quad \text{pour tout } i \in \mathbb{N}.$$

Comme ε ne dépend pas de i, on a donc $\sum_0^\infty \mathbb{P}(E_i) = \infty$. Donc $\mathbb{P}(\overline{\lim}\, E_i) = 1$ par Borel-Cantelli. Autrement dit : pour presque tout $\omega \in \Omega$, il existe une infinité d'entiers i tels que $X_{n_i}(\omega) = \cdots = X_{n_i+K}(\omega) = 1$. □

EXERCICE. Montrer que si (E_i) est une suite d'évènements quelconque, alors

$$\mathbb{P}(\overline{\lim}\, E_i) \geqslant \overline{\lim}\, \mathbb{P}(E_i).$$

En particulier, si $\inf_i \mathbb{P}(E_i) > 0$ alors $\overline{\lim}\, E_i \neq \varnothing$.

3. Un peu plus de définitions

3.1. Tribu engendré par une va ou une famille de va.

DÉFINITION 3.1. *Soit $(\Omega, \mathfrak{A}, \mathbb{P})$ un espace de probabilité.*

(1) *Si $X : (\Omega, \mathfrak{A}, \mathbb{P}) \to (\Lambda, \mathfrak{B})$ est une va à valeurs dans un certain univers $(\Lambda, \mathfrak{B})$, la **tribu engendrée par** X, notée $\sigma(X)$, est la famille de toutes les parties E de Ω de la forme $E = \{X \in A\} = X^{-1}(A)$, où $A \in \mathfrak{B}$.*

(2) *Si $(X_i)_{i\in}$ est une famille dénombrable de va, $X_i : (\Omega, \mathfrak{A}, \mathbb{P}) \to (\Lambda_i, \mathfrak{B}_i)$, la **tribu engendrée par les** X_i est la tribu engendrée par la va produit $X = (X_i)_{i\in I}$. Cette tribu se note $\sigma\big(X_i, i \in I\big)$. S'il n'y a qu'un nombre fini de va $X_1, \ldots, X_d$, on écrit $\sigma(X_1, \ldots, X_d)$.*

SIGNIFICATION INTUITIVE. Dire qu'un évènement E appartient à $\sigma(X_i, i \in I)$ signifie ceci : le fait que E soit réalisé ou non pour un certain $\omega \in \Omega$ "dépend uniquement des $X_i(\omega)$".

Remarque 1. Si X est une va, alors $\sigma(X)$ est bien une tribu (exo) ; et $\sigma(X) \subseteq \mathfrak{A}$.

Remarque 2. Si $(X_i)_{i\in}$ est une famille dénombrable de va et si $J \subseteq I$, alors $\sigma(X_i, i \in J) \subseteq \sigma(X_i, i \in I)$.

Démonstration. Chaque X_i est à valeurs dans $(\Lambda_i, \mathfrak{B}_i)$. Soit $X = (X_i)_{i\in I}$ et $Y = (X_i)_{i\in J}$. Alors X est à valeurs dans $\Lambda_I = \prod_{i\in I} \Lambda_i$ et Y est à valeurs dans $\Lambda_J = \prod_{i\in J} \Lambda_i$. On veut montrer que pour tout tout ensemble mesurable $A \subseteq \Lambda_J$, l'ensemble $\{Y \in A\}$ appartient à la tribu $\sigma(X_i, i \in I)$.

On a $Y = \Phi(X) = \Phi \circ X$, où $\Phi : \Lambda_I \to \Lambda_J$ est définie par $\Phi\big((\omega_i)_{i\in I}\big) := (\omega_i)_{i\in J}$. Donc $\{Y \in A\} = \{X \in \Phi^{-1}(A)\}$; et donc $\{Y \in A\} \in \sigma(X)$ car Φ est mesurable (exo). □

EXEMPLE. Si $(X_i)_{i\in\mathbb{N}}$ est une suite de va réelles, alors l'évènement "la suite (X_i) converge" appartient à la tribu $\sigma(X_i, i \in \mathbb{N})$.

Démonstration. Notons E l'évènement en question. Un point $\omega \in \Omega$ appartient à E si et seulement si la suite $(X_i(\omega))$ est de Cauchy, ce qui peut s'écrire ainsi :

$$\forall k \in \mathbb{N}^* \; \exists N \; \forall p, q \geqslant N \;:\; |X_q(\omega) - X_p(\omega)| < \frac{1}{k}.$$

On a donc

$$E = \bigcap_{k\in\mathbb{N}^*} \bigcup_{N\in\mathbb{N}} \bigcap_{p,q\geqslant N} E_{p,q,k}, \quad \text{où } E_{p,q,k} = \big\{|X_q - X_p| < 1/k\big\}.$$

Pour p, q, k fixés, on peut écrire $E_{p,q,k} = X^{-1}(A_{p,q,k})$, où X est la va produit $(X_i)_{i\in\mathbb{N}}$ et $A_{p,q,k} = \{x = (x_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}};\ |x_q - x_p| < 1/k\}$. De plus, $A_{p,q,k} = \Phi_{p,q}^{-1}(]-1/k, 1/k[)$, où

$\Phi_{p,q} : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ est définie par $\Phi_{p,q}((x_i)_{i\in\mathbb{N}}) := x_q - x_p$. La fonction $\Phi_{p,q}$ est mesurable car les applications coordonnées $x \mapsto x_q$ et $x \mapsto x_q$ le sont ; donc $A_{p,q,k}$ est mesurable, et donc $E_{p,q,k} = X^{-1}(A_{p,q,k}) \in \sigma(X_i, i \in \mathbb{N})$. □

Exercice. Soit $(\Omega_i, \mathfrak{A}_i)_{i\in I}$ une famille d'univers probabilisables, et soit $\Omega := \prod_{i\in I} \Omega_i$. Montrer que la tribu produit $\otimes_{i\in I}\mathfrak{A}_i$ est engendrée par les "projections canoniques" $\pi_i : \Omega \to \Omega_i$.

3.2. Mesurabilité par rapport à une sous-tribu.

DÉFINITION 3.2. *Soit $(\Omega, \mathfrak{A}, \mathbb{P})$ un espace de probabilité, et soit $\widetilde{\mathfrak{A}}$ une sous-tribu de $\mathfrak{A}$. On dit qu'une va $Y : (\Omega, \mathfrak{A}, \mathbb{P}) \to (M, \mathfrak{T})$ à valeurs dans un univers $(M, \mathfrak{T})$ est $\widetilde{\mathfrak{A}}$-**mesurable** si Y est mesurable de $(\Omega, \widetilde{\mathfrak{A}})$ dans $(M, \mathfrak{T})$; autrement dit, si, pour tout $A \in \mathfrak{T}$, l'ensemble $\{Y \in A\}$ appartient à la sous-tribu $\widetilde{\mathfrak{A}}$.*

Exemple. Soit $E \in \mathfrak{A}$, et soit $\mathfrak{A}_E := \{\varnothing, \Omega, E, E^c\}$. Une va Y est $\mathfrak{A}_E$-mesurable si et seulement si elle est constante sur E et sur E^c.

Démonstration. Supposons que Y soit $\mathfrak{A}_E$-mesurable. Soit $\omega_0 \in E$ (on suppose $E \neq \varnothing$), et soit $\alpha_0 = Y(\omega_0)$. Alors l'ensemble $\{Y = \alpha_0\}$ appartient à $\mathfrak{A}_E$ et contient un point de E (à savoir ω_0), donc $\{Y = \alpha_0\} = E$ ou Ω ; et donc Y est constante sur E. De même, Y est constante sur E^c.
Réciproque : exo. □

FAIT 3.3. *Soit $X : (\Omega, \mathfrak{A}, \mathbb{P}) \to (\Lambda, \mathfrak{B})$ une va. Si $\Phi : (\Lambda, \mathfrak{B}) \to (M, \mathfrak{T})$ est une application mesurable, alors $Y = \Phi(X) = \Phi \circ X$ est une va $\sigma(X)$-mesurable.*

Démonstration. Pour tout $A \in \mathfrak{T}$, on a $\{Y \in A\} = \{\Phi(X) \in A\} = \{X \in \Phi^{-1}(A)\}$, donc $\{Y \in A\} \in \sigma(X)$ puisque $\Phi^{-1}(A) \in \mathfrak{B}$. □

REMARQUE. On peut montrer que ce fait admet une réciproque : si $X : (\Omega, \mathfrak{A}, \mathbb{P}) \to (\Lambda, \mathfrak{B})$ est une va et si $Y : (\Omega, \mathfrak{A}, \mathbb{P}) \to \mathbb{R}$ est une va *réelle* $\sigma(X)$-mesurable, alors Y est de la forme $Y = \Phi(X)$, pour une certaine fonction mesurable $\Phi : \Lambda \to \mathbb{R}$.

Exercice. Soit X une va réelle *étagée*, et soient $x_1, \dots, x_N$ les valeurs distinctes prises par X. Pour $i = 1, \dots, N$, on pose $E_i := \{X = x_i\}$. Montrer qu'une va réelle Y est $\sigma(X)$-mesurable si et seulement si elle est constante sur chaque ensemble E_i.

FAIT 3.4. *Soit $(X_i)_{i\in I}$ une famille dénombrable de va indépendantes, et pour tout $i \in I$, soit Y_i une va $\sigma(X_i)$-mesurable. Alors les Y_i sont indépendantes.*

Démonstration. Pour $i \in I$, la va X_i est à valeurs dans un univers $(\Lambda_i, \mathfrak{B}_i)$, et la va Y_i est à valeurs dans un univers $(M_i, \mathfrak{T}_i)$. Si $i_1, \dots, i_r \in I$ sont deux à deux distincts et si $A_{i_1}, \dots, A_{i_r} \in \mathfrak{T}_i$, on peut écrire $\{Y_{i_k} \in A_{i_k}\} = \{X_{i_k} \in E_{i_k}\}$ pour $k = 1, \dots, r$, où $E_{i_k} \in \mathfrak{B}_{i_k}$; donc

$$\begin{aligned}
\mathbb{P}\big(Y_{i_1} \in A_{i_1}, \dots, Y_{i_r} \in A_{i_r}\big) &= \mathbb{P}\big(X_{i_1} \in E_{i_1}, \dots, X_{i_r} \in E_{i_r}\big) \\
&= \mathbb{P}(X_{i_1} \in E_{i_1}) \cdots \mathbb{P}(X_{i_r} \in E_{i_r}) \quad \text{par indépendance des } X_i \\
&= \mathbb{P}(Y_{i_1} \in A_{i_1}) \cdots \mathbb{P}(Y_{i_r} \in A_{i_r}).
\end{aligned}$$

□

3.3. Un lemme utile.

LEMME 3.5. *Soit $(X_i)_{i\in I}$ une famille dénombrable de va indépendantes. Si $I^1, I^2 \subseteq I$ et $I^1 \cap I^2 = \varnothing$, alors les va produits $X^1 = (X_i)_{i\in I^1}$ et $X^2 = (X_i)_{i\in I^2}$ sont indépendantes.* Par conséquent, *si Y^1 et Y^2 sont deux va telles que Y^1 est $\sigma(X_i, i \in I^1)$-mesurable et Y^2 est $\sigma(X_i, i \in I^2)$-mesurable, alors Y^1 et Y^2 sont indépendantes.*

Démonstration. Chaque X_i est à valeurs dans un univers $(\Lambda_i, \mathcal{B}_i)$.
Posons $\Lambda^1 := \prod_{i\in I^1} \Lambda_i$ et $\Lambda^2 := \prod_{i\in I^2} \Lambda_i$, et notons $\mathfrak{A}^1$ et $\mathfrak{A}^2$ les tribus produits sur Λ^1 et Λ^2. Il s'agit de montrer qu'on a

$$\mathbb{P}(X^1 \in A, X^2 \in B) = \mathbb{P}(X^1 \in A)\,\mathbb{P}(X^2 \in B) \quad \text{pour tous } A \in \mathfrak{A}^1,\ B \in \mathfrak{A}^2.$$

□

CAS 1. A et B sont des cylindres.

On écrit $A = \ll A_{i_1} \times \cdots \times A_{i_r} \gg$ et $B = \text{`` } \ll B_{j_1} \times \cdots \times B_{j_s} \gg$ où $i_1, \ldots, i_r \in I^1$ sont deux à deux distincts et $j_1, \ldots j_s \in I^2$ sont deux à deux distincts. Alors $i_1, \ldots, i_r, j_1, \ldots, j_s$ sont tous distincts car $I^1 \cap I^2 = \varnothing$. Donc, par indépendance des X_i, on a

$$\begin{aligned}\mathbb{P}(X^1 \in A, X^2 \in B) &= \mathbb{P}\big(X_{i_1} \in A_{i_1}, \ldots, X_{i_r} \in A_{i_r}, X_{j_1} \in B_{j_1}, \ldots, X_{j_s} \in B_{j_s}\big)\\ &= \mathbb{P}(X_{i_1} \in A_{i_1}) \cdots \mathbb{P}(X_{i_r} \in A_{i_r})\,\mathbb{P}(X_{j_1} \in B_{j_1}) \cdots \mathbb{P}(X_{j_s} \in B_{j_s})\\ &= \mathbb{P}(X^1 \in A)\,\mathbb{P}(X^2 \in B).\end{aligned}$$

CAS 2. A est un cylindre et B est quelconque.

On fixe un cylindre $A \in \Lambda^1$ et on considère les mesures μ et μ' sur $(\Lambda^2, \mathfrak{A}^2)$ définies par

$$\mu(B) := \mathbb{P}(X^1 \in A, X^2 \in B) = \mathbb{P}_{(X^1, X^2)}(A \times B)$$

et

$$\mu'(B) = \mathbb{P}(X^1 \in A)\,\mathbb{P}(X^2 \in B) = \mathbb{P}_{X^1}(A)\,\mathbb{P}_{X^2}(B).$$

(Exo : vérifier que μ et μ' sont effectivement des mesures.)
Les mesures μ et μ' sont finies, avec $\mu(\Lambda^2) = \mathbb{P}(X^1 \in A) = \mu'(\Lambda^2)$. De plus, on a $\mu(B) = \mu'(B)$ pour tout *cylindre* $B \subseteq \Lambda^2$ d'après le Cas 1. Donc $\mu = \mu'$ d'après le Théorème des classes monotones ; ce qui est la conclusion souhaitée.

CAS 3. A et B sont quelconques.

Même raisonnement, en utilisant le Cas 2 et le Théorème des classes monotones (exo).

Exemple 1. Si X_1, X_2, X_3, X_4 sont des va réelles indépendantes, alors $X_1 + X_3$ et X_2X_4 sont indépendantes.

Démonstration. C'est évident par le Lemme 3.5 puisque $X_1 + X_3$ est $\sigma(X_1, X_2)$-mesurable et X_2X_4 est $\sigma(X_2, X_4)$-mesurable. □

Exemple 2. Si $(X_i)_{i\in\mathbb{N}}$ est une suite de va indépendantes à valeurs dans $[0, 1]$, alors $Y_1 = \sum_{k=0}^{\infty} \frac{X_{2k}}{3^k}$ et $Y_2 = \sum_{k=0}^{\infty} \frac{X_{2k+1}}{5^k}$ sont des va bien définies et indépendantes.

Démonstration. Exo. □

REMARQUE 3.6. Le Lemme 3.5 se généralise comme suit : si $(I^k)_{k\in K}$ est une famille de parties de I deux à deux disjointes, alors les va produits $X^k = (X_i)_{i\in I^k}$ sont indépendantes.

Démonstration. Identique à celle du Lemme 3.5, avec des notations plus lourdes. □

4. Loi du 0-1

DÉFINITION 4.1. *Soit $(X_i)_{i\in\mathbb{N}}$ une suite de va définies sur un même espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$. On dit qu'un évènement $E \in \mathfrak{A}$ est* ***asymptotique relativement aux*** *X_i si $E \in \sigma\big(X_i,\, i \geqslant N\big)$ pour tout $N \in \mathbb{N}$.*

SIGNIFICATION INTUITIVE. Dire que E est asymptotique relativement aux X_i signifie ceci : le fait que E soit réalisé ou non pour un certain $\omega \in \Omega$ "ne dépend que de la suite $(X_i(\omega))_{i\in\mathbb{N}}$ et ne dépend pas des premiers termes de cette suite".

Exemple 1. Si les X_i sont des va réelles, l'évènement "la suite (X_i) converge" est asymptotique relativement aux X_i.

Démonstration. C'est intuitivement clair puisque le fait qu'une suite converge ne dépend pas de ses premiers termes. Pour une preuve un peu plus détaillée, posons $E := \{\omega \in \Omega;\ \text{la suite } (X_i(\omega)) \text{ converge}\}$. On a vu au début du chapitre que E appartient à la tribu $\sigma(X_i,\, i \in \mathbb{N})$. Mais le fait qu'une suite converge ne dépend pas de ses premiers termes ; donc, pour tout $N \in \mathbb{N}$, on a aussi $E = \{\omega \in \Omega;\ \text{la suite } (X_i(\omega))_{i\geqslant N} \text{converge}\}$, et donc $E \in \sigma(X_i,\, i \geqslant N)$. □

Exemple 2. Si les X_i sont des va réelles, l'évènement "$\frac{X_1+\cdots+X_n}{n} \to 0$ quand $n \to \infty$" est asymptotique relativement aux X_i.

Démonstration. C'est intuitivement clair car le fait que $\frac{x_1+\cdots+x_n}{n}$ tende vers 0 ne dépend pas des premiers termes de la suite (x_i). La preuve détaillée est laissée en exo. □

Exemple 3. Si les X_i sont des va réelles, l'évènement "$X_{35} + X_{47} \leqslant 8$" n'est *a priori* pas asymptotique relativement aux X_i.

Démonstration. C'est à nouveau intuitivement clair. Prenons par exemple $\Omega = [0, 1]$ muni de sa tribu borélienne, $X_i(\omega) = 8\omega$ pour $0 \leqslant i \leqslant 47$ et $X_i(\omega) = 0$ pour $i > 47$. Alors $\{X_{35} + X_{47} \leqslant 8\} = [0, 1/2]$, et $\sigma(X_i,\, i > 47) = \{\varnothing, [0, 1]\}$ (exo) ; donc $\{X_{35} + X_{47} \leqslant 8\} \notin \sigma(X_i,\, i > 47)$. □

THÉORÈME 4.2. (Loi du 0 - 1 de Kolmogorov)
Soit $(X_i)_{i\in\mathbb{N}}$ une suite de va indépendantes *définies sur $(\Omega, \mathfrak{A}, \mathbb{P})$. Si $E \in \mathfrak{A}$ est un évènement asymptotique relativement aux X_i, alors $\mathbb{P}(E) = 0$ ou 1.*

Démonstration. Pour $n \in \mathbb{N}$, posons $\mathfrak{A}_n := \sigma(X_0, \ldots, X_n)$. Posons aussi $\mathfrak{A}_\infty := \sigma\big(X_i, i \geqslant 0\big)$.

FAIT 1. La tribu $\mathfrak{A}_\infty = \sigma\big(X_i, i \geqslant 0\big)$ est engendrée par $\bigcup_{n\in\mathbb{N}} \mathfrak{A}_n$.

Preuve du Fait 1. On la laisse en exo. □

FAIT 2. On a $\mathbb{P}(A \cap E) = \mathbb{P}(A)\mathbb{P}(E)$ pour tout $A \in \bigcup_{n\in\mathbb{N}} \mathfrak{A}_n$.

Preuve du Fait 2. Soit $A \in \bigcup_{n\in\mathbb{N}} \mathfrak{A}_n$, et soit n tel que $A \in \mathfrak{A}_n = \sigma(X_0, \ldots, X_n)$. Alors la va $\mathbf{1}_A$ est $\sigma(X_0, \ldots, X_n)$-mesurable. Par ailleurs, E est $\sigma(X_i, i \geqslant n + 1)$-mesurable, donc la va $\mathbf{1}_E$ est $\sigma(X_i, i \geqslant n + 1)$-mesurable. Par le Lemme 3.5, on en déduit que les va $\mathbf{1}_A$ et $\mathbf{1}_E$ sont indépendantes. Donc les évènements A et E sont indépendants. □

FAIT 3. On a $\mathbb{P}(A \cap E) = \mathbb{P}(A)\mathbb{P}(E)$ pour tout $A \in \mathfrak{A}_\infty$.

Preuve du Fait 3. Par le Fait 1, la tribu $\mathfrak{A}_\infty$ est engendré par $\mathcal{C} = \bigcup_{n\in\mathbb{N}} \mathfrak{A}_n$. De plus, $\mathcal{C}$ est un π-système car c'est une réunion *croissante* de tribus (exo) ; et par le Fait 2, on a $\mathbb{P}(A \cap E) = \mathbb{P}(A)\mathbb{P}(E)$ pour tout $A \in \mathcal{C}$. Donc OK en considérant les mesures μ et μ' sur $(\Omega, \mathfrak{A}_\infty)$ définies par $\mu(A) := \mathbb{P}(A \cap E)$ et $\mu'(A) := \mathbb{P}(E)\,\mathbb{P}(A)$ et en appliquant le Théorème des classes monotones (exo). □

Si on applique le Fait 3 avec $A = E$ (qui appartient bien à $\mathfrak{A}_\infty = \sigma(X_i, i \geqslant 0)$), on obtient

$$\mathbb{P}(E) = \mathbb{P}(E)^2,$$

et donc $\mathbb{P}(E) = 0$ ou 1. □

Exercice. Montrer que si (X_k) est une suite de va réelles indépendantes, alors ou bien la série $\sum X_k$ converge presque sûrement, ou bien cette série diverge presque sûrement.

Chapitre 6

Espérance, variance, moments

Toutes les va sont supposées définies sur un même espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$.

Pour $p \in [1, \infty]$, on pose $L^p = L^p(\Omega, \mathbb{P})$.

Comme $\mathbb{P}(\Omega) = 1$, on a $L^\infty \subseteq L^p \subseteq L^1$ pour tout p, et $\| \cdot \|_1 \leqslant \| \cdot \|_p \leqslant \| \cdot \|_\infty$ (conséquence de Hölder).

1. Espérance d'une variable aléatoire

Définition 1.1. *Pour toute va X à valeurs complexes, ou bien intégrable ou bien $\geqslant 0$, on pose $\mathbb{E}(X) = \int_\Omega X \, d\mathbb{P}$. On dit que $\mathbb{E}(X)$ est l'*espérance *de X, ou la* moyenne *de X, ou encore la "valeur attendue" de X.*

Remarque 1. On a $X \in L^1 \iff \mathbb{E}(|X|) < \infty$, et dans ce cas $\|X\|_1 = \mathbb{E}(|X|)$. De même, si $p < \infty$, alors $X \in L^p \iff \mathbb{E}(|X|^p) < \infty$, et dans ce cas $\|X\|_p = \mathbb{E}(|X|^p)^{1/p}$.

Remarque 2. Si X et Y sont des va réelles et dans L^2, alors $\langle X, Y \rangle_{L^2} = \mathbb{E}(XY)$.

Remarque 3. Par définition, l'espérance dépend *linéairement de la va* : si $X, Y \in L^1$ et $a, b \in \mathbb{C}$, alors

$$\mathbb{E}\big(aX + bY\big) = a\,\mathbb{E}(X) + b\,\mathbb{E}(Y).$$

De plus, l'espérance est *croissante* : si X et Y sont des va réelles et dans L^1 avec $X \leqslant Y$, alors $\mathbb{E}(X) \leqslant \mathbb{E}(Y)$; et même $\mathbb{E}(X) < \mathbb{E}(Y)$ sauf si $X = Y$ ps (exo).

Remarque 4. On dit qu'une va $X \in L^1$ est **centrée** si $\mathbb{E}(X) = 0$.

Proposition 1.2. *Soit X une va réelle, et soit $f : \mathbb{R} \to \mathbb{R}$ une fonction borélienne.*

(1) *Si f est $\geqslant 0$, alors*

$$(*) \qquad \mathbb{E}(f(X)) = \int_{\mathbb{R}} f(x) \, d\mathbb{P}_X(x).$$

(2) *On a $f(X) \in L^1$ si et seulement si $\int_{\mathbb{R}} |f(x)| \, d\mathbb{P}_X(x) < \infty$, et dans ce cas $(*)$ est vraie.*

Démonstration. (1) D'après le Théorème de transfert,

$$\mathbb{E}(f(X)) = \int_\Omega f(X) \, d\mathbb{P} = \int_{\mathbb{R}} f \, d\mathbb{P}_X.$$

(2) Exo. □

Corollaire 1.3. *Si X est une va réelle ou bien $\geqslant 0$ ou bien dans L^1, alors*

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

Corollaire 1.4. *Si X et Y sont deux va réelles telles que $\mathbb{P}_X = \mathbb{P}_Y$, alors $X \in L^1 \iff Y \in L^1$, et dans ce cas $\mathbb{E}(X) = \mathbb{E}(Y)$.*

Démonstration. C'est évident par la proposition. □

COROLLAIRE 1.5. *Soit X une va réelle, et soit $f : \mathbb{R} \to \mathbb{R}$ une fonction borélienne ou bien $\geqslant 0$, ou bien telle que $f(X) \in L^1$.*

(a) *Si X est à densité, de densité associée $\rho_X : \mathbb{R} \to \mathbb{R}$, alors*

$$\mathbb{E}(f(X)) = \int_{\mathbb{R}} f(x)\, \rho_X(x) dx.$$

(b) *Si X est une va discrète à valeurs dans un ensemble dénombrable $\Lambda \subseteq \mathbb{R}$, alors*

$$\mathbb{E}(f(X)) = \sum_{a \in \Lambda} f(a)\, \mathbb{P}(X = a).$$

Démonstration. Exo. □

EXEMPLE 0. On a $\mathbb{E}(c) = c$ pour toute va *constante* $X \equiv c$.

Démonstration. C'est évident puisque $\mathbb{P}(\Omega) = 1$. □

REMARQUE. On en déduit que si $X \in L^1$, alors la va $X - \mathbb{E}(X)$ est centrée.

EXEMPLE 1. Si X est une va suivant une loi normale $\mathcal{N}(m, \sigma^2)$, alors $X \in L^1$ et $\mathbb{E}(X) = m$.

Démonstration. On a $\mathbb{E}(|X|) = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} |x|\, e^{-\frac{(x-m)^2}{2\sigma^2}} dx < \infty$ car $f(x) := |x| e^{-\frac{(x-m)^2}{2\sigma^2}}$ est continue sur $\mathbb{R}$ avec $f(x) = O(1/x^2)$ quand $x \to \pm\infty$; donc $X \in L^1$. Ensuite,

$$\begin{aligned}
\mathbb{E}(X) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} x\, e^{-\frac{(x-m)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} (u+m)\, e^{-\frac{u^2}{2\sigma^2}} du \\
&= \frac{1}{\sqrt{2\pi}\sigma} \left(\int_{\mathbb{R}} u\, e^{-\frac{u^2}{2\sigma^2}} du + m \int_{\mathbb{R}} e^{-\frac{u^2}{2\sigma^2}} du \right).
\end{aligned}$$

La première intégrale vaut 0 car on intègre sur $\mathbb{R}$ une fonction impaire, et la deuxième intégrale vaut $\sqrt{2\pi}\sigma$. Donc $\mathbb{E}(X) = m$. □

EXEMPLE 2. Si X est une va suivant une loi de Bernoulli $\mathcal{B}(p)$, alors $\mathbb{E}(X) = p$.

Démonstration. Comme X est à valeurs dans $\{0, 1\}$, on a

$$\mathbb{E}(X) = 1 \times \mathbb{P}(X = 1) + 0 \times \mathbb{P}(X = 0) = p.$$

□

EXEMPLE 3. Si X est une va suivant une loi de Poisson $\mathcal{P}(\lambda)$, alors $\mathbb{E}(X) = \lambda$.

Démonstration. Comme X est à valeurs dans $\mathbb{N}$, on a

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=0}^{\infty} k\,\mathbb{P}(X=k)\\ &= \sum_{k=0}^{\infty} k\,\frac{\lambda^k}{k!}\,e^{-\lambda}\\ &= e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}\\ &= \lambda e^{-\lambda}\underbrace{\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}}_{e^{\lambda}} = \lambda.\end{aligned}$$

□

Exemple 4. Si X est une va suivant une loi de Cauchy, alors $X \notin L^1$.

Démonstration. On a $\mathbb{E}(|X|) = \frac{1}{\pi}\int_{\mathbb{R}} \frac{|x|}{1+x^2}\,dx = \infty$ car $\frac{|x|}{1+x^2} \sim \frac{1}{|x|}$ en $\pm\infty$. □

2. Variance d'une va réelle

Lemme 2.1. *Si X est une une va réelle appartenant à L^2, alors la va $\mathbb{E}(X)\mathbf{1}$ est la va constante la plus proche de X en norme L^2.*

Démonstration. Soit $\mathcal{E} = \mathbb{R}\,\mathbf{1} = \{\text{va constantes}\} \subseteq L^2$. Il suffit de montrer que $X - \mathbb{E}(X)\mathbf{1}$ est orthogonale à $\mathcal{E}$ au sens du produit scalaire de L^2 ; ce qui est facile : si $Z = a\mathbf{1} \in \mathcal{E}$, alors

$$\langle X - \mathbb{E}(X)\mathbf{1}, a\mathbf{1}\rangle_{L^2} = a\langle X, \mathbf{1}\rangle_{L^2} - a\mathbb{E}(X)\langle \mathbf{1}, \mathbf{1}\rangle_{L^2} = a\mathbb{E}(X) - a\mathbb{E}(X)\underbrace{\mathbb{E}(\mathbf{1})}_{=1} = 0.$$

□

Définition 2.2. *Soit X une va réelle appartenant à L^2. La **variance** de X est le nombre*

$$\mathbb{V}(X) = \|X - \mathbb{E}(X)\mathbf{1}\|_{L_2}^2 = \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)^2\Big).$$

*L'**écart type** de X est le nombre $\sigma(X) = \sqrt{\mathbb{V}(X)} = \|X - \mathbb{E}(X)\mathbf{1}\|_{L_2}$. On a donc $\mathbb{V}(X) = \sigma(X)^2$, et on écrit plutôt $\mathbb{V}(X) = \sigma^2(X)$.*

Remarque 1. D'après le Lemme 2.1, on a

$$\sigma(X) = \mathrm{dist}_{L^2}(X, \{\text{va constantes}\})$$

En particulier,

$$\begin{aligned}\sigma^2(X) = 0 &\iff X \text{ est ps égale à une constante}\\ &\iff X = \mathbb{E}(X) \quad \text{ps.}\end{aligned}$$

Remarque 2. La variance dépend "quadratiquement" de la va : pour toute va $X \in L^2$ et pour tout $a \in \mathbb{R}$, on a $\sigma^2(aX) = a^2\sigma^2(X)$.

Démonstration. Exo. □

Remarque 3. La variance "ne voit pas les constantes" : pour toute va $X \in L^2$ et pour toute constante c, on a $\sigma^2(X + c) = \sigma^2(X)$.

Démonstration. Comme $\mathbb{E}(c) = c$, on a $(X+c)-\mathbb{E}(X+c) = X-\mathbb{E}(X)$ par linéarité de l'espérance. □

FAIT 2.3. *Pour toute va réelle $X \in L^2$, on a $\sigma^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.*

Démonstration. On développe la norme au carré :

$$\begin{aligned}\sigma^2(X) = \|X - \mathbb{E}(X)\mathbf{1}\|_{L_2}^2 &= \|X\|_2^2 - 2\langle X, \mathbb{E}(X)\mathbf{1}\rangle_{L^2} + \|\mathbb{E}(X)\mathbf{1}\|_2^2 \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)\langle X, \mathbf{1}\rangle_{L^2} + \mathbb{E}(X)^2\|\mathbf{1}\|_2^{L^2} \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2.\end{aligned}$$

□

EXEMPLE 1. Si X est une va réelle suivant une loi normale $\mathcal{N}(m, \sigma^2)$, alors $X \in L^2$ et $\sigma^2(X) = \sigma^2$.

Démonstration. Le fait que X appartienne à L^2 est laissé en exo. On sait déjà que $\mathbb{E}(X) = m$; donc

$$\begin{aligned}\sigma^2(X) &= \mathbb{E}\big((X-m)^2\big) \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx \qquad \text{(Théorème de transfert)} \\ &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} u^2 e^{-\frac{u^2}{2\sigma^2}}\frac{du}{\sigma} \\ &= \frac{\sigma^2}{\sqrt{2\pi}}\int_{\mathbb{R}} v^2 e^{-\frac{v^2}{2}}\,dv.\end{aligned}$$

Ensuite, on a

$$v^2 e^{-\frac{v^2}{2}} = -v\,\frac{d}{dv}\big(e^{-\frac{v^2}{2}}\big).$$

donc, en intégrant par parties :

$$\int_{\mathbb{R}} v^2 e^{-\frac{v^2}{2}}\,dv = \underbrace{\Big[-ve^{-\frac{v^2}{2}}\Big]_{-\infty}^{\infty}}_{=0} + \int_{\mathbb{R}} e^{-\frac{v^2}{2}}\,dv = \sqrt{2\pi};$$

et donc $\sigma^2(X) = \sigma^2$. □

EXEMPLE 2. Si X suit une loi de Bernoulli $\mathcal{B}(p)$, alors $\sigma^2(X) = p(1-p)$.

Démonstration. On sait déjà que $\mathbb{E}(X) = p$. Donc $\mathbb{E}(X^2) = p$ car $X = X^2$ (la va X est à valeurs dans $\{0,1\}$), et donc $\sigma^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = p - p^2$. □

EXEMPLE 3. Si X suit une loi de Poisson $\mathcal{P}(\lambda)$, alors $\sigma^2(X) = \lambda$.

Démonstration. On sait déjà que $\mathbb{E}(X) = \lambda$. Comme X est à valeurs dans $\mathbb{N}$, on a donc

$$\begin{aligned}\sigma^2(X) &= \mathbb{E}(X^2) - \lambda^2 \\ &= \sum_{k=0}^{\infty} k^2\,\mathbb{P}(X=k) - \lambda^2 \qquad \text{(Théorème de transfert)} \\ &= \sum_{k=1}^{\infty} k^2\,\frac{\lambda^k}{k!}e^{-\lambda} - \lambda^2.\end{aligned}$$

Ensuite, on écrit $k^2 \frac{\lambda^k}{k!} = \lambda k \frac{\lambda^{k-1}}{(k-1)!}$, ce qui donne

$$\begin{aligned}\sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} &= \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \\ &= \lambda \sum_{k=0}^{\infty} (k+1) \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \lambda \left(\sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} + \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \right) \\ &= \lambda\big(\mathbb{E}(X) + 1\big) = \lambda^2 + \lambda.\end{aligned}$$

Donc $\sigma^2(X) = (\lambda^2 + \lambda) - \lambda^2 = \lambda$. □

3. Sommes et produits

FAIT 3.1. *Si $X_1, \dots, X_n$ sont des va réelles appartenant à L^1, alors*

$$\mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n).$$

Démonstration. C'est évident par linéarité de l'espérance. □

COROLLAIRE 3.2. *Si X est une va suivant une loi binomiale $\mathcal{B}(n,p)$, alors $\mathbb{E}(X) = np$.*

Démonstration. Soient $X_1, \dots, X_n$ des va indépendantes suivant la loi de Bernoulli $\mathcal{B}(p)$. On sait que $S = X_1 + \cdots + X_n$ suit la loi $\mathcal{B}(n,p)$; donc X a la même loi que S, et donc $\mathbb{E}(X) = \mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = np$ puisque $\mathbb{E}(X_i) = p$ pour $i = 1, \dots, n$. □

Exercice. Démontrer *directement* le Corollaire 3.2 (sans rien utiliser d'autre que la définition de la loi binomiale).

PROPOSITION 3.3. *Si X et Y sont des va réelles indépendantes et appartenant à L^1, alors $XY \in L^1$ et $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$.*

Démonstration. Par le Théorème de transfert et l'indépendance des va $|X|$ et $|Y|$, on a

$$\begin{aligned}\mathbb{E}(|XY|) = \int_\Omega |XY|\, d\mathbb{P} &= \int_{\mathbb{R}^2} |xy|\, d\mathbb{P}_{(X,Y)}(x,y) \\ &= \int_{\mathbb{R}^2} |xy|\, d\mathbb{P}_X(x) d\mathbb{P}_Y(y) \\ &= \int_{\mathbb{R}} |x| \underbrace{\left(\int_{\mathbb{R}} |y|\, d\mathbb{P}_Y(y) \right)}_{\mathbb{E}(|Y|)} d\mathbb{P}_X(x) \\ &= \mathbb{E}(|Y|) \int_{\mathbb{R}} |x|\, d\mathbb{P}_X(x) \\ &= \mathbb{E}(|X|)\,\mathbb{E}(|Y|).\end{aligned}$$

En particulier $\mathbb{E}(|XY|) < \infty$, donc $XY \in L^1$. Ensuite, le même calcul (en enlevant les valeurs absolues) donne $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$. □

REMARQUE 1. On montre de la même façon que si X et Y sont des va *positives* et indépendantes, alors $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$.

REMARQUE 2. La proposition se généralise à un nombre fini quelconque de va : si $X_1, \dots, X_d$ sont des va réelles indépendantes et appartenant à L^1, alors $X_1 \cdots X_d \in L^1$ et

$$\mathbb{E}(X_1 \cdots X_d) = \mathbb{E}(X_1) \cdots \mathbb{E}(X_d).$$

Démonstration. Identique. □

COROLLAIRE 3.4. *Si X et Y sont des va réelles appartenant à L^2,* indépendantes et centrées*, alors X et Y sont* orthogonales dans L^2.

Démonstration. Par la proposition, $\langle X, Y\rangle_{L^2} = \mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y) = 0$. □

Remarque 1. En fait, il suffit que l'*une des deux* va X ou Y soit centrée.

Remarque 2. Si X et Y sont deux va réelles appartenant L^2, la **covariance** e X et Y est le nombre $\mathrm{Cov}(X, Y)$ défini par

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \langle X - \mathbb{E}(X)\mathbf{1}, Y - \mathbb{E}(Y)\mathbf{1}\rangle_{L^2} \\ &= \mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)\big).\end{aligned}$$

Les va X et Y sont dites **non-corrélées** si on a $\mathrm{Cov}(X, Y) = 0$. Donc :

$$\text{indépendance} \implies \text{non-corrélation}.$$

Exercice. Établir l'identité $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y)$.

COROLLAIRE 3.5. *Si $X_1, \dots, X_n$ sont des va* deux à deux indépendantes *et appartenant à L^2, alors*

$$\sigma^2(X_1 + \cdots + X_n) = \sigma^2(X_1) + \cdots + \sigma^2(X_n).$$

Démonstration. Si on pose $Y_i = X_i - \mathbb{E}(X_i)$, alors les Y_i sont centrées par définition, et deux à deux indépendantes car les X_i le sont (micro-exo). Donc les Y_i sont orthogonales dans L^2. Par le *Théorème de Pythagore*, on a donc

$$\|Y_1 + \cdots + Y_n\|_2^2 = \|Y_1\|_2^2 + \cdots + \|Y_n\|_2^2;$$

donc OK car $\|Y_i\|_2^2 = \|X_i - \mathbb{E}(X_i)\|_2^2 = \sigma^2(X_i)$ pour tout i et $\|Y_1 + \cdots + Y_n\|_2^2 = \|(X_1 + \cdots + X_n) - \mathbb{E}(X_1 + \cdots + X_n)\|_2^2 = \sigma^2(X_1 + \cdots + X_n)$. □

COROLLAIRE 3.6. *Soient $n \in \mathbb{N}$ et $p \in [0, 1]$. Si X est une va suivant la loi binomiale $\mathcal{B}(n, p)$, alors $\sigma^2(X) = np(1 - p)$.*

Démonstration. La va X a la même loi que $S = X_1 + \cdots + X_n$, où les X_i sont indépendantes et suivent la loi de Bernoulli $\mathcal{B}(p)$; donc $\sigma^2(X) = \sigma^2(X_1) + \cdots + \sigma^2(X_n) = np(1 - p)$. □

Exercice. Démontrer directement ce résultat.

PROPOSITION 3.7. *Soient $X_1, \dots, X_d$ des va réelles. Les propriétés suivantes sont équivalentes.*

(1) *Les X_i sont indépendantes.*
(2) *Pour toutes fonctions boréliennes $f_1, \dots, f_d : \mathbb{R} \to \mathbb{R}$ telles que les $f_i(X_i)$ sont dans L^1, on a $\mathbb{E}\big(f_1(X_1) \cdots f_d(X_d)\big) = \mathbb{E}(f_1(X_1)) \cdots \mathbb{E}(f_d(X_d))$.*

(2') *La même formule pour toutes fonctions boréliennes positives* $f_1, \dots, f_d$.

Démonstration. Pour (1) $\Longrightarrow$ (2) et (1) $\Longrightarrow$ (2'), on applique la Proposition 3.3 aux va $\widetilde{X}_i = f_i(X_i)$, qui sont indépendantes si les X_i le sont.

Inversement, si (2) ou (2') est vérifiée, on a pour tous boréliens $A_1, \dots, A_d \subseteq \mathbb{R}$:

$$\begin{aligned}
\mathbb{P}(X_1 \in A_1, \dots, X_d \in A_d) &= \mathbb{E}\big(\mathbf{1}_{\{X_1 \in A_1, \dots, X_d \in A_d\}}\big) \\
&= \mathbb{E}\big(\big(\mathbf{1}_{A_1}(X_1) \cdots \mathbf{1}_{A_d}(X_d)\big)\big) \\
&= \mathbb{E}(\mathbf{1}_{A_1}(X_1)) \cdots \mathbb{E}(\mathbf{1}_{A_d}(X_d)) \qquad \text{par (2) - (2') avec } f_i = \mathbf{1}_{A_i} \\
&= \mathbb{E}\big(\mathbf{1}_{\{X_1 \in A_1\}}\big) \cdots \mathbb{E}\big(\mathbf{1}_{\{X_d \in A_d\}}\big) \\
&= \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_d \in A_d);
\end{aligned}$$

donc les X_i sont indépendantes. □

4. Formules et inégalités (très) utiles

4.1. "Intégration par parties".

Proposition 4.1. *Soit Z une va réelle positive, et soit $\phi : [0, \infty[\to \mathbb{R}$ une fonction croissante, positive et de classe $\mathcal{C}^1$. Alors*

$$\begin{aligned}
\mathbb{E}(\phi(Z)) &= \phi(0) + \int_0^\infty \mathbb{P}\big(Z > t\big)\, \phi'(t)\, dt \\
&= \phi(0) + \int_0^\infty \mathbb{P}\big(Z \geqslant t\big)\, \phi'(t)\, dt.
\end{aligned}$$

Démonstration. La va $\phi(Z)$ est positive, donc $\mathbb{E}(\phi(Z))$ a un sens. Comme ϕ est de classe $\mathcal{C}^1$, on peut écrire $\phi(z) = \phi(0) + \int_0^z \phi'(t)\, dt$ pour tout $z \geqslant 0$. Donc

$$\begin{aligned}
\mathbb{E}(\phi(Z)) &= \int_\Omega \phi(Z(\omega))\, d\mathbb{P}(\omega) \\
&= \int_\Omega \left(\phi(0) + \int_0^{Z(\omega)} \phi'(t)\, dt \right) d\mathbb{P}(\omega) \\
&= \int_\Omega \left(\phi(0) + \int_{[0, Z(\omega)[} \phi'(t)\, dt \right) d\mathbb{P}(\omega) \\
&= \phi(0) + \int_\Omega \left(\int_0^\infty \mathbf{1}_{]t, \infty[}(Z(\omega)) \phi'(t)\, dt \right) d\mathbb{P}(\omega) \qquad \text{car } \mathbb{P}(\Omega) = 1 \\
&= \phi(0) + \int_0^\infty \phi'(t) \left(\int_\Omega \mathbf{1}_{\{Z > t\}}(\omega)\, d\mathbb{P}(\omega) \right) dt \qquad \text{par "Fubini } \geqslant 0\text{"} \\
&= \phi(0) + \int_0^\infty \mathbb{P}(Z > t)\, \phi'(t)\, dt.
\end{aligned}$$

Idem pour la 2è formule. □

Corollaire 4.2. *Si X est une va réelle, on a pour tout $1 \leqslant p < \infty$:*

$$\|X\|_p^p = \mathbb{E}\big(|X|^p\big) = p \int_0^\infty t^{p-1} \mathbb{P}\big(|X| > t\big)\, dt.$$

En particulier :

$$\|X\|_1 = \mathbb{E}\big(|X|\big) = \int_0^\infty \mathbb{P}\big(|X| > t\big)\, dt;$$

et

$$X \in L^p \iff \int_0^\infty t^{p-1}\mathbb{P}(|X| > t)\,dt < \infty.$$

Démonstration. On applique la proposition avec $Z = |X|$ et $\phi(t) = t^p$. □

COROLLAIRE 4.3. *Si X est une va à valeurs dans $\mathbb{N}$, alors*

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} \mathbb{P}(X \geqslant n).$$

En particulier, $X \in L^1 \iff \sum_{n=1}^{\infty} \mathbb{P}(X \geqslant n) < \infty$.

Démonstration. La va X est positive, donc $\mathbb{E}(X)$ a un sens. Par le Corollaire 4.2 avec $p = 1$ et comme $X \geqslant 0$ ps, on a

$$\begin{aligned}\mathbb{E}(X) &= \int_0^\infty \mathbb{P}(X > t)\,dt \\ &= \sum_{k=0}^{\infty} \int_{[k,k+1[} \mathbb{P}(X > t)\,dt \\ &= \sum_{k=0}^{\infty} \mathbb{P}(X \geqslant k+1).\end{aligned}$$

(La dernière ligne vient du fait que X est à valeurs dans $\mathbb{N}$, ce qui entraine qu'on a $\{X > t\} = \{X \geqslant k+1\}$ pour tout $k \in \mathbb{N}$ et pour tout $t \in [k, k+1[$.) □

Exercice. Démontrer le Corollaire 4.3 directement, *i.e.* sans utiliser la formule d'intégration par parties.

4.2. Markov et Tchebitchev.

PROPOSITION 4.4. *Pour toute va positive Z et pour tout $\alpha > 0$, on a l'inégalité suivante, qu'on appelle l'**inégalité de Markov** :*

$$\mathbb{P}(Z \geqslant \alpha) \leqslant \frac{1}{\alpha}\,\mathbb{E}(Z).$$

Démonstration. On a $\mathbb{P}(Z \geqslant \alpha) = \int_{\{Z \geqslant \alpha\}} d\mathbb{P}$. De plus, sur l'ensemble $\{Z \geqslant \alpha\}$, on a par définition $\frac{Z}{\alpha} \geqslant 1$. Comme $Z \geqslant 0$, on en déduit

$$\mathbb{P}(Z \geqslant \alpha) = \int_{\{Z \geqslant \alpha\}} 1\,d\mathbb{P} \leqslant \int_{\{Z \geqslant \alpha\}} \frac{Z}{\alpha}\,d\mathbb{P} \leqslant \int_\Omega \frac{Z}{\alpha}\,d\mathbb{P} = \frac{1}{\alpha}\mathbb{E}(Z).$$

□

COROLLAIRE 4.5. (inégalité de Tchebitchev)
Soit X une va réelle dans L^2, et soit $m = \mathbb{E}(X)$. Alors, pour tout $\varepsilon > 0$, on a

$$\mathbb{P}\big(|X - m| \geqslant \varepsilon\big) \leqslant \frac{\sigma^2(X)}{\varepsilon^2}.$$

Démonstration. Comme la fonction $u \mapsto u^2$ est strictement croissante sur $\mathbb{R}^+$, on peut écrire

$$\mathbb{P}\big(|X - m| \geqslant \varepsilon\big) = \mathbb{P}\big((X - m)^2 \geqslant \varepsilon^2\big);$$

et donc, par l'inégalité de Markov :

$$\mathbb{P}\big(|X-m|\geqslant\varepsilon\big)\leqslant\frac{\mathbb{E}\big((X-m)^2\big)}{\varepsilon^2}.$$

□

COROLLAIRE 4.6. *Pour toute va réelle X et pour tous $\varepsilon,\lambda>0$, on a*

$$\mathbb{P}\big(X\geqslant\varepsilon\big)\leqslant e^{-\lambda\varepsilon}\,\mathbb{E}\big(e^{\lambda X}\big).$$

Démonstration. Comme $\mathbb{P}(X\geqslant\varepsilon)=\mathbb{P}\big(e^{\lambda X}\geqslant e^{\lambda\varepsilon}\big)$ car $\lambda>0$, il suffit d'appliquer Markov à $Z=e^{\lambda X}$. □

EXEMPLE. ("Loi des grands nombres" pour un jeu de pile ou face infini)
Si on joue à pile ou face une infinité de fois, alors il est presque certain que la proportion de "piles" obtenus après n lancers tend vers $1/2$ quand $n\to\infty$.

Démonstration. On modélise la situation par une suite $(X_i)_{i\geqslant1}$ de va indépendantes suivant une loi de Bernoulli de paramètre $1/2$, en convenant que 1 correspond à "pile" et 0 à "face". L'évènement "on obtient pile au i-ème lancer" est donc $\{X_i=1\}$. La proportion de piles obtenus après n lancers est

$$Z_n=\frac{X_1+\cdots+X_n}{n};$$

et il s'agit de montrer que $Z_n\to1/2$ presque sûrement.

"Par Borel-Cantelli" (*cf* l'Exemple 1 après le Théorème 2.1 du Chapitre 5), il suffit de montrer que pour tout $\varepsilon>0$, on a

$$\sum_{n=1}^{\infty}\mathbb{P}\Big(\big|Z_n-\frac12\big|\geqslant\varepsilon\Big)<\infty.$$

De plus, comme $\big\{|Z_n-\frac12|\geqslant\varepsilon\big\}=\big\{Z_n\geqslant\frac12+\varepsilon\big\}\cup\big\{Z_n\leqslant-\frac12-\varepsilon\big\}$ et que la réunion est disjointe, on a

$$\mathbb{P}\Big(\big|Z_n-\frac12\big|\geqslant\varepsilon\Big)=\mathbb{P}\Big(Z_n\geqslant\frac12+\varepsilon\Big)+\mathbb{P}\Big(Z_n\leqslant-\frac12-\varepsilon\Big);$$

et par conséquent, il suffit de montrer que

$$\sum_{n=1}^{\infty}\mathbb{P}\Big(Z_n\geqslant\frac12+\varepsilon\Big)<\infty\qquad\text{et}\qquad\sum_{n=1}^{\infty}\mathbb{P}\Big(Z_n\leqslant-\frac12-\varepsilon\Big)<\infty.$$

On a pour tout $\lambda>0$:

$$\begin{aligned}\mathbb{P}\Big(Z_n\geqslant\frac12+\varepsilon\Big)&=\mathbb{P}\Big(X_1+\cdots+X_n\geqslant n(\frac12+\varepsilon)\Big)\\&\leqslant e^{-\lambda\,n(\frac12+\varepsilon)}\,\mathbb{E}\big(e^{\lambda(X_1+\cdots+X_n)}\big)\qquad\text{par Markov}\\&=e^{-\lambda\,n(\frac12+\varepsilon)}\prod_{i=1}^{n}\mathbb{E}\big(e^{\lambda X_i}\big)\qquad\text{par indépendance.}\end{aligned}$$

De plus, $\mathbb{E}\big(e^{\lambda X_i}\big) = \frac{1}{2}e^{\lambda\times 0} + \frac{1}{2}e^{\lambda\times 1} = \frac{1}{2}(1+e^{\lambda})$ pour $i = 1, \dots, n$. Donc on obtient

$$\begin{aligned}\mathbb{P}\Big(Z_n \geqslant \frac{1}{2} + \varepsilon\Big) &\leqslant e^{-\lambda\, n(\frac{1}{2}+\varepsilon)} \left(\frac{1+e^{\lambda}}{2}\right)^n \\ &= e^{-n\varepsilon\lambda} \left(\frac{e^{-\frac{\lambda}{2}} + e^{\frac{\lambda}{2}}}{2}\right)^n \\ &= e^{-n\varepsilon\lambda} \big(\mathrm{ch}(\lambda/2)\big)^n.\end{aligned}$$

Ensuite, en développant le cosinus hyperbolique en en série entière on voit qu'on a

$$0 \leqslant \mathrm{ch(t)} \leqslant e^{t^2/2} \qquad \text{pour tout } t \in \mathbb{R}.$$

En effet :

$$\mathrm{ch}(t) = \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!} \leqslant \sum_{k=0}^{\infty} \frac{t^{2k}}{2^k k!} = e^{\frac{t^2}{2}}.$$

On obtient donc

$$\mathbb{P}\Big(Z_n \geqslant \frac{1}{2} + \varepsilon\Big) \leqslant e^{-n\varepsilon\lambda}\, e^{n\frac{\lambda^2}{8}} = e^{-n\varepsilon\lambda + n\frac{\lambda^2}{8}} \quad \text{pour tout } \lambda > 0.$$

En étudiant $\varphi(\lambda) := -n\varepsilon\lambda + n\frac{\lambda^2}{8}$ (trinôme du second degré...), on voit que le "meilleur" λ, *i.e.* celui pour lequel $-n\varepsilon\lambda + n\frac{\lambda^2}{8}$ est minimal, est $\lambda = 4\varepsilon$; et que pour ce λ on a $-n\varepsilon\lambda + n\frac{\lambda^2}{8} = -2\varepsilon^2 n$. Ainsi,

$$\mathbb{P}\Big(Z_n \geqslant \frac{1}{2} + \varepsilon\Big) \leqslant e^{-2\varepsilon^2 n};$$

ce qui montre que la série $\sum \mathbb{P}\Big(Z_n \geqslant \frac{1}{2} + \varepsilon\Big)$ est convergente.

On montre de même que la série $\sum \mathbb{P}\Big(Z_n \leqslant -\frac{1}{2} - \varepsilon\Big)$ est convergente, ce qui termine la démonstration. □

4.3. Jensen.

Rappel. Si I est un intervalle de $\mathbb{R}$, une fonction $f : I \to \mathbb{R}$ est dite **convexe sur** I si

$$\forall x, y \in I\ \forall \lambda \in [0,1]\ :\ f\big((1-\lambda)x + \lambda y\big) \leqslant (1-\lambda)\, f(x) + \lambda f(y).$$

Par exemple, la fonction exponentielle est convexe sur $\mathbb{R}$, et la fonction $x \mapsto x^p$ est convexe sur $[0, \infty[$ si $p \geqslant 1$.

Proposition 4.7. (inégalité de Jensen)
Soit X une va réelle appartenant à L^1, à valeurs dans un intervalle $I \subseteq \mathbb{R}$, et soit $\phi : I \to \mathbb{R}$ une fonction convexe telle que $\phi(X) \geqslant 0$ ou $\phi(X) \in L^1$. Alors $\mathbb{E}(X) \in I$ et

$$\phi\big(\mathbb{E}(X)\big) \leqslant \mathbb{E}\big(\phi(X)\big).$$

Démonstration.

Fait 1. On a $\mathbb{E}(X) \in I$, et $\mathbb{E}(X) \in \mathring{I}$ si X n'est pas presque sûrement égale à une constante.

Preuve du Fait 1. Supposons par exemple que l'intervalle I soit de la forme $I = [a, b]$, avec $a < b < \infty$. Alors $a \leqslant X \leqslant b$ presque partout, donc $a = \mathbb{E}(a) \leqslant \mathbb{E}(X) \leqslant \mathbb{E}(b) = b$ par croissance de l'espérance, *i.e.* $\mathbb{E}(X) \in I$. De plus, si $\mathbb{E}(X) \notin \mathring{I}$, alors on a par exemple $\mathbb{E}(X) = a$, donc $\mathbb{E}(X - a) = 0$, et donc $X = a$ ps car $X - a \geqslant 0$. Ainsi, $\mathbb{E}(X) \in \mathring{I}$ si X n'est pas presque sûrement égale à une constante. La preuve est la même pour tout autre type d'intervalle. □

Fait 2. Pour tout $x_0 \in \mathring{I}$, on peut trouver une fonction *affine* $\alpha : \mathbb{R} \to \mathbb{R}$ telle que $\alpha(x_0) = \phi(x_0)$ et $\alpha(x) \leqslant \phi(x)$ pour tout $x \in I$.

Preuve du Fait 2. C'est une propriété "bien connue" des fonctions convexes. □

Revenons à l'inégalité de Jensen. Si X est presque sûrement égale à une constante c, alors le résultat est évident : on a $\mathbb{E}(\phi(X)) = \mathbb{E}(\phi(c)) = \phi(c) = \phi(\mathbb{E}(X))$. On peut donc supposer que X n'est pas presque sûrement égale à une constante. Alors $\mathbb{E}(X) \in \mathring{I}$ d'après le Fait 1. Par le Fait 2, on peut donc trouver une fonction affine $\alpha(x) = ax + b$ telle que $\alpha(\mathbb{E}(X)) = \phi(\mathbb{E}(X))$ et $\alpha(x) \leqslant \phi(x)$ pour tout $x \in I$. Ainsi,

$$\begin{aligned}
\phi(\mathbb{E}(X)) &= \alpha(\mathbb{E}(X)) \\
&= a\,\mathbb{E}(X) + b \\
&= \mathbb{E}(aX + b) \qquad \text{car } \mathbb{E}(b) = b \\
&= \mathbb{E}(\alpha(X)) \\
&\leqslant \mathbb{E}(\phi(X)) \qquad \text{par croissance de l'espérance.}
\end{aligned}$$

□

Remarque 1. L'inégalité de Jensen permet de retrouver le fait que $\|\cdot\|_{p_1} \leqslant \|\cdot\|_{p_2}$ si $p_1 \leqslant p_2$.

Démonstration. Supposons $1 \leqslant p_1 \leqslant p_2 < \infty$ (le cas $p_2 = \infty$ est laissé en exo). Soit X une va positive,. On a $\mathbb{E}\big(X^{p_2}\big) = \mathbb{E}\big((X^{p_1})^{p_2/p_1}\big) = \mathbb{E}\big(\phi(X^{p_1})\big)$, où $\phi(t) = t^{p_2/p_1}$. La fonction ϕ est convexe sur $\mathbb{R}^+$ car $p_2/p_1 \geqslant 1$; donc $\mathbb{E}\big(\phi(X^{p_1})\big) \geqslant \phi\big(\mathbb{E}(X^{p_1})\big)$, autrement dit $\mathbb{E}(X^{p_2}) \geqslant \mathbb{E}(X^{p_1})^{p_2/p_1}$, et donc $\|X\|_{p_2} = \mathbb{E}(X^{p_2})^{1/p_2} \geqslant \mathbb{E}(X^{p_1})^{1/p_1} = \|X\|_{p_1}$. □

Exercice. Montrer que si $X \in L^\infty$, alors $\|X\|_\infty = \lim_{p\to\infty} \|X\|_p$.

Remarque 2. L'inégalité de Jensen entraine ce qu'on appelle parfois l'**inégalité de Jensen discrète** : si $\phi : I \to \mathbb{R}$ est une fonction convexe alors, pour tous $x_1, \dots, x_n \in I$ et pour tous $\lambda_1, \dots, \lambda_n \geqslant 0$ vérifiant $\sum_1^n \lambda_i = 1$, on a

$$\phi\left(\sum_{i=1}^n \lambda_i x_i\right) \leqslant \sum_{i=1}^n \lambda_i\, \phi(x_i).$$

Démonstration. On peut supposer que les x_i sont tous différents (exo). Par hypothèse sur les λ_i, on définit une loi de probabilité μ sur $\{x_1, \dots, x_n\}$ en posant $\mu(\{x_i\}) = \lambda_i$ pour $i = 1, \dots, n$. Donc il existe une va réelle X à valeurs dans $\{x_1, \dots, x_n\}$ telle que

$$\mathbb{P}(X = x_i) = \lambda_i \qquad \text{pour } i = 1, \dots, n.$$

On a alors

$$\sum_{i=1}^n \lambda_i x_i = \mathbb{E}(X) \qquad \text{et} \qquad \sum_{i=1}^n \lambda_i \phi(x_i) = \mathbb{E}(\phi(X));$$

d'où le résultat par l'inégalité de Jensen. □

5. Moments d'ordres supérieurs

DÉFINITION 5.1. *Soit X une va réelle, et soit $k \in \mathbb{N}^*$. On dit que X* ***admet un moment d'ordre*** *k si $X \in L^k$, i.e. $\mathbb{E}(|X|^k) < \infty$. Si c'est le cas ; le nombre $\mathbb{E}(X^k)$ s'appelle le* ***k-ième moment*** *de X.*

Remarque. Par convention, toute va réelle possède un moment d'ordre 0, et $\mathbb{E}(X^0) = 1$ (puisque $X^0 = \mathbf{1}$).

EXEMPLE 1. Si $X \in L^\infty$, alors X admet des moments de tous ordres.

EXEMPLE 2. S'il existe une constante $\lambda > 0$ tel que $e^{\lambda|X|} \in L^1$, alors X admet des moments de tous ordres.

Démonstration. On veut montrer que $X \in L^p$ pour tout $p < \infty$. D'après l'inégalité de Markov, on a pour tout $t > 0$:

$$\mathbb{P}(|X| > t) = \mathbb{P}\big(e^{\lambda|X|} > e^{\lambda t}\big) \leqslant C\, e^{-\lambda t}, \quad \text{où } C = \mathbb{E}(e^{\lambda|X|}).$$

Donc

$$\int_0^\infty t^{p-1}\mathbb{P}(|X| > t)\, dt \leqslant C \int_0^\infty t^{p-1}e^{-\lambda t} dt < \infty,$$

et donc $X \in L^p$ par le Corollaire 4.2. □

EXEMPLE 3. Soit X une va suivant une loi normale $\mathcal{N}(m, \sigma^2)$. Alors X n'appartient pas à L^∞, mais vérifie cependant l'hypothèse de l'Exemple 2 pour tout $\lambda > 0$.

Démonstration. exo. □

PROPOSITION 5.2. ("Théorème des moments")
Les moments d'une va réelle bornée *déterminent sa loi. Autrement dit : si X et Y sont deux va réelles appartenant à L^∞ telles que $\mathbb{E}(X^k) = \mathbb{E}(Y^k)$ pour tout $k \in \mathbb{N}$, alors $\mathbb{P}_X = \mathbb{P}_Y$.*

Démonstration. Fixons X et Y comme dans l'énoncé.

FAIT 1. On a $\mathbb{E}(f(X)) = \mathbb{E}(f(Y))$ pour toute fonction continue $f : \mathbb{R} \to \mathbb{R}$.

Preuve du Fait 1. Comme X et Y sont dans L^∞, on peut trouver un intervalle compact $[a, b]$ tel que X et Y sont (presque sûrement) à valeurs dans $[a, b]$. Alors, si $f : \mathbb{R} \to \mathbb{R}$ est continue, elle est bornée sur $[a, b]$, donc les va $f(X)$ et $f(Y)$ sont dans L^∞, et en particulier dans L^1. Donc : l'énoncé a un sens !

Par hypothèse et par linéarité de l'espérance, on a $\mathbb{E}(P(X)) = \mathbb{E}(P(Y))$ pour tout *polynôme* P. Maintenant, si $f : \mathbb{R} \to \mathbb{R}$ est continue, on peut trouver une suite de polynômes (P_n) telle que $P_n \to f$ uniformément sur $[a, b]$, où $[a, b]$ est comme plus haut. Alors $P_n(X) \to f(X)$ pour la norme $\|\cdot\|_\infty$; en particulier $\|P_n(X) - f(X)\|_1 \to 0$ puisque $\|\cdot\|_1 \leqslant \|\cdot\|_\infty$, et donc $\mathbb{E}(P_n(X)) \to \mathbb{E}(f(X))$. De même $\mathbb{E}(P_n(Y)) \to \mathbb{E}(f(Y))$. D'où le résultat puisque $\mathbb{E}(P_n(X)) = \mathbb{E}(P_n(Y))$ pour tout n. □

FAIT 2. Soient μ_1 et μ_2 deux mesures boréliennes finies sur $\mathbb{R}$. Si on a $\int_{\mathbb{R}} f\, d\mu_1 = \int_{\mathbb{R}} f\, d\mu_2$ pour toute fonction *continue bornée* $f : \mathbb{R} \to \mathbb{R}$, alors $\mu_1 = \mu_2$.

Preuve du Fait 2. D'après le Corollaire 1.5 du Chapitre 3, il suffit de montrer qu'on a $\mu_1(]-\infty, b]) = \mu_2(]-\infty, b])$ pour tout $b \in \mathbb{R}$.

Pour $n \in \mathbb{N}^*$, soit $f_n : \mathbb{R} \to \mathbb{R}$ la fonction définie comme suit : $f_n(t) \equiv 1$ sur $]-\infty, b]$, $f_n(t) \equiv 0$ sur $[b + \frac{1}{n}, \infty[$, et f_n est affine sur $[b, b + \frac{1}{n}]$. (*Dessiner le graphe*

de f_n.) Alors les f_n sont continues, $f_n \to \mathbf{1}_{]-\infty,b]}$ simplement (exo), et $0 \leqslant f_n \leqslant 1$. Comme les mesures μ_1 et μ_2 sont finies, on en déduit par convergence dominée qu'on a

$$\mu_1(]-\infty,b]) = \lim_{n\to\infty} \int_{\mathbb{R}} f_n \, d\mu_1 \qquad \text{et} \qquad \mu_2(]-\infty,b]) = \lim_{n\to\infty} \int_{\mathbb{R}} f_n \, d\mu_2;$$

donc $\mu_1(]-\infty,b]) = \mu_2(]-\infty,b])$ puisque $\int_{\mathbb{R}} f_n \, d\mu_1 = \int_{\mathbb{R}} f_n \, d\mu_2$ pour tout n (les f_n étant continues bornées). □

La preuve du Théorème des moments est maintenant terminée : par le Fait 1, on a $\int_{\mathbb{R}} f \, d\mathbb{P}_X = \int_{\mathbb{R}} f \, d\mathbb{P}_Y$ pour toute fonction continue bornée $f : \mathbb{R} \to \mathbb{R}$; donc $\mathbb{P}_X = \mathbb{P}_Y$ par le Fait 2. □

Chapitre 7

Convergence des variables aléatoires

1. Convergence presque sûre et convergence L^p

Définition 1.1. *Soit (Z_n) une suite de va réelles définies sur le même $(\Omega, \mathfrak{A}, \mathbb{P})$. On dit que (Z_n) **converge presque sûrement** s'il existe une va Z telle que $Z_n(\omega) \to Z(\omega)$ pour presque tout $\omega \in \Omega$. On écrit alors $Z_n \xrightarrow{\text{ps}} Z$.*

Exemple 1.2. Si on a $\sum_{n=0}^{\infty} \mathbb{P}\big(|Z_n - Z| \geqslant \varepsilon\big) < \infty$ pour tout $\varepsilon > 0$, alors $Z_n \to Z$ presque sûrement.

Démonstration. C'est une conséquence (de la partie triviale) du Lemme Borel-Cantelli ; *cf* le Chapitre 5. □

Définition 1.3. *Soit (Z_n) une suite de va réelles définies sur le même $(\Omega, \mathfrak{A}, \mathbb{P})$ et appartenant toutes à L^p pour un certain $p \geqslant 1$. On dit que (Z_n) **converge en norme** L^p s'il existe une va $Z \in L^p$ telle que $\|Z_n - Z\|_p \to 0$.*

Exemple 1.4. Soit (X_k) une suite de va deux à deux indépendantes, appartenant à L^2 et centrées. Si on a $\sum_{k=0}^{\infty} \sigma^2(X_k) < \infty$, alors la série $\sum X_k$ converge en norme L^2.

Démonstration. Comme L^2 est un espace complet, suffit de montrer que les sommes partielles $Z_n = \sum_{k=0}^{n} X_k$ forment une *suite de Cauchy* dans L^2.

Comme les X_k sont deux à deux indépendantes et centrées, elles sont *orthogonales dans* L^2. D'après le théorème de Pythagore, on en déduit que pour tous $p < q$, on a

$$\|S_q - S_p\|_2^2 = \Big\| \sum_{k=p+1}^{q} X_k \Big\|_2^2 = \sum_{k=p+1}^{q} \|X_k\|_2^2 = \sum_{k=p+1}^{q} \sigma^2(X_k).$$

Donc $\|S_q - S_p\|_2 \to 0$ quand $p, q \to \infty$ puisque la séries $\sum \sigma^2(X_k)$ est convergente. □

Remarque 1. Si une suite de va réelles (Z_n) converge en norme L^{p_0} pour un certain $p_0 \geqslant 1$, alors elle converge en norme L^p pour tout $1 \leqslant p \leqslant p_0$, donc en particulier en norme L^1.

Démonstration. C'est clair puisque $L^{p_0} \subseteq L^p$ et $\| \cdot \|_p \leqslant \| \cdot \|_{p_0}$. □

Remarque 2. Si $Z_n \to Z$ en norme L^1, alors $\mathbb{E}(Z_n) \to \mathbb{E}(Z)$.

Démonstration. C'est évident puisque

$$|\mathbb{E}(Z_n) - \mathbb{E}(Z)| = |\mathbb{E}(Z_n - Z)| \leqslant \mathbb{E}(|Z_n - Z|) = \|Z_n - Z\|_1.$$

□

Remarque 3. La convergence presque sûre et la convergence en norme L^p ne sont pas comparables (si $p < \infty$). Autrement dit : aucune n'entraine l'autre. (On le verra en TD.)

PROPOSITION 1.5. *Soit (Z_n) une suite de va réelles. Si (Z_n) converge presque sûrement vers une va Z et s'il existe une va $G \geqslant 0$ telle que $G \in L^p$ et $\forall n : |Z_n| \leqslant G$ presque sûrement, alors (Z_n) tend vers Z en norme L^p.*

Démonstration. Vu en principe dans le cours de théorie de l'intégration ("convergence dominée dans L^p"). La preuve consiste simplement à appliquer le Théorème de convergence dominée usuel à $f_n := |Z_n - Z|^p$ (exo). □

PROPOSITION 1.6. *Soit (Z_n) une suite de va réelles. Si (Z_n) converge en norme L^1 vers une va Z, alors (Z_n) possède une sous-suite qui tend presque sûrement vers Z.*

Démonstration. Vu aussi, en principe, dans le cours de théorie de l'intégration. On redémontrera ce résultat un peu plus loin, sous une forme plus générale (Corollaire 2.4). □

2. Convergence en probabilité

DÉFINITION 2.1. *Soit (Z_n) une suite de va réelles, toutes définies sur le même $(\Omega, \mathfrak{A}, \mathbb{P})$. On dit que la suite (Z_n)* ***converge en probabilité*** *s'il existe une va Z telle que*

$$\mathbb{P}\big(|Z_n - Z| \geqslant \varepsilon\big) \xrightarrow{n\to\infty} 0 \qquad \text{pour tout } \varepsilon > 0.$$

On écrit alors $Z_n \xrightarrow{\text{proba}} Z$.

REMARQUE 1. On a "unicité de la limite" : si $Z_n \xrightarrow{\text{proba}} Z$ et $Z_n \xrightarrow{\text{proba}} Z'$, alors $Z' = Z$ presque sûrement.

Démonstration. Comme $|Z' - Z| \leqslant |Z' - Z_n| + |Z_n - Z|$, on a

$$\big\{|Z' - Z| \geqslant \varepsilon\big\} \subseteq \big\{|Z' - Z_n| \geqslant \varepsilon/2\big\} \cup \big\{|Z_n - Z| \geqslant \varepsilon/2\big\}$$

pour tout $\varepsilon > 0$ et pour tout n, et donc

$$\mathbb{P}\big(|Z' - Z| \geqslant \varepsilon\big) \leqslant \mathbb{P}\big(|Z' - Z_n| \geqslant \varepsilon/2\big) + \mathbb{P}\big(|Z' - Z_n| \geqslant \varepsilon/2\big).$$

En faisant tendre n vers l'infini, on en déduit que

$$\mathbb{P}\big(|Z' - Z| \geqslant \varepsilon\big) = 0 \qquad \text{pour tout } \varepsilon > 0.$$

Donc $\mathbb{P}(Z \neq Z') = \mathbb{P}\Big(\bigcup_{k\in\mathbb{N}^*} \big\{|Z' - Z| \geqslant 1/k\big\}\Big) = 0$, autrement dit $Z' = Z$ ps. □

REMARQUE 2. Si (Z_n) converge en probabilité, alors elle est *de Cauchy en probabilité* : pour tout $\varepsilon > 0$,

$$\mathbb{P}\big(|Z_q - Z_p| \geqslant \varepsilon\big) \to 0 \qquad \text{quand } p, q \to \infty.$$

Démonstration. En notant Z la va limite, on a comme plus haut

$$\mathbb{P}\big(|Z_q - Z_p| \geqslant \varepsilon\big) \leqslant \mathbb{P}\big(|Z_q - Z| \geqslant \varepsilon/2\big) + \mathbb{P}\big(|Z - Z_p| \geqslant \varepsilon/2\big),$$

ce qui donne immédiatement le résultat. □

REMARQUE 3. La réciproque est vraie : si une suite de va (Z_n) est de Cauchy en probabilité, alors elle converge en probabilité vers une certaine va Z.

Démonstration. En TD. □

PROPOSITION 2.2. *La convergence presque sûre et la convergence en norme L^1 entrainent chacune la convergence en probabilité.*

Démonstration. Supposons que $Z_n \xrightarrow{\text{ps}} Z$, et fixons $\varepsilon > 0$. Pour tout $n \in \mathbb{N}$, on a

$$\mathbb{P}\big(|Z_n - Z| \geqslant \varepsilon\big) \leqslant \mathbb{P}\Big(\underbrace{\bigcup_{k \geqslant n} \{|Z_k - Z| \geqslant \varepsilon\}}_{A_n}\Big).$$

Donc il suffit de montrer que $\mathbb{P}(A_n) \to 0$ quand $n \to \infty$.

Comme la suite (A_n) est décroissante, on a

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\Big(\bigcap_{n\in\mathbb{N}} A_n\Big).$$

Mais si $\omega \in \bigcap_n A_n$, alors $|Z_k(\omega) - Z(\omega)| \geqslant \varepsilon$ pour une infinité de k, et donc $Z_k(\omega)$ ne tend pas vers $Z(\omega)$. Donc $\mathbb{P}(\bigcap_n A_n) = 0$ puisque $Z_k \to Z$ ps.

Supposons maintenant que $Z_n \xrightarrow{L^1} Z$. D'après l'inégalité de Markov, on a

$$\mathbb{P}\big(|Z_n - Z| \geqslant \varepsilon\big) \leqslant \frac{\mathbb{E}(|Z_n - Z|)}{\varepsilon} = \frac{\|Z_n - Z\|_1}{\varepsilon}$$

pour tout $\varepsilon > 0$ et pour tout n. Donc $\mathbb{P}\big(|Z_n - Z| \geqslant \varepsilon\big) \to 0$ pour tout $\varepsilon > 0$. □

Remarque. Les réciproques sont fausses. (On le verra en TD.)

Proposition 2.3. *Soit (Z_n) une suite de va réelles, et soit Z une va réelle. Alors $Z_n \xrightarrow{\text{proba}} Z$ si et seulement si toute sous-suite (Z'_n) de (Z_n) possède une sous-suite (Z'_{n_k}) telle que $Z'_{n_k} \xrightarrow{\text{ps}} Z$.*

Démonstration. Supposons que $Z_n \xrightarrow{\text{proba}} Z$, et soit (Z'_n) une sous-suite quelconque de (Z_n). Alors $Z'_n \xrightarrow{\text{proba}} Z$; donc $\mathbb{P}\big(|Z'_n - Z| \geqslant 1/k\big) \to 0$ quand $n \to \infty$ pour tout $k \in \mathbb{N}^*$. On peut donc trouver une suite strictement croissante d'entiers $(n_k)_{k\geqslant 1}$ telle que

$$\forall k\, \forall n \geqslant n_k \ : \ \mathbb{P}\big(|Z'_n - Z| \geqslant 1/k\big) \leqslant 2^{-k}.$$

Pour $\varepsilon > 0$ donné, on a alors

$$\forall k \geqslant 1/\varepsilon \ : \ \mathbb{P}\big(|Z'_{n_k} - Z| \geqslant \varepsilon\big) \leqslant \mathbb{P}\big(|Z'_{n_k} - Z| \geqslant 1/k\big) \leqslant 2^{-k}.$$

Donc

$$\sum_{k=1}^{\infty} \mathbb{P}\big(|Z'_{n_k} - Z| \geqslant \varepsilon\big) < \infty \qquad \text{pour tout } \varepsilon > 0,$$

et donc $Z'_{n_k} \xrightarrow{\text{ps}} Z$.

Inversement, supposons que (Z_n) ne converge pas en probabilité vers Z : il s'agit alors de trouver une sous-suite (Z'_n) de (Z_n) telle qu'aucune sous-suite de (Z'_n) ne converge presque sûrement vers Z.

Par hypothèse, on peut trouver $\varepsilon_0 > 0$ tel que $\mathbb{P}(|Z_n - Z| \geqslant \varepsilon_0)$ ne tend pas vers 0 quand $n \to \infty$. Cela signifie qu'il existe $\eta > 0$ tel que $\mathbb{P}(|Z_n - Z| \geqslant \varepsilon_0) \geqslant \eta$ pour une infinité d'entiers n, autrement dit qu'il existe $\eta > 0$ et une sous-suite (Z'_n) de (Z_n) telle que

$$\forall n \in \mathbb{N} \ : \ \mathbb{P}(|Z'_n - Z| \geqslant \varepsilon_0) \geqslant \eta.$$

Alors aucune sous-suite de (Z'_n) ne peut converger en probabilité vers Z, donc *a fortiori* aucune sous-suite de (Z'_n) ne tend presque sûrement vers Z. □

COROLLAIRE 2.4. *Si (Z_n) est une suite de va réelles qui converge en probabilité (par exemple, si (Z_n) converge en norme L^1), alors (Z_n) possède une sous-suite qui converge presque sûrement.*

Démonstration. C'est évident par la propostion. □

EXERCICE. Montrer que si $Z_n \xrightarrow{\text{proba}} Z$, alors $f(Z_n) \xrightarrow{\text{proba}} f(Z)$ pour toute fonction *continue* $f : \mathbb{R} \to \mathbb{R}$.

3. Séries de va indépendantes

Vu ce qu'on a raconté dans les sections précédentes, le résultat suivant est *a priori* assez surprenant. On le donne à titre culturel (c'est un des très nombreux "théorèmes de Paul Lévy").

THÉORÈME 3.1. *Soit (X_k) une suite de va réelles indépendantes, et pour $n \in \mathbb{N}$, soit $S_n := \sum_{k=0}^{n} X_k$. Si la suite (S_n) converge en probabilité, alors elle converge presque sûrement. Autrement dit : pour une série de va indépendantes, la convergence en probabilité est équivalente à la convergence presque sûre.*

Démonstration. *Cf* TD, ou pas.

□

COROLLAIRE 3.2. *Soit (X_k) une suite de va réelles indépendantes, appartenant à L^2 et centrées. Si on a $\sum_{k=0}^{\infty} \sigma^2(X_k) < \infty$, alors la série $\sum X_k$ converge presque sûrement.*

Démonstration. L'hypothèse sur la série $\sum \sigma^2(X_k)$ et l'orthogonalité des va X_k entrainent la convergence en norme L^2 de la série $\sum X_k$ (Exemple 1.4). Comme la convergence L^2 entraine la convergence en probabilité, on en déduit que la série $\sum X_k$ converge en probabilité, *et donc* presque sûrement d'après le théorème puisque les X_k sont indépendantes. □

COROLLAIRE 3.3. *Soit (a_k) une suite de nombres réels. Si la série $\sum a_k^2$ est convergente, alors on peut trouver une suite de choix de signes $\pm$ telle que la série $\sum \pm a_k$ soit convergente.*

Démonstration. Soit (ε_k) une suite de va indépendantes (définies sur un certain $(\Omega, \mathfrak{A}, \mathbb{P})$) suivant toutes une **loi de Rademacher**, *i.e* les ε_k sont à valeurs dans $\{-1, 1\}$ avec

$$\mathbb{P}(\varepsilon_k = 1) = \frac{1}{2} = \mathbb{P}(\varepsilon_k = -1).$$

On applique le corollaire précédent avec $X_k := a_k \varepsilon_k$. Les X_k sont indépendantes, et on vérifie très facilement qu'elles sont dans L^2, centrées, avec $\sigma^2(X_k) = a_k^2$. Par hypothèse sur les a_k, on en déduit que la série $\sum X_k = \sum a_k \varepsilon_k$ converge presque sûrement. En particulier, il existe au moins un $\omega \in \Omega$ tel que la série $\sum \varepsilon_k(\omega) a_k$ soit convergente, ce qui démontre le résultat souhaité. □

4. Convergence en loi

4.1. Convergence étroite des mesures.

4.1.1. *Généralités.* Dans ce qui suit, on se donne un espace métrique (Λ, d).

NOTATIONS. On note $\mathrm{M}_+(\Lambda)$ l'ensemble des mesures boréliennes *finies* sur Λ, et $\mathcal{C}_b(\Lambda)$ l'ensemble des fonctions *continues bornées* $f : \Lambda \to \mathbb{R}$.

LEMME 4.1. *$\mathcal{C}_b(\Lambda)$ sépare les points de $\mathrm{M}_+(\Lambda)$: si $\mu, \mu' \in M_+(\Lambda)$ et $\mu \neq \mu'$, alors il existe $f \in \mathcal{C}_b(\Lambda)$ telle que $\int_\Lambda f\,d\mu \neq \int_\Lambda f\,d\mu'$. Autrement dit, si μ et μ' sont deux mesures telles que $\int_\Lambda f\,d\mu = \int_\Lambda f\,d\mu'$ pour toute fonction $f \in \mathcal{C}_b(\Lambda)$, alors $\mu = \mu'$.*

Démonstration. Pour $\Lambda = \mathbb{R}$, on a déjà démontré ce résultat quand on a fait la preuve du théorème des moments (Proposition 5.2 du Chapitre 6). Dans le cas général, la démonstration est très similaire.

Par le théorème des classes monotones, il suffit de montrer qu'on a $\mu(F) = \mu'(F)$ pour tout fermé $F \subseteq \Lambda$ (exo).

FAIT. Il existe une suite décroissante (f_k) de fonctions lipschitziennes bornées (donc appartenant à $C_b(\Lambda)$) telle que $f_k(x) \to \mathbf{1}_F(x)$ pour tout $x \in \Lambda$.

Preuve du Fait. Pour $k \in \mathbb{N}^*$, soit $\theta_k : \mathbb{R} \to \mathbb{R}$ la fonction définie comme suit $= \theta_k(0) = 1$, $\theta_k(t) \equiv 0$ sur $]-\infty, -1/k] \cup [1/k, \infty[$, et θ_k est affine sur $[-1/k, 0]$ et sur $[0, 1/k]$. (*Faire un dessin.*) Alors les θ_k sont lipschitziennes bornées, la suite (θ_k) est décroissante, et $\theta_k(t) \to \mathbf{1}_{\{0\}}(t)$ pour tout $t \in \mathbb{R}$ (exo). Si maintenant on définit $f_k : \Lambda \to \mathbb{R}$ par $f_k(x) = \theta_k(\mathrm{dist}(x, F))$, alors les f_k sont lipschitziennes bornées par composition, la suite (f_k) est décroissante, et $f_k \to \mathbf{1}_{\{x \in \Lambda\,:\,\mathrm{dist}(x,F)=0\}}$; autrement dit $f_k \to \mathbf{1}_F$ puisque F est un fermé de Λ. □

Supposons que $\int_\Lambda f\,d\mu = \int_\Lambda f\,d\mu'$ pour toute $f \in \mathcal{C}_b(\Lambda)$. Soit (f_k) comme dans le Fait. Comme μ et μ' sont des mesures finies et $0 \leqslant f_k \leqslant f_1$ pour tout $k \geqslant 1$, on a $\mu(F) = \lim_{k\to\infty} \int_\Lambda f_k\,d\mu$ par convergence dominée ; et de même $\mu'(F) = \lim_{k\to\infty} \int_\Lambda f_k\,d\mu'$. Donc $\mu(F) = \mu'(F)$ puisque $\int_\Lambda f_k\,d\mu = \int_\Lambda f_k\,d\mu$ pour tout k. □

DÉFINITION 4.2. *Soit (μ_n) une suite dans $\mathrm{M}_+(\Lambda)$, et soit $\mu \in \mathrm{M}_+(\Lambda)$. On dit que la suite (μ_n) **converge étroitement vers** μ, et on écrit $\mu_n \twoheadrightarrow \mu$, si*

$$\int_\Lambda f\,d\mu_n \xrightarrow{n\to\infty} \int_\Lambda f\,d\mu \qquad \text{pour toute } f \in \mathcal{C}_b(\Lambda).$$

REMARQUE. On a *unicité de la limite* : si $\mu_n \twoheadrightarrow \mu$ et $\mu_n \twoheadrightarrow \mu'$, alors $\mu = \mu'$.

Démonstration. Par définition, on a $\int_\Lambda f\,d\mu = \lim_{n\to\infty} \int_\Lambda f\,d\mu_n = \int_\Lambda f\,d\mu'$ pour toute $f \in \mathcal{C}_b(\Lambda)$, donc $\mu = \mu'$ par le Lemme 4.1. □

EXERCICE. Montrer que si $\mu_n \twoheadrightarrow \mu$, alors $\mu_n(\Lambda) \to \mu(\Lambda)$.

EXEMPLE 1. Si (a_n) est une suite de points de Λ convergeant vers un point $a \in \Lambda$, alors $\delta_{a_n} \twoheadrightarrow \delta_a$.

Démonstration. Exo. □

EXEMPLE 2. Soit ρ une densité lebesguienne sur $\mathbb{R}^d$, et soit $(C_n) \subseteq]0, \infty[$ une suite tendant vers ∞. Alors $\mu_n = C_n\rho(C_n x)dx$ converge étroitement vers $\mu = \delta_0$.

Démonstration. Pour toute $f \in \mathcal{C}_b(\mathbb{R})$, on a

$$\begin{aligned}\int_{\mathbb{R}} f\, d\mu_n &= \int_{\mathbb{R}} f(x)\, C_n \rho(C_n x)\, dx \\ &= \int_{\mathbb{R}} f\left(\frac{u}{C_n}\right) \rho(u)\, du.\end{aligned}$$

Comme f est continue en 0 et bornée, on en déduit par convergence dominée que

$$\int_{\mathbb{R}} f\, d\mu_n \xrightarrow{n\to\infty} \int_{\mathbb{R}} f(0)\, \rho(u)\, du = f(0) = \int_{\mathbb{T}} f\, d\delta_0$$

□

PROPOSITION 4.3. ("Théorème porte-manteau")
Soit (μ_n) une suite dans $\mathrm{M}_+(\Lambda)$, et soit $\mu \in \mathrm{M}_+(\Lambda)$. Les propriétés suivantes sont équivalentes :

(1) $\mu_n \twoheadrightarrow \mu$;
(2) $\int_\Lambda f\, d\mu_n \to \int_\Lambda f\, d\mu$ *pour toute fonction* lipschitzienne bornée $f : \Lambda \to \mathbb{R}$;
(3) $\mu_n(\Lambda) \to \mu(\Lambda)$ *et* $\overline{\lim}\, \mu_n(F) \leqslant \mu(F)$ *pour tout* fermé $F \subseteq \Lambda$;
(3') $\mu_n(\Lambda) \to \mu(\Lambda)$ *et* $\underline{\lim}\, \mu_n(O) \geqslant \mu(O)$ *pour tout* ouvert $O \subseteq \Lambda$;
(4) $\mu_n(B) \to \mu(B)$ *pour tout borélien* $B \subseteq \Lambda$ *tel que* $\mu(\partial B) = 0$.

Démonstration. (1) $\implies$ (2) est évident.

(2) $\implies$ (3). Supposons (2) vérifiée, et montrons (3).

Soit F un fermé de Λ. On a vu dans la preuve du Lemme 4.1 qu'il existe une suite décroissante (f_k) de fonctions lipschitziennes bornées, $f_k : \Lambda \to \mathbb{R}$, telle que $f_k(x) \to \mathbf{1}_F(x)$ pour tout $x \in \Lambda$. Comme $\mathbf{1}_F \leqslant f_k$, on a pour tout $n \in \mathbb{N}$ et pour tout k :

$$\mu_n(F) = \int_\Lambda \mathbf{1}_F\, d\mu_n \leqslant \int_\Lambda f_k\, d\mu_n.$$

De plus, $\int_\Lambda f_k\, d\mu_n \xrightarrow{n\to\infty} \int_\Lambda f_k\, d\mu$ puisque $\mu_n \twoheadrightarrow \mu$, donc

$$\overline{\lim}\, \mu_n(F) \leqslant \int_\Lambda f_k\, d\mu \qquad \text{pour tout } k \geqslant 1.$$

Mais $\int_\Lambda f_k\, d\mu \to \int_\Lambda \mathbf{1}_F\, d\mu = \mu(F)$ quand $k \to \infty$, par convergence dominée ; donc on obtient bien $\overline{\lim}\, \mu_n(F) \leqslant \mu(F)$.

(3) $\iff$ (3′) : c'est un exo utilisant juste le fait que $\mu_n(O) = \mu_n(\Lambda) - \mu_n(\Lambda \backslash O)$ et $\mu(O) = \mu(\Lambda) - \mu(\Lambda \backslash O)$.

(3), (3′) $\implies$ (4). Si $B \subseteq \Lambda$ est un borélien tel que $\mu(\partial B) = 0$, alors $\mu(\overline{B}) = \mu(\mathring{B}) = \mu(B)$ car $\mu(\partial B) = \mu(\overline{B} \backslash \mathring{B}) = \mu(\overline{B}) - \mu(\mathring{B})$, donc en fait $\mu(\overline{B}) = \mu(\mathring{B}) = \mu(B)$ puisque $\mathring{B} \subseteq B \subseteq \overline{B}$. En appliquant (3) et (3') avec $F = \overline{B}$ et $O = \mathring{B}$, on obtient

$$\overline{\lim}\, \mu_n(B) \leqslant \overline{\lim}\, \mu_n(\overline{B}) \leqslant \mu(\overline{B}) = \mu(B) = \mu(\mathring{B}) \leqslant \underline{\lim}\, \mu_n(\mathring{B}) \leqslant \underline{\lim}\, \mu_n(B);$$

ce qui montre que $\mu_n(B) \to \mu(B)$.

(4) $\implies$ (1). C'est plus long, et on l'*admettra*.

□

COROLLAIRE 4.4. *Si $\mu_n(B) \to \mu(B)$ pour tout borélien $B \subseteq \Lambda$, alors $\mu_n \twoheadrightarrow \mu$.*

Démonstration. C'est une conséquence immédiate de la proposition, puisque l'hypothèse faite est plus forte que (4). Cependant, comme on n'a pas montré que (4) entraine la convergence en loi, on va donner une preuve directe de ce corollaire.

On remarque d'abord que $\int_\Lambda \varphi\, d\mu_n \to \int_\Lambda \varphi\, d\mu$ pour toute fonction borélienne *étagée* $\varphi : \Lambda \to \mathbb{R}$. En effet, c'est vrai pour φ de la forme $\mathbf{1}_B$ par hypothèse, donc vrai également pour φ étagée quelconque par linéarité (micro-exo).

Soit maintenant $f \in \mathcal{C}_b(\Lambda)$ quelconque, et soit $\varepsilon > 0$. Comme f est borélienne bornée, on sait qu'on peut trouver une suite de fonction étagées qui converge uniformément vers f. Donc, on peut fixer une fonction φ étagée telle que

$$\forall x \in \Lambda \ : \ |\varphi(x) - f(x)| \leqslant \varepsilon.$$

On a alors

$$\left|\int_\Lambda \varphi\, d\mu - \int_\Lambda f\, d\mu\right| \leqslant \int_\Lambda |\varphi - f|\, d\mu \leqslant \varepsilon\, \mu(\Lambda);$$

et de même

$$\left|\int_\Lambda \varphi\, d\mu_n - \int_\Lambda f\, d\mu_n\right| \leqslant \varepsilon\, \mu_n(\Lambda) \qquad \text{pour tout } n \in \mathbb{N}$$

D'après l'inégalité triangulaire, on voit donc (exo) que

$$\forall n \in \mathbb{N} \ : \ \left|\int_\Lambda f\, d\mu_n - \int_\Lambda f\, d\mu\right| \leqslant \varepsilon\, \mu_n(\Lambda) + \left|\int_\Lambda \varphi\, d\mu_n - \int_\Lambda \varphi\, d\mu_n\right| + \varepsilon\, \mu(\Lambda).$$

Comme $\int_\Lambda \varphi\, d\mu_n \to \int_\Lambda \varphi\, d\mu$ et $\mu_n(\Lambda) \to \mu(\Lambda)$, on en déduit

$$\overline{\lim} \left|\int_\Lambda f\, d\mu_n - \int_\Lambda f\, d\mu\right| \leqslant 2\varepsilon\, \mu(\Lambda);$$

et ceci étant vrai pour tout $\varepsilon > 0$, on en conclut que $\int_\Lambda f\, d\mu_n \to \int_\Lambda f\, d\mu$. □

Exercice. Trouver une suite $(\mu_n) \subseteq \mathrm{M}_+(\mathbb{R})$ convergeant étroitement vers δ_0 mais telle que $\mu_n(\{0\})$ ne tende pas vers $1 = \delta_0(\{0\})$.

4.1.2. *Cas où Λ est dénombrable.*

Proposition 4.5. *Supposons que l'espace métrique Λ soit* dénombrable. *Soit (μ_n) une suite dans $\mathrm{M}_+(\Lambda)$, et soit $\mu \in \mathrm{M}_+(\Lambda)$. Si $\mu_n(\Lambda) \to \mu(\Lambda)$ et $\mu_n(\{x\}) \to \mu(\{x\})$ pour tout $x \in \Lambda$, alors $\mu_n(B) \to \mu(B)$ pour tout ensemble $B \subseteq \Lambda$, et donc $\mu_n \twoheadrightarrow \mu$.*

Démonstration. On remarque d'abord que $\mu_n(F) \to \mu(F)$ pour tout ensemble *fini* $F \subseteq \Lambda$: en effet, si $F = \{a_1, \dots, a_K\}$, alors $\mu_n(F) = \sum_{k=1}^K \mu_n(\{a_k\}) \to \sum_{k=1}^K \mu(\{a_k\}) = \mu(F)$ quand $n \to \infty$ par (5).

Fait. Pour tout $\varepsilon > 0$, on peut trouver un ensemble fini $F \subseteq \Lambda$ et un entier N tels que $\mu(\Lambda \backslash F) < \varepsilon$ et $\mu_n(\Lambda \backslash F) < \varepsilon$ pour tout $n \geqslant N$.

Preuve du Fait. Comme Λ est dénombrable, il est réunion d'une suite croissante d'ensemble finis $(F_k)_{k\in\mathbb{N}}$. Alors $\mu(\Lambda) = \lim \mu(F_k)$; donc, comme μ est une mesure finie, on peut trouver un entier k tel que $F := F_k$ vérifie $\mu(F) > \mu(\Lambda) - \varepsilon$. On a alors $\mu(\Lambda \backslash F) < \varepsilon$. Comme $\mu_n(F) \to \mu(F)$ et $\mu_n(\Lambda) \to \mu(\Lambda)$ par (5), on peut trouver un entier N tel que $\mu_n(F) > \mu_n(\Lambda) - \varepsilon$ pour tout $n \geqslant N$. Pour $n \geqslant N$, on a alors $\mu_n(\Lambda \backslash F) = \mu_n(\Lambda) - \mu_n(F) < \varepsilon$. □

Maintenant, soit $B \subseteq \Lambda$ quelconque. Soit aussi $\varepsilon > 0$, et soient F et N comme dans le Fait. Pour $n \geqslant N$, on a

$$\begin{aligned} |\mu_n(B) - \mu(B)| &\leqslant |\mu_n(B) - \mu_n(B \cap F)| + |\mu_n(B \cap F) - \mu(B \cap F)| \\ &\qquad + |\mu(B \cap F) - \mu(B)| \\ &< 2\varepsilon + |\mu_n(B \cap F) - \mu(B \cap F)|. \end{aligned}$$

Comme $\mu_n(B \cap F) \to \mu(B \cap F)$ puisque $B \cap F$ est fini, on en déduit

$$\overline{\lim}\, |\mu_n(B) - \mu(B)| \leqslant 2\varepsilon \quad \text{pour tout } \varepsilon > 0,$$

ce qui prouve que $\mu_n(B) \to \mu(B)$. □

4.1.3. *Cas où* $\Lambda = \mathbb{R}$.

Notations. On note $\mathcal{C}_{00}(\mathbb{R})$ l'ensemble des fonctions $f : \mathbb{R} \to \mathbb{R}$ continues à support compact, et $\mathcal{C}^\infty_{00}(\mathbb{R})$ l'ensemble des fonctions de classe $\mathcal{C}^\infty$ à support compact. On a ainsi

$$\mathcal{C}^\infty_{00}(\mathbb{R}) \subseteq \mathcal{C}_{00}(\mathbb{R}) \subseteq \mathcal{C}_b(\mathbb{R}).$$

Proposition 4.6. *Supposons que* $\Lambda = \mathbb{R}$. *Soit* (μ_n) *une suite dans* $\mathrm{M}_+(\mathbb{R})$, *et soit* $\mu \in \mathrm{M}_+(\mathbb{R})$. *Les propriétés suivantes sont équivalentes :*

(1) $\mu_n \twoheadrightarrow \mu$;
(5) $\mu_n(\mathbb{R}) \to \mu(\mathbb{R})$ *et* $\int_{\mathbb{R}} \varphi \, d\mu_n \to \int_{\mathbb{R}} \varphi \, d\mu$ *pour toute fonction* $\varphi \in \mathcal{C}_{00}(\mathbb{R})$.
(6) $\mu_n(\mathbb{R}) \to \mu(\mathbb{R})$ *et* $\int_{\mathbb{R}} \varphi \, d\mu_n \to \int_{\mathbb{R}} \varphi \, d\mu$ *pour toute fonction* $\varphi \in \mathcal{C}^\infty_{00}(\mathbb{R})$.

Démonstration. Évidemment, (1) $\Longrightarrow$ (5) $\Longrightarrow$ (6). De plus, on *sait* (*cf* le cours d'intégration ; approximation par convolution) que $\mathcal{C}^\infty_{00}(\mathbb{R})$ est dense dans $\mathcal{C}_{00}(\mathbb{R})$ pour la norme $\| \cdot \|_\infty$; autrement dit, pour toute $\varphi \in \mathcal{C}_{00}(\mathbb{R})$ et pour tout $\varepsilon > 0$ on peut trouver une fonction $\widetilde{\varphi} \in \mathcal{C}^\infty_{00}(\mathbb{R})$ telle que

$$\forall x \in \mathbb{R} \ : \ |\widetilde{\varphi}(x) - \varphi(x)| \leqslant \varepsilon.$$

Comme dans la preuve du Corollaire 4.4, on en déduit que (6) $\Longrightarrow$ (5).

Il reste à montrer que (5) $\Longrightarrow$ (1).

Supposons (5) vérifiée, et fixons $f \in \mathcal{C}_b(\mathbb{R})$: on veut montrer que $\int_{\mathbb{R}} f \, d\mu_n \to \int_{\mathbb{R}} f \, d\mu$ quand $n \to \infty$.

Fait. Pour tout $\varepsilon > 0$, il existe une fonction $\chi \in \mathcal{C}_{00}(\mathbb{R})$ vérifiant $0 \leqslant \chi \leqslant 1$ et un entier N, tels que $\int_{\mathbb{R}} (1 - \chi) \, d\mu < \varepsilon$ et $\int_{\mathbb{R}} (1 - \chi) \, d\mu_n < \varepsilon$ pour tout $n \geqslant N$.

Preuve du Fait. Comme la mesure μ est finie et comme $\mathbb{R}$ est réunion d'une suite croissante d'intervalles compacts, on peut trouver un intervalle compact I tel que $\mu(I_0) > \mu(\mathbb{R}) - \varepsilon$ (exo). Soit $\chi \in \mathcal{C}_{00}(\mathbb{R})$ telle que $0 \leqslant \chi \leqslant 1$ et $\chi \equiv 1$ sur I (exo : dessiner le graphe d'une telle fonction χ). Alors $\int_{\mathbb{R}} \chi \, d\mu > \mu(\mathbb{R}) - \varepsilon$ car $\chi \geqslant \mathbf{1}_I$, et donc $\int_{\mathbb{R}} (1 - \chi) \, d\mu = \mu(\mathbb{R}) - \int_{\mathbb{R}} \chi \, d\mu < \varepsilon$. De plus, comme $\mu_n(\mathbb{R}) \to \mu(\mathbb{R})$ et $\int_{\mathbb{R}} \chi \, d\mu_n \to \int_{\mathbb{R}} \chi \, d\mu$ par (5), on peut trouver un entier N tel que $\int_{\mathbb{R}} \chi \, d\mu_n > \mu_n(\mathbb{R}) - \varepsilon$ pour tout $n \geqslant N$; et alors $\int_{\mathbb{R}} (1 - \chi) \, d\mu_n < \varepsilon$ pour $n \geqslant N$. □

Soit $\varepsilon > 0$, et soient $\chi \in \mathcal{C}_{00}(\mathbb{R})$ et N comme dans le Fait. Alors $\varphi := \chi f$ appartient à $\mathcal{C}_{00}(\mathbb{R})$. Comme $f - \varphi = (1 - \chi) f$ et $1 - \chi \geqslant 0$, on a

$$\int_{\mathbb{R}} |f - \varphi| \, d\mu = \int_{\mathbb{R}} (1 - \chi) \, |f| \, d\mu \leqslant \|f\|_\infty \int_{\mathbb{R}} (1 - \chi) \, d\mu \leqslant \varepsilon \, \|f\|_\infty ;$$

et de même

$$\int_{\mathbb{R}} |\varphi - f|\, d\mu_n \leqslant \varepsilon \, \|f\|_\infty \qquad \text{pour tout } n \geqslant N.$$

D'après l'inégalité triangulaire, on en déduit

$$\left|\int_{\mathbb{R}} f\, d\mu_n - \int_{\mathbb{R}} f\, d\mu\right| \leqslant 2\varepsilon\, \|f\|_\infty + \left|\int_{\mathbb{R}} \varphi\, d\mu_n - \int_{\mathbb{R}} \varphi\, d\mu\right| \quad \text{pour tout } n \in \mathbb{N}.$$

Comme $\int_{\mathbb{R}} \varphi\, d\mu_n \to \int_{\mathbb{R}} \varphi\, d\mu$ par (5) et comme $\varepsilon > 0$ est quelconque, on peut donc conclure que $\int_{\mathbb{R}} f\, d\mu_n \to \int_{\mathbb{R}} f\, d\mu$. □

4.2. Convergence en loi d'une suite de va. Dans ce qui suit, on se donne une suite de va $(Z_n)_{n\in\mathbb{N}}$ et une va Z, toutes à valeurs dans un espace métrique (Λ, d).

Définition 4.7. *On dit que la suite (Z_n)* ***converge en loi vers*** *Z si la suite $(\mathbb{P}_{Z_n})$ converge étroitement vers $\mathbb{P}_Z$. On écrit alors $Z_n \xrightarrow{\mathcal{L}} Z$.*

Reformulation. Par définition, (Z_n) converge en loi vers Z si et seulement si

$$\mathbb{E}(f(Z_n)) \to \mathbb{E}(f(Z)) \qquad \text{pour toute } f \in \mathcal{C}_b(\Lambda).$$

Remarque 1. Les Z_n ne sont pas forcément définies sur le même $(\Omega, \mathfrak{A}, \mathbb{P})$.

Remarque 2. On a "unicité de la limite" au sens des lois : si $Z_n \xrightarrow{\mathcal{L}} Z$ et $Z_n \xrightarrow{\mathcal{L}} Z'$, alors $\mathbb{P}_Z = \mathbb{P}_{Z'}$.

Démonstration. C'est évident puisqu'on a unicité de la limite pour la convergence étroite. □

Exemple. Si les Z_n sont des va réelles suivant des lois $\mathcal{N}(0, \sigma_n)$ avec $\lim_{n\to\infty} \sigma_n = 0$, alors $Z_n \xrightarrow{\mathcal{L}} 0$.

Démonstration. Si on note $\rho : \mathbb{R} \to \mathbb{R}$ la "densité gaussienne standard", *i.e.* $\rho(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, alors $\mathbb{P}_{Z_n} = \frac{1}{\sigma_n}\rho(\frac{x}{\sigma_n})dx$ pour tout $n \in \mathbb{N}$. Donc $\mathbb{P}_{Z_n} \twoheadrightarrow \delta_0$ puisque $C_n := \frac{1}{\sigma_n} \to \infty$ (*cf* l'Exemple 2 après la définition de la convergence étroite). □

Exercice. Montrer que si Z et les Z_n sont des va réelles et si $Z_n \xrightarrow{\mathcal{L}} Z$, alors $aZ_n + b \xrightarrow{\mathcal{L}} aZ + b$ pour tous $a, b \in \mathbb{R}$. Plus généralement, montrer que si $Z_n \xrightarrow{\mathcal{L}} Z$, alors $\phi(Z_n) \xrightarrow{\mathcal{L}} \phi(Z)$ pour toute fonction continue $\phi : \Lambda \to \mathbb{R}$.

Proposition 4.8. *La convergence en probabilité entraine la convergence en loi : si Z et les Z_n sont des va réelles définies sur le même $(\Omega, \mathfrak{A}, \mathbb{P})$ et si $Z_n \xrightarrow{\text{proba}} Z$, alors $Z_n \xrightarrow{\mathcal{L}} Z$.*

Démonstration. Supposons que $Z_n \xrightarrow{\text{proba}} Z$, et fixons $f \in \mathcal{C}_b(\mathbb{R})$: il s'agit de montrer que $\mathbb{E}(f(Z_n)) \to \mathbb{E}(f(Z))$. Supposons que ce ne soit pas le cas. Alors certainement $\|f(Z_n) - f(Z)\|_1 \not\to 0$, donc on peut trouver $\varepsilon > 0$ et une sous-suite (Z'_n) de (Z_n) tels que $\forall n \in \mathbb{N} \ : \ \|f(Z'_n) - f(Z)\|_1 \geqslant \varepsilon$. Comme $Z_n \xrightarrow{\text{proba}} Z$, la suite (Z'_n) possède une sous-suite (Z'_{n_k}) telle que $Z'_{n_k} \xrightarrow{\text{ps}} Z$. Alors $f(Z'_{n_k}) \xrightarrow{\text{ps}} f(Z)$ car f est continue. De plus, la suite $(f(Z'_{n_k}))$ est uniformément bornée car f est bornée. Donc $\|f(Z'_{n_k}) - f(z)\|_1 \to 0$ par convergence dominée ; ce qui n'est pas vrai puisque $\|f(Z'_{n_k}) - f(Z)\|_1 \geqslant \varepsilon$ pour tout k. □

REMARQUE. La réciproque est fausse (*cf* TD).

PROPOSITION 4.9. *Une suite (Z_n) de va réelles converge en loi vers une va Z si et seulement si $\mathbb{P}(Z_n \in B) \to \mathbb{P}(Z \in B)$ pour tout borélien $B \subseteq \Lambda$ tel que $\mathbb{P}(Z \in \partial B) = 0$.*

Démonstration. C'est une conséquence de la Proposition 4.3. □

COROLLAIRE 4.10. *Supposons que Z et les Z_n soient des va réelles. Si $Z_n \xrightarrow{\mathcal{L}} Z$ et si de plus $\mathbb{P}_Z$ est une loi* diffuse *(par exemple, si Z est une va* à densité*), alors $\mathbb{P}(Z_n \in I) \to \mathbb{P}(Z \in I)$ pour tout intervalle $I \subseteq \mathbb{R}$.*

Démonstration. ∂I contient au plus 2 points, donc $\mathbb{P}_Z(\partial I) = 0$. □

COROLLAIRE 4.11. *Si Z et les Z_n sont des va réelles à densité et si $\rho_{Z_n}(x) \to \rho_Z(x)$ presque partout, alors $Z_n \xrightarrow{\mathcal{L}} Z$.*

Démonstration. Montrons qu'en fait $\mathbb{P}(Z_n \in B) \to \mathbb{P}(Z \in B)$ *pour tout borélien* $B \subseteq \mathbb{R}$. On a

$$\begin{aligned} |\mathbb{P}(Z_n \in B) - \mathbb{P}(Z \in B)| &= \left| \int_B \rho_{Z_n}(x)\,dx - \int_B \rho_Z(x)\,dx \right| \\ &\leqslant \int_B |\rho_{Z_n}(x) - \rho_Z(x)|\,dx \\ &\leqslant \|\rho_{Z_n} - \rho_Z\|_1. \end{aligned}$$

Donc il suffit de montrer que $\|\rho_{Z_n} - \rho_Z\|_1 \to 0$.

Posons $f_n := |\rho_{Z_n}| + |\rho_Z| - |\rho_{Z_n} - \rho_Z| = \rho_{Z_n} + \rho_Z - |\rho_{Z_n} - \rho_Z|$. Les f_n sont $\geqslant 0$ et $f_n \to 2\rho_Z$ presque partout. De plus, $\int_{\mathbb{R}} f_n = 2 - \int_{\mathbb{R}} |\rho_{Z_n} - \rho_Z|$ car $\int_{\mathbb{R}} \rho_Z = 1 = \int_{\mathbb{R}} \rho_{Z_n}$. Par le *Lemme de Fatou*, on a donc

$$2 = \int_{\mathbb{R}} 2\rho_Z = \int_{\mathbb{R}} \underline{\lim}\, f_n \leqslant \underline{\lim} \int_{\mathbb{R}} f_n = 2 - \overline{\lim}\, \|\rho_{Z_n} - \rho_Z\|_1.$$

On en déduit $\overline{\lim}\, \|\rho_{Z_n} - \rho_Z\|_1 = 0$, ce qui est le résultat souhaité. □

EXEMPLE. Si Z_n suit une loi normale $\mathcal{N}(m_n, \sigma_n^2)$ et si $m_n \to m \in \mathbb{R}$ et $\sigma_n \to \sigma > 0$, alors (Z_n) converge en loi vers une va $Z \sim \mathcal{N}(m, \sigma^2)$.

Démonstration. Micro-exo. □

EXERCICE. Soit $(X, \mathfrak{B}, \mu)$ un espace mesuré, et soit (f_n) une suite dans $L^1(\mu)$. On suppose que f_n converge presque partout vers une fonction $f \in L^1(\mu)$, et que de plus $\|f_n\|_1 \to \|f\|_1$. Montrer que f_n tend vers f en norme L^1.

COROLLAIRE 4.12. *Si les Z_n sont des va réelles définies sur le même $(\Omega, \mathfrak{A}, \mathbb{P})$ et si les Z_n convergent en loi vers une va Z* constante*, alors $Z_n \xrightarrow{\text{proba}} Z$.*

Démonstration. Soit a l'unique valeur prise par Z. Pour $\varepsilon > 0$, on a $\mathbb{P}(|Z_n - Z| \geqslant \varepsilon) = \mathbb{P}(|Z_n - a| \geqslant \varepsilon) = \mathbb{P}(Z_n \in B_\varepsilon)$, où $B_\varepsilon =]-\infty, a-\varepsilon] \cup [a+\varepsilon, \infty[$. De plus, comme $\mathbb{P}_Z = \delta_a$, on a $\mathbb{P}(Z \in \partial B_\varepsilon) = \delta_a(\{a-\varepsilon\} \cup \{a+\varepsilon\}) = 0$. Donc $\mathbb{P}(|Z_n - a| \geqslant \varepsilon) = \mathbb{P}(Z_n \in B_\varepsilon) \to \mathbb{P}(Z \in B_\varepsilon) = \delta_a(B_\varepsilon) = 0$. □

PROPOSITION 4.13. *Supposons que Z et les Z_n soient discrètes, à valeurs dans un espace métrique Λ dénombrable. Si $\mathbb{P}(Z_n = x) \to \mathbb{P}(Z = x)$ pour tout $x \in \Lambda$, alors $\mathbb{P}(Z_n \in B) \to \mathbb{P}(Z \in B)$ pour tout $B \subseteq \Lambda$ et donc $Z_n \xrightarrow{\mathcal{L}} Z$.*

Démonstration. C'est une conséquence de la Proposition 4.5. □

COROLLAIRE 4.14. (Théorème de Poisson)
Supposons que Z_n suive une loi binomiale $\mathfrak{B}(n, p_n)$, avec $p_n \sim \frac{\lambda}{n}$ pour un certain $\lambda > 0$. Alors (Z_n) converge en loi vers une va Z suivant la loi de Poisson $\mathcal{P}(\lambda)$.

Démonstration. Les Z_n sont toutes à valeurs dans $\Lambda = \mathbb{N}$. Pour $k \in \mathbb{N}$ fixé et pour tout $n > k$, on a

$$\begin{aligned}\mathbb{P}(Z_n = k) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \frac{1}{k!}\, n(n-1)\cdots(n-k+1) p_n^k (1-p_n)^{n-k};\end{aligned}$$

et donc

$$\mathbb{P}(Z_n = k) \sim \frac{1}{k!}\, n^k p_n^k (1-p_n)^{n-k} \quad \text{quand } n \to \infty.$$

De plus, comme $p_n \sim \frac{\lambda}{n}$, on voit que $n^k p_n^k \to \lambda^k$; et comme $p_n \to 0$, on voit aussi que $(1-p_n)^{n-k} = e^{((n-k)\log(1-p_n)} \to e^{-\lambda}$ quand $n \to \infty$ (car $\log(1-p_n) \sim -p_n \sim -\frac{\lambda}{n}$). Donc

$$\mathbb{P}(Z_n = k) \xrightarrow{n\to\infty} \frac{\lambda^k}{k!}\, e^{-\lambda} \qquad \text{pour tout } k \in \mathbb{N}.$$

□

PROPOSITION 4.15. *Supposons que Z et les Z_n soient des va réelles. Alors $Z_n \xrightarrow{\mathcal{L}} Z$ si et seulement si $\mathbb{E}(\varphi(Z_n)) \to \mathbb{E}(\varphi(Z))$ pour toute fonction $\varphi \in \mathcal{C}^\infty_{00}(\mathbb{R})$.*

Démonstration. C'est une conséquence de la Proposition 4.6. □

PROPOSITION 4.16. *Supposons que Z et les Z_n soient des va réelles. Alors les propriétés suivantes sont équivalentes :*

(i) *$Z_n \xrightarrow{\mathcal{L}} Z$;*

(ii) *$F_{Z_n} \to F_Z(t)$ en tout point $t \in \mathbb{R}$ où la fonction F_Z est continue.*

Démonstration. On notera D l'ensemble des points de *discontinuité* de F_Z.

FAIT. On a $D = \{t \in \mathbb{R};\ \mathbb{P}_Z(\{t\}) > 0\}$.

Preuve du Fait. On l'a vu au Chapitre 3 (c'était le Fait 6.3). □

Par le Fait, un nombre réel t est un point de continuité de F_Z si et seulement si $\mathbb{P}_Z(\partial I_t) = 0$, où $I_t =]-\infty, t]$. Donc, si (i) est vérifiée, alors $\mathbb{P}_{Z_n}(I_t) \to \mathbb{P}_Z(I_t)$ en tout point t où F_Z est continue d'après la Proposition 4.3 ; autrement dit (ii) est vérifiée.

Inversement, supposons (ii) vérifiée. Pour montrer que $Z_n \xrightarrow{\mathcal{L}} Z$, il suffit de vérifier que $\mathbb{E}(\varphi(Z_n)) \to \mathbb{E}(\varphi(Z))$ pour toute fonction $\varphi \in \mathcal{C}^\infty_{00}(\mathbb{R})$.

Par définition, $\mathbb{E}(\varphi(Z_n)) = \int_\Omega \varphi(Z_n(\omega))\, d\mathbb{P}(\omega)$. De plus, comme $\varphi(t) \to 0$ quand $t \to -\infty$, on a aussi $\varphi((Z_n(\omega)) = \int_{-\infty}^{Z_n(\omega)} \varphi'(t)\, dt$ pour tout $\omega \in \Omega$, ce qui peut encore

s'écrire $\varphi((Z_n(\omega)) = \int_{\mathbb{R}} \mathbf{1}_{]t,\infty[}(Z_n(\omega))\varphi'(t)dt$. En appliquant le Théorème de Fubini, on obtient donc

$$\begin{aligned}\mathbb{E}\big(\varphi(Z_n)\big) &= \int_\Omega \left(\int_{\mathbb{R}} \mathbf{1}_{]t,\infty[}(Z_n(\omega))\, \varphi'(t)\, dt \right) d\mathbb{P}(\omega) \\ &= \int_{\mathbb{R}} \varphi'(t)\mathbb{P}(Z_n > t)\, dt \\ &= \int_{\mathbb{R}} \varphi'(t)\big(1 - F_{Z_n}(t)\big)\, dt;\end{aligned}$$

et de même

$$\mathbb{E}\big(\varphi(Z)\big) = \int_{\mathbb{R}} \varphi'(t)\big(1 - F_Z(t)\big)\, dt.$$

(Exo : justifier l'utilisation du Théorème de Fubini.)

Par (ii), on sait que $1 - F_{Z_n}(t) \to 1 - F_Z(t)$ en tout point $t \notin D$. De plus, $D = \{t;\ \mathbb{P}_Z(\{t\}) > 0\}$ est dénombrable, donc en particulier λ_1-négligeable ; et donc $1 - F_{Z_n}(t) \to 1 - F_Z(t)$ *presque partout*. Comme $\varphi' \in L^1(\mathbb{R})$ (car φ' est continue à support compact) et $|(1-F_{Z_n}(t))\varphi'(t)| \leqslant |\varphi'(t)|$ pour tout $n \in \mathbb{N}$, on en déduit par convergence dominée que

$$\mathbb{E}\big(\varphi(Z_n)\big) \xrightarrow{n\to\infty} \int_{\mathbb{R}} \varphi'(t)\big(1 - F_Z(t)\big)\, dt = \mathbb{E}\big(\varphi(Z)\big).$$

□

Corollaire 4.17. *Si $F_{Z_n}(t) \to F_Z(t)$ pour tout $t \in \mathbb{R}$, alors $Z_n \xrightarrow{\mathcal{L}} Z$.*

Démonstration. C'est évident par la proposition. □

Corollaire 4.18. *Si $\mathbb{P}(Z_n \in I) \to \mathbb{P}(Z \in I)$ pour tout intervalle $I \subseteq \mathbb{R}$, alors $Z_n \xrightarrow{\mathcal{L}} Z$.*

Démonstration. C'est une conséquence du corollaire précédent (micro-exo). □

Exercice. Montrer que si $\mathbb{P}(a < Z_n \leqslant b) \to \mathbb{P}(a < Z \leqslant b)$ pour tous $a, b \in \mathbb{R}$, alors $Z_n \xrightarrow{\mathcal{L}} Z$.

Chapitre 8

Fonctions caractéristiques

1. Transformation de Fourier dans $L^1(\mathbb{R})$

RAPPEL. Si $f : \mathbb{R} \to \mathbb{C}$ est une fonction intégrable, sa **transformée de Fourier** est la fonction $\widehat{f} : \mathbb{R} \to \mathbb{C}$ définie par

$$\widehat{f}(x) := \int_{\mathbb{R}} f(t)\, e^{-ixt}\, dt \qquad \text{pour tout } x \in \mathbb{R}.$$

REMARQUE. Si $f \in L^1(\mathbb{R})$, alors $\widehat{f}$ est une fonction *continue bornée*, avec

$$\|\widehat{f}\|_\infty \leqslant \|f\|_1.$$

Démonstration. La continuité de $\widehat{f}$ est une conséquence du théorème de continuité pour les intégrales à paramètres, applicable car la fonction $F(t,x) = f(t)e^{-ixt}$ est continue par rapport à x, intégrable par rapport à t, et $|F(t,x)| = |f(t)|$, fonction intégrable sur $\mathbb{R}$ indépendante de x. Enfin,

$$|\widehat{f}(x)| \leqslant \int_{\mathbb{R}} |f(t)e^{-ixt}|\, dt = \int_{\mathbb{R}} |f(t)|\, dt = \|f\|_1 \quad \text{pour tout } x \in \mathbb{R},$$

donc $\widehat{f}$ est bornée et $\|\widehat{f}\|_\infty \leqslant \|f\|_1$. □

EXEMPLE 1.1. Soit $g : \mathbb{R} \to \mathbb{R}$ la fonction définie par

$$g(t) = e^{-\frac{t^2}{2}}.$$

Pour tout $x \in \mathbb{R}$, on a

$$\widehat{g}(x) = \sqrt{2\pi}\, e^{-\frac{x^2}{2}}.$$

Démonstration. Par définition,

$$\widehat{g}(x) = \int_{\mathbb{R}} e^{-ixt} e^{-\frac{t^2}{2}}\, dt$$

En utilisant le théorème de dérivation des intégrales à paramètres (exo), on voit que $\widehat{g}$ est dérivable sur $\mathbb{R}$ avec

$$\widehat{g}\,'(x) = -i \int_{\mathbb{R}} t e^{-ixt} e^{-\frac{t^2}{2}}\, dt.$$

Puis, en remarquant que $te^{-\frac{t^2}{2}} = -\frac{d}{dt}\big(e^{-\frac{t^2}{2}}\big)$ et faisant une intégration par parties, on obtient (exo)

$$\widehat{g}\,'(x) = i \int_{\mathbb{R}} ixe^{-ixt} e^{-\frac{t^2}{2}}\, dt = -x\, \widehat{g}(x).$$

Donc $\widehat{g}(x) = Ce^{-x^2/2}$ pour une certaine constante C ; qui se calcule : $C = \widehat{g}(0) = \int_{\mathbb{R}} e^{-t^2/2} dt = \sqrt{2\pi}$. □

Exercice. Montrer que si $\alpha > 0$, alors on a pour tout $x \in \mathbb{R}$:

$$\int_{\mathbb{R}} e^{-\frac{\alpha^2 t^2}{2}}\, e^{-ixt} dt = \frac{\sqrt{2\pi}}{\alpha}\, e^{-\frac{x^2}{2\alpha^2}}.$$

THÉORÈME 1.2. (Formule d'inversion de Fourier)
Soit $f \in L^1(\mathbb{R})$. Si f est continue et si la fonction $\widehat{f}$ est intégrable sur $\mathbb{R}$, alors on a pour tout $t \in \mathbb{R}$:

$$f(t) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(x) e^{itx} dx.$$

Démonstration. On va la faire en supposant de plus que la fonction f est *bornée*.
Soit $t \in \mathbb{R}$ quelconque. On veut montrer que

$$\int_{\mathbb{R}} \widehat{f}(x) e^{itx} dx = 2\pi\, f(t).$$

Le point de départ est d'observer que

$$\int_{\mathbb{R}} \widehat{f}(x) e^{itx} dx = \lim_{\alpha \to 0^+} \int_{\mathbb{R}} \widehat{f}(x) e^{itx}\, e^{-\frac{\alpha^2 x^2}{2}}\, dx;$$

ce qui se montre par convergence dominée (exo). Ensuite, on a pour tout $\alpha > 0$,

$$\begin{aligned}
\int_{\mathbb{R}} \widehat{f}(x) e^{itx}\, e^{-\frac{\alpha^2 x^2}{2}}\, dx &= \int_{\mathbb{R}} e^{-\frac{\alpha^2 x^2}{2}}\, e^{itx} \left(\int_{\mathbb{R}} f(u) e^{-ixu} du \right) dx \\
&= \int_{\mathbb{R}} f(u) \left(e^{-\frac{\alpha^2 x^2}{2}} e^{-i(u-t)x} dx \right) du \\
&= \int_{\mathbb{R}} f(u) \times \frac{\sqrt{2\pi}}{\alpha}\, e^{-\frac{(u-t)^2}{2\alpha^2}}\, du \\
&= \sqrt{2\pi} \int_{\mathbb{R}} f(t + \alpha v)\, e^{-\frac{v^2}{2}}\, dv,
\end{aligned}$$

où on a utilisé le Théorème de Fubini, l'Exemple 1.1 et le changement de variable $v = \frac{u-t}{\alpha}$.

Comme f est supposé continue et bornée, on en déduit par convergence dominée que

$$\lim_{\alpha \to 0^+} \int_{\mathbb{R}} \widehat{f}(x) e^{itx}\, e^{-\frac{\alpha^2 x^2}{2}}\, dx = \sqrt{2\pi} \int_{\mathbb{R}} f(t)\, e^{-\frac{v^2}{2}}\, dv = 2\pi\, f(t);$$

ce qui termine la démonstration. □

COROLLAIRE 1.3. *Toute fonction $\varphi : \mathbb{R} \to \mathbb{C}$ de classe $\mathcal{C}^\infty$ à support compact est une transformée de Fourier, i.e. $\varphi = \widehat{f}$ pour une certaine $f \in L^1(\mathbb{R})$.*

Démonstration. Soit $\varphi \in \mathcal{C}^\infty_{00}(\mathbb{R})$, et soit $A < \infty$ telle que $\varphi \equiv 0$ en dehors de $[-A, A]$. Par continuité de φ et de ses dérivées, on a alors $\varphi^{(k)}(A) = 0 = \varphi^{(k)}(-A)$ pour tout $k \in \mathbb{N}$.

Si $x \neq 0$, une intégration par parties donne

$$\begin{aligned}\widehat{\varphi}(x) &= \int_{-A}^{A} \varphi(t) e^{-ixt} dt \\ &= \Big[\frac{-1}{ix} e^{-ixt}\, \varphi(t)\Big]_{-A}^{A} + \frac{1}{ix}\int_{-A}^{A} \varphi'(t) e^{-ixt} dt \\ &= \frac{1}{ix}\,\widehat{\varphi'}\,(x).\end{aligned}$$

En intégrant par parties une 2ème fois, on obtient de même

$$\widehat{\varphi}(x) = -\frac{1}{x^2}\,\widehat{\varphi''}(x).$$

Comme la fonction $\widehat{\varphi''}$ est bornée sur $\mathbb{R}$, on en déduit que $|\widehat{\varphi}(x)| = O(1/x^2)$ quand $x \to \pm\infty$; et comme $\widehat{\varphi}$ est continue, donc localement intégrable, cela montre que $\widehat{\varphi}$ est intégrable sur $\mathbb{R}$.

Par la formule d'inversion de Fourier, on en déduit que pour tout $t \in \mathbb{R}$, on a

$$\begin{aligned}\varphi(t) &= \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{\varphi}(x) e^{itx} dx \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{\varphi}(-x) e^{-itx}.\end{aligned}$$

Autrement dit :

$$\varphi = \widehat{f} \qquad \text{où} \qquad f(x) = \frac{1}{2\pi}\,\widehat{\varphi}(-x).$$

□

2. Transformée de Fourier d'une mesure

NOTATION. On note $\mathrm{M}_+(\mathbb{R})$ l'ensemble des mesures boréliennes *finies* sur $\mathbb{R}$.

DÉFINITION 2.1. *Si $\mu \in \mathrm{M}_+(\mathbb{R})$, la **transformée de Fourier** de μ est la* fonction $\widehat{\mu} : \mathbb{R} \to \mathbb{C}$ *définie par*

$$\widehat{\mu}(t) = \int_{\mathbb{R}} e^{-itx} d\mu(x) \qquad \textit{pour tout } x \in \mathbb{R}.$$

REMARQUE. La définition a un sens, et $\widehat{\mu}$ est toujours une fonctions *continue bornée*, avec $\|\widehat{\mu}\|_\infty = \mu(\mathbb{R}) = \widehat{\mu}(0)$.

Démonstration. La définition a un sens car la fonction $x \mapsto e^{-itx}$ est bornée et μ est une mesure finie. La continuité est laissée en exo. Enfin, on a $\widehat{\mu}(0) = \int_{\mathbb{R}} e^{i0x}\, d\mu(x) = \mu(\mathbb{R})$, et

$$|\widehat{\mu}(t)| \leqslant \int_{\mathbb{R}} |e^{-itx}|\, d\mu(x) = \int_{\mathbb{R}} \mathbf{1}\, d\mu(x) = \mu(\mathbb{R}) \quad \text{pour tout } t \in \mathbb{R}.$$

□

EXEMPLE. Soit $g \in L^1(\mathbb{R})$, avec $g \geqslant 0$. Si on note μ_g la mesure définie par $\mu_g(A) = \int_A g(t)\, dt$ pour tout borélien $A \subseteq \mathbb{R}$, alors $\widehat{\mu_g} = \widehat{g}$.

Démonstration. Tout repose sur le fait suivant (déjà essentiellement démontré au Chapitre 2 quand on a parlé de lois à densité) : si $\phi : \mathbb{R} \to \mathbb{C}$ est une fonction ou bien positive, ou bien intégrable par rapport à μ_g, alors

$$\int_{\mathbb{R}} \phi \, d\mu_g = \int_{\mathbb{R}} \phi(x) g(x) \, dx.$$

Pour le montrer, on commence par le cas où ϕ et étagée, puis on utilise le théorème de convergence monotone si $\phi \geqslant 0$, et on conclut par linéarité.

En prenant $\phi_t(x) = e^{-itx}$ pour $t \in \mathbb{R}$ fixé, on obtient immédiatement le résultat :

$$\widehat{\mu_g}(t) = \int_{\mathbb{R}} \phi_t \, d\mu_g = \int_{\mathbb{R}} g(x) \phi_t(x) \, dx = \widehat{g}(t).$$

□

LEMME 2.2. *Si $\mu \in \mathrm{M}_+(\mathbb{R})$ et $f \in L^1(\mathbb{R})$ alors la fonction $f\widehat{\mu}$ est intégrable sur $\mathbb{R}$, la fonction $\widehat{f}$ est intégrable par rapport à μ, et on a la* ***formule d'échange***

$$\int_{\mathbb{R}} f(t) \, \widehat{\mu}(t) \, dt = \int_{\mathbb{R}} \widehat{f}(x) \, d\mu(x).$$

Démonstration. La fonction $f\widehat{\mu}$ est intégrable sur $\mathbb{R}$ car $f \in L^1(\mathbb{R})$ et $\widehat{\mu}$ est bornée ; et $\widehat{f}$ est est intégrable par rapport à μ car elle est bornée et μ est une mesure finie. Ensuite, on applique Fubini :

$$\int_{\mathbb{R}} f(t) \, \widehat{\mu}(t) \, dt = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{-itx} d\mu(x) \right) dt = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(t) e^{-ixt} dt \right) d\mu(x) = \int_{\mathbb{R}} \widehat{f} \, d\mu.$$

La justification du calcul formel est laissée en exercice. □

THÉORÈME 2.3. *Soit (μ_n) une suite dans $\mathrm{M}_+(\mathbb{R})$, et soit $\mu \in \mathrm{M}_+(\mathbb{R})$. Les propriétés suivantes sont équivalentes :*

(a) *(μ_n) converge étroitement vers μ ;*

(b) *$\widehat{\mu_n}(t) \to \widehat{\mu}(t)$ pour tout $t \in \mathbb{R}$.*

Démonstration. Pour $t \in \mathbb{R}$ fixé, la fonction ϕ_t définie par $\phi_t(x) = e^{-itx}$ est continue bornée sur $\mathbb{R}$. Donc, si $\mu_n \twoheadrightarrow \mu$, alors $\widehat{\mu_n}(t) = \int_{\mathbb{R}} \phi_t \, d\mu_n \to \int_{\mathbb{R}} \phi_t \, d\mu = \widehat{\mu}(t)$. Ainsi, (a) entraine (b).

Inversement, supposons (b) vérifiée. Alors $\mu_n(\mathbb{R}) = \widehat{\mu_n}(0) \to \widehat{\mu}(0) = \mu(\mathbb{R})$. Pour montrer que $\mu_n \twoheadrightarrow \mu$, il suffit donc, d'après la Proposition 4.6 du Chapitre 7, de montrer que $\int_{\mathbb{R}} \varphi \, d\mu_n \to \int_{\mathbb{R}} \varphi \, d\mu$ pour toute fonction $\varphi \in \mathcal{C}^\infty_{00}(\mathbb{R})$.

Par le Corollaire 1.3, on peut trouver une fonction $f \in L^1(\mathbb{R})$ telle que $\varphi = \widehat{f}$. En utilisant la formule d'échange, on en déduit ceci :

$$\begin{aligned} \int_{\mathbb{R}} \varphi \, d\mu_n &= \int_{\mathbb{R}} \widehat{f} \, d\mu_n \\ &= \int_{\mathbb{R}} f(t) \, \widehat{\mu_n}(t) \, dt \\ &\xrightarrow{n\to\infty} \int_{\mathbb{R}} f(t) \, \widehat{\mu}(t) \, dt = \int_{\mathbb{R}} \widehat{f} \, d\mu. \end{aligned}$$

Le passage à la limite est justifié par le théorème de convergence dominée car $|f(t) \, \widehat{\mu_n}(t)| \leqslant \mu_n(\mathbb{R}) \, |f(t)| \leqslant C \, |f(t)|$ pour une certaine constante C (la suite convergente $(\mu_n(\mathbb{R}))$ est bornée) et $f \in L^1(\mathbb{R})$. □

COROLLAIRE 2.4. *La transformée de Fourier* caractérise la mesure *: si* $\mu, \mu' \in \mathrm{M}_+(\mathbb{R})$ *vérifient* $\widehat{\mu} = \widehat{\mu'}$*, alors* $\mu = \mu'$*.*

Démonstration. On applique le théorème avec $\mu_n = \mu$ pour tout n. Alors d'une part $\mu_n \twoheadrightarrow \mu$ (!) ; et d'autre part $\mu_n \twoheadrightarrow \mu'$ par le théorème puisque $\widehat{\mu_n} = \widehat{\mu} = \widehat{\mu'}$ pour tout n. Donc $\mu = \mu'$ par "unicité de la limite" pour la convergence étroite. □

3. Fonctions caractéristiques et applications

3.1. Définition et exemples.

DÉFINITION 3.1. *Soit X une va réelle. La* ***fonction caractéristique*** *de X est la fonction $\varphi_X : \mathbb{R} \to \mathbb{C}$ définie par*

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

Autrement dit :

$$\varphi_X(t) = \mathbb{E}\left(e^{itX}\right).$$

REMARQUE ESSENTIELLE. La fonction caractéristique détermine la loi : si X et X' sont deux va telles que $\varphi_X = \varphi_{X'}$, alors $\mathbb{P}_X = \mathbb{P}_{X'}$.

Démonstration. C'est évident puisque "la transformée de Fourier caractérise la mesure". □

FAIT 3.2. *On a $\varphi_X(0) = 1$ et $|\varphi_X(t)| \leqslant 1$ pour tout $t \in \mathbb{R}$.*

Démonstration. On a vu que pour toute mesure $\mu \in M_+(\mathbb{R})$, on a $|\widehat{\mu}(t)| \leqslant \mu(\mathbb{R}) = \widehat{\mu}(0)$; et ici $\mu = \mathbb{P}_X$ vérifie $\mathbb{P}_X(\mathbb{R}) = 1$. □

EXEMPLE 1. Si X suit une loi normale $\mathcal{N}(m, \sigma^2)$, alors

$$\varphi_X(t) = e^{imt} e^{-\sigma^2 \frac{t^2}{2}}.$$

Démonstration. Supposons d'abord que X suive la loi $\mathcal{N}(0,1)$. Alors

$$\varphi_X(t) = \mathbb{E}\big(e^{itX}\big) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{itx} e^{-\frac{x^2}{2}}\, dx = \frac{1}{\sqrt{2\pi}} \times \sqrt{2\pi} e^{-\frac{t^2}{2}} = e^{-\frac{t^2}{2}},$$

d'après l'Exemple 1.1

Dans le cas général, si on pose $Z := \frac{X-m}{\sigma}$, alors Z suit la loi $\mathcal{N}(0,1)$ (exo déjà posé) ; et comme $X = \sigma Z + m$, on a

$$\varphi_X(t) = \mathbb{E}\big(e^{it(\sigma Z + m)}\big) = e^{imt}\mathbb{E}\big(e^{i\sigma t Z}\big) = e^{imt}\varphi_Z(\sigma t) = e^{imt} e^{-\sigma^2 \frac{t^2}{2}}.$$

□

EXEMPLE 2. Si X suit une loi de Poisson de paramètre $\lambda > 0$, alors

$$\varphi_X(t) = e^{\lambda(e^{it}-1)}.$$

Démonstration. Comme X est une va à valeurs dans $\mathbb{N}$, on a

$$\begin{aligned}\varphi_X(t) = \mathbb{E}\big(e^{itX}\big) &= \sum_{k=0}^{\infty} \mathbb{P}(X=k)\, e^{itk} \\ &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} e^{itk} \\ &= \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} e^{-\lambda} = e^{\lambda e^{it}} e^{-\lambda}.\end{aligned}$$

□

Exemple 3. Si X suit une loi de Bernoulli de paramètre p, alors

$$\varphi_X(t) = pe^{it} + 1 - p.$$

Démonstration. Exo. □

Exemple 4. Si X suit une loi binômiale $\mathcal{B}(n,p)$, alors

$$\varphi_X(t) = \big(pe^{it} + 1 - p\big)^n.$$

Démonstration. On a

$$\begin{aligned}\varphi_X(t) = \mathbb{E}\big(e^{itX}\big) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{itk} \\ &= \sum_{k=0}^{n} \binom{n}{k} (pe^{it})^k (1-p)^{n-k} \\ &= \big(pe^{it} + 1 - p\big)^n \qquad \text{par la formule du binôme.}\end{aligned}$$

□

Exemple 5. Si X suit une loi de Cauchy de paramètre $a > 0$, *i.e.* $\mathbb{P}_X = \frac{a}{\pi} \frac{dx}{a^2+x^2}$, alors

$$\varphi_X(t) = \pi\, e^{-a\,|t|}.$$

Démonstration. On utilise la formule d'inversion de Fourier. Soit $f : \mathbb{R} \to \mathbb{R}$ la fonction définie par

$$f(t) := e^{-a\,|t|}.$$

La fonction f est clairement intégrable sur $\mathbb{R}$, et sa transformée de Fourier n'est pas difficile à calculer :

$$\begin{aligned}\widehat{f}(x) &= \int_{\mathbb{R}} e^{-a|t|}\, e^{-ixt}\, dt \\ &= \int_{-\infty}^{0} e^{(a-ix)t}\, dt + \int_{0}^{\infty} e^{(-a-ix)t}\, dt \\ &= \frac{1}{a-ix} + \frac{1}{a+ix} \\ &= \frac{2a}{a^2+x^2}.\end{aligned}$$

Donc $\widehat{f}$ est intégrable sur $\mathbb{R}$; et comme f est de plus continue, on a par la formule d'inversion :

$$f(t) = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(x)\, e^{itx}\, dx \qquad \text{pour tout } t \in \mathbb{R},$$

autrement dit $f(t) = \frac{a}{\pi}\int_{\mathbb{R}} \frac{e^{itx}}{a^2+x^2}\, dx = \varphi_X(t)$. □

La remarque suivant est évidente mais souvent utile.

REMARQUE 3.3. Si X est une va réelle, alors

$$\forall a, t \in \mathbb{R} \ : \ \varphi_{aX}(t) = \varphi_X(at).$$

Démonstration. Exo. □

3.2. Fonctions caractéristiques et sommes.

PROPOSITION 3.4. *Si X et Y sont deux va réelles indépendantes, alors $\varphi_{X+Y} = \varphi_X\, \varphi_Y$.*

Démonstration. C'est évident :

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

où on a utilisé le fait que les va e^{itX} et e^{itY} sont indépendantes. □

EXEMPLE 1. Si X et Y sont deux va réelles indépendantes suivant des lois normales $\mathcal{N}(m_X, \sigma_X^2)$ et $\mathcal{N}(m_Y, \sigma_Y^2)$, alors $X + Y$ suit la loi $\mathcal{N}(m, \sigma^2)$, où $m = m_X + m_Y$ et $\sigma^2 = \sigma_X^2 + \sigma_Y^2$.

Démonstration. On a

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) = e^{im_X t}e^{-\sigma_X^2\frac{t^2}{2}}\, e^{im_Y t}e^{-\sigma_Y^2\frac{t^2}{2}} = e^{imt}e^{-\sigma^2\frac{t^2}{2}},$$

qui est la fonction caractéristique d'une loi $\mathcal{N}(m, \sigma^2)$; d'où le résultat puisque la fonction caractéristique détermine la loi. □

Remarque. Cette preuve est quand même nettement plus limpide que celle donnée au Chapitre 5 !

EXEMPLE 2. Si X et Y sont des va indépendantes suivant des lois de Poisson $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$, alors $X + Y$ suit la loi $\mathcal{P}(\lambda + \mu)$.

Démonstration. On a

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) = e^{\lambda(e^{it}-1)}e^{\mu(e^{it}-1)} = e^{(\lambda+\mu)(e^{it}-1)},$$

qui est la fonction caractéristique d'une loi $\mathcal{P}(\lambda + \mu)$. □

EXEMPLE 3. Si $X_1, \ldots, X_n$ sont des va indépendantes suivant une loi de Bernoulli de paramètre p, alors $X_1 + \cdots + X_n$ suit la loi binômiale $\mathcal{B}(n, p)$.

Démonstration. Cette fois, en posant $S = X_1 + \cdots + X_n$, on a

$$\varphi_S(t) = \prod_{k=1}^{n} \varphi_{X_k}(t) = \big(pe^{it} + 1 - p\big)^n,$$

qui est la fonction caractéristique d'une loi $\mathcal{B}(n, p)$. □

EXEMPLE 4. Si X et Y sont des va indépendantes suivant des lois de Cauchy de paramètres a et b, alors $X + Y$ suit la loi de Cauchy de paramètre $a + b$.

Démonstration. Exo. □

3.3. Régularité.

PROPOSITION 3.5. *Soit X une va réelle.*

(1) *φ_X est une fonction* continue bornée, *avec $\|\varphi_X\|_\infty = 1 = \varphi_X(0)$.*

(2) *Si X admet un moment d'ordre $N \geqslant 1$, alors φ_X est de classe $\mathcal{C}^N$, avec*

$$\varphi_X^{(k)}(t) = i^k\,\mathbb{E}\big(X^k e^{itX}\big) \qquad \textit{pour } k = 1, \dots, N.$$

En particulier,

$$\varphi^{(k)}(0) = i^k\,\mathbb{E}(X^k).$$

Démonstration. La partie (1) a déjà été vue ; et pour (2), on utilise le théorème de dérivation des intégrales à paramètres (Faire l'exo soigneusement). □

Remarque. En général, une fonction caractéristique n'a pas de raisons d'être "mieux que continue". Par exemple, si X suit une loi de Cauchy de paramètre $a > 0$, alors $\varphi_X(t) = e^{-a|t|}$ n'est pas dérivable en 0.

COROLLAIRE 3.6. *Si X est une va réelle* bornée, *alors φ_X est développable en série entière sur $\mathbb{R}$.*

Démonstration. Comme X est bornée, elle admet des moments de tous ordres, donc φ_X est de classe $\mathcal{C}^\infty$ sur $\mathbb{R}$.

Par la formule de Taylor, il suffit de montrer que pour tout $t \in \mathbb{R}$,

$$R_n(t) := \int_0^1 \frac{(1-s)^n}{n!}\varphi^{(n+1)}(st)\,t^{n+1}\,ds \quad \text{tend vers 0 quand } n \to \infty.$$

Si on pose $C = \|X\|_\infty$, alors on a pour tout $k \in \mathbb{N}$ et pour tout $u \in \mathbb{R}$:

$$|\varphi_X^{(k)}(u)| = |\mathbb{E}(X^k e^{iuX})| \leqslant \mathbb{E}(|X|^k) \leqslant C^k.$$

Donc

$$|R_n(t)| \leqslant \int_0^1 \frac{(1-s)^n}{n!}\,C^{n+1}\,|t|^{n+1}\,ds = \frac{(C|t|)^{n+1}}{(n+1)!};$$

ce qui donne le résultat souhaité car $\frac{A^n}{n!} \to 0$ quand $n \to \infty$, pour tout $A \in \mathbb{R}$. □

Autre preuve. On écrit la définition de $\varphi_X(t)$ et on développe l'exponentielle en série entière.

$$\begin{aligned}
\varphi_X(t) &= \int_\Omega e^{itX}\,d\mathbb{P} \\
&= \int_\Omega \sum_{k=0}^{\infty} \frac{(itX)^k}{k!}\,d\mathbb{P} \\
&= \sum_{k=0}^{\infty} \int_\Omega \frac{(itX)^k}{k!}\,d\mathbb{P} \\
&= \sum_{k=0}^{\infty} c_k t^k \qquad \text{où } c_k := \tfrac{i^k}{k!}\mathbb{E}(X^k).
\end{aligned}$$

Exo : justifier soigneusement l'interversion $\sum$-$\int$. □

EXEMPLE. Théorème des moments

Soient X et Y deux va réelles bornées telles $\mathbb{E}(X^k) = \mathbb{E}(Y^k)$ pour tout $k \in \mathbb{N}$. Montrons que X et Y ont la même loi (*cf* le chapitre "Espérance, variance, moments").

Par la proposition, φ_X et φ_Y sont développables en série entière sur $\mathbb{R}$, et $\varphi_X^{(k)}(0) = i^k\mathbb{E}(X^k) = i^k\mathbb{E}(Y^k) = \varphi_Y^{(k)}(0)$ pour tout $k \in \mathbb{N}$. Donc $\varphi_X = \varphi_Y$, et donc X et Y ont la même loi ! À savoir refaire.

3.4. Fonctions caractéristiques et convergence en loi.

THÉORÈME 3.7. *Soit (Z_n) une suite de va réelles, et soit Z une va réelle. Alors (Z_n) converge en loi vers Z si et seulement si $\varphi_{Z_n}(t) \to \varphi_Z(t)$ pour tout $t \in \mathbb{R}$.*

Démonstration. C'est la traduction du Théorème 2.3. □

EXEMPLE. Théorème de Poisson.

Soit $(Z_n)_{n\geqslant 1}$ telle que Z_n suit une loi $\mathcal{B}(n, p_n)$ avec $p_n \sim \frac{\lambda}{n}$ pour un certain $\lambda > 0$. Montrons que $Z_n \xrightarrow{\mathcal{L}} Z$ avec $Z \sim \mathcal{P}(\lambda)$.
On a pour tout n et pour tout $t \in \mathbb{R}$:

$$\varphi_{Z_n}(t) = (p_n e^{it} + 1 - p_n)^n = \big(1 + p_n(e^{it} - 1)\big)^n.$$

Comme $p_n \to 0$, on a $|p_n(e^{it} - 1)| < 1$ pour n assez grand. Donc $1 + p_n(e^{it} - 1) \in D(1, 1) \subseteq \mathbb{C}\backslash\mathbb{R}^-$, et donc on peut écrire

$$\varphi_{Z_n}(t) = e^{n\log(1+p_n(e^{it}-1))},$$

où log est la *détermination principale du logarithme*, qui est holomorphe sur $\mathbb{C}\backslash\mathbb{R}^-$. Comme $\log(1 + u) \sim u$ quand $u \to 0$ et comme $p_n \to 0$, on voit que

$$\log(1 + p_n(e^{it} - 1)) \sim p_n(e^{it} - 1) \sim \frac{\lambda}{n}(e^{it} - 1).$$

Donc $n\log(1 + p_n(e^{it} - 1)) \to \lambda(e^{it} - 1)$, et donc $\varphi_{Z_n}(t) \to e^{\lambda(e^{it}-1)}$, fonction caractéristique de la loi de Poisson $\mathcal{P}(\lambda)$. À savoir refaire.

3.5. Fonction caractéristique d'une va à valeurs dans $\mathbb{R}^d$. Si X est une va à valeurs dans $\mathbb{R}^d$, sa fonction caractéristique est la fonction $\varphi_X : \mathbb{R}^d \to \mathbb{C}$ définie par

$$\varphi_X(t) = \mathbb{E}\big(e^{i\langle t, X\rangle}\big) \qquad \text{pour tout } t \in \mathbb{R}^d,$$

où $\langle \cdot, \cdot \rangle$ est le produit scalaire usuel sur $\mathbb{R}^d$.

Autrement dit, si $t = (t_1, \dots, t_d) \in \mathbb{R}^d$ et si on écrit $X = (X_1, \dots, X_d)$, alors

$$\begin{aligned}\varphi_X(t) &= \mathbb{E}\big(e^{i(t_1X_1+\cdots+t_dX_d)}\big)\\ &= \int_\Omega e^{i(t_1X_1+\cdots+t_dX_d)}d\mathbb{P}\\ &= \int_{\mathbb{R}^d} e^{i(t_1x_1+\cdots+t_dx_d)}d\mathbb{P}_{(X_1,\dots,X_d)}(x_1, \dots, x_d).\end{aligned}$$

On *admettra* que les résultats essentiels concernant les fonctions caractéristiques sont encore valables pour des va à valeurs dans $\mathbb{R}^d$. En particuier :

- la fonction caractéristique détermine la loi ;
- une suite (Z_n) de va à valeurs dans $\mathbb{R}^d$ converge en loi vers une va Z si et seulement si $\varphi_{Z_n}(t) \to \varphi_Z(t)$ pour tout $t \in \mathbb{R}^d$.

Comme illustration, on va démontrer le résultat suivant.

PROPOSITION 3.8. *Soient $X_1, \dots, X_d$ des va réelles définies sur le même $(\Omega, \mathfrak{A}, \mathbb{P})$. Alors $X_1, \dots, X_d$ sont indépendantes* si et seulement si

$$(*) \qquad \varphi_{a_1X_1+\cdots+a_dX_d}(s) = \varphi_{X_1}(a_1 s) \cdots \varphi_{X_d}(a_d s)$$

pour tous $a_1, \dots, a_d \in \mathbb{R}$ et pour tout $s \in \mathbb{R}$.

Démonstration. Si $X_1, \dots, X_d$ sont indépendantes, alors $a_1X_1, \dots, a_dX_d$ aussi, donc $\varphi_{a_1X_1+\cdots+a_dX_d} = \varphi_{a_1X_1} \cdots \varphi_{a_dX_d}$, ce qui donne $(*)$.

Inversement, supposons $(*)$ vérifiée. Soit $X = (X_1, \dots, X_d)$, qui est une va à valeurs dans $\mathbb{R}^d$. Par $(*)$, si $t = (t_1, \dots, t_d) \in \mathbb{R}^d$, alors

$$\begin{aligned} \varphi_X(t) &= \mathbb{E}\big(e^{i(t_1X_1+\cdots+t_dX_d)}\big) \\ &= \varphi_{t_1X_1+\cdots+t_dX_d}(1) \\ &= \varphi_{X_1}(t_1) \cdots \varphi_{X_d}(t_d). \end{aligned}$$

Maintenant, soient $X'_1, \dots, X'_d$ des va *indépendantes* et de mêmes lois que $X_1, \dots, X_d$, et soit $X' = (X'_1, \dots, X'_d)$. En utilisant l'indépendance, on voit que si $t = (t_1, \dots, t_d) \in \mathbb{R}^d$, alors

$$\varphi_{X'}(t) = \prod_{k=1}^{d} \varphi_{X'_k}(t_k),$$

et donc $\varphi_{X'} = \varphi_X$ puisque $\varphi_{X'_k} = \varphi_{X_k}$ pour $k = 1, \dots, d$. Comme la fonction caractéristique détermine la loi, on en déduit que $\mathbb{P}_X = \mathbb{P}_{X'}$; autrement dit $\mathbb{P}_X = \mathbb{P}_{X'_1} \otimes \cdots \otimes \mathbb{P}_{X'_d}$ car les X'_k sont indépendantes. Comme $\mathbb{P}_{X'_k} = \mathbb{P}_{X_k}$ pour tout k, on obtient ainsi $\mathbb{P}_{(X_1,\dots,X_d)} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_d}$, ce qui prouve que $X_1, \dots, X_d$ sont indépendantes. □

EXEMPLE. Soient X et Y deux va réelles indépendantes suivant la loi normale $\mathcal{N}(0,1)$. Montrons que les va $X+Y$ et $X-Y$ sont indépendantes.
Par la proposition, il suffit de montrer que

$$\forall a, b, s \in \mathbb{R} \ : \ \varphi_{a(X+Y)+b(X-Y)}(s) = \varphi_{X+Y}(as)\, \varphi_{X-Y}(bs).$$

Il suffit de faire le calcul : on a d'une part

$$\begin{aligned} \varphi_{a(X+Y)+b(X-Y)}(s) &= \varphi_{(a+b)X+(a-b)Y}(s) \\ &= \varphi_{(a+b)X}(s)\, \varphi_{(a-b)Y}(s) \qquad \text{par indépendance de } X \text{ et } Y \\ &= e^{-(a+b)^2 \frac{s^2}{2}}\, e^{-(a-b)^{\frac{s^2}{2}}}; \end{aligned}$$

et d'autre part (toujours par indépendance de X et Y) :

$$\begin{aligned} \varphi_{X+Y}(as)\, \varphi_{X-Y}(bs) &= \varphi_X(as)\varphi_Y(as)\varphi_X(bs)\varphi_{-Y}(bs) \\ &= e^{-\frac{(as)^2}{2}}\, e^{-\frac{(as)^2}{2}}\, e^{-\frac{(bs)^2}{2}}\, e^{-\frac{(bs)^2}{2}}. \end{aligned}$$

Après simplification, on voit (exo) que les deux choses sont bien égales. À savoir refaire.

Chapitre 9

Théorèmes limites

1. Loi des grands nombres

RAPPEL. Si X est une va suivant une loi de Bernoulli $\mathcal{B}(1/2)$ et si $(X_i)_{i\geqslant 1}$ est une suite de va indépendantes et de même loi que X, alors

$$\frac{X_1+\cdots+X_n}{n} \xrightarrow{\text{ps}} \frac{1}{2} = \mathbb{E}(X).$$

Le théorème suivant montre que ceci n'est pas du tout spécifique à la loi $\mathcal{B}(1/2)$, mais est au contraire un phénomène complètement général.

THÉORÈME 1.1. (Loi des grands nombres)
Soit X une va réelle appartenant à L^1, et soit $(X_i)_{i\geqslant 1}$ une suite de va indépendantes et de même loi que X. Pour $n\geqslant 1$, on pose $S_n = \sum_{i=1}^{n} X_i$. Alors

$$\frac{S_n}{n} \xrightarrow{\text{ps}} \mathbb{E}(X).$$

Démonstration. Soit $m = \mathbb{E}(X)$. Quitte à remplacer X par $X-m$ et X_i par $X_i - m$, on peut supposer que $m=0$, *i.e.* que X est centrée. Il s'agit alors de montrer que $\frac{S_n}{n} \xrightarrow{\text{ps}} 0$; ce qu'on va faire en prouvant que

$$\overline{\lim}\,\frac{S_n}{n} \leqslant 0 \quad \text{et} \quad \underline{\lim}\,\frac{S_n}{n} \geqslant 0 \quad \text{ps.}$$

FAIT 1. Soit $Y\in L^1$, et soit $(Y_i)_{i\geqslant 1}$ une suite de va indépendantes et de même loi que Y. Pour $n\geqslant 1$, on pose $\Sigma_n = Y_1+\cdots+Y_n$. Si $\mathbb{E}(Y)<0$, alors $\sup_{n\geqslant 1}\Sigma_n(\omega)<\infty$ presque sûrement.

Preuve du Fait 1. Soit $E = \{\omega\in\Omega;\ \sup_{n\geqslant 1}\Sigma_n(\omega)<\infty\}$. On veut montrer que $\mathbb{P}(E)=1$ (sous l'hypothèse $\mathbb{E}(Y)<0$).

L'évènement E est *asymptotique* relativement à la suite (Y_i) (exo), donc $\mathbb{P}(E)$ vaut 0 ou 1 par la loi du 0-1. Il suffit donc de montrer que $\mathbb{P}(E)>0$. On raisonne par l'absurde en supposant que $\mathbb{P}(E)=0$; autrement dit, qu'on a

$$(*) \qquad \sup_{n\geqslant 1}\Sigma_n(\omega)=\infty \quad \text{presque sûrement.}$$

Pour $n\geqslant 1$, posons

$$M_n = \max(\Sigma_1,\dots,\Sigma_n) = \max(Y_1, Y_1+Y_2,\dots,Y_1+\cdots+Y_n).$$

La suite de va (M_n) est croissante, et $M_n\to\infty$ ps par $(*)$.

Posons maintenant

$$M_n^* = \max(Y_2, Y_2+Y_3,\dots,Y_2+\cdots+Y_{n+1}).$$

Alors M_n^* a la même loi que M_n : en effet, $M_n = \Phi(Y_1, \dots, Y_n)$ où $\Phi(u_1, \dots, u_n) = \max(u_1, u_1 + u_2, \dots, u_1 + \cdots + u_n)$, $M_n^* = \Phi(Y_2, \dots, Y_{n+1})$ pour la même fonction Φ, et les va $(Y_1, \dots, Y_n)$ et $(Y_2, \dots, Y_{n+1})$ ont la même loi, à savoir $\mathbb{P}_Y \otimes \cdots \otimes \mathbb{P}_Y$. En particulier,

$$\mathbb{E}(M_n^*) = \mathbb{E}(M_n).$$

De plus la suite (M_n^*) est croissante, et on vérifie que $M_n^* \to \infty$ ps (exo).

Par définition de M_n^*, on a

$$Y_1 + M_n^* = \max(\Sigma_2, \dots, \Sigma_{n+1});$$

et donc

$$M_{n+1} = \max(\Sigma_1, Y_1 + M_n^*) = \max(Y_1, Y_1 + M_n^*).$$

Autrement dit,

$$\begin{aligned} M_{n+1} &= Y_1 + \max(M_n^*, 0) \\ &= Y_1 + M_n^* - \min(0, M_n^*). \end{aligned}$$

De plus, comme la suite (M_n) est croissante, on a $\mathbb{E}(M_{n+1}) \geqslant \mathbb{E}(M_n)$, ce qui s'écrit $\mathbb{E}(M_n^*) + \mathbb{E}(Y_1) - \mathbb{E}\big(\min(0, M_n^*)\big) \geqslant \mathbb{E}(M_n)$. Comme $\mathbb{E}(M_n^*) = \mathbb{E}(M_n)$, on obtient donc

$$\mathbb{E}(Y_1) \geqslant \mathbb{E}\big(\min(0, M_n^*)\big).$$

Mais si on pose $\Phi_n = \min(0, M_n^*)$, alors la suite (Φ_n) est croissante, et tend presque sûrement vers 0 car $M_n^* \to \infty$ ps. De plus, on a (par croissance) $\Phi_1 \leqslant \Phi_n \leqslant 0$ pour tout $n \geqslant 1$. Comme $\Phi_1 = \min(0, M_1^*) = \min(0, Y_2) \in L^1$, on en déduit, par convergence dominée, que

$$\mathbb{E}\big(\min(0, M_n^*)\big) \to 0.$$

On obtient ainsi $\mathbb{E}(Y_1) \geqslant 0$, ce qui est la contradiction souhaitée. □

Fait 2. Pour tout $\varepsilon > 0$ fixé, on a

$$\sup_{n \geqslant 1} \big(S_n(\omega) - n\varepsilon\big) < \infty \quad \text{presque sûrement.}$$

Preuve du Fait 2. On applique le Fait 1 avec $Y = X - \varepsilon$ et $Y_i = X_i - \varepsilon$. Comme $\mathbb{E}(X) = 0$, on a bien $\mathbb{E}(Y) < 0$. Les Y_i sont indépendantes et de même loi que Y car les X_i sont indépendantes et de même loi que X ; et $\Sigma_n = \sum_{i=1}^n (X_i - \varepsilon) = S_n - n\varepsilon$. Donc le résultat est immédiat. □

Par le Fait 2, on a $\sup_{n \geqslant 1} \big(S_n - n\varepsilon\big) < \infty$ presque sûrement, pour tout $\varepsilon > 0$. En écrivant

$$\frac{S_n}{n} = \frac{1}{n}\big(S_n - n\varepsilon\big) + \varepsilon,$$

on voit donc que pour tout $\varepsilon > 0$ fixé, on a

$$\overline{\lim}\ \frac{S_n}{n} \leqslant \varepsilon \quad \text{presque sûrement.}$$

En prenant $\varepsilon = 1/k$ avec $k \in \mathbb{N}^*$ et se souvenant qu'une conjonction dénombrable de propriétés presque sûres est encore presque sûre, on en déduit que

$$\text{presque sûrement} \quad : \quad \Big(\overline{\lim}\ \frac{S_n}{n} \leqslant \frac{1}{k} \quad \text{pour tout } k \in \mathbb{N}^*\Big).$$

Autrement dit :

$$\overline{\lim}\ \frac{S_n}{n} \leqslant 0 \quad \text{presque sûrement.}$$

En appliquant ceci à la suite $(-X_i)$, on en déduit qu'on a aussi $\overline{\lim}\,\frac{(-S_n)}{n} \leqslant 0$ ps, d'où

$$\underline{\lim}\,\frac{S_n}{n} = -\overline{\lim}\,\frac{(-S_n)}{n} \geqslant 0 \quad \text{ps.}$$

□

Remarque. On peut montrer que l'hypothèse "$X \in L^1$" est *nécessaire* pour avoir la convergence presque sûre de $\frac{S_n}{n}$.

Illustration. Soient f et g deux fonctions continues à valeurs réelles sur un intervalle $[a,b]$, avec $g(x) > 0$ pour tout $x \in [a,b]$. On va utiliser la loi des grands nombres pour montrer que

$$\frac{1}{(b-a)^n}\int_{[a,b]^n} \frac{f(x_1)+\cdots+f(x_n)}{g(x_1)+\cdots+g(x_n)}\,dx_1\cdots dx_n \xrightarrow{n\to\infty} \frac{\int_a^b f(x)\,dx}{\int_a^b g(x)\,dx}.$$

Soit X une variable aléatoire uniformément distribuée sur $[a,b]$, et soit $(X_i)_{i\geqslant 1}$ une suite de va indépendantes et de même loi que X, définies sur $(\Omega,\mathfrak{A},\mathbb{P})$. Pour $n \geqslant 1$, posons

$$Z_n := \frac{f(X_1)+\cdots+f(X_n)}{g(X_1)+\cdots+g(X_n)}.$$

Les va $f(X_i)$ sont indépendantes et de même loi que $f(X)$, qui est bornée (car f est bornée sur $[a,b]$), donc dans L^1. De même, les $g(X_i)$ sont indépendantes et de même loi que $g(X)$, qui est dans L^1 et vérifie $\mathbb{E}\big(g(X)\big) > 0$ car $g(x) > 0$ sur $[a,b]$. Par la loi des grands nombres, on voit donc que

$$Z_n := \frac{f(X_1)+\cdots+f(X_n)}{n} \times \frac{n}{g(X_1)+\cdots+g(X_n)} \xrightarrow{\text{ps}} \frac{\mathbb{E}\big(f(X)\big)}{\mathbb{E}\big(g(X)\big)}.$$

De plus, f est bornée sur $[a,b]$, et comme g est continue et > 0 sur $[a,b]$, il existe une constante $c > 0$ telle que $g(x) \geqslant c$ pour tout $x \in [a,b]$. Pour $\omega \in \Omega$, on a donc

$$\forall n \geqslant 1 \,:\, |Z_n(\omega)| \leqslant \frac{n \times \|f\|_\infty}{n \times c} = \frac{\|f\|_\infty}{c} =: M.$$

La constante M étant dans L^1, on en déduit, par convergence dominée, que

$$\mathbb{E}(Z_n) \to \mathbb{E}\left(\frac{\mathbb{E}(f(X))}{\mathbb{E}\big(g(X)\big)}\right) = \frac{\mathbb{E}(f(X))}{\mathbb{E}\big(g(X)\big)}.$$

Mais par le Théorème de transfert, l'indépendance des X_i et le fait que $\mathbb{P}_{X_i} = \frac{1}{b-a}\,\mathbf{1}_{[a,b]}(x)dx$, on a aussi (micro-exo)

$$\mathbb{E}(Z_n) = \frac{1}{(b-a)^n}\int_{[a,b]^n} \frac{f(x_1)+\cdots+f(x_n)}{g(x_1)+\cdots+g(x_n)}\,dx_1\cdots dx_n,$$

et de même

$$\mathbb{E}\big(f(X)\big) = \frac{1}{b-a}\int_a^b f(x)\,dx \qquad \text{et} \qquad \mathbb{E}\big(g(X)\big) = \frac{1}{b-a}\int_a^b g(x)\,dx.$$

Donc on obtient le résultat annoncé.

2. Loi des grands nombres L^2

DÉFINITION 2.1. *Soient X et Y deux va réelles définies sur $(\Omega, \mathfrak{A}, \mathbb{P})$ et appartenant à L^2. La **covariance** de X et Y est le nombre $\mathrm{Cov}(X,Y)$ défini par*

$$\mathrm{Cov}(X,Y) := \langle X - \mathbb{E}(X), Y - \mathbb{E}(Y)\rangle_{L^2} = \mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)\big).$$

REMARQUE 1. On a $\mathrm{Cov}(X,X) = \sigma^2(X)$ pour toute va $X \in L^2$.

REMARQUE 2. On dit que X et Y sont **non corrélées** si on a $\mathrm{Cov}(X,Y) = 0$.

REMARQUE 3. Si X et Y sont indépendantes, alors elles sont non corrélées.

Démonstration. Les va $X - \mathbb{E}(X)$ et $Y - \mathbb{E}(Y)$ sont indépendantes et centrées, donc orthogonales dans L^2. □

PROPOSITION 2.2. *Soit $(X_i)_{i\geqslant 1}$ une suite de va réelles définies sur $(\Omega, \mathfrak{A}, \mathbb{P})$ et appartenant à L^2. Pour $n \geqslant 1$, on pose*

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

On fait les hypothèses suivantes :

(i) $\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(X_i)$ *admet une limite m quand $n \to \infty$;*

(ii) $\sum_{i,j=1}^{n} \mathrm{Cov}(X_i, X_j) = o(n^2)$ *quand $n \to \infty$.*

Alors $\frac{S_n}{n} \to m$ en norme L^2, et donc en probabilité.

Démonstration. Par (i), il suffit de montrer que

$$\frac{S_n}{n} - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(X_i) \xrightarrow{L^2} 0.$$

On a

$$\begin{aligned}\left\| \frac{S_n}{n} - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(X_i) \right\|_{L^2}^2 &= \frac{1}{n^2} \left\| \sum_{i=1}^{n} \big(X_i - \mathbb{E}(X_i)\big) \right\|_{L^2}^2 \\ &= \frac{1}{n^2} \sum_{i,j=1}^{n} \big\langle X_i - \mathbb{E}(X_i), X_j - \mathbb{E}(X_j)\big\rangle_{L^2} \\ &= \frac{1}{n^2} \sum_{i,j=1}^{n} \mathrm{Cov}(X_i, X_j);\end{aligned}$$

Donc le résultat est immédiat par (ii) ! □

COROLLAIRE 2.3. *Si (X_i) est une suite de va appartenant à L^2,* deux à deux non corrélées, *de même moyenne m et de même variance σ^2, alors $\frac{S_n}{n} \to m$ en norme L^2 et donc en probabilité.*

Démonstration. La condition (i) est évidemment satisfaite. De plus, comme par hypothèse $\mathrm{Cov}(X_i, X_j) = 0$ si $i \neq j$ et $\mathrm{Cov}(X_i, X_i) = \sigma^2(X_i) = \sigma^2$ on a

$$\sum_{i,j=1}^{n} \mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{n} \sigma^2(X_i) = n\sigma^2 \quad \text{pour tout } n \geqslant 1;$$

donc (ii) est également satisfaite □

3. Théorème limite central

REMARQUE 3.1. Soient $X_1, \dots, X_n$ des va réelles, et posons $S_n = X_1 + \cdots + X_n$. Si les X_i sont indépendantes et suivent des lois normales $\mathcal{N}(m_i, \sigma_i^2)$, alors la va

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\sigma(S_n)}$$

suit la loi $\mathcal{N}(0,1)$. En particulier, si les X_i sont indépendantes et suivent toutes une même loi $\mathcal{N}(m, \sigma^2)$, alors $Z_n = \frac{S_n - nm}{\sqrt{n}\sigma}$ suit la loi $\mathcal{N}(0,1)$.

Démonstration. On sait que S_n suit la loi $\mathcal{N}(m_0, \sigma_0^2)$, où $m_0 = m_1 + \cdots + m_n$ et $\sigma_0^2 = \sigma_1^2 + \cdots + \sigma_n^2 = \sigma^2(S_n)$. Donc (*cf* un exo déjà posé au Chapitre 3) $Y_n = S_n - \mathbb{E}(S_n) = S_n - m_0$ suit la loi $\mathcal{N}(0, \sigma_0^2)$ et $Z_n = \frac{Y_n}{\sigma_0}$ suit la loi $\mathcal{N}(0, \frac{\sigma_0^2}{\sigma_0^2}) = \mathcal{N}(0,1)$. Dans le cas particulier où les X_i suivent toutes la loi $\mathcal{N}(m, \sigma)$, on a $\mathbb{E}(S_n) = nm$ et $\sigma^2(S_n) = n\sigma^2$, donc $Z_n = \frac{S_n - nm}{\sqrt{n}\sigma}$. □

THÉORÈME 3.2. *Soit X une va réelle appartenant à L^2, de moyenne $\mathbb{E}(X) = m$ et de variance $\sigma^2(X) = \sigma^2$. Soit également (X_i) une suite va va indépendantes et de même loi que X. Pour $n \geqslant 1$, on pose*

$$S_n = \sum_{i=1}^{n} X_i \qquad \text{et} \qquad Z_n = \frac{S_n - nm}{\sqrt{n}\,\sigma}.$$

Alors la suite (Z_n) converge en loi vers une va Z suivant la loi normale $\mathcal{N}(0,1)$. Par conséquent, pour tout intervalle $I \subseteq \mathbb{R}$,

$$\mathbb{P}\left(\frac{S_n - nm}{\sqrt{n}\,\sigma} \in I\right) \xrightarrow{n\to\infty} \frac{1}{\sqrt{2\pi}} \int_I e^{-\frac{x^2}{2}}\, dx.$$

REMARQUE 1. La deuxième partie du théorème découle de la première car la loi normale est à densité et donc $\mathbb{P}(Z \in \partial I) = 0$ (*cf* le Corollaire 4.10 du Chapitre 7).

REMARQUE 2. Z_n est la "version centrée et normalisée dans L^2 de S_n". En effet, on a $Z_n = \frac{1}{\sqrt{n}\sigma} \sum_{i=1}^{n} (X_i - \mathbb{E}(X_i))$, donc

$$\mathbb{E}(Z_n) = 0 \qquad \text{et} \qquad \sigma^2(Z_n) = \frac{1}{n\sigma^2} \sum_{i=1}^{n} \sigma^2(X_i) = 1.$$

REMARQUE 3. Comme $\frac{S_n}{n} - m = \frac{1}{\sqrt{n}} \frac{S_n - nm}{\sqrt{n}}$ et comme "$\frac{S_n - nm}{\sqrt{n}}$ a une limite" d'après le TLC, on voit que moralement, $\frac{S_n}{n} - m$ "tend vers 0 comme $\frac{1}{\sqrt{n}}$" (précision par rapport à la loi des grands nombres)

Preuve du théorème. Posons

$$Z_n = \frac{S_n - nm}{\sqrt{n}\sigma} = \frac{X_1 + \cdots + X_n - nm}{\sqrt{n}\sigma}.$$

On veut montrer que $Z_n \xrightarrow{\mathcal{L}} Z$, où $Z \sim \mathcal{N}(0,1)$; et pour cela il suffit de prouver que

$$\varphi_{Z_n}(t) \to \varphi_Z(t) = e^{-\frac{t^2}{2}} \qquad \text{pour tout } t \in \mathbb{R}.$$

Il n'y a rien à démontrer si $t = 0$ puisque $\varphi_{Z_n}(0) = 1 = \varphi_Z(0)$ pour tout n ; donc on supposera dans la suite que $t \neq 0$.

Quitte à remplacer X par $X' = \frac{X-m}{\sigma}$ et X_k par $X'_k = \frac{X_k-m}{\sigma}$, on peut supposer que

$$\mathbb{E}(X) = 0 \qquad \text{et} \qquad \mathbb{E}(X^2) = 1;$$

de sorte que

$$Z_n = \frac{S_n}{\sqrt{n}} = \frac{X_1 + \cdots + X_n}{\sqrt{n}}\cdot$$

Comme les X_k sont indépendantes, on a

$$\varphi_{Z_n}(t) = \mathbb{E}\Big(\prod_{k=1}^{n} e^{it\frac{X_k}{\sqrt{n}}}\Big) = \prod_{k=1}^{n} \mathbb{E}\big(e^{i\frac{t}{\sqrt{n}}X_k}\big);$$

et comme les X_k ont la même loi que X, on en déduit

$$\varphi_{Z_n}(t) = \varphi_X\Big(\frac{t}{\sqrt{n}}\Big)^n.$$

De plus, φ_X est de classe $\mathcal{C}^2$ sur $\mathbb{R}$ car $X \in L^2$, avec $\varphi_X(0) = 1$, $\varphi'_X(0) = \mathbb{E}(X) = 0$ et $\varphi''_X(0) = i^2\mathbb{E}(X^2) = -1$. Donc φ_X admet au voisinage de 0 le développement limité suivant :

$$\varphi_X(u) = 1 - \frac{u^2}{2} + o(u^2).$$

On en déduit que pour tout $t \in \mathbb{R}$, on peut écrire

$$\varphi_X\Big(\frac{t}{\sqrt{n}}\Big) = 1 - \frac{t^2}{2n} + \xi_n, \qquad \text{où} \quad |\xi_n| = o\big(\tfrac{1}{n}\big).$$

Donc $\Big|\varphi_X\big(\frac{t}{\sqrt{n}}\big) - 1\Big| < 1$ pour n assez grand, et on peut écrire

$$\varphi_{Z_n}(t) = \left(1 - \frac{t^2}{2n} + \xi_n\right)^n = \exp\left(n\log\Big(1 - \frac{t^2}{2n} + \xi_n\Big)\right),$$

où log est la *détermination principale du logarithme*, qui est bien définie et holomorphe dans le disque $D(0,1)$. Comme $\log'(1) = 1$, on a $\log(1+u) \sim u$ quand $u \to 0$; et comme $t \neq 0$, on en déduit que $\log\Big(1 - \frac{t^2}{2n} + \xi_n\Big) \sim -\frac{t^2}{2n} + \xi_n \sim -\frac{t^2}{2n}$ quand $n \to \infty$, ce qui donne le résultat souhaité :

$$\varphi_{Z_n}(t) \xrightarrow{n\to\infty} e^{-\frac{t^2}{2}}.$$

□

Remarque. On peut éviter le recours au logarithme complexe en utilisant le fait suivant : si $u, v \in \mathbb{C}$ vérifient $|u|, |v| \leqslant 1$ et si $n \in \mathbb{N}^*$, alors

$$(*) \qquad |u^n - v^n| \leqslant n\,|u - v|;$$

ce qui se démontre en écrivant $u^n - v^n = (u - v)(u^{n-1} + u^{n-2}v + \cdots + uv^{n-2} + v^{n-1})$. En appliquant $(*)$ avec $u = \varphi_X(\frac{t}{\sqrt{n}})$ et $v = e^{-\frac{t^2}{2n}}$, on obtient

$$\big|\varphi_{Z_n}(t) - e^{-\frac{t^2}{2}}\big| \leqslant n\left|\varphi_X\Big(\frac{t}{\sqrt{n}}\Big) - e^{-\frac{t^2}{2n}}\right|;$$

et on conclut en utilisant les développements limités (à l'ordre 2) de φ_X et de l'exponentielle en 0.

COROLLAIRE 3.3. *Avec les notations du théorème, la suite* $\frac{S_n - nm}{\sqrt{n}}$ *converge en loi vers une va suivant la loi* $\mathcal{N}(0, \sigma^2)$.

Démonstration. On a $\frac{S_n - nm}{\sqrt{n}} = \sigma Z_n$, donc $\frac{S_n - nm}{\sqrt{n}} \xrightarrow{\mathcal{L}} \sigma Z$ (exo déjà posé) ; et $\sigma Z \sim \mathcal{N}(0, \sigma^2)$ car $Z \sim \mathcal{N}(0, 1)$. □

COROLLAIRE 3.4. *Avec les notations du théorème,* $\mathbb{E}(Z_n^+) \to \frac{1}{\sqrt{2\pi}}$ *quand* $n \to \infty$.

Démonstration. On rappelle que si $x \in \mathbb{R}$, alors $x^+ = \max('x, 0)$. Par la formule d'intégration par parties, on a

$$\mathbb{E}(Z_n^+) = \int_0^\infty \mathbb{P}(Z_N^+ > t)\, dt = \int_0^\infty \mathbb{P}(Z_n > t)\, dt,$$

et de même

$$\mathbb{E}(Z^+) = \int_0^\infty \mathbb{P}(Z > t)\, dt.$$

Cela étant, $\mathbb{P}(Z_n > t) = 1 - F_{Z_n}(t) \to 1 - F_Z(t) = \mathbb{P}(Z > t)$ pour tout $t \in \mathbb{R}$ par le TLC. De plus par l'inégalité de Tchebitchev et comme les Z_n sont centrées, on a

$$\forall n\ \forall t > 1\ :\ \mathbb{P}(Z_n > t) \leqslant \frac{\mathbb{E}(Z_n^2)}{t^2} = \frac{1}{t^2};$$

et aussi

$$\forall n\ \forall t \in [0, 1]\ :\ \mathbb{P}(Z_n > t) \leqslant 1.$$

Par convergence dominée, on en déduit que $\mathbb{E}(Z_n^+) \to \mathbb{E}(Z^+)$ quand $n \to \infty$. Enfin, $\mathbb{E}(Z^+)$ se calcule très bien :

$$\begin{aligned}
\mathbb{E}(Z^+) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^+ e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_0^\infty x\, e^{-\frac{x^2}{2}} dx \\
&= \left[-\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right] = \frac{1}{\sqrt{2\pi}}.
\end{aligned}$$

□

REMARQUE 3.5. Tel qu'il est énoncé le TLC ne donne pas de précision sur la vitesse de convergence de $\mathbb{P}(Z_n \in I)$ vers $\mathbb{P}(Z \in I)$ quand I est un intervalle de $\mathbb{R}$. Il se trouve qu'il est possible d'estimer cette convergence *si on suppose que* $X \in L^3$. Sous cette hypothèse, pour tout intervalle $I \subseteq \mathbb{R}$ et pour tout $n \in \mathbb{N}^*$, on a

$$|\mathbb{P}(Z_n \in I) - \mathbb{P}(Z \in I)| \leqslant \frac{C(X)}{\sqrt{n}},$$

où

$$C(X) = \frac{\mathbb{E}\big(|X - m|^3\big)}{2\sigma^3}.$$

Ce résultat s'appelle le **Théorème de Berry-Esseen**. Il montre en particulier que la convergence de $\mathbb{P}(Z_n \in I)$ vers $\mathbb{P}(Z \in I)$ est *uniforme* par rapport à l'intervalle I. La preuve est un peu trop délicate pour être faite ici.

ILLUSTRATION. On va utiliser le Théorème Limite Central pour démontrer la **Formule de Stirling** :

$$n! \sim n^n e^{-n}\sqrt{2\pi n} \qquad \text{quand } n \to \infty.$$

Soit $(X_i)_{i\geqslant 1}$ une suite de variables indépendantes suivant toutes la loi de Poisson de paramètre $\lambda := 1$. Pour $n \geqslant 1$, posons comme d'habitude $S_n := X_1 + \cdots + X_n$, et

$$Z_n := \frac{S_n - n}{\sqrt{n}}\cdot$$

Comme l'espérance et la variance d'une v.a. suivant la loi $\mathcal{P}(1)$ sont égales à 1, le Corollaire 3.4 nous dit que

$$\mathbb{E}(Z_n^+) \to \frac{1}{\sqrt{2\pi}}\cdot$$

Calculons alors $\mathbb{E}(Z_n^+)$. On sait que S_n suit la loi de Poisson $\mathcal{P}(n)$; et on a $Z_n^+ = \varphi(S_n)$, où $\varphi(k) := \frac{(k-n)^+}{\sqrt{n}}\cdot$ Donc

$$\begin{aligned}
\mathbb{E}(Z_n^+) &= \sum_{k=0}^{\infty} \frac{(k-n)^+}{\sqrt{n}}\, \frac{n^k}{k!}\, e^{-n} \\
&= \frac{e^{-n}}{\sqrt{n}} \sum_{k=n}^{\infty} (k-n)\, \frac{n^k}{k!} \\
&= \frac{e^{-n}}{\sqrt{n}} \left(\sum_{k=n}^{\infty} \frac{n^k}{(k-1)!} - \sum_{k=n}^{\infty} \frac{n^{k+1}}{k!} \right) \\
&= \frac{e^{-n}}{\sqrt{n}}\, \frac{n^n}{(n-1)!}\cdot
\end{aligned}$$

Ainsi, on peut dire que

$$\frac{e^{-n}}{\sqrt{n}}\, \frac{n^n}{(n-1)!} \to \frac{1}{\sqrt{2\pi}} \qquad \text{quand } n \to \infty.$$

Autrement dit : $(n-1)! \sim \sqrt{2\pi} \times \frac{n^n e^{-n}}{\sqrt{n}}$; d'où la Formule de Stirling en multipliant tout par n.