Probabilités

Licence de Mathématiques, 3ème année

Table des matières

Chapitre 1

Bla-bla introductif

1. Ce dont il s'agit

De façon très schématique, on peut dire que le but de la théorie des probabilités est d'étudier mathématiquement des situations réelles ou fictives "où intervient le hasard" ; autrement dit : où on ne sait pas prévoir *avec certitude* ce qui va se passer.

Étant donné une telle situation à étudier (et toujours de façon très schématique) la démarche générale est la suivante.

(i) On essaye de **modéliser** la situation, *i.e.* de trouver un cadre mathématique qui la décrit de façon à peu près satisfaisante.

(ii) On fait des mathématiques dans le cadre du modèle choisi.

En réalité, dans ce cours on va surtout s'occuper de l'aspect (ii), qui est le plus simple car il ne s'agit que de mathématiques. Mais il est quand même intéressant de parler un peu de l'aspect "modélisation" (i) pour tenter justifier le cadre mathématique dans lequel on va se placer.

L'exemple suivant, volontairement très simple (et sans intérêt), sera le "fil directeur" de ce bla-bla introductif.

Exemple guide. On jette deux dés, et on veut savoir s'il y a plus de chances que la somme des chiffres fasse 7 ou qu'elle fasse 8.

2. Vocabulaire intuitif

2.1. Expériences aléatoires et évènements.

• *Expérience aléatoire* : n'importe quelle "expérience" (qu'on peut soit réaliser effectivement, soit imaginer) 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 préférée.

• *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 1. Le mot "évènement" est à prendre dans le sens le plus large possible : un évènement est "n'importe quoi qui est susceptible de se produire (ou pas)".

Remarque 2. 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 savez quelle bière vous avez préférée (et si vous êtes capable d'identifier 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 $E_1, \dots, E_n$ 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. Un exemple important : 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.

Fait. *S'il n'y a qu'un nombre fini de résultats possibles*, alors tout évènement E est équivalent à une disjonction d'évènements élémentaires.

Démonstration. C'est évident : E est par définition équivalent à $E_{\omega_1} \vee \cdots \vee E_{\omega_N}$, où $\omega_1, \dots, \omega_N$ sont les résultats de l'expérience pour lesquels E est réalisé. □

Remarque. Si E est équivalent à $E_{\omega_1} \vee \cdots \vee E_{\omega_N}$ avec des ω_i tous différents, alors les ω_i sont nécessairement ceux introduits dans la preuve du Fait (micro-exo). On dit que les évènements élémentaires E_{ω_i} "composent E".

Exemple guide. (On jette 2 dés.)

- Le résultat est : "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 $E_{i,j}$ et $E_{j,i}$ désignent le même évènement.)
- Les évènements qui nous intéressent : “la somme des chiffres fait 8” (noté S_8) ; “la somme des chiffres fait 7” (noté S_7). Visiblement, S_7 est équivalent à $E_{1,6} \vee E_{2,5} \vee E_{3,4}$ et S_8 est équivalent à $E_{2,6} \vee E_{3,5} \vee E_{4,4}$.

2.2. 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 :

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

Remarques. Ce n'est pas une définition précise, même si “tout le monde” a une intuition de ce que cela peut signifier. De plus, la probabilité dépend fortement de l'“observateur” : deux personnes différentes peuvent très bien ne pas être du tout d'accord sur le degré de plausibilité d'un évènement. Enfin, il n'est absolument pas clair qu'un observateur soit en mesure d'attribuer de façon exacte un degré de plausibilité à n'importe quel évènement.

Quoi qu'il en soit, on considèrera que les faits suivants sont “intuitivement raisonnables”.

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*, la probabilité d'un évènement E est la somme des probabilités des évènements élémentaires qui le composent. Autrement dit : si E est équivalent à une disjonction d'évènements élémentaires $E_{\omega_1} \vee \cdots \vee E_{\omega_N}$ avec des ω_i tous différents, alors $\mathrm{Prob}(E) = \mathrm{Prob}(E_{\omega_1}) + \cdots + \mathrm{Prob}(E_{\omega_N})$. En particulier, en prenant un évènement E certain, on voit que *la somme des probabilités de* tous *les évènements élémentaires doit être égale à* 1.

Exemple guide. (On jette 2 dés.)
- Il est naturel de considérer qu'il y a 2 fois plus de chances d'obtenir un résultat mixte (deux chiffres différents) plutôt qu'un “double”, que tous les “doubles” ont la même probabilité p, et que tous les résultats mixtes ont la même probabilité $q = 2p$. Comme il y a 6 doubles possibles et 15 résultats mixtes possibles (Exo ; énumérer...), et que la somme de toutes les probabilités des évènements élémentaires doit faire 1, on a donc $6 \times p + 15 \times 2p = 1$, d'où $p = \frac{1}{36}\cdot$ Ainsi $\mathrm{Prob}(E_{i,i}) = \frac{1}{36}$ et $\mathrm{Prob}(E_{i,j}) = \frac{1}{18}$ si $i \neq j$.
- La question qui nous intéresse : d'après ce qui précède on a $\mathrm{Prob}(S_7) = \mathrm{Prob}(E_{1,6}) + \mathrm{Prob}(E_{2,5}) + \mathrm{Prob}(E_{3,4}) = 3 \times \frac{1}{18} = \frac{1}{6}$, et $\mathrm{Prob}(S_8) = \mathrm{Prob}(E_{2,6}) + \mathrm{Prob}(E_{3,5}) + \mathrm{Prob}(E_{4,4}) = 2 \times \frac{1}{18} + \frac{1}{36} = \frac{5}{36}\cdot$ Conclusion : il y plus de chances d'obtenir un 7 qu'un 8.

L'exemple guide étant particulièrement simple, on a pu le traiter de manière "semi-mathématique", *i.e.* en raisonnant de façon logique mais sans fixer un cadre mathématique précis. Dans des situations plus compliquées, on ne s'en sortirait pas aussi facilement sans "mathématisation" précise.

3. "Modélisation"

On considère une certaine expérience aléatoire $\boldsymbol{\mathcal{E}}$, que l'on cherche à modéliser. 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.

Exemple guide. On peut identifier un résultat possible "i et j" avec la paire $(i,j) \in [\![1,6]\!] \times [\![1,6]\!]$. Mais il faut tenir compte du fait que "i et j" et "j et i" sont le même résultat, donc ne 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.* ne prendre que les (i,j) tels que $i \leqslant j$. Bref : on peut prendre $\Omega = \big\{(i,j) \in [\![1,6]\!] \times [\![1,6]\!];\ i \leqslant j\big\}$. (Mais on n'est pas obligé : par exemple, on pourrait aussi prendre pour Ω l'ensemble des parties de $[\![1,6]\!]$ possèdant 1 ou 2 éléments.)

• 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" ?

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

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

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

Cette définition a un sens, *i.e.* il n'y a pas d'ambiguité : si $\widetilde{E} = \widetilde{E'}$, alors les deux évènements E et E' sont équivalents et donc $\mathrm{Prob}(E) = \mathrm{Prob}(E')$.

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

Cas particulier. On suppose qu'il n'y a qu'un nombre fini de résultats possibles, et donc que l'ensemble Ω est fini. Alors $\mathfrak{A} = \mathcal{P}(\Omega)$ car $\mathfrak{A}$ contient tous les singletons et est stable par réunions finies. Et si $A \subseteq \Omega$, alors

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \sum_{\omega \in A} \mathrm{Prob}(E_\omega).$$

En effet : si on écrit $A = \{\omega_1, \dots, \omega_N\}$ où les ω_i sont deux à deux différents, alors A est la réunion disjointe des singletons $\{\omega_i\}$, et donc $\mathbb{P}(A) = \sum_{i=1}^N \mathbb{P}(\{\omega_i\}) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$.

Cas très particulier. On suppose qu'il n'y a qu'un nombre fini de résultats possibles et qu'ils sont tous **équiprobables**, *i.e.* ils ont tous la même probabilité. Alors

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

car $\mathbb{P}(\Omega) = 1 = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\})$. Donc, pour $A \subseteq \Omega$, on a

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega}.$$

Autrement dit, pour tout évènement E, on a

$$\mathrm{Prob}(E) = \frac{\text{“nombre de cas favorables”}}{\text{“nombre de cas possibles”}}.$$

Exemple guide. (On jette 2 dés.)

On a pris $\Omega := \{(i,j) \in [\![1,6]\!] \times [\![1,6]\!];\ i \leqslant j\}$. On a vu que

$$\mathbb{P}(\{(i,i)\}) = \frac{1}{36} \quad \text{pour tout } i \qquad \text{et} \qquad \mathbb{P}(\{(i,j)\}) = \frac{1}{18} \quad \text{pour } i < j;$$

ce qui permet de calculer $\mathbb{P}(A)$ pour tout $A \subseteq \Omega$. Ainsi

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

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

Remarque. Il est parfois avantageux de remplacer l'expérience par une expérience “équivalente”, pour que la modélisation soit plus simple.

Exemple guide. Pour ce qui nous intéresse, il reviendrait au même de lancer les dés l'un après l'autre. L'avantage est que de cette façon, on a *différencié* les deux dés : on peut donc 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". Mais maintenant, tous les évènements élémentaires deviennent équiprobables ; donc $\mathbb{P}(\{(i,j)\}) = \frac{1}{36}$ pour tout $(i,j) \in \Omega$. Du coup, le calcul de $\mathrm{Prob}(S_7)$ devient

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

et de même

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

4. Hypothèses supplémentaires

Pour que le modèle $(\Omega, \mathfrak{A}, \mathbb{P})$ soit mathématiquement plus "performant", on va en fait en 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 $(\Omega, \mathfrak{A})$.

Lorsque l'ensemble Ω est fini, ces hypothèses sont automatiquement satisfaites, comme le montre le fait suivant.

Fait 4.1. *Si Ω un ensemble fini, alors toute algèbre $\mathfrak{A}$ de parties de Ω est en fait une tribu, et toute application finiment additive $\mu : \mathfrak{A} \to [0,\infty]$ telle que $\mu(\varnothing) = 0$ est en fait une mesure.*

Démonstration. Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathfrak{A}$. Si on pose $B_0 := A_0$ et $B_n := A_n \backslash (A_0 \cup \cdots \cup A_{n-1})$ pour $n \geqslant 1$, alors les B_n appartiennent à $\mathfrak{A}$ puisque $\mathfrak{A}$ est une algèbre, et on a $B_0 \cup \cdots \cup B_N = A_0 \cup \cdots \cup A_N$ pour tout $N \in \mathbb{N}$, donc $\bigcup_{n=0}^\infty A_n = \bigcup_{n=0}^\infty B_n$. De plus, les B_n sont deux à deux disjoints par définition ; et comme Ω est supposé fini, on en déduit qu'il n'y a qu'un nombre fini de B_n non vides ; ainsi $B_n = \varnothing$ à partir d'un certain rang N. Donc $\bigcup_{n=0}^\infty A_n = \bigcup_{n=0}^N B_n$, et donc $\bigcup_{n=0}^\infty A_n \in \mathfrak{A}$. On a ainsi montré que $\mathfrak{A}$ est stable par réunions dénombrables ; donc $\mathfrak{A}$ est une tribu puisque $\varnothing \in \mathfrak{A}$ et $\mathfrak{A}$ est stable par complémentation.

Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'éléments de $\mathfrak{A}$ deux à deux disjoints, alors $A_n = \varnothing$ à partir d'un certain rang N car Ω est fini, et donc $\bigcup_{n=0}^\infty A_n = \bigcup_{n=0}^N A_n$. Comme μ est finiment additive et $\mu(\varnothing) = 0$, on en déduit $\mu(\bigcup_{n=0}^\infty A_n) = \mu(\bigcup_{n=0}^N A_n) = \sum_{n=0}^N \mu(A_n) = \sum_{n=0}^\infty \mu(A_n)$; donc μ est une mesure. □

Lorsque Ω est infini, les algèbres de parties de Ω n'ont aucune raison d'être des tribus, et l'additivité dénombrable ne découle plus de l'additivité finie. On impose donc des conditions vraiment plus fortes.

Remarque 1. Il est difficile de justifier "intuitivement" l'hypothèse d'additivité dénombrable. En fait, il y a même des situations où elle est franchement contre-intuitive. Par exemple, il est intuitivement clair que si on choisit un entier "au hasard", alors la probabilité que cet entier soit égal à un certain entier a fixé à l'avance est égale à 0. Cependant, comme $\mathbb{N}$ est dénombrable, il n'existe pas de mesure $\mathbb{P}$ sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ telle que $\mathbb{P}(\mathbb{N}) = 1$ et $\mathbb{P}(\{a\}) = 0$ pour tout $a \in \mathbb{N}$ (exo).

Remarque 2. Si on y réfléchit trop, il peut même sembler délicat de justifier le fait que la famille $\mathfrak{A}$ soit stable par réunions dénombrables. En effet, si $(E_i)_{i\in\mathbb{N}}$ est une suite infinie d'évènements relatifs à une certaine expérience et si on note $E = \bigvee_{i\in\mathbb{N}} E_i$ la "disjonction" des E_i (*i.e.* E se produit si et seulement si l'un des E_i se produit), alors on peut vérifier "en temps fini" que E se produit *si c'est effectivement le cas* (il suffit de considérer les E_i l'un après l'autre et de s'arrêter au premier i tel que E_i se produit), mais il est *a priori* douteux qu'on puisse vérifier en temps fini que E ne se produit pas si en effet il ne se produit pas. Donc on pourrait légitimement rechigner à considérer E comme un évènement relatif à l'expérience ; autrement dit, on pourrait estimer douteux que $\widetilde{E} = \bigcup_{i\in\mathbb{N}} \widetilde{E_i}$ appartienne à la famille $\mathfrak{A}$.

Remarque 3. Le même type d'argument pourrait conduire à remettre en cause le fait que la famille $\mathfrak{A}$ contienne tous les singletons. Imaginons par exemple qu'on choisisse un nombre réel au hasard, donc qu'on prenne $\Omega := \mathbb{R}$ et $\mathfrak{A} \subseteq \mathcal{P}(\mathbb{R})$. Alors, si $a \in \mathbb{R}$ est donné, il n'est pas clair du tout qu'on puisse vérifier en temps fini que le nombre réel ω choisi est égal à a, car on n'a *a priori* pas accès en temps fini à toutes les décimales de ω et de a. Donc, il n'est pas clair que le singleton $\{a\}$ apartienne à $\mathfrak{A}$.

La morale de ces élucubrations est peut-être qu'il ne faut pas se poser trop de questions. Comme on le verra, le "modèle" $(\Omega, \mathfrak{A}, \mathbb{P})$ permet de faire des mathématiques intéressantes et d'obtenir sans trop de difficultés des résultats parfois spectaculaires ; et on s'en tiendra là...

Exercice. Soit $\Omega := \mathbb{N}$. On note $\mathfrak{A}$ la famille de toutes les parties A de $\mathbb{N}$ vérifiant la propriété suivante : ou bien A est un ensemble fini, ou bien A^c est fini. Pour $A \in \mathfrak{A}$, on pose $\mathbb{P}(A) := 0$ si A est fini, et $\mathbb{P}(A) := 1$ si A^c est fini.

(1) Montrer que $\mathfrak{A}$ est une algèbre de parties de $\mathbb{N}$, et que l'application $\mathbb{P}$ est finiment additive.
(2) Montrer que $\mathfrak{A}$ n'est pas une tribu, et que $\mathbb{P}$ n'est pas dénombrablement additive.

Chapitre 2

Espaces de probabilité

1. Généralités

1.1. Lois de probabilité. Ce qui a été dit dans le bla-bla introductif conduit aux définitions suivantes.

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. On n'est pas du tout obligé d'imposer que la tribu $\mathfrak{A}$ contienne tous les singletons : il s'agit plutôt d'une commodité "technique".

Remarque 1. Si $(\Omega, \mathfrak{A})$ est un univers probabilisable, alors la tribu $\mathfrak{A}$ contient tous les ensembles *dénombrables*.

Remarque 2. Lorsque 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. Il faut garder en tête que si E est un évènement, alors

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

Remarque 4. Il faut également se souvenir de la **continuité monotone** de la loi de probabilité $\mathbb{P}$:

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

et, comme $\mathbb{P}$ est une mesure finie,

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

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 $\mathbb{P}(E) = 1$; ou, de mani ere quivalente, si $\mathbb{P}(E^c) = 0$. 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. La plupart du temps, 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 $\int_\Omega |f|\, d\mathbb{P} \leqslant \|f\|_\infty \times \mathbb{P}(\Omega) < \infty$. □

Remarque. Il ne faut pas oublier non plus les deux faits suivants, qui sont très souvent utiles.

• Si $f : \Omega \to \mathbb{R}$ est une fonction mesurable positive telle que $\int_\Omega f\, d\mathbb{P} < \infty$, alors $f(\omega) < \infty$ presque sûrement.

• Si $f : \Omega \to \mathbb{R}$ est une fonction mesurable positive telle que $\int_\Omega f\, d\mathbb{P} = 0$, alors $f(\omega) = 0$ presque sûrement.

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 $\mathcal{M}$ est stable par* réunions dénombrables croissantes *et par* différences propres, *et si de plus $\Omega \in \mathcal{M}$. Autrement dit, $\mathcal{M}$ est une classe monotone si*

- $\Omega \in \mathcal{M}$;
- $(A, B \in \mathcal{M}$ *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 un* ***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. □

Le résultat suivant jouera un rôle très important dans plusieurs démonstrations à venir.

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. Elle a été faite dans le cours d'intégration. Voici cependant les grandes lignes. 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. Comme les mesures μ_1 et μ_2 sont finies et qu'on a de plus $\mu_1(\mathbb{R}) = \lim_{n\to\infty} \mu_1(]-\infty, n]) = \lim_{n\to\infty} \mu_2(]-\infty, n]) = \mu_2(\mathbb{R})$, la famille

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

est une classe monotone. Par hypothèse, on peut appliquer le Théorème des classes monotones avec $\mathcal{M}$ et le π-système $\mathcal{C} := \big\{]-\infty, t];\ t \in \mathbb{R}\big\}$: la conclusion est 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 $\sum\limits_{\omega\in\Omega} \rho(\omega) = 1$.*

Remarque. Dit autrement, ρ est une fonction positive 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 $\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = \sum_{\omega \in D} \mathbb{P}(\{\omega\}) = 1$. Ainsi, $\rho(\omega) = \mathbb{P}(\{\omega\})$ est une densité discrète sur Ω.

(2) Soit $\rho : \Omega \to \mathbb{R}$ une densité discrète sur Ω. Pour $A \in \mathfrak{A}$, posons

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

Par définition, 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 on a par définition $\mathbb{P}(\{\omega\}) = \rho(\omega)$ pour tout $\omega \in \Omega$.

Montrons maintenant que $\mathbb{P}$ est une loi discrète. Cela repose sur le fait suivant, dont on se re-servira.

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 $\omega_1, \dots, \omega_N$. On a alors

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

donc *a fortiori* $\mu(\Omega) \geqslant \frac{N}{k}$ pour tout $N \in \mathbb{N}^*$, ce qui est 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 ; et 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 $\mathbb{P}(\Omega \backslash D) = 1 - \mathbb{P}(D) = 0$. Ainsi, $\mathbb{P}$ est portée par D et est donc une loi discrète.

Montrons enfin que $\mathbb{P}$ est la *seule* loi de probabilité sur $(\Omega, \mathfrak{A})$ telle que $\mathbb{P}(\{\omega\}) = \rho(\omega)$ pour tout $\omega \in \Omega$. Soit $\mathbb{P}'$ une autre loi vérifiant cette propriété. En posant à nouveau $D := \{\omega \in \Omega;\ \rho(\omega) \neq 0\}$, on a $\mathbb{P}'(D) = \mathbb{P}'\big(\bigcup_{\omega \in D} \{\omega\}\big) = \sum_{\omega \in D} \mathbb{P}'(\{\omega\})$ car D est dénombrable ; autrement dit $\mathbb{P}'(D) = \sum_{\omega \in D} \rho(\omega) = \sum_{\omega \in \Omega} \rho(\omega) = 1$. Donc $\mathbb{P}'(\Omega \backslash D) = 1 - \mathbb{P}'(D) = 0$. 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 de 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. Par la proposition précédente, si $\mathbb{P}$ est une loi discrète sur $(\Omega, \mathfrak{A})$ de densité associée ρ, on a pour tout ensemble $A \in \mathfrak{A}$:

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

où μ_c est la mesure de comptage sur Ω. Le membre de droite a un sens pour n'importe quelle partie A de Ω, et définit une mesure sur $(\Omega, \mathcal{P}(\Omega))$; donc on voit que la loi $\mathbb{P}$ peut se prolonger 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. Il s'agit de montrer que la fonction $\rho : \Omega \to \mathbb{R}^+$ définie par $\rho(\omega) := \frac{1}{\#\Omega}$ est une densité discrète, autrement dit que $\sum_{\omega \in \Omega} \rho(\omega) = 1$; ce qui est évident. □

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 $k \in [\![0, n]\!]$, on a par la formule du binôme

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

donc ρ est une densité discrète sur $\Omega = [\![0, n]\!]$. □

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 $k \in \mathbb{N}$, on a

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

donc ρ est une densité discrète sur $\mathbb{N}$. □

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, $f = \mathbf{1}_A$, le résultat est vrai 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, on en déduit que le résultat est encore vrai pour toute fonction f *étagée* positive ; et on conclut par convergence monotone.

(ii) La première partie découle de (i) appliqué à la fonction positive $|f|$; et $(*)$ s'obtient en écrivant $f = f^+ - f^-$ et en appliquant (i) à f^+ et f^-. □

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. C'est 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.

DÉFINITION 2.9. *Soit $d \in \mathbb{N}^*$. Une* **densité lebesguienne** *sur $\mathbb{R}^d$ est une fonction λ_d-mesurable $\rho : \mathbb{R}^d \to \mathbb{R}$ (i.e. ρ est définie λ_d-presque partout et presque partout égale à une fonction borélienne) telle que $\rho \geqslant 0$ presque partout et $\int_{\mathbb{R}^d} \rho(x)dx = 1$.*

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 $\rho_1(x) = \rho_2(x)$ λ_d-presque partout.

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 $\int_E (\rho_1 - \rho_2)\, d\lambda_d > 0$. Ainsi $\int_E \rho_1(x)\, dx > \int_E \rho_2(x)\, dx$, 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 $\lambda_d(K) > 0$, 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 est souhaitable de savoir redémontrer que l'intégrale $I := \int_{\mathbb{R}} e^{-u^2/2} du$ est égale à $\sqrt{2\pi}$. La preuve la plus simple consiste à écrire I^2 comme une intégrale double et à passer en coordonnées polaires (exo).

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 $\int_{\mathbb{R}} \frac{dx}{1+x^2} = \left[\arctan(x)\right]_{-\infty}^{\infty} = \pi$. □

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.**

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

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

Dans cette section, on se donne un espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$, et on suppose que $(\Omega, \mathfrak{A}, \mathbb{P})$ modélise une "expérience réelle". Par conséquent, on identifiera complètement les évènements mathématiques (*i.e.* les $E \in \mathfrak{A}$) avec des "évènements réels", et on s'autorisera des raisonnements "intuitifs".

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 devrait être intuitivement clair que le résultat du 1er jet n'influe pas sur les chances d'obtenir "pile" au 2ème jet, et donc qu'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 (et que le tirage est "sans remise" : on ne remet pas la 1ère carte tirée dans le jeu), 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. Il est raisonnable de considérer que la "probabilité conditionnelle" possède les propriétés suivantes.

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

(iii) Si la réalisation d'un évènement A implique celle de E, autrement dit si $A \subseteq E$, alors $\mathbb{P}(A \,|\, E)$ 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, puisque de toutes façons A entraîne E.)

Ces propriétés suffisent à déterminer $\mathbb{P}(A \,|\, E)$ pour n'importe quel évènement A. En effet, 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 (ii) et (o), la constante c vaut $1/\mathbb{P}(E)$; d'où la formule suivante :

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

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

Le raisonnement qui a conduit à cette formule peut être généralisé, moyennant une hypothèse sur l'espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$.

REMARQUE 3.2. Supposons que l'espace $(\Omega, \mathfrak{A}, \mathbb{P})$ possède la propriété suivante : pour tout $E \in \mathfrak{A}$ et pour tout $x \in [0, \mathbb{P}(E)]$, il existe $A \subseteq E$ tel que $\mathbb{P}(A) = x$. Supposons également que soit donnée, pour tout $E \in \mathfrak{A}$ vérifiant $\mathbb{P}(E) \neq 0$, une application $\mathbb{P}_E : \mathfrak{A} \to [0,1]$ vérifiant les conditions suivantes :

(i) $\mathbb{P}_E$ est une loi de probabilité ;

(ii) $\mathbb{P}_E(E) = 1$;

(iii') si $A \subseteq E$, alors $\mathbb{P}_E(A)$ ne dépend que de $\mathbb{P}(A)$.

Alors $\mathbb{P}_E(A) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)} = \mathbb{P}(A \mid E)$ pour tout E tel que $\mathbb{P}(E) \neq 0$ et pour tout $A \in \mathfrak{A}$.

Démonstration. Fixons E tel que $\mathbb{P}(E) > 0$, et posons $c := 1/\mathbb{P}(E)$.

Par hypothèse sur $(\Omega, \mathfrak{A}, \mathbb{P})$, on peut, pour tout $x \in [0, \mathbb{P}(E)]$, trouver $A_x \subseteq E$ tel que $\mathbb{P}(A_x) = x$; et par (iii'), on sait que $\mathbb{P}_E(A_x)$ ne dépend pas de l'évènement A_x vérifiant ces propriétés. On peut donc définir une fonction $\varphi : [0, \mathbb{P}(E)] \to [0,1]$ en posant

$$\varphi(x) := \mathbb{P}_E(A) \qquad \text{pour n'importe quel } A \subseteq E \text{ tel que } \mathbb{P}(A) = x.$$

Il suffit alors de prouver que

$$(*) \qquad \varphi(x) = c\,x \qquad \text{pour tout } x \in [0, \mathbb{P}(E)].$$

En effet, pour $A \in \mathfrak{A}$ quelconque, on aura alors (en utilisant le fait que $\mathbb{P}_E(E) = 1$ et donc $\mathbb{P}_E(\Omega \backslash E) = 0$) :

$$\mathbb{P}_E(A) = \mathbb{P}_E(A \cap E) = \varphi\big(\mathbb{P}(A \cap E)\big) = c\,\mathbb{P}(A \cap E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)}.$$

FAIT 1. Si $0 \leqslant x, y \leqslant \mathbb{P}(E)$ et $x + y \leqslant \mathbb{P}(E)$, alors $\varphi(x+y) = \varphi(x) + \varphi(y)$.

Preuve du Fait 1. Par hypothèse sur $(\Omega, \mathfrak{A}, \mathbb{P})$, on peut trouver $C \subseteq E$ tel que $\mathbb{P}(C) = x + y$; puis, comme $x \leqslant \mathbb{P}(C)$, on peut trouver $A \subseteq C$ tel que $\mathbb{P}(A) = x$. Alors $B := C \backslash A$ vérifie $\mathbb{P}(B) = \mathbb{P}(C) - \mathbb{P}(A) = y$; et par (i), on a

$$\varphi(x+y) = \mathbb{P}_E(C) = \mathbb{P}_E(A) + \mathbb{P}_E(B) = \varphi(x) + \varphi(y).$$

□

FAIT 2. La fonction φ est croissante.

Preuve du Fait 2. Soient x, y tels que $0 \leqslant x \leqslant y \leqslant \mathbb{P}(E)$. Par hypothèse sur $(\Omega, \mathfrak{A}, \mathbb{P})$, on peut trouver $B \subseteq E$ tel que $\mathbb{P}(B) = y$, puis $A \subseteq B$ tel que $\mathbb{P}(A) = x$. Par (i), on a donc $\varphi(x) = \mathbb{P}_E(A) \leqslant \mathbb{P}_E(B) = \varphi(y)$. □

FAIT 3. On a $\varphi(0) = 0$ et $\varphi(\mathbb{P}(E)) = 1$.

Preuve du Fait 3. Cela découle de (i) et (ii) : $\varphi(0) = \mathbb{P}_E(\varnothing) = 0$, et $\varphi(\mathbb{P}(E)) = \mathbb{P}_E(E) = 1$. □

La fin de la preuve est maintenant "classique".

En utilisant le Fait 1, on montre qu'on a $\varphi(px) = p\varphi(u)$ pour tout $u \in [0, \mathbb{P}(E)]$ et pour tout $p \in \mathbb{N}$ tel que $pu \leqslant \mathbb{P}(E)$; puis que $\varphi(rv) = r\varphi(v)$ pour tout $v \in [0, \mathbb{P}(E)]$ et tout *rationnel* $r \geqslant 0$ tel que $rv \leqslant \mathbb{P}(E)$. En prenant $v := \mathbb{P}(E) = c^{-1}$, on obtient donc $\varphi(rc^{-1}) = r$ pour tout $r \in \mathbb{Q} \cap [0,1]$, autrement dit $\varphi(x) = c\,x$ pour tout $x \in \mathcal{D} := [0, \mathbb{P}(E)] \cap c^{-1}\mathbb{Q}$.

Pour $x \in]0, \mathbb{P}(E)[$ quelconque, on choisit alors deux suites (s_n) et (t_n) de points de $\mathcal{D}$ tendant vers x, avec $s_n \leqslant x \leqslant t_n$. Par le Fait 2, on a $\varphi(s_n) \leqslant \varphi(x) \leqslant \varphi(t_n)$, autrement dit $c\,s_n \leqslant \varphi(x) \leqslant c\,t_n$ pour tout $n \in \mathbb{N}$, d'où $\varphi(x) = c\,x$ en passant à la limite. Ainsi, $\varphi(x) = c\,x$ pour tout $x \in]0, \mathbb{P}(E)[$, et ceci est encore vrai pour $x = 0$ et $x = \mathbb{P}(E)$ car $\varphi(0) = 0$ et $\varphi(\mathbb{P}(E)) = 1$ par le Fait 3. □

Exercice 1. Montrer que l'hypothèse sur $(\Omega, \mathfrak{A}, \mathbb{P})$ est satisfaite si $\Omega = \mathbb{R}$ et si $\mathbb{P}$ est une loi diffuse. (Commencer par montrer que pour tout borélien $E \subseteq \mathbb{R}$, la fonction ϕ définie par $\phi(t) = \mathbb{P}\big(E \cap]-\infty, t]\big)$ est continue sur $\mathbb{R}$.)

Exercice 2. Montrer que la Remarque 3.2 vaut encore si on remplace l'hypothèse sur $(\Omega, \mathfrak{A}, \mathbb{P})$ par la suivante : Ω est un ensemble fini et $\mathbb{P}$ est la probabilité uniforme. (Commencer par décrire l'ensemble $\mathcal{R} := \{\mathbb{P}(A);\ A \subseteq \Omega\}$. Montrer ensuite que si $x, y \in \mathcal{R}$ sont tels que $x + y \in \mathcal{R}$, et si $C \in \mathcal{P}(\Omega)$ est tel que $\mathbb{P}(C) = x + y$, alors on peut trouver $A, B \subseteq C$ tels que $\mathbb{P}(A) = x$ et $\mathbb{P}(B) = y$. Puis adapter la preuve de la Remarque 3.2.)

Notation. Si $A, E_1, \dots, E_n$ sont des évènements et si $\mathbb{P}(E_1 \cap \cdots \cap E_n) \neq 0$, on pose
$$\mathbb{P}(A \,|\, E_1, \dots, E_n) := \mathbb{P}(A \,|\, E_1 \cap \cdots \cap E_n).$$
C'est la “probabilité que A soit réalisé sachant que E_1, E_2 ... et E_n sont réalisés”.

Proposition 3.3. *Si $E_1, \dots, E_N$ 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*** *:*
$$\mathbb{P}(E_1 \cap \cdots \cap E_N) = \mathbb{P}(E_1)\,\mathbb{P}(E_2 \,|\, E_1)\,\mathbb{P}(E_3 \,|\, E_1, E_2) \cdots \mathbb{P}(E_N \,|\, E_1, \dots, E_{N-1}).$$

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 des probabilités en cascade 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 (et sans remise). Si on note E_i l'évènement “la i-ème carte tirée est un as” (pour $i = 1, 2, 3, 4$) alors
$$\begin{aligned} q &= \mathbb{P}(E_1)\,\mathbb{P}(E_2 \,|\, E_1)\,\mathbb{P}(E_3 \,|\, E_1, E_2)\,\mathbb{P}(E_4 \,|\, E_1, E_2, E_3) \\ &= \frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} \times \frac{1}{49}, \end{aligned}$$
d'où $p = \frac{1}{20825}$ après simplification.

Proposition 3.4. (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 :*
$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \,|\, E_i)\,\mathbb{P}(E_i).$$

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

Exemple. Une petite fille aimant les bonbons au citron mais pas les bonbons à la banane a devant elle trois boîtes opaques, la 1ère contenant 10 bonbons au citron et 30 bonbons à la banane, la 2ème contenant 20 bonbons au citron et 20 bonbons à la banane, et la 3ème contenant 20 bonbons au citron et 30 bonbons à la banane. Elle choisit une boîte 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 boîtes et qu'on note E_i l'évènement "la petite fille choisit la boîte 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,

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

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

Proposition 3.5. (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*

$$\mathbb{P}(E\,|\,E') = \frac{\mathbb{P}(E'\,|\,E)\,\mathbb{P}(E)}{\mathbb{P}(E')}\cdot$$

(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*

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

Démonstration. (1) est évident (micro-exo).

(2) Par (1), on a

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

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

Exemple. Dans un certain ensemble d'objets, on a de bonnes raisons d'estimer 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 ; et il y a cependant $\varepsilon\,\%$ de chances qu'un objet non défectueux soit déclaré défectueux par le test. On veut savoir quelle est la **fiabilité** du test, autrement dit la probabilité qu'un objet déclaré défectueux par le test soit effectivement défectueux.

Formellement, l'expérience consiste à tirer un objet au hasard et à le tester. Notons D l'évènement "l'objet est défectueux" et DD l'évènement "l'objet est déclaré défectueux". La probabilité cherchée est $p := \mathbb{P}(D\,|\,DD)$. On applique la théorème de

Bayes avec la partition (D, D^c) : cela donne

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

autrement dit

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

Application numérique : avec $N = 1000$ et $\varepsilon = 1$, on obtient $p = \frac{100}{1099} < \frac{1}{10}$. Donc, un test apparemment très raisonnable s'avère n'être même pas "fiable à 10%". La morale pourrait être la suivante : si on sait déjà qu'il y a très peu d'objets défectueux, autant ne rien tester du tout (!)

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
$$E \text{ et } F \text{ indépendants} \iff \mathbb{P}(E \mid F) = \mathbb{P}(E) \iff \mathbb{P}(F \mid E) = \mathbb{P}(F).$$
Autrement dit : savoir que F est réalisé n'a aucune influence sur le degré de plausibilité qu'on accorde à E, et *vice versa.*

Démonstration. Comme $\mathbb{P}(E \cap F) = \mathbb{P}(E \mid F)\mathbb{P}(F)$, on a $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$ si et seulement si $\mathbb{P}(E \mid F) = \mathbb{P}(E)$; et le résultat "symétrique" en découle en échangeant les rôles de E et F. □

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.

ATTENTION. Lorsqu'il y a plus de 2 évènements E_i, l'indédendance **deux à deux**, i.e le fait que E_i et E_j soient indépendants pour tout couple (i, j) tel que $i \neq j$, ne suffit pas à assurer l'indépendance de la famille $(E_i)_{i\in I}$. Par exemple, si on lance 2 dés l'un après l'autre, les évènements $E_1 :=$"le 1er chiffre obtenu est pair", $E_2 :=$"le 2ème chiffre obtenu est pair" et $E_3 :=$"la somme des chiffres est paire" sont deux à deux indépendants, mais la famille (E_1, E_2, E_3) n'est pas indépendante (exo).

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

(i) si à chaque étape on remet le dragibus tiré dans la boîte, les évènements $E_1, \ldots, E_n$ sont indépendants ;

(ii) si on mange les dragibus tirés à chaque étape, les évènements $E_1, \ldots, E_n$ ne sont pas indépendants.

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

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

Preuve formelle. On modélise par $\Omega := D^n$, où D est l'ensemble des dragibus que contient la boîte : un résultat est donc une suite $\omega = (d_1, \ldots, d_n)$, où les d_i sont dans D. Tous les tirages sont équiprobables, donc on prend pour $\mathbb{P}$ la loi uniforme sur Ω. Appelons B l'ensemble des dragibus noirs, et R l'ensemble des dragibus rouges. Pour $i = 1, \ldots, n$, on a alors $E_i = D \times \cdots \times B \times \cdots D$, où B apparaît à la place i. Donc $\mathbb{P}(E_i) = \frac{\# E_i}{\# \Omega} = \frac{N_1 \times N^{n-1}}{N^n} = N_1/N$. Maintenant, soient $i_1, \ldots, i_r \in [\![1, n]\!]$ deux à deux différents ; pour simplifier on suppose que $i_1 = 1, \ldots, i_r = r$. Alors $E_1 \cap \cdots \cap E_r = B \times \cdots \times B \times D \times \cdots \times D$. Donc $\mathbb{P}(E_1 \cap \cdots \cap E_r) = \frac{N_1^r \times N^{n-r}}{N^n} = (N_1/N)^r$; et ainsi $\mathbb{P}(E_1 \cap \cdots \cap E_r) = \mathbb{P}(E_1) \cdots \mathbb{P}(E_r)$.

(ii) Preuve intuitive. On a "évidemment" $\mathbb{P}(E_i) = N_1/N$ pour tout i, mais "tout aussi évidemment" $\mathbb{P}(E_2 \,|\, E_1) = (N_1 - 1)/(N - 1)$, et donc $\mathbb{P}(E_2 \,|\, E_1) \neq \mathbb{P}(E_2)$. Donc E_1 et E_2 ne sont déjà pas indépendants et, *a fortiori*, $E_1, \ldots, E_n$ ne le sont pas non plus.

Preuve formelle. On peut modéliser en prenant pour Ω l'ensemble de toutes les suites $(d_1, \ldots, d_n) \in D^n$ où les d_i sont *deux à deux différents*, avec toujours des évènements élémentaires équiprobables. On a donc $\#\Omega = N(N-1) \cdots (N-n+1) = \frac{N!}{(N-n)!}$. Il y a N_1 façons de choisir le premier dragibus si on veut qu'il soit noir, et $\frac{(N-1)!}{\big((N-1)-(n-1)\big)!} = \frac{(N-1)!}{(N-n)!}$ de choisir ensuite les autres. Donc $\mathbb{P}(E_1) = N_1 \times \frac{(N-1)!}{(N-n)!} \times \frac{(N-n)!}{N!} = \frac{N_1}{N}$. De même, $\mathbb{P}(E_2) = \frac{N_1}{N}$. Mais $\mathbb{P}(E_1 \cap E_2) = \frac{N_1}{N} \times \frac{N_1 - 1}{N-1}$ par un raisonnement analogue, et donc $\mathbb{P}(E_1 \cap E_2) \neq \mathbb{P}(E_1)\, \mathbb{P}(E_2)$. $\square$

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 2ème 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. $\square$

Le fait suivant est souvent utile.

Fait 4.2. *Si $(E_i)_{i \in I}$ est une famille indépendante d'évènements, alors les E_i^c sont indépendants. Plus généralement, si pour $i \in I$ on pose $F_i := E_i$ ou E_i^c, alors les F_i sont indépendants.*

Démonstration. Tout repose sur un

Sous-fait. Si $(A_i)_{i\in I}$ est une famille indépendante d'évènements et si $j_0 \in I$, alors la famille $\{A_{j_0}^c\} \cup (A_i)_{i\neq j_0}$ est indépendante.

Preuve du sous-fait. On sait déjà que si $i_1, \ldots, i_r \in I$ sont deux à deux distincts et différents de j_0, alors $\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_r}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_r})$; donc il suffit de montrer que si $i_1, \ldots, i_r$ sont deux à deux distincts et différents de j_0, alors

$$\mathbb{P}(A_{j_0}^c \cap A_{i_1} \cap \cdots \cap A_{i_r}) = \mathbb{P}(A_{j_0}^c)\,\mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_r}).$$

Pour cela, on écrit

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

□

Soient maintenant $i_1, \ldots, i_r \in I$ deux à deux distincts. En appliquant r fois le sous-fait, on voit que la famille $(F_{i_1}, E_{i_2}, \ldots, E_{i_r})$ est indépendante, puis que la famille $(F_{i_1}, F_{i_2}, E_{i_3}, \ldots, E_{i_r})$ est indépendante, et ainsi de suite jusqu'à $(F_{i_1}, \ldots, F_{i_r})$. Ainsi, toute sous-famille finie de la famille $(F_i)_{i\in I}$ est indépendante, ce qui termine la preuve.

□

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 l'identité de la tribu $\mathfrak{B}$, on parle de variable aléatoire *à valeurs dans* Λ. Ainsi, une variable aléatoire à valeurs dans Λ est "un élément de Λ qui dépend du hasard".

Remarque 2. Pour aller plus vite, 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. En termes semi-formels, le résultat de l'expérience est une suite $(a_1, \dots, a_n)$, où a_i est le mot "pile" ou le mot "face". Plus formellement, on peut modéliser 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 $\left\{X \in \bigcup_{k\in\mathbb{N}} A_k\right\} = \bigcup_{k\in\mathbb{N}}\{X \in A_k\}$ où la réunion est disjointe, donc $\mathbb{P}_X\left(\bigcup_{k\in\mathbb{N}} A_k\right) = \mathbb{P}\left(\bigcup_{k\in\mathbb{N}}\{X \in A_k\}\right) = \sum_{k=0}^{\infty}\mathbb{P}(X \in A_k) = \sum_{k=0}^{\infty}\mathbb{P}_X(A_k)$. Donc $\mathbb{P}_X$ est une mesure ; et on a $\mathbb{P}_X(\Lambda) = \mathbb{P}(X \in \Lambda) = \mathbb{P}(\Omega) = 1$. $\square$

Remarque. Pour alléger, on écrira souvent $\mathbb{P}(X \in A)$ au lieu de $\mathbb{P}(\{X \in A\})$; et donc

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

EXEMPLE 1. Soient $(\Omega, \mathfrak{A}, \mathbb{P})$ 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 $\mathbb{P}_X(\{0\}) = \mathbb{P}(E^c) = 1 - p$. $\square$

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 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 $k \in [\![0, n]\!]$:

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

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

REMARQUE. Plus généralement, supposons qu'on répète n fois une expérience "binaire" 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 $\mathcal{B}(n, p)$.

Démonstration. Pour $i = 1, \dots, n$, notons E_i l'évènement "succès au i-ème coup" ; et pour $A \subseteq [\![1, n]\!]$, notons E^A l'évènement "succès au i-ème coup pour tout $i \in A$ et échec au j-ième coup pour tout $j \notin A$", autrement dit

$$E^A = \bigcap_{i\in A} E_i \cap \bigcap_{j\in[\![1,n]\!]\setminus A} E_j^c.$$

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

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

$\square$

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. Il suffit de prendre $(\Omega, \mathfrak{A}, \mathbb{P}) := (\Lambda, \mathfrak{B}, \mu)$ et de poser $X(\omega) := \omega$ 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”

Le résultat suivant est fondamental pour toute la suite.

Théorème 3.1. (Théorème de transfert)
Soit X une variable aléatoire 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 $\mu=\mathbb{P}_X$. $\square$

Remarque. Au lieu de $f\circ X$, on écrira souvent $f(X)$. Avec cette notation, le Théorème de transfert devient

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

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. On applique le Théorème de transfert et le Lemme 2.6 du Chapitre 2 (exo). $\square$

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. $\square$

Exemple 1. 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 est presque sûrement >0. Comme X est presque sûrement à valeurs dans $]0,1[$, on a pour toute fonction borélienne $f:\mathbb{R}\to\mathbb{R}^+$:

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

où $g:\mathbb{R}\to\mathbb{R}^+$ est la fonction obtenue en posant $g(x):=f(-\log(x))$ si $x\in\,]0,1[$ et, par exemple, $g(x):=0$ partout ailleurs. En appliquant le Théorème de transfert à X, on obtient

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

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. $\square$

EXEMPLE 2. Soit Z une va à valeurs dans $\mathbb{R}^2$, qu'on écrit $Z = (X,Y)$. On suppose que Z est une va à densité, de densité $\rho : \mathbb{R}^2 \to \mathbb{R}$. Alors X et Y sont des va à densité, de densités ρ_X et ρ_Y données par

$$\rho_X(x) = \int_{\mathbb{R}} \rho(x,y)\,dy \qquad \text{et} \qquad \rho_Y(y) = \int_{\mathbb{R}} \rho(x,y)\,dx.$$

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

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

où $g : \mathbb{R}^2 \to \mathbb{R}^+$ est la fonction définie par $g(x,y) := f(x)$. Donc, par le Théorème de transfert,

$$\begin{aligned}\int_\Omega f(X)\,d\mathbb{P} &= \int_{\mathbb{R}^2} g(x,y)\,d\mathbb{P}_Z(x,y)\\ &= \int_{\mathbb{R}^2} f(x)\,\rho(x,y)\,dxdy\\ &= \int_{\mathbb{R}} f(x)\left(\int_{\mathbb{R}} \rho(x,y)\,dy\right)dx \qquad \text{par “Fubini positif”.}\end{aligned}$$

Donc X est à densité avec la bonne formule pour ρ_X ; et de même pour Y. □

Exercice 1. Démontrer le résultat de l'Exemple 2 en utilisant uniquement la définition de “va à densité”.

Exercice 2. Montrer que si X est une va 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'.

Le lemme suivant permet de déterminer “mécaniquement” la loi de la va $\Phi(X)$ en fonction de celle de X.

LEMME 4.1. *Soit $X : (\Omega,\mathfrak{A},\mathbb{P}) \to \Lambda$ une va à valeurs dans Λ, et soit $\Phi : \Lambda \to \Lambda'$ une application mesurable.*

(a) *La loi $\mathbb{P}_{\Phi(X)}$ est l'*image par Φ *de la loi de X : pour tout ensemble mesurable $A \subseteq \Lambda'$, on a*

$$\mathbb{P}_{\Phi(X)}(A) = \mathbb{P}_X\big(\Phi^{-1}(A)\big).$$

(b) *Pour toute fonction mesurable $f : \Lambda' \to \mathbb{C}$ ou bien $\geqslant 0$ ou bien intégrable par rapport à $\mathbb{P}_{\Phi(X)}$, on a*

$$\int_{\Lambda'} f\,d\mathbb{P}_{\Phi(X)} = \int_\Lambda (f\circ\Phi)\,d\mathbb{P}_X.$$

Démonstration. La partie (a) est immédiate puisque $\Phi(X) \in A \iff X \in \Phi^{-1}(A)$. Pour (b), on applique (2 fois) le Théorème de transfert :

$$\int_{\Lambda'} f \, d\mathbb{P}_{\Phi(X)} = \int_{\Omega} f(\Phi(X)) \, d\mathbb{P} = \int_{\Omega} (f \circ \Phi)(X) \, d\mathbb{P} = \int_{\Lambda} (f \circ \Phi) \, d\mathbb{P}_X .$$

□

Corollaire 4.2. *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. On a $\mathbb{P}(\Phi(X) = j) = \mathbb{P}_{\Phi(X)}(\{j\}) = \mathbb{P}_X\big(\Phi^{-1}(\{j\})\big)$ par le Lemme 4.1 ; et donc, comme X est une va discrète :

$$\mathbb{P}(\Phi(X) = j) = \sum_{i \in \Phi^{-1}(\{j\})} \mathbb{P}(X = i).$$

Voici une autre façon d'écrire la même chose : d'après la formule des probabilités totales (et en supposant que $\mathbb{P}(X = i) \neq 0$ pour tout $i \in I$), on a

$$\mathbb{P}(\Phi(X) = j) = \sum_{i \in I} \mathbb{P}\big(\Phi(X) = j \,|\, X = i\big) \, \mathbb{P}(X = i);$$

d'où le résultat car $\mathbb{P}\big(\Phi(X) = j \,|\, X = i\big)$ vaut 1 si $\Phi(i) = j$ et 0 si $\Phi(i) \neq j$. □

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 le Corollaire 4.2 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

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

□

Corollaire 4.3. *Soit $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ un difféomorphisme, et soit $X : (\Omega, \mathfrak{A}, \mathbb{P}) \to \mathbb{R}^d$ une va à densité à valeurs dans $\mathbb{R}^d$, de densité $\rho_X : \mathbb{R}^d \to \mathbb{R}^+$. Alors $Y := \Phi(X)$ est une va à densité, et sa densité ρ_Y est donnée par*

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

Démonstration. Bien entendu, $J_\psi(y)$ désigne le *déterminant jacobien* de ψ au point y. La preuve du corollaire est une application "automatique" du Lemme 4.1 et de la formule de changement de variable : pour toute fonction borélienne $f : \mathbb{R}^d \to \mathbb{R}^+$, on a

$$\begin{aligned}\int_{\mathbb{R}^d} f\, d\mathbb{P}_{\Phi(X)} &= \int_{\mathbb{R}^d} (f \circ \Phi)\, d\mathbb{P}_X \\ &= \int_{\mathbb{R}^d} f(\Phi(x))\, \rho_X(x) dx \\ &= \int_{\mathbb{R}^d} f(y)\, \rho_X(\Psi(y)) |J_\Psi(y)| dy \qquad \text{en posant } y = \Phi(x),\ i.e.\ x = \Psi(y);\end{aligned}$$

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

Remarque. Il n'est pas indispensable d'apprendre par coeur la formule pour ρ_Y. Ce qui est important, en revanche, est de retenir que ρ_Y se calcule mécaniquement en appliquant le Théorème de transfert et la formule de changement de variable ; et de savoir refaire le calcul !

Exemple. Soit Z une va à densité à valeurs dans $\mathbb{R}^d$, dont la densité associée $\rho_Z : \mathbb{R}^d \to \mathbb{R}$ est de la forme

$$\rho_Z(x) = \alpha(x_1^2 + \cdots + x_d^2),$$

pour une certaine fonction borélienne $\alpha : [0, \infty[\to \mathbb{R}^+$. Alors la loi de Z est **invariante par isométries** : pour toute isométrie euclidienne $\Phi : \mathbb{R}^d \to \mathbb{R}^d$, on a $\mathbb{P}_{\Phi(Z)} = \mathbb{P}_Z$.

Démonstration. D'après le Corollaire 4.3, on a $\mathbb{P}_{\Phi(Z)} = \rho_Z(\Psi(y))|J_\Psi(y)|dy$, où $\Psi := \Phi^{-1}$. Comme Ψ est une isométrie et $\rho_Z(x) = \alpha(\|x\|^2)$ pour tout $x \in \mathbb{R}^d$ (où $\|\cdot\|$ est la norme euclidienne sur $\mathbb{R}^d$), on a $|J_\Psi(y)| = 1$ et $\rho_Z(\Psi(y)) = \phi(\|\Psi(y)\|^2) = \alpha(\|y\|^2) = \rho_Z(y)$ pour tout $y \in \mathbb{R}^d$; donc $\mathbb{P}_{\Phi(Z)} = \rho_Z(y)dy = \mathbb{P}_Z$. □

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, 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. Il est clair que si les X_i sont indépendantes, alors $(*)$ est vraie. Inversement, supposons que $(*)$ ait lieu. Soient $i_1, \ldots, i_r \in \{1, \ldots, d\}$ deux à deux distincts, et soient $A_{i_1} \in \mathfrak{B}_{i_1}, \ldots, A_{i_r} \in \mathfrak{B}_{i_r}$. Si on pose $A_i := \Lambda_i$ pour $i \neq i_1, \ldots, i_r$, alors

$$\{X_{i_1} \in A_{i_1}, \ldots, X_{i_r} \in A_{i_r}\} = \{X_1 \in A_1, \ldots, X_d \in A_d\}.$$

Par $(*)$, on a donc

$$\begin{aligned}\mathbb{P}(X_{i_1} \in A_{i_1}, \ldots, X_{i_r} \in A_{i_r}) &= \mathbb{P}(X_1 \in A_1, \ldots, X_d \in A_d)\\ &= \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_d \in A_d)\\ &= \mathbb{P}(X_{i_1} \in A_{i_1}) \cdots \mathbb{P}(X_{i_r} \in A_{i_r})\end{aligned}$$

car $\mathbb{P}(X_i \in A_i) = \mathbb{P}(X_i \in \Lambda_i) = 1$ si $i \neq i_1, \ldots, i_r$. □

Exemple 1. On joue n fois à pile ou face. Pour $i = 1, \ldots, n$, 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 $X_1, \ldots, X_n$ 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 $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$.

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

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

Donc, par définition de la loi uniforme,

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

autrement dit

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

Par ailleurs, on a aussi $\{X_i = A_i\} = \{0,1\} \times \cdots \times A_i \times \cdots \times \{0,1\}$, 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 $\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n)$. □

Généralisation. Soient $\Omega_1, \ldots, \Omega_n$ des ensembles finis, et soit $\Omega := \Omega_1 \times \cdots \times \Omega_n$ muni de la loi uniforme. Pour $i = 1, \ldots, n$, soit $X_i : \Omega \to \Omega_i$ la va définie par $X_i(\omega_1, \ldots, \omega_n) := \omega_i$. 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. (Observer que si $i \in I$ et $A_i \subseteq \{0,1\}$, alors $\{X_i \in A_i\} = \varnothing$, Ω, E_i ou E_i^c, et utiliser le Fait 4.2 du Chapitre 2.) □

Le lemme suivant, spécifique aux va discrètes, est souvent bien utile.

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. Alors X et Y sont indépendantes si et seulement si*

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

Démonstration. Supposons $(*)$ vérifiée, et soient $A \subseteq I$ et $B \subseteq J$ quelconques. Pour $(i,j) \in I \times J$, posons $E_{i,j} := \{X = i, Y = j)\}$. Les $E_{i,j}$ forment une partition dénombrable de Ω ; donc, par la formule des probabilités totales, on a

$$\begin{aligned}\mathbb{P}(X \in A, Y \in B) &= \sum_{(i,j) \in I \times J} \mathbb{P}\big(X \in A, Y \in B \,|\, E_{i,j}\big)\,\mathbb{P}(E_{i,j}) \\ &= \sum_{(i,j) \in I \times J} \mathbb{P}\big((X,Y) \in A \times B \,|\, X = i, Y = j\big)\,\mathbb{P}(X = i, Y = j).\end{aligned}$$

Comme $\mathbb{P}\big((X,Y) \in A \times B \,|\, X = i, Y = j\big)$ vaut 1 si $(i,j) \in A \times B$ et 0 sinon, on en déduit

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

Donc X et Y sont indépendantes. □

6. Fonction de répartition d'une va

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, 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 presque sûrement égale à une certaine constante $a \in \mathbb{R}$. Alors $F_X(t) = 0$ si $t < a$ et $F_X(t) = 1$ si $t \geqslant a$. Autrement dit : si $\mathbb{P}_X = \delta_a$, alors $F_X = \mathbf{1}_{[a,\infty[}$.

Démonstration. C'est évident par définition : $F_X(t) = \mathbb{P}(X \leqslant t)$ vaut 0 si $t < a$ et 1 si $t \geqslant a$. □

Exercice. Soit X une va à valeurs dans $\mathbb{N}$, qu'on considère comme une va réelle. Montrer que $F_X(t) = 0$ si $t < 0$, et que si $t \geqslant 0$, alors

$$F_X(t) = \sum_{k=0}^{E(t)} \mathbb{P}(X = k).$$

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 $t \in \mathbb{R}$:*

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

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 $\mathbb{P}_X(\{t\}) = 0$. ☐

PROPOSITION 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.

Refaisons quand même la preuve pour bien insister sur le fait que ce résultat est non trivial. Soit $\mathcal{M} := \{A \in \mathfrak{B}(\mathbb{R});\ \mathbb{P}_{X_1}(A) = \mathbb{P}_{X_2}(A)\}$. Alors $\mathcal{M}$ est une classe monotone car $\mathbb{P}_{X_1}$ et $\mathbb{P}_{X_2}$ sont des lois de probabilité. Maintenant, soit $\mathcal{C} := \{\,]-\infty, t];\ t \in \mathbb{R}\}$. Alors $\mathcal{C}$ est un π-système. De plus, on a $\mathcal{C} \subseteq \mathcal{M}$ puisque $F_{X_1} = F_{X_2}$. Par le Théorème des classes monotones, on en déduit que $\mathcal{M}$ contient $\sigma(\mathcal{C})$; mais $\sigma(\mathcal{C}) = \mathfrak{B}(\mathbb{R})$, donc $\mathcal{M} = \mathfrak{B}(\mathbb{R})$, *i.e.* $\mathbb{P}_{X_1} = \mathbb{P}_{X_2}$. ☐

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. De plus, les hypothèses faites sur F_X entraînent que pour tous $a, b \in \mathbb{R}$, on a

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

En effet, en supposant $a < b$, en notant $t_1 < \cdots < t_{N-1}$ les points de K compris entre a et b et en posant $t_0 := a$ 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 $k = 0, \dots, N-1$ car F est $\mathcal{C}^1$ sur $]t_k, t_{k+1}[$ et continue sur $[t_k, t_{k+1}]$ (exo). D'où $(*)$ en sommant ces égalités.

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. La preuve est très courte si on suppose que F est *continue* et *strictement croissante*. En effet, 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. Donc $F_X = F$.

Dans le cas général, F n'est pas une bijection de $\mathbb{R}$ sur $]0,1[$, mais on peut tout de même trouver un “substitut” à F^{-1} et adapter la preuve précédente.

Soit $F^- :]0,1[\,\to\mathbb{R}$ la fonction définie comme suit :

$$\forall u\in\,]0,1[\,:\;\; F^-(u) := \inf\{x\in\mathbb{R};\; F(x)>u\}.$$

(Exo : montrer que si F est continue et strictement croissante, alors $F^- = F^{-1}$.)

Fait. Pour $u\in\,]0,1[$ et $t\in\mathbb{R}$, on a l'équivalence

$$F^-(u)\leqslant t \iff \forall n\in\mathbb{N}^*\;:\; F\Big(t+\frac{1}{n}\Big)>u.$$

Preuve du Fait. Si $F^-(u)\leqslant t$, alors on a certainement $F^-(u)<t+\frac{1}{n}$ pour tout $n\in\mathbb{N}^*$; donc, par définition de F^-, on peut trouver $x<t+\frac{1}{n}$ tel que $F(x)>u$; et comme F est croissante, on en déduit que $F\big(t+\frac{1}{n}\big)>u$, pour tout $n\in\mathbb{N}^*$. Inversement, si $F\big(t+\frac{1}{n}\big)>u$ pour tout $n\in\mathbb{N}^*$, alors on a $F^-(u)\leqslant t+\frac{1}{n}$ pour tout $n\in\mathbb{N}^*$ par définition de F^-, et donc $F^-(u)\leqslant t$. $\square$

Maintenant, soit U une va réelle uniformément distribuée sur $]0,1[$. Comme F^- est une fonction borélienne (car elle est croissante, exo), on définit une va réelle X en posant $X := F^-(U) = F^-\circ U$. Montrons que $F_X = F$.

Soit $t\in\mathbb{R}$ fixé. Par le Fait, on a

$$\{X\leqslant t\} = \{F^-(U)\leqslant t\} = \bigcap_{n\in\mathbb{N}^*}\Big\{U<F\Big(t+\frac{1}{n}\Big)\Big\} =: \bigcap_{n\in\mathbb{N}^*} E_n.$$

Comme la fonction F est croissante, la suite (E_n) est décroissante ; donc on a

$$\mathbb{P}(X\leqslant t) = \lim_{n\to\infty}\mathbb{P}(E_n) = \lim_{n\to\infty}\mathbb{P}\left(U<F\Big(t+\frac{1}{n}\Big)\right).$$

Mais $\mathbb{P}\left(U<F\big(t+\frac{1}{n}\big)\right) = F\big(t+\frac{1}{n}\big)$ par définition de la loi uniforme U, puisque $0\leqslant F\Big(t+\frac{1}{n}\Big)\leqslant 1$. De plus, F est continue à droite, donc $F\big(t+\frac{1}{n}\big)\to F(t)$ quand $n\to\infty$. Ainsi, on obtient

$$F_X(t) = \lim_{n\to\infty} F\Big(t+\frac{1}{n}\Big) = F(t).$$

$\square$

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 $i_1, \dots, i_r \in I$ *et* $A_{i_k} \in \mathfrak{A}_{i_k}$ *pour* $k = 1, \dots, r$.

(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

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

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 $\bigotimes_{i\in I} \mathfrak{A}_i$ ne contient pas nécessairement tous les singletons.

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 $A \in \mathfrak{A}_i$, ce qui montre que π_i est mesurable. □

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

- 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 il suffit de poser $A_i := \Omega_i$ 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. On l'a vu dans le cours d'intégration. □

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

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 $\mathfrak{A} = \otimes_{i\in I}\mathfrak{A}_i$.

Notation. Si $i_1,\dots,i_r\in I$ sont deux à deux distincts et si $A_{i_1}\in\mathfrak{A}_{i_1},\dots,A_{i_r}\in\mathfrak{A}_{i_r}$, on pose

$$\text{“}A_{i_1}\times\cdots\times A_{i_r}\text{”} := \{\omega = (\omega_i)_{i\in I}\in\Omega;\ \omega_{i_1}\in A_{i_1},\dots,\omega_{i_r}\in A_{i_r}\}.$$

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

$$(*)\qquad \mathbb{P}(\text{“}A_{i_1}\times\cdots\times A_{i_r}\text{”}) = \mathbb{P}_{i_1}(A_{i_1})\cdots\mathbb{P}_{i_r}(A_{i_r})$$

*pour tout cylindre $C = \text{“}A_{i_1}\times\cdots\times A_{i_r}\text{”}\subseteq\Omega$. 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 dans le cas d'un nombre fini d'espaces $(\Omega_i,\mathfrak{A}_i,\mathbb{P}_i)$.

Pour alléger la typographie, on va supposer qu'il n'y a que *deux* espaces $(\Omega_1,\mathfrak{A}_1,\mathbb{P}_1)$ et $(\Omega_2,\mathfrak{A}_2,\mathbb{P}_2)$. Soit comme plus haut $\mathcal{C}$ la famille de tous les cylindres de $\Omega := \Omega_1\times\Omega_2$. Un $C\in\mathcal{C}$ typique est donc de la forme

$$C = A_1\times A_2,\qquad \text{où}\quad A_1\in\mathfrak{A}_1 \text{ et } A_2\in\mathfrak{A}_2.$$

Fait 1. La famille $\mathcal{C}$ est un **semi-anneau** ; autrement dit :

- $\mathcal{C}$ est stable par intersections finies ;
- si $C, C'\in\mathcal{C}$, alors $C\backslash C'$ est réunion d'un nombre fini d'éléments de $\mathcal{C}$ deux à deux disjoints.

Démonstration. On a déjà observé que $\mathcal{C}$ est stable par intersections finies. De plus, si $C = A_1 \times A_2$ et $C' = A_1' \times A_2'$, alors

$$C \backslash C' = \Big[(A_1 \backslash A_1') \times A_2\Big] \cup \Big[(A_1 \cap A_1') \times (A_2 \backslash A_2')\Big];$$

donc $C \backslash C'$ est réunion de 2 éléments de $\mathcal{C}$ disjoints. □

Fait 2. Soit $\alpha : \mathcal{C} \to \mathbb{R}^+$ la "fonction d'ensembles" définie par

$$\alpha(C) := \mathbb{P}_1(A_1)\,\mathbb{P}_2(A_2) \qquad \text{pour tout cylindre} \quad C = A_1 \times A_2.$$

Alors α possède les propriétés suivantes.

(i) α est *finiment additive* : si C et C' sont deux cylindres disjoints tels que $C \cup C'$ est encore un cylindre, alors $\alpha(C \cup C') = \alpha(C) + \alpha(C')$.

(ii) α est *dénombrablement sous-additive* : si $(C_k)_{k \in \mathbb{N}}$ est une suite de cylindres et si C est un cylindre tel que $C \subseteq \bigcup_{k=0}^{\infty} C_k$, alors

$$\alpha(C) \leqslant \sum_{k=0}^{\infty} \alpha(C_k).$$

Démonstration. (i) Écrivons $C = A_1 \times A_2$ et $C' = A_1' \times A_2'$. Posons également $E := C \cup C'$, et écrivons $E = B_1 \times B_2$ (on suppose que E est un cylindre). Comme C et C' sont disjoints, on a $\mathbf{1}_E = \mathbf{1}_C + \mathbf{1}_{C'}$; autrement dit

$$\mathbf{1}_{B_1}(x)\mathbf{1}_{B_2}(y) = \mathbf{1}_{A_1}(x)\mathbf{1}_{A_2}(y) + \mathbf{1}_{A_1'}(x)\mathbf{1}_{A_2'}(y) \qquad \text{pour tout } (x,y) \in \Omega_1 \times \Omega_2.$$

En fixant x et en intégrant en y sur Ω_2 (par rapport à $\mathbb{P}_2$), on en déduit

$$\mathbf{1}_{B_1}(x)\mathbb{P}_2(B_2) = \mathbf{1}_{A_1}(x)\,\mathbb{P}_2(A_2) + \mathbf{1}_{A_1'}(x)\,\mathbb{P}_2(A_2') \qquad \text{pour tout } x \in \Omega_1;$$

d'où, en intégrant maintenant en x sur Ω_1 (par rapport à $\mathbb{P}_1$) :

$$\mathbb{P}_1(B_1)\,\mathbb{P}_2(B_2) = \mathbb{P}_1(A_1)\,\mathbb{P}_2(A_2) + \mathbb{P}_1(A_1')\,\mathbb{P}_2(A_2'),$$

autrement dit $\alpha(E) = \alpha(C) + \alpha(C')$.

(ii) Comme $C \subseteq \bigcup_0^{\infty} C_k$, on a

$$\mathbf{1}_C \leqslant \sum_{k=0}^{\infty} \mathbf{1}_{C_k}.$$

Autrement dit, en écrivant $C = A_1 \times A_2$ et $C_k = A_{k,1} \times A_{k,2}$:

$$\mathbf{1}_{A_1}(x)\mathbf{1}_{A_2}(y) \leqslant \sum_{k=0}^{\infty} \mathbf{1}_{A_{k,1}}(x)\mathbf{1}_{A_{k,2}}(y) \qquad \text{pour tout } (x,y) \in \Omega_1 \times \Omega_2.$$

En raisonnant comme dans (i) et en appliquant (deux fois) le théorème de convergence monotone (sous la forme "interversion série/intégrale"), on en déduit

$$\mathbb{P}_1(A_1)\,\mathbb{P}_2(A_2) \leqslant \sum_{k=0}^{\infty} \mathbb{P}_1(A_{k,1})\,\mathbb{P}_2(A_{k,2}),$$

ce qui est le résultat souhaité. □

Conclusion. Par un théorème général vu dans le cours d'intégration, il existe une mesure $\mathbb{P}$ sur $(\Omega, \mathfrak{A})$ telle que $\mathbb{P}(C) = \alpha(C)$ pour tout $C \in \mathcal{C}$; autrement dit, $\mathbb{P}(A_1 \times A_2) = \mathbb{P}_1(A_1)\,\mathbb{P}_2(A_2)$ pour tout cylindre $A_1 \times A_2 \subseteq \Omega_1 \times \Omega_2$. On a $\mathbb{P}(\Omega) = \mathbb{P}_1(\Omega_1)\mathbb{P}_2(\Omega_2) = 1$, donc $\mathbb{P}$ est une loi de probabilité.

□

Preuve de l'existence dans le cas d'une infinité d'espaces $(\Omega_i, \mathfrak{A}_i, \mathbb{P}_i)$.

On va supposer que $I = \mathbb{N}^*$, et donc $\Omega = \prod_{i \geqslant 1} \Omega_i$. La preuve suit le même schéma que dans le cas d'un nombre fini d'espaces, mais les détails sont un peu plus techniques.

Notons encore $\mathcal{C}$ la famille de tous les cylindres de Ω. Comme $I = \mathbb{N}^*$, un $C \in \mathcal{C}$ peut toujours s'écrire sous la forme

$$C = \text{“}A_1 \times \cdots \times A_N\text{”} \quad \text{où } N \geqslant 1 \text{ et } A_i \in \mathfrak{A}_i \text{ pour } i = 1, \ldots, N.$$

De manière équivalente,

$$C = \prod_{i \geqslant 1} A_i,$$

où $A_i \in \mathfrak{A}_i$ pour tout $i \geqslant 1$ et $A_i = \Omega_i$ sauf pour un nombre fini de i.

On montre comme plus haut que la famille $\mathcal{C}$ est un semi-anneau : si C et C' sont deux cylindres de Ω, on peut écrire $C = \text{“}A_1 \times \cdots \times A_N\text{”}$ et $C' = \text{“}A'_1 \times \cdots \times A'_N\text{”}$ pour un même entier N (micro-exo), et on a alors

$$\begin{aligned} C \backslash C' = \text{“}(A_1 \backslash A'_1) \times A_2 \times \cdots \times A_N\text{”} \cup \text{“}(A_1 \cap A'_1) \times (A_2 \backslash A'_2) \times A_3 \times \cdots \times A_N\text{”} \\ \cup \cdots \cup \text{“}(A_1 \cap A'_1) \times \cdots \times (A_{N-1} \cap A'_{N-1}) \times (A_N \backslash A'_N)\text{”}, \end{aligned}$$

ce qui montre que $C \backslash C'$ est réunion de N cylindres disjoints.

Soit $\alpha : \mathcal{C} \to [0,1]$ la fonction d'ensembles définie comme on imagine :

$$\alpha(C) := \prod_{i=1}^{\infty} \mathbb{P}_i(A_i) \qquad \text{pour tout cylindre } C = \prod_{i \geqslant 1} A_i.$$

Cette définition a bien un sens : en effet, dans l'écriture $C = \prod_{i \geqslant 1} A_i$ tous les A_i sauf un nombre fini sont égaux à Ω_i, donc $\mathbb{P}_i(A_i) = 1$ pour tout i sauf un nombre fini, et donc le “produit infini ” $\prod_{i=1}^{\infty} \mathbb{P}_i(A_i)$ est en fait un produit fini.

Le point clé est de prouver l'analogue du Fait 2 (dont l'énoncé est rigoureusement identique). La partie (i) se démontre exactement comme plus haut (exo), donc on va se concentrer sur (ii).

Soient $(C_k)_{k \in \mathbb{N}}$ une suite de cylindres et C un cylindre tels que $C \subseteq \bigcup_{k=0}^{\infty} C_k$. Écrivons $C = \prod_{i \geqslant 1} A_i$ et $C_k = \prod_{i \geqslant 1} A_{k,i}$. Il s'agit de montrer que $\alpha(C) \leqslant \sum_{k=0}^{\infty} \alpha(C_k)$, autrement dit que

$$\prod_{i=1}^{\infty} \mathbb{P}(A_i) \leqslant \sum_{k=0}^{\infty} \prod_{i=1}^{\infty} \mathbb{P}(A_{k,i}). \tag{2.1}$$

Supposons que (2.1) soit faux, autrement dit qu'on ait

$$\sum_{k=0}^{\infty} \prod_{i=1}^{\infty} \mathbb{P}(A_{k,i}) < \prod_{i=1}^{\infty} \mathbb{P}(A_i). \tag{$*$}$$

Alors on peut trouver $\omega_1 \in \Omega_1$ tel que

$$\sum_{k=0}^{\infty} \mathbf{1}_{A_{k,1}}(\omega_1) \prod_{i>1} \mathbb{P}(A_{k,i}) < \mathbf{1}_{A_1}(\omega_1) \prod_{i>1} \mathbb{P}(A_i). \tag{$*_1$}$$

En effet, si on avait

$$\forall \omega_1 \in \Omega_1 \ : \ \sum_{k=0}^{\infty} \mathbf{1}_{A_{k,1}}(\omega_1) \prod_{i>1} \mathbb{P}(A_{k,i}) \geqslant \mathbf{1}_{A_1}(\omega_1) \prod_{i>1} \mathbb{P}(A_i),$$

on obtiendrait une contradiction avec $(*)$ en intégrant sur Ω_1 par rapport à $\mathbb{P}_1$.

Ayant fixé ω_1 vérifiant $(*_1)$, on peut ensuite trouver $\omega_2 \in \Omega_2$ tel que

$$(*_2) \qquad \sum_{k=0}^{\infty} \mathbf{1}_{A_{k,1}}(\omega_1)\,\mathbf{1}_{A_{k,2}}(\omega_2) \prod_{i>2} \mathbb{P}(A_{k,i}) < \mathbf{1}_{A_1}(\omega_1)\,\mathbf{1}_{A_2}(\omega_2) \prod_{i>2} \mathbb{P}(A_i);$$

et ainsi de suite. De façon précise, on construit par récurrence une suite $(\omega_i)_{i\geqslant 1}$ avec $\omega_i \in \Omega_i$ pour tout i, telle que

$$(**) \qquad \forall s \geqslant 1\ :\ \sum_{k=0}^{\infty} \prod_{i=1}^{s} \mathbf{1}_{A_{k,i}}(\omega_i) \prod_{i>s} \mathbb{P}(A_{k,i}) < \prod_{i=1}^{s} \mathbf{1}_{A_i}(\omega_i) \prod_{i>s} \mathbb{P}(A_i).$$

Posons alors $\omega := (\omega_i)_{i\geqslant 1}$, de sorte que ω est un point de l'espace produit $\Omega = \prod_{i\geqslant 1} \Omega_i$.

Par $(**)$, on a $\mathbf{1}_{A_i}(\omega_i) > 0$ pour tout $i \geqslant 1$, *i.e.* $\omega_i \in A_i$. Donc ω appartient à $\prod_{i\geqslant 1} A_i = C$; et donc $\omega \in C_{k_0}$ pour un certain $k_0 \in \mathbb{N}$ puisque $C \subseteq \bigcup_0^\infty C_k$.

Choisissons $i_0 \geqslant 1$ tel que $A_{k_0,i} = \Omega_i$ pour tout $i > i_0$. Alors $\mathbb{P}_i(A_{k_0,i}) = 1$ pour $i > i_0$; et comme $\omega_i \in A_{k_0,i}$ pour tout i, on en déduit que

$$\prod_{i=1}^{i_0} \mathbf{1}_{A_{k_0,i}}(\omega_i) \prod_{i>i_0} \mathbb{P}_i(A_{k_0,i}) = 1.$$

Par conséquent

$$\sum_{k=0}^{\infty} \prod_{i=1}^{i_0} \mathbf{1}_{A_{k,i}}(\omega_i) \prod_{i>i_0} \mathbb{P}_i(A_{k,i}) \geqslant 1,$$

ce qui contredit $(**)$ pour $s := i_0$.

□

Remarque. Le Théorème 2.1 est en fait valable pour une famille quelconque (possiblement non dénombrable) d'espaces de probabilité $(\Omega_i, \mathfrak{A}_i, \mathbb{P}_i)$. La preuve n'est pas beaucoup plus compliquée que celle donnée plus haut, mais il y a quand même un peu de travail en plus.

Exemple 1. Soient $\Omega_1, \dots, \Omega_d$ des ensembles finis, et pour $i = 1, \dots, d$, 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 ρ est définie par $\rho(x,y) := \rho_1(x)\rho_2(y)$.

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. □

Remarque. Si $(\Omega_1, \mathfrak{A}_1, \mu_1)$, ..., $(\Omega_d, \mathfrak{A}_d, \mu_d)$ sont des espaces mesurés (en nombre fini), on peut démontrer l'existence et l'unicité de la mesure produit $\mu_1 \otimes \cdots \otimes \mu_d$ sans qu'il soit nécessaire de supposer que $\mu_1, \dots, \mu_d$ sont des mesures *de probabilité* : il suffit que les μ_i soient *σ-finies*. (Une mesure ν sur un espace mesurable $(M, \mathfrak{T})$ est dite σ-finie si l'espace M est réunion d'une suite $(M_n)_{n\in\mathbb{N}}$ d'ensembles mesurables vérifiant $\nu(M_n) < \infty$. Par exemple, la mesure de Lebesgue sur $\mathbb{R}^d$ est σ-finie ; et toute mesure *finie* est évidemment σ-finie.) La preuve est essentiellement identique à celle donnée plus haut. En revanche, dans le cas d'une infinité d'espaces mesurés $(\Omega_i, \mathfrak{A}_i, \mu_i)$, il est essentiel de supposer que l'on a affaire à des mesures de probabilité.

2.2. Théorème de Fubini. Il s'agit du résultat suivant, d'usage constant.

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 $\mathbb{P}_1 \otimes \mathbb{P}_2$, 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. Elle est tout à fait non triviale, mais identique à celle faite dans le cours d'intégration pour le Théorème de Fubini sur $\mathbb{R}^p \times \mathbb{R}^q$ (muni de la mesure de Lebesgue). □

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 σ-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 $X = (X_i)_{i \in I}$, alors les X_i sont des va. Les lois $\mathbb{P}_{X_i}$ s'appellent les **lois 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 $C =$ "$A_{i_1} \times \cdots \times A_{i_r}$". Mais ceci est évident : on a

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

et donc $X^{-1}(\text{“}A_{i_1} \times \cdots \times A_{i_r}\text{”}) \in \mathfrak{A}$ 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 la 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 "A_i" $\subseteq \Omega$; 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 a déjà démontré ce résultat pour $d = 2$ au Chapitre 3 (Exemple 2 après le Théorème de transfert 3.1). La preuve pour d quelconque est identique, aux notations près. □

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

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

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

$$\mathbb{P}_X(\text{“}A_{i_1} \times \cdots \times A_{i_r}\text{”}) = \mathbb{P}_{X_{i_1}}(A_{i_1}) \cdots \mathbb{P}_{X_{i_r}}(A_{i_r})$$

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

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 3.2 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 le Lemme 3.2 entraîne que les X_i sont indépendantes. $\square$

Corollaire 3.4. *Soient X_1, X_2 deux va réelles, et soient $\mathbb{P}_1$, $\mathbb{P}_2$ deux lois de probabilité sur $\mathbb{R}$. 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 3.2. 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 $\mathbb{P}_{(X_1,X_2)} = \mathbb{P}_X = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$, et le Lemme 3.2 entraîne que X_1 et X_2 sont indépendantes. $\square$

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. $\square$

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 $\int_{\mathbb{R}} \rho = \left(\int_{\mathbb{R}} \rho_1\right)\left(\int_{\mathbb{R}} \rho_2\right) = 1$ (exo). Donc ρ est effectivement une densité lebesguienne.

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

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

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. 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

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

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

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

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 $\mathcal{N}(0,\sigma^2)$.

Supposons maintenant m_X et m_Y quelconques. Alors $\widetilde{X}:=X_1-m_X$ et $\widetilde{Y}:=Y-m_Y$ sont toujours indépendantes, et suivent les lois $\mathcal{N}(0,\sigma_X^2)$ et $\mathcal{N}(0,\sigma_Y^2)$ (exo déjà posé). 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 $\mathcal{N}(m,\sigma^2)$. □

Remarque. Le calcul précédent n'est pas particulièrement lumineux. Au chapitre 9, on donnera une preuve beaucoup plus claire du Corollaire 1.2.

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

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

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 4.2 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 $X_1, \ldots, X_n$ 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 on a posé

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

Maintenant, un $\bar{\varepsilon} \in \{0, 1\}^n$ s'identifie à $A_{\bar{\varepsilon}} = \{i \in [\![1, n]\!];\ \varepsilon_i = 1\} \subseteq [\![1, n]\!]$ (précisément : $\bar{\varepsilon} = \mathbf{1}_{A_{\bar{\varepsilon}}}$). 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.

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

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

REFORMULATION. "Avec des mots", le Lemme de Borel-Cantelli s'énonce comme suit.

(1) Si la série $\sum \mathbb{P}(E_i)$ converge, alors il est presque sûr que : à partir d'un certain, rang l'évènement E_i n'a pas lieu.
(2) Si la série $\sum \mathbb{P}(E_i)$ diverge et si les E_i sont indépendants, alors il est presque sûr que : E_i a lieu pour une infinité de i.

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 $\log \mathbb{P}\Big(\bigcap_{i\geqslant n} E_i^c\Big) = -\infty$. (On étend la fonction log à $[0, \infty]$ en posant $\log(\infty) = \infty$ et $\log(0) = -\infty$.)
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 que la série $\sum \mathbb{P}(E_i)$ diverge, on obtient donc le résultat souhaité :

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

$\square$

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 facile.

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. $\square$

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

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

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 qui 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 en déduit que $\sum_0^{\infty} \mathbb{P}(E_i) = \infty$. 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 $X_{n_i}(\omega) = \cdots = X_{n_i+K}(\omega) = 1$. $\square$

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 $X_1, \dots, X_d$, on écrit $\sigma(X_1, \dots, X_d)$.*

Remarque 1. La définition a un sens : $\sigma(X)$ est bien une tribu (exo) ; et c'est une *sous-tribu* de $\mathfrak{A}$, *i.e.* $\sigma(X) \subseteq \mathfrak{A}$.

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 $(\Lambda_i, \mathfrak{B}_i)$. Posons $X := (X_i)_{i\in I}$ 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 ainsi $\{Y \in A\} \in \sigma(X)$ puisque Φ est mesurable (exo). □

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 $X_i(\omega)$".

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. C'est clair intuitivement ; mais faisons cependant une preuve détaillée. 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 ainsi $E_{p,q,k} = X^{-1}(A_{p,q,k}) \in \sigma(X_i, i \in \mathbb{N})$. $\square$

Remarque. La preuve précédente est presque trop détaillée. En fait, tout ce qu'il importe de retenir est le "slogan" suivant : si un évènement E peut s'écrire à l'aide d'unions et d'intersections dénombrables faisant intervenir uniquement les va X_i, $i \in I$, alors $E \in \sigma(X_i, i \in I)$.

Exercice. Soit $(\Omega_i, \mathfrak{A}_i)_{i\in I}$ une famille d'univers probabilisables, et soit $\Omega := \otimes_{i\in I}\Omega_i$. 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 ainsi Y est constante sur E. De même, Y est constante sur E^c.

La réciproque est laissée en exo. $\square$

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 $\Phi^{-1}(A) \in \mathfrak{B}$. $\square$

Remarque. 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 $\Phi : \Lambda \to \mathbb{R}$.

Démonstration. Comme les va Y^+ et Y^- sont $\sigma(X)$-mesurables (exo), on se ramène au cas où $Y \geqslant 0$.

Supposons d'abord que Y soit $\sigma(X)$-étagée. Alors Y s'écrit $Y = \sum_{i=1}^N \alpha_i \mathbf{1}_{E_i}$, où $E_i \in \sigma(X)$, *i.e.* $E_i = X^{-1}(A_i)$ pour un certain $A_i \in \mathfrak{B}$. Alors $\mathbf{1}_{E_i} = \mathbf{1}_{A_i} \circ X = \mathbf{1}_{A_i}(X)$ par définition, donc $Y = \Phi(X)$ pour $\Phi = \sum_{i=1}^N \alpha_i \mathbf{1}_{A_i}$.

Si Y est quelconque, on peut trouver une suite de va $\sigma(X)$-étagées positives (Y_n) qui croît vers Y. D'après ce qui précède, on peut écrire $Y_n = \Phi_n(X)$, pour une certaine fonction mesurable $\Phi_n : \Lambda \to \mathbb{R}^+$. On pose alors $\Psi := \limsup \Phi_n$ et $\Phi := \Psi\, \mathbf{1}_{\Psi<\infty}$; et on vérifie que $Y = \Phi(X)$ (exo). $\square$

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. Indépendance et disjointude. Le lemme suivant est très intuitif, mais sa preuve n'est pas immédiate. C'est par ailleurs un résultat extrêmement 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.* En particulier *: si E^1 et E^2 sont deux évènements tels que $E^1 \in \sigma(X_i, i \in I^1)$ et $E^2 \in \sigma(X_i, i \in I^2)$, alors E^1 et E^2 sont indépendants.*

Démonstration. Chaque X_i est à valeurs dans un univers $(\Lambda_i, \mathcal{B}_i)$.

Posons $\Lambda^1 := \otimes_{i \in I^1} \Lambda_i$ et $\Lambda^2 := \otimes_{i \in I^2} \Lambda_i$, 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 $A =$ "$A_{i_1} \times \cdots \times A_{i_r}$" et $B =$ "$B_{j_1} \times \cdots \times B_{j_s}$" où $i_1, \dots, i_r \in I^1$ sont deux à deux distincts et $j_1, \dots j_s \in I^2$ sont deux à deux distincts. Alors $i_1, \dots, i_r, j_1, \dots, j_s$ sont tous distincts car $I^1 \cap I^2 = \varnothing$. Donc, par indépendance des X_i, on a

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

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

$$\begin{aligned} \mu(B) &:= \mathbb{P}(X^1 \in A, X^2 \in B) = \mathbb{P}_{(X^1, X^2)}(A \times B) \quad \text{et} \\ \mu'(B) &:= \mathbb{P}(X^1 \in A)\, \mathbb{P}(X^2 \in B) = \mathbb{P}_{X^1}(A)\, \mathbb{P}_{X^2}(B). \end{aligned}$$

(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.

La preuve est identique à celle du Cas 2, 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. $\square$

Exemple 2. Soit $(X_i)_{i\in\mathbb{N}}$ 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. $\square$

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. $\square$

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 à la suite** (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 à (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 à la section 3.1 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 ainsi $E \in \sigma(X_i,\, i \geqslant N)$. $\square$

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 à (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. $\square$

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 à (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 $\{X_{35} + X_{47} \leqslant 8\} \notin \sigma(X_i,\, i > 47)$. $\square$

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 à la suite* (X_i)*, alors* $\mathbb{P}(E) = 0$ *ou* 1*.*

Démonstration. Pour $n \in \mathbb{N}$, posons $\mathfrak{A}_n := \sigma(X_0, \dots, X_n)$. 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. Chaque X_i est à valeurs dans un univers $(\Lambda_i, \mathfrak{B}_i)$. Notons X la va produit $(X_i)_{i\in\mathbb{N}}$, de sorte que $\mathfrak{A}_\infty = \sigma(X)$.

Pour $n \in \mathbb{N}$ fixé, la va $Z^n := (X_0, \dots, X_n)$ est $\sigma(X)$-mesurable car les X_i le sont. Donc $\mathfrak{A}_n = \sigma(X_0, \dots, X_n) \subseteq \sigma(X) = \mathfrak{A}_\infty$ pour tout $n \in \mathbb{N}$, et ainsi la tribu engendrée par $\bigcup_{n\in\mathbb{N}} \mathfrak{A}_n$ est contenue dans $\mathfrak{A}_\infty$.

Inversement, notons $\mathfrak{B}$ la tribu produit sur $\Lambda = \prod_{i\in\mathbb{N}} \Lambda_i$ et posons $\mathfrak{T} = \sigma\big(\bigcup_{n\in\mathbb{N}} \mathfrak{A}_n\big)$. Il s'agit de montrer que pour tout $B \in \mathfrak{B}$, l'ensemble $X^{-1}(B)$ appartient à la tribu $\mathfrak{T}$. Autrement dit, on veut montrer que l'application $X : \Omega \to \Lambda$ est $(\mathfrak{T}, \mathfrak{B})$-mesurable. Comme les cylindres engendrent la tribu produit $\mathfrak{B}$, il suffit de vérifier que $X^{-1}(C) \in \mathfrak{T}$ pour tout cylindre $C =$ “$A_{i_1} \times \dots \times A_{i_r}$” $\subseteq \Lambda$. Mais ceci est évident : comme $X^{-1}(C) = \{X_{i_1} \in A_{i_1}, \dots, X_{i_r} \in A_{i_r}\}$, on voit que $X^{-1}(C) \in \sigma(X_{i_1}, \dots, X_{i_r})$, donc $X^{-1}(C) \in \bigcup_{n\in\mathbb{N}} \mathfrak{A}_n$, et donc $X^{-1}(C) \in \mathfrak{T}$. □

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 $A \in \mathfrak{A}_n = \sigma(X_0, \dots, X_n)$. Comme E est asymptotique relativement à (X_i), on a $E \in \sigma(X_i, i \geqslant n+1)$; donc, par le Lemme 3.5, 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 π-systèmes (exo) ; et par le Fait 2, on a $\mathbb{P}(A \cap E) = \mathbb{P}(A)\mathbb{P}(E)$ pour tout $A \in \mathcal{C}$. Le résultat en découle 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 1. 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.

Exercice 2. Soit $(X_i)_{i\in\mathbb{N}}$ une suite de va réelles indépendantes, et soit X une va réelle. On suppose que X est $\sigma(X_i, i \geqslant N)$-mesurable pour tout $N \in \mathbb{N}$. Montrer X est *presque sûrement constante* ; autrement dit : qu'il existe une constante c telle que $X = c$ ps. (Commencer par montrer que la fonction de répartition de X est une fonction indicatrice.)

Exercice 3. Soit $(a_n)_{n\in\mathbb{N}}$ une suite de va indépendantes à valeurs complexes. Montrer que le rayon de convergence de la série entière $\sum a_n z^n$ est presque sûrement constant.

Chapitre 6

Espérance, variance, moments

Dans ce chapitre, toutes les va sont supposées définies sur un même espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$ fixé une fois pour toutes.

Pour $p \in [1, \infty]$, on pose $L^p := L^p(\Omega, \mathbb{P})$, où on ne considère que des fonctions à valeurs réelles.

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

1. Espérance d'une va réelle

Définition 1.1. *On dit qu'un va réelle X* ***admet une espérance****, ou* ***admet une moyenne****, si $X \in L^1$; et dans ce cas 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 (ce qui correspond plus ou moins au mot anglais "expectation"). Plus généralement, si $X = (X_1, \dots, X_d)$ est une va à valeurs dans $\mathbb{R}^d$, on dit que X admet une espérance si les X_i sont dans L^1, et on pose $\mathbb{E}(X) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_d))$.*

Remarque 1. On pose aussi $\mathbb{E}(X) := \int_\Omega X\, d\mathbb{P}$ pour toute va *positive* X. Donc, une va $X \geqslant 0$ "admet une espérance" si et seulement si $\mathbb{E}(X) < \infty$.

Remarque 2. Une va réelle X appartient à L^1 si et seulement si $\mathbb{E}(|X|) < \infty$, et dans ce cas $\|X\|_1 = \mathbb{E}(|X|)$. Plus généralement, une va X appartient à L^p pour un certain $p < \infty$ si et seulement si $\mathbb{E}(|X|^p) < \infty$, et dans ce cas $\|X\|_p = \mathbb{E}(|X|^p)^{1/p}$. Enfin, si $X, Y \in L^2$, alors leur produit scalaire au sens de L^2 est $\langle X, Y \rangle_{L^2} = \mathbb{E}(XY)$.

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

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

De plus, l'espérance est *croissante* : si $X, Y \in L^1$ et $X \leqslant Y$, 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 $\mathbb{E}(X) = 0$.

Exercice. Soit X une va à valeurs dans $\mathbb{R}^d$. Montrer que X admet une espérance si et seulement si $\mathbb{E}(\|X\|) < \infty$, où $\| \cdot \|$ est n'importe quelle norme sur $\mathbb{R}^d$.

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. $\square$

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

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

Démonstration. On applique la proposition avec $f(x) := x$. $\square$

COROLLAIRE 1.4. *La loi détermine l'espérance : 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. $\square$

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. $\square$

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

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

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 la fonction f apparaissant sous l'intégrale est continue sur $\mathbb{R}$ avec $f(x) = O(1/x^2)$ quand $x \to \pm\infty$; donc $X \in L^1$. Ensuite,

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

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

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

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

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 $\pm\infty$. $\square$

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 $\mathcal{E} := \mathbb{R}\,\mathbf{1} \subseteq L^2$ le sous-espace vectoriel (de dimension 1) formé par les va constantes. 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
$$\langle X - \mathbb{E}(X)\mathbf{1}, a\mathbf{1}\rangle_{L^2} = a\,\langle X, \mathbf{1}\rangle_{L^2} - a\mathbb{E}(X)\,\langle \mathbf{1},\mathbf{1}\rangle_{L^2} = a\mathbb{E}(X) - a\mathbb{E}(X)\,\underbrace{\mathbb{E}(\mathbf{1})}_{=1} = 0.$$
$\square$

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, $\sigma(X)$ est donc la distance de X au sous-espace de L^2 constitué par les va constantes. Intuitivement, $\sigma^2(X)$ mesure la "dispersion" de X par rapport à sa moyenne. En particulier, on a $\sigma^2(X) = 0$ si et seulement si X est presque sûrement égale à une constante, et si et seulement si $X = \mathbb{E}(X)$ ps.

REMARQUE 2. La variance ne dépend pas linéairement de la va, mais "quadratiquement" : 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. □

Le lemme suivant permet souvent de simplifier un calcul de variance.

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

Démonstration. On développe la norme au carré :

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

□

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

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

Pour calculer cette dernière intégrale, on intégre par parties en remarquant que

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

On obtient ainsi

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

d'où finalement $\sigma^2(X) = \sigma^2$. □

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

PROPOSITION 3.1. *Si $X_1, \ldots, X_n$ 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 $X_1, \ldots, X_n$ 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 $i = 1, \ldots, n$. □

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

THÉORÈME 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

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

En particulier $\mathbb{E}(|XY|) < \infty$, i.e. $XY \in L^1$. 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. La preuve est 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. On a $\langle X, Y\rangle_{L^2} = \mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y) = 0$. □

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 dans L^2, on appelle **covariance** de X et Y 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

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

ce qui est le résultat souhaité 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 $\sigma^2(X) = \sigma^2(S) = \sigma^2(X_1) + \cdots + \sigma^2(X_n) = np(1-p)$. □

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') $\mathbb{E}\big(f_1(X_1) \cdots f_d(X_d)\big) = \mathbb{E}(f_1(X_1)) \cdots \mathbb{E}(f_d(X_d))$ *pour toutes fonctions boréliennes positives $f_1, \dots, f_d$.*

Démonstration. Les implication $(1) \implies (2)$ et $(1) \implies (2')$ découlent du Théorème 3.3 appliquée 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, alors on a pour tous boréliens $A_1, \dots, A_d \subseteq \mathbb{R}$:

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

donc les X_i sont indépendantes. □

4. Formules et inégalités (très) utiles

4.1. Intégration par parties. Le résultat suivant porte souvent le nom de "formule d'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

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

La preuve de la 2ème formule est identique. On peut aussi déduire cette formule de la première en observant que $\mathbb{P}(Z > t) = \mathbb{P}(Z \geqslant t)$ pour tout $t \in \mathbb{R}$ tel que $\mathbb{P}(Z = t) = 0$, donc *presque partout* relativement à la mesure de Lebesgue car $\{t;\ \mathbb{P}(Z = t) \neq 0\}$ est dénombrable. □

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,

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

Démonstration. On applique la Proposition 4.1 avec $Z = |X|$ et $\phi(t) = t^p$. □

Corollaire 4.3. *Soit $1 \leqslant p < \infty$. Une va X réellle est dans L^p si et seulement si*

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

Démonstration. Conséquence immédiate du Corollaire 4.2 □

Corollaire 4.4. *Soit $a \in \mathbb{N}$. Si X est une va discrète à valeurs dans l'ensemble $\mathbb{N}_a := \{a, a+1, a+2, \dots\}$, alors*

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

Démonstration. La va X est positive, donc $\mathbb{E}(X)$ a un sens. Par le Corollaire 4.3 avec $p := 1$ et comme $X \geqslant a$ ps, on a

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

(La dernière ligne vient du fait que X est à valeurs dans $\mathbb{N}$: on a $\{X > t\} = \{X > k\}$ pour tout $k \in \mathbb{N}$ et pour tout $t \in [k, k+1[$.) □

Exercice. Montrer que si X est une va réelle quelconque, alors

$$X \in L^1 \iff \sum_{n=0}^{\infty} \mathbb{P}\big(|X| > n\big) < \infty.$$

4.2. Inégalité de Markov et applications. Le résultat suivant est très simple, mais remarquablement utile.

PROPOSITION 4.5. *Pour toute va positive Z et pour tout $\alpha > 0$, on a l'inégalité suivante, qu'on appelle **inégalité de Markov** :*

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

Plus généralement, pour tout $\alpha > 0$ et pour tout $p \geqslant 1$, on a

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

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

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

Pour la deuxième inégalité, il suffit d'observer qu'on a $\mathbb{P}(Z \geqslant \alpha) = \mathbb{P}(Z^p \geqslant \alpha^p)$ car la fonction $u \mapsto u^p$ est strictement croissante sur $\mathbb{R}^+$, et d'appliquer l'inégalité de Markov à $\widetilde{Z} := Z^p$ et $\widetilde{\alpha} := \alpha^p$. □

COROLLAIRE 4.6. (inégalité de Bienaymé-Tchebitchev)
Soit X une va réelle dans L^2, et soit $m := \mathbb{E}(X)$. Alors, pour tout $\varepsilon > 0$, on a

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

Démonstration. Par l'inégalité de Markov "d'ordre 2" appliquée à $Z := |X - m|$, on a

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

□

Corollaire 4.7. *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 la fonction $x \mapsto e^{\lambda x}$ est strictement croissante sur $\mathbb{R}$, on a $\mathbb{P}(X \geqslant \varepsilon) = \mathbb{P}\big(e^{\lambda X} \geqslant e^{\lambda\varepsilon}\big)$. Il suffit donc d'appliquer Markov à $Z := e^{\lambda X}$. □

Remarque. L'inégalité du Corollaire 4.7 pourrait s'appeler "inégalité de Markov exponentielle".

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 sûr 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 $(X_i)_{i\geqslant 1}$ 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 $Z_n \to 1/2$ 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

$$\sum_{n=1}^{\infty} \mathbb{P}\Big(\big|Z_n - \frac{1}{2}\big| \geqslant \varepsilon\Big) < \infty.$$

Comme $\big\{|Z_n - \frac{1}{2}| \geqslant \varepsilon\big\} = \big\{Z_n \geqslant \frac{1}{2} + \varepsilon\big\} \cup \big(Z_n \leqslant \frac{1}{2} - \varepsilon\big\}$, union disjointe, on a

$$\mathbb{P}\Big(\big|Z_n - \frac{1}{2}\big| \geqslant \varepsilon\Big) = \mathbb{P}\Big(Z_n \geqslant \frac{1}{2} + \varepsilon\Big) + \mathbb{P}\Big(Z_n \leqslant \frac{1}{2} - \varepsilon\Big).$$

Par conséquent, il suffit de montrer que

$$\sum_{n=1}^{\infty} \mathbb{P}\Big(Z_n \geqslant \frac{1}{2} + \varepsilon\Big) < \infty \qquad \text{et} \qquad \sum_{n=1}^{\infty} \mathbb{P}\Big(Z_n \leqslant \frac{1}{2} - \varepsilon\Big) < \infty.$$

On a pour tout $\lambda > 0$:

$$\begin{aligned}
\mathbb{P}\Big(Z_n \geqslant \frac{1}{2} + \varepsilon\Big) &= \mathbb{P}\Big(X_1 + \cdots + X_n \geqslant n(\frac{1}{2} + \varepsilon)\Big) \\
&\leqslant e^{-\lambda\, n(\frac{1}{2}+\varepsilon)}\, \mathbb{E}\big(e^{\lambda(X_1+\cdots+X_n)}\big) \qquad \text{par Markov "exponentiel"} \\
&= e^{-\lambda\, n(\frac{1}{2}+\varepsilon)} \prod_{i=1}^{n} \mathbb{E}\big(e^{\lambda X_i}\big) \qquad \text{par indépendance.}
\end{aligned}$$

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 $i = 1, \ldots, n$. 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 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

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

En étudiant $\varphi(\lambda) := -n\varepsilon\lambda + n\frac{\lambda^2}{8}$ (qui est après tout un gentil trinôme du second degré), on constate 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. Inégalité de Jensen. Rappelons que 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 4.8. (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 entraîne 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 deux à deux distincts (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 $\lambda > 0$ telle 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 $t > 0$:

$$\mathbb{P}(|X| > t) = \mathbb{P}\big(e^{\lambda|X|} > e^{\lambda t}\big) \leqslant C\, e^{-\lambda t}, \quad \text{avec } C := \mathbb{E}(e^{\lambda|X|}).$$

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 4.3. □

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. L'énoncé a donc 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 (Théorème d'approximation de Weierstrass). 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. Il suffit de montrer que $\mu_1([a, b]) = \mu_2([a, b])$ pour tout intervalle compact $[a, b] \subseteq \mathbb{R}$. En effet, ceci entraîne que $\mu_1(]-\infty, t]) = \lim_{n\to\infty} \mu_1([-n, t]) = \lim_{n\to\infty} \mu_2([-n, t]) = \mu_2(]-\infty, t]$ pour tout $t \in \mathbb{R}$, et donc $\mu_1 = \mu_2$ d'après le Corollaire 1.5 du Chapitre 3. On peut aussi utiliser directement le Théorème des classes monotones, ce qui est sans doute plus naturel et en tous cas plus "formateur". (Faire l'exo.)

Pour $n \in \mathbb{N}^*$, soit $f_n : \mathbb{R} \to \mathbb{R}$ la fonction définie comme suit : $f_n(t) \equiv 1$ sur $[a, b]$, $f_n(t) \equiv 0$ en dehors de $[a - \frac{1}{n}, b + \frac{1}{n}]$, et f_n est affine sur $[a - \frac{1}{n}, a]$ et sur $[b, b + \frac{1}{n}]$. (*Dessiner le graphe de f_n.*) Alors les f_n sont continues, $f_n \to \mathbf{1}_{[a,b]}$ 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

$$\mu_1([a, b]) = \lim_{n\to\infty} \int_{\mathbb{R}} f_n\, d\mu_1 \qquad \text{et} \qquad \mu_2([a, b]) = \lim_{n\to\infty} \int_{\mathbb{R}} f_n\, d\mu_2;$$

donc $\mu_1([a,b]) = \mu_2([a,b])$ 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). $\square$

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. $\square$

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 espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$. 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 $Z_n \xrightarrow{\text{ps}} Z$.

Démonstration. C'est une conséquence (de la partie triviale) du Lemme de Borel-Cantelli ; *cf* le Chapitre 5. □

DÉFINITION 1.3. *Soit (Z_n) une suite de va réelles définies sur le même espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$ 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 $\|Z_n - Z\|_p \to 0$. On écrit alors $Z_n \xrightarrow{L^p} Z$.*

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 $S_n := \sum_{k=0}^{n} X_k$ 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'entraîne l'autre.

Exemple 1. Soit $(Z_n)_{n\geqslant 1}$ une suite de va (réelles) telles que $\mathbb{P}(Z_n = n^3) = \frac{1}{n^2}$ et $\mathbb{P}(Z_n = 0) = 1 - \frac{1}{n^2}$ pour tout $n \geqslant 1$. Alors $Z_n \xrightarrow{\text{ps}} 0$ par l'Exemple 1.2 , mais $\|Z_n\|_1 \to \infty$ (exo).

Exemple 2. Soit $(Z_n)_{n\geqslant 1}$ une suite de va *indépendantes* telles que $\mathbb{P}(Z_n = 1) = \frac{1}{n}$ et $\mathbb{P}(Z_n = 0) = 1 - \frac{1}{n}$ pour tout $n \geqslant 1$. Alors $\|Z_n\|_1 \to 0$, mais $\mathbb{P}\big(\overline{\lim}\,\{Z_n = 1\}\big) = 1$ par Borel-Cantelli, donc Z_n ne tend pas vers 0 presque sûrement (exo).

Il y a cependant des liens entre la convergence presque sûre et la convergence L^p, comme le montrent les deux résultats suivants.

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. Ce résultat a été vu dans le cours d'intégration, et porte le nom de **Théorème de 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). $\square$

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. Ce résultat a également été vu dans le cours d'intégration. Supposons que Z_n tende vers Z en norme L^1. Alors, (Z_n) possède une sous-suite (Z_{n_k}) telle que $\|Z_{n_k} - Z\|_1 \leqslant 2^{-k}$ pour tout $k \in \mathbb{N}$. Si on pose $f := \sum_{k=0}^{\infty} |Z_{n_k} - Z|$, alors f est une fonction mesurable positive bien définie, et $\int_\Omega f\, d\mathbb{P} = \sum_0^\infty \int_\Omega |Z_{n_k} - Z|\, d\mathbb{P} = \sum_0^\infty \|Z_{n_K} - Z\|_1 < \infty$. Donc $f < \infty$ presque sûrement, autrement dit la série $\sum (Z_{n_k} - Z)$ est presque sûrement absolument convergente, et en particulier $Z_{n_k} - Z \xrightarrow{\text{ps}} 0$. $\square$

2. Convergence en probabilité

Définition 2.1. *Soit (Z_n) une suite de va réelles, toutes définies sur le même espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$. On dit que la suite (Z_n) **converge en probabilité** s'il existe une va Z telle que*

$$\mathbb{P}\big(|Z_n - Z| \geqslant \varepsilon\big) \xrightarrow{n\to\infty} 0 \qquad \textit{pour tout } \varepsilon > 0.$$

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. $\square$

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.

Preuve abrégée. Si (Z_n) est de Cauchy en probabilité, elle possède une sous-suite (Z_{n_k}) telle que $\mathbb{P}\big(|Z_{n_{k+1}} - Z_{n_k}| \geqslant 2^{-k}\big) \leqslant 2^{-k}$ pour tout $k \in \mathbb{N}$ (exo 1). Par Borel-Cantelli, on a alors $|Z_{n_{k+1}}(\omega) - Z_{n_k}(\omega)| < 2^{-k}$ à partir d'un certain rang pour presque tout $\omega \in \Omega$ (exo 2) ; donc la série $\sum |Z_{n_{k+1}} - Z_{n_k}|$ converge presque sûrement, et par conséquent la *suite* (Z_{n_k}) converge presque sûrement. Ainsi (Z_n) possède une sous-suite (Z_{n_k}) qui converge presque sûrement, et donc en probabilité (*cf* plus bas), vers une certaine va Z. Comme (Z_n) est toujours de Cauchy en probabilité, on en déduit que $Z_n \xrightarrow{\text{proba}} Z$ (exo 3). □

Proposition 2.2. *La convergence presque sûre et la convergence en norme L^1 entraînent chacune la convergence en probabilité.*

Démonstration. Supposons que $Z_n \xrightarrow{\text{ps}} Z$, et fixons $\varepsilon > 0$. Comme $\mathbb{P}(\overline{\lim}\, E_n) \geqslant \overline{\lim}\, \mathbb{P}(E_n)$ pour toute suite d'évènements (E_n) (exo déjà posé, qu'il est temps de faire), on a

$$\overline{\lim}\ \mathbb{P}\big(|Z_n - Z| \geqslant \varepsilon\big) \leqslant \mathbb{P}\big(\overline{\lim}\, \{|Z_n - Z| \geqslant \varepsilon\}\big).$$

Mais comme $Z_n \xrightarrow{\text{ps}} Z$, on a $|Z_n(\omega) - Z(\omega)| < \varepsilon$ à partir d'un certain rang pour presque tout $\omega \in \Omega$, et donc $\mathbb{P}\big(\overline{\lim}\, \{|Z_n - Z| \geqslant \varepsilon\}\big) = 0$. Donc $\overline{\lim}\, \mathbb{P}\big(|Z_n - Z| \geqslant \varepsilon\big) = 0$, autrement dit $\mathbb{P}\big(|Z_n - Z| \geqslant \varepsilon\big) \to 0$ pour tout $\varepsilon > 0$. Ainsi $Z_n \xrightarrow{\text{proba}} Z$.

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. Par exemple, si $(Z_n)_{n \geqslant 1}$ est une suite de va indépendantes avec $\mathbb{P}(Z_n = n) = \frac{1}{n}$ et $\mathbb{P}(Z_n = 0) = 1 - \frac{1}{n}$ pour tout $n \geqslant 1$, alors $Z_n \xrightarrow{\text{proba}} 0$, mais Z_n ne tend vers 0 ni en norme L^1, ni presque sûrement (exo).

La proposition suivante caractérise complètement la converge en probabilité en termes de convergence presque sûre.

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 par conséquent $Z'_{n_k} \xrightarrow{\text{ps}} Z$.

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, et *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, alors (Z_n) possède une sous-suite qui converge presque sûrement.*

Démonstration. C'est évident par la propostion. □

COROLLAIRE 2.5. *Si (Z_n) est une suite de va* bornée dans L^∞ *et si $Z_n \xrightarrow{\text{proba}} Z$, alors $Z \in L^\infty$ et $Z_n \xrightarrow{L^p} Z$ pour tout $1 \leqslant p < \infty$.*

Démonstration. Soit $M := \sup_{n\in\mathbb{N}} \|Z_n\|_\infty$, qui est $< \infty$ par hypothèse. Par définition de $\|\cdot\|_\infty$, on a $|Z_n(\omega)| \leqslant M$ ps pour tout $n \in \mathbb{N}$, donc

$$(*) \qquad \big(\forall n \in \mathbb{N} \ : \ |Z_n(\omega)| \leqslant M\big) \quad \text{ps},$$

car une conjonction dénombrable de propriétés presque sûres est encore presque sûre. Par ailleurs, comme $Z_n \xrightarrow{\text{proba}} Z$, on peut trouver une sous-suite (Z_{n_k}) de (Z_n) qui converge presque sûrement vers Z. Par $(*)$, on en déduit que $|Z(\omega)| \leqslant M$ ps, et donc $Z \in L^\infty$.

Fixons $p < \infty$, et supposons que (Z_n) ne tende pas vers Z en norme L^p. Alors on peut trouver $\varepsilon_0 > 0$ et une sous-suite (Z'_n) de (Z_n) tels que

$$(2.1) \qquad \|Z'_n - Z\|_p \geqslant \varepsilon_0 \qquad \text{pour tout } n \in \mathbb{N}.$$

Comme $Z_n \xrightarrow{\text{proba}} Z$, la suite (Z'_n) possède une sous-suite (Z'_{n_k}) qui converge presque sûrement vers Z. De plus, on a $|Z'_{n_k}| \leqslant M$ ps pour tout $k \in \mathbb{N}$, et la constante M appartient à L^p. Par convergence dominée dans L^p (Proposition 1.5), on en déduit que $Z'_{n_k} \to Z$ en norme L^p ; ce qui contredit (2.1).

□

Corollaire 2.6. *Si* $Z_n \xrightarrow{\text{proba}} Z$*, alors* $f(Z_n) \xrightarrow{\text{proba}} f(Z)$ *pour toute fonction* continue $f : \mathbb{R} \to \mathbb{R}$.

Démonstration. Posons $Y_n := f(Z_n)$. Si (Y'_n) est une sous-suite de (Y_n), alors $Y'_n = f(Z'_n)$ pour une certaine sous-suite (Z'_n) de (Z_n). Comme $Z_n \xrightarrow{\text{proba}} Z$, la suite (Z'_n) possède une sous-suite (Z'_{n_k}) telle que $Z'_{n_k} \xrightarrow{\text{ps}} Z$; et comme f est continue, $Y'_{n_k} = f(Z'_{n_k}) \xrightarrow{\text{ps}} f(Z)$. Ainsi, toute sous-suite de (Y_n) possède une sous-suite qui tend presque sûrement vers $f(Z)$, et donc $Y_n \xrightarrow{\text{proba}} f(Z)$. □

3. Séries de va indépendantes

Vu ce qu'on a raconté dans les sections précédentes, le résultat suivant est *a priori* assez surprenant. 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. On la présente sous la forme d'un long exercice.

(1) Soit $\varepsilon > 0$, et soit $n \in \mathbb{N}$. On définit une va $\tau_{n,\varepsilon} : \Omega \to \mathbb{N} \cup \{\infty\}$ en posant
$$\tau_{n,\varepsilon}(\omega) := \min \big\{p > n;\ |S_p(\omega) - S_n(\omega)| \geqslant \varepsilon\big\},$$
avec la convention $\min \varnothing = \infty$. Le but de cette question est de montrer que
$$\mathbb{P}(\tau_{n,\varepsilon} < \infty) \to 0 \qquad \text{quand } n \to \infty.$$

   (a) Pour $\eta > 0$, justifier l'existence d'un entier N_η tel que
   $$\forall m, m' \geqslant N_\eta\ :\ \mathbb{P}\big(|S_{m'} - S_m| \geqslant \varepsilon/2\big) \leqslant \eta.$$
   (b) Avec les notations de (a), montrer que pour tous $q > n \geqslant N_\eta$, on a
   $$\sum_{p=n+1}^{q} \mathbb{P}\big(\tau_{n,\varepsilon} = p\,,\ |S_q - S_p| < \varepsilon/2\big) \leqslant \eta.$$
   (c) Montrer que si $n < p \leqslant q$, alors les deux évènements $(\tau_{n,\varepsilon} = p)$ et $(|S_q - S_p| < \varepsilon/2)$ sont indépendants.
   (d) Déduire des questions précédentes que si $0 < \eta < 1$ et $n \geqslant N_\eta$, alors
   $$\eta \geqslant (1 - \eta)\, \mathbb{P}(\tau_{n,\varepsilon} < \infty).$$
   (e) Démontrer le résultat souhaité.

(2) Pour $k \in \mathbb{N}^*$ et $n \in \mathbb{N}$, on pose
$$E_{n,k} := \big\{\omega \in \Omega;\ \exists p > n\ :\ |S_p(\omega) - S_n(\omega)| \geqslant 1/k\big\}.$$
Montrer que pour tout k fixé, $\mathbb{P}(E_{n,k}) \to 0$ quand $n \to \infty$.

(3) Montrer que la suite (S_n) vérifie presque sûrement le critère de Cauchy, et conclure.

□

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 entraînent la convergence en norme L^2 de la série $\sum X_k$ (Exemple 1.4). Comme la convergence L^2 entraîne la convergence en probabilité, on en déduit que la série $\sum X_k$ converge en probabilité, *et donc* presque sûrement d'après le Théorème 3.1 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. 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*

$$\int_\Lambda f\, d\mu_n \xrightarrow{n\to\infty} \int_\Lambda f\, d\mu \qquad \textit{pour toute } f \in \mathcal{C}_b(\Lambda).$$

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 $(\lambda_n) \subseteq]0, \infty[$ une suite tendant vers ∞. Alors $\mu_n := \lambda_n^d \rho(\lambda_n x)dx$ converge étroitement vers $\mu := \delta_0$.

Démonstration. Si on pose $\rho_n(x) = \lambda_n \rho(\lambda_n x)$, alors (ρ_n) est une *suite de Dirac* dans $L^1(\mathbb{R}^d)$ (*cf* le cours d'intégration). Donc $\rho_n * g(x) \to g(x)$ pour toute $g \in \mathcal{C}_b(\mathbb{R}^d)$ et pour tout $x \in \mathbb{R}^d$. En particulier, si $f \in \mathcal{C}_b(\mathbb{R}^d)$ et si on pose $f^-(t) := f(-t)$, alors

$$\int_{\mathbb{R}^d} f\, d\mu_n = \int_{\mathbb{R}^d} f^-(-t)\, \rho_n(t)dt = \rho_n * f^-(0) \xrightarrow{n\to\infty} f^-(0) = f(0) = \int_{\mathbb{R}^d} f\, d\delta_0.$$

□

Remarque. Sans invoquer aucun résultat sur la convolution, on peut simplement écrire

$$\begin{aligned}\int_{\mathbb{R}^d} f\, d\mu_n &= \int_{\mathbb{R}^d} f(x)\, \rho(\lambda_n x)\, \lambda_n^d dx \\ &= \int_{\mathbb{R}^d} f\Big(\frac{u}{\lambda_n}\Big)\rho_n(u)du \quad \text{en posant } u := \lambda_n x\end{aligned}$$

et conclure, par continuité de f en 0 et convergence dominée, que

$$\int_{\mathbb{R}^d} f\, d\mu_n \to \int_{\mathbb{R}^d} f(0)\, \rho(u)du = f(0).$$

Le résultat suivant s'appelle assez bizarrement le **Théorème porte-manteau**.

Proposition 4.3. *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 une 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′) $\implies$ (3). 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). Supposons (4) vérifiée, et fixons $f \in \mathcal{C}_b(\Lambda)$: on veut montrer que $\int_\Lambda f \, d\mu_n \to \int_\Lambda f \, d\mu$.

Fait 1. L'ensemble $D := \big\{t \in \mathbb{R};\ \mu(f^{-1}(\{t\})) > 0\big\}$ est dénombrable.

Preuve du Fait 1. On définit une mesure borélienne ν sur $\mathbb{R}$ en posant $\nu(A) := \mu(f^{-1}(A))$ (exo), et la mesure ν est *finie* car μ est finie. Donc $D = \{t \in \mathbb{R};\ \nu(\{t\}) > 0\}$ est dénombrable par le Fait 2.4 du Chapitre 2. □

Fait 2. Pour tout $\varepsilon > 0$ donné, on peut trouver des nombres réels $\alpha_0 < \alpha_1 < \ldots < \alpha_N$ vérifiant les propriétés suivantes :

(i) $\mu\big(f^{-1}(\{\alpha_k\})\big) = 0$ pour $k = 0, \ldots, N-1$;

(ii) $\alpha_{k+1} - \alpha_k < \varepsilon$ pour $k = 0, \ldots, N-1$;

(iii) $\alpha_0 \leqslant f(x) < \alpha_N$ pour tout $x \in \Lambda$.

Preuve du Fait 2. Soit M une constante telle que $-M \leqslant f \leqslant M$. (Une telle constante M existe puisque f est bornée.) Par le Fait 1 et comme l'intervalle $]-\infty, M]$ est non dénombrable, on peut trouver $\alpha_0 \leqslant M$ tel que $\mu(f = \alpha_0) = 0$. Ensuite, toujours par le Fait 1, on peut trouver α_1 tel que $\alpha_0 + \varepsilon/2 < \alpha_1 < \alpha_0 + \varepsilon$ et $\mu(f = \alpha_1) = 0$; puis α_2 tel que $\alpha_1 + \varepsilon/2 < \alpha_2 < \alpha_1 + \varepsilon$, et ainsi de suite. Si N est assez grand, alors $\alpha_N > \alpha_0 + N\varepsilon/2 \geqslant M$, donc (i), (ii) *et aussi* (iii) sont vérifiées. □

Fait 3. Pour tout $\varepsilon > 0$, on peut trouver une fonction borélienne bornée $\varphi : \Lambda \to \mathbb{R}$ telle que

$$|\varphi(x) - f(x)| < \varepsilon \quad \text{pour tout } x \in \Lambda \qquad \text{et} \qquad \int_\Lambda \varphi \, d\mu_n \xrightarrow{n\to\infty} \int_\Lambda \varphi \, d\mu.$$

Preuve du Fait 3. Soient $\alpha_0 < \ldots < \alpha_N$ donnés par le Fait 2. Posons

$$\varphi := \sum_{k=0}^{N-1} \alpha_k \mathbf{1}_{B_k} \quad , \qquad \text{avec} \quad B_k := f^{-1}\big([\alpha_k, \alpha_{k+1}[\big).$$

Comme $\alpha_0 \leqslant f < \alpha_N$, les B_k forment une partition de Λ ; et par définition de φ, on a $0 \leqslant f(x) - \varphi(x) < \alpha_{k+1} - \alpha_k < \varepsilon$ pour tout k et pour tout $x \in B_k$. Donc

$$|\varphi(x) - f(x)| < \varepsilon \qquad \text{pour tout } x \in \Lambda.$$

De plus, comme f est *continue*, on a $\overline{B_k} \subseteq f^{-1}([\alpha_k, \alpha_{k+1}])$ et $f^{-1}(]\alpha_k, \alpha_{k+1}[) \subseteq \mathring{B}_k$ pour tout $k \in [\![0, N-1]\!]$. Donc $\partial B_k \subseteq f^{-1}(\{\alpha_k\}) \cup f^{-1}(\{\alpha_{k+1}\})$, et par conséquent $\mu(\partial B_k) = 0$. Par (4), on en déduit que

$$\int_\Lambda \varphi \, d\mu_n = \sum_{k=0}^{N-1} \alpha_k \mu_n(B_k) \xrightarrow{n\to\infty} \sum_{k=0}^{N-1} \alpha_k \mu(B_k) = \int_\Lambda \varphi \, d\mu.$$

□

On peut maintenant conclure la preuve. Soit $\varepsilon > 0$ quelconque. Par le Fait 3, on peut trouver une fonction $\varphi : \Lambda \to \mathbb{R}$ telle que

$$|\varphi(x) - f(x)| < \varepsilon \quad \text{pour tout } x \in \Lambda \qquad \text{et} \qquad \int_\Lambda \varphi \, d\mu_n \xrightarrow{n\to\infty} \int_\Lambda \varphi \, d\mu.$$

On a alors $\int_\Lambda |\varphi - f| d\mu \leqslant \varepsilon \mu(\Lambda)$, et $\int_\Lambda |\varphi - f| d\mu_n \leqslant \varepsilon \mu_n(\Lambda)$ pour tout $n \in \mathbb{N}$. En écrivant

$$\int_\Lambda f \, d\mu_n - \int_\Lambda f \, d\mu = \int_\Lambda (f - \varphi) \, d\mu_n + \left(\int_\Lambda \varphi \, d\mu_n - \int_\Lambda \varphi \, d\mu \right) + \int_\Lambda (\varphi - f) \, d\mu,$$

on en déduit

$$\left| \int_\Lambda f \, d\mu_n - \int_\Lambda f \, d\mu \right| \leqslant \varepsilon \mu_n(\Lambda) + \left| \int_\Lambda \varphi \, d\mu_n - \int_\Lambda \varphi \, d\mu \right| + \varepsilon \mu(\Lambda). \tag{4.1}$$

Enfin, $\mu_n(\Lambda) \to \mu(\Lambda)$ par (4) à nouveau, car $\partial\Lambda = \varnothing$. En faisant tendre n vers l'infini dans (4.1), on en déduit

$$\overline{\lim} \left| \int_\Lambda f \, d\mu_n - \int_\Lambda f \, d\mu \right| \leqslant 2\varepsilon \mu(\Lambda) \qquad \text{pour tout } \varepsilon > 0,$$

ce qui prouve que $\int_\Lambda f \, d\mu_n \to \int_\Lambda f \, d\mu$. □

Corollaire 4.4. *Si $\mu_n(B) \to \mu(B)$ pour tout borélien $B \subseteq \Lambda$, alors $\mu_n \twoheadrightarrow \mu$.*

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 d'un espace Λ discret.*

Définition 4.5. *L'espace métrique Λ est dit **topologiquement discret** si toute suite convergente d'éléments de Λ est constante à partir d'un certain rang.*

Exemples. Tout espace métrique fini est topologiquement discret ; $\mathbb{N}$ est topologiquement discret ; $\mathbb{R}$ et $\mathbb{Q}$ ne sont pas topologiquement discrets (exo).

Remarque 4.6. Λ est topologiquement discret si et seulement si tout ensemble $B \subseteq \Lambda$ est à la fois ouvert et fermé dans Λ.

Démonstration. Si Λ est topologiquement discret, alors on voit immédiatement que tout ensemble $B \subseteq \Lambda$ est fermé (et donc aussi ouvert puisque $\Lambda \backslash B$ est fermé) : si (x_n) est une suite de points de B convergeant vers $x \in \Lambda$, alors $x_n = x$ à partir d'un certain rang, et donc $x \in B$. Inversement, supposons que tout $B \subseteq \Lambda$ soit ouvert. Si (x_n) est une suite d'éléments de Λ convergeant vers $x \in \Lambda$, alors, comme le singleton $\{x\}$ est ouvert et contient x (!), on a $x_n \in \{x\}$ à partir d'un certain rang, *i.e.* $x_n = x$ à partir d'un certain rang. □

PROPOSITION 4.7. *Supposons que l'espace métrique* Λ *soit* dénombrable *et* topologiquement discret. *Soit* (μ_n) *une suite dans* $\mathrm{M}_+(\Lambda)$*, et soit* $\mu \in \mathrm{M}_+(\Lambda)$*. Alors les propriétés suivantes sont équivalentes :*

(1) $\mu_n \twoheadrightarrow \mu$ *;*

(4') $\mu_n(B) \to \mu(B)$ *pour tout ensemble* $B \subseteq \Lambda$ *;*

(5) $\mu_n(\Lambda) \to \mu(\Lambda)$ *et* $\mu_n(\{x\}) \to \mu(\{x\})$ *pour tout* $x \in \Lambda$*.*

Démonstration. Comme Λ est topologiquement discret, tout ensemble $B \subseteq \Lambda$ est ouvert et fermé. Donc $\partial B = \overline{B} \backslash \mathring{B} = \varnothing$ pour tout $B \subseteq \Lambda$. Donc (4') est en fait la même condition que (4) dans la Proposition 4.3, et donc (1) $\Longleftrightarrow$ (4'). Évidemment (4') $\Longrightarrow$ (5), donc il reste à montrer que (5) $\Longrightarrow$ (4').

Supposons (5) vérifiée. Alors $\mu_n(F) \to \mu(F)$ pour tout ensemble *fini* $F \subseteq \Lambda$: en effet, si $F = \{a_1, \ldots, a_K\}$, 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 $\mu_n(B) \to \mu(B)$. □

REMARQUE. L'implication (5) $\Longrightarrow$ (4') est valable dans tout espace métrique dénombrable Λ (pas forcément topologiquement discret), et (4') $\Longrightarrow$ (1) vaut dans n'importe quel espace métrique Λ.

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 $\mathcal{C}_{00}^{\infty}(\mathbb{R})$ l'ensemble des fonctions de classe $\mathcal{C}^{\infty}$ à support compact. On a ainsi

$$\mathcal{C}_{00}^{\infty}(\mathbb{R}) \subseteq \mathcal{C}_{00}(\mathbb{R}) \subseteq \mathcal{C}_b(\mathbb{R}).$$

PROPOSITION 4.8. *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) $\mu_n \twoheadrightarrow \mu$;

(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* $\varphi \in \mathcal{C}_{00}(\mathbb{R})$.

(7) $\mu_n(\mathbb{R}) \to \mu(\mathbb{R})$ *et* $\int_{\mathbb{R}} \varphi\, d\mu_n \to \int_{\mathbb{R}} \varphi\, d\mu$ *pour toute fonction* $\varphi \in \mathcal{C}_{00}^{\infty}(\mathbb{R})$.

Démonstration. Évidemment, (1) $\implies$ (6) $\implies$ (7). De plus, on *sait* (*cf le cours d'intégration*) que $\mathcal{C}_{00}^{\infty}(\mathbb{R})$ 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 $\widetilde{\varphi} \in \mathcal{C}_{00}^{\infty}(\mathbb{R})$ telle que

$$|\widetilde{\varphi}(x) - \varphi(x)| < \varepsilon \qquad \text{pour tout } x \in \mathbb{R}.$$

En raisonnant comme dans la preuve de (4) $\implies$ (1) de la Proposition 4.3, on en déduit que (7) $\implies$ (6). Donc il reste à montrer que (6) $\implies$ (1).

Supposons (6) 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 que $\mathbb{R}$ 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 (6), 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 (comme dans la preuve de (4) $\implies$ (1) de la Proposition 4.3)

$$\left| \int_{\mathbb{R}} f\, d\mu_n - \int_{\mathbb{R}} f\, d\mu \right| \leqslant 2\varepsilon\, \|f\|_\infty + \left| \int_{\mathbb{R}} \varphi\, d\mu_n - \int_{\mathbb{R}} \varphi\, d\mu \right| \qquad \text{pour tout } n \in \mathbb{N}.$$

Comme $\int_{\mathbb{R}} \varphi\, d\mu_n \to \int_{\mathbb{R}} \varphi\, d\mu$ par (6) 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 4.9. *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 espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$.

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 $\lambda_n := \frac{1}{\sigma_n} \to \infty$ (*cf* l'Exemple 2 après la définition de la convergence étroite). □

EXERCICE. On suppose que Z et les Z_n sont des va réelles. Montrer que 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 $\phi(Z_n) \xrightarrow{\mathcal{L}} \phi(Z)$ pour toute fonction continue $\phi : \mathbb{R} \to \mathbb{R}$.

PROPOSITION 4.10. *La convergence en probabilité entraîne la convergence en loi : si Z et les Z_n sont des va réelles définies sur le même espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$ 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 $\mathbb{E}(f(Z_n)) \to \mathbb{E}(f(Z))$. Par le Corollaire 2.6, on sait que $f(Z_n) \xrightarrow{\text{proba}} f(Z)$. Ensuite, on applique le Corollaire 2.5 : comme la fonction f est bornée, la suite $(f(Z_n))$ est bornée dans L^∞ ; et comme $f(Z_n) \xrightarrow{\text{proba}} f(Z)$, on en déduit que $f(Z_n) \to f(Z)$ en norme L^1. En particulier, $\mathbb{E}(f(Z_n)) \to \mathbb{E}(f(Z))$. □

Exercice. Soit Z une va suivant une loi de Bernoulli $\mathcal{B}(1/2)$, et soit $Z_n := 1 - Z$ pour tout $n \in \mathbb{N}$. Montrer que (Z_n) converge vers Z en loi mais pas en probabilité.

PROPOSITION 4.11. *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 4.12. *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 $\mathbb{P}_Z(\partial I) = 0$. $\square$

COROLLAIRE 4.13. *Si les Z_n sont des va réelles définies sur le même espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$ 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 $\mathbb{P}(|Z_n - a| \geqslant \varepsilon) = \mathbb{P}(Z_n \in B_\varepsilon) \to \mathbb{P}(Z \in B_\varepsilon) = \delta_a(B_\varepsilon) = 0$. $\square$

COROLLAIRE 4.14. *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. $\square$

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. $\square$

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.

PROPOSITION 4.15. *Supposons que Z et les Z_n soient à valeurs dans un espace métrique Λ dénombrable et topologiquement discret. Alors $Z_n \xrightarrow{\mathcal{L}} Z$ si et seulement si $\mathbb{P}(Z_n = x) \to \mathbb{P}(Z = x)$ pour tout $x \in \Lambda$; et dans ce cas $\mathbb{P}(Z_n \in B) \to \mathbb{P}(Z \in B)$ pour tout $B \subseteq \Lambda$.*

Démonstration. Comme $\mathbb{P}_{Z_n}(\Lambda) = 1 = \mathbb{P}_Z(\Lambda)$ pour tout $n \in \mathbb{N}$, c'est une conséquence de la Proposition 4.7. $\square$

COROLLAIRE 4.16. (Théorème de Poisson)
Supposons que Z_n suive une loi binomiale $\mathfrak{B}(n, p_n)$, où $p_n \sim \frac{\lambda}{n}$. 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 4.17. *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. Conséquence de la Proposition 4.8. □

Voici pour finir un peu de nouveauté.

Proposition 4.18. *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 ainsi $1 - F_{Z_n}(t) \to 1 - F_Z(t)$ *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 4.19. *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 évident par la proposition et la définition d'une fonction de répartition. □

Chapitre 8

Théorèmes limites

1. Loi des grands nombres

On a vu au Chapitre 6 que 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. Ce résultat s'appelle la **Loi des grands nombres**.

Théorème 1.1. *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 $M_n\xrightarrow{\text{ps}}\infty$ 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, on a $M_n = \Phi(Y_1, \ldots, Y_n)$ où $\Phi : \mathbb{R}^n \to \mathbb{R}$ est la fonction définie par $\Phi(u_1, \ldots, u_n) := \max(u_1, u_1 + u_2, \ldots, u_1 + \cdots + u_n)$, et $M_n^* = \Phi(Y_2, \ldots, Y_{n+1})$ pour la même fonction Φ. Comme les va $(Y_1, \ldots, Y_n)$ et $(Y_2, \ldots, Y_{n+1})$ ont la même loi, à savoir $\mathbb{P}_Y \otimes \cdots \otimes \mathbb{P}_Y$, cela montre bien que $\mathbb{P}_{M_n^*} = \mathbb{P}_{M_n}$. 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

$$Y_1 + M_n^* = \max(\Sigma_2, \ldots, \Sigma_{n+1});$$

et donc

$$M_{n+1} = \max(\Sigma_1, Y_1 + M_n^*) = \max(Y_1, Y_1 + M_n^*).$$

Autrement dit,

$$M_{n+1} = Y_1 + M_n^* - \min(0, M_n^*).$$

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

$$\mathbb{E}(Y_1) \geqslant \mathbb{E}(\min(0, M_n^*)\big).$$

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. $\square$

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. $\square$

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

$$\text{presque sûrement} \quad : \quad \left(\overline{\lim}\, \frac{S_n}{n} \leqslant \frac{1}{k} \quad \text{pour tout } k \in \mathbb{N}^*\right).$$

Autrement dit :

$$\overline{\lim}\, \frac{S_n}{n} \leqslant 0 \qquad \text{ps.}$$

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

$$\underline{\lim}\,\frac{S_n}{n} = -\overline{\lim}\,\frac{(-S_n)}{n} \geqslant 0 \qquad \text{ps.}$$

□

REMARQUE. On peut montrer que l'hypothèse "$X \in L^1$" est *nécessaire* pour avoir la convergence presque sûre de $\frac{S_n}{n}$.

2. Lois des grands nombres L^2

Le "défaut" du Théorème 1.1 est qu'il ne s'applique qu'à des va X_i *indépendantes et de même loi.* Dans cette section, on va démontrer des résultats du même type où on s'affranchit un peu de ces hypothèses. Le prix à payer est qu'on doit supposer que les X_i sont dans L^2 ; et, surtout, on n'obtient plus de la convergence presque sûre mais "seulement" de la convergence en norme L^2 (et donc, quand même, de la convergence en probabilité). Pour cette raison, on parle parfois de lois "faibles" des grands nombres.

Rappelons la définition de la covariance de deux va, dont il a déjà été question au Chapitre 6.

DÉFINITION 2.1. *Soient X et Y deux va réelles définies sur le même espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$ et appartenant à L^2. La **covariance** de X et Y est le nombre $\mathrm{Cov}(X,Y)$ défini par*

$$\mathrm{Cov}(X,Y) := \langle X - \mathbb{E}(X), Y - \mathbb{E}(Y)\rangle_{L^2} = \mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big).$$

*On dit que X et Y sont **non corrélées** si on a $\mathrm{Cov}(X,Y) = 0$.*

REMARQUE 1. On a $\mathrm{Cov}(X,X) = \sigma^2(X)$ pour toute va $X \in L^2$.

REMARQUE 2. 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. Supposons d'abord que les X_i sont *centrées.* Alors $m = 0$, donc il s'agit de montrer que $\frac{S_n}{n} \xrightarrow{L^2} 0$.

On a

$$\begin{aligned}\|S_n\|_{L^2}^2 &= \sum_{i,j=1}^{n} \langle X_i, X_j\rangle_{L^2} \\ &= \sum_{i,j=1}^{n} \operatorname{Cov}(X_i, X_j) \quad \text{car les } X_i \text{ sont centrées.}\end{aligned}$$

Donc le résultat est immédiat par (ii) !

Pour le cas général, on applique le cas “centré” aux va $Y_i := X_i - \mathbb{E}(X_i)$, qui vérifient (ii) car $\operatorname{Cov}(Y_i, Y_j) = \operatorname{Cov}(X_i, X_j)$, et qui vérifient (i) avec $m = 0$. La conclusion est que

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(X_i) \xrightarrow{L^2} 0;$$

d'où le résultat puisque $\frac{1}{n}\sum\limits_{i=1}^{n} \mathbb{E}(X_i) \to m$.

□

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 $\operatorname{Cov}(X_i, X_j) = 0$ si $i \neq j$ et $\operatorname{Cov}(X_i, X_i) = \sigma^2(X_i) = \sigma^2$, on a

$$\sum_{i,j=1}^{n} \operatorname{Cov}(X_i, X_j) = \sum_{i=1}^{n} \sigma^2(X_i) = n\sigma^2 \quad \text{pour tout } n \geqslant 1;$$

donc (ii) est également satisfaite □

Définition 2.4. *Soient X et Y deux va réelles définies sur le même espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$ et appartenant à L^2. Le* ***coefficient de corrélation*** *de X et Y est le nombre $\operatorname{Cor}(X, Y)$ défini par*

$$\operatorname{Cor}(X, Y) := \frac{\operatorname{Cov}(X, Y)}{\sigma(X)\sigma(Y)},$$

en convenant que $\operatorname{Cor}(X, Y) = 0$ si $\sigma(X) = 0$ ou $\sigma(Y) = 0$ (ce qui revient à dire que X ou Y est presque sûrement égale à une constante).

Remarque. On a toujours $|\operatorname{Cor}(X, Y)| \leqslant 1$ par l'inégalité de Cauchy-Schwarz ; et $\operatorname{Cor}(X, X) = 1$ pour toute va X non presque sûrement égale à une constante.

Théorème 2.5. *Soit $(X_i)_{i\geqslant 1}$ une suite de va réelles définies sur le même espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$ et appartenant à L^2. Pour $n \geqslant 1$, on pose*

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

On fait les hypothèses suivantes :

(a) $\frac{1}{n}\sum\limits_{i=1}^{n} \mathbb{E}(X_i)$ *admet une limite m quand $n \to \infty$;*

(b) $\sum_{i=1}^{n} \sigma^2(X_i) = O(n)$ *quand* $n \to \infty$;

(c) *pour* $i \neq j$, *on a une majoration de la forme* $|\mathrm{Cor}(X_i, X_j)| \leqslant r(|i-j|)$, *où* $r : \mathbb{N} \to \mathbb{R}^+$ *vérifie que* $\sum_{k=1}^{n-1} r(k) = o(n)$ *quand* $n \to \infty$.

Alors $\frac{S_n}{n} \to m$ *en norme* L^2, *et donc en probabilité.*

Démonstration. D'après la Proposition 2.2, il suffit de montrer que

$$\sum_{i,j=1}^{n} \mathrm{Cov}(X_i, X_j) = o(n^2) \qquad \text{quand } n \to \infty.$$

Par définition, on a $\mathrm{Cov}(X_i, X_i) = \sigma^2(X_i)$ pour tout i, et

$$\begin{aligned} |\mathrm{Cov}(X_i, X_j)| &= |\sigma(X_i)\sigma(X_j)\mathrm{Cor}(X_i, X_j)| \\ &\leqslant \sigma(X_i)\sigma(X_j)\, r(|i-j|) \qquad \text{si } i \neq j. \end{aligned}$$

Donc

$$\begin{aligned} \left|\sum_{i,j=1}^{n} \mathrm{Cov}(X_i, X_j)\right| &\leqslant \sum_{i=1}^{n} \sigma^2(X_i) + \sum_{i \neq j} \sigma(X_i)\sigma(X_j)\, r(|i-j|) \\ &= \sum_{i=1}^{n} \sigma^2(X_i) + \sum_{k=1}^{n-1} r(k) \sum_{|i-j|=k} \sigma(X_i)\sigma(X_j) \\ &= \sum_{i=1}^{n} \sigma^2(X_i) + 2\sum_{k=1}^{n-1} r(k) \sum_{i=1}^{n-k} \sigma(X_i)\sigma(X_{i+k}). \end{aligned}$$

De plus, par l'inégalité de Cauchy-Schwarz on a aussi

$$\begin{aligned} \sum_{i=1}^{n-k} \sigma(X_i)\sigma(X_{i+k}) &\leqslant \Big(\sum_{i=1}^{n-k} \sigma^2(X_i)\Big)^{1/2} \Big(\sum_{i=1}^{n-k} \sigma^2(X_{i+k})\Big)^{1/2} \\ &\leqslant \sum_{i=1}^{n} \sigma^2(X_i) \qquad \text{pour tout } k \in [\![1, \dots, n-1]\!]. \end{aligned}$$

En posant $\sigma_n^2 := \sum_{i=1}^{n} \sigma^2(X_i)$, on obtient donc

$$\left|\sum_{i,j=1}^{n} \mathrm{Cov}(X_i, X_j)\right| \leqslant \sigma_n^2 + 2\sigma_n^2 \sum_{k=1}^{n-1} r(k);$$

et par (b) et (c), on en déduit que $\sum_{i,j=1}^{n} \mathrm{Cov}(X_i, X_j) = o(n^2)$. ☐

3. Théorème des évènements rares

Le résultat suivant généralise le Théorème de Poisson vu au Chapitre 7 (Corollaire 4.16).

THÉORÈME 3.1. *Pour tout* $n \in \mathbb{N}^*$, *soient* $X_{n,1}, \dots, X_{n,M_n}$ *des va indépendantes à valeurs dans* $\{0, 1\}$, *chaque* $X_{n,i}$ *suivant une loi de Bernoulli* $\mathcal{B}(p_{n,i})$. *On fait les hypothèses suivantes :*

(i) $\lambda_n := \sum_{i=1}^{M_n} p_{n,i}$ *admet une limite* $\lambda > 0$ *quand* $n \to \infty$ *;*

(ii) $\varepsilon_n := \max(p_{n,1}, \ldots, p_{n,M_n})$ *tend vers* 0 *quand* $n \to \infty$.

Alors, si on pose $S_n := X_{n,1} + \cdots + X_{n,M_n}$*, la suite* (S_n) *converge en loi vers une va* Z *suivant la loi de Poisson* $\mathcal{P}(\lambda)$*.*

Exercice. Montrer que le théorème de Poisson est effectivement une conséquence de ce résultat.

Preuve du théorème. On aura besoin de la notation suivante : si μ et μ' sont deux lois de probabilité sur $\mathbb{N}$, on pose

$$\|\mu - \mu'\| := \sum_{k=0}^{\infty} \big|\mu(\{k\}) - \mu'(\{k\})\big|.$$

Autrement dit, si Xet X' sont deux va à valeurs dans $\mathbb{N}$, alors

$$\|\mathbb{P}_X - \mathbb{P}_{X'}\| = \sum_{k=0}^{\infty} \big|\mathbb{P}(X = k) - \mathbb{P}(X' = k)\big|.$$

Fait 1. Soient $X_1, \ldots, X_M$, $X'_1, \ldots, X'_M$ des va à valeurs dans $\mathbb{N}$. On suppose que les X_i sont indépendantes, et que les X'_i sont indépendantes. Alors, si on pose $S := X_1 + \cdots + X_M$ et $S' := X'_1 + \cdots + X'_M$, on a

$$\|\mathbb{P}_S - \mathbb{P}_{S'}\| \leqslant \sum_{i=1}^{M} \|\mathbb{P}_{X_i} - \mathbb{P}_{X'_i}\|.$$

Preuve du Fait 1. Il suffit de traiter le cas où $M = 2$: le cas général s'en déduit par récurrence (exo). On a donc affaire à 4 va X_1, X_2, X'_1, X'_2, avec $S = X_1 + X_2$ et $S' = X'_1 + X'_2$. Pour alléger les écritures, on posera pour $i = 1, 2$ et $k \in \mathbb{N}$:

$$\rho_{i,k} := \mathbb{P}(X_i = k) \qquad \text{et} \qquad \rho'_{i,k} := \mathbb{P}(X'_i = k).$$

Comme X_1 et X_2 sont indépendantes, on sait que pour tout $k \in \mathbb{N}$, on a

$$\mathbb{P}(S = k) = \sum_{s=0}^{k} \mathbb{P}(X_1 = s)\, \mathbb{P}(X_2 = k - s) = \sum_{s=0}^{k} \rho_{1,s}\rho_{2,k-s};$$

et de même

$$\mathbb{P}(S' = k) = \sum_{s=0}^{k} \rho'_{1,s}\rho'_{2,k-s}.$$

Donc

$$\begin{aligned}
\|\mathbb{P}_S - \mathbb{P}_{S'}\| &= \sum_{k=0}^{\infty} \Big| \sum_{s=0}^{k} \rho_{1,s}\rho_{2,k-s} - \rho'_{1,s}\rho'_{2,k-s} \Big| \\
&\leqslant \sum_{k=0}^{\infty} \sum_{s=0}^{k} \big|\rho_{1,s}\rho_{2,k-s} - \rho'_{1,s}\rho'_{2,k-s}\big|.
\end{aligned}$$

De plus, pour tous $0 \leqslant s \leqslant k$, on peut majorer $|\rho_{1,s}\rho_{2,k-s} - \rho'_{1,s}\rho'_{2,k-s}|$ comme suit :

$$\Big|\rho_{1,s}\rho_{2,k-s} - \rho'_{1,s}\rho'_{2,k-s}\Big| \leqslant \rho_{1,s}\, |\rho_{2,k-s} - \rho'_{2,k-s}| + \rho'_{2,k-s}\, |\rho_{1,s} - \rho'_{1,s}|.$$

On obtient donc

$$\|\mathbb{P}_S - \mathbb{P}_{S'}\| \leqslant \sum_{k=0}^{\infty}\sum_{s=0}^{k} \rho_{1,s}\,|\rho_{2,k-s} - \rho'_{2,k-s}| + \sum_{k=0}^{\infty}\sum_{s=0}^{k} |\rho_{1,s} - \rho'_{1,s}|\,\rho'_{2,k-s}.$$

Mais d'après le théorème classique sur les produits de séries à termes positifs, on a

$$\begin{aligned}\sum_{k=0}^{\infty}\sum_{s=0}^{k} \rho_{1,s}\,|\rho_{2,k-s} - \rho'_{2,k-s}| &= \underbrace{\left(\sum_{k=0}^{\infty} \rho_{1,k}\right)}_{=1}\underbrace{\left(\sum_{k=0}^{\infty} |\rho_{2,k} - \rho'_{2,k}|\right)}_{=\|\mathbb{P}_{X_2} - \mathbb{P}_{X'_2}\|}\\ &= \|\mathbb{P}_{X_2} - \mathbb{P}_{X'_2}\|;\end{aligned}$$

et de même

$$\sum_{k=0}^{\infty}\sum_{s=0}^{k} |\rho_{1,s} - \rho'_{1,s}|\,\rho'_{2,k-s} = \|\mathbb{P}_{X_1} - \mathbb{P}_{X'_1}\|.$$

Donc $\|\mathbb{P}_S - \mathbb{P}_{S'}\| \leqslant \|\mathbb{P}_{X_1} - \mathbb{P}_{X'_1}\| + \|\mathbb{P}_{X_2} - \mathbb{P}_{X'_2}\|$, comme attendu. □

Remarque. On peut écrire la même preuve d'une façon un peu plus "conceptuelle" (ou pédante...), comme suit. Si on identifie une mesure finie μ sur $\mathbb{N}$ avec la suite $\big(\mu(\{k\})\big)_{k\in\mathbb{N}}$, on voit que $M_+(\mathbb{N})$ s'identifie à une partie de $\ell^1(\mathbb{N})$, et que $\|\mu - \nu\| = \|\mu - \nu\|_1$ pour toutes $\mu, \nu \in M_+(\mathbb{N})$. De plus, $\ell^1(\mathbb{N})$ est muni d'un "produit" naturel : si $a, b \in \ell^1(\mathbb{N})$, alors $c = a * b$ est la suite définie par $c_k := \sum_{s=0}^{k} a_s b_{s-k}$. En utilisant le théorème classique sur les produits de séries absolument convergentes, on montre sans difficulté que pour toutes $a, b \in \ell^1(\mathbb{N})$, on a

$$\|a * b\|_1 \leqslant \|a\|_1\,\|b\|_1.$$

Par ailleurs, les formules pour $\mathbb{P}(X_1 + X_2 = k)$ et $\mathbb{P}(X'_1 + X'_2 = k)$ montrent que

$$\mathbb{P}_{X_1+X_2} = \mathbb{P}_{X_1} * \mathbb{P}_{X_2} \qquad \text{et} \qquad \mathbb{P}_{X'_1+X'_2} = \mathbb{P}_{X'_1} * \mathbb{P}_{X'_2}.$$

Donc on obtient

$$\begin{aligned}\|\mathbb{P}_S - \mathbb{P}_{S'}\| &= \|\mathbb{P}_{X_1} * \mathbb{P}_{X_2} - \mathbb{P}_{X'_1} * \mathbb{P}_{X'_2}\|_1\\ &\leqslant \|\mathbb{P}_{X_1} * (\mathbb{P}_{X_2} - \mathbb{P}_{X'_2})\|_1 + \|(\mathbb{P}_{X_1} - \mathbb{P}_{X'_1}) * \mathbb{P}_{X'_2}\|_1\\ &\leqslant \|\mathbb{P}_{X_1}\|_1\,\|\mathbb{P}_{X_2} - \mathbb{P}_{X'_2}\|_1 + \|\mathbb{P}_{X_1} - \mathbb{P}_{X'_1}\|_1\,\|\mathbb{P}_{X'_2}\|_1\\ &= \|\mathbb{P}_{X_2} - \mathbb{P}_{X'_2}\|_1 + \|\mathbb{P}_{X_1} - \mathbb{P}_{X'_1}\|_1.\end{aligned}$$

FAIT 2. Soit $p \in [0,1]$. Si X et X' sont deux va suivant respectivement la loi de Bernoulli $\mathcal{B}(p)$ et la loi de Poisson $\mathcal{P}(p)$, alors $\|\mathbb{P}_X - \mathbb{P}_{X'}\| \leqslant 2p^2$.

Preuve du Fait 2. Comme $\mathbb{P}(X = k) = 0$ pour $k \geqslant 2$, on a

$$\begin{aligned}\|\mathbb{P}_X - \mathbb{P}_{X'}\| &= |\mathbb{P}(X = 0) - \mathbb{P}(X' = 0)| + |\mathbb{P}(X = 1) - \mathbb{P}(X' = 1)| + \sum_{k=2}^{\infty} \mathbb{P}(X' = k)\\ &= |(1-p) - e^{-p}| + |p - pe^{-p}| + \sum_{k=2}^{\infty} \frac{p^k}{k!}\,e^{-p}\\ &= |e^{-p} - (1-p)| + p(1 - e^{-p}) + 1 - (e^{-p} + pe^{-p}).\end{aligned}$$

Comme $e^{-p} \geqslant 1-p$, on obtient donc

$$\begin{aligned}\|\mathbb{P}_X - \mathbb{P}_{X'}\| &= e^{-p} - (1-p) + p\,(1-e^{-p}) + 1 - (1+p)e^{-p} \\ &= 2p(1-e^{-p}) \\ &\leqslant 2p^2.\end{aligned}$$

□

Fait 3. Pour $n \in \mathbb{N}^*$, soient $X'_{n,1}, \dots, X'_{n,M_n}$ des va indépendantes, chaque $X'_{i,n}$ suivant la loi de Poisson $\mathcal{P}(p_{n,i})$, et soit $S'_n := X'_{n,1} + \cdots + X'_{n,M_n}$. Alors (S'_n) converge en loi vers une va $Z \sim \mathcal{P}(\lambda)$; autrement dit $\mathbb{P}(S'_n = k) \xrightarrow{n\to\infty} \frac{\lambda^k}{k!}e^{-\lambda}$ pour tout $k \in \mathbb{N}$.

Preuve du Fait 3. Comme les $X'_{n,i}$ sont indépendantes, on sait que S'_n suit la loi de Poisson $\mathcal{P}(\lambda_n)$, avec $\lambda_n := \sum_{i=1}^{M_n} p_{n,i}$. Autrement dit $\mathbb{P}(S'_n = k) = \frac{\lambda_n^k}{k!}e^{-\lambda_n}$ pour tout $k \in \mathbb{N}$; d'où le résultat puisque $\lambda_n \to \lambda$. □

On peut maintenant démontrer le théorème. Il s'agit de prouver que

$$\mathbb{P}(S_n = k) \xrightarrow{n\to\infty} \frac{\lambda^k}{k!} \qquad \text{pour tout } k \in \mathbb{N}.$$

Pour $n \in \mathbb{N}^*$, soient $X'_{n,1}, \dots, X'_{n,M_n}$ comme dans le Fait 3, et soit $S'_n := X'_{n,1} + \cdots + X'_{n,M_n}$. Par le Fait 3, il suffit de montrer que $\mathbb{P}(S_n = k) - \mathbb{P}(S'_n = k) \xrightarrow{n\to\infty} 0$ pour tout $k \in \mathbb{N}$; et pour cela, il suffit évidemment de montrer que $\|\mathbb{P}_{S_n} - \mathbb{P}_{S'_n}\| \to 0$.

Par les Faits 1 et 2, on a pour tout $n \in \mathbb{N}$:

$$\begin{aligned}\|\mathbb{P}_{S_n} - \mathbb{P}_{S'_n}\| &\leqslant \sum_{i=1}^{M_n} \|\mathbb{P}_{X_{n,i}} - \mathbb{P}_{X'_{n,i}}\| \\ &\leqslant 2\sum_{i=1}^{M_n} p_{n,i}^2.\end{aligned}$$

De plus,

$$\sum_{i=1}^{M_n} p_{n,i}^2 \leqslant \max(p_{n,1}, \dots, p_{n,M_n}) \sum_{i=1}^{n} p_{n,i} = \varepsilon_n \lambda_n \xrightarrow{n\to\infty} 0 \quad \text{par (i) et (ii).}$$

Donc en effet $\|\mathbb{P}_{S_n} - \mathbb{P}_{S'_n}\| \to 0$. □

4. Théorème limite central

Pour "motiver" le résultat principal de cette section, commençons par démontrer un fait très simple montrant que si on "centre" et si on "normalise dans L^2" une somme de va gaussiennes indépendantes, on obtient une va suivant la loi normale $\mathcal{N}(0,1)$.

Remarque 4.1. 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 $\mathcal{N}(m_0, \sigma_0^2)$, avec $m_0 := m_1 + \cdots + m_n$ et $\sigma_0^{\cdot} = \sigma_1^2 + \cdots + \sigma_n^2 = \sigma^2(S_n)$. 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 $Z_n = \frac{S_n - nm}{\sqrt{n}\sigma}$. □

Ce fait presque évident devrait on l'espère rendre "intuitif" le résultat très général suivant, qu'on appelle le **Théorème limite central**. Comme son nom l'indique, il s'agit d'un résultat "central" de la théorie des probabilités.

Théorème 4.2. *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 (*cf* le Corollaire 4.12 du Chapitre 7).

Remarque 2. Il faut voir Z_n comme 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. Par la loi des grands nombres, $\frac{S_n}{n} \xrightarrow{\text{ps}} m$. Le Théorème limite central donne une sorte de précision : 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 théorème, on voit que moralement, $\frac{S_n}{n} - m$ "tend vers 0 comme $\frac{1}{\sqrt{n}}$".

Preuve du Théorème 4.2. Quitte à remplacer X par $\frac{X-m}{\sigma}$ et X_i par $\frac{X_i - m}{\sigma}$, on se ramène au cas où

$$\mathbb{E}(X) = 0 \qquad \text{et} \qquad \sigma^2(X) = 1.$$

On doit alors montrer que

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} Z,$$

où Z suit la loi $\mathcal{N}(0,1)$. On va le faire en supposant de plus que

$$X \in L^3.$$

Une autre preuve, qui fonctionne dans le cas "général" où on suppose seulement que la va X appartient à L^2, sera donnée au Chapitre 9.

D'après la Proposition 4.17 du Chapitre 7, il suffit de montrer que

$$\mathbb{E}\left[\varphi\Big(\frac{X_1+\cdots+X_n}{\sqrt{n}}\Big)\right] \to \mathbb{E}[\varphi(Z)] \quad \text{pour toute } \varphi \in \mathcal{C}^\infty_{00}(\mathbb{R}).$$

Soit (Z_i) une suite de va indépendantes, indépendantes des X_i, et de loi $\mathcal{N}(0,1)$. Alors $\frac{Z_1+\cdots+Z_n}{\sqrt{n}}$ suit la loi $\mathcal{N}(0,1)$ par la Remarque 4.1. Il suffit donc de montrer que

$$\mathbb{E}\left[\varphi\Big(\frac{X_1+\cdots+X_n}{\sqrt{n}}\Big)\right] - \mathbb{E}\left[\varphi\Big(\frac{Z_1+\cdots+Z_n}{\sqrt{n}}\Big)\right] \xrightarrow{n\to\infty} 0;$$

ou encore, en posant

$$Y_i := \frac{X_i}{\sqrt{n}} \qquad \text{et} \qquad Y_i' := \frac{Z_i}{\sqrt{n}},$$

que

$$\mathbb{E}\Big(\varphi(Y_1+\cdots+Y_n)\Big) - \mathbb{E}\Big(\varphi(Y_1'+\cdots+Y_n')\Big) \to 0. \tag{4.1}$$

On aura pour cela besoin du Fait suivant.

Fait. Soient V, Y et Y' des va réelles indépendantes, appartenant à L^3, telles que $\mathbb{E}(Y)=\mathbb{E}(Y')$ et $\mathbb{E}(Y^2)=\mathbb{E}(Y'^2)$. Soit également $\varphi\in\mathcal{C}^3_{00}(\mathbb{R})$. Alors

$$\big|\mathbb{E}\big(\varphi(V+Y)\big) - \mathbb{E}\big(\varphi(V+Y')\big)\big| \leqslant \frac{\|\varphi^{(3)}\|_\infty}{6}\big(\mathbb{E}(|Y|^3)+\mathbb{E}(|Y'|^3)\big).$$

Preuve du Fait. D'après l'inégalité de Taylor-Lagrange, on peut écrire pour tout $\omega\in\Omega$:

$$\varphi\big(V(\omega)+Y(\omega)\big) = \varphi(V(\omega)) + \varphi'(V(\omega))Y(\omega) + \frac{\varphi''(V(\omega))}{2}Y(\omega)^2 + R(\omega)Y(\omega)^3,$$

où $|R(\omega)|\leqslant \frac{\|\varphi^{(3)}\|_\infty}{6}$. Comme V et Y sont indépendantes, on en déduit

$$\begin{aligned}\mathbb{E}\big(\varphi(V+Y)\big) &= \mathbb{E}(\varphi(V)) + \mathbb{E}(\varphi'(V)Y) + \frac{1}{2}\mathbb{E}(\varphi''(V)Y^2) + \mathbb{E}(RY^3)\\ &= \mathbb{E}(\varphi(V)) + \mathbb{E}(\varphi'(V))\mathbb{E}(Y) + \frac{1}{2}\mathbb{E}(\varphi''(V))\mathbb{E}(Y^2) + \mathbb{E}(RY^3).\end{aligned}$$

De même,

$$\mathbb{E}\big(\varphi(V+Y')\big) = \mathbb{E}(\varphi(V)) + \mathbb{E}(\varphi'(V))\mathbb{E}(Y') + \frac{1}{2}\mathbb{E}(\varphi''(V))\mathbb{E}(Y'^2) + \mathbb{E}(R'Y'^3),$$

où R' est une va vérifiant $|R'|\leqslant \frac{\|\varphi^{(3)}\|_\infty}{6}$.

Comme $\mathbb{E}(Y)=\mathbb{E}(Y')$ et $\mathbb{E}(Y^2)=\mathbb{E}(Y'^2)$, on obtient donc

$$\mathbb{E}\big(\varphi(V+Y)\big) - \mathbb{E}\big(\varphi(V+Y')\big) = \mathbb{E}(RY^3) - \mathbb{E}(R'Y'^3)\,,$$

et par conséquent

$$\begin{aligned}\big|\mathbb{E}\big(\varphi(V+Y)\big) - \mathbb{E}\big(\varphi(V+Y')\big)\big| &\leqslant \mathbb{E}(|RY^3|) + \mathbb{E}(|R'Y'^3|)\\ &\leqslant \frac{\|\varphi^{(3)}\|_\infty}{6}\big(\mathbb{E}(|Y|^3)+\mathbb{E}(|Y'|^3)\big).\end{aligned}$$

□

Revenons maintenant à la preuve du Théorème limite central. Pour démontrer (4.1), l'idée est d'écrire

$$\begin{aligned}\varphi(Y_1+\cdots+Y_n)-\varphi(Y_1'+\cdots+Y_n') = \\ &\varphi(Y_1+\cdots+Y_n)-\varphi(Y_1+\cdots+Y_{n-1}+Y_n') \\ &+\varphi(Y_1+\cdots+Y_{n-1}+Y_n')-\varphi(Y_1+\cdots+Y_{n-2}+Y_{n-1}'+Y_n') \\ &+\cdots \\ &+\varphi(Y_1+Y_2'+\cdots+Y_n')-\varphi(Y_1'+\cdots+Y_n');\end{aligned}$$

autrement dit

$$\varphi(Y_1+\cdots+Y_n)-\varphi(Y_1'+\cdots+Y_n')=\sum_{i=1}^n\Big(\varphi(U_i)-\varphi(U_{i-1})\Big),$$

où on a posé $U_0 := Y_1'+\cdots+Y_n'$ et

$$U_i := Y_1+\cdots+Y_i+Y_{i+1}'+\cdots+Y_n' \quad \text{pour } 1\leqslant i\leqslant n.$$

On a ainsi

$$\mathbb{E}\Big(\varphi(Y_1+\cdots+Y_n)\Big)-\mathbb{E}\Big(\varphi(Y_1'+\cdots+Y_n')\Big)=\sum_{i=1}^n\Big(\mathbb{E}\big(\varphi(U_i)\big)-\mathbb{E}\big(\varphi(U_{i-1})\big)\Big). \tag{4.2}$$

De plus,

$$U_i = V_i+Y_i \qquad \text{et} \qquad U_{i-1}=V_i+Y_i',$$

avec $V_i := Y_1+\cdots+Y_{i-1}+Y_{i+1}'+\cdots+Y_n'$.

Comme $\mathbb{E}(Y_i)=0=\mathbb{E}(Y_i')$ et $\mathbb{E}(Y_i^2)=\frac{1}{n}=\mathbb{E}((Y_i')^2)$ et comme les va V_i, Y_i et Y_i' sont indépendantes, on en déduit, d'après le Fait, qu'on a

$$\begin{aligned}\big|\mathbb{E}\big(\varphi(U_i)\big)-\mathbb{E}\big(\varphi(U_{i-1})\big)\big| &\leqslant \frac{\|\varphi^{(3)}\|_\infty}{6}\left(\mathbb{E}(|Y_i|^3)+\mathbb{E}(|Y_i'|^3)\right)\\ &= \frac{\|\varphi^{(3)}\|_\infty}{6n^{3/2}}\left(\mathbb{E}(|X|^3)+\mathbb{E}(|Z|^3)\right)\\ &= \frac{C}{n^{3/2}},\end{aligned}$$

où C est une constante indépendante de i et de n.

Par (4.2), on obtient donc

$$\Big|\mathbb{E}\Big(\varphi(Y_1+\cdots+Y_n)\Big)-\mathbb{E}\Big(\varphi(Y_1'+\cdots+Y_n')\Big)\Big|\leqslant n\times\frac{C}{n^{3/2}}=\frac{C}{\sqrt{n}},$$

ce qui termine la démonstration. □

COROLLAIRE 4.3. *Avec les notations du Théorème 4.2, 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 $Z\sim\mathcal{N}(0,1)$. □

ILLUSTRATION. On va utiliser le Théorème limite central pour démontrer la **Formule de Stirling** :

$$n!\sim n^n e^{-n}\sqrt{2\pi n} \qquad \text{quand } n\to\infty.$$

On commence par démontrer le fait suivant

Fait. Soit $(Z_n)_{n\geqslant 1}$ une suite de v.a. rélles convergeant en loi vers une va Z. On suppose que les Z_n sont dans L^2, et qu'il existe une constante C telle que $\forall n \geqslant 1$: $\mathbb{E}(Z_n^2) \leqslant C$. Alors $\mathbb{E}\big(Z_n^+\big) \to \mathbb{E}(Z^+)$ quand $n \to \infty$.

Preuve du Fait. Par la formule d'intégration par parties, on a

$$\mathbb{E}\big(Z_n^+\big) = \int_0^\infty \mathbb{P}\big(Z_N^+ > t\big)\, dt = \int_0^\infty \mathbb{P}(Z_n > t)\, dt,$$

et de même

$$\mathbb{E}\big(Z^+\big) = \int_0^\infty \mathbb{P}(Z > t)\, dt.$$

Cela étant, on a $\mathbb{P}(Z_n > t) = 1 - F_{Z_n}(t) \to 1 - F_Z(t) = \mathbb{P}(Z > t)$ en tout point de continuité de F_Z, donc *presque partout.* De plus par l'inégalité de Markov, on a

$$\forall n\ \forall t > 1\ :\ \mathbb{P}(Z_n > t) \leqslant \frac{\mathbb{E}(Z_n^2)}{t^2} \leqslant \frac{C}{t^2};$$

et aussi

$$\forall n\ \forall t \in [0,1]\ :\ \mathbb{P}(Z_n > t) \leqslant 1.$$

Par convergence dominée, on en déduit que $\int_0^\infty \mathbb{P}(Z_n > t)\, dt \to \int_0^\infty \mathbb{P}(Z > t)\, dt$; d'où le résultat. □

Maintenant, 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 Théorème limite central nous dit que

$$Z_n \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0,1).$$

De plus, on a $\mathbb{E}(Z_n^2) = 1$ pour tout $n \geqslant 1$. Par le Fait, on en déduit que

$$\mathbb{E}(Z_n^+) \to \mathbb{E}(Z^+) \qquad \text{quand } n \to \infty.$$

Calculons alors $\mathbb{E}(Z_n^+)$ et $\mathbb{E}(Z^+)$. On sait que S_n suit la loi de Poisson $\mathcal{P}(n)$; et on a $Z_n^+ = \varphi(S_n)$, où $\varphi : \mathbb{N} \to \mathbb{R}$ est la fonction définie par $\varphi(k) := \frac{(k-n)^+}{\sqrt{n}}\cdot$ Donc

$$\begin{aligned}
\mathbb{E}(Z_n^+) &= \sum_{k=0}^\infty \frac{(k-n)^+}{\sqrt{n}}\, \frac{n^k}{k!}\, e^{-n} \\
&= \frac{e^{-n}}{\sqrt{n}} \sum_{k=n}^\infty (k-n)\, \frac{n^k}{k!} \\
&= \frac{e^{-n}}{\sqrt{n}} \left(\sum_{k=n}^\infty \frac{n^k}{(k-1)!} - \sum_{k=n}^\infty \frac{n^{k+1}}{k!} \right) \\
&= \frac{e^{-n}}{\sqrt{n}}\, \frac{n^n}{(n-1)!}\cdot
\end{aligned}$$

Quant à $\mathbb{E}(Z^+)$, le calcul se fait tout seul :

$$\begin{aligned}\mathbb{E}(Z^+) &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} x^+\, e^{-\frac{x^2}{2}}\, dx \\ &= \frac{1}{\sqrt{2\pi}}\int_0^\infty x\, e^{-\frac{x^2}{2}}\, dx \\ &= \left[-\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\right] = \frac{1}{\sqrt{2\pi}}\cdot\end{aligned}$$

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.

5. Version précisée

La preuve du Théorème limite central donnée plus haut fournit une information supplémentaire : si on suppose que $X \in L^3$ et si on pose comme dans l'énoncé du théorème

$$Z_n := \frac{S_n - nm}{\sqrt{n}\,\sigma},$$

alors, pour toute fonction $\varphi \in \mathcal{C}^\infty_{00}(\mathbb{R})$, il existe une constante C dépendant uniquement de $\|\varphi^{(3)}\|_\infty$ et de $\mathbb{E}\big(|\frac{X-m}{\sigma}|^3\big)$ telle que

$$\forall n \geqslant 1\ :\ \big|\mathbb{E}\big(\varphi(Z_n)\big) - \mathbb{E}\big(\varphi(Z)\big)\big| \leqslant \frac{C}{\sqrt{n}}\cdot$$

La "conjecture" suivante paraît donc plausible : toujours en supposant $X \in L^3$, il existe une constante $C(X)$ telle que, pour tout $n \geqslant 1$ et pour tout intervalle $I \subseteq \mathbb{R}$, on ait

$$|\mathbb{P}(Z_n \in I) - \mathbb{P}(Z \in I)| \leqslant \frac{C(X)}{\sqrt{n}}\cdot$$

C'est effectivement le cas : ce résultat s'appelle le **Théorème de Berry-Esseen**.

Théorème 5.1. *Soient X et Z_n comme dans l'énoncé du Théorème limite central, et supposons de plus que $X \in L^3$. Soit également $Z \sim \mathcal{N}(0,1)$. Alors, pour tout intervalle $I \subseteq \mathbb{R}$ et pour tout $n \in \mathbb{N}^*$, on a*

$$|\mathbb{P}(Z_n \in I) - \mathbb{P}(Z \in I)| \leqslant \frac{C(X)}{\sqrt{n}},$$

avec

$$C(X) := \frac{\mathbb{E}\big(|X-m|^3\big)}{2\sigma^3}\cdot$$

Remarque 1. Ce résultat montre en particulier que $\mathbb{P}(Z_n \in I) - \mathbb{P}(Z \in I)$ tend vers 0 *uniformément* par rapport à l'intevalle I.

Remarque 2. Dans la version "originale" du théorème (qui date de 1942), la constante $C(X)$ vaut $c\,\mathbb{E}\big(|X-m|^3\big)/\sigma^3$, avec $c := 7,59$. Le fait qu'on puisse prendre $c \leqslant \frac{1}{2}$ est beaucoup plus récent (2010). On ne sait pas quel est le meilleur c possible.

La preuve du Théorème de Berry-Esseen est un peu trop délicate pour être donnée ici. En revanche, rien n'empêche d'utiliser le résultat !

Corollaire 5.2. *Soit $(X_i)_{i\geqslant 1}$ une suite de va indépendantes suivant une loi de Bernoulli de paramètre p. Pour $n\in\mathbb{N}^*$, posons $S_n := X_1+\cdots+X_n$. Soit également Z une va suivant la loi normale $\mathcal{N}(0,1)$. Pour tout $0<\varepsilon<1$ et pour tout $n\in\mathbb{N}^*$, on a*

$$\mathbb{P}\left(\left|\frac{S_n}{n}-p\right|\leqslant\varepsilon\right)\geqslant\mathbb{P}\big(Z\in[-2\varepsilon\sqrt{n}\,,\,2\varepsilon\sqrt{n}]\big)-\frac{C_p}{\sqrt{n}},\qquad \textit{avec}\quad C_p:=\tfrac{p^2+(1-p)^2}{2\sqrt{p(1-p)}}\cdot$$

Démonstration. Si X suit la loi $\mathcal{B}(p)$, alors $X\in L^3$ avec $m=p$, $\sigma^2=p(1-p)$ et $\mathbb{E}(|X-m|^3)=p(1-p)^3+(1-p)p^3=p(1-p)\big(p^2+(1-p)^2\big)=\sigma^2\big(p^2+(1-p)^2\big)$. D'après le Théorème de Berry-Esseen, on a donc pour tout $n\in\mathbb{N}^*$ et pour tout intervalle $I\subseteq\mathbb{R}$:

$$\left|\mathbb{P}\left(\frac{S_n-np}{\sqrt{n}\sigma}\in I\right)-\mathbb{P}(Z\in I)\right|\leqslant\frac{p^2+(1-p)^2}{2\sqrt{n}\sigma}=\frac{C_p}{\sqrt{n}};\tag{5.1}$$

et donc

$$\mathbb{P}\left(\frac{S_n-np}{\sqrt{n}\sigma}\in I\right)\geqslant\mathbb{P}(Z\in I)-\frac{C_p}{\sqrt{n}}\cdot$$

Comme

$$\mathbb{P}\left(\left|\frac{S_n}{n}-p\right|\leqslant\varepsilon\right)=\mathbb{P}\left(\frac{S_n-np}{\sqrt{n}\sigma}\in\underbrace{\left[-\frac{\varepsilon\sqrt{n}}{\sigma},\frac{\varepsilon\sqrt{n}}{\sigma}\right]}_{=:I_n}\right),$$

on en déduit

$$\mathbb{P}\left(\left|\frac{S_n}{n}-p\right|\leqslant\varepsilon\right)\geqslant\mathbb{P}\left(Z\in\left[-\frac{\varepsilon\sqrt{n}}{\sigma},\frac{\varepsilon\sqrt{n}}{\sigma}\right]\right)-\frac{C_p}{\sqrt{n}}\cdot$$

De plus, comme $0\leqslant p\leqslant 1$, on a $p(1-p)\leqslant\frac14$ (exo), et donc $\sigma=\sqrt{p(1-p)}\leqslant\frac12$. Par conséquent

$$\mathbb{P}\left(Z\in\left[-\frac{\varepsilon\sqrt{n}}{\sigma},\frac{\varepsilon\sqrt{n}}{\sigma}\right]\right)\geqslant\mathbb{P}\big(Z\in[-2\varepsilon\sqrt{n}\,,\,2\varepsilon\sqrt{n}]\big);$$

d'où le résultat. □

Corollaire 5.3. *Avec les notations précédentes, on a*

$$\mathbb{P}\left(\left|\frac{S_n}{n}-p\right|\leqslant\varepsilon\right)=1-O\Big(\frac{1}{\sqrt{n}}\Big)\qquad\textit{quand } n\to\infty.$$

Démonstration. Il suffit de montrer que $\mathbb{P}(Z\in[-2\varepsilon\sqrt{n},2\varepsilon\sqrt{n}])=1-O(\frac{1}{\sqrt{n}})$; ce qui n'est pas difficile : on a

$$\begin{aligned}\mathbb{P}(Z\in[-2\varepsilon\sqrt{n},2\varepsilon\sqrt{n}])&=1-2\mathbb{P}(Z>2\varepsilon\sqrt{n})\\&=1-\sqrt{\frac{2}{\pi}}\int_{2\varepsilon\sqrt{n}}^{\infty}e^{-\frac{x^2}{2}}\,dx,\end{aligned}$$

d'où le résultat car $\int_A^\infty e^{-\frac{x^2}{2}}\,dx=O(1/A)$ quand $A\to\infty$ (exo ; utiliser par exemple le fait que $e^{-x^2/2}\leqslant e^{-x}$ si x est assez grand). □

Pour illustrer le Corollaire 5.2, on va détailler une “application numérique”.

Exemple. Soit $(X_i)_{i\geqslant 1}$ une suite de va indépendantes suivant une loi de Bernoulli de paramètre p, et pour $n\in\mathbb{N}^*$, posons $S_n:=X_1+\cdots+X_n$. Soit également $0<\varepsilon<1$.

(i) Supposons qu'on ait $\frac{1}{4} \leqslant p \leqslant \frac{3}{4}$. Alors, si $n \geqslant \max\left(\left(\frac{1{,}055}{\varepsilon}\right)^2, 2315\right)$, on a $\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon$ avec une probabilité au moins égale à 95%.

(ii) Supposons qu'on ait $0,4 \leqslant p \leqslant 0,6$. Alors $\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon$ avec une probabilité au moins égale à 95% dès que $n \geqslant \max\left(\left(\frac{1{,}03}{\varepsilon}\right)^2, 2817\right)$

Démonstration. (i) Si $\frac{1}{4} \leqslant p \leqslant \frac{3}{4}$, alors $\sigma = \sqrt{p(1-p)} \geqslant \sqrt{\frac{1}{4}(1-\frac{1}{4})} = \frac{\sqrt{3}}{4}$ et $p^2 + (1-p)^2 \leqslant \frac{9}{16} + \frac{1}{16} = \frac{5}{8}$ (exo). Donc

$$\frac{p^2 + (1-p)^2}{2\sqrt{n}\sigma} \leqslant \frac{5}{4\sqrt{3n}},$$

et le Corollaire 5.2 donne

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right) \geqslant \mathbb{P}\big(Z \in [-2\varepsilon\sqrt{n}\,,\, 2\varepsilon\sqrt{n}]\big) - \frac{5}{4\sqrt{3n}}\cdot$$

Pour avoir $\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right) \geqslant 0,95$, il suffit donc que

$$\mathbb{P}\big(Z \in [-2\varepsilon\sqrt{n}, 2\varepsilon\sqrt{n}]\big) \geqslant 0,965 \qquad \text{et} \qquad \frac{5}{4\sqrt{3n}} \leqslant 0,015.$$

En consultant une *table de la loi normale* (trouvée par exemple sur internet, *cf* plus bas), on constate que la 1ère condition est réalisée si $2\varepsilon\sqrt{n} \geqslant 2,11$, autrement dit $\sqrt{n} \geqslant \frac{1{,}055}{\varepsilon}\cdot$ Comme la 2ème condition est équivalente à $\sqrt{3n} \geqslant \frac{500}{6}$ et comme $\frac{1}{3}\left(\frac{500}{6}\right)^2 \simeq 2314,8$, on en déduit le résultat annoncé : $\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right) \geqslant 0,95$ dès que $n \geqslant \max\left(\left(\frac{1{,}055}{\varepsilon}\right)^2, 2315\right)$.

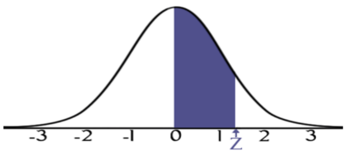
(ii) Si $0,4 \leqslant p \leqslant 0,6$, alors $\sigma = \sqrt{p(1-p)} \geqslant \sqrt{0,4(1-0,4)} = \frac{\sqrt{6}}{5}$ et $p^2+(1-p)^2 \leqslant (0,4)^2 + (0,6)^2 = 0,52$. Donc le Corollaire 5.2 donne cette fois

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right) \geqslant \mathbb{P}\big(Z \in [-2\varepsilon\sqrt{n}\,,\, 2\varepsilon\sqrt{n}]\big) - \frac{1,3}{\sqrt{6n}}\cdot$$

Pour avoir $\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leqslant \varepsilon\right) \geqslant 0,95$, il suffit donc que

$$\mathbb{P}\big(Z \in [-2\varepsilon\sqrt{n}, 2\varepsilon\sqrt{n}]\big) \geqslant 0,96 \qquad \text{et} \qquad \frac{1,3}{\sqrt{6n}} \leqslant 0,01.$$

D'après la table de la loi normale, la 1ère condition est réalisée si $2\varepsilon\sqrt{n} \geqslant 2,06$, *i.e.* $n \geqslant \left(\frac{1{,}03}{\varepsilon}\right)^2$; et comme $\frac{1}{6}\left(\frac{1{,}3}{0{,}01}\right)^2 \simeq 2816,7$ la 2ème l'est si $n \geqslant 2817$. □



STANDARD NORMAL TABLE (Z)

Entries in the table give the area under the curve between the mean and z standard deviations above the mean. For example, for $z = 1.25$ the area under the curve between the mean (0) and z is 0.3944.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998

SUITE DE L'EXEMPLE. Supposons que $0,4 \leqslant p \leqslant 0,6$.

Si on prend $\varepsilon := 0,01$, alors $\left(\frac{1,03}{\varepsilon}\right)^2 = 10609 > 2817$; et si on prend $\varepsilon := 0,015$, alors $\left(\frac{1,03}{\varepsilon}\right)^2 \simeq 4715,1 > 2817$. Donc

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leqslant 0,01\right) \geqslant 95\% \qquad \text{si } n \geqslant 10609,$$

et

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leqslant 0,015\right) \geqslant 95\% \qquad \text{si } n \geqslant 4716.$$

Cela peut s'interpréter comme suit : si on réalise un sondage auprès de n personnes pour estimer le pourcentage P de voix qu'obtiendra un candidat à une élection, et si on estime *a priori* que P est compris entre 40% et 60% (c'est la condition $0,4 \leqslant p \leqslant 0,6$), ce qui est raisonnable par exemple pour un 2ème tour d'élection présidentielle qu'on estime *a priori* assez serré, alors le sondage donnera un pourcentage "correct à 1% près avec un taux de fiabilité d'au moins 95%" si $n \geqslant 10609$, et un pourcentage "correct à $1,5$% près avec un taux de fiabilité d'au moins 95%" si $n \geqslant 4716$.

Prenons maintenant $n := 1600$ (ce qui est sans doute une taille d'échantillon plus "réaliste"), et $\varepsilon := 0,01$. Alors

$$\begin{aligned}\mathbb{P}\left(\left|\frac{S_n}{n}-p\right| \leqslant 0,01\right) &\geqslant \mathbb{P}\big(Z \in [-0,02\sqrt{n}\,,\,0,02\sqrt{n}]\big) - \frac{1,3}{\sqrt{6n}}\\ &= \mathbb{P}\big(Z \in [-0,8\,,\,0,8]\big) - \frac{0,0325}{\sqrt{6}}\\ &\geqslant \mathbb{P}\big(Z \in [-0,8\,,\,0,8]\big) - 0,0133\cdot\end{aligned}$$

D'après la table de la loi normale, $\mathbb{P}\big(Z \in [-1,2\,,\,1,2]\big) \geqslant 0,5762$; donc

$$\mathbb{P}\left(\left|\frac{S_n}{n}-p\right| \leqslant 0,01\right) \geqslant 0,5629.$$

Ainsi, l'affirmation "le sondage donne un pourcentage P correct à 1%" près est fiable à au moins $56,2$% ; ce qui n'est franchement pas satisfaisant.

Si on avait pris $n := 400$, on aurait obtenu

$$\mathbb{P}\left(\left|\frac{S_n}{n}-p\right| \leqslant 0,015\right) \geqslant \mathbb{P}\big(Z \in [-0,4\,,\,0,4]\big) - \frac{0,065}{\sqrt{6}} \geqslant 0,3108 - 0,0266 = 0,2842,$$

donc un taux de fiabilité au moins égal à 28,4%, ce qui est cette fois totalement ridicule.

Évidemment, on peut se dire qu'il doit y avoir moyen de minorer plus intelligemment $\mathbb{P}\left(\left|\frac{S_n}{n}-p\right| \leqslant 0,01\right)$. Mais on ne pourra quand même pas faire de miracles, car l'inégalité (5.1) donne quoi qu'il arrive

$$\mathbb{P}\left(\left|\frac{S_n}{n}-p\right| \leqslant \varepsilon\right) \leqslant \mathbb{P}\left(Z \in \left[-\frac{\varepsilon\sqrt{n}}{\sigma}, \frac{\varepsilon\sqrt{n}}{\sigma}\right]\right) + \frac{p^2+(1-p)^2}{2\sqrt{n}\sigma}\cdot$$

En prenant toujours $0,4 \leqslant p \leqslant 0,6$, on a $p^2+(1-p)^2 \leqslant 0,52$ et $\sigma = \sqrt{p(1-p)} \geqslant \frac{\sqrt{6}}{5}\cdot$ Donc, pour $\varepsilon := 0,01$,

$$\begin{aligned}\mathbb{P}\left(\left|\frac{S_n}{n}-p\right| \leqslant 0,01\right) &\leqslant \mathbb{P}\left(Z \in \left[-\frac{0,05}{\sqrt{6}}\sqrt{n}\,,\,\frac{0,05}{\sqrt{6}}\sqrt{n}\right]\right) + \frac{1,3}{\sqrt{6n}}\\ &\leqslant \mathbb{P}\left(Z \in [-0,0205\sqrt{n}\,,\,0,0205\sqrt{n}]\right) + \frac{1,3}{\sqrt{6n}}\cdot\end{aligned}$$

Pour $n := 400$, cela donne

$$\begin{aligned}\mathbb{P}\left(\left|\frac{S_n}{n}-p\right| \leqslant 0,01\right) &\leqslant \mathbb{P}\big(Z \in [-0,41\,,\,0,41]\big) + 0,027\\ &\leqslant 0,319 + 0,027 = 0,346 \qquad \text{d'après la table de la loi } \mathcal{N}(0,1).\end{aligned}$$

Donc, l'affirmation "le sondage donne un pourcentage P correct à 1% près" n'est pas fiable à plus de $34,6$%. Moralité : il ne faut absolument pas se fier à un sondage effectué auprès de 400 personnes, si on veut que la précision soit de 1%.

Pour $n := 1600$, on obtiendrait

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leqslant 0,01\right) \leqslant \mathbb{P}\big(Z \in [-0,82\,,\,0,82]\big) + 0,0135 \leqslant 0,5879 + 0,0135 = 0,6014,$$

donc une fiabilité inférieure à $60,2\%$; ce qui n'est toujours pas satisfaisant !

Chapitre 9

Fonctions caractéristiques

1. Transformée de Fourier d'une fonction intégrable

DÉFINITION 1.1. *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*

$$\widehat{f}(x) := \int_{\mathbb{R}} f(t)\, e^{-ixt}\, dt \qquad \textit{pour tout } x \in \mathbb{R}.$$

Remarque 1. Cette définition a bien un sens : comme $|f(t)\, e^{-ixt}| = |f(t)|$, la fonction $t \mapsto f(t)\, e^{-ixt}$ est intégrable sur $\mathbb{R}$, et donc $\widehat{f}(x)$ est bien défini pour tout $x \in \mathbb{R}$.

Remarque 2. On peut aussi écrire $\mathcal{F}f$ au lieu de $\widehat{f}$. En accord avec cette notation, on pose

$$\mathcal{F}L^1(\mathbb{R}) := \{\widehat{f};\ f \in L^1(\mathbb{R})\}.$$

Remarque 3. 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 $|F(t,x)| = |f(t)|$ est une 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. Soit $g : \mathbb{R} \to \mathbb{R}$ la fonction définie par

$$g(t) := e^{-\frac{t^2}{2}}.$$

Alors on a pour tout $x \in \mathbb{R}$:

$$\widehat{g}(x) = \sqrt{2\pi}\, e^{-\frac{x^2}{2}}.$$

Démonstration. Une manière de faire consiste à introduire la fonction $\phi : \mathbb{C} \to \mathbb{C}$ définie par

$$\phi(z) := \int_{\mathbb{R}} e^{-\frac{t^2}{2}} e^{-itz}\, dt.$$

La fonction ϕ est bien définie, car si $z \in \mathbb{C}$ est fixé, on a $|e^{-\frac{t^2}{2}} e^{-itz}| \leqslant e^{-\frac{t^2}{2}} e^{|z|\,|t|}$, fonction de t intégrable sur $\mathbb{R}$. De plus, si K est un compact de $\mathbb{C}$, on a une constante C_K telle que $|z| \leqslant C_K$ pour tout $z \in K$, et donc $|e^{-\frac{t^2}{2}} e^{-itz}| \leqslant e^{-\frac{t^2}{2}} e^{C_K\,|t|} =: g_K(t)$ pour $z \in K$ et $t \in \mathbb{R}$. La fonction g_K étant intégrable sur $\mathbb{R}$, on en déduit que ϕ est *holomorphe sur* $\mathbb{C}$, par le théorème d'holomorphie pour les intégrales à paramètres.

Si $y \in \mathbb{R}$, alors

$$\begin{aligned}\phi(iy) &= \int_{\mathbb{R}} e^{-\frac{t^2}{2}} e^{ty} dt \\ &= \int_{\mathbb{R}} e^{-\frac{1}{2}\left(t^2-2ty\right)} dt \\ &= \int_{\mathbb{R}} e^{-\frac{1}{2}\left((t-y)^2-y^2)\right)} dt \\ &= e^{\frac{y^2}{2}} \int_{\mathbb{R}} e^{-\frac{1}{2}(t-y)^2} dt \\ &= e^{\frac{y^2}{2}} \int_{\mathbb{R}} e^{-\frac{u^2}{2}} du = \sqrt{2\pi}\, e^{\frac{y^2}{2}}.\end{aligned}$$

Autrement dit :

$$\forall z = iy \in i\mathbb{R} \ : \ \phi(z) = \sqrt{2\pi}\, e^{-\frac{z^2}{2}}.$$

Comme $\phi(z)$ et $\psi(z) := \sqrt{2\pi}e^{-z^2/2}$ sont holomorphes sur $\mathbb{C}$, on en déduit qu'on a en fait $\phi(z) = \sqrt{2\pi}e^{-z^2/2}$ *pour tout* $z \in \mathbb{C}$, d'après le principe des zéros isolés. En particulier, $\widehat{g}(x) = \phi(x) = \sqrt{2\pi}e^{-x^2/2}$ pour tout $x \in \mathbb{R}$. □

On admet provisoirement le résultat fondamental suivant, qu'on appelle la **formule d'inversion de Fourier**. La preuve de ce résultat n'ayant absolument rien à voir avec les probabilités, on la fera seulement à la fin du chapitre (voir le Théorème 4.2).

THÉORÈME 1.2. *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 $t \in \mathbb{R}$:*

$$f(t) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(x) e^{itx} dx.$$

On va en déduire le résultat d'apparence anodine suivant, qui est en fait la seule chose dont on aura réellement besoin concernant la transformation de Fourier.

PROPOSITION 1.3. *Toute fonction $\varphi : \mathbb{R} \to \mathbb{C}$* de classe $\mathcal{C}^\infty$ à support compact *est une transformée de Fourier. Autrement dit :*

$$\mathcal{C}^\infty_{00}(\mathbb{R}) \subseteq \mathcal{F}L^1(\mathbb{R}).$$

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

$$\begin{aligned}\widehat{\varphi}(x) &= \int_{-A}^{A} \varphi(t) e^{-ixt} dt \\ &= \left[\frac{-1}{ix} e^{-ixt}\, \varphi(t)\right]_{-A}^{A} + \frac{1}{ix} \int_{-A}^{A} \varphi'(t) e^{-ixt} dt \\ &= \frac{1}{ix}\, \widehat{\varphi'}\,(x).\end{aligned}$$

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 :

$$\varphi = \widehat{f} \qquad \text{avec} \qquad f(x) := \frac{1}{2\pi}\,\widehat{\varphi}(-x).$$

□

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 de $\widehat{\mu}$ est laissée en exercice. 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 borélienne 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. □

Pour ce qui est des probabilités, l'intérêt principal de la transformation de Fourier des mesures tient au résultat suivant, qui est (un cas particulier d')un autre “théorème de Paul Lévy”.

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) entraîne (b).

Inversement, supposons (b) vérifiée. Alors $\mu_n(\mathbb{R}) = \widehat{\mu_n}(0) \to \widehat{\mu}(0) = \mu(\mathbb{R})$. De plus, par la formule d'échange, si $f \in L^1(\mathbb{R})$ alors

$$\int_{\mathbb{R}} \widehat{f}\,d\mu_n = \int_{\mathbb{R}} f(t)\,\widehat{\mu_n}(t)\,dt \xrightarrow{n\to\infty} \int_{\mathbb{R}} f(t)\,\widehat{\mu}(t)\,dt = \int_{\mathbb{R}} \widehat{f}\,d\mu.$$

Le passage à la limite est justifié par le théorème de convergence dominée, puisque $|f(t)\,\widehat{\mu_n}(t)| \leqslant \mu_n(\mathbb{R})\,|f(t)| \leqslant C\,|f(t)|$ pour une certaine constante C (la suite convergente $(\mu_n(\mathbb{R}))$ est bornée) et $f \in L^1(\mathbb{R})$.

Ainsi, on voit que $\int_{\mathbb{R}} \varphi\,d\mu_n \to \int_{\mathbb{R}} \varphi\,d\mu$ pour toute fonction $\varphi = \widehat{f} \in \mathcal{F}L^1(\mathbb{R})$, donc en particulier pour toute fonction $\varphi \in \mathcal{C}^\infty_{00}(\mathbb{R})$ d'après la Proposition 1.3. Donc $\mu_n \twoheadrightarrow \mu$ par la Proposition 4.8 du Chapitre 7, puisqu'on a déjà observé que $\mu_n(\mathbb{R}) \to \mu(\mathbb{R})$. □

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. □

REMARQUE. Comme indiqué plus haut, le Théorème 2.3 est un cas particulier du “Théorème de Paul Lévy”, lequel peut s'énoncer comme suit : *une suite $(\mu_n) \subseteq \mathrm{M}(\mathbb{R})$ converge étroitement si et seulement si la suite $(\widehat{\mu_n})$ converge simplement vers une fonction $\varphi : \mathbb{R} \to \mathbb{C}$ continue en 0, et dans ce cas $\varphi = \widehat{\mu}$ où μ est la mesure limite.* La preuve de ce résultat demande plus de travail que celle du Théorème 2.3, et nécessite en particulier de connaître un peu d'analyse fonctionnelle.

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”. □

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. Par définition, $\mathbb{P}_X = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$; donc

$$\begin{aligned}
\varphi_X(t) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} e^{itx} \, e^{-\frac{(x-m)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} e^{-\frac{u^2}{2\sigma^2}} e^{it(u+m)} du \\
&= \frac{1}{\sqrt{2\pi}} \times e^{imt} \int_{\mathbb{R}} e^{-\frac{u^2}{2\sigma^2}} \, e^{itu} \, \frac{du}{\sigma} \\
&= \frac{1}{\sqrt{2\pi}} \times e^{imt} \int_{\mathbb{R}} e^{-\frac{v^2}{2}} \, e^{i\sigma t v} dv = \frac{1}{\sqrt{2\pi}} \times e^{imt} \times \sqrt{2\pi} \, e^{-\frac{(\sigma t)^2}{2}}.
\end{aligned}$$

□

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

$$\varphi_X(t) = e^{-a\,|t|}.$$

Démonstration. Par définition,

$$\varphi_X(t) = \frac{a}{\pi}\int_{\mathbb{R}} \frac{e^{itx}}{a^2+x^2}\,dx.$$

Une façon de faire est d'appliquer le Théorème des résidus à la fonction f définie par $f(z) := \frac{e^{itz}}{a^2+z^2}$. Visiblement, f est holomorphe sur $\mathbb{C}\backslash\{\pm ia\}$ et possède un pôle simple en $\pm ia$, avec

$$\mathrm{Res}(f,\alpha) = \frac{e^{it\alpha}}{2\alpha} \qquad \text{pour } \alpha = \pm ia.$$

Si $t \geqslant 0$, alors e^{itz} est borné dans le demi-plan $\{\mathrm{Im}(z) > 0\}$, et une application "facile" du Théorème des résidus donne

$$\int_{\mathbb{R}} \frac{e^{itx}}{a^2+x^2}\,dx = 2i\pi \times \mathrm{Res}(f, ia) = 2i\pi \times \frac{e^{-ta}}{2ia} = \frac{\pi}{a}\,e^{-at}.$$

On obtient donc

$$\varphi_X(t) = e^{-at} \qquad \text{pour } t \geqslant 0.$$

Si $t < 0$ on trouve $\varphi_X(t) = e^{ta} = e^{-a|t|}$ par le même argument ; ou bien on peut observer que φ_X est une fonction paire (micro-exo), et donc $\varphi_X(t) = \varphi_X(-t) = e^{(-a)\times(-t)} = e^{-a|t|}$. □

Autre démonstration. On va utiliser la formule d'inversion de Fourier plutôt que le Théorème des résidus. 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 $f(t) = \frac{a}{\pi}\int_{\mathbb{R}} \frac{e^{itx}}{a^2+x^2}\, dx = \varphi_X(t)$. □

EXERCICE. Soit X une va réelle, et soient $a, b \in \mathbb{R}$. Montrer que pour tout $t \in \mathbb{R}$, on a

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

3.2. Fonctions caractéristiques et sommes.

PROPOSITION 3.2. *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 $\mathcal{N}(m, \sigma^2)$, avec $m := m_X + m_Y$ et $\sigma^2 := \sigma_X^2 + \sigma_Y^2$.

Démonstration. On a

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) = e^{im_Xt}e^{-\sigma_X^2\frac{t^2}{2}}\, e^{im_Yt}e^{-\sigma_Y^2\frac{t^2}{2}} = e^{imt}e^{-\sigma^2\frac{t^2}{2}},$$

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

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 3.3. *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*

$$\varphi_X^{(k)}(t) = i^k\,\mathbb{E}\big(X^k e^{itX}\big) \qquad \text{pour } k = 1, \dots, N.$$

En particulier,

$$\varphi^{(k)}(0) = i^k\,\mathbb{E}(X^k).$$

Démonstration. La partie (1) a déjà été vue (on a montré que pour toute mesure $\mu \in M_+(\mathbb{R})$, la fonction $\widehat{\mu}$ est continue bornée avec $\|\widehat{\mu}\|_\infty = \mu(\mathbb{R}) = \widehat{\mu}(0)$) ; et pour (2), on utilise le théorème de dérivation des intégrales à paramètres (exo). □

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 3.4. *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. □

Autre preuve. On utilise directement le développement en série entière de l'exponentielle :

$$\begin{aligned}\varphi_X(t) &= \mathbb{E}\left(e^{itX}\right)\\ &= \mathbb{E}\left(\sum_{k=0}^{\infty}\frac{(itX)^k}{k!}\right)\\ &= \sum_{k=0}^{\infty}\frac{i^k\mathbb{E}(X^k)}{k!}\,t^k.\end{aligned}$$

Pour justifier ce calcul formel, il suffit de vérifier qu'on a

$$\mathbb{E}\left(\sum_{k=0}^{\infty}\frac{|itX|^k}{k!}\right) < \infty;$$

ce qui n'est pas difficile en utilisant le fait que $X \in L^\infty$:

$$\mathbb{E}\left(\sum_{k=0}^{\infty}\frac{|itX|^k}{k!}\right) \leqslant \mathbb{E}\left(\sum_{k=0}^{\infty}\frac{(|t|\,\|X\|_\infty)^k}{k!}\right) = \mathbb{E}\left(e^{|t|\,\|X\|_\infty}\right) = e^{|t|\,\|X\|_\infty} < \infty.$$

□

Illustration. Théorème des moments (*cf* Chapitre 6).

Soient X et Y deux va réelles bornées telles que $\mathbb{E}(X^k) = \mathbb{E}(Y^k)$ pour tout $k \in \mathbb{N}$. Alors 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 $k \in \mathbb{N}$. Ainsi $\varphi_X = \varphi_Y$, et donc X et Y ont la même loi !

Proposition 3.5. *Soit X une va réelle, et soit $k \in \mathbb{N}^*$. Si φ_X est $2k$ fois dérivable en 0, alors X admet un moment d'ordre $2k$.*

Démonstration. On va la faire seulement pour $k = 1$. On suppose donc que φ_X est deux fois dérivable en 0, et il s'agit de montrer que $X \in L^2$.

Par hypothèse, φ_X admet un développement limité d'ordre 2 en 0 :

$$\varphi_X(t) = a + bt + ct^2 + o(t^2), \qquad \text{avec } a = \varphi_X(0) = 1.$$

On en déduit que

$$\frac{\varphi_X(t) + \varphi_X(-t) - 2}{2t^2} \to c \qquad \text{quand } t \to 0.$$

Par ailleurs, on a

$$\frac{\varphi_X(t) + \varphi_X(-t)}{2} = \mathbb{E}\Big(\frac{e^{itX} + e^{-itX}}{2}\Big) = \mathbb{E}\big(\cos(tX)\big).$$

Donc

$$\frac{\varphi_X(t) + \varphi_X(-t) - 2}{2t^2} = \mathbb{E}\Big(\frac{\cos(tX) - 1}{t^2}\Big),$$

et par conséquent

$$\mathbb{E}\Big(\frac{1 - \cos(tX)}{t^2}\Big) \xrightarrow{t\to 0} -c.$$

Soit alors (t_n) une suite de réels strictement positifs tendant vers 0, et posons

$$Y_n := \frac{1-\cos(t_n X)}{t_n^2}\cdot$$

Les (Y_n) sont des va *positives*, et en utilisant le développement limité de $\cos(u)$ en 0, on voit que

$$Y_n \xrightarrow{\text{ps}} \frac{1}{2}X^2 \qquad \text{quand } n\to\infty.$$

D'après le *lemme de Fatou*, on en déduit qu'on a

$$\mathbb{E}\Big(\frac{1}{2}X^2\Big) \leqslant \underline{\lim}\,\mathbb{E}(Y_n) = -c < \infty;$$

et donc $X \in L^2$. □

COROLLAIRE 3.6. *Si φ_X est infiniment dérivable en 0, alors X admet des moments de tous ordres.*

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 vu au Chapitre 7. □

ILLUSTRATION 1. Preuve du Théorème limite central avec “hypothèse minimale”.

Soit X une va réelles *appartenant à L^2*, de moyenne m et de variance σ^2. Soit (X_k) une suite de va indépendantes et de même loi que X, et soit

$$Z_n := \frac{S_n - nm}{\sqrt{n}\sigma} = \frac{X_1+\cdots+X_n - nm}{\sqrt{n}\sigma}\cdot$$

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

$$\varphi_{Z_n}(t) = \mathbb{E}\left(\prod_{k=1}^n e^{it\frac{X_k}{\sqrt{n}}}\right) = \prod_{k=1}^n \mathbb{E}\big(e^{i\frac{t}{\sqrt{n}}X_k}\big);$$

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 $\varphi_X''(0) = \mathbb{E}(i^2X^2) = -1$. Donc φ_X admet le développement limité suivant au voisinage de 0 :

$$\varphi_X(u) = 1 - \frac{u^2}{2} + o(u^2).$$

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 $\left|\varphi_X\big(\frac{t}{\sqrt{n}}\big) - 1\right| < 1$ 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.

Illustration 2. Loi “faible” des grands nombres dans L^1.

Soit X une va réelle appartenant à L^1, de moyenne m, et soit $(X_k)_{k\geqslant 1}$ une suite de va indépendantes et de même loi que X. Pour $n \in \mathbb{N}^*$, posons comme d'habitude

$$S_n := X_1 + \cdots + X_n.$$

On va utiliser les fonctions caractéristiques pour montrer que

$$\frac{S_n}{n} \longrightarrow m \qquad \textit{en probabilité}.$$

Comme la convergence en probabilité vers une *constante* est équivalente à la convergence en loi vers cette même constante (Corollaire 4.13 du Chapitre 7), il suffit de montrer que $\frac{S_n}{n}$ converge en loi vers m, autrement dit que

$$\varphi_{\frac{S_n}{n}}(t) \xrightarrow{n\to\infty} \varphi_{m\mathbf{1}}(t) = e^{imt} \qquad \text{pour tout } t \in \mathbb{R}.$$

On procède comme pour la preuve du Théorème limite central. Comme les X_k sont indépendantes et de même loi que X, on trouve

$$\varphi_{\frac{S_n}{n}}(t) = \varphi_X\Big(\frac{t}{n}\Big)^n.$$

De plus, comme $X \in L^1$, on sait que φ_X est dérivable en 0 avec $\varphi_X'(0) = \mathbb{E}(X) = m$; donc φ_X admet un développement limité à l'ordre 1 en 0 :

$$\varphi_X(u) = 1 + mu + o(u),$$

On en déduit alors facilement le résultat, en utilisant le logarithme complexe ou en s'en passant (*cf* la remarque suivant la preuve du Théorème limite central). Exo : écrire les détails.

ILLUSTRATION 3. Une autre preuve du Théorème des évènements rares.

Pour $n \in \mathbb{N}^*$, soient $X_{n,1}, \dots, X_{n,M_n}$ des va indépendantes à valeurs dans $\mathbb{N}$, chaque $X_{n,j}$ suivant une loi de Bernoulli $\mathcal{B}(p_{n,j})$. On suppose que

(i) $\lambda_n := \sum_{j=1}^{M_n}$ admet une limite $\lambda > 0$ quand $n \to \infty$;

(ii) $\varepsilon_n := \max(p_{n,1}, \dots, p_{n,M_n})$ tend vers 0 quand $n \to \infty$;

et on veut montrer que $S_n := X_{n,1} + \cdots + X_{n,M_n}$ converge en loi vers une va Z suivant la loi de Poisson $\mathcal{P}(\lambda)$.

Fixons $t \in \mathbb{R}$. Pour tout $n \in \mathbb{N}^*$, on a par indépendance des $X_{n,j}$:

$$\varphi_{S_n}(t) = \prod_{j=1}^{M_n} \varphi_{X_{n,j}}(t) = \prod_{j=1}^{M_n} \big(1 + p_{n,j}(e^{it} - 1)\big).$$

Par (ii), on peut trouver un entier n_0 tel que $p_{n,j} \leqslant 1/4$ pour tout $n \geqslant n_0$ et pour $j = 1, \dots, M_n$. Si $n \geqslant n_0$, on a alors $|p_{n,j}(e^{it} - 1)| \leqslant 1/2 < 1$ pour $j = 1, \dots, M_n$ car $|e^{it} - 1| \leqslant 2$, donc on peut écrire

$$1 + p_{n,j}(e^{it} - 1) = \exp\Big(\log\big(1 + p_{n,j}(e^{it} - 1)\big)\Big),$$

où log est la détermination principale du logarithme. Ainsi

$$\varphi_{S_n}(t) = \exp\left(\sum_{j=1}^{M_n} \log\big(1 + p_{n,j}(e^{it} - 1)\big)\right) \qquad \text{pour } n \geqslant n_0.$$

De plus, comme $f(u) := \log(1+u)$ est holomorphe dans le disque $D(0,1)$ avec $f(0) = 0$ et $f'(0) = 1$, on peut écrire $\log(1+u) = u + u^2 g(u)$ où g est une fonction continue sur le disque $D(0,1)$; et comme $p_{n,j}(e^{it} - 1) \in \overline{D}(0, 1/2)$ si $n \geqslant n_0$ et $j = 1, \dots, M_n$, on en déduit que si $n \geqslant n_0$, alors

$$\log\big(1 + p_{n,j}(e^{it} - 1)\big) =: p_{n,j}(e^{it} - 1) + \xi_{n,j},$$

avec

$$|\xi_{n,j}| \leqslant |p_{n,j}(e^{it} - 1)|^2 \times \sup\big\{|g(u)|;\ |u| \leqslant 1/2\big\} = C\, p_{n,j}^2,$$

la constante C étant indépendante de n et j. Donc, pour $n \geqslant n_0$,

$$\begin{aligned}\varphi_{S_n}(t) &= \exp\left(\sum_{j=1}^{M_n} \big(p_{n,j}(e^{it} - 1) + \xi_{n,j}\big)\right)\\ &= e^{\lambda_n(e^{it}-1)} \exp\Big(\sum_{j=1}^{M_n} \xi_{n,j}\Big).\end{aligned}$$

Comme $\lambda_n \to \lambda$ quand $n \to \infty$ et que

$$\Big|\sum_{j=1}^{M_n} \xi_{n,j}\Big| \leqslant C \sum_{j=1}^{M_n} p_{n,j}^2 \leqslant C\,\varepsilon_n \lambda_n \longrightarrow 0,$$

on peut donc conclure que

$$\varphi_{S_n}(t) \xrightarrow{n\to\infty} e^{\lambda(e^{it}-1)}.$$

On reconnaît dans le membre de droite la fonction caractéristique de la loi de Poisson de paramètre λ ; donc (S_n) converge en loi vers une va $Z \sim \mathcal{P}(\lambda)$.

Illustration 4. Séries de Rademacher.

Comme dernière illustration de l'utilité des fonctions caractéristiques, on va démontrer le résultat suivant.

Proposition 3.8. *Soit $(\varepsilon_k)_{k\geqslant 0}$ une suite de va indépendantes à valeurs dans $\{-1,1\}$ et suivant une loi de Rademacher, i.e.*

$$\mathbb{P}(\varepsilon_k = 1) = \frac{1}{2} = \mathbb{P}(\varepsilon_k = -1).$$

Soit également $(a_k)_{k\in\mathbb{N}}$ une suite de nombres réels. Alors la série $\sum a_k\varepsilon_k$ converge presque sûrement si et seulement si $\sum_0^\infty a_k^2 < \infty$.

Démonstration. On a vu au Chapitre 7 que si $\sum_0^\infty a_k^2 < \infty$, alors la série $\sum a_k\varepsilon_k$ converge presque sûrement : *cf* le Corollaire 3.2 (prendre $X_k := a_k\varepsilon_k$).

La preuve de la réciproque est donnée sous forme d'exercice. Dans ce qui suit, on suppose que la série $\sum a_k\varepsilon_k$ converge presque sûrement. On pose $S := \sum_{k=0}^\infty a_k\varepsilon_k$ (qui est une va bien définie presque sûrement), et $S_n := \sum_{k=0}^n a_k\varepsilon_k$ pour tout $n \in \mathbb{N}$.

(1) Montrer que a_k tend vers 0 quand $k \to \infty$.

(2) Justifier qu'il existe $t > 0$ tel que $\varphi_S(t) \neq 0$.

(3) Exprimer $\varphi_{S_n}(t)$ à l'aide de φ_X, puis montrer que la série $\sum \log|\varphi_X(a_k t)|$ est convergente.

(4) Justifier que φ_X admet un développement limité d'ordre 2 au voisinage de 0, et que ce développement est de la forme

$$\varphi_X(u) = 1 - c\,u^2 + o(u^2),$$

où $c > 0$ est à déterminer explicitement.

(5) En utilisant (1), (3) et (4), montrer que la série $\sum a_k^2$ est convergente.

□

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 particulier :

- 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 3.9. *Soient $X_1, \dots, X_d$ des va réelles définies sur le même espace de probabilité $(\Omega, \mathfrak{A}, \mathbb{P})$. Alors $X_1, \dots, X_d$ sont indépendantes si et seulement si*

$$(*) \qquad \varphi_{a_1X_1+\cdots+a_dX_d}(s) = \varphi_{X_1}(a_1s)\cdots\varphi_{X_d}(a_ds)$$

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. □

ILLUSTRATION. Si X et Y sont des va indépendantes suivant la loi $\mathcal{N}(0,1)$ et si $a, b \in \mathbb{R}$, alors les va $aX + bY$ et $bX - aY$ sont indépendantes.

Démonstration. Posons $X_1 := aX + bY$ et $X_2 := bX - aY$. Si $a_1, a_2 \in \mathbb{R}$, alors

$$a_1X_1 + a_2X_2 = (a_1a + a_2b)X + (a_1b - a_2a)Y.$$

Par indépendance de X et Y, on a donc (pour tout $s \in \mathbb{R}$)

$$\begin{aligned}\varphi_{a_1X_1+a_2X_2}(s) &= \varphi_{(a_1a+a_2b)X}(s)\,\varphi_{(a_1b-a_2a)Y}(s)\\ &= e^{-(a_1a+a_2b)^2\frac{s^2}{2}}e^{-(a_1b-a_2a)^2\frac{s^2}{2}}\\ &= e^{-(a_1^2+a_2^2)(a^2+b^2)\frac{s^2}{2}}.\end{aligned}$$

Mais toujours par indépendance de X et Y, on a aussi

$$\begin{aligned}\varphi_{X_1}(a_1s)\varphi_{X_2}(a_2s) &= \varphi_X(a_1as)\varphi_Y(a_1bs)\varphi_X(a_2bs)\varphi_Y(-a_2as)\\ &= e^{-(a_1^2+a_2^2)(a^2+b^2)\frac{s^2}{2}}.\end{aligned}$$

Donc X_1 et X_2 sont indépendantes par la proposition. □

Exercice. Re-démontrer ce résultat sans utiliser les fonctions caractéristiques.

4. Appendice : formule d'inversion de Fourier

Dans cette section, on démontre la formule d'inversion de Fourier, ce qui achèvera la preuve de la Proposition 1.3. La présentation est volontairement minimaliste : on pourrait (devrait) en dire beaucoup plus sur la transformation de Fourier !

Démontrons d'abord un lemme d'"approximation".

LEMME 4.1. *Soit $k \in L^1(\mathbb{R})$ vérifiant $\int_{\mathbb{R}} k(t)\,dt = 1$, et soit (λ_n) une suite de réels strictement positifs tendant vers l'infini. Pour $n \in \mathbb{N}$, on pose $k_n(t) := \lambda_n k(\lambda_n t)$. Alors, pour toute fonction $f \in L^1(\mathbb{R})$, la suite $(k_n * f)$ tend vers f en norme L^1.*

Démonstration. Pour tout $n \in \mathbb{N}$ et presque tout $x \in \mathbb{R}$, on a

$$\begin{aligned}k_n * f(x) &= \int_{\mathbb{R}} f(x-t)k(\lambda_n t)\,\lambda_n dt\\ &= \int_{\mathbb{R}} f\Big(x-\frac{u}{\lambda_n}\Big)k(u)\,du.\end{aligned}$$

De plus, comme $\int_{\mathbb{R}} k(u)\,du = 1$, on peut écrire

$$f(x) = \int_{\mathbb{R}} f(x)\,k(u)\,du,$$

et donc

$$k_n * f(x) - f(x) = \int_{\mathbb{R}} \Big[f\Big(x-\frac{u}{\lambda_n}\Big) - f(x)\Big]\,k(u)\,du.$$

On en déduit (en utilisant Fubini) :

$$\begin{aligned}\|k_n * f - f\|_1 &\leqslant \int_{\mathbb{R}}\left(\int_{\mathbb{R}}\Big|f\Big(x-\frac{u}{\lambda_n}\Big) - f(x)\Big|\,|k(u)|\,du\right)dx\\ &= \int_{\mathbb{R}}\left(\Big|f\Big(x-\frac{u}{\lambda_n}\Big) - f(x)\Big|\,dx\right)|k(u)|\,du\\ &= \int_{\mathbb{R}} \|\tau_{u/\lambda_n} f - f\|_1 |k(u)|\,du,\end{aligned}$$

où $\tau_a f$ est la "translatée de f par a".

Comme $u/\lambda_n \to 0$, on sait (*continuité des translations*) que $\|\tau_{u/\lambda_n} f - f\|_1 \to 0$ quand $n \to \infty$, pour tout $u \in \mathbb{R}$. De plus, on a $\|\tau_{u/\lambda_n} f - f\|_1 \leqslant C := 2\|f\|_1$, et donc

$\|\tau_{u/\lambda_n} f - f\|_1 \leqslant C\,|k(u)|$ pour tout n et pour tout $u \in \mathbb{R}$. Comme $k \in L^1(\mathbb{R})$, on en déduit que $\|k_n * f - f\|_1 \to 0$ par convergence dominée. $\square$

Remarque. On dira qu'une suite $(k_n) \subseteq L^1(\mathbb{R})$ vérifiant la conclusion du Lemme 4.1 est une **unité approchée pour** $L^1(\mathbb{R})$.

Voici maintenant la formule d'inversion de Fourier.

Théorème 4.2. *Si $f \in L^1(\mathbb{R})$ est telle que $\widehat{f}$ appartient également à $L^1(\mathbb{R})$, alors*

$$f(t) = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(x)\, e^{itx}\, dx \quad \textit{presque partout}.$$

Démonstration. Pour toute fonction $g \in L^1(\mathbb{R})$, on notera $\breve{g} : \mathbb{R} \to \mathbb{C}$ la fonction définie par

$$\breve{g}\,(t) := \int_{\mathbb{R}} g(x)\, e^{itx}\, dx\,.$$

(On dit parfois que $\breve{g}$ est la **transformée de Fourier inverse** de la fonction g.)

On a besoin de deux faits préliminaires.

Fait 1. Si $f, g \in L^1(\mathbb{R})$, alors $g\widehat{f} \in L^1(\mathbb{R})$ et $\breve{\widehat{g\widehat{f}}} = \breve{g} * f$.

Preuve du Fait 1. On a $g\widehat{f} \in L^1$ car $g \in L^1$ et $\widehat{f}$ est bornée. Ensuite, le calcul se fait tout seul :

$$\begin{aligned}
\breve{(g\widehat{f})}\,(t) &= \int_{\mathbb{R}} g(x)\widehat{f}\,(x)\, e^{itx}\, dt \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(s) e^{-ixs} ds\right) g(x) e^{itx} dx \\
&= \int_{\mathbb{R}} f(s) \left(\int_{\mathbb{R}} g(x) e^{ix(t-s)}\right) ds \qquad \text{par Fubini} \\
&= \int_{\mathbb{R}} f(s)\, \breve{g}\,(t-s)\, ds \\
&= f * \breve{g}\,(t) = \breve{g} * f\,(t)\,.
\end{aligned}$$

(La justification de l'emploi du Théorème de Fubini est laissée en exercice.) $\square$

Fait 2. Il existe une suite $(g_n) \subseteq L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ vérifiant les propriétés suivantes :

(a) (g_n) est uniformément bornée et $g_n(x) \to 1$ pour tout $x \in \mathbb{R}$;

(b) $\breve{g_n} \in L^1(\mathbb{R})$ pour tout n, et la suite $\left(\frac{1}{2\pi}\breve{g_n}\right)$ est une unité approchée pour $L^1(\mathbb{R})$.

Preuve du Fait 2. Soit $g : \mathbb{R} \to \mathbb{R}$ la fonction gaussienne définie par

$$g(x) := e^{-x^2/2}\,.$$

La fonction g est intégrable sur $\mathbb{R}$, et on a vu qu'on a

$$\breve{g}(t) = \int_{\mathbb{R}} e^{-x^2/2} e^{itx} dt = \sqrt{2\pi}\, e^{-t^2/2} \quad \text{pour tout } t \in \mathbb{R}\,.$$

En particulier, $\breve{g} \in L^1(\mathbb{R})$ et

$$\int_{\mathbb{R}} \breve{g}(t)\, dt = \sqrt{2\pi} \int_{\mathbb{R}} e^{-t^2/2}\, dt = 2\pi\,.$$

Posons alors

$$g_n(x) := g(\varepsilon_n x)\,,$$

où (ε_n) est n'importe quelle suite de réels strictement positifs tendant vers 0.

Pour tout $t \in \mathbb{R}$, on a

$$\begin{aligned}\breve{g}_n(t) &= \int_{\mathbb{R}} g(\varepsilon_n x) e^{itx} dx \\ &= \int_{\mathbb{R}} g(u) e^{it\frac{u}{\varepsilon_n}} \frac{du}{\varepsilon_n} \\ &= \frac{1}{\varepsilon_n}\, \breve{g}\left(\frac{t}{\varepsilon_n}\right).\end{aligned}$$

Donc la suite $\left(\frac{1}{2\pi}\breve{g}_n\right)$ est une unité approchée pour $L^1(\mathbb{R})$ puisque $\int_{\mathbb{R}} \frac{1}{2\pi}\,\breve{g} = 1$ et $1/\varepsilon_n \to \infty$.

De plus, la suite (g_n) est uniformément bornée car g est bornée, et $g_n(x) \to g(0) = 1$ pour tout $x \in \mathbb{R}$ car g est continue en 0 et $\varepsilon_n \to 0$. $\square$

On peut maintenant démontrer la formule d'inversion de Fourier. Soit $f \in L^1(\mathbb{R})$ telle que $\widehat{f} \in L^1$. Si la suite (g_n) est comme dans le Fait 2, on a d'après le Fait 1 :

$$\widetilde{g_n \widehat{f}} = \widetilde{g_n} * f \qquad \text{pour tout } n \in \mathbb{N}\,.$$

Comme (g_n) est uniformément bornée et tend simplement vers 1, et comme $\widehat{f} \in L^1$, on voit que

$$\widetilde{g_n \widehat{f}}\,(t) = \int_{\mathbb{R}} g_n(x) \widehat{f}(x)\, e^{itx}\, dt \to \int_{\mathbb{R}} \widehat{f}(x) e^{itx}\, dx \quad \text{pour tout } t \in \mathbb{R}\,, \tag{4.1}$$

par le théorème de convergence dominée.

De plus, d'après le Fait 2, $\widetilde{g_n} * f$ tend vers $2\pi f$ *en norme* L^1 ; donc on peut trouver une suite strictement croissante d'entiers (n_k) telle que

$$\widetilde{g_{n_k}} * f\,(t) \to 2\pi f(t) \qquad \text{presque partout}\,. \tag{4.2}$$

De (4.1) et (4.2), on déduit immédiatement le résultat souhaité :

$$2\pi\, f(t) = \int_{\mathbb{R}} \widehat{f}(x)\, e^{itx}\, dx \qquad \text{presque partout}\,.$$

$\square$

Corollaire 4.3. *La transformation de Fourier est* injective *sur* $L^1(\mathbb{R})$ *: si* $f_1, f_2 \in L^1(\mathbb{R})$ *vérifient* $\widehat{f_1} = \widehat{f_2}$*, alors* $f_1 = f_2$ *presque partout.*

Démonstration. Si $f \in L^1$ vérifie $\widehat{f} = 0$, alors $f = 0$ presque partout d'après la formule d'inversion ; d'où le résultat par linéarité. $\square$

Corollaire 4.4. *Si* $f : \mathbb{R} \to \mathbb{C}$ *est une fonction* continue *intégrable sur* $\mathbb{R}$ *telle que* $\widehat{f} \in L^1(\mathbb{R})$*, alors*

$$f(t) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(x)\, e^{itx}\, dx \qquad \textit{pour } \text{tout } t \in \mathbb{R}\,.$$

Démonstration. On a $f = \frac{1}{2\pi}\check{\widehat{f}}$ presque partout, donc *partout* car ces deux fonctions sont continues (exo). $\square$

Chapitre 10

Fonctions génératrices

Ce dernier chapitre concerne exclusivement les variables aléatoires *à valeurs dans* $\mathbb{N}$. Pour une telle va X, on va définir et étudier brièvement un objet qui rend exactement les mêmes services que la fonction caractéristique.

1. Définition et exemples

DÉFINITION 1.1. *Soit X une va* à valeurs dans $\mathbb{N}$. *La* ***fonction génératrice*** *de X est la fonction $G_X : [-1,1] \to \mathbb{R}$ définie par*

$$G_X(s) := \sum_{k=0}^{\infty} \mathbb{P}(X = k)\, s^k.$$

Remarque 1. La définition a un sens, et la fonction G_X est continue sur $[-1,1]$: en effet, comme la série (à termes positifs) $\sum \mathbb{P}(X = k)$ converge, la série définissant G_X converge normalement sur $[-1,1]$. De plus, comme $\sum_{k=0}^{\infty} \mathbb{P}(X = k) = 1$, on a $|G_X(s) \leqslant 1$ pour tout $s \in [-1,1]$.

Remarque 2. On peut également écrire $G_X(s)$ comme suit (ce qui est peut-être un peu déroutant) :

$$G_X(s) = \mathbb{E}(s^X).$$

EXEMPLE 1. l Si X suit une loi de Bernoulli $\mathfrak{B}(p)$, alors

$$G_X(s) = 1 - p + ps.$$

Démonstration. Exo. □

EXEMPLE 2. Si X suit une loi binomiale $\mathfrak{B}(n,p)$, alors

$$G_X(s) = \big(1 - p + ps\big)^n.$$

Démonstration. Le calcul se fait tout seul : on a

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

□

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

$$G_X(s) = e^{\lambda(s-1)}.$$

Démonstration. À nouveau, le calcul se fait tout seul :

$$G_X(s) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}\, s^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} = e^{-\lambda} e^{\lambda s}.$$

□

2. Propriétés de base

Proposition 2.1. *Si X est une va à valeurs entières, alors sa fonction génératrice G_X est de classe $\mathcal{C}^\infty$ sur $]-1,1[$, et on a*

$$\mathbb{P}(X=k) = \frac{G_X^{(k)}(0)}{k!} \qquad \text{pour tout } k \in \mathbb{N}.$$

Démonstration. C'est évident puisque G_X est la somme d'une série entière qui converge sur $]-1,1[$. □

Corollaire 2.2. *La fonction génératrice caractérise la loi : si X et Y sont deux va à valeurs dans $\mathbb{N}$ telles que $G_X = G_Y$, alors $\mathbb{P}_X = \mathbb{P}_Y$.*

Démonstration. Par la proposition, $\mathbb{P}(X=k) = \frac{G_X^{(k)}(0)}{k!} = \frac{G_Y^{(k)}(0)}{k!} = \mathbb{P}(Y=k)$ pour tout $k \in \mathbb{N}$. □

Proposition 2.3. *Si $X_1, \dots, X_n$ sont des va indépendantes à valeurs dans $\mathbb{N}$, alors*

$$G_{X_1+\cdots+X_n} = G_{X_1} \cdots G_{X_n}.$$

Démonstration. Il suffit de traiter le cas de deux va indépendantes X et Y, le cas général s'en déduisant par récurrence. Par indépendance, on a pour tout $k \in \mathbb{N}$:

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

Donc la série entière $\sum \mathbb{P}(X+Y=k)\,s^k$ est la série entière "produit" des séries entières $\sum \mathbb{P}(X=k)\,s^k$ et $\sum \mathbb{P}(Y=k)\,s^k$; et comme ces deux séries convergent normalement sur $[-1,1]$, on en déduit que pour tout $s \in [-1,1]$,

$$\sum_{k=0}^{\infty} \mathbb{P}(X+Y=k)\,s^k = \left(\sum_{k=0}^{\infty} \mathbb{P}(X=k)\,s^k\right)\left(\sum_{k=0}^{\infty} \mathbb{P}(Y=k)\,s^k\right).$$

Autrement dit, $G_{X+Y}(s) = G_X(s)\,G_Y(s)$. □

"Autre" démonstration. On a

$$G_{X_1+\cdots+X_n}(s) = \mathbb{E}\big(s^{X_1+\cdots+X_n}\big) = \mathbb{E}\big(s^{X_1}\cdots s^{X_n}\big)$$

et donc, par indépendance des X_i :

$$G_{X_1+\cdots+X_n}(s) = \mathbb{E}\big(s^{X_1}\big)\cdots\mathbb{E}\big(s^{X_n}\big) = G_{X_1}(s)\cdots G_{X_n}(s).$$

□

Exemple 1. 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 de Poisson $\mathcal{P}(\lambda+\mu)$.

Démonstration. Par la Proposition 2.3, on a

$$G_{X+Y}(s) = G_X(s)\,G_Y(s) = e^{\lambda(s-1)}e^{\mu(s-1)} = e^{(\lambda+\mu)(s-1)}.$$

On reconnaît la fonction génératrice d'une loi de Poisson de paramètre $\lambda+\mu$. Donc $X+Y \sim \mathcal{P}(\lambda+\mu)$ puisque la fonction génératrice caractérise la loi. □

Exemple 2. Si $X_1, \dots, X_n$ sont des va indépendantes suivant une loi de Bernoulli $\mathcal{B}(p)$, alors $X_1+\cdots+X_n$ suit la loi binomiale $\mathcal{B}(n,p)$.

Démonstration. Exo. □

3. Fonctions génératrices et moments

PROPOSITION 3.1. *Soit X une va à valeurs dans $\mathbb{N}$, et soit $p \in \mathbb{N}^*$. Alors X admet un moment d'ordre p si et seulement si G_X est p fois dérivable (à gauche) en 1 ; et dans ce cas, on a*

$$\begin{aligned} G^{(p)}(1) &= \sum_{k=p}^{\infty} k(k-1)\cdots(k-p+1)\,\mathbb{P}(X=k) \\ &= \mathbb{E}\big(X(X-1)\cdots(X-p+1)\big). \end{aligned}$$

En particulier, $X \in L^1$ si et seulement si G_X est dérivable à gauche en 1, et on a alors

$$\mathbb{E}(X) = G'(1).$$

Démonstration. Tout repose sur le fait suivant.

FAIT. Soit $\sum a_k s^k$ une série entière à coefficients a_k *positifs* tels que $\sum_0^\infty a_k < \infty$, et soit $f : [0,1] \to \mathbb{R}$ la fonction définie par $f(s) := \sum_0^\infty a_k s^k$. Alors

$$\frac{f(s)-f(1)}{s-1} \longrightarrow \sum_{k=1}^{\infty} k a_k \quad \text{quand } s \to 1^-.$$

Preuve du Fait. Pour tout $s \in [0,1[$, on a

$$\begin{aligned} \frac{f(s)-f(1)}{s-1} &= \frac{1}{s-1}\sum_{k=0}^{\infty} a_k(s^k-1) \\ &= \sum_{k=1}^{\infty} a_k\,\frac{s^k-1}{s-1} \qquad \text{car } s^k-1=0 \text{ si } k=0 \\ &= \sum_{k=1}^{\infty} \underbrace{a_k\,(1+s+\cdots+s^{k-1})}_{=:b_k(s)}. \end{aligned}$$

Comme les a_k sont $\geqslant 0$, on voit que chaque $b_k(s)$ est une fonction positive et croissante de $s \in [0,1[$, et $b_k(s) \to k a_k$ quand $s \to 1^-$. Par le *Théorème de convergence monotone pour les sommes*, on en déduit que

$$\frac{f(s)-f(1)}{s-1} \xrightarrow{s\to 1^-} \sum_{k=1}^{\infty} \lim_{s\to 1^-} b_k(s) = \sum_{k=1}^{\infty} k a_k.$$

$\square$

Si on applique le Fait avec $f(s) := G_X(s)$, *i.e.* $a_k := \mathbb{P}(X=k)$, on obtient

$$\lim_{s\to 1^-} \frac{G_X(s)-G_X(1)}{s-1} = \sum_{k=1}^{\infty} k\,\mathbb{P}(X=k) = \mathbb{E}(X).$$

Donc G_X est dérivable à gauche en 1 si et seulement si $\mathbb{E}(X) < \infty$, avec la bonne formule pour $G_X'(1)$. Ainsi, la proposition est démontrée pour $p=1$.

Si maintenant on suppose que G_X est dérivable à gauche en 1 et si on applique le Fait avec $f(s) = G'_X(s)$, *i.e.* $a_k = (k+1)\mathbb{P}(X = k+1)$, on obtient de même

$$\begin{aligned}\lim_{s\to 1^-} \frac{G'_X(s) - G'_X(1)}{s-1} &= \sum_{k=1}^{\infty} k(k+1)\mathbb{P}(X = k+1) \\ &= \sum_{k=2}^{\infty} k(k-1)\mathbb{P}(X = k) = \mathbb{E}\big(X(X-1)\big).\end{aligned}$$

Donc G_X est deux fois dérivable à gauche en 1 si et seulement si $\mathbb{E}(X) < \infty$ (pour que $G'_X(1)$ existe) et $\mathbb{E}\big(X(X-1)\big) < \infty$, avec la bonne formule pour $G''_X(1)$. De plus, comme $X^2 = X(X-1) + X$ et $X \geqslant 0$, on a $\Big(\mathbb{E}(X) < \infty \text{ et } \mathbb{E}(X(X-1)) < \infty\Big)$ si et seulement si $\mathbb{E}(X^2) < \infty$. Ainsi, la proposition est démontrée pour $p = 2$.

Pour p quelconque, on procède par récurrence (exo). □

4. Fonctions génératrices et convergence en loi

Le théorème suivant est l'analogue du théorème de Paul Lévy reliant convergence en loi et fonctions caractéristiques.

THÉORÈME 4.1. *Soit $(Z_n)_{n\in\mathbb{N}}$ une suite de va à valeurs dans $\mathbb{N}$, et soit Z une va à valeurs dans $\mathbb{N}$. La suite (Z_n) converge en loi vers Z si et seulement si $G_{Z_n}(s) \to G_Z(s)$ pour tout $s \in [-1, 1]$.*

Démonstration. Pour tout $s \in [-1, 1]$, on a $G_{Z_n}(s) = \mathbb{E}\big(\varphi_s(Z_n)\big)$ et $G_Z(s) = \mathbb{E}\big(\varphi_s(Z)\big)$, où $\varphi_s : \mathbb{N} \to \mathbb{R}$ est la fonction définie par $\varphi_s(k) := s^k$. La fonction φ_s est (continue) bornée sur $\mathbb{N}$ car $|s| \leqslant 1$; donc, si $Z_n \xrightarrow{\mathcal{L}} Z$, alors $G_{Z_n}(s) \to G_Z(s)$ pour tout $s \in [-1, 1]$.

La preuve de la réciproque est plus délicate. Dans la suite, on suppose que $G_{Z_n}(s) \to G_Z(s)$ pour tout $s \in [-1, 1]$.

Pour $k \in \mathbb{N}$, on posera

$$p_k := \mathbb{P}(Z = k) \qquad \text{et} \qquad p_{n,k} := \mathbb{P}(Z_n = k) \quad \text{pour } n \in \mathbb{N}.$$

Par définition, on a donc pour tout $s \in [-1, 1]$:

$$G_Z(s) = \sum_{k=0}^{\infty} p_k s^k \qquad \text{et} \qquad G_{Z_n}(s) = \sum_{k=0}^{\infty} p_{n,k} s^k.$$

Comme Z et les Z_n sont des va à valeurs dans $\mathbb{N}$, il suffit de montrer que

$$p_{n,k} \xrightarrow{n\to\infty} p_k \qquad \text{pour tout } k \in \mathbb{N}.$$

FAIT 0. Soit $\mathbb{D} := \{z \in \mathbb{C};\ |z| < 1\}$ le disque unité de $\mathbb{C}$. Pour tout $z \in \mathbb{D}$, on peut définir

$$G(z) := \sum_{k=0}^{\infty} p_k z^k \qquad \text{et} \qquad G_n(z) := \sum_{k=0}^{\infty} p_{n,k} z^k;$$

et les fonctions G et G_n sont holomorphes sur $\mathbb{D}$.

Démonstration. Comme $0 \leqslant p_k, p_{n,k} \leqslant 1$ pour tout $k \in \mathbb{N}$, les séries entières $\sum p_k z^k$ et $\sum p_{n,k} z^k$ ont des rayons de convergence au moins égaux à 1 ; donc le résultat est clair. □

Le point clé est le fait suivant, qui est un cas très particulier d'un résultat d'analyse complexe qu'on appelle le **Théorème de Montel**.

Fait 1. Toute sous-suite (f_n) de (G_n) possède une sous-suite (f_{n_i}) qui converge uniformément sur tout compact de $\mathbb{D}$.

Preuve du Fait 1. On va faire la preuve pour $(f_n) = (G_n)$, ce qui simplifiera les notations.

Étape 1. Il existe une suite strictement croissante d'entiers $(n_i)_{i\in\mathbb{N}}$ et une suite $(q_k)_{k\in\mathbb{N}} \subseteq [0,1]$ tels que

$$p_{n_i,k} \to q_k \qquad \text{pour tout } k \in \mathbb{N}.$$

Démonstration. Le résultat est "évident" si on connaît assez de topologie pour savoir que l'espace $[0,1]^{\mathbb{N}}$ est compact et métrisable pour la topologie produit. Refaisons cependant la preuve.

La suite $(p_{n,0})_{n\in\mathbb{N}}$ étant bornée, on peut en extraire une sous-suite convergente par le théorème de Bolzano-Weierstrass. Autrement dit, on peut trouver un ensemble infini $\mathbf{N}_0 \subseteq \mathbb{N}$ et un nombre $q_0 \in [0,1]$ tels que

$$p_{n,0} \to q_0 \qquad \text{quand } n \to \infty, \quad n \in \mathbf{N}_0.$$

De même, la suite $(p_{n,1})_{n\in\mathbf{N}_0}$ est bornée, donc admet une sous-suite convergente. Il existe donc un ensemble infini $\mathbf{N}_1 \subseteq \mathbf{N}_0$ et un nombre réel $q_1 \in [0,1]$ tels que $p_{n,1} \to q_1$ quand $n \in \mathbf{N}_1$ tend vers l'infini. Et ainsi de suite : par récurrence, on construit une suite décroissante $(\mathbf{N}_k)_{k\in\mathbb{N}}$ de parties infinies de $\mathbb{N}$ et une suite $(q_k) \subseteq [0,1]$ telles que pour tout $k \in \mathbb{N}$, la suite $(p_{n,k})_{n\in\mathbf{N}_k}$ converge vers q_k.

On définit alors une suite strictement croissante d'entiers $(n_i)_{i\in\mathbb{N}}$ en posant

$$n_0 := \min \mathbf{N}_0 \qquad \text{et} \qquad n_i := \min\{n \in \mathbf{N}_i;\ n > n_{i-1}\} \quad \text{pour } i \geqslant 1.$$

Par définition, pour tout $k \in \mathbb{N}$ fixé, la suite $(p_{n_i,k})_{i\geqslant k}$ est extraite de $(p_{n,k})_{k\in\mathbf{N}_k}$. Donc

$$p_{n_i,k} \xrightarrow{i\to\infty} q_k \qquad \text{pour tout } k \in \mathbb{N}.$$

□

Remarque. Le raisonnement utilisé (extraction d'une infinité de sous-suites emboitées les unes dans les autres, puis d'une seule sous-suite qui "fait le travail pour tout le monde") porte le nom de **procédé diagonal**.

Étape 2. La suite (G_{n_i}) converge uniformément sur tout compact de $\mathbb{D}$ vers la fonction f définie par $f(z) := \sum_{k=0}^{\infty} q_k z^k$.

Démonstration. Soit K un compact de $\mathbb{D}$, et soit $\varepsilon > 0$. Soit également $r < 1$ tel que $K \subseteq \overline{D}(0,1)$. Comme $r < 1$, la série $\sum r^k$ converge ; donc on peut trouver un entier N tel que

$$\sum_{k>N} r^k < \varepsilon.$$

Pour $z \in K$ et $i \in \mathbb{N}$, on a alors

$$
\begin{aligned}
|G_{n_i}(z) - f(z)| &= \left| \sum_{k=0}^{\infty} (p_{n_i,k} - q_k) z^k \right| \\
&\leqslant \sum_{k=0}^{N} |p_{n_i,k} - q_k|\, |z|^k + \sum_{k>N} |p_{n_i,k} - q_k|\, |z|^k \\
&\leqslant \sum_{k=0}^{N} |p_{n_i,k} - q_k| + 2 \sum_{k>N} r^k \qquad \text{car } 0 \leqslant p_{n_i,k}, q_k \leqslant 1 \\
&\leqslant \sum_{k=0}^{N} |p_{n_i,k} - q_k| + 2\varepsilon.
\end{aligned}
$$

Comme $p_{n_i,k} \to q_k$ pour $k = 0, \dots, N$, on peut maintenant trouver $i_0 \in \mathbb{N}$ tel que $\sum_{k=0}^{N} |p_{n_i,k} - q_k| \leqslant \varepsilon$. On obtient ainsi

$$\forall i \geqslant i_0\ \forall z \in K\ :\ |G_{n_i}(z) - f(z)| \leqslant 3\varepsilon;$$

ce qui termine la démonstration. □

La preuve du Fait 1 est maintenant terminée. □

Remarque. Pour l'Étape 2, on aurait pu aller plus vite en invoquant le *Théorème de convergence dominée pour les séries.* Exo : écrire les détails.

Fait 2. La suite (G_n) converge vers G uniformément sur tout compact de $\mathbb{D}$.

Preuve du Fait 2. Supposons qu'il existe un compact $K \subseteq \mathbb{D}$ tel que (G_n) ne converge pas uniformément vers G sur K. Alors, en posant

$$\|\varphi\|_K := \sup \big\{ |\varphi(z)|;\ z \in K \big\}$$

pour toute fonction $\varphi : \mathbb{D} \to \mathbb{C}$, on peut trouver un $\varepsilon > 0$ et une sous-suite (f_n) de (G_n) tels que

$$(*) \qquad \|f_n - G\|_K \geqslant \varepsilon \qquad \text{pour tout } n \in \mathbb{N}.$$

Par le Fait 1, la suite (f_n) possède une sous-suite (f_{n_i}) qui converge uniformément sur tout compact vers une fonction $f : \mathbb{D} \to \mathbb{C}$. La fonction f est holomorphe sur $\mathbb{D}$ car les f_{n_i} le sont. De plus, on a $f(s) = G(s)$ pour tout $s \in]-1,1[$ car (f_{n_i}) est une sous-suite de (G_n) et $G_n(s) = G_{Z_n}(s) \to G_Z(s) = G(s)$. Par le *principe des zéros isolés*, on en déduit que $f = G$.

Ainsi $f_{n_i} \to G$ uniformément sur K ; ce qui contredit visiblement $(*)$. □

On peut maintenant terminer la preuve du théorème. Par le Fait 2 et comme on a affaire à des fonctions holomorphes, on voit que pour tout $k \in \mathbb{N}$, la suite $(G_n^{(k)})$ converge vers $G^{(k)}$ uniformément sur tout compact de $\mathbb{D}$. En particulier

$$p_{n,k} = \frac{G^{(k)}(0)}{k!} \xrightarrow{n\to\infty} \frac{G^{(k)}(0)}{k!} = p_k \qquad \text{pour tout } k \in \mathbb{N};$$

et donc (Z_n) converge en loi vers Z. □

Illustration. Encore une preuve du Théorème des évènements rares.

Pour tout $n \in \mathbb{N}^*$, soient $X_{n,1}, \dots, X_{n,M_n}$ des va indépendantes à valeurs dans $\mathbb{N}$, chaque $X_{n,j}$ suivant une loi de Bernoulli $\mathcal{B}(p_{n,j})$. On suppose que

(i) $\lambda_n := \sum_{j=1}^{M_n}$ admet une limite $\lambda > 0$ quand $n \to \infty$;

(ii) $\varepsilon_n := \max(p_{n,1}, \dots, p_{n,M_n})$ tend vers 0 quand $n \to \infty$.

Il s'agit de montrer que $S_n := X_{n,1} + \dots + X_{n,M_n}$ converge en loi vers une va Z suivant la loi de Poisson $\mathcal{P}(\lambda)$.

Fixons $s \in [-1, 1]$. Pour tout $n \in \mathbb{N}^*$, on a par indépendance des $X_{n,j}$:

$$G_{S_n}(t) = \prod_{j=1}^{M_n} G_{X_{n,j}}(s) = \prod_{j=1}^{M_n} \big(1 - p_{n,j} + p_{n,j}s\big).$$

Par (ii), on a $1 - p_{n,j} + p_{n,j}s = 1 + (s-1)p_{n,j} > 0$ pour $j = 1, \dots, M_n$ si n est assez grand, donc on peut écrire

$$G_{S_n}(t) = \exp\left(\sum_{j=1}^{M_n} \log\big(1 + (s-1)p_{n,j}\big)\right).$$

De plus, toujours par (ii), on a

$$\log\big(1 + (s-1)p_{n,j}\big) = (s-1)p_{n,j} + o(p_{n,j}) \quad \text{quand } n \to \infty,$$

où le "o" est *uniforme* par rapport à $j \in \{1, \dots, M_n\}$. (Exo : écrire précisément ce que cela signifie.) Donc

$$\begin{aligned}
\sum_{j=1}^{M_n} \log\big(1 + (s-1)p_{n,j}\big) &= (s-1)\sum_{j=1}^{M_n} p_{n,j} + o\Big(\sum_{j=1}^{M_n} p_{n,j}\Big) \\
&= (s-1)\lambda_n + o(\lambda_n) \\
&= (s-1)\lambda_n + o(1) \qquad \text{car } \lambda_n \text{ reste borné.}
\end{aligned}$$

Comme $\lambda_n \to \lambda$, on en déduit que

$$G_{S_n}(s) \xrightarrow{n\to\infty} e^{(s-1)\lambda}.$$

On reconnaît au second membre la fonction génératrice de la loi de Poisson $\mathcal{P}(\lambda)$; donc (S_n) converge en loi vers une va $Z \sim \mathcal{P}(\lambda)$.