Introduction à la Théorie des Probabilités: notes de cours

Sébastien Ott

17 septembre 2024

Remarques générales

Si vous avez des remarques/trouvez des erreurs ou des typos/avez des suggestions d'amélioration des notes, envoyez moi un email à : sebastien.ott@epfl.ch.

Un point important (qui est inhérent à étudier les probabilités avant de faire de la théorie de la mesure) est qu'il sera plusieurs fois nécessaire d'utiliser des résultats qui ne seront démontrés qu'en Analyse IV ou en théorie de la mesure. Ces points seront spécifiés dans le texte et des précisions seront apportées en bas de page.

Je me suis basé sur les polycopiés de J. Aru (EPFL), I. Manolescu (Fribourg) et Y. Velenik (Genève) pour la rédaction de ces notes et pour le contenu du cours. [1]

1. Deux des polycopiés sont disponibles sur les pages web des auteurs : `https://homeweb.unifr.ch/manolesc/Pub/teaching/proba_stat_cours.pdf` et `https://www.unige.ch/math/folks/velenik/Cours/2018-2019/PS1819/probastat.pdf`

Table des matières

Chapitre 0

Introduction

0.1 Notations récurrentes

$\mathbb{R}$	L'ensemble des nombres réels
$\mathbb{Q}$	L'ensemble des nombres rationnels
$\mathbb{Z}$	L'ensemble des nombres entiers relatifs
$\mathbb{N}$	L'ensemble des nombres entiers strictement positifs $\mathbb{N} = \{1, 2, 3, \dots\}$
$\mathbb{R}_+$	L'ensemble des nombres réels positifs : $\mathbb{R}_+ = \{x \in \mathbb{R} : \ x \geq 0\}$
$\mathbb{Z}_+$	L'ensemble des nombres entiers positifs : $\mathbb{Z}_+ = \{n \in \mathbb{Z} : \ n \geq 0\}$
$\lfloor x \rfloor$	La partie entière inférieure de x
$\lceil x \rceil$	La partie entière supérieure de x
$\lvert A \rvert$	Le cardinal de A (le nombre d'éléments dans A)
B^A	L'ensemble des fonctions de A dans B : $B^A := \{f : A \to B\}$
$\mathcal{P}(\Omega)$	L'ensemble des parties de Ω : $\mathcal{P}(\Omega) = \{A : \ A \subset \Omega\}$
$\varnothing$	L'ensemble vide
$A_i \nearrow A$	Limite croissante d'ensembles : si $A_i \subset A_{i+1}$, $\lim_{i\to\infty} A_i = \bigcup_{i\geq 1} A_i \equiv A$
$A_i \searrow A$	Limite décroissante d'ensembles : si $A_i \supset A_{i+1}$, $\lim_{i\to\infty} A_i = \bigcap_{i\geq 1} A_i \equiv A$
$A \sqcup B$	Union disjointe ($A \sqcup B = A \cup B$ mais n'est utilisé que si $A \cap B = \varnothing$)
$\mathbb{1}_A$	Fonction caractéristique de A : $\mathbb{1}_A(x) = 1$ si $x \in A$ et 0 sinon
$\{X \in A\}$	Raccourci pour $\{\omega \in \Omega : \ X(\omega) \in A\}$
$P(X \in A)$	Raccourci pour $P\big(\{\omega \in \Omega : \ X(\omega) \in A\}\big)$, où $A \in \mathcal{B}(\mathbb{R})$ et X une variables aléatoire sur $(\Omega, \mathcal{F}, P)$.
$P(a \leq X \leq b)$	Raccourci pour $P(X \in [a, b])$.
$P(A, B)$	Notation alternative de $P(A \cap B)$.

0.2 Motivation : modélisation de phénomènes aléatoires

L'origine de la théorie des probabilité se trouve dans le besoin de décrire des *phénomènes aléatoires*. Ces derniers peuvent être divisé en deux catégories : les phénomènes *intrinsèquement aléatoire*, tels que ceux observés en physique quantique, et les phénomènes ou le terme "aléatoire" indique notre incapacité à faire des prédictions significatives (par exemple : pile ou face, lancer de dés, roulette, loterie, évolution du cours d'une action, trajectoire d'une poussière à la surface d'un lac...).

Dans tous les cas, on aimerait pouvoir modéliser ces phénomènes. La théorie des probabilité est le cadre formel qui permet de décrire les modèles liés aux phénomènes aléatoires. On va commencer par voir quelques tentatives "incomplètes" de donner du sens à l'étude de phénomènes aléatoires, avant de se tourner vers la version générale moderne. Pour la discussion informelle qui va suivre, il sera utile d'avoir en tête quelques exemples de problèmes que l'on aimerait modéliser. On notera EXP une *expérience* et on notera $\Omega = \Omega_{\text{EXP}}$ l'ensemble des *résultats possibles* de l'expérience. Quelques exemples d'expérience et les résultats associés :

- Expérience : pile ou face, $\Omega = \{\text{P}, \text{F}\}$;
- Expérience : lancé d'un dé à six faces, $\Omega = \{1, 2, 3, 4, 5, 6\}$;
- Expérience : regarder la couleur de la première voiture prenant la sortie Nyon après 8 :00, $\Omega = \{\text{rouge}, \text{bleu}, \text{noir}, \dots\}$;
- Expérience : mesurer la taille du 108èm client de la journée dans la Migros de l'EPFL, $\Omega = (0, \infty)$.

Pour une expérience EXP donnée, on notera Result(EXP) le résultat de l'expérience [1].

Pour une expérience donnée, on appellera *événement* une propriété qui peut être vérifiée ou non selon le résultat de l'expérience. Par exemple :

- dans le cas d'un lancé de dé à six faces : un événement possible est "le résultat est paire" ;
- dans le cas de la taille du client : un événement possible est "le client fait plus de 1m73" ;

Dans tous les cas, un événement est naturellement associé à un sous-ensemble des résultats.

0.3 Expériences répétables ou non, interprétation d'une "probabilité"

Essayons de préciser ce que l'on aimerait qu'une "probabilité" modélise. Dans les expériences que l'on veut traiter, certaines sont *répétables* : par exemple, un lancer de pièce ou de dé. Dans ce cas, on aimerait que la *probabilité d'un événement* représente

1. Dans un sens, Result(EXP) est l'objet qui est *aléatoire* et que l'on cherche à décrire.

sa *fréquence d'occurrence*[2] : pour donner du sens à la fréquence d'occurrence d'un événement $A \subset \Omega$, on procède comme suit

— on répète n fois l'expérience EXP, ce qui donne lieu à n expériences $\mathrm{EXP}_1, \ldots, \mathrm{EXP}_n$, remarquons que l'on peut aussi voir ceci comme une plus grande expérience ;

— on regarde la proportion des fois où A est vérifié :

$$X_n(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_A\big(\mathrm{Result}(\mathrm{EXP}_i)\big);$$

— on aimerait que la probabilité de A, $P(A)$, reflète la valeur de $X_n(A)$ quand n est pris grand.

Dans d'autre situations, telles que prédire un crash boursier ou des tremblements de terre, ces événements sont "rares" et difficilement reproductibles, la probabilité d'un événement dans ce cas devrait représenter le *risque d'occurrence* de cet événement[3].

0.4 Tirage uniforme (modèle de Laplace)

Une première approche à la modélisation de phénomènes aléatoires est d'imaginer que

— notre expérience n'admet qu'un nombre fini de résultats possibles ;

— aucun des résultats n'est favorisé par rapport aux autres.

C'est par exemple le cas quand on veut modéliser un lancer de pièce ou de dés équilibrés, mais cela devient faux si le dé est truqué (par exemple en lestant une des face). Si l'on se trouve dans ce cas, on peut définir la probabilité d'un événement comme suit :

1. On note Ω les résultats possibles de notre expérience ($\Omega = \{1, 2, 3, 4, 5, 6\}$ dans le cas d'un dés à six faces).
2. On voit notre événement A comme un ensemble

$$A \equiv \{\omega \in \Omega : \ A \text{ est vérifié}\},$$

(par exemple, l'ensemble associé à "tirer un nombre paire" dans le cas du dés est $\{2, 4, 6\}$).

3. On définit la *probabilité de l'événement* A par

$$P(A) = \frac{|A|}{|\Omega|}. \tag{1}$$

2. Historiquement, ça a même été la manière de *définir* une probabilité. On verra dans la section consacrée à la Loi dans Grand Nombres que cette propriété est une *conséquence* de l'approche moderne.

3. Il est beaucoup plus difficile de démontrer un résultat qui dirait qu'un modèle probabiliste est un bon modèle pour de tels événements.

La motivation pour ce modèle est est la probabilité d'un événement est donnée par sa fréquence de réalisation : si on effectue la même expérience beaucoup de fois, la proportion des fois où A est réalisé va converger vers $P(A)$ [4].

Ce modèle possède déjà de bonnes propriétés qui ré-apparaîtront par la suite :

Lemme 0.4.1. *Soit Ω un ensemble fini (l'ensemble des résultats). Soit $P : \{A \subset \Omega\} \to [0,1]$ définit par (1). Alors,*

- $P(\varnothing) = 0$;
- *si $A, B \subset \Omega$ ne peuvent par être vérifiés en même temps (i.e. : $A \cap B = \varnothing$), alors $P(A \cup B) = P(A) + P(B)$ [5] ;*
- *si $A \subset \Omega$, alors $P(\Omega \setminus A) = 1 - P(A)$.*

La preuve est laissée en exercice.

0.5 Modèles de probabilité discrets

Le modèle de Laplace possède deux limitations sévères (qui sont les hypothèses décrites au début de la section précédente). Une première amélioration est de regarder des *modèles de probabilité discrets* :

Définition 0.5.1 (Mesure de probabilité sur espace discret)**.** Soit Ω un ensemble dénombrable. Une *densité de probabilité* sur Ω est une fonction $f : \Omega \to [0,1]$ telle que $\sum_{\omega \in \Omega} f(\omega) = 1$. On définit alors la *mesure de probabilité*, $P : \mathcal{P}(\Omega) \to [0,1]$, associée par

$$P(A) = \sum_{\omega \in A} f(\omega), \quad A \subset \Omega.$$

On peut voir cette généralisation comme suit : on se donne la probabilité des *événements élémentaires* (i.e. : des résultats de notre expérience) via la fonction f et on construit la probabilité d'événements généraux (événements composites) à partir de celle-ci. Ici encore, on montrera que si notre expérience est répétable, la fréquence d'occurrence de A sera bien donnée par $P(A)$.

0.6 Une famille de modèles de probabilité continus

Les exemples précédents reposent sur le fait qu'il est dans chaque cas possible de définir la probabilité *de chaque résultat possible* de notre expérience. Si on considère des expériences telles que "une goutte de pluie qui tombe sur un parking" et que l'on se demande quelle est la probabilité que la goutte tombe sur la place 173, on a intuitivement envie de dire que cette probabilité doit être proportionelle à la surface de la place. Mais quand on se demande quelle est la probabilité que la goutte tombe *à un point précis*, on voit vite que cette probabilité doit être 0 : chaque point devrait avoir la même probabilité mais il y a une infinité de points (non-dénombrable qui plus est). On est amené à définir

4. On démontrera ceci quand on discutera de la Loi des Grands Nombres.
5. On appelle cette propriété *l'additivité finie*.

Définition 0.6.1. Soit $f : \mathbb{R}^d \to \mathbb{R}$ une fonction Riemann-intégrable telle que $f \geq 0$ et $\int_{\mathbb{R}^d} f(x)dx = 1$. On définit la probabilité de $A \subset \mathbb{R}^d$ par

$$P(A) = \int_{\mathbb{R}^d} dx \mathbb{1}_A(x) f(x),$$

quand cette intégrale fait sens.

On peut voir que cette notion de "probabilité" possède les mêmes propriétés que les précédentes. On remarque aussi que l'on ne peut pas donner du sens à $P(A)$ pour *tout* $A \subset \mathbb{R}^d$. En effet, prendre $d = 1$ et $f = \mathbb{1}_{[0,1]}$ définit une probabilité au sens ci-dessus, mais l'ensemble $A = \mathbb{Q} \cap [0, 1]$ n'a pas une probabilité bien définie [6].

6. On verra que pour cet exemple précis, on peut donner du sens à l'intégrale de $\mathbb{1}_{\mathbb{Q}\cap[0,1]}$. En revanche, l'existence d'ensembles pour lesquels on ne peut pas définir une probabilité est un vrai problème qui ne peut en général pas être évité dans le cas non-dénombrable, du moins tant que l'on travail en acceptant l'axiome du choix.

Chapitre 1

Espaces de probabilité généraux

On se tourne maintenant vers la structure générale qui contient les tentatives précédentes de définir une probabilité comme cas particulier. La structure prend en compte le problème mentionné dans le dernier exemple de la section précédente : il n'est en général pas possible de définir la probabilité de tous les sous ensembles de notre ensemble de résultats.

1.1 Mesures de probabilité et espaces d'événements

Définition 1.1.1 (Tribu)**.** Soit Ω un ensemble. Une *tribu* (ou *σ-algèbre*) sur Ω un ensemble $\mathcal{F} \subset \mathcal{P}(\Omega)$ qui satisfait

1. $\mathcal{F}$ contient l'ensemble vide ($\varnothing \in \mathcal{F}$).
2. $\mathcal{F}$ est stable par passage au complémentaire ($A \in \mathcal{F} \implies \Omega \setminus A \in \mathcal{F}$).
3. $\mathcal{F}$ est stable par union dénombrable (si pour tout $i \in \mathbb{N}$, $A_i \in \mathcal{F}$, alors $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{F}$).

Définition 1.1.2 (Espace probabilisable)**.** Un *espace probabilisable* (ou *espace mesurable*) est une paire $(\Omega, \mathcal{F})$ où Ω est un ensemble et $\mathcal{F}$ est une tribu sur Ω.

Exemple 1.1.1. Deux tribus jouent un rôle important pour garantir l'existence de certain objets : la *tribu triviale* : $\mathcal{F} = \{\varnothing, \Omega\}$ et la *tribu discrète* : $\mathcal{F} = \mathcal{P}(\Omega)$.

Une tribu est aussi appelée "ensemble d'événements" : il s'agira de tout les ensembles desquels on pourra mesurer la probabilité. De la définition, on déduit

Théorème 1.1.1 (Propriétés élémentaires des tribus)**.** *Soit Ω un ensemble et $\mathcal{F}$ une tribu sur Ω. Alors les points suivants sont tous vérifiés.*

1. *$\Omega \in \mathcal{F}$.*
2. *Stabilité par intersections finies : si $A, B \in \mathcal{F}$, alors $A \cap B \in \mathcal{F}$.*
3. *Si $A, B \in \mathcal{F}$, alors $A \setminus B \in \mathcal{F}$.*
4. *Stabilité par intersections dénombrables : si $A_1, A_2, \cdots \in \mathcal{F}$, alors $\bigcap_{i\geq 1} A_i \in \mathcal{F}$.*
5. *Stabilité par limites croissantes : soit $A_i \in \mathcal{F}, i \geq 1$ une suite croissante d'ensembles. Alors, $A \in \mathcal{F}$, où $A_i \nearrow A$.*

6. *Stabilité par limites décroissantes : soit $A_i \in \mathcal{F}, i \geq 1$ une suite décroissante d'ensembles. Alors, $A \in \mathcal{F}$, où $A_i \searrow A$.*

Démonstration. Pour simplifier, on note $A^c = \Omega \setminus A$. Le premier point suit de $\Omega = \varnothing^c$. Le second point de $A \cap B = (A^c \cup B^c)^c$. Le troisième de $A \setminus B = ((A \cup B)^c \cup B)^c$. Le quatrième de $\bigcap_{i\geq 1} A_i = \Big(\bigcup_{i\geq 1} A_i^c\Big)^c$. Finalement, une limite croissante est une union dénombrable et une limite décroissante est une intersection dénombrable. □

Définition 1.1.3 (Mesure de probabilité)**.** Soit Ω un ensemble et $\mathcal{F}$ une tribu sur Ω. Une *mesure de probabilité* (ou *loi de probabilité* ou simplement *probabilité*) sur $(\Omega, \mathcal{F})$ est une fonction $P : \mathcal{F} \to [0, 1]$ telle que

1. $P(\Omega) = 1$.
2. Si $A_i \in \mathcal{F}, i \in \mathbb{N}$ est une famille dénombrable d'ensembles deux à deux disjoints [1], alors [2]
$$P\Big(\bigsqcup_{i\in\mathbb{N}} A_i\Big) = \sum_{i\in\mathbb{N}} P(A_i).$$

Plus généralement, une *mesure* μ sur $(\Omega, \mathcal{F})$ est une fonction $\mu : \mathcal{F} \to \mathbb{R}_+ \cup \{+\infty\}$ qui satisfait $\mu(\varnothing) = 0$ et le second point de la définition ci-dessus. Une mesure de probabilité est simplement une mesure telle que $\mu(\Omega) = 1$.

Définition 1.1.4 (Espace de probabilité)**.** Un *espace de probabilité* est un triplet $(\Omega, \mathcal{F}, P)$ où Ω est un ensemble, $\mathcal{F}$ est une tribu sur Ω et P est une mesure de probabilité sur $(\Omega, \mathcal{F})$.

Les mesures de probabilités possèdent plusieurs propriétés qui vont être clé quand on voudra les manipuler.

Théorème 1.1.2 (Propriétés élémentaires des mesures de probabilité)**.** *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Alors,*

1. *$P(\varnothing) = 0$;*
2. *pour tout $A \in \mathcal{F}$, $P(\Omega \setminus A) = 1 - P(A)$;*
3. *si $A, B \in \mathcal{F}$ sont tels que $A \subset B$, alors $P(B) = P(A) + P(B \setminus A)$, en particulier $P(A) \leq P(B)$;*
4. *σ-additivité finie : si $A_1, \ldots, A_n \in \mathcal{F}$, $n \geq 1$ sont deux à deux disjoints alors*
$$P\Big(\bigsqcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} P(A_i);$$
5. *σ-sous-additivité : pour I un ensemble fini ou dénombrable et $A_i \in \mathcal{F}$, $i \in I$ une collection d'événements,*
$$P\Big(\bigcup_{i\in I} A_i\Big) \leq \sum_{i\in I} P(A_i) \ ;$$

1. $i \neq j \implies A_i \cap A_j = \varnothing$
2. Cette propriété généralise l'additivité finie et est appelée *additivité dénombrable.*

6. *pour une suite d'événements* $A_i \in \mathcal{F}, i \geq 1$ *telle que* $A_i \subset A_{i+1}$, $P(A_i) \xrightarrow{i\to\infty} P(A)$ *où* $A = \bigcup_{i\geq 1} A_i$;
7. *pour une suite d'événements* $A_i \in \mathcal{F}, i \geq 1$ *telle que* $A_i \supset A_{i+1}$, $P(A_i) \xrightarrow{i\to\infty} P(A)$ *où* $A = \bigcap_{i\geq 1} A_i$;
8. *pour* $A, B \in \mathcal{F}$, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Démonstration. L'additivité dénombrable implique l'additivité finie en prenant $A_i = \varnothing$ pour $i > n$. Les deux premiers points suivent : le premier point suit du second avec $A = \varnothing$, et comme $\Omega \setminus A \in \mathcal{F}$ et $A \cap (\Omega \setminus A) = \varnothing$, $1 = P(\Omega) = P(A \cup (\Omega \setminus A)) = P(A) + P(\Omega \setminus A)$. Le troisième point suit de $B \setminus A \in \mathcal{F}$ et donc pour $A \subset B$, $P(B) = P(A) + P(B \setminus A) \geq P(A)$. Dans le même esprit, le dernier point suit de $A \cup B = (A \cap B) \sqcup (A \setminus (A \cap B)) \sqcup (B \setminus (A \cap B))$ et donc, par le troisième point,

$$\begin{aligned} P(A \cup B) = P(A \cap B) + P(B \setminus (A \cap B)) + P(A \setminus (A \cap B)) = \\ = P(A \cap B) + P(B) - P(A \cap B) + P(A) - P(A \cap B). \end{aligned}$$

On se tourne vers la sous-additivité (on peut se restreindre au cas dénombrable, le cas fini suit en prenant une infinité d'ensembles $= \varnothing$) : soient $A_i \in \mathcal{F}$, $i \geq 1$. On définit $B_1 = A_1$ et $B_n = A_n \setminus \left(\bigcup_{i=1}^{n-1} A_i\right) \in \mathcal{F}$ pour $n \geq 2$. Alors, $\bigcup_{i\geq 1} A_i = \bigcup_{i\geq 1} B_i$ et $B_1, B_2, \ldots$ sont deux à deux disjoints. En particulier, par σ-additivité et monotonie de P,

$$P\Big(\bigcup_{i\geq 1} A_i\Big) = P\Big(\bigsqcup_{i\geq 1} B_i\Big) = \sum_{i\geq 1} P(B_i) \leq \sum_{i\geq 1} P(A_i).$$

Considérons maintenant les limites croissante : soit $A_1 \subset A_2 \subset \ldots$ une suite croissante dans $\mathcal{F}$. Notons $A = \bigcup_{i\geq 1} A_i$ et introduisons $B_1 = A_1$, $B_n = A_n \setminus A_{n-1} \in \mathcal{F}$. Comme $B_1, B_2, \ldots$ sont deux-à-deux disjoints,

$$1 \geq P(A) = P\Big(\bigsqcup_{i\geq 1} B_i\Big) = \sum_{i\geq 1} P(B_i),$$

ce qui implique que $\sum_{i>n} P(B_i) = P(\cup_{i>n} B_i) = P(A \setminus A_n)$ tend vers 0 quand n tend vers l'infini. Donc, $P(A) - P(A_n) = P(A \setminus A_n) \xrightarrow{n\to\infty} 0$.

Pour le cas décroissant, on se ramène au cas croissant en introduisant $B_i = \Omega \setminus A_i$ qui est une suite croissante d'ensembles. On a donc

$$\lim_{i\to\infty} P(A_i) = \lim_{i\to\infty} [1 - P(B_i)] = 1 - P\Big(\bigcup_{i\geq 1} B_i\Big) = P\Big(\Omega \setminus \bigcup_{i\geq 1} B_i\Big) = P\Big(\bigcap_{i\geq 1} A_i\Big).$$

□

Une généralisation du dernier point est le principe d'inclusion-exclusion :

Lemme 1.1.3. *Soit* $(\Omega, \mathcal{F}, P)$ *un espace de probabilité,* $n \geq 1$ *et* $A_1, \ldots, A_n \in \mathcal{F}$. *On définit*

$$a_k = \sum_{1\leq i_1<i_2<\cdots<i_k\leq n} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}).$$

Alors,

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{k=1}^{n} (-1)^{k+1} a_k.$$

De plus, pour $1 \leq l \leq n/2$ *entier*[3],

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) \begin{cases} \leq \sum_{k=1}^{2l-1} (-1)^{k+1} a_k \\ \geq \sum_{k=1}^{2l} (-1)^{k+1} a_k \end{cases} .$$

Démonstration. Laissée en exercice. Indice : procéder par induction. □

Finalement, un fait très utile et élémentaire est une forme de convexité de l'ensemble des mesures de probabilité :

Lemme 1.1.4. *Soit* $(\Omega, \mathcal{F})$ *un espace probabilisable. Soit* I *un ensemble dénombrable ou fini. Soient* $P_i, i \in I$ *des mesures de probabilité sur* $(\Omega, \mathcal{F})$*. Soit* $p : I \to [0,1]$ *tel que* $\sum_{i \in I} p(i) = 1$*. Alors, la fonction* $P : \mathcal{F} \to \mathbb{R}$ *définie par*

$$P(A) = \sum_{i \in I} p(i) P_i(A),$$

est une mesure de probabilité sur $(\Omega, \mathcal{F})$*. On notera* $P = \sum_{i \in I} p(i) P_i$ [4]*.*

Démonstration. On a bien que $P(\varnothing) = \sum_{i \in I} p(i) P_i(\varnothing) = \sum_{i \in I} p(i) 0 = 0$. Pour $A_1, A_2, \cdots \in \mathcal{F}$ deux à deux disjoints,

$$\begin{aligned} P\Big(\bigsqcup_{j \geq 1} A_j\Big) &= \sum_{i \in I} p(i) P_i\Big(\bigsqcup_{j \geq 1} A_j\Big) \\ &= \sum_{i \in I} p(i) \sum_{j \geq 1} P_i(A_j) \\ &= \sum_{j \geq 1} \sum_{i \in I} p(i) P_i(A_j) = \sum_{j \geq 1} P(A_j) \end{aligned}$$

où on a utilisé que tous les termes sont positifs pour pouvoir échanger les sommes. □

1.2 Pourquoi les tribus ?

On peut (à juste titre) se demander pourquoi utiliser une autre tribu que $\mathcal{P}(\Omega)$ (et donc, pourquoi utiliser des tribus tout court). Il y a plusieurs raison à cela.

La première est que l'on veut souvent que la mesure que probabilité que l'on considère ait certaines propriétés (comme respecter certaines symétries, ou donner une probabilité proportionnelle au volume...). Le soucis est qu'il n'est en général pas possible de définir une mesure de probabilité avec une propriété requise sur tout $\mathcal{P}(\Omega)$.

La seconde est que les tribus donnent un bon formalisme pour décrire la notion *d'information partielle.*

3. Ces majorations/minorations successives sont appelées *inégalités de Bonferroni.*

4. Remarquons que p définit juste une mesure de probabilité sur $(I, \mathcal{P}(I))$). Ce lemme dit en essence que le processus de choisir une mesure de probabilité au hasard puis échantillonner sous cette mesure donne a nouveau une mesure de probabilité. Il est possible de généraliser à des ensembles I plus grands via l'intégration contre une mesure plutôt que la sommation.

1.2.1 Suite infinie de pile ou face

On aimerait modéliser l'expérience "lancer une infinité de fois une pièce équilibrée". L'ensemble des résultats est alors $\Omega = \{0,1\}^{\mathbb{N}}$ (on note 0 pour pile et 1 pour face). Réfléchissons maintenant à quelles propriétés de la mesure de probabilité seraient naturelles. On aimerait que la probabilité de faire pile au lancer n soit $1/2$ (pièce équilibrée) et que cette probabilité ne dépende pas du résultat des autres lancés (lancés indépendants). On peut formaliser ceci par : pour tout événement A et pour tout $n \geq 1$,

$$P(A) = P(T_n A), \tag{1.1}$$

où $T_n : \mathcal{P}(\Omega) \to \mathcal{P}(\Omega)$ est définie par

$$T_n A = \{T_n(\omega) : \ \omega \in A\},$$
$$T_n(\omega) = (\omega_1, \dots, \omega_{n-1}, 1 - \omega_n, \omega_{n+1}, \dots).$$

En mots : T_n inverse le résultat du $n$ième lancé. Cet exemple nous donne un cas dans lequel notre mesure P ne peut pas être définie sur $\mathcal{P}(\Omega)$ tout entier :

Lemme 1.2.1. *Il n'existe pas de mesure P sur $(\Omega, \mathcal{P}(\Omega))$ telle que $P(T_n A) = P(A)$ pour tout $A \subset \Omega$ et tout $n \geq 1$.*

Démonstration. On procède par l'absurde et on suppose l'existence de P comme dans l'énoncé. On va construire un ensemble $A \in \mathcal{P}(\Omega)$ qui va avoir un comportement pathologique. On définit la relation d'équivalence suivante sur Ω : $\omega \sim \omega'$ si $\omega(i) = \omega'(i)$ pour tout i assez grand :

$$\omega \sim \omega' \iff (\exists N \in \mathbb{N} : \ \omega(i) = \omega(i)' \ \forall i \geq N).$$

On note alors $\mathcal{C}$ l'ensemble des classes d'équivalence sous $\sim$. L'axiome du choix nous dit qu'il est alors possible de choisir un représentant $\omega_C \in C \subset \Omega$ pour chaque $C \in \mathcal{C}$ simultanément. Fixons un tel choix. On définit

$$A = \{\omega_C : \ C \in \mathcal{C}\},$$

l'ensemble des représentants. Définissons aussi

$$\mathcal{J} = \{J \subset \mathbb{N} : \ |J| < \infty\} = \{\varnothing\} \sqcup \Big(\bigsqcup_{M \geq 1} \{J \subset \mathbb{N} : \ \max J = M\} \Big),$$

l'ensemble des sous-ensembles de $\mathbb{N}$ de taille finie. La dernière égalité nous garantit que $\mathcal{J}$ est dénombrable (union dénombrable d'ensembles finis). Pour $J \in \mathcal{J}$, on dénote alors $T_J = T_{n_1} \circ T_{n_2} \circ \cdots \circ T_{n_k}$ où $J = \{n_1, n_2, \dots, n_k\}$ (notez que $T_n \circ T_m = T_m \circ T_n$ quand $n \neq m$). On peut alors considérer les ensembles

$$T_J A = \{\omega \in \Omega : \ \exists C \in \mathcal{C}, T_J(\omega_C) = \omega\}.$$

On observe que

— si $J \neq I$, alors $T_J A \cap T_I A = \varnothing$. En effet, supposons que $T_J A \cap T_I A \neq \varnothing$. Alors, il existe $C_1, C_2 \in \mathcal{C}$ tels que $T_J(\omega_{C_1}) = T_I(\omega_{C_2})$. Mais, par définition de $\sim$, comme I, J sont finis, on a que $\omega_{C_1} \sim \omega_{C_2}$ et donc que $\omega_{C_1} = \omega_{C_2}$ ce qui implique $I = J$.

— $\bigcup_{J\in\mathcal{J}} T_J A = \Omega$. En effet, pour tout $\omega \in \Omega$, il existe $C \in \mathcal{C}$ tel que $\omega_C \sim \omega$. En particulier, il existe $J \in \mathcal{J}$ tel que $T_J(\omega_C) = \omega$.

De ces observations, on déduit

$$1 = P(\Omega) = P\Big(\bigsqcup_{J\in\mathcal{J}} T_J A\Big) = \sum_{J\in\mathcal{J}} P(T_J A) = \sum_{J\in\mathcal{J}} P(A),$$

ce qui est impossible (soit $P(A) = 0$ et on a $1 = 0$, soit $P(A) > 0$ et on a $1 = \infty$) [5]. □

1.2.2 Autre utilité des tribus : formalisation de l'information partielle

Les tribus, au delà de donner un ensemble d'ensembles sur lequel il est possible de définir des mesures, sont aussi une manière d'encoder l'information que l'on a sur un système/une expérience. On va voir ceci sur un exemple.

Considérons l'expérience "lancer trois pièces de monnaie distinctes". L'ensemble des résultats est alors

$$\Omega = \{000, 001, 010, 100, 011, 101, 110, 111\},$$

où on a noté 0 pour pile et 1 pour face. On se met maintenant dans trois situations d'observation :

1. on peut voir les trois pièces devant nous ;
2. les deux premières pièces sont tombées dans un ravin, on ne voit donc pas leurs résultat ;
3. un ami regarde les pièces et nous dit seulement si il y a plus de piles ou plus de faces.

À chacune de ces situation, on peut associer une tribu qui encode la précision de l'observation effectuée :

1. on a une information complète, la tribu naturelle est $\mathcal{F} = \mathcal{P}(\Omega)$;
2. on ne connaît pas les résultats des deux premiers lancés, les événements observables ne doivent donc pas différencier si ceux-ci sont pile ou face, on peut encoder ce fait en prenant la tribu
$$\mathcal{F} = \big\{\varnothing, \{000, 100, 010, 110\}, \{001, 101, 011, 111\}, \Omega\big\};$$
3. en suivant la même idée qu'au point précédent, on peut encoder l'information observable en prenant la tribu
$$\mathcal{F} = \big\{\varnothing, \{000, 001, 010, 100\}, \{011, 101, 110, 111\}, \Omega\big\}.$$

1.3 Quelques tribus omniprésentes

Il existe de plusieurs procédés pour construire des tribus ayant de bonnes propriétés. On va voir les plus fondamentaux ici.

5. Il est à noter que l'axiome du choix (dans sa version la plus forte) est requis pour construire un ensemble pour lequel (1.1) conduit à une pathologie. Il est possible de montrer que l'axiome du choix est nécessaire pour que le Lemme soit vrai.

1.3.1 Tribu engendrée par un ensemble

Le procédé le plus standard est la construction d'une tribu contenant une famille prescrite d'événements. Il repose sur l'observation

Lemme 1.3.1. *Soit Ω, I deux ensembles (quelconques). Soient $\mathcal{F}_i, i \in I$ une famille de tribus sur Ω. Alors, $\mathcal{F} = \bigcap_{i\in I} \mathcal{F}_i$ est une tribu sur Ω.*

Démonstration. Remarquons d'abord que $\varnothing \in \mathcal{F}_i$ pour tout $i \in I$ (car $\mathcal{F}_i$ est une tribu). Par conséquent, $\varnothing$ est dans leur intersection. Ensuite, si $A \in \mathcal{F}$, alors $A\mathcal{F}_i$ pour tout $i \in I$. Donc $\Omega \setminus A \in \mathcal{F}_i$ pour tout $i \in I$, et donc $\Omega \setminus A \in \mathcal{F}$. Finalement, si $A_1, A_2, \cdots \in \mathcal{F}$, alors $A_1, A_2, \cdots \in \ \mathcal{F}_i$ pour tout $i \in I$. En particulier, $\bigcup_{j\geq 1} A_j \in \mathcal{F}_i$ pour tout $i \in I$ et donc $\bigcup_{j\geq 1} A_j \in \mathcal{F}$. □

On définit alors :

Définition 1.3.1 (Tribu engendrée)**.** Soit $A \subset \mathcal{P}(\Omega)$. On définit la *tribu engendrée par A* via

$$\sigma(A) = \bigcap_{\mathcal{F} \text{ tribu sur } \Omega:\ A\subset\mathcal{F}} \mathcal{F}.$$

Comme $\mathcal{P}(\Omega)$ est une tribu, l'ensemble $\{\mathcal{F} \subset \mathcal{P}(\Omega) :\ \mathcal{F} \text{ tribu}, A \subset \mathcal{F}\}$ est non-vide. Par le Lemme précédent, $\sigma(A)$ est donc bien définit comme tribu.

1.3.2 Tribu produit

La tribu produit est une premier exemple d'utilisation de la tribu engendrée.

Définition 1.3.2. Soient $(\Omega_i, \mathcal{F}_i)$, $i = 1, 2$ deux espaces probabilisables. On définit la *tribu produit* $\mathcal{F}_1 \otimes \mathcal{F}_2$ sur l'espace produit $\Omega_1 \times \Omega_2$ par

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma\big(\{A \times B :\ A \in \mathcal{F}_1, B \in \mathcal{F}_2\}\big).$$

1.3.3 Tribu borélienne sur $\mathbb{R}^d$

Une *tribu borélienne* peut être définie pour tout espace topologique $(\Omega, \mathcal{T})$. Dans ce cas, la tribu borélienne associée à $(\Omega, \mathcal{T})$ est simplement $\sigma(\mathcal{T})$, la tribu engendrée par les ouverts. Dans le cas de $\mathbb{R}^d$, on a une caractérisation plus simple.

Définition 1.3.3 (Tribu borélienne sur $\mathbb{R}^d$)**.** Notons $\mathcal{I}$ l'ensemble des pavés fermés :

$$\mathcal{I} = \{[a_1, b_1] \times \cdots \times [a_d, b_d] :\ a_1, b_1, \ldots, a_d, b_d \in \mathbb{R}\}.$$

La *tribu borélienne sur* $\mathbb{R}^d$ est alors $\sigma(\mathcal{I})$.

On peut de la même façon obtenir la tribu borélienne à partir des pavés ouverts. On peut également se restreindre à $a_1, b_1, \ldots, a_d, b_d \in \mathbb{Q}$, ce qui a l'avantage de générer la tribu borélienne en utilisant un nombre dénombrable d'ensembles.

1.3.4 Tribu restreinte

Le dernier exemple est de restreindre une tribu sur Ω pour obtenir une tribu sur $\Omega' \subset \Omega$.

Lemme 1.3.2 (Restriction de tribu)**.** *Soit $(\Omega, \mathcal{F})$ un espace mesurable. Soit $\Omega' \in \mathcal{F}$. Alors, l'ensemble*

$$\mathcal{F}|_{\Omega'} := \{A \in \mathcal{F} : A \subset \Omega'\},$$

est une tribu sur Ω' : la restriction *de $\mathcal{F}$ à Ω'.*

Démonstration. On sait que $\varnothing \in \mathcal{F}|_{\Omega'}$ car $\varnothing \in \mathcal{F}$ et $\varnothing \subset \Omega'$. Si $A \in \mathcal{F}$ et $B \subset \Omega'$, on a aussi que $\Omega' \setminus A \in \mathcal{F}$ et $\Omega' \setminus A \subset \Omega'$ donc $\Omega' \setminus A \in \mathcal{F}|_{\Omega'}$. Finalement, si $A_1, A_2, \cdots \in \mathcal{F}$ sont tous inclus dans Ω', alors $\bigcup_{i\geq 1} A_i \subset \Omega'$ et donc $\bigcup_{i\geq 1} A_i \in \mathcal{F}|_{\Omega'}$. □

1.4 (Non-examinable) Digression : imports de théorie de la mesure

Il y aura plusieurs résultats de théorie de la mesure que l'on aura envie d'utiliser dans ce cours. Malheureusement, leurs preuves demandent de développer plus en détail la théorie générale. On se contentera donc d'énoncer les résultats comme "boîtes noires". Les preuves de ces différents points seront vu en Analyse IV et en théorie de la mesure.

1.4.1 Mesure et intégrale de Lebesgue

On fait une rapide digression pour introduire une notion qui sera utilisée régulièrement mais qui ne sera introduite que dans le cours d'Analyse IV : la *mesure de Lebesgue*. Dans la pratique, pour tous les calculs que nous ferons, l'intégrale de Riemann sera suffisante, mais elle n'est pas suffisante pour définir la mesure dont on aura besoin. On utilisera la boîte noire suivante :

Théorème 1.4.1 (Mesure de Lebesgue sur $\mathbb{R}^d$)**.** *Il existe une unique mesure, ℓ, sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ telle que*

$$\ell([a_1, b_1] \times \cdots \times [a_d, b_d]) = \prod_{i=1}^{d} |a_i - b_i|$$

pour tout $a_1, b_1, \ldots, a_d, b_d \in \mathbb{R}$.

Théorème 1.4.2 (Intégrale de Lebesgue sur $\mathbb{R}^d$)**.** *On peut construire une théorie de l'intégration*

$$f \mapsto \int_{\mathbb{R}^d} f d\ell$$

définie pour toute fonction f mesurable[6] *satisfaisant $f \geq 0$ (auquel cas l'intégrale peut être infinie) ou $\int_{\mathbb{R}^d} |f| d\ell < \infty$, telle que*

— *$f \mapsto \int_{\mathbb{R}^d} f d\ell$ est $\mathbb{R}$-linéaire ;*

6. On verra cette notion un peu plus tard, les fonctions continues ou Riemann-intégrables sont des cas particuliers.

— $\int_{\mathbb{R}^d} \mathbb{1}_A d\ell = \ell(A),\ \forall A \in \mathcal{B}(\mathbb{R}^d)$;
— $\int_{\mathbb{R}^d} f d\ell$ *coïncide avec l'intégrale de Riemann de* f *pour tout* f *Riemann-intégrable* ;
— $\int_{\mathbb{R}^d} \big(\sum_{i\geq 1} f_i\big) d\ell = \sum_{i\geq 1} \int_{\mathbb{R}^d} f_i d\ell$ *quand* $f_i \geq 0$ *pour tout* i *ou quand* $\int_{\mathbb{R}^d} |f_i| d\ell < \infty$ *pour tout* i.

1.4.2 Critère d'égalité entre mesures

Théorème 1.4.3. *Soit* $(\Omega, \mathcal{F})$ *un espace probabilisable. Soient* P, Q *deux mesures de probabilité sur* $(\Omega, \mathcal{F})$. *Soit* $\mathcal{A} \subset \mathcal{F}$ *tel que*

— $A, B \in \mathcal{A} \implies A \cap B \in \mathcal{A}$;
— $\sigma(\mathcal{A}) = \mathcal{F}$;
— *il existe une suite croissante d'ensembles* $D_1, D_2, \cdots \in \mathcal{A}$ *telle que* $D_i \nearrow \Omega$.

Alors, si $P(A) = Q(A)$ *pour tout* $A \in \mathcal{A}$, $P(A) = Q(A)$ *pour tout* $A \in \mathcal{F}$ *(i.e. :* $P = Q$*).*

1.4.3 Théorème d'extension de Carathéodory

Théorème 1.4.4 (Théorème d'extension de Carathéodory, version probabiliste)**.** *Soit* Ω *un ensemble. Soit* $\mathcal{A} \subset \mathcal{P}(\Omega)$ *une* algèbre d'ensembles *:*

— $\varnothing \in \mathcal{A}$;
— *si* $A \in \mathcal{A}$, *alors* $\Omega \setminus A \in \mathcal{A}$;
— *si* $A, B \in \mathcal{A}$, *alors* $A \cup B \in \mathcal{A}$.

Soit $\rho : \mathcal{A} \to [0, 1]$ *une* pré-mesure de probabilité *:*

— $\rho(\varnothing) = 0$, $\rho(\Omega) = 1$;
— *pour* $A_1, A_2, \cdots \in \mathcal{A}$ *deux à deux disjoints tels que* $\bigcup_{i\geq 1} A_i \in \mathcal{A}$, *on a*

$$\rho\Big(\bigcup_{i\geq 1} A_i\Big) = \sum_{i\geq 1} \rho(A_i).$$

Alors, il existe une unique mesure de probabilité P *sur* $\sigma(\mathcal{A})$ *telle que* $P(A) = \rho(A)$ *pour tout* $A \in \mathcal{A}$.

1.5 Quelques exemples d'espaces de probabilité

1.5.1 Mesure de Dirac

Un des exemple les plus triviaux de mesure de probabilité est aussi un des plus fondamentaux de par son apparition dans de nombreuses situations en physique.

Lemme 1.5.1 (Mesure de Dirac)**.** *Soit* $(\Omega, \mathcal{F})$ *un espace probabilisable. Soit* $\omega_0 \in \Omega$, *la fonction* $\delta_{\omega_0} : \mathcal{F} \to \mathbb{R}$ *donnée par*

$$\delta_{\omega_0}(A) = \begin{cases} 0 & \text{si } \omega_0 \notin A \\ 1 & \text{si } \omega_0 \in A \end{cases},$$

est une mesure de probabilité, appelée la mesure de Dirac en ω_0.

Démonstration. On a que $\omega_0 \in \Omega$, donc $\delta_{\omega_0}(\Omega) = 1$.

Soient $A_1, A_2, \cdots \in \mathcal{F}$ des ensembles deux à deux disjoints. Alors, soit ω_0 appartient à leur union, auquel cas il y a un unique $j \in \mathbb{N}$ tel que $\omega_0 \in A_j$ et on a

$$\delta_{\omega_0}(\bigsqcup_{i\geq 1} A_i) = 1, \quad \sum_{i\geq 1} \delta_{\omega_0}(A_i) = \delta_{\omega_0}(A_j) = 1,$$

soit ω_0 n'appartient à aucun des A_is, auquel cas

$$\delta_{\omega_0}(\bigsqcup_{i\geq 1} A_i) = 0 = \sum_{i\geq 1} 0 = \sum_{i\geq 1} \delta_{\omega_0}(A_i).$$

□

Cette mesure de probabilité (qui formalise simplement une expérience qui donne toujours le même résultat...) permet d'en construire bien d'autres en utilisant le Lemme 1.1.4.

1.5.2 Espaces de probabilité discrets classiques

Dans ce cas, la tribu sera toujours $\mathcal{P}(\Omega)$. L'additivité dénombrable implique que la mesure de probabilité P est entièrement définie par $\big(P(\{\omega\})\big)_{\omega\in\Omega}$ car

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

On notera $P(\omega) = P(\{\omega\})$ (ce qui est un léger abus de notation).

La loi de Bernoulli

Cette loi prend un paramètre $p \in [0,1]$. L'espace de probabilité associé est $(\{0,1\}, \mathcal{P}(\{0,1\}), P)$ où P est donnée par

$$P(1) = 1 - P(0) = p.$$

On notera cette loi $\mathrm{Bern}(p)$.

La loi Binomiale

Cette loi prend deux paramètres : $p \in [0,1]$ et $n \in \mathbb{Z}_+$. L'espace de probabilité associé est $(\{0,1,\ldots,n\}, \mathcal{P}(\{0,1,\ldots,n\}), P)$ où P est donnée par

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

On notera cette loi $\mathrm{Bin}(n,p)$.

La loi géométrique

Cette loi prend un paramètre : $p \in [0,1]$. L'espace de probabilité associé est $(\mathbb{N}, \mathcal{P}(\mathbb{N}), P)$ où P est donnée par

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

On notera cette loi $\mathrm{Geo}(p)$.

La loi de Poisson

Cette loi prend un paramètre $\lambda \in \mathbb{R}_+$. L'espace de probabilité associé est $(\mathbb{Z}_+, \mathcal{P}(\mathbb{Z}_+), P)$ où P est donnée par

$$P(k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

On notera cette loi Poi(λ).

La loi uniforme sur un ensemble fini

Cette loi prend comme paramètre un ensemble fini Ω. L'espace de probabilité associé est $(\Omega, \mathcal{P}(\Omega), P)$ où P est donnée par

$$P(\omega) = \frac{1}{|\Omega|},\ \forall \omega \in \Omega.$$

On notera cette loi Uni(Ω).

Exercice 1. Vérifier que les exemples ci-dessus sont bien des espaces de probabilité.

On reviendra sur ces exemples dans la section consacrée aux variables aléatoires.

1.5.3 Espaces de probabilité continus classiques

Mesures à densité sur $\mathbb{R}$

On a besoin de l'intégrale de Lebesgue pour tomber dans notre cadre général, n'y prêtez pas une attention trop grande : pour toutes les applications que nous verrons (et même pour la plupart des applications tout court...), comprendre les choses avec l'intégrale de Riemann et oublier qu'il y a des problèmes de définition pour certains ensembles pathologiques est amplement suffisant...

Définition 1.5.1. Soit $f : \mathbb{R} \to \mathbb{R}_+$ une fonction intégrable [7] telle que $\int_{-\infty}^{\infty} dx f(x) = 1$. La tribu considérée sera

$$\mathcal{F} = \mathcal{B}(\mathbb{R}).$$

On définit alors la *mesure de probabilité avec densité* f par

$$P(A) = \int_{-\infty}^{\infty} dx f(x) \mathbb{1}_A(x),$$

pour tout $A \in \mathcal{F}$ [8]. La fonction f est appelée la *densité par rapport à la mesure de Lebesgue* (ou juste *densité*) de P.

Exercice 2. Vérifier que l'espace $(\mathbb{R}, \mathcal{F}, P)$ ainsi définit est bien un espace de probabilité. Utiliser le Théorème 1.4.2.

Les mesures à densité incluent un grand nombre de mesures d'importance. En voici quelques exemples.

7. On supposera ici "Riemann-intégrable". Une notion d'intégrabilité plus générale sera introduite en théorie de la mesure.

8. Pour tomber dans le cadre introduit précédemment (i.e. : pour que l'intégrale fasse sens pour tout $A \in \mathcal{F}$), l'intégrale doit être comprise au sens de Lebesgue. Mais pour toutes les applications et pour tous les calculs que l'on verra, il suffit de comprendre cette intégrale au sens de Riemann.

Loi Uniforme

La mesure uniforme sur $[0,1]$: elle est obtenue avec le choix de densité $f(x) = \mathbb{1}_{[0,1]}(x)$. On la notera Uni($[0,1]$).

Plus généralement, la mesure uniforme sur un intervalle $[a,b]$ $(a < b)$ est obtenue avec le choix de densité $f(x) = (b-a)^{-1}\mathbb{1}_{[a,b]}(x)$. On la notera Uni($[a,b]$).

Loi Normale standard

La Loi Normale standard (ou loi gaussienne centrée réduite) est obtenue avec le choix de densité $f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. On notera cette loi $\mathcal{N}(0,1)$.

Loi Exponentielle

La Loi Exponentielle de paramètre $\lambda > 0$ est obtenue avec le choix de densité $f(x) = \lambda e^{-\lambda x}\mathbb{1}_{[0,\infty)}(x)$. On notera cette loi Exp(λ).

Loi de Cauchy

La Loi de Cauchy de paramètres $x_0 \in \mathbb{R}, \alpha > 0$ est obtenue avec le choix de densité $f(x) = \frac{\alpha}{\pi((x-x_0)^2+\alpha^2)}$. On notera cette loi Cauchy(x_0, α).

Exercice 3. Vérifier que les fonctions f données ci-dessus sont bien des densités de probabilité.

Comme dans le cas discret, on reviendra sur ces exemples dans la section consacrée aux variables aléatoires.

1.5.4 La marche aléatoire à temps fini

On peut regarder des exemples plus évolués : un modèle classique est la marche aléatoire simple symétrique sur $\mathbb{Z}$. Soit $n \geq 1$, on veut modéliser l'expérience suivante : on se met sur une échelle (que l'on prend infiniment longue) et, pendant n secondes, à chaque seconde on lance une pièce de monnaie et on monte d'un échelon si le résultat est pile, ou descend d'un échelon si le résultat est face. Les échelons sont modélisés par $\mathbb{Z}$. Un résultat de notre expérience est le chemin que l'on a parcouru durant nos n pas (on suppose que l'on part de l'échelon 0) :

$$\Omega_n = \{(s_0, s_1, \ldots, s_n) \in \{0\} \times \mathbb{Z}^n : \ |s_i - s_{i+1}| = 1\}.$$

Comme à chaque étape monter ou descendre est équiprobable, tous les chemins on la même probabilité :

$$P_n((s_0, s_1, \ldots, s_n)) = |\Omega_n|^{-1} = 2^{-n}.$$

La tribu associée est la tribu discrète : $\mathcal{P}(\Omega_n)$.

L'espace de probabilité associé est donc $(\Omega_n, \mathcal{P}(\Omega_n), P_n)$. Cet exemple est un bon cadre pour introduire la notion de *filtration* qui est une manière de formaliser le fait que l'information sur un système peut être obtenue petit à petit.

Notez qu'il y a une surjection naturelle de Ω_n dans Ω_k quand $k < n$:

$$\varphi_{n,k}(s_1, \ldots, s_n) = (s_0, \ldots, s_k).$$

Cette surjection donne lieu à une surjection de $\mathcal{P}(\Omega_n)$ dans $\mathcal{P}(\Omega_k)$:

$$\varphi_{n,k}(A) = \{\varphi_{n,k}(s_0, \ldots, s_n) : \ (s_0, \ldots, s_n) \in A\}.$$

On a vu dans le Lemme 1.6.2 que $\mathcal{F}_n^{(k)} = \varphi_{n,k}^{-1}\big(\mathcal{P}(\Omega_k)\big)$ est une tribu sur Ω_n. On a de plus que $\varphi_{n,k'} = \varphi_{k,k'} \circ \varphi_{n,k}$ pour $n > k > k'$. Ceci nous amène au fait que

$$\mathcal{P}(\Omega_n) = \mathcal{F}_n^{(n)} \supset \mathcal{F}_n^{(n-1)} \supset \cdots \supset \mathcal{F}_n^{(1)}.$$

Une telle suite monotone de tribu est appelée une *filtration*. Dans notre exemple, $\mathcal{F}_n^{(k)}$ continent tous les événements dans $\mathcal{P}(\Omega_n)$ qui ne dépendent que des k premiers pas de la marche.

1.6 Variables aléatoires

Les variables aléatoires sont des objets centraux dans la théorie des probabilité. Elles représentent toutes les mesures/observations possibles que l'on peut effectuer sur un système donné.

1.6.1 Fonctions mesurables et variables aléatoires

Définition 1.6.1 (Fonction mesurable)**.** Soit $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2)$ des espaces probabilisables. Une fonction $g : \Omega_1 \to \Omega_2$ est dite *mesurable* si pour tout $A \in \mathcal{F}_2$, $g^{-1}(A) \in \mathcal{F}_1$.

Remarque 1.6.1. *Il est important de noter que la fonction g ne dépend pas des tribus considérées, mais sa mesurabilité en dépend fortement : une même fonction peut être mesurable pour un choix de tribu et non-mesurable pour un autre choix. Si on voulait être vraiment prudent, on devrait noter et dire "g est $(\mathcal{F}_1, \mathcal{F}_2)$-mesurable"...*

Lemme 1.6.1 (Composition de fonctions mesurables)**.** *Soient $(\Omega_i, \mathcal{F}_i), i = 1, 2, 3$ des espaces probabilisables. Soient $g : \Omega_1 \to \Omega_2$, $h : \Omega_2 \to \Omega_3$ deux fonctions mesurables. Alors, $h \circ g$ est mesurable.*

Démonstration. On veut montrer que $(h \circ g)^{-1}(A) \in \mathcal{F}_3$ pour tout $\in \mathcal{F}_1$. Comme h est mesurable, $h^{-1}(A) \in \mathcal{F}_2$. Donc, comme g est mesurable, $(h \circ g)^{-1}(A) = g^{-1}(h^{-1}(A)) \in \mathcal{F}_1$. □

Lemme 1.6.2 (Tribu image-réciproque)**.** *Soit Ω_1 un ensemble et $(\Omega_2, \mathcal{F})$ un espace probabilisable. Soit $g : \Omega_1 \to \Omega_2$ une fonction. Alors,*

$$g^{-1}(\mathcal{F}) := \{g^{-1}(A) : \ A \in \mathcal{F}\}$$

est une tribu sur Ω_1, la tribu engendrée par g.

Démonstration. On a d'abord que $\varnothing \in g^{-1}(\mathcal{F})$ car $g^{-1}(\varnothing) = \varnothing$ vu que g est une fonction. Soit $A' \in g^{-1}(\mathcal{F})$. Il existe $A \in \mathcal{F}$ tel que $g^{-1}(A) = A'$. Comme g est une fonction, on a que $g^{-1}(\Omega_2 \setminus A) = \Omega_1 \setminus A'$. Mais $\Omega_2 \setminus A \in \mathcal{F}$ (car $\mathcal{F}$ est une tribu) et donc $\Omega_1 \setminus A \in g^{-1}(\mathcal{F})$. Finalement, si $A'_1, A'_2, \cdots \in g^{-1}(\mathcal{F})$, il existe $A_1, A_2, \cdots \in \mathcal{F}$ tels que

$g^{-1}(A_i) = A_i'$. On a alors que $\bigcup_{i\geq 1} A_i \in \mathcal{F}$ et donc que $g^{-1}\Big(\bigcup_{i\geq 1} A_i\Big) \in g^{-1}(\mathcal{F})$. Mais la pré-image d'une union est égale à l'union des pré-images, donc

$$\bigcup_{i\geq 1} A_i' = \bigcup_{i\geq 1} g^{-1}(A_i) = g^{-1}\Big(\bigcup_{i\geq 1} A_i\Big) \in g^{-1}(\mathcal{F}).$$

□

Lemme 1.6.3 (Tribu image)**.** *Soit $(\Omega_1, \mathcal{F})$ un espace probabilisable et soit Ω_2 un ensemble. Soit $g : \Omega_1 \to \Omega_2$ une fonction. Alors,*

$$g(\mathcal{F}) := \{A \subset \Omega_2 : \ g^{-1}(A) \in \mathcal{F}\}$$

est une tribu sur Ω_2, la tribu image par g.

Démonstration. Comme $g^{-1}(\varnothing) = \varnothing$, $\varnothing \in g(\mathcal{F}_1)$.

Ensuite, si $A \in g(\mathcal{F}_1)$, $g^{-1}(A) \in \mathcal{F}_2$. Donc, $\Omega_1 \setminus g^{-1}(A) \in \mathcal{F}_1$. Or, $g^{-1}(\Omega_2 \setminus A) = \Omega_1 \setminus g^{-1}(A)$ (la pré-image du complémentaire est le complémentaire de la pré-image). Donc, $\Omega_2 \setminus A \in g(\mathcal{F}_1)$.

Finalement, soient $A_1, A_2, \cdots \in g(\mathcal{F}_1)$. On a alors $\bigcup_{i\geq 1} g^{-1}(A_i) \in \mathcal{F}_1$. Mais comme l'union des pré-images est égale à la pré-image de l'union, $g^{-1}\Big(\bigcup_{i\geq 1} A_i\Big) \in \mathcal{F}_1$ et donc $\bigcup_{i\geq 1} A_i \in g(\mathcal{F}_1)$. □

Remarque 1.6.2.
1. *La tribu engendrée par g est la plus petite tribu sur Ω_1 qui rend g mesurable.*
2. *g est automatiquement mesurable si on muni Ω_2 de la tribu image par g.*

Lemme 1.6.4 (Critère de mesurabilité)**.** *Soient $(\Omega_i, \mathcal{F}_i), i = 1, 2$ des espaces probabilisables. Soit $\mathcal{A} \subset \mathcal{P}(\Omega)$ tel que $\sigma(\mathcal{A}) = \mathcal{F}_2$. Soit $g : \Omega_1 \to \Omega_2$. Alors g est mesurable si et seulement si $g^{-1}(A) \in \mathcal{F}_1$ pour tout $A \in \mathcal{A}$.*

Démonstration. Une implication est triviale, on se concentre donc sur l'autre. Supposons que $g^{-1}(A) \in \mathcal{F}_1$ pour tout $A \in \mathcal{A}$. Ceci est équivalent à $\mathcal{A} \subset \{A \in \mathcal{F}_2 : g^{-1}(A) \in \mathcal{F}_1\} = \mathcal{G}$. On a montré dans le Lemme 1.6.3 que $\mathcal{G}$ est une tribu sur Ω_2. En particulier, $\sigma(\mathcal{A}) \subset \mathcal{G}$ car $\sigma(\mathcal{A})$ est la plus petite tribu contenant $\mathcal{A}$. Mais, par hypothèse sur $\mathcal{A}$ et comme $\mathcal{G} \subset \mathcal{F}_2$, $\mathcal{F}_2 = \sigma(A) \subset \mathcal{G} \subset \mathcal{F}_2$, donc $\mathcal{F}_2 = \mathcal{G}$ et g est par conséquent mesurable. □

Définition 1.6.2 (Variable aléatoire)**.** Soit $(\Omega, \mathcal{F})$ un espace probabilisable. Une *variable aléatoire*[9] est une fonction $g : \Omega \to \mathbb{R}$ mesurable si on muni $\mathbb{R}$ de la tribu $\mathcal{B}(\mathbb{R})$. De manière équivalente, une fonction $g : \Omega \to \mathbb{R}$ est une variable aléatoire si $g^{-1}([a, b]) \in \mathcal{F}$ pour tout $a, b \in \mathbb{R}$ ou si $g^{-1}((a, b)) \in \mathcal{F}$ pour tout $a, b \in \mathbb{R}$[10].

Plus généralement, si $(\Omega', \mathcal{F}')$ est un espace mesurable (probabilisable), une *variable aléatoire à valeurs dans Ω'* est une fonction mesurable $g : \Omega \to \Omega'$[11].

9. En statistiques, une variable aléatoire est aussi appelée une *statistique*.

10. La raison pour cette équivalence est que $\mathcal{B}(\mathbb{R})$ est la tribu engendrée par les intervalles fermés (ou ouverts). On peut alors appliquer le Lemme 1.6.4

11. On ne considérera dans ce cours que des variables aléatoires réelles, mais il est utile de savoir que cette notion plus générale existe.

En mots : une variable aléatoire est une fonction g telle que les ensembles du type $\{\omega \in \Omega : \ g(\omega) \in [a,b]\}$ soient des événements (i.e. : on peut mesurer si $g(\omega)$ appartient à un intervalle).

Exemple 1.6.1. Quelques exemples de variables aléatoires liés à la marche aléatoire (voir section 1.5.4) :

— Le maximum de la marche :

$$X(s_0, s_1, \ldots, s_n) = \max\{s_0, \ldots, s_n\}.$$

— Le nombre de pas vers le haut :

$$X(s_0, s_1, \ldots, s_n) = \sum_{k=0}^{n-1} \frac{1}{2}(s_{k+1} - s_k + 1).$$

— La hauteur au temps $t \leq n$:

$$X(s_0, s_1, \ldots, s_n) = s_t.$$

Un choix "judicieux" de variable aléatoire permet de "réduire" les questions que l'on se pose sur $(\Omega, \mathcal{F}, P)$ (qui peut être extrêmement compliqué) à des questions sur $\mathbb{R}$ (que l'on comprend mieux).

Une manière de voir ceci est de regarder comment les variables aléatoires "transforment" l'espace de probabilité $(\Omega, \mathcal{F}, P)$ en l'espace $(\mathbb{R}, \mathcal{B}(\mathbb{R}), X_*P)$, où X_*P est "l'image de P par X" :

Définition 1.6.3 (Espace de probabilité lié à une variable aléatoire)**.** Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire. On définit X_*P (la *mesure image de P par X*) via

$$X_*P(A) = P(X^{-1}(A)).$$

L'espace de probabilité associé à $(\Omega, \mathcal{F}, P)$ par X est alors $(\mathbb{R}, \mathcal{B}(\mathbb{R}), X_*P)$.

Notez que pour que la définition ci-dessus fasse du sens, il nous faut

Lemme 1.6.5. *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité et $(\Omega', \mathcal{F}')$ un espace probabilisable. Soit $X : \Omega \to \Omega'$ une fonction mesurable. Alors, $X_*P : \mathcal{F}' \to [0,1]$ définie par*

$$X_*P(A) = P(X^{-1}(A)), \ A \in \mathcal{F}',$$

est une mesure de probabilité sur $(\Omega', \mathcal{F}')$.

Démonstration. On a que $X_*P(\varnothing) = P(X^{-1}(\varnothing)) = P(\varnothing) = 0$ car X est une fonction (donc tout élément de Ω_1 est envoyé sur un élément de Ω_2). De la même manière, $X_*P(\Omega_2) = P(X^{-1}(\Omega_2)) = P(\Omega_1) = 1$.

Reste à vérifier la σ-additivité. Soient $A_1, A_2, \cdots \in \mathcal{F}'$ deux à deux disjoints. On a

$$X^{-1}\Big(\bigsqcup_{i\geq 1} A_i\Big) = \bigsqcup_{i\geq 1} X^{-1}(A_i),$$

et donc

$$X_*P\Big(\bigsqcup_{i\geq 1} A_i\Big) = P\Big(X^{-1}\Big(\bigsqcup_{i\geq 1} A_i\Big)\Big) = P\Big(\bigsqcup_{i\geq 1} X^{-1}(A_i)\Big)$$
$$= \sum_{i\geq 1} P(X^{-1}(A_i)) = \sum_{i\geq 1} X_*P(A_i).$$

□

Remarque 1.6.3. *Notez que toute mesure de probabilité P sur l'espace probabilisable $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est la loi d'une variable aléatoire : la variable identité,* $\mathrm{Id} : \mathbb{R} \to \mathbb{R}$, $\mathrm{Id}(x) = x$ *sur l'espace $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$. On a alors que* $\mathrm{Id}_*P = P$.

Un point notation

Par abus de notation on notera souvent

$$P(X \in A) = P(\{\omega \in \Omega :\ X(\omega) \in A\}),$$

et similairement pour $P(a \leq X \leq b)$, $\{X \in A\} = \{\omega \in \Omega :\ X(\omega) \in A\}$, etc..

On termine cette section par vérifier que les opérations standards sur les fonctions préservent la mesurabilité.

Lemme 1.6.6 (Opérations sur les variables aléatoires)**.** *Soit $(\Omega, \mathcal{F})$ un espace probabilisable. Soient $X, Y, X_1, X_2, \ldots : \Omega \to \mathbb{R}$ des fonctions mesurables ($\mathcal{F}, \mathcal{B}(\mathbb{R})$-mesurables). Alors les fonctions suivantes sont mesurables*

1. $\sup_{n\geq 1} X_n$ *et* $\inf_{n\geq 1} X_n$ [12] *;*
2. $\limsup_{n\to\infty} X_n$ *et* $\liminf_{n\to\infty} X_n$*. En particulier, si la suite X_n converge ponctuellement vers X_∞, alors X_∞ est mesurable ;*
3. aX *pour* $a \in \mathbb{R}$ *;*
4. $X + Y$*,* XY*,* $\max(X, Y)$ *et* $\min(X, Y)$ *;*
5. $\mathbb{1}_A$ *pour* $A \in \mathcal{F}$*.*

Démonstration. On procède dans l'ordre. On se rappelle du Lemme 1.6.4 et que $\mathcal{B}(\mathbb{R})$ est engendré par $\{(-\infty, t] :\ t \in \mathbb{R}\}$ et par $\{(x, y) :\ x \leq y \in \mathbb{R}\}$.

1. On a que $X_n^{-1}\big((-\infty, t]\big) \in \mathcal{F}$ par hypothèse, de plus

$$\big(\sup_n X_n\big)^{-1}\big((-\infty, t]\big) = \bigcap_{n\geq 1} X_n^{-1}\big((-\infty, t]\big) \in \mathcal{F}$$

car les tribus sont stables par intersections dénombrables. L'inf suit de la même manière avec une union plutôt qu'une intersection.

12. Si X prend des valeurs dans $[-\infty, +\infty]$, on regardera la tribu $\{A \subset [-\infty, +\infty] :\ A \setminus \{-\infty, +\infty\} \in \mathcal{B}(\mathbb{R})\}$.

2. On note $X'_k = \sup_{n\geq k} X_n$ qui est une variable aléatoire (par le premier point). De plus, $X'_k \geq X'_{k+1}$, donc $(X'_k)^{-1}\big((-\infty,t]\big)$ est une suite croissante d'ensembles, donc

$$\Big(\limsup_{n\to\infty} X_n\Big)^{-1}\big((-\infty,t]\big) = \Big(\lim_{k\to\infty} X'_k\Big)^{-1}\big((-\infty,t]\big) = \bigcup_{k\geq 1}(X'_k)^{-1}\big((-\infty,t]\big) \in \mathcal{F}.$$

   La liminf se traite similairement avec des intersections. Si la limite existe, elle est égale à la limsup/liminf, ce qui donne sa mesurabilité.

3. Si $a > 0$, $(aX)^{-1}\big((x,y)\big) = X^{-1}\big((x/a, y/a)\big) \in \mathcal{F}$. $a = 0$ donne une fonction constante, la préimage d'un ensemble est donc soit $\varnothing$ soit Ω qui sont tous deux dans $\mathcal{F}$. Finalement, $a < 0$ donne $(aX)^{-1}\big((x,y)\big) = X^{-1}\big((-y/|a|, -x/|a|)\big) \in \mathcal{F}$.

4. On a vu dans le Lemme 1.6.1 que la composition de fonctions mesurables est mesurable. On regarde alors $Z : \mathbb{R}^2 \to \mathbb{R}$ une des fonctions suivantes :

$$Z(x,y) = x+y,\ Z(x,y) = xy,\ Z(x,y) = \max(x,y),\ Z(x,y) = \min(x,y).$$

   Toutes sont continues, donc $(\mathcal{B}(\mathbb{R}^2), \mathcal{B}(\mathbb{R}))$-mesurables. Pour obtenir le résultat voulu, il suffit donc de montrer que $\omega \mapsto (X(\omega), Y(\omega))$ est $(\mathcal{F}, \mathcal{B}(\mathbb{R}^2))$-mesurable. Mais on a aussi que $\mathcal{B}(\mathbb{R}^2)$ est engendré par $\big\{(a_1,b_1)\times(a_2,b_2) :\ a_1,a_2,b_1,b_2 \in \mathbb{R}\big\}$. Comme X, Y sont $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-mesurables, la pré-image de ces pavés sont bien des ensembles mesurables : $X^{-1}\big((a_1,b_1)\big) \cap Y^{-1}\big(a_2,b_2\big)$, ce qui permet de conclure.

5. $(\mathbb{1}_A)^{-1}(B)$ est soit A si $1 \in B, 0 \notin B$, soit A^c si $0 \in B, 1 \notin B$, soit Ω si $0,1 \in B$, soit $\varnothing$ si $0,1 \notin B$. Comme $A \in \mathcal{F}$, tous ces ensembles sont dans $\mathcal{F}$.

□

1.6.2 Espérance de variables aléatoires

Un première information que l'on peut vouloir sur une variable aléatoire est sa *valeur moyenne*.

Cas discret

Définition 1.6.4 (variable aléatoire discrète)**.** On dit qu'une variable aléatoire X (sur un espace $(\Omega, \mathcal{F})$ est *discrète* si $X(\Omega) = \{x \in \mathbb{R} :\ \exists \omega \in \Omega, X(\omega) = x\}$ est dénombrable (i.e. : elle ne prend qu'un nombre dénombrable de valeurs).

On trouvera plusieurs exemples de telles variables dans la section 1.8.2.

Définition 1.6.5 (Espérance, cas discret)**.** Soit $(\Omega, \mathcal{F})$ un espace probabilisable. Soit P une mesure de probabilité sur $(\Omega, \mathcal{F})$. Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire discrète.

On dira que X est *P-intégrable* si la série $\sum_{x\in X(\Omega)} P(X = x)|x|$ converge. Dans ce cas, on définit *l'espérance de X sous P* par

$$E_P(X) := \sum_{x\in X(\Omega)} P(X = x)x.$$

On notera $L^1_{P,\mathrm{dic}}(\Omega)$ l'ensemble des variables aléatoires discrètes P-intégrables de Ω dans $\mathbb{R}$.

Si on voulait être vraiment exact, on devrait plutôt noter $L^1_{P,\mathrm{dic}}(\Omega,\mathcal{F})$: par définition des variables aléatoires, l'espace image est $\mathbb{R}$ muni de la tribu $\mathcal{B}(\mathbb{R})$. On devrait garder trace de la tribu dont on muni Ω pour vérifier la mesurabilité d'une fonction donnée...

Exemple 1.6.2. 1. On peut regarder la variable aléatoire $X(\omega)=\omega$ sur l'espace associé à la loi uniforme sur $\{1,2,\dots,n\}$: $\Omega=\{1,\dots,n\}$, $\mathcal{F}=\mathcal{P}(\Omega)$, $P(\{k\})=\frac{1}{n}$ pour tout $k=1,\dots,n$. On a alors que $X(\Omega)=\{1,\dots,n\}\subset\mathbb{R}$ est fini. De plus

$$\{X=k\}=\{\omega\in\Omega:\ X(\omega)=k\}=\{k\}.$$

Donc,

$$E_P(X)=\sum_{x\in X(\Omega)}xP(X=x)=\sum_{x\in\{1,\dots,n\}}xP(\{x\})=\sum_{k=1}^{n}\frac{k}{n}=\frac{n+1}{2}.$$

2. Un cas plus évolué, toujours sur $\Omega=\{1,\dots,n\}$, $\mathcal{F}=\mathcal{P}(\Omega)$, $P(\{k\})=\frac{1}{n}$, est regarder la variable aléatoire

$$X(\omega)=\omega \mod 2,$$

(où $m \mod 2$ est le reste de la division euclidienne de m par 2). On a alors que $X(\Omega)=\{0,1\}\subset\mathbb{R}$ est fini. De plus, pour $k=0,1$,

$$\{X=k\}=\{\omega\in\Omega:\ X(\omega)=k\}=\{i\in\{1,\dots,n\}:\ i \mod 2=k\}.$$

On a que

$$|\{i\in\{1,\dots,n\}:\ i \mod 2=1\}|=\lceil\frac{n}{2}\rceil,$$
$$|\{i\in\{1,\dots,n\}:\ i \mod 2=0\}|=\lfloor\frac{n}{2}\rfloor.$$

En donc

$$P(X=1)=\frac{|\{i\in\{1,\dots,n\}:\ i \mod 2=1\}|}{n}=\frac{1}{n}\lceil\frac{n}{2}\rceil,$$
$$P(X=0)=\frac{|\{i\in\{1,\dots,n\}:\ i \mod 2=0\}|}{n}=\frac{1}{n}\lfloor\frac{n}{2}\rfloor.$$

L'espérence de X est alors donnée par

$$E_P(X)=0P(X=0)+1P(X=1)=P(X=1)=\frac{1}{n}\lceil\frac{n}{2}\rceil.$$

Remarque 1.6.4. *Quelques observations dans le cas discret :*

- *La condition d'absolue sommabilité (P-intégrabilité) garanti que l'espérance est bien définit (ne dépend pas de l'ordre de sommation).*
- *Les variable aléatoires bornées sont toujours intégrables.*
- *L'espérance d'une variable aléatoire constante 1 vaut 1 : $E_P(1)=1$.*
- *Les variables aléatoires à support fini sont toujours intégrables (car elles sont bornées).*

— *$E_P(X)$ ne dépend que de X_*P, on le voit en remarquant que :*

$$E_P(X) = \sum_{x\in X(\Omega)} P(X=x)x = \sum_{x\in X(\Omega)} xP(X^{-1}(x)) = \sum_{x\in X(\Omega)} xX_*P(x).$$

Une famille d'exemples de variables discrètes est construite comme suit :

— On se donne une partition de Ω : $A_1 \sqcup A_2 \sqcup \cdots = \Omega$ avec $A_1, A_2, \cdots \in \mathcal{F}$.

— On se donne $a_1, a_2, \cdots \in \mathbb{R}$ des nombre réels.

— On définit une variable aléatoire $X : \Omega \to \mathbb{R}$ via

$$X(\omega) = \sum_{i\geq 1} a_i \mathbb{1}_{A_i}(\omega).$$

L'espérance possède de nombreuses propriétés qui devraient rappeler celles de l'intégrale :

Théorème 1.6.7 (Propriétés de l'espérance : cas discret)**.** *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient $X, Y \in L^1_{P,\mathrm{dic}}(\Omega)$ deux variables aléatoires discrètes P-intégrables, et $a, b \in \mathbb{R}$. Alors,*

1. *linéarité de l'espérance : $E_P(aX+bY) = aE_P(X) + bE_P(Y)$;*
2. *si $P(X \geq Y) = 1$, $E_P(X) \geq E_P(Y)$. En particulier,*
   — *si $P(X \geq 0) = 1$, $E_P(X) \geq 0$;*
   — *si $P(a \leq X \leq b) = 1$, $a \leq E_P(X) \leq b$;*
   — *$|E_P(X)| \leq E_P(|X|)$.*

Démonstration. Montrons d'abord que $Z = aX + bY \in L^1_{P,\mathrm{dic}}(\Omega)$. On a que $Z(\Omega) = \{ax + by : \ x \in X(\Omega), y \in Y(\Omega)\} = \bigcup_{x\in X(\Omega)} \{ax + by : \ y \in Y(\Omega)\}$ est dénombrable (union dénombrable d'ensembles dénombrables). Ensuite,

$$\begin{aligned}
\sum_{z\in Z(\Omega)} P(Z=z)|z| &= \sum_{z\in Z(\Omega)} \sum_{x\in X(\Omega)} \sum_{y\in Y(\Omega)} P(Z=z, X=x, Y=y)|z| \\
&= \sum_{z\in Z(\Omega)} \sum_{x\in X(\Omega)} \sum_{y\in Y(\Omega)} \mathbb{1}_{ax+by=z} P(X=x, Y=y)|ax+by| \\
&\leq \sum_{x\in X(\Omega)} \sum_{y\in Y(\Omega)} P(X=x, Y=y)(|ax| + |by|) \\
&= |a| \sum_{x\in X(\Omega)} P(X=x)|x| + |b| \sum_{y\in Y(\Omega)} P(Y=y)|y| < \infty,
\end{aligned}$$

où on a utilisé l'inégalité triangulaire et $\sum_{y\in Y(\Omega)} P(X=x, Y=y) = P(X=x)$ ainsi que l'égalité similaire avec les rôles de X, Y échangés [13]. On utilise ensuite la linéarité

13. Cette égalité suit de
— les événements $\{Y=y\} = \{\omega \in \Omega : \ Y(\omega) = y\}$ pour $y \in Y(\Omega)$ sont deux à deux disjoints,
— $\bigcup_{y\in Y(\Omega)} \{Y=y\} = \{\omega \in \Omega : \ Y(\omega) \in Y(\Omega)\} = \Omega$.
Donc les événements $\{X=x, Y=y\}, y \in Y(\Omega)$ sont deux à deux disjoints et $\bigcup_{y\in Y(\Omega)} \{X=x, Y=y\} = \{X=x\}$.

de la somme pour obtenir celle de l'espérance : en procédant comme avant,

$$\sum_{z\in Z(\Omega)} P(Z=z)z = \sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega)} P(X=x,Y=y)(ax+by) =$$
$$= a\sum_{x\in X(\Omega)} P(X=x)x + b\sum_{y\in Y(\Omega)} P(Y=y)y.$$

Passons au second point. Si $P(X \geq Y) = 1$,

$$\begin{aligned} E_P(X) = \sum_{x\in X(\Omega)} P(X=x)x &= \sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega)} P(X=x,Y=y)x \\ &= \sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega):y\leq x} P(X=x,Y=y)x \\ &\geq \sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega):y\leq x} P(X=x,Y=y)y \\ &= \sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega)} P(X=x,Y=y)y \\ &= \sum_{y\in Y(\Omega)} P(Y=y)y = E_P(Y), \end{aligned}$$

où on a utilisé deux fois (deuxième et quatrième lignes) que $P(X = x, Y = y) = 0$ si $x < y$. Les deux premières conséquences du second point suivent en utilisant les variables aléatoires constantes valant 0, a et b. La troisième suit de $-|X| \leq X \leq |X|$. □

On a utilisé deux fois des conditions du type "si $P(X \in A) = 1$" plutôt que de demander la condition plus forte "$X(\omega) \in A$ pour tout $\omega \in \Omega$". C'est l'un des points importants quand on travaille avec des probabilité : on a pas besoin de contrôler que quelque chose arrive *dans tous les cas* mais seulement *dans un ensemble de cas qui arrive avec probabilité* 1. On utilise la dénomination suivante.

Définition 1.6.6. Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. On dit qu'une propriété est vraie *presque sûrement* si l'ensemble des résultats qui satisfont la propriété, B, contient un ensemble mesurable A de probabilité 1 :

$$B \subset \Omega \text{ est vérifié } P\text{-presque sûrement} \iff \exists A \in \mathcal{F} \text{ tel que } A \subset B,\ P(A) = 1.$$

Exemple 1.6.3. On regarde l'expérience "lancé de trois pièces distinctes". On a $\Omega = \{0,1\}^3$. On suppose que le résultat de l'expérience nous est communiqué par un ordinateur peu performant qui nous dit seulement si il y a plus de 1 ou plus de 0. On a donc $\mathcal{F} = \big\{\varnothing, \Omega, \{000, 001, 010, 100\}, \{111, 011, 110, 101\}\big\}$. On suppose maintenant que les pièces sont fortement corrélées : la première pièce est uniforme sur $\{0,1\}$, la seconde donne 1 si la première a donné 0 et est uniforme sinon, la troisième donne 1 si soit la première soit la seconde ont donné 0 et est uniforme sinon. On obtient alors que la probabilité de nos événements est

$$P(\varnothing) = 0,\ P(\Omega) = 1,\ P(\{000, 001, 010, 100\}) = 0,$$
$$P(\{111, 011, 110, 101\}) = 1.$$

On a alors que "au moins une des deux premières pièces donne 1" n'est pas un événement mesurable car $\{010, 011, 100, 101, 110, 111\} \notin \mathcal{F}$. En revanche, "au moins une des deux premières pièces donne 1" se produit presque sûrement car $\{111, 011, 110, 101\} \subset \{010, 011, 100, 101, 110, 111\}$.

Extension au cas général

Pour passer au cas général, on va approximer les variables générales par des variables discrètes. Pour une fonction $X : \Omega \to \mathbb{R}$ (Ω un ensemble), on définit les discrétisations :

$$X^{(n)}(\omega) = 2^{-n}\lfloor 2^n X(\omega)\rfloor = \sum_{k\in\mathbb{Z}} k2^{-n}\mathbb{1}_{[k2^{-n},(k+1)2^{-n})}(X(\omega)). \tag{1.2}$$

Si X est une variable aléatoire sur $(\Omega, \mathcal{F})$, on a bien que $X^{(n)}$ est une variable aléatoire discrète (aussi sur $(\Omega, \mathcal{F})$, prenant des valeurs dans $2^{-n}\mathbb{Z}$). L'intérêt des $X^{(n)}$ est

Lemme 1.6.8. *Soit Ω un ensemble et soit $X : \Omega \to \mathbb{R}$. Alors les fonctions $X^{(n)}$ définies par* (1.2) *satisfont*

$$X^{(n)}(\omega) \leq X(\omega) \leq X^{(n)}(\omega) + 2^{-n}. \tag{1.3}$$

En particulier, la suite de fonctions $(X^{(n)})_{n\geq 0}$ converge ponctuellement vers X.

Démonstration. Comme $\lfloor a\rfloor \leq a$, on a directement que $X^{(n)}(\omega) \leq X(\omega)$. De plus, $\lfloor a\rfloor \geq a - 1$, donc $X^{(n)}(\omega) \geq X(\omega) - 2^{-n}$. □

L'espérance de variables aléatoires générales est alors définie par

Théorème 1.6.9 (Espérance de variables aléatoires)**.** *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soit X une variable aléatoire. Soit $(X^{(n)})_{n\geq 0}$ la suite de variables aléatoires discrètes définies par* (1.2)*.*

Si $X^{(0)}$ est P-intégrable ($E_P(|X^{(0)}|) < \infty$), alors il existe $C > 0$ tel que $E_P(|X^{(n)}|) \leq C$ pour tout $n \geq 0$ (en particulier, $X^{(n)}$ est P-intégrable pour tout $n \geq 0$). On dit alors que X est P-intégrable et on définit son espérance par

$$E_P(X) := \lim_{n\to\infty} E_P(X^{(n)}).$$

Démonstration. (Non-éligible à l'examen) Supposons $E_P(|X^{(0)}|) < \infty$ et prenons $C = E_P(|X^{(0)}|) + 1$. Le Lemme 1.6.8 nous donne que $X^{(n)} \leq X \leq X^{(0)} + 1$ et $X^{(n)} + 1 \geq X \geq X^{(0)}$ et donc $X^{(0)} - 1 \leq X^{(n)} \leq X^{(0)} + 1$ ce qui implique $|X^{(n)}| \leq |X^{(0)}| + 1$. Par le Théorème 1.6.7, on a que $E_P(|X^{(n)}|) \leq E_P(|X^{(0)}|) + 1 = C$. En particulier, $X^{(n)}$ est P-intégrable pour $n \geq 0$. Il reste à montrer que la suite $a_n = E_P(X^{(n)})$ converge. On va montrer qu'elle est de Cauchy : pour $m \geq n$, on a par le Théorème 1.6.7 que

$$|a_n - a_m| = |E_P(X^{(n)} - X^{(m)})| \leq E_P(|X^{(n)} - X^{(m)}|).$$

Mais, par le Lemme 1.6.8, on a que $|X^{(n)} - X^{(m)}| \leq |X^{(n)} - X| + |X - X^{(m)}| \leq 2^{-n} + 2^{-m} \leq 2^{1-n}$. Ceci implique que $|a_n - a_m| \leq 2^{1-n} \xrightarrow{n\to\infty} 0$, qui donne que $(a_n)_{n\geq 0}$ est de Cauchy et donc converge. □

Comme dans le cas discret, on notera $L^1_P(\Omega)$ l'ensemble des variables aléatoires P-intégrables.

Un exercice recommandé est de vérifier que cette définition de l'intégrale correspond à celle donnée précédemment dans le cas des variables aléatoires discrètes.

Du Théorème 1.6.7 on déduit

Théorème 1.6.10 (Propriétés de l'espérance : cas général)**.** *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient $X, Y \in L^1_P(\Omega)$ deux variables aléatoires P-intégrables, et $a, b \in \mathbb{R}$. Alors,*

1. *linéarité de l'espérance : $E_P(aX + bY) = aE_P(X) + bE_P(Y)$;*
2. *si $P(X \geq Y) = 1$, $E_P(X) \geq E_P(Y)$. En particulier,*
   - *si $P(X \geq 0) = 1$, $E_P(X) \geq 0$;*
   - *si $P(a \leq X \leq b) = 1$, $a \leq E_P(X) \leq b$;*
   - *$|E_P(X)| \leq E_P(|X|)$.*

Démonstration. (Non-éligible à l'examen) La preuve suit de la définition de $E_P(X)$ et du Lemme 1.6.8. Les détails sont laissés en exercice. □

Remarque 1.6.5. *On a choisit une discrétisation particulière des variables aléatoires. Il est possible de montrer qu'en fait n'importe quelle discrétisation P-intégrable qui converge ponctuellement convient et donne la même définition de l'espérance. Vous verrez ceci en Analyse IV et en théorie de la mesure.*

Remarque 1.6.6. *Il suit de la preuve que dans le cas de variables aléatoires presque sûrement positives, soit $E_P(X)$ fait du sens, soit la suite $(E_P(X^{(n)}))$ tend vers $+\infty$. Il est naturel dans ce cas de définir $E_P(X) = +\infty$.*

Les derniers résultats que nous verrons sur l'espérance sont des outils techniques très puissants. Tous deux sont d'important résultats de théorie de la mesure, mais comme l'usage que nous en aurons dans ce cours est restreint, ils ne sont pas éligibles à l'examen. Il est néanmoins bon de se rappeler qu'ils existent car leur usage est omniprésent dés que l'on veut pouvoir échanger des limites et des intégrales.

Théorème 1.6.11 (Convergence monotone (non-éligible à l'examen))**.** *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient $X_1, X_2, \cdots \in L^1_P(\Omega)$ une famille de variables aléatoires P-intégrables et positives ($X_n : \Omega \to \mathbb{R}_+$). Supposons qu'il existe $X \in L^1_P(\Omega)$ telle que*

$$X_1(\omega) \leq X_2(\omega) \leq \cdots \leq X(\omega), \quad X_n(\omega) \xrightarrow{n\to\infty} X(\omega)$$

pour tout $\omega \in \Omega$. On a alors que

$$E_P(X_n) \xrightarrow{n\to\infty} E_P(X).$$

Démonstration. La monotonicité de l'espérance nous donne $E_P(X_n) \leq E_P(X)$ (et donc $\limsup_{n\to\infty} E_P(X_n) \leq E_P(X)$). Montrons l'autre borne. Soit $X^{(k)}, k \geq 0$ les variables définies par (1.2). Fixons $\epsilon > 0$. On introduit les ensembles

$$A_{n,k} = \{\omega \in \Omega : \ X_n(\omega) \geq (1-\epsilon)X^{(k)}\}.$$

Comme $X_n(\omega) \nearrow X$ et que $X^{(k)} \leq X$, on a que $A_{n,k} \nearrow \Omega$ (quand $n \to \infty$) pour tout $k \geq 0$. On a alors que

$$\begin{aligned} E_P(X_n) \geq E_P(X_n \mathbb{1}_{A_{n,k}}) &\geq (1-\epsilon) E_P(X^{(k)} \mathbb{1}_{A_{n,k}}) \\ &= (1-\epsilon) \sum_{m \in \mathbb{Z}} m 2^{-k} P\big(\{m \leq 2^k X < (m+1)\} \cap A_{n,k}\big). \end{aligned}$$

Mais par la convergence monotone des probabilité, $P\big(\{m \leq 2^k X < (m+1)\} \cap A_{n,k}\big) \nearrow P\big(\{m \leq 2^k X < (m+1)\}\big)$ quand $n \to \infty$. En particulier, en prenant la liminf,

$$\liminf_{n \to \infty} E_P(X_n) \geq (1-\epsilon) \sum_{m \in \mathbb{Z}} m 2^{-k} P\big(\{m \leq 2^k X < (m+1)\}\big) = (1-\epsilon) E_P(X^{(k)}).$$

k et ϵ étant arbitraires, on peut prendre $k \to \infty$ suivit de $\epsilon \to 0$ pour obtenir $\liminf_{n \to \infty} E_P(X_n) \geq E_P(X)$. □

Ce théorème nous dit que la convergence ponctuelle monotone de variables aléatoires positives donne la convergence des espérances.

Théorème 1.6.12 (Convergence dominée (non-éligible à l'examen))**.** *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient $X_1, X_2, \ldots$ une famille de variables aléatoires. Si*

— *la suite X_n converge ponctuellement vers une fonction X ;*

— *il existe une variable aléatoire P-intégrable Y telle que $|X_n| \leq Y$ pour tout $n \geq 1$;*

alors X est P-intégrable et

$$\lim_{n \to \infty} E_P(|X_n - X|) = 0.$$

En particulier,

$$\lim_{n \to \infty} E_P(X_n) = E_P\big((\lim_{n \to \infty} X_n)\big) = E_P(X).$$

Démonstration. Pour $n \geq 1$, on introduit les variables aléatoires positives

$$Z_n = |X - X_n|, \quad Z'_n = \sup_{k \geq n} Z_k, \quad \alpha_n = 2Y - Z'_n.$$

Comme on a requis que $Y \geq |X_n|$ pour tout $n \geq 1$, $\alpha_n \geq 0$ (par définition de Z_n et inégalité triangulaire). De plus, comme Z'_n est une suite décroissante, α_n est une suite croissante qui converge ponctuellement vers $2Y$ (X_n converge ponctuellement vers X et donc Z_n, Z'_n convergent ponctuellement vers 0). On peut donc appliquer le théorème de convergence monotone (Théorème 1.6.11) pour obtenir

$$\lim_{n \to \infty} E_P(\alpha_n) = E_P(2Y).$$

En particulier, $\lim_{n \to \infty} E_P(Z'_n) = 0$ et donc (comme $Z_n \leq Z'_n$) $\lim_{n \to \infty} E_P(Z_n) = 0$. De ce dernier point, on conclue que

$$\big|E_P(X_n) - E_P(X)\big| = \big|E_P(X_n - X)\big| \leq E_P\big(|X_n - X|\big) \xrightarrow{n \to \infty} 0.$$

□

Convention de notation pour le reste du cours

Pour l'instant, on a gardé explicite la loi P dans la notation de l'espérance. On a vu dans le cas discret que P n'est pas réellement important pour définir l'espérance : seule X_*P l'est (la loi de la variable aléatoire). Il est possible de remarquer la même chose dans le cas général, cela motive la notation $E(X)$ pour l'espérance de X : on suppose simplement que la variable X est définie sur un espace de probabilité $(\Omega, \mathcal{F}, P)$ de sorte à ce que X_*P soit la loi de X. On notera donc simplement $E(X)$ pour l'espérance dans la suite du cours.

1.6.3 Fonction de répartition

L'information sur (la loi de) une variable aléatoire X est encodée dans la probabilité de tomber dans un intervalle donné. Ces probabilités sont elles-même données pas les probabilités des événements $\{X \leq a\}, a \in \mathbb{R}$ [14]. Cet ensemble de nombres est ce que l'on appelle la *fonction de répartition* de la variable X.

Définition 1.6.7. Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soit X une variable aléatoire. On appelle *fonction de répartition* de X la fonction $F_X : \mathbb{R} \to [0,1]$ donnée par

$$F_X(t) = P(X \leq t)\big(= P(X^{-1}((-\infty, t]))\big).$$

Exemple 1.6.4. On peut regarder la cas de ce que l'on appellera une variable aléatoire de Bernoulli : on considère l'espace de probabilité $\Omega = \{0,1\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, $P(\{0\}) = 1 - P(\{1\}) = p$. On regarde alors la variable aléatoire $X(\omega) = \omega$. On a que $P(X = 1) = P(\{1\}) = p$ et que $P(X = 0) = P(\{0\}) = 1 - p$. Sa fonction de répartition est donnée par

$$F_X(t) = (1-p)\mathbb{1}_{[0,1)}(t) + \mathbb{1}_{[1,+\infty)}(t).$$

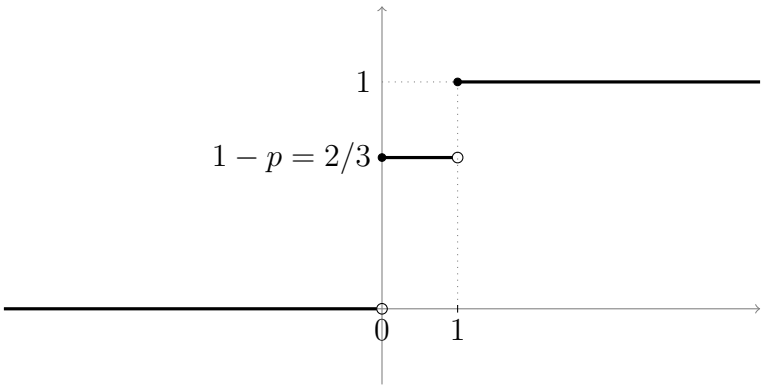

FIGURE 1.1 – La fonction de répartition d'une variable aléatoire de Bernoulli de paramètre $p = 1/3$.

14. On prend un petit raccourcis de notation :

$$\{X \leq a\} = \{\omega \in \Omega : \ X(\omega) \leq a\}.$$

Les propriétés élémentaires de cette fonction sont

Lemme 1.6.13. *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité et soit X une variable aléatoire. Alors,*

- *F_X est croissante ;*
- *F_X est continue à droite ;*
- *$\lim_{t\to-\infty} F_X(t) = 0$ et $\lim_{t\to+\infty} F_X(t) = 1$.*

Démonstration. F_X est croissante par monotonicité de P et $\{\omega : \ X(\omega) \leq t\} \subset \{\omega : X(\omega) \leq t'\}$ quand $t \leq t'$.

F_X est continue à droite car pour toute suite $t_i, i \geq 1$ telle que $t_i \searrow t$, $\{X \leq t_i\} \searrow \{X \leq t\}$, et donc

$$F_X(t_i) = P(X \leq t_i) \searrow P(X \leq t) = F_X(t).$$

Notez qu'il est important que $F_X(t)$ soit définit avant une inégalité non-stricte : sinon la continuité à droite devient fausse (et on obtient à la place la continuité à gauche).

$\lim_{t\to-\infty} F_X(t) = 0$ car pour toute suite $t_i, i \geq 1$ telle que $t_i \searrow -\infty$, on a $\{\omega : X(\omega) \leq t_i\} \searrow \varnothing$ et donc

$$P(X \leq t_i) \searrow P(\varnothing) = 0.$$

On procède de la même manière pour montrer $\lim_{t\to+\infty} F_X(t) = 1$. □

L'intérêt des fonctions de répartition est dans le Lemme suivant.

Lemme 1.6.14. *Soit $(\Omega_i, \mathcal{F}_i, P_i), i = 1, 2$ deux espaces de probabilité et soit $X_i : \Omega_i \to \mathbb{R}, i = 1, 2$ deux variables aléatoires. Alors, si $F_{X_1} = F_{X_2}$, on a que X_1 et X_2 ont la même loi : $(X_1)_* P_1 = (X_2)_* P_2$. En d'autres termes, pour tout $A \in \mathcal{B}(\mathbb{R})$,*

$$P_1(X_1 \in A) = P_2(X_2 \in A).$$

Démonstration. (Non-éligible à l'examen) Supposons $F_{X_1} = F_{X_2}$. Notons $P = (X_1)_* P_1$ et $Q = (X_2)_* P_2$. On introduit

$$\mathcal{I} = \{(-\infty, a] : \ a \in \mathbb{R}, a\},$$

l'ensemble des intervalles infini à gauche, fermés à droite (on utilise la convention $(a, a] = \varnothing$). On a déjà vu (en exercice) que $\sigma(\mathcal{I}) = \mathcal{B}(\mathbb{R})$. Comme $F_{X_1} = F_{X_2}$, on a que pour tout a,

$$P((-\infty, a]) = F_{X_1}(a) = F_{X_2}(a) = Q((-\infty, a]).$$

De plus, si $A, B \in \mathcal{I}$, on a aussi que $A \cap B \in \mathcal{I}$. Finalement, $(-\infty, n] \xrightarrow{n\to\infty} \mathbb{R}$. On a donc que $\mathcal{I}$ vérifie toutes les hypothèse du Théorème 1.4.3, on peut donc l'appliquer pour déduire que $P = Q$. □

L'avantage de ce Lemme est que l'on peut prendre l'égalité entre fonctions de répartition comme *la définition* de l'égalité en loi pour des variables aléatoires, ce que l'on fera dans le reste du cours :

Définition 1.6.8 (Égalité en loi)**.** Soit $(\Omega_i, \mathcal{F}_i, P_i), i = 1, 2$ deux espaces de probabilité. Soient $X_i : \Omega_i \to \mathbb{R}, i = 1, 2$ deux variables aléatoires. On dit que X_1 et X_2 sont *égales en loi*, noté $X_1 \overset{\text{Loi}}{=} X_2$, si $F_{X_1} = F_{X_2}$.

1.6.4 Variables aléatoires à densité

Une classe de variables aléatoires avec laquelle il est particulièrement agréable de travailler du point de vue analytique est les variables aléatoires à densité. On trouvera plusieurs exemples de telles variables dans la section 1.8.2.

Définition 1.6.9 (Variables aléatoires à densité)**.** Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité et $X : \Omega \to \mathbb{R}$ une variable aléatoire. On dit que X est une *variable aléatoire à densité* si il existe une fonction $f_X : \mathbb{R} \to \mathbb{R}_+$ (la *densité* de X) telle que pour tout $a < b$ [15],

$$P(a \leq X \leq b) = P(a < X < b) = \int_a^b f_X(x)dx.$$

Attention, $f_X(x)$ n'est pas la probabilité que $X = x$, f_X peut d'ailleurs prendre des valeurs plus grandes que 1. On peut en revanche parfois interpréter $f_X(x)dx$ comme "la probabilité que $X = x$" : par exemple, si f_X est continue, on a

$$P(x < X < x + \epsilon) = \int_x^{x+\epsilon} f_X(x)dx = f_X(x)\epsilon + o(\epsilon).$$

Dans le cas des variables aléatoires à densité, l'espérance revient à l'intégrale habituelle :

$$E(X) = \int_{-\infty}^{\infty} f_X(x)xdx.$$

Un bon exercice est de le démontrer dans le cas où f_X est Riemann intégrable (ou même simplement continue et à support compact). Pour faire ceci, il est intéressant de noter que les variables aléatoires dont la densité est constante par morceaux sont des cas particuliers de variables aléatoires discrètes. On peut aussi remarquer que la fonction de répartition est donnée par

$$F_X(t) = \int_{-\infty}^{t} f_X(x)dx.$$

Remarque 1.6.7. *On remarquera que dans le cas des espaces de probabilité continus vu dans la section 1.5, la fonction identité,* $\mathrm{Id} : \mathbb{R} \to \mathbb{R}$, $\mathrm{Id}(x) = x$, *est une variable aléatoire à densité. Une densité pour* Id *est donnée par celle utilisée pour définir la loi.*

1.6.5 Théorème de transfert pour l'espérance

Le Théorème de transfert (parfois appelée *loi du statisticien inconscient*) est, du point de vue de la mesure, une observation : on peut le résumer par

$$E_P(g(X)) = E_{X_*P}(g).$$

Néanmoins, cette forme n'est pas immédiatement pratique pour faire des calculs. On utilisera le cas particulier suivant :

Lemme 1.6.15 (Théorème de transfert)**.** *Soit* $(\Omega, \mathcal{F}, P)$ *un espace de probabilité et* $X : \Omega \to \mathbb{R}$ *une variable aléatoire. Alors,*

15. L'intégrale est comprise au sens de Lebesgue si f n'est pas Riemann-intégrable, mais les exemples que nous verrons auront tous f_X Riemann-intégrable.

1. *si X est une variable aléatoire discrète, pour toute $g : \mathbb{R} \to \mathbb{R}$, $g \circ X$ est une variable aléatoire et*

$$E_P(g(X)) = \sum_{x \in X(\Omega)} g(x) P(X = x)$$

dès que la somme converge absolument ;

2. *si X est une variable aléatoire à densité, pour toute $g : \mathbb{R} \to \mathbb{R}$ mesurable*

$$E_P(g(X)) = \int_{-\infty}^{+\infty} f_X(x) g(x) dx$$

dès que l'intégrale converge absolument.

Démonstration. (Non examinable)
Pour le cas discret, comme $X(\Omega)$ est au plus dénombrable, $g(X(\Omega))$ est au plus dénombrable donc $g \circ X$ est automatiquement mesurable et est donc une variable aléatoire. Par définition de l'espérance on a alors

$$\begin{aligned} E_P(g(X)) &= \sum_{y \in g(X(\Omega))} y P(g \circ X = y) \\ &= \sum_{y \in g(X(\Omega))} \sum_{x \in X(\Omega)} y P\big(g \circ X = y, X = x\big) \\ &= \sum_{x \in X(\Omega)} \sum_{y \in g(X(\Omega)) \colon g(x) = y} y P\big(g \circ X = y, X = x\big) \\ &= \sum_{x \in X(\Omega)} g(x) P\big(X = x\big) \end{aligned}$$

ou on a utilisé la σ-additivité dans la seconde ligne, la sommabilité absolue pour échanger l'ordre des sommes et le fait que $P\big(g \circ X = y, X = x\big) = 0$ si $g(x) \neq y$ et $P\big(g \circ X = y, X = x\big) = P\big(X = x\big)$ si $g(x) = y$.

Pour le cas à densité, on a que comme g est mesurable, $g \circ X$ l'est aussi et donc $g \circ X$ est une variable aléatoire. Si g est une fonction discrète ($g(\mathbb{R})$ est fini ou dénombrable), on a alors par définition

$$\begin{aligned} E_P(g \circ X) &= \sum_{y \in g(\mathbb{R})} y P(g \circ X = y) \\ &= \sum_{y \in g(\mathbb{R})} y P(X \in g^{-1}(y)) \\ &= \sum_{y \in g(\mathbb{R})} y \int_{-\infty}^{+\infty} \mathbb{1}_{g^{-1}(y)}(x) f_X(x) dx \\ &= \int_{-\infty}^{+\infty} g(x) f_X(x) dx \end{aligned}$$

où on a utilisé l'intégrabilité absolue pour échanger la somme et l'intégrale. Le cas général suit en approximant g mesurable par g discrète comme dans la construction de l'espérance (on peut utiliser la même discrétisation). □

1.6.6 Vecteurs aléatoires, variables aléatoires complexes

Vecteurs aléatoires

Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Un *vecteur aléatoire* de dimension $d \geq 1$ est une fonction

$$X : \Omega \to \mathbb{R}^d, \quad X = (X_1, \ldots, X_d),$$

tel que $X_i : \Omega \to \mathbb{R}$ est une variable aléatoire pour tout $1 \leq i \leq d$. La *fonction de répartition* du vecteur aléatoire X est donnée par

$$F_X(t_1, \ldots, t_d) = P\Big(\bigcap_{i=1}^{d} \{X_i \leq t_i\}\Big).$$

Exemple 1.6.5. Un exemple de vecteur aléatoire est donné par les hauteurs successives de la marche aléatoire (voir section 1.5.4) : on a que la hauteur au temps k est une variables aléatoire :

$$S_k(s_0, \ldots, s_n) = s_k.$$

On peut "ranger" toutes ces variables dans un vecteur aléatoire :

$$\bar{S} = \big(S_0, S_1, \ldots, S_n\big) : \Omega_n \to \mathbb{R}^{n+1}.$$

Variables aléatoires complexes

Dans le même esprit, une *variable aléatoire complexe* Z est une fonction

$$Z : \Omega \to \mathbb{C}, \quad Z = X + \mathrm{i}Y,$$

telle que $X, Y : \Omega \to \mathbb{R}$ sont des variables aléatoires.

Vecteurs à densité

De la même manière que l'on a définit les variables aléatoires à densité, on dit qu'un vecteur aléatoire $X = (X_1, \ldots, X_d)$ est un *vecteur aléatoire à densité* si il existe $f_X : \mathbb{R}^d \to \mathbb{R}_+$ mesurable telle que

$$\begin{aligned} P\big(X \in [a_1, b_1] \times \cdots \times [a_d, b_d]\big) &= P\big(X \in (a_1, b_1) \times \cdots \times (a_d, b_d)\big) \\ &= \int_{a_1}^{b_1} dx_1 \cdots \int_{a_d}^{b_d} dx_d f_X(x_1, \ldots, x_d), \end{aligned}$$

pour tous $a_1, b_1, \ldots, a_d, b_d \in \mathbb{R}$ avec $a_i \leq b_i$. En particulier, $\int_{\mathbb{R}^d} dx f_X(x) = 1$.

1.6.7 Processus stochastiques

Un processus stochastique généralise la notion de vecteur aléatoire à un nombre infini de variables aléatoires. L'intérêt de cette notion est que *toutes les variables aléatoires vivent sur le même espace de probabilité.*

Définition 1.6.10 (Processus stochastique)**.** Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soit T un ensemble. Un *processus stochastique indexé par* T est une famille $(X_t)_{t \in T}$ telle que X_t est une variable aléatoire pour tout $t \in T$.

On peut par exemple imaginer que $T = \mathbb{R}_+$ est le temps et que X_t est la température au centre-ville de Mumbaï à l'instant t.

1.6.8 Lois marginales

Cas général

Définition 1.6.11 (Loi marginale)**.** Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soit $X : \Omega \to \mathbb{R}^d$ un vecteur aléatoire. Soit $i \in \{1, \dots, d\}$. La *loi marginale* de X pour l'indice i est la loi de $\pi_i(X)$, $(\pi_i \circ X)_* P$, où π_i est la projection sur la $i$ème coordonnée :

$$\pi_i : \mathbb{R}^d \to \mathbb{R}, \quad \pi_i(x_1, \dots, x_d) = x_i.$$

Remarque 1.6.8. *On peut similairement définir la "loi marginale" de $(X_i, i \in I)$ pour $I \subset \{1, \dots, d\}$.*

Cas des vecteurs à densité

Dans le cas des vecteurs à densité, la loi marginale est donnée en intégrant les autres coordonnées : la loi marginale de X_i, $P_i = (\pi_i \circ X)_* P$, est donnée par

$$P_i(X_i \in A) = \int_{\mathbb{R}} dx_1 \cdots \int_A dx_i \cdots \int_{\mathbb{R}^d} dx_d f_X(x_1, \dots, x_d).$$

X_i est alors une variable aléatoire à densité, et sa densité est donnée par

$$f_{X_i}(x_i) = \int_{\mathbb{R}} dx_1 \cdots \int_{\mathbb{R}} dx_{i-1} \int_{\mathbb{R}} dx_{i+1} \cdots \int_{\mathbb{R}^d} dx_d f_X(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_d).$$

1.6.9 Changement de variables

L'espace de probabilité lié à une variable aléatoire ou à un vecteur aléatoire est en particulier... un espace de probabilité ! On peut donc regarder des variables/vecteurs aléatoires sur cet espace, ce qui revient à regarder une composition de variables aléatoires. On peut alors utiliser les règles de changements de variables usuelles pour calculer les densités des forces en présence.

On rappelle quelques définitions. Soit $d \geq 1$, $U \subset \mathbb{R}^d$ un ouvert et $\varphi : U \to \mathbb{R}^d$, $\varphi(y) = (\varphi_1(y), \dots, \varphi_d(y))$, $y = (y_1, \dots, y_d)$. Alors,

- on dit que φ est *continûment différentiable* sur U si les dérivées partielles $\frac{\partial \varphi_i}{\partial y_j}$ existent et sont continues sur U ;
- on définit $D_\varphi(y)$ la *matrice jacobienne* de φ en y :

$$D_\varphi(y) = \begin{pmatrix} \frac{\partial \varphi_1}{\partial y_1}(y) & \dots & \frac{\partial \varphi_1}{\partial y_d}(y) \\ \vdots & \dots & \vdots \\ \frac{\partial \varphi_d}{\partial y_1}(y) & \dots & \frac{\partial \varphi_d}{\partial y_d}(y) \end{pmatrix};$$

- on note det le déterminant.

On a alors la formule de changement de variables :

Lemme 1.6.16. *Soient $d \geq 1$, $U \subset \mathbb{R}^d$ un ouvert, $V \subset \mathbb{R}^d$ et $\varphi : U \to V$ une application bijective et continûment différentiable telle que $\det D_\varphi(y) \neq 0$ pour tout $y \in U$. Alors, pour toute fonction intégrable*[16] *$f : V \to \mathbb{R}$, on a*

$$\int_U dy_1 \dots dy_d f\big(\varphi(y)\big) |\det D_\varphi(y)| = \int_V dx_1 \dots dx_d f(x).$$

Démonstration. Le cas f suffisamment régulière a normalement été fait en Analyse I ou II. Le cas énoncé ici (f intégrable) sera fait en Analyse IV ou en théorie de la mesure. □

De ce Lemme on déduit

Corollaire 1.6.17. *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité et $X : \Omega \to \mathbb{R}^d$ un vecteur aléatoire avec densité f_X à valeurs dans un ouvert U. Soit $\varphi : U \to \mathbb{R}^d$ satisfaisant les hypothèses du Lemme 1.6.16. On a alors que $Y = \varphi \circ X : \Omega \to \mathbb{R}^d$ est un vecteur aléatoire avec densité donnée par*

$$f_Y(y) = f_X\big(\varphi^{-1}(y)\big) |\det D_{\varphi^{-1}}(y)| = \frac{1}{|\det D_\varphi(y)|} f_X\big(\varphi^{-1}(y)\big).$$

1.7 Indépendance et lois conditionnelles

1.7.1 Loi conditionnelle

Intuitivement, les probabilités conditionnelles formalise le fait que si on regarde une expérience (dont le résultat est aléatoire : par exemple, lancer trois pièces de monnaie numérotées 1,2,3), et que d'une manière ou d'une autre on obtient une information à propos du résultat de l'expérience (par exemple : en regardant le résultat de la pièce numéro 1), on se retrouve dans une situation dans laquelle la loi de probabilité de départ n'est plus la bonne (on a par exemple une probabilité 0 de faire trois piles si on a vu que la pièce no 1 donnait face alors que cette même probabilité était de 1/8 sans l'information supplémentaire). Mais il reste de l'aléa, car notre information n'est que partielle, la loi conditionnelle est exactement la loi de probabilité qui modélise cette situation.

Définition 1.7.1 (Loi conditionnelle)**.** Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé. Soit $A \in \mathcal{F}$ un événement tel que $P(A) > 0$. On définit[17] alors la *mesure de probabilité P conditionnellement à A*, notée $P(\ \mid A)$, par

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}, \quad \forall B \in \mathcal{F}.$$

On définit alors *l'espérance conditionnelle* d'une variable aléatoire X par

$$E_P(X \mid A) = E_{P_A}(X),$$

où on a noté $P_A = P(\ \mid A)$.

16. I.e. : f mesurable (pour $\mathbb{R}^d$ et $\mathbb{R}$ muni des tribus boréliennes $\mathcal{B}(\mathbb{R}^d)$ et $\mathcal{B}(\mathbb{R})$) et $\int_{\mathbb{R}^d} |f(x)| d\ell(x) < \infty$.

17. Il est facile de vérifier que la fonction donnée est bien définie et remplit les conditions pour être une mesure de probabilité : $P(\Omega \mid A) = P(A)/P(A) = 1$, et $P(\sqcup_{i\geq 1} A_i \mid A) = \frac{1}{P(A)} P(\sqcup_{i\geq 1}(A_i \cap A)) = \sum_{i\geq 1} P(A_i \mid A)$.

Pour obtenir un peu d'intuition sur la notion de probabilité conditionnelle, on peut revenir au modèle de Laplace (le tirage uniforme d'un élément dans un ensemble fini). Dans ce cas, on a

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

En d'autre termes : la loi uniforme conditionnellement à A devient juste la loi uniforme sur A.

On peut aussi voir la quantité $P(A \mid B)$ dans l'interprétation fréquentielle des probabilités : $P(A)$ donne la proportion de fois où A est réalisé parmi *toutes les réalisations* de notre expérience, alors que $P(A \mid B)$ donne la proportion de fois où A est réalisé parmi *les réalisations de notre expérience où B est vérifiée.*

Lemme 1.7.1 (Formule de Bayes)**.** *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé. Soient $A, B \in \mathcal{F}$ deux événements tels que $P(A), P(B) > 0$. Alors,*

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

Démonstration. Par définition, on a

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

□

Lemme 1.7.2 (Formule des probabilités totales)**.** *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé. Soit I un ensemble fini ou dénombrable. Soit $(A_i)_{i \in I} \in \mathcal{F}^I$ une collection d'événements tels que*

- *si $i \neq j$, alors $A_i \cap A_j = \varnothing$;*
- *$\bigcup_{i \in I} A_i = \Omega$* [18].

Supposons de plus que $P(A_i) > 0$ pour tout $i \in I$. Alors, pour tout $B \in \mathcal{F}$,

$$P(B) = \sum_{i \in I} P(B \cap A_i) = \sum_{i \in I} P(B \mid A_i)P(A_i).$$

Démonstration. Il suffit de remarquer que $B = B \cap \bigsqcup_{i \in I} A_i = \bigsqcup_{i \in I}(A_i \cap B)$, et que les ensembles $A_i \cap B, i \in I$ sont deux à deux disjoints, puis d'utiliser l'additivité dénombrable de P. □

Remarque 1.7.1. *Il existe une autre façon de voir cette loi en utilisant l'espérance. Si on a que $A_1, A_2, \ldots$ forment une partition de Ω (deux à deux disjoints et $\bigcup_i A_i = \Omega$), on a que pour toute variable aléatoire X P-intégrable*

$$E(X) = E(X \cdot 1) = E\Big(\sum_{i \geq 1} \mathbb{1}_{A_i} X\Big) = \sum_{i \geq 1} E(\mathbb{1}_{A_i} X)$$

18. Une telle collection est parfois appelée système exhaustif ou système complet. On peut relaxer la condition $\bigcup_{i \in I} A_i = \Omega$ en $P\Big(\bigcup_{i \in I} A_i\Big) = 1$

car pour tout $\omega \in \Omega$, $\sum_{i\geq 1} \mathbb{1}_{A_i}(\omega) = 1$ *(on peut utiliser le Théorème 1.6.12 pour justifier l'échange somme-intégrale sans faire de calculs). Si on a maintenant que les* A_i *ont tous une probabilité strictement positive,*

$$\sum_{i\geq 1} E(\mathbb{1}_{A_i} X) = \sum_{i\geq 1} E(X \mid A_i) P(A_i)$$

par définition de l'espérance conditionnelle. On peut alors utiliser la variable $X = \mathbb{1}_B$ *pour retrouver le Lemme précédent.*

Exemple 1.7.1. La formule de Bayes est quasiment triviale d'un point de vue "mathématique". Mais ce qu'elle nous dit d'un point de vue pratique est beaucoup plus profond. On va essayer d'avoir une idée de l'utilité de cette formule dans un cas pratique. Un exemple classique est de vouloir tester l'efficacité d'un test médical et de déterminer ce que l'on peut réellement inférer du résultat du test.

On regarde la situation suivante : on a une population dans laquelle circule une maladie. On estime qu'une proportion $\rho \in (0, 1)$ de la population est infectée par la maladie. Il existe un test très simple et peu cher pour vérifier si un individu est infecté ou non, mais ce test peut parfois rater. Le test peut rater de deux manières :

— faux positif : le test dit "infecté" (notons ceci $T = 1$) alors que la personne était saine (que nous noterons $I = 0$). On estime que 0.9% des personnes saines sont testées positives ;

— faux négatif : le test dit "sain" (notons ceci $T = 0$) alors que la personne était infectée (que nous noterons $I = 1$). On estime 0.1% des personnes infectées sont testées négatives.

On se demande maintenant ce que l'on peut inférer du résultat si un individu va passer le test.

On modélise la situation comme suit : on a aucune information a priori sur l'individu, on peut donc supposer qu'il est tiré uniformément dans la population. Ceci revient à dire qu'il a une probabilité ρ d'être infecté. Un résultat complet de notre expérience et la donnée de si l'individu est infecté ou non et de si le test est positif ou non. On introduit donc l'espace de résultats

$$\Omega = \{0, 1\} \times \{0, 1\},$$

et $(a, b) \in \Omega$ correspond à la situation $I = a, T = b$. La tribu peut être prise comme étant $\mathcal{P}(\Omega)$, et une mesure probabilité est donnée par $P(a, b) = p_{a,b} \in [0, 1]$ pour $(a, b) \in \Omega$. On introduit les événements "individu sain", "individu infecté", "test positif" et "test négatif"

$$\begin{aligned} \mathrm{IS} &= \{(0, 1), (0, 0)\}, \quad \mathrm{II} = \{(1, 1), (1, 0)\}, \\ \mathrm{TP} &= \{(1, 1), (0, 1)\}, \quad \mathrm{TN} = \{(1, 0), (0, 0)\}. \end{aligned}$$

On connaît :

— $p_{11} + p_{10} + p_{01} + p_{00} = 1$;

— $P(\mathrm{II}) = p_{11} + p_{10} = \rho = 1 - P(\mathrm{IS})$;

— $P(\mathrm{TP} \mid \mathrm{IS}) = 0.9\%$ et $P(\mathrm{TN} \mid \mathrm{II}) = 0.1\%$.

On peut d'abord se demander si notre test est "fiable" : avec quelle probabilité donne-t-il le mauvais résultat (i.e. : individu sain et test positif ou individu infecté et test négatif) ?

Par la formule de probabilité totale,

$$P\big((\mathrm{IS}\cap\mathrm{TP})\cup(\mathrm{II}\cap\mathrm{TN})\big) = P(\mathrm{TP}\,|\,\mathrm{IS})P(\mathrm{IS}) + P(\mathrm{TN}\,|\,\mathrm{II})P(\mathrm{II}) = \frac{0.9}{100}(1-\rho) + \frac{0.1}{100}\rho \leq 0.9\%.$$

Notre test est donc fiable à 99.1% *indépendamment de la proportion de gens infectés*. On peut donc considérer le test comme très fiable.

On se pose maintenant une question "classique" : si l'individu est testé positif, quelle est alors la probabilité qu'il ait effectivement contracté la maladie ? Cette question revient à calculer $P(\mathrm{II}\,|\,\mathrm{TP})$. On peut alors appliquer la formule de Bayes et la formule de probabilité totale pour obtenir

$$P(\mathrm{II}\,|\,\mathrm{TP}) = \frac{P(\mathrm{TP}\,|\,\mathrm{II})P(\mathrm{II})}{P(\mathrm{TP})} = \frac{P(\mathrm{TP}\,|\,\mathrm{II})P(\mathrm{II})}{P(\mathrm{TP}\,|\,\mathrm{II})P(\mathrm{II}) + P(\mathrm{TP}\,|\,\mathrm{IS})P(\mathrm{IS})}.$$

On utilise alors ce que l'on connaît pour avoir

$$P(\mathrm{TP}\,|\,\mathrm{IS}) = 0.9\%,\ P(\mathrm{II}) = \rho,\ P(\mathrm{IS}) = 1-\rho,$$
$$P(\mathrm{TP}\,|\,\mathrm{II}) = 1 - P(\mathrm{TN}\,|\,\mathrm{II}) = 1 - 0.1\% = 99.9\%.$$

On obtient alors

$$P(\mathrm{II}\,|\,\mathrm{TP}) = \frac{99.9\rho}{99\rho + 0.9}.$$

On peut alors voir que si une grande fraction de la population est infectée (ρ proche de 1), cette probabilité est proche de 1, et donc le résultat d'un test positif reflète souvent la réalité. À l'inverse, si ρ est très petit (maladie rare), le résultat d'un test positif est très souvent faux ! Il est donc important quand on veut interpréter le résultat du test de regarder aussi dans quelle situation on se trouve.

Le fait qu'un test très fiable semble "rater souvent" vient d'un biais cognitif : la *négligence de la taille de l'échantillon* (ou *oubli de la fréquence de base*) : on travaille dans l'exemple sous l'hypothèse que l'événement rare "le test est positif" se produise. C'est assez clair quand on fait les choses formellement, mais c'est souvent beaucoup moins évident quand on regarde une situation pratique sans faire attention.

1.7.2 Indépendance d'événements

Un concept fondamental en probabilité est celui d'indépendance. Dans l'application des probabilités, *l'indépendance est souvent une hypothèse que l'on fait* il est donc extrêmement important de comprendre ce que l'on entend par indépendance pour être capable de jauger si cette hypothèse est justifiée dans une situation donnée.

Définition 1.7.2 (Événements indépendants)**.** Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé. Soient $A, B \in \mathcal{F}$ deux événements. On dit que A et B sont *indépendants* si

$$P(A\cap B) = P(A)P(B).$$

On peut voir que l'indépendance nous dit que la réalisation de A ne donne aucune information sur la réalisation de B : la probabilité de B conditionnellement à A, $P(B \mid A)$, est la même que la probabilité de B (dans le cas $P(A) > 0$) :

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

Définition 1.7.3 (Événements deux-à-deux indépendants)**.** Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé. Soit I un ensemble. Soient $A_i \in \mathcal{F}, i \in I$ une famille d'événements indexée par I. On dit que les $A_i, i \in I$ sont *deux à deux indépendants* si

$$P(A_i \cap B_j) = P(A_i)P(A_j), \ \forall i, j \in I, i \neq j.$$

Définition 1.7.4 (Famille indépendante d'événements)**.** Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé. Soit I un ensemble. Soient $A_i \in \mathcal{F}, i \in I$ une famille d'événements indexée par I. On dit que les $A_i, i \in I$ sont *mutuellement indépendants* (ou *forment une famille indépendante*) si

$$P\Big(\bigcap_{i \in J} A_i\Big) = \prod_{i \in J} P(A_i), \ \forall J \subset I, \text{ fini.}$$

Exemple 1.7.2. On va voir un exemple d'une famille d'événements qui sont deux à deux indépendants mais qui ne forment pas une famille indépendante. On regarde l'expérience suivante : on lance deux dés à six faces (numérotés 1 et 2, équilibrés). On regarde alors les événements

— A_1 : le dé numéro 1 fait un nombre paire ;

— A_2 : le dé numéro 2 fait un nombre paire ;

— A_2 : la somme des deux dés est un nombre impaire.

On a alors que

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{2},$$

et que

$$P(A_1 \cap A_2) = P(A_2 \cap A_3) = P(A_3 \cap A_1) = \frac{1}{4}.$$

Donc les A_i sont deux à deux indépendants. Mais,

$$P(A_1 \cap A_2 \cap A_3) = 0 \neq \frac{1}{8} = P(A_1)P(A_2)P(A_3),$$

donc ils ne forment pas une famille indépendante.

1.7.3 Construction d'un espace produit

On a maintenant définit ce que l'on voulait dire par indépendance : du point de vue de la mesure, c'est une forme de *factorisation* de la mesure. On peut naturellement se demander si, étant donné deux espaces de probabilité, on peut en construire un nouveau dans lequel vivent une copie de chacun des espaces de départ, de sorte à ce que les événements ne dépendant que de l'espace 1 soient indépendants des événements ne dépendant que de l'espace 2. C'est intuitivement ce que l'on aimerait faire pour

modéliser sur un même espace de probabilité deux expériences qui sont isolées l'une de l'autre, comme faire deux lancés de pièce ou lancer une pièce et un dé.

On commence donc avec deux espaces de probabilité $(\Omega_1, \mathcal{F}_1, P_1)$ et $(\Omega_2, \mathcal{F}_2, P_2)$. Les résultats seront la donnée d'un résultat pour l'expérience 1 et d'un résultat pour l'expérience 2, donc $\Omega = \Omega_1 \times \Omega_2$. Pour la tribu, on aimerait pouvoir mesurer (au moins) la réalisation simultanée d'un événement de l'expérience 1 et d'un événement de l'expérience 2. On a déjà rencontré la tribu nous permettant de faire ceci, c'est la tribu produit : $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$.

Reste à construire une mesure appropriée, P. Formalisons ce que l'on attend d'elle : on aimerait que la probabilité la réalisation simultanée de l'événement A pour la première expérience (qui est donné par l'événement $A \times \Omega$ dans $\mathcal{F}$) et de l'événement B pour la seconde expérience (qui est donné par l'événement $\Omega \times B$ dans $\mathcal{F}$) soit égale à la probabilité de l'événement A dans la première expérience isolée multipliée par la probabilité de l'événement B dans la seconde expérience isolée :

$$P(A \times B) = P\big((A \times \Omega) \cap (\Omega \times B)\big) = P_1(A)P_2(B), \ \forall A \in \mathcal{F}_1, B \in \mathcal{F}_2.$$

Une mesure ayant cette propriété existe bel et bien, et est de plus *unique*. C'est l'objet du Théorème suivant.

Théorème 1.7.3. *Soient $(\Omega_1, \mathcal{F}_1, P_1)$ et $(\Omega_2, \mathcal{F}_2, P_2)$ deux espaces de probabilité. Alors, il existe une unique mesure $P = P_1 \otimes P_2$ sur $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ telle que*

$$P(A \times B) = P_1(A)P_2(B), \ \forall A \in \mathcal{F}_1, B \in \mathcal{F}_2.$$

Démonstration. (Esquisse, non-éligible à l'examen)

On va utiliser le Théorème d'extension de Carathéodory (Théorème 1.4.4). On veut donc se placer dans le bon cadre. On définit

$$\mathcal{C} = \{A \times B : \ A \in \mathcal{F}_1, B \in \mathcal{F}_2\};$$
$$\mathcal{A} = \bigcup_{n \geq 1} \big\{C_1 \sqcup \cdots \sqcup C_n : \ C_i \in \mathcal{C} \text{ sont 2 à 2 disjoints}\big\}.$$

On peut alors montrer que $\mathcal{A}$ est une algèbre d'ensembles (c.f. Théorème 1.4.4, exercice pour les motivés).

On définit alors $\rho : \mathcal{A} \to \mathbb{R}_+$ par

$$\rho\Big(\bigsqcup_{i=1}^{n}(A_i \times B_i)\Big) := \sum_{i=1}^{n} P_1(A_i)P_2(B_i).$$

L'additivité de P_1, P_2 permet alors de vérifier que ρ est bien définie et est finiement additive. On a alors que

— $\rho(A \times B) = P_1(A)P_2(B), \ \forall A \in \mathcal{F}_1, B \in \mathcal{F}_2$ comme voulu ;

— $\rho(\varnothing) = P_1(\varnothing)P_2(\varnothing) = 0$;

— $\rho(\Omega_1 \times \Omega_2) = P_1(\Omega_1)P_2(\Omega_2) = 1$.

Il nous manque donc la σ-additivité dénombrable de ρ pour conclure que ρ est une pré-mesure (c.f. Théorème 1.4.4). On peut ensuite conclure par le Théorème 1.4.4 que

ρ s'étend de manière unique à $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{A}) = \sigma(\mathcal{C})$. La preuve de la σ-additivité dénombrable est un peu plus pénible et repose sur l'identité suivante :

$$\rho(A \times B) = E_{P_1}\big(E_{P_2}(\mathbb{1}_A(\omega_1)\mathbb{1}_B(\omega_2))\big),$$

le fait que si $C = \bigsqcup_{i\geq 1} C_i$, on a

$$\mathbb{1}_C = \sum_{i\geq 1} \mathbb{1}_{C_i}$$

et l'usage du Théorème 1.6.11 pour échanger la somme infinie et les espérances.

□

1.7.4 Indépendance de variables aléatoires, covariance, variance, corrélation

Des variables aléatoires sont indépendantes si les événements "la variable i prend valeur dans A_i" sont indépendants pour différents i.

Définition 1.7.5. Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soit I un ensemble. Soient $X_i, i \in I$ une famille de variables aléatoires. On dit que $(X_i)_{i\in I}$ sont *mutuellement indépendantes* (ou *forment une famille indépendante*) si pour tout $J \subset I$ fini et tout $A_i \in \mathcal{B}(\mathbb{R}), i \in J$,

$$P\Big(\bigcap_{i\in J}\{X_i \in A_i\}\Big) = \prod_{i\in J} P(X_i \in A_i).$$

Un critère pour l'indépendance de variables aléatoires est lié aux fonctions de répartitions et aux fonctions de densité.

Lemme 1.7.4. *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient $n \geq 1$ et $X_1, \ldots, X_n$ une famille de variables aléatoires. Alors, les points suivants sont équivalents.*

1. *$(X_i)_{i=1}^n$ forment une famille indépendante.*
2. *$F_X(t_1, \ldots, t_n) = \prod_{i=1}^n F_{X_i}(t_i)$ pour tout $t_1, \ldots t_n \in \mathbb{R}$, où F_X est la fonction de répartition du vecteur $X = (X_1, \ldots, X_n)$.*
3. *Pour toutes fonctions $\varphi_1, \ldots \varphi_n : \mathbb{R} \to \mathbb{R}$, telle que $\varphi_i \circ X_i \in L^1_P(\Omega)$,*

$$E_P\big(\varphi_1(X_1)\ldots\varphi_n(X_n)\big) = \prod_{i=1}^n E_P\big(\varphi_i(X_i)\big).$$

Si de plus X est un vecteur aléatoire à densité, ces points sont équivalents à

— *la densité de X s'écrit sous forme produit :*

$$f_X(x_1, \ldots, x_n) = \prod_{i=1}^n g_i(x_i),$$

où les g_i sont des fonctions positives intégrables. Dans ce cas, $f_i(x_i) = \frac{g_i(x_i)}{\int_{\mathbb{R}} du g_i(u)}$ est une densité pour X_i ;

— *les X_i sont des variables aléatoires à densité et la fonction $f(x_1, \ldots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$ est une densité pour X.*

Démonstration. Sera vue en Théorie de la mesure. Notez que 1. $\implies$ 2., 3. $\implies$ 1. et 3. $\implies$ 2. sont directs (en prenant des indicatrices bien choisies). On peut obtenir 1. $\implies$ 3. en approximant les fonctions φ_i par des fonctions discrètes, un peu dans l'esprit de ce que l'on a fait pour l'espérance. 2. $\implies$ 3. suit la même idée, mais le type d'approximation doit être restreint à une classe bien choisie de fonctions. □

Notons que l'indépendance de deux variables aléatoires X, Y nous donne en particulier que

$$E_P(XY) = E_P(X)E_P(Y).$$

On peut donc voir la quantité

$$\mathrm{Cov}_P(X,Y) := E_P(XY) - E_P(X)E_P(Y) = E_P\big((X - E_P(X))(Y - E_P(Y))\big), \quad (1.4)$$

comme une forme plus ou moins valide de "à quel point X et Y s'influencent mutuellement". La quantité $\mathrm{Cov}_P(X,Y)$ est appelée la *covariance entre X et Y sous P*. On déduit de la définition que

$$(X,Y) \mapsto \mathrm{Cov}_P(X,Y),$$

est une application bilinéaire (voir ci-dessous). Cette quantité est souvent utilisée (à tord) comme étant systématiquement une mesure de l'indépendance entre X et Y. Bien que cela soit parfois justifié, c'est en général faux, comme le montre l'exemple suivant.

Exemple 1.7.3. Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité et X' une variable aléatoire de Bernoulli de paramètre $1/2$. Soit $X = 1 - 2X'$. Soit Z une variable aléatoire quelconque (par exemple binomiale) indépendante de X. On définit alors $Y = XZ$. On a clairement que Y et Z ne sont pas indépendantes (on a même que $|Y| = |Z|$). En revanche,

$$\begin{aligned}\mathrm{Cov}_P(Z,Y) &= E_P(ZY) - E_P(Z)E_P(Y)\\ &= E_P(Z^2X) - E_P(Z)E_P(XZ)\\ &= E_P(X)E_P(Z^2) - E_P(Z)^2E_P(X) = 0.\end{aligned}$$

La covariance d'une variable aléatoire avec elle même est appelée *variance*

$$\mathrm{Var}_P(X) = E_P(X^2) - E_P(X)^2 = E_P\big((X - E_P(X))^2\big). \quad (1.5)$$

Elle quantifie à quel point une variable à tendance à prendre des valeurs loin de sa valeur moyenne.

La covariance et la variances satisfont :

Lemme 1.7.5. *Soient X_1, X_2, X, Y des variables aléatoires telles que $E(X_1^2), E(X_2^2), E(X^2), E(Y^2) < \infty$. Alors, pour tout $a, b \in \mathbb{R}$,*

- *symétrie :* $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$;
- *bi-linéarité I :* $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1,Y) + \mathrm{Cov}(X_2,Y)$;
- *bi-linéarité II :* $\mathrm{Cov}(aX, bY) = ab\mathrm{Cov}(X,Y)$.

En particulier,

$$\mathrm{Var}(aX) = a^2\mathrm{Var}(X), \quad \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y).$$

Démonstration. Calcule direct, en exercice. □

En statistiques, un coefficient déduit à partir de la covariance et de la variance est fréquement utilisé : le *coefficient de corrélation* (plus précisément, le *coefficient de corrélation linéaire de Bravais-Pearson*) [19] :

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \in [-1,1].$$

Il est à noter que ce coefficient requiert que la variance soit bien définie (et donc que $E(X^2), E(Y^2) < \infty$). La covariance mesure à quel point le fait que X prenne des valeurs grandes positives/grandes négatives pousse la variable Y à prendre des valeurs grandes positives/grandes négatives. La quantité

$$\sigma_X = \sqrt{\mathrm{Var}(X)},$$

est appelée *écart type* et mesure à quel point une variable aléatoire va *typiquement* dévier de sa valeur moyenne (i.e. la magnitude de la déviation typique). Le coefficient de corrélation mesure donc "les chances que X influence Y" ainsi que si cette influence et un encouragement mutuel ($\rho_{XY} > 0$) ou un découragement mutuel ($\rho_{XY} < 0$), et non pas la magnitude de cette influence.

Il faut faire attention avec la notion de corrélation : on a vu que "ne pas être corrélé" ne veut pas dire "être indépendant". On va maintenant voir un exemple du fait qu'être fortement corréler n'implique pas qu'il y ait un lien de causalité entre deux variables (raccourcis trop souvent fait...).

Exemple 1.7.4 (Corrélation n'implique pas causalité)**.** Soit $X \sim \mathrm{Uni}(\{0,1,\dots,30\})$ une variable aléatoire (qui représente, par exemple, le nombre de jours de pluie au mois de novembre à Lausanne). On définit alors deux variables aléatoires :

$$Y = aX, \quad Z = bX, \quad a,b \in (0,+\infty).$$

On peut imaginer que Y représente le nombre de parapluies vendus à Lausanne pendant le mois et Z le nombre d'heures qu'un lausannois typique va passer devant Netflix durant le mois. On calcule alors la corrélation entre Y et Z :

$$\begin{gathered}\mathrm{Var}(Y) = a^2\mathrm{Var}(X), \quad \mathrm{Var}(Z) = b^2\mathrm{Var}(X),\\ \mathrm{Cov}(Y,Z) = ab\mathrm{Cov}(X,X) = ab\mathrm{Var}(X),\\ \rho_{XY} = \frac{\mathrm{Cov}(Y,Z)}{\sqrt{\mathrm{Var}(Y)\mathrm{Var}(Z)}} = \frac{ab\mathrm{Var}(X)}{\sqrt{a^2\mathrm{Var}(X)b^2\mathrm{Var}(X)}} = \frac{ab}{|ab|} = 1.\end{gathered}$$

On a donc que ces deux variables sont fortement corrélées, mais leur corrélation vient d'une *causalité commune* plutôt que d'une *causalité mutuelle*. On ne peut par exemple pas conclure que l'augmentation du temps passé sur Netflix augmente notre envie d'acheter un parapluie ou vice-versa...

19. Le fait que $|\rho_{XY}| \leq 1$ est non trivial, cela suit de l'inégalité de Cauchy-Schwartz que l'on verra plus tard dans le cours.

1.7.5 Produit infini

De la même façon qu'on a pu construire un espace produit fini, on peut construire un espace produit infini. Néanmoins, la construction est un peu plus compliquée que dans le cas du produit fini et sera vue en théorie de la mesure (la clé étant aussi l'usage du Théorème d'extension de Carathéodory). On se contentera d'énoncer le Théorème dont on aura besoin durant le cours.

La tribu naturelle pour les produits infinis d'espaces de probabilité est la tribu qui permet d'observer un nombre fini quelconque de ces espaces simultanément. C'est ce que l'on appelle la *tribu cylindrique*. Elle est définie par :

$$C_J = \big\{A_1 \times A_2 \times \ldots : \ A_i \in \mathcal{F}_i, i \in J, A_i = \Omega_i, i \notin J\big\}, \ J \subset \mathbb{N} \text{ fini },$$
$$\bigotimes_{i\geq 1} \mathcal{F}_i = \sigma\Big(\bigcup_{J\subset\mathbb{N}:|J|<\infty} C_J\Big).$$

Théorème 1.7.6 (Théorème d'extension de Kolmogorov, version produit)**.** *Soient $(\Omega_i, \mathcal{F}_i, P_i), i \geq 1$ des espaces de probabilité. Alors, il existe une unique mesure de probabilité P sur $\big(\bigtimes_{i\geq 1}\Omega_i, \bigotimes_{i\geq 1}\mathcal{F}_i\big)$ telle que pour tout $J \subset \mathbb{N}$ fini et tout $A_i \in \mathcal{F}_i, i \in J$,*

$$P(A) = \prod_{i\in J} P_i(A_i),$$

où $A = A_1 \times A_2 \times \ldots$ et $A_i = \Omega_i, i \notin J$.

Un corollaire immédiat est

Corollaire 1.7.7. *Soient $X_1, X_2, \ldots$ des variables aléatoires (chacune définie sur un espace $(\Omega_i, \mathcal{F}_i, P_i)$). Alors, il existe un espace $(\Omega, \mathcal{F}, P)$ et des variables aléatoires $Y_1, Y_2, \ldots : \Omega \to \mathbb{R}$ tels que*

- *$(Y_i)_* P = (X_i)_* P_i$ pour tout $i \geq 1$;*
- *$(Y_i)_{i\geq 1}$ forme une famille indépendante.*

On utilisera souvent ce corollaire pour garantir l'existence d'une famille indépendante de variables aléatoires ayant des marginales prescrites. L'exemple le plus courant étant

Définition 1.7.6 (Suite i.i.d.)**.** Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Une *suite Indépendante Identiquement Distribuée* (suite i.i.d.) est une famille indépendante de variables aléatoires $X_1, X_2, \ldots$ ayant toutes la même loi.

Comme le Corollaire 1.7.7 nous garantit l'existence d'un espace sur lequel nos variables peuvent vivre simultanément, on ne mentionnera pas toujours ce dernier dans les énoncés.

1.8 Quelques variables aléatoires classiques et leurs propriétés

1.8.1 Identifications des résultats avec des réels

La fonction évaluation

Dans le cas $\Omega \subset \mathbb{R}$, on peut considérer la variable aléatoire $X : \Omega \to \mathbb{R}$ donnée par

$$X(\omega) = \omega.$$

Cette variable aléatoire peut sembler inutile au premier abord, mais elle est fondamentale (et son usage est généralement implicite). En effet, quand on fait une expérience dont les issues possibles sont des nombres (par exemple : jeter un dé ou jouer à la roulette), on identifie intuitivement le *résultat de l'expérience* avec la *valeur réelle associée*. La différence (subtile) entre les deux est que le premier est juste un élément dans un ensemble (en particulier, sa valeur moyenne n'est pas définie) alors que le second est un élément d'un espace vectoriel (même d'un corps) dont la structure est beaucoup plus riche. La variable aléatoire évaluation est simplement la variable aléatoire formalisant cette identification intuitive. Dans la pratique, il est fréquent d'omettre l'usage de cette variable aléatoire, ou de le garder implicite en faisant un abus de notation.

La fonction coordonnée

De la même manière que l'on a vu la fonction évaluation si $\Omega \subset \mathbb{R}$, on définit les fonctions coordonnées si $\Omega \subset \mathbb{R}^d$ par

$$X_i(\omega) = \omega_i,\ i = 1, \dots, d,$$

où $\omega = (\omega_1, \dots, \omega_d)$ avec $\omega_i \in \mathbb{R}$.

1.8.2 Variables aléatoires liées aux exemples de la section 1.5

Dans chaque exemples de la section 1.5, l'espace des résultats est un sous-ensemble de $\mathbb{R}$. On peut donc regarder l'image de ces différentes mesures de probabilité par la fonction évaluation, ce qui nous donne une nouvelle mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Dans tous les exemples qui vont suivre, on notera X la variable évaluation :

$$X : \Omega \to \mathbb{R}, \quad X(\omega) = \omega.$$

Variables aléatoires discrètes

Variable aléatoire de Bernoulli

On regarde la loi de Bernoulli de paramètre p, P_p, sur l'espace $(\{0,1\}, \mathcal{P}(\{0,1))$. Soit $(\Omega', \mathcal{F}', Q)$ un espace de probabilité et $Y : \Omega' \to \mathbb{R}$ une variable aléatoire. On dit que Y est *une variable aléatoire de Bernoulli de paramètre* p

si [20]

$$Y_*Q = X_*P_p.$$

On notera $Y \sim \text{Bern}(p)$ pour "Y est une variable aléatoire de Bernoulli de paramètre p".

Variable aléatoire binomiale

On regarde cette fois la loi binomiale de paramètres $n \in \mathbb{N}, p \in [0,1]$, notée $P_{n,p}$, sur $(\{0,1,\dots,n\}, \mathcal{P}(\{0,1,\dots,n\}))$. Comme pour la variable aléatoire de Bernoulli, si $(\Omega', \mathcal{F}', Q)$ est un espace de probabilité et $Y : \Omega' \to \mathbb{R}$ une variable aléatoire, on dira que Y est *une variable aléatoire binomiale de paramètres* n, p, noté $Y \sim \text{Bin}(n,p)$, si

$$Y_*Q = X_*P_{n,p}.$$

La loi binomiale représente le nombre de 1 obtenus si on effectue n expériences de Bernoulli indépendantes.

Lemme 1.8.1. *Soient $n \in \mathbb{N}, p \in [0,1]$. Soient $X_1, \dots, X_n$ une famille indépendante de variables de Bernoulli de paramètre p. On définit la variable aléatoire*

$$Y = \sum_{k=1}^{n} X_k.$$

Alors, $Y \sim \text{Bin}(n,p)$.

Démonstration. En exercice. □

Variable aléatoire géométrique

On procède comme dans les cas précédents. Soit $p \in [0,1]$. On considère l'espace de probabilité $(\mathbb{N}, \mathcal{P}(\mathbb{N}), P)$ où P est donnée par

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

On dira qu'une variable aléatoire Y (définie sur un espace de probabilité $(\Omega, \mathcal{F}, P')$) est *une variable aléatoire géométrique*, noté $Y \sim \text{Geo}(p)$, si $X_*P = Y_*P'$.

Une variable aléatoire géométrique est le nombres d'expériences de Bernoulli indépendantes nécessaire pour obtenir un 1.

Lemme 1.8.2. *Soient $X_1, X_2, \dots$ une famille indépendante dénombrable de variables aléatoires de Bernoulli de paramètre p. On définit*

$$Y = 1 + \sum_{n \geq 1} \prod_{i=1}^{n} (1 - X_i),$$

le nombre de lancers nécessaires pour obtenir un 1. Alors, $Y \sim \text{Geo}(p)$.

20. Cela peut paraître idiot de mettre autant de formalisme pour quelque chose d'aussi élémentaire, mais ce formalisme nous permet de gagner en flexibilité : une variable aléatoire de Bernoulli peut être obtenue par des expériences bien plus complexes que de jeter une pièce. Par exemple, si on regarde l'expérience qui est "prédire la météo sur l'année 3067" l'espace de probabilité est très complexe (beaucoup plus que celui de la loi de Bernoulli), mais si l'on pose la question "fera-t-il beau le 26 mars 3067 ?" (qui se formalise par une variable aléatoire prenant valeur 1 si oui et 0 si non), on se retrouve avec une variable aléatoire de Bernoulli (de paramètre à définir).

Démonstration. En exercice. □

Lemme 1.8.3 (Perte de mémoire)**.** *Soit $Y \sim \mathrm{Geo}(p)$ une variable aléatoire géométrique. Notons P sa loi. Alors, pour tout $n > k \in \mathbb{N}$,*

$$P(Y = n \,|\, Y > k) = P(Y = n - k).$$

En particulier, sous la loi $P(\ |Y > k)$, $Y - k$ suit une loi géométrique de paramètre p[21]*.*

Démonstration. On commence par calculer $P(Y > a)$ pour $a \geq 0$ (entier) :

$$P(Y > a) = \sum_{k>a} P(Y = k) = \sum_{k>a} p(1-p)^{k-1} = p \sum_{k \geq 0} (1-p)^{k+a} = (1-p)^a.$$

On a alors que pour $n > k$,

$$P(Y = n \,|\, Y > k) = \frac{P(Y = n)}{P(Y > k)} = \frac{p(1-p)^{n-1}}{(1-p)^k} = p(1-p)^{n-k-1} = P(Y = n - k).$$

□

On peut voir cette perte de mémoire intuitivement de la manière suivante : une géométrique est le nombre de lancés de pièce indépendants qu'il faut pour faire un 1. Si on fait une pause après k lancés et que les k premières pièces ont toutes fait 0, les pièces suivantes étant indépendantes des k premières, on se retrouve juste avec une suite de lancés indépendants, exactement comme quand on a commencé.

Variable aléatoire de Poisson

On procède comme dans les cas précédents. Soit $\lambda \in \mathbb{R}_+$. On considère l'espace de probabilité $(\mathbb{Z}_+, \mathcal{P}(\mathbb{Z}_+), P)$ où P est donnée par

$$P(k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

On dira qu'une variable aléatoire Y (définie sur un espace de probabilité $(\Omega, \mathcal{F}, P')$) est *une variable aléatoire de Poisson de paramètre* λ, noté $Y \sim \mathrm{Poi}(\lambda)$, si $X_* P = Y_* P'$. Une caractérisation utile des variables de Poisson est

Lemme 1.8.4. *Soit Y une variable aléatoire (sur un espace de probabilité $(\Omega, \mathcal{F}, P)$). Alors les points suivants sont équivalents :*

1. *$Y \sim \mathrm{Poi}(\lambda)$;*
2. *$P(Y = 0) = e^{-\lambda}$ et pour tout $k \in \mathbb{Z}_+$,*

$$\frac{P(Y = k+1)}{P(Y = k)} = \frac{\lambda}{k+1}.$$

Démonstration. En exercice. □

21. Pour énoncer les choses très formellement : on se donne $(\Omega, \mathcal{F}, P)$ un espace de probabilité et $Y : \Omega \to \mathbb{R}$ une variable aléatoire géométrique de paramètre p. Soit $k \geq 0$. Notons P_k la loi conditionnelle $P(\ |Y > k)$ (toujours sur $(\Omega, \mathcal{F})$). Notons Z la variable aléatoire $Z = Y - k$. L'énoncé nous dit alors que $Z_* P_k = Y_* P$.

Variable aléatoire uniforme (cas fini)

On procède comme dans les cas précédents. Soit $n \in \mathbb{N}$. On considère l'espace de probabilité $(\{1,\dots,n\}, \mathcal{P}(\{1,\dots,n\}), P)$ où $P(k) = \frac{1}{n}$ pour tout $k \in \{1,\dots,n\}$. On dira qu'une variable aléatoire Y (définie sur un espace de probabilité $(\Omega, \mathcal{F}, P')$) est *une variable aléatoire uniforme sur* $\{1,\dots,n\}$, noté $Y \sim \mathrm{Uni}(\{1,\dots,n\})$, si $X_*P = Y_*P'$.

Une propriété utile des variables uniformes est que les lois conditionnelles sont elles-même uniformes :

Lemme 1.8.5. *Soit $Y \sim \mathrm{Uni}(\{1,\dots,n\})$ une variable aléatoire uniforme. Notons P sa loi. Soit $A \subset \{1,\dots,n\}$. Alors, pour tout $B \subset A$,*

$$P(Y \in B \,|\, Y \in A) = \frac{|B|}{|A|}.$$

Démonstration. En exercices. □

Variables aléatoires à densité

Variable aléatoire uniforme sur un intervalle

Soient $a < b \in \mathbb{R}$. On considère l'espace de probabilité $[a,b], \mathcal{B}([a,b]), P$ où $P(A) = \frac{1}{b-a}\int_a^b \mathbb{1}_A(x)d\ell(x)$. On dira qu'une variable aléatoire Y (sur un espace $(\Omega, \mathcal{F}, P')$) suit une *loi uniforme sur l'intervalle* $[a,b]$, noté $Y \sim \mathrm{Uni}([a,b])$, si $X_*P = Y_*P'$.

Une variable aléatoire uniforme est une variable a densité. Une densité est donnée par $\frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$. Voir remarque 1.6.7.

Comme dans le cas fini, les variables uniformes ont de bonne propriétés face au conditionnement.

Lemme 1.8.6. *Soient $a < b < c < d \in \mathbb{R}$. Alors, si $Y \in \mathrm{Uni}([a,d])$,*

$$P(t_1 \le Y \le t_2 \,|\, b \le Y \le c) = \frac{t_2 - t_1}{c-b},\ \forall b \le t_1 < t_2 \le c$$

Démonstration. En exercice. □

Variables aléatoires gaussiennes

On considère l'espace de probabilité $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ où P est donnée par

$$P(A) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} dx e^{-x^2/2}\mathbb{1}_A(x).$$

On dira qu'une variable aléatoire Y (sur un espace $(\Omega, \mathcal{F}, P')$) est une *variable aléatoire normale standard* (ou *variable aléatoire gaussienne centrée réduite*), noté $Y \sim \mathcal{N}(0,1)$, si $X_*P = Y_*P'$.

Plus généralement, on définit la loi normale de moyenne $\mu \in \mathbb{R}$ et de variance σ^2, $\sigma \in \mathbb{R}_+$, via

$$P_{\mu,\sigma^2}(A) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{+\infty} dx e^{-(x-\mu)^2/(2\sigma^2)}\mathbb{1}_A(x).$$

On dira qu'une variable aléatoire Y (sur un espace $(\Omega, \mathcal{F}, P')$) est une *variable aléatoire normale de moyenne μ et de variance σ^2*, noté $Y \sim \mathcal{N}(\mu, \sigma^2)$, si $X_* P_{\mu,\sigma^2} = Y_* P'$.

L'intérêt et les propriétés d'universalité des variables aléatoires gaussiennes viennent de

Lemme 1.8.7. *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soit $Y_1, Y_2 : \Omega \to \mathbb{R}$ deux variables aléatoires indépendantes. Supposons que $Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et que $Y_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Alors,*

1. *la variable $\tilde{Y} = (Y_1 - \mu_1)/\sigma_1$ est une variable aléatoire gaussienne centrée réduite : $\tilde{Y} \sim \mathcal{N}(0,1)$;*
2. *la variable $Z = Y_1 + Y_2$ est une variable aléatoire gaussienne : $Z \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.*

Démonstration. Dans chacun des deux cas, il suffit de calculer les fonctions de répartition associées. On commence par $\tilde{Y}$:

$$F_{\tilde{Y}}(t) = P(Y_1 \leq \sigma_1 t + \mu_1) = \int_{-\infty}^{\sigma_1 t + \mu_1} \frac{dx}{\sqrt{2\pi\sigma_1^2}} e^{-(x-\mu_1)^2/(2\sigma_1^2)} = \int_{-\infty}^{t} \frac{dy}{\sqrt{2\pi}} e^{-y^2/2}$$

où on a utilisé le changement de variable $y = \frac{x-\mu_1}{\sigma_1}$. La dernière intégrale est bien la fonction de répartition d'une $\mathcal{N}(0,1)$.

On se tourne vers Z : par indépendance entre Y_1 et Y_2 (Lemme 1.7.4),

$$\begin{aligned} F_{Y_1+Y_2}(t) &= E\big(\mathbb{1}_{(-\infty,t]}(Y_1 + Y_2)\big) \\ &= \int_{-\infty}^{+\infty} dx \frac{e^{-(x-\mu_1)^2/(2\sigma_1^2)}}{\sqrt{2\pi\sigma_1^2}} \int_{-\infty}^{+\infty} dy \frac{e^{-(y-\mu_2)^2/(2\sigma_2^2)}}{\sqrt{2\pi\sigma_2^2}} \mathbb{1}_{(-\infty,t]}(x+y). \end{aligned}$$

On fait alors deux fois le même changement de variable qu'avant pour obtenir

$$\int_{-\infty}^{+\infty} dx \frac{e^{-x^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dy \frac{e^{-y^2/2}}{\sqrt{2\pi}} \mathbb{1}_{(-\infty,t]}(x\sigma_1 + \mu_1 + y\sigma_2 + \mu_2).$$

En ré-organisant, on obtient

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy e^{-(x^2+y^2)/2} \mathbb{1}_{(-\infty, t-\mu_1-\mu_2]}(x\sigma_1 + y\sigma_2).$$

On fait alors le changement de variables

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \sigma_1 & \sigma_2 \\ -\sigma_2 & \sigma_1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Ce qui donne

$$\frac{1}{2\pi(\sigma_1^2 + \sigma_2^2)} \int_{-\infty}^{+\infty} dx' \int_{-\infty}^{+\infty} dy' e^{-((x')^2+(y')^2)/(2(\sigma_1^2+\sigma_2^2))} \mathbb{1}_{(-\infty, t-\mu_1-\mu_2]}(x'),$$

car

$$\begin{aligned} x^2 + y^2 &= \frac{1}{(\sigma_1^2+\sigma_2^2)^2} \Big((\sigma_1 x' - \sigma_2 y')^2 + (\sigma_2 x' + \sigma_1 y')^2 \Big) \\ &= \frac{1}{(\sigma_1^2+\sigma_2^2)^2} \Big(\sigma_1^2 (x')^2 + \sigma_2^2 (y')^2 + \sigma_2^2 (x')^2 + \sigma_1^2 (y')^2 \Big) = \frac{1}{\sigma_1^2+\sigma_2^2} \big((x')^2 + (y')^2 \big). \end{aligned}$$

Cette dernière identité vient du fait que l'on a simplement fait un changement d'une base orthogonal de $\mathbb{R}^2$ vers une autre. Maintenant, on peut effectuer l'intégrale sur y' : l'indicatrice ne dépend que de x'. Le changement de variable $y = \frac{y'}{\sqrt{\sigma_1^2+\sigma_2^2}}$ ainsi que le fait que l'intégrale gaussienne vaille $\sqrt{2\pi}$ donne qu'elle vaut $\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}$. En notant $x' \equiv x$, on obtient alors

$$\frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} \int_{-\infty}^{+\infty} dx e^{-x^2/(2(\sigma_1^2+\sigma_2^2))} \mathbb{1}_{(-\infty,t]}(x+\mu_1+\mu_2).$$

En translatant x par $\mu_1+\mu_2$, on obtient bien la fonction caractéristique d'une $\mathcal{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$. □

En itérant ce Lemme, on remarque que si $X_1, \ldots, X_n$ forment une famille i.i.d. de $\mathcal{N}(0,1)$, alors leur somme renormalisée par $\sqrt{n}$ est aussi une $\mathcal{N}(0,1)$:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \sim \mathcal{N}(0,1).$$

Cette propriétés des gaussiennes est une des raisons derrière le Théorème Central Limite.

Variable aléatoire Exponentielle

On considère l'espace de probabilité $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ où P est donnée par

$$P(A) = \lambda \int_{-\infty}^{+\infty} dx e^{-\lambda x} \mathbb{1}_{[0,\infty)}(x) \mathbb{1}_A(x).$$

On dira qu'une variable aléatoire Y (sur un espace $(\Omega, \mathcal{F}, P')$) est une *variable aléatoire exponentielle de paramètre* λ, noté $Y \sim \mathrm{Exp}(\lambda)$, si $X_* P = Y_* P'$.

Les variables exponentielles sont le pendant continu des variables géométriques. On a la même propriété de perte de mémoire :

Lemme 1.8.8. *Soit $Y \sim \mathrm{Exp}(\lambda)$, $\lambda > 0$. Alors, pour tout $0 < a < b$,*

$$P(Y \geq b \,|\, Y \geq a) = P(Y \geq b-a).$$

Démonstration. En exercices. □

Variable aléatoire de Cauchy

On considère l'espace de probabilité $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ où P est donnée par

$$P(A) = \int_{-\infty}^{+\infty} dx \frac{\alpha}{\pi((x-x_0)^2+\alpha^2)} \mathbb{1}_A(x).$$

On dira qu'une variable aléatoire Y (sur un espace $(\Omega, \mathcal{F}, P')$) est une *variable aléatoire de Cauchy de paramètres* (x_0, α), noté $Y \sim \mathrm{Cauchy}(x_0, \alpha)$, si $X_* P = Y_* P'$.

On peut remarquer qu'une variable aléatoire de Cauchy n'admet pas d'espérance. On regarde le cas $x_0 = 0$ et $\alpha = 1$,

$$\int_{-\infty}^{+\infty} dx \frac{|x|}{\pi(x^2+1)} \geq \frac{1}{\pi} \int_{1}^{+\infty} dx \frac{1}{x+1} = \infty.$$

1.8.3 Tableau récapitulatif des variables aléatoires classiques

Variable	Espérance	Variance	Propriété
Bern(p)	p	$p(1-p)$	-
Bin(n,p)	np	$np(1-p)$	nb. de 1 parmi n Bern(p) indép.
Geo(p)	$\frac{1}{p}$	$\frac{1-p}{p^2}$	nb. de Bern(p) indép. pour faire un 1
Uni($\{0,1,\dots,n\}$)	$\frac{n}{2}$	$\frac{n^2+2}{12}$	Lois conditionnelles uniformes
Uni($\{1,\dots,n\}$)	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	Lois conditionnelles uniformes
Poi(λ)	λ	λ	-
Uni($[a,b]$)	$\frac{b-a}{2}$	$\frac{(b-a)^2}{12}$	Lois conditionnelles uniformes
$\mathcal{N}(0,1)$	0	1	$\mathcal{N}(a,b) \overset{\text{Loi}}{=} a+\sqrt{b}\mathcal{N}(0,1)$
Exp(λ)	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	Perte de mémoire
Cauchy(x_0,α)	N.D.	N.D.	N'a pas de premier moment

Chapitre 2

Notions de convergence et lois limites

2.1 Modes de convergence

2.1.1 Les différents modes de convergence

On aimerait formaliser l'idée qu'une suite de variables aléatoire "ressemble de plus en plus" à une variable aléatoire donnée. Cette notion est celle de *convergence*, déjà rencontrée de nombreuses fois dans d'autres cours... la question est donc "qu'entend-t-on par convergence de variables aléatoires ?". Il existe plusieurs manières de formaliser cette convergence, qui permettent de déduire plus ou moins d'information sur les propriétés de la limite à partir des propriétés de la suite. On va voir ici les notions principalement utilisées de convergence pour les variables aléatoires.

Définition 2.1.1 (Convergence en loi)**.** Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient $X, X_1, X_2, \ldots : \Omega \to \mathbb{R}$ des variables aléatoires. On dit que la suite $(X_i)_{i\geq 1}$ *converge en loi* vers X, noté $X_n \xrightarrow{\text{Loi}} X$, si $F_{X_n}(t) \to F_X(t)$ pour tout $t \in \mathbb{R}$ tel que F_X est continue en t [1].

La convergence en loi peut être définie de plusieurs façons équivalentes. Le Théorème regroupant ces définitions est appelé le *Théorème porte-manteau* :

Théorème 2.1.1 (Théorème porte-manteau (non-examinable))**.** *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient $X, X_1, X_2, \ldots : \Omega \to \mathbb{R}$ des variables aléatoires. Alors, les assertions suivantes sont équivalentes.*

1. *$F_{X_n}(t) \to F_X(t)$ pour tout $t \in \mathbb{R}$ tel que F_X est continue en t.*
2. *Pour toute fonction $\varphi : \mathbb{R} \to \mathbb{R}$ continue et bornée, $E_P(\varphi(X_n)) \xrightarrow{n\to\infty} E_P(\varphi(X))$.*
3. *Pour toute fonction $\varphi : \mathbb{R} \to \mathbb{R}$ bornée et uniformément continue, $E_P(\varphi(X_n)) \xrightarrow{n\to\infty} E_P(\varphi(X))$.*
4. *Pour tout fermé $F \subset \mathbb{R}$, $\limsup_{n\to\infty} P(X_n \in F) \leq P(X \in F)$.*
5. *Pour tout ouvert $O \subset \mathbb{R}$, $\liminf_{n\to\infty} P(X_n \in O) \geq P(X \in O)$.*
6. *Pour tout borélien $A \subset \mathbb{R}$ tel que $P(X \in \partial A) = 0$, $\lim_{n\to\infty} P(X_n \in A) = P(X \in A)$.*

1. Il est plus standard de voir la définition : $X_n \xrightarrow{\text{Loi}} X$ si $E_P(\varphi(X_n)) \xrightarrow{n\to\infty} E_P(\varphi(X))$ pour toute fonction $\varphi : \mathbb{R} \to \mathbb{R}$ continue et bornée. Cette définition est équivalente à celle donnée.

On ne démontrera pas ce Théorème, ce sera fait en théorie de la mesure. Les intéressés pourront jeter un oeil à la page wikipedia dédiée : `https://fr.wikipedia.org/wiki/Th%C3%A9or%C3%A8me_porte-manteau#D%C3%A9monstration_du_th%C3%A9or%C3%A8me_porte-manteau`. Certaines des implications sont tout à fait démontrables avec ce que l'on a déjà vu. On utilise principalement l'équivalence entre 1. et 2..

Définition 2.1.2 (Convergence en probabilité)**.** Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient $X, X_1, X_2, \ldots : \Omega \to \mathbb{R}$ des variables aléatoires. On dit que la suite $(X_i)_{i\geq 1}$ *converge en probabilité* vers X, noté $X_n \xrightarrow{\text{Proba}} X$, si pour tout $\epsilon > 0$, $\lim_{i\to\infty} P(|X - X_i| \geq \epsilon) = 0$.

Définition 2.1.3 (Convergence presque sûre)**.** Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient $X, X_1, X_2, \ldots : \Omega \to \mathbb{R}$ des variables aléatoires. On dit que la suite $(X_i)_{i\geq 1}$ *converge presque sûrement* vers X, noté $X_n \xrightarrow{\text{p.s.}} X$, si $P(\lim_{n\to\infty} X_n = X) = 1$.

Remarque 2.1.1. *Notez que la convergence en loi fait du sens même si les variables aléatoires ne sont pas toutes définies sur le même espace de probabilité. À l'opposé, il est très important pour la convergence en probabilité et pour la convergence presque sûre que toutes les variables vivent sur le même espace !*

Lemme 2.1.2. $P(\lim_{n\to\infty} X_n = X) = 1$ *est équivalent à*

$$\forall \epsilon > 0, \ \lim_{n\to\infty} P(\sup_{m\geq n} |X_m - X| \geq \epsilon) = 0.$$

Démonstration. La convergence presque sûre de X_n vers X est le fait que l'ensemble

$$\{\omega \in \Omega : \ \exists 1 > \epsilon > 0, \limsup_{n\to\infty} |X_n(\omega) - X(\omega)| \geq \epsilon\}$$

est négligeable. Mais cet ensemble peut se ré-écrire

$$\bigcup_{\epsilon>0}\{\omega \in \Omega : \ \inf_{m\geq 1} \sup_{n\geq m} |X_n(\omega) - X(\omega)| \geq \epsilon\}$$
$$= \bigcup_{l\geq 1} \bigcap_{m\geq 1} \{\omega \in \Omega : \ \sup_{n\geq m} |X_n(\omega) - X(\omega)| \geq \frac{1}{l}\}.$$

où on a utilisé que si $1 > \epsilon > 0$, il existe $l \in \mathbb{N}$ tel que $1/(l+1) \leq \epsilon \leq 1/l$. L'union sur l est une limite croissante, on peut donc utiliser le Théorème 1.1.2 pour obtenir

$$P\Big(\bigcup_{l\geq 1} \bigcap_{m\geq 1} \{\omega \in \Omega : \ \sup_{n\geq m} |X_n(\omega) - X(\omega)| \geq \frac{1}{l}\}\Big) =$$
$$= \sup_{l\geq 1} P\Big(\bigcap_{m\geq 1} \{\omega \in \Omega : \ \sup_{n\geq m} |X_n(\omega) - X(\omega)| \geq \frac{1}{l}\}\Big).$$

On a donc que la convergence presque sûre est équivalente à : pour tout $l \geq 1$,

$$P\Big(\bigcap_{m\geq 1} \{\omega \in \Omega : \ \sup_{n\geq m} |X_n(\omega) - X(\omega)| \geq \frac{1}{l}\}\Big) = 0,$$

en utilisant que l'intersection sur m est une limite décroissante et que pour tout $1 > \epsilon > 0$ il existe $l \in \mathbb{N}$ tel que $1/(l+1) \leq \epsilon \leq 1/l$, la convergence presque sûre est alors équivalente à

$$\forall \epsilon > 0, \ \lim_{m \to \infty} P\Big(\{\omega \in \Omega : \ \sup_{n \geq m} |X_n(\omega) - X(\omega)| \geq \epsilon\}\Big) = 0,$$

qui est le résultat voulu. □

Ces trois premiers modes de convergence sont vraiment liés à la notion de mesure : ce sont des modes de convergence pour la loi de X_n vers la loi de X. Les modes de convergence suivants sont plus "analyse fonctionnelle" en essence : ils sont liés à la notion de convergence dans des espaces de fonctions et traite de la convergence de la fonction X_n vers la fonction X.

Définition 2.1.4 (Convergence essentiellement uniforme)**.** Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient $X, X_1, X_2, \ldots : \Omega \to \mathbb{R}$ des variables aléatoires. On dit qu'une variable aléatoire Y est essentiellement bornée s'il existe $M > 0$ tel que $P(|Y| < M) = 1$. On note alors

$$\|Y\|_{P,\infty} = \inf\{M > 0 : \ P(|Y| < M) = 1\}.$$

On dit que la suite $(X_i)_{i \geq 1}$ *converge essentiellement uniformément* vers X, noté $X_n \xrightarrow{L^\infty} X$, si

— $X, X_1, X_2, \ldots$ sont essentiellement bornées ;

— $\lim_{n \to \infty} \|X - X_n\|_{P,\infty} = 0$.

On peut voir la convergence essentiellement uniforme comme une limite $p \to \infty$ du mode de convergence suivant. On le nomme aussi parfois *convergence dans* L^∞. On notera $L^\infty_P(\Omega)$ l'ensemble des variables aléatoires essentiellement bornées pour P.

Définition 2.1.5 (Convergence dans L^p)**.** Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient $X, X_1, X_2, \ldots : \Omega \to \mathbb{R}$ des variables aléatoires. Soit $p \in (0, +\infty)$. On dit que la suite $(X_i)_{i \geq 1}$ *converge dans* L^p vers X, noté $X_n \xrightarrow{L^p} X$, si

— $E_P[|X_i|^p] < \infty$ pour tout i suffisamment grand ;

— $E_P[|X|^p] < \infty$;

— $\lim_{i \to \infty} E_P[|X_i - X|^p] = 0$.

Noter que ce dernier mode de convergence requiert une condition de la forme $|X|^p$ est P-intégrable, ce que l'on a dénoté $|X|^p \in L^1_P(\Omega)$. On généralise la notation utilisée pour l'espace des variables aléatoires intégrables comme suit

Définition 2.1.6 (Espace L^p)**.** Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soit $p \in (0, +\infty)$. On dira qu'une variable aléatoire $X : \Omega \to \mathbb{R}$ *possède un moment d'ordre* p si $E_P(|X|^p) < \infty$. On notera $L^p_P(\Omega)$ l'ensemble des variables aléatoires ayant un moment d'ordre p.

On réservera la terminologie "avoir un moment d'ordre p" au cas $p \geq 1$. Dans le cas $p < 1$, on parlera de "moments fractionnaires". Le plus souvent, on considérera $p \in \mathbb{N}$.

On rencontre souvent une condition plus forte que d'avoir des moments d'ordre p pour tout $p \in \mathbb{N}$: la condition de moments exponentiels.

Définition 2.1.7 (Moments exponentiels)**.** Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. On dira qu'une variable aléatoire $X : \Omega \to \mathbb{R}$ *possède des moments exponentiels* si il existe $\delta > 0$ tel que

$$E_P(e^{tX}) < \infty, \quad \forall t \in (-\delta, \delta).$$

La condition "avoir des moments exponentiels" est souvent appelée "condition de Cramér"[2] et est fondamentale en théorie des grandes déviations (qui est l'étude des questions du type : "quelle est la probabilité qu'une somme de variables aléatoires prenne des valeurs atypiquement grandes ?").

2.2 Inégalités classiques

2.2.1 Inégalité de Markov

Lemme 2.2.1. *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité et $X : \Omega \to \mathbb{R}$ une variable aléatoire presque sûrement positive*[3]*. Alors, pour tout $a > 0$,*

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

Démonstration. Comme X est presque sûrement positive, l'ensemble $B = \{\omega \in \Omega : X \geq 0\}$ a probabilité 1. Notons $A = \{\omega \in \Omega : \ X(\omega) \geq a\}$. On a alors les inégalités

$$\mathbb{1}_A(\omega) a \leq X(\omega)\mathbb{1}_A(\omega) \leq X(\omega), \ \forall \omega \in B.$$

En prenant l'espérance de ces variables, on a par le Théorème 1.6.10 que

$$aP(A) \leq E(X\mathbb{1}_A) \leq E(X),$$

ce qui est le résultat voulu. □

2.2.2 Inégalité de Tchebychev

L'inégalité de Tchebychev est juste une re-formulation de l'inégalité de Markov.

Lemme 2.2.2. *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité et $X : \Omega \to \mathbb{R}$ une variable aléatoire. Soit $g : \mathbb{R} \to \mathbb{R}_+$ une fonction croissante mesurable. Alors,*

$$P\big(X \geq a\big) \leq \frac{E(g(X))}{g(a)}, \ \forall a \in \mathbb{R}. \tag{2.1}$$

En particulier, on a

— *Inégalité de Tchebychev version moment d'ordre p : pour tout $p \in (0, +\infty)$,*

$$P\big(X \geq a\big) \leq \frac{E(|X|^p)}{a^p}, \ \forall a > 0. \tag{2.2}$$

— *Inégalité de Tchebychev exponentielle : pour tout $\delta \in \mathbb{R}_+$,*

$$P\big(X \geq a\big) \leq e^{-\delta a} E(e^{\delta X}), \ \forall a > 0. \tag{2.3}$$

2. D'après le mathématicien suédois Harald Cramér.
3. I.e. : $P(X \geq 0) = 1$.

Démonstration. On applique l'inégalité de Markov à la variable aléatoire presque-sûrement positive $g \circ X$ pour obtenir

$$P\big(g(X) \geq g(a)\big) \leq \frac{E(g(X))}{g(a)}, \ \forall a \in \mathbb{R}.$$

On remarque alors que, comme g est croissante, $g(X) \geq g(a)$ si et seulement si $X \geq a$. Les deux cas particuliers correspondent aux choix $g(x) = |x|^p$ et $g(x) = e^{\delta x}$. □

Une application simple de l'inégalité de Tchebychev est une preuve de la loi faible des grands nombres pour des variables aléatoires non corrélées avec une variance finie.

Théorème 2.2.3 (Loi faible des grands nombres, version second moment)**.** *Soient $X_1, X_2, \ldots$ une suite de variables aléatoires identiquement distribuées et non corrélées :*

$$\mathrm{Cov}(X_i, X_j) = 0, \ i \neq j.$$

Supposons que $E(X_1) = \mu$ et $\mathrm{Var}(X_1) = \sigma^2$. On note $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Alors, pour tout $\epsilon > 0$,

$$P\big(|\bar{X}_n - \mu| \geq \epsilon\big) \leq \frac{\sigma^2}{\epsilon^2 n}.$$

En particulier, $\bar{X}_n$ converge en probabilité vers μ.

Démonstration. En utilisant l'inégalité de Tchebychev avec $g(x) = x^2$ sur la variable aléatoire $|\bar{X}_n - \mu|$, on obtient

$$P\big(|\bar{X}_n - \mu| \geq \epsilon\big) \leq \frac{E((\bar{X}_n - \mu)^2)}{\epsilon^2}.$$

Maintenant,

$$\begin{aligned} E((\bar{X}_n - \mu)^2) &= \frac{1}{n^2} E\Big(\big(\sum_{i=1}^n X_i - E(X_i)\big)^2\Big) \\ &= \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n E\big((X_i - E(X_i))(X_j - E(X_j))\big) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) = \frac{\sigma^2}{n}. \end{aligned}$$

En insérant ceci dans l'expression précédente, on obtient le résultat voulu. □

2.2.3 Inégalité de Jensen

Cette inégalité est fortement connectée à l'analyse convexe.

Lemme 2.2.4 (Inégalité de Jensen)**.** *Soit $I \subset \mathbb{R}$ un intervalle. Soit $g : I \to \mathbb{R}$ une fonction convexe. Alors, pour toute variable aléatoire X à valeur dans I,*

$$E\big(g(X)\big) \geq g\big(E(X)\big).$$

De plus, pour toute $h : I \to \mathbb{R}$ concave,

$$E\big(h(X)\big) \leq h\big(E(X)\big).$$

Démonstration. Comme g est convexe, il existe au moins un $a \in \mathbb{R}$ (une "pente" de g en $E(X)$) tel que pour tout $y \in I$,

$$g(y) \geq g(E(X)) + a(y - E(X)).$$

On peut donc appliquer ceci et la monotonicité + linéarité de l'espérance pour obtenir

$$E(g(X)) \geq E\big(g(E(X)) + a(X - E(X))\big) = g(E(X)) + a(E(X) - E(X)) = g(E(X)).$$

L'affirmation pour les fonctions concaves se déduit de celle pour les fonctions convexes : si h est concave, $-h$ est convexe. □

2.2.4 Inégalités de Cauchy-Schwarz et de Hölder

L'inégalité de Cauchy-Schwarz est une propriété classique des espaces de Hilbert. On retrouve la structure hilbertienne en remarquant que $L^2_P(\Omega)$ est un espace vectoriel réel et que

$$(X, Y) \mapsto E_P(XY),$$

est un produit scalaire sur $L^2_P(\Omega)$.

Lemme 2.2.5 (Inégalité de Cauchy-Schwarz)**.** *Pour toutes variables aléatoires X, Y,*

$$E(XY)^2 \leq E(X^2)E(Y^2).$$

Démonstration. Soient X, Y deux variables aléatoires (on peut supposer X, Y ont un second moment, sinon le côté droite de l'inégalité vaut $+\infty$). On définit $h(t) = E\big((tX + Y)^2\big) \geq 0$. En développant le carré et en utilisant la linéarité, on obtient

$$h(t) = t^2 E(X^2) + 2tE(XY) + E(Y^2),$$

qui est un polynôme de degré 2 en t. Comme ce polynôme ne change pas de signe, il a au plus une racine réelle, et donc le discriminant $4E(XY)^2 - 4E(X^2)E(Y^2)$ est négatif. Ce qui est l'inégalité voulue. □

Une application directe est que le coefficient de corrélation définit dans la section 1.7.4 est bien à valeurs dans $[-1, 1]$.

Lemme 2.2.6. *Soient X, Y deux variables aléatoires. Alors,*

$$|\mathrm{Cov}(X, Y)| \leq \sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}.$$

En particulier $\rho_{XY} \in [-1, 1]$.

Démonstration. On utilise

$$\begin{aligned}\mathrm{Cov}(X, Y)^2 &= E\big((X - E(X))(Y - E(Y))\big)^2 \\ &\leq E\big((X - E(X))^2\big)E\big((Y - E(Y))^2\big) = \mathrm{Var}(X)\mathrm{Var}(Y),\end{aligned}$$

par Cauchy-Schwarz appliqué aux variables $X - E(X)$ et $Y - E(Y)$. En prenant la racine de chaque côté, on obtient le résultat voulu. □

Une généralisation (que l'on ne prouvera pas) de cette inégalité est l'inégalité de Hölder, dont Cauchy-Schwarz est le cas $p = q = 2$.

Lemme 2.2.7 (Inégalité de Hölder)**.** *Soient $p, q \in (1, +\infty)$ tels que $\frac{1}{p} + \frac{1}{q} = 1$. Pour toutes variables aléatoires X, Y,*

$$E(|XY|) \leq E(|X|^p)^{1/p} E(|Y|^q)^{1/q}.$$

2.3 Lemmes de Borel-Cantelli

En plus des inégalités de la section précédente, deux résultats élémentaires sont particulièrement utiles en probabilité.

Lemme 2.3.1 (Lemme de Borel-Cantelli)**.** *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient $A_1, A_2, \cdots \in \mathcal{F}$. Soit $A = \bigcap_{n\geq 1}\Big(\bigcup_{k\geq n} A_k\Big)$ l'événement "une infinité des A_i sont réalisés". Alors, si $\sum_{i\geq 1} P(A_i) < \infty$, $P(A) = 0$.*

Démonstration. La suite $B_n = \bigcup_{k\geq n} A_k$ est décroissante. Donc

$$P(A) = P\Big(\bigcap_{n\geq 1} B_n\Big) = \lim_{n\to\infty} P(B_n) \leq \lim_{n\to\infty} \sum_{k\geq n} P(A_k) = 0,$$

car la somme $\sum_{k\geq 1} P(A_k) < \infty$. □

Lemme 2.3.2 (Loi 0-1 de Borel)**.** *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient $A_1, A_2, \cdots \in \mathcal{F}$ une famille* indépendante *d'événements. Soit $A = \bigcap_{n\geq 1}\Big(\bigcup_{k\geq n} A_k\Big)$ l'événement "une infinité des A_i sont réalisés". Alors, si $\sum_{i\geq 1} P(A_i) = \infty$, $P(A) = 1$.*

On appelle aussi souvent ce Lemme le *second Lemme de Borel-Cantelli.*

Démonstration. On a que $A^c = \bigcup_{n\geq 1}\Big(\bigcap_{k\geq n} A_k^c\Big)$. De plus,

$$\begin{aligned} P\Big(\bigcap_{k\geq n} A_k^c\Big) &= \lim_{N\to\infty} P\Big(\bigcap_{k=n}^{N} A_k^c\Big) \\ &= \lim_{N\to\infty} \prod_{k=n}^{N} (1 - P(A_k)) \\ &\leq \lim_{N\to\infty} \prod_{k=n}^{N} e^{-P(A_k)} = e^{-\sum_{k=n}^{\infty} P(A_k)} = 0 \end{aligned}$$

ou on a utilisé l'indépendance dans la seconde ligne, l'inégalité $1 - x \leq e^{-x}$ pour $x \geq 0$, la continuité de la fonction exponentielle et la divergence de la somme dans la troisième. On a de plus que $\bigcap_{k\geq n} A_k^c$ pour $n \geq 1$ est une suite croissante et donc

$$P(A^c) = \lim_{n\to\infty} P\Big(\bigcap_{k\geq n} A_k^c\Big) = 0.$$

□

2.4 Hiérarchie des modes de convergence

On commence par la hiérarchie entre les modes "probabiliste" de convergence

Théorème 2.4.1. *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient $X, X_1, X_2, \ldots$ une suite de variables aléatoires. Alors,*

$$X_n \xrightarrow{\text{p.s.}} X \implies X_n \xrightarrow{\text{Proba}} X \implies X_n \xrightarrow{\text{Loi}} X.$$

Démonstration. On utilise la caractérisation de la convergence presque sûre donnée dans le Lemme 2.1.2. On obtient que pour tout $\epsilon > 0$,

$$0 = \lim_{m\to\infty} P(\sup_{n\geq m} |X_n - X| \geq \epsilon) \geq \lim_{m\to\infty} P(|X_m - X| \geq \epsilon),$$

qui est la convergence en probabilité.

Si on suppose maintenant que $X_n \xrightarrow{\text{Proba}} X$. Soit $t \in \mathbb{R}$. Soit $\epsilon > 0$, pour tout $\delta > 0$ il existe $n_\delta \geq 1$ tel que $P(|X_n - X| \geq \epsilon) \leq \delta$ pour tout $n \geq n_\delta$. Comme

$$P(X_n \leq t) = P(X_n \leq t, |X_n - X| < \epsilon) + P(X_n \leq t, |X_n - X| \geq \epsilon),$$

on obtient que pour $n \geq n_\delta$,

$$P(X \leq t - \epsilon) - \delta \leq P(X_n \leq t) \leq P(X \leq t + \epsilon) + \delta.$$

En laissant $n \to \infty$ et $\delta \to 0$, on obtient que pour tout $\epsilon > 0$,

$$F_X(t - \epsilon) \leq \liminf_{n\to\infty} F_{X_n}(t) \leq \limsup_{n\to\infty} F_{X_n}(t) \leq F_X(t + \epsilon).$$

Si F_X est continue en t on a alors bien que

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

□

On remarque que ces implications ne peuvent pas être des équivalences :

Exemple 2.4.1. — Soient $Y_1, Y_2, \ldots$ une famille indépendante de variables de Bernoulli : $Y_n \sim \text{Bern}(1/n)$. On définit $X_n = nY_n$ pour tout $n \geq 1$. On a alors que la suite X_n converge en probabilité vers 0 : pour $\epsilon > 0$,

$$P(|X_n| \geq \epsilon) \leq P(X_n = n) = \frac{1}{n} \xrightarrow{n\to\infty} 0.$$

En revanche, cette suite ne converge pas presque sûrement vers 0 :

$$\sum_{n\geq 1} P(X_n = n) = \sum_{n\geq 1} \frac{1}{n} = +\infty.$$

Donc, par la loi du 0-1 de Borel (Lemme 2.3.2), on a que (comme les $X_n, n \geq 1$ forment une famille indépendante)

$$P(\limsup_{n\to\infty} X_n = +\infty) = 1,$$

car avec probabilité 1, il existe une infinité d'indices n tels que $X_n = n$.

— On peut regarder la situation $X \sim \text{Bern}(1/2)$ et $X_n = 1 - X$. On a alors que $X_n \sim \text{Bern}(1/2)$ pour tout $n \geq 1$ (en particulier, $X_n \xrightarrow{\text{Loi}} X$). Mais, $|X_n - X| = 1$ P-presque sûrement et donc X_n ne converge pas en probabilité vers X. On peut remarquer ici la différence cruciale entre le fait que la convergence en loi ne voit pas comment les variables sont corrélées entre elles sous la mesure P et le la convergence en probabilité qui elle est sensible à cette dépendance.

On continue avec la hiérarchie entre les modes "analyse fonctionelle" de convergence.

Théorème 2.4.2. *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient $X, X_1, X_2, \dots$ une suite de variables aléatoires.*

1. *Si $X, X_1, X_2, \cdots \in L_P^\infty(\Omega)$ on a que $X, X_1, X_2, \cdots \in L_P^p(\Omega)$ pour tout $p \in (0, +\infty)$ et que*
$$X_n \xrightarrow{L^\infty} X \implies X_n \xrightarrow{L^p} X.$$
2. *Soient $p \geq q \geq 1$. Si $X, X_1, X_2, \cdots \in L_P^p(\Omega)$ on a que $X, X_1, X_2, \cdots \in L_P^q(\Omega)$ et que*
$$X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{L^q} X.$$

Démonstration. 1. On remarque que si $X_n, X \in L_P^\infty(\Omega)$,
$$E(|X_n - X|^p) \leq \|X_n - X\|_{P,\infty}^p,$$
car $|Y| \leq \|Y\|_{P,\infty}$ P-presque sûrement et donc $|Y|^p \leq \|Y\|_{P,\infty}^p$ P-presque sûrement.

2. On a que pour $p > q \geq 1$,
$$E(|X_n - X|^q) = E(|X_n - X|^{pq/p}) \leq E(|X_n - X|^p)^{q/p}$$
par l'inégalité de Jensen (Lemme 2.2.4) : $q/p < 1$ et donc $x \mapsto x^{q/p}$ est concave sur $[0, +\infty)$.

□

Ici encore, les implications ne peuvent pas être transformées en équivalences.

Exemple 2.4.2. — On peut regarder une suite de variables de Bernoulli : $X_n \sim \text{Bern}(1/n)$ pour $n \geq 1$. Alors on a que pour tout $p > 0$, $E(|X_n - 0|^p) \leq 1/n$ et donc $X_n \xrightarrow{L^p} 0$. Mais, $\|X_n\|_{P,\infty} = 1$ et donc X_n ne converge pas essentiellement uniformément vers 0.

— Pour $p > q \geq 1$, il suffit de prendre une suite constante de variables aléatoires admettant un moment d'ordre q mais pas de moment d'ordre p.

On termine par les implications entre les deux types de mode de convergence.

Théorème 2.4.3. *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient $X, X_1, X_2, \dots$ une suite de variables aléatoires.*

1. *Si $X, X_1, X_2, \cdots \in L_P^p(\Omega)$ pour un $p \geq 1$ et que $X_n \xrightarrow{L^p} X$, alors $X_n \xrightarrow{\text{Proba}} X$.*
2. *Si $X, X_1, X_2, \cdots \in L_P^\infty(\Omega)$ et que $X_n \xrightarrow{L^\infty} X$, alors $X_n \xrightarrow{\text{p.s.}} X$.*

Démonstration. 1. Si $X_n \xrightarrow{L^p} X$, $X_n, n \geq 1$ et X possèdent un moment d'ordre p. On peut donc utiliser l'inégalité de Tchebychev pour obtenir
$$P(|X_n - X| \geq \epsilon) \leq \frac{E(|X_n - X|^p)}{\epsilon^p} \xrightarrow{n \to \infty} 0,$$
car $X_n \xrightarrow{L^p} X$.

2. Comme $X_n \xrightarrow{L^\infty} X$, la variable $|X_n - X|$ est essentiellement bornée et on a que pour tout $\epsilon > 0$, il existe $n_\epsilon \geq 1$ tel que $\|X_n - X\|_{P,\infty} \leq \epsilon$ pour tout $n \geq n_\epsilon$. En particulier, pour tout $n \geq n_\epsilon$,

$$P(|X_n - X| < \epsilon) = 1,$$

et donc

$$P(\sup_{n > n_\epsilon} |X_n - X| \geq \epsilon) \leq \sum_{n > n_\epsilon} P(|X_n - X| \geq \epsilon) = 0.$$

On obtient donc que pour tout $\epsilon > 0$,

$$\lim_{m \to \infty} P(\sup_{n \geq m} |X_n - X| \geq \epsilon) = 0,$$

qui est la caractérisation de la convergence presque sûre du Lemme 2.1.2.

□

Là encore, il n'est pas possible de retourner les implications.

Exemple 2.4.3. — On considère $X_1, X_2, \ldots$ comme dans le premier cas de l'exemple 2.4.1. On a $X_n \xrightarrow{\text{Proba}} 0$ mais,

$$E(|X_n|^p) = \frac{n^p}{n} = n^{p-1} \xrightarrow{n \to \infty} \infty,$$

si $p > 1$.

— Soit $Y \sim \text{Uni}([0,1])$ et $X_n = \mathbb{1}_{[0,1/n]}(Y)$. Alors $X_n \xrightarrow{\text{p.s.}} 0$, mais $\|X_n\|_{P,\infty} = 1$ pour tout n donc X_n ne converge pas essentiellement uniformément vers 0.

En revanche, il existe des conditions additionnelles (pouvant être très restrictives...) pour rendre vrai les implications inverses dans chaque cas.

Le Théorème suivant regroupe certaines de ces "implications inverses".

Théorème 2.4.4. *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient $X, X_1, X_2, \ldots$ une suite de variables aléatoires. Alors,*

1. *si $X_n \xrightarrow{\text{Loi}} c \in \mathbb{R}$, alors $X_n \xrightarrow{\text{Proba}} c$;*
2. *si $X_n \xrightarrow{\text{Proba}} X$ et qu'il existe $C \geq 0$ tel que $P(|X_n| \leq C) = 1$ pour tout $n \geq 1$, alors $X_n \xrightarrow{L^p} X$ pour tout $p \geq 1$;*
3. *si $\sum_{n \geq 1} P(|X_n - X| > \epsilon) < \infty$ pour tout $\epsilon > 0$, alors $X_n \xrightarrow{\text{p.s.}} X$.*

Démonstration. On procède dans l'ordre.

1. Soit $\epsilon > 0$. Alors,

$$P(|X_n - c| \geq \epsilon) = P(X_n \in \mathbb{R} \setminus (c - \epsilon, c + \epsilon)).$$

Mais cette dernière probabilité doit tendre vers 0 quand $n \to \infty$ si $X_n \xrightarrow{\text{Loi}} c$.

2. On commence par remarquer que comme $X_n \xrightarrow{\text{Proba}} X$, on a en particulier que $X_n \xrightarrow{\text{Loi}} X$ et donc que

$$P(-C \leq X \leq C) = \lim_{n\to\infty} P(-C \leq X_n \leq C) = 1,$$

par hypothèse. On introduit l'événement $A_{\epsilon,n} = \{\omega : \ |X_n(\omega) - X(\omega)| > \epsilon\}$. On a alors que pour P-presque tout $\omega \in \Omega$,

$$|X_n(\omega) - X(\omega)|^p \leq \epsilon^p \mathbb{1}_{A^c_{\epsilon,n}}(\omega) + \mathbb{1}_{A_{\epsilon,n}}(\omega)(2C)^p,$$

car $|X_n| + |X| \leq 2C$ P-presque sûrement. En prenant l'espérance de chaque côtés, on obtient

$$E(|X_n(\omega) - X(\omega)|^p) \leq \epsilon^p + (2C)^p P(A_{\epsilon,n}).$$

En prenant $n \to \infty$, on obtient $\lim_{n\to\infty} E(|X_n(\omega) - X(\omega)|^p) \leq \epsilon^p$. ϵ étant arbitraire, on a le résultat voulu.

3. On introduit $A_m = \bigcup_{k\geq 1} \bigcap_{n\geq k} \{\omega : \ |X_n(\omega) - X(\omega)| \leq m^{-1}\}$. On a alors que, par convergence monotone des probabilité,

$$P(\lim_{n\to\infty} X_n = X) = P\Big(\bigcap_{m\geq 1} A_m\Big) = \lim_{m\to\infty} P(A_m).$$

On a de plus que $A^c_m = \bigcap_{k\geq 1} \bigcup_{n\geq k} \{\omega : \ |X_n(\omega) - X(\omega)| > m^{-1}\}$. Par notre hypothèse, on a que

$$\sum_{n\geq 1} P(|X_n - X| > m^{-1}) < \infty.$$

Donc, par le Lemme de Borel-Cantelli (Lemme 2.3.1), on a que $P(A^c_m) = 0$. En injectant ceci dans la première équation, on obtient le résultat voulu.

□

2.4.1 Diagramme récapitulatif

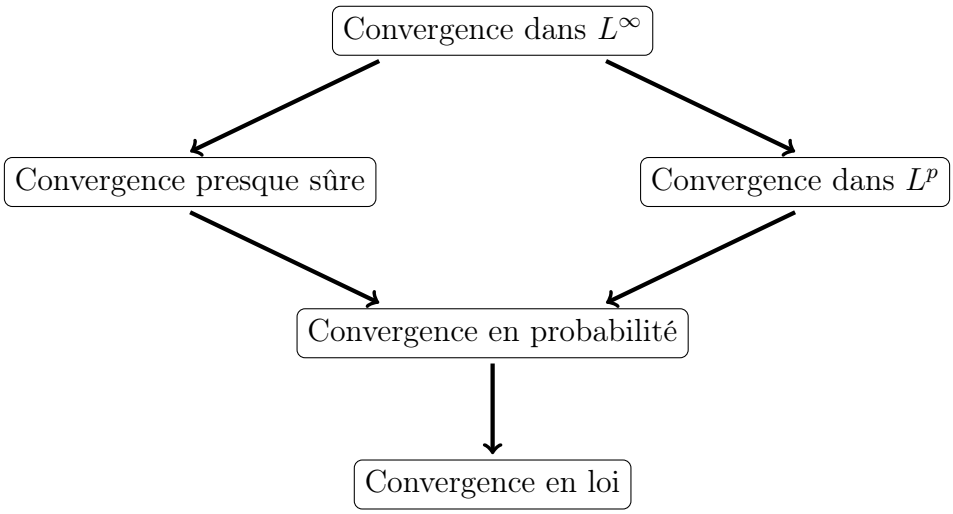

FIGURE 2.1 – Implications entre les différents modes de convergence.

2.5 Théorèmes limites

2.5.1 Cadre

Dans cette section, on travaillera avec un espace de probabilité $(\Omega, \mathcal{F}, P)$ fixé, que l'on supposera être capable de supporter toutes les variables aléatoires que l'on veut définir dessus [4]. $X_1, X_2, \cdots \in L^1_P(\Omega)$ dénotera une famille indépendante identiquement distribuée de variables aléatoires avec

$$E(X_1) = 0.$$

On définira les sommes partielles

$$S_n = \sum_{k=1}^{n} X_k.$$

Le but des Théorèmes limites que l'on va voir est d'étudier le comportement de S_n quand n devient très grand. On verra deux résultat principaux : la Loi des Grands Nombres qui dit comment S_n se comporte *en moyenne* : $S_n = o(n)$ avec grande probabilité ; et le Théorème Central Limite qui dit comment S_n fluctue autour de sa valeur moyenne : $S_n = \Theta(\sqrt{n})$ avec grande probabilité.

2.5.2 Loi forte des grands nombres

Le premier résultat est que la moyenne empirique des variables converge (dans un sens fort) vers l'espérance.

Théorème 2.5.1 (Loi forte des grands nombres)**.**

$$\frac{1}{n} S_n \xrightarrow{\text{p.s.}} 0.$$

Remarque 2.5.1. *On pourrait se dire que ce résultat est plus faible que ce que l'on voudrait : on a supposé que les variables étaient d'espérance* 0*. Si* $Y_1, Y_2, \ldots$ *sont i.i.d. mais n'ont pas espérance* 0*, on considère la suite* $X_i = Y_i - E(Y_i)$ *qui elle a* $E(X_i) = 0$*. On applique le Théorème à la suite* X_i *pour obtenir*

$$\frac{1}{n} \sum_{k=1}^{n} (Y_i - E(Y_i)) \xrightarrow{\text{p.s.}} 0.$$

Mais,

$$\frac{1}{n} \sum_{k=1}^{n} (Y_i - E(Y_i)) = \frac{1}{n} \sum_{k=1}^{n} Y_i - E(Y_1),$$

ce qui implique que

$$P\Big(\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} Y_i = E(Y_1) \Big) = P\Big(\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (Y_i - E(Y_i)) = 0 \Big) = 1.$$

4. Notez que cela est rendu possible par le Corollaire 1.7.7.

Le résultat optimal demandant passablement de travail, on va se contenter d'un résultat légèrement plus faible, que l'on a déjà essentiellement prouvé.

Théorème 2.5.2. *Si* $E(X_1^4) < \infty$,

$$\frac{1}{n} S_n \xrightarrow{\text{p.s.}} 0.$$

Démonstration. On a montré dans le Théorème 2.2.3 que la suite $\frac{1}{n}S_n$ convergeait en probabilité vers 0 sous l'hypothèse d'un second moment. On va utiliser le moment d'ordre 4 pour obtenir une borne d'une puissance plus élevée en n^{-1} et déduire la convergence presque-sûre via le Théorème 2.4.4.

On déduit de $E(X_1^4) < \infty$ et de l'inégalité de Tchebychev, Lemme 2.2.2, que

$$P(|\frac{1}{n} S_n| \geq \epsilon) \leq \frac{E(S_n^4)}{n^4 \epsilon^4}.$$

On calcule alors

$$E(S_n^4) = \sum_{i,j,k,l=1}^{n} E(X_i X_j X_k X_l) = \sum_{i=1}^{n} E(X_i^4) + \sum_{i,j=1}^{n} E(X_i^2) E(X_j^2)$$
$$= nE(X_1^4) + n^2 E(X_1^2)^2,$$

car $E(X_i X_j X_k X_l) = 0$ si un des indices n'apparaît qu'une fois par indépendance et $E(X_i) = 0$. Comme $E(X_1^2)^2 \leq E(X_1^4)$ par l'inégalité de Jensen (voir Lemme 2.2.4), on obtient que $E(S_n^4) \leq 2n^2 E(X_1^4)$ et donc que

$$P(|\frac{1}{n} S_n| \geq \epsilon) \leq \frac{2E(X_1^4)}{n^2 \epsilon^4}.$$

Cette dernière quantité est sommable en n pour tout $\epsilon > 0$, on peut donc appliquer le Théorème 2.4.4 pour conclure. □

2.5.3 Théorème central limite

Théorème 2.5.3 (Théorème central limite)**.** *Soit* $Z \sim \mathcal{N}(0,1)$ *une variable gaussienne centrée réduite. Alors, si* $E(X_1^2) = 1$,

$$\frac{1}{\sqrt{n}} S_n \xrightarrow{\text{Loi}} Z.$$

Remarque 2.5.2. *Comme dans le cas de la loi des grands nombres, on peut déduire que si* $E(Y_i) = \mu$ *et* $\mathrm{Var}(Y_i) = \sigma^2$, *on a*

$$\frac{1}{\sqrt{n}} \Big(\sum_{k=1}^{n} Y_k - n\mu \Big) \xrightarrow{\text{Loi}} \mathcal{N}(0, \sigma^2),$$

en appliquant le Théorème à la suite $X_i = \frac{Y_i - \mu}{\sigma}$.

Comme pour la loi des grands nombres, la version optimale du Théorème central limite demande plus de travail (ou plus d'outils). On se contentera de prouver une version avec des hypothèses plus restrictives. Notez que le Théorème 2.5.4 n'est pas plus faible que le Théorème 2.5.3 : il contient une estimée quantitative [5] qui n'est pas vraie dans le cadre général du Théorème 2.5.3.

Théorème 2.5.4. *Soit $Z \sim \mathcal{N}(0,1)$ une variable gaussienne centrée réduite. Il existe deux constantes $c_1, c_2 > 0$ universelles telles que, si $E(X_1^2) = 1$ et $E(|X_1|^3) < \infty$, alors*

$$\sup_{t\in\mathbb{R}} \Big| P\Big(\frac{1}{\sqrt{n}} S_n \leq t\Big) - F_Z(t)\Big| \leq \frac{c_1 E(|X_1|^3) + c_2}{n^{1/8}}.$$

En particulier, $\frac{1}{\sqrt{n}} S_n \xrightarrow{\text{Loi}} Z$.

Démonstration. La preuve utilise la méthode de Lindeberg [6] : l'idée est d'utiliser le fait que les sommes de gaussiennes sont naturellement des gaussiennes et de remplacer une à une les variables $X_1, \ldots, X_n$ par des gaussiennes en contrôlant l'erreur à chaque étape.

Soient $Z_1, Z_2, \ldots$ une suite i.i.d. de variables aléatoires gaussiennes : $Z_1 \sim \mathcal{N}(0,1)$, indépendante des X_is. On introduit

$$\tilde{S}_n = \frac{1}{\sqrt{n}} S_n, \quad T_{n,k} = \frac{1}{\sqrt{n}} \Big(\sum_{i=1}^{k} Z_i + \sum_{i=k+1}^{n} X_i \Big).$$

Par le Lemme 1.8.7, $T_{n,n} \sim \mathcal{N}(0,1)$ pour tout $n \geq 1$. De plus, $T_{n,0} = \tilde{S}_n$. Soit $h : \mathbb{R} \to \mathbb{R}$ une fonction décroissante, trois fois continûment différentiable, et telle que $h(x) = 1$ quand $x \leq 0$ et $h(x) = 0$ quand $x \geq 1$ [7]. On introduit alors, pour $\delta \in [0,1]$ et $t \in \mathbb{R}$,

$$h_\delta^+(x) = h\Big(\frac{x-t}{\delta}\Big),$$
$$h_\delta^-(x) = h\Big(1 + \frac{x-t}{\delta}\Big).$$

Ces fonctions sont des approximations lisses de la fonction $\mathbb{1}_{(-\infty,t]}$ et satisfont $h_\delta^+ \geq \mathbb{1}_{(-\infty,t]} \geq h_\delta^-$ et

$$\begin{aligned} |h_\delta^+(x) - \mathbb{1}_{(-\infty,t]}(x)| &\leq \mathbb{1}_{[t,t+\delta]}(x), \\ |h_\delta^+(x) - \mathbb{1}_{(-\infty,t]}(x)| &\leq \mathbb{1}_{[t-\delta,t]}(x). \end{aligned} \tag{2.4}$$

5. La borne sur la différence entre les fonctions de répartition dans le Théorème 2.5.4 est une "version faible" de l'*inégalité de Berry-Esséen*, qui donne une borne par $\frac{E(|X_1|^3)}{\sqrt{n}}$. Cette dernière est souvent utilisée dans les applications du TCL car elle quantifie l'erreur commise en fonction de n.

6. Jarl Waldemar Lindeberg, 1876-1932, mathématicien finlandais

7. Une telle fonction existe : on peut par exemple prendre

$$\varphi(x) = \frac{e^{-\frac{1}{x(1-x)}}}{\int_0^1 e^{-\frac{1}{s(1-s))}} ds} \mathbb{1}_{[0,1]}(x), \quad h(x) = \int_{-\infty}^{\infty} \varphi(y) \mathbb{1}_{(-\infty,0]}(x-y) dy,$$

qui est même C^∞.

En particulier, elles permettent d'approcher $F_{\tilde{S}_n}(t)$:

$$P(\tilde{S}_n \leq t) = E\big(\mathbb{1}_{(-\infty,t]}(\tilde{S}_n)\big) \begin{cases} \leq E\big(h_\delta^+(\tilde{S}_n)\big) \\ \geq E\big(h_\delta^-(\tilde{S}_n)\big) \end{cases} .$$

Comme $T_{n,n} \sim \mathcal{N}(0,1)$ pour tout n, on a

$$F_Z(t) = P(Z \leq t) = P(T_{n,n} \leq t) = E\big(\mathbb{1}_{(-\infty,t]}(T_{n,n})\big).$$

On a alors la borne

$$P(\tilde{S}_n \leq t) - F_Z(t) \leq \underbrace{E\big(h_\delta^+(\tilde{S}_n)\big) - E\big(h_\delta^+(T_{n,n})\big)}_{I} + \underbrace{E\big(h_\delta^+(T_{n,n})\big) - E\big(\mathbb{1}_{(-\infty,t]}(T_{n,n})\big)}_{II}.$$

Par (2.4) et le fait que $T_{n,n} \sim \mathcal{N}(0,1)$, on a

$$II \leq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dx e^{-x^2/2} |h_\delta^+(x) - \mathbb{1}_{(-\infty,t]}(x)| \leq \frac{1}{\sqrt{2\pi}} \int_t^{t+\delta} dx e^{-x^2/2} \leq \frac{\delta}{\sqrt{2\pi}}. \quad (2.5)$$

On veut maintenant borner I. On procède en télescopant la somme :

$$I = E\big(h_\delta^+(T_{n,0})\big) - E\big(h_\delta^+(T_{n,n})\big) = \sum_{k=0}^{n-1} \big[E\big(h_\delta^+(T_{n,k})\big) - E\big(h_\delta^+(T_{n,k+1})\big)\big]. \quad (2.6)$$

Pour estimer le terme numéro k dans la somme, on introduit

$$W_k = \frac{1}{\sqrt{n}}\Big(\sum_{i=1}^{k} Z_i + \sum_{i=k+2}^{n} X_i\Big).$$

On a alors

1. $T_{n,k} = W_k + \frac{1}{\sqrt{n}} X_{k+1}$ et $T_{n,k+1} = W_k + \frac{1}{\sqrt{n}} Z_{k+1}$;
2. W_k, X_{k+1}, Z_{k+1} forment une famille indépendante.

On développe alors h_δ^+ en série de Taylor autour de W_k [8] pour obtenir pour tout $z \in \mathbb{R}$,

$$h_\delta(W_k + z) = h_\delta(W_k) + h'_\delta(W_k) z + \frac{1}{2} h''_\delta(W_k) z^2 + \frac{1}{6} h'''_\delta(Y) z^3$$

où $Y \in [W_k, W_k + z]$ et on a noté $h_\delta \equiv h_\delta^+$. En évaluant ce polynôme de Taylor en $\frac{Z_{k+1}}{\sqrt{n}}$ et en $\frac{X_{k+1}}{\sqrt{n}}$, on obtient

$$\begin{aligned} h_\delta(T_{n,k}) - h_\delta(T_{n,k+1}) = h'_\delta(W_k) \frac{X_{k+1} - Z_{k+1}}{\sqrt{n}} + \frac{1}{2} h''_\delta(W_k) \frac{X_{k+1}^2 - Z_{k+1}^2}{n} \\ + \frac{1}{6} h'''_\delta(Y) \frac{X_{k+1}^3}{n^{3/2}} - \frac{1}{6} h'''_\delta(Y') \frac{Z_{k+1}^3}{n^{3/2}} \end{aligned} \quad (2.7)$$

8. Pour être précis : toutes nos variables vivent sur un espace de probabilité $(\Omega, \mathcal{F}, P)$. Pour chaque $\omega \in \Omega$, on a alors que $W_k(\omega) \in \mathbb{R}$. Le développement de Taylor se fait autour de ce nombre, et est donc un "polynôme aléatoire". On évalue ensuite ce polynôme en $\frac{Z_{k+1}(\omega)}{\sqrt{n}}$ et en $\frac{X_{k+1}(\omega)}{\sqrt{n}}$ (composition de fonctions).

où $Y \in [W_k, n^{-1/2}X_{k+1}]$ et $Y' \in [W_k, n^{-1/2}Z_{k+1}]$. En prenant l'espérance de chaque côté et en utilisant $E(X_{k+1}^2) = 1 = E(Z_{k+1}^2)$ et $E(X_{k+1}) = 0 = E(Z_{k+1})$ on obtient

$$E\big(h_\delta(T_{n,k})\big) - E\big(h_\delta(T_{n,k+1})\big) = \frac{1}{6n^{3/2}} E\big(h_\delta'''(Y)X_{k+1}^3 - h_\delta'''(Y')Z_{k+1}^3\big).$$

En introduisant

$$a = \sup_{y\in\mathbb{R}} |h'''(y)| = \delta^3 \sup_{y\in\mathbb{R}} |h_\delta'''(y)| < \infty,$$

on a que

$$\begin{aligned} E\big(h_\delta(T_{n,k})\big) - E\big(h_\delta(T_{n,k+1})\big) &\leq \frac{a}{6n^{3/2}\delta^3} E\big(|X_{k+1}|^3 + |Z_{k+1}|^3\big) \\ &\leq \frac{a}{6n^{3/2}\delta^3}\big(E(|X_1|^3) + 3\big). \end{aligned}$$

car $E(|Z_{k+1}|^3) \leq 3$ [9]. On peut alors injecter cette borne dans (2.6) pour obtenir

$$I \leq \frac{a}{6n^{1/2}\delta^3}\big(E(|X_1|^3) + 3\big). \tag{2.8}$$

Les bornes (2.5) et (2.8) sont valides pour tout $\delta \in (0,1)$. On prend alors $\delta = n^{-1/8}$ et on utilise (2.5) et (2.8) pour obtenir

$$P(\tilde{S}_n \leq t) - F_Z(t) \leq \frac{1}{n^{1/8}}\Big(\frac{1}{\sqrt{2\pi}} + \frac{a}{6}\big(E(|X_1|^3) + 3\big)\Big).$$

On procède de la même manière en remplaçant h_δ^+ par h_δ^- pour obtenir la borne opposée et le résultat voulu. □

Remarque 2.5.3. *On peut remarque que la preuve ci-dessus ne dépend pas de la loi de chacun des $X_1, X_2, \ldots$: la même preuve marche si on remplace les hypothèses par*

1. *$E(X_i) = 0$ et $E(X_i^2) = 1$ pour tout $i \geq 1$;*
2. *$\sup_{i\geq 1} E(|X_i|^3) < +\infty$;*
3. *$X_1, X_2, \ldots$ forment une famille indépendante.*

9. En effet : par l'inégalité de Jensen et si $Z \sim \mathcal{N}(0,1)$, $E(|Z|^3) = E(|Z|^{4\frac{3}{4}}) \leq E(Z^4)^{3/4} = 3^{3/4} \leq 3$

Chapitre 3

Quelques modèles classiques

Dans ce chapitre on va voir quelques modèles classiques/sujets étudiés en probabilités et utiliser le formalisme et les outils développé dans le cours pour dire quelque chose sur ces modèles. Dans chaque case, on se donne un espace de probabilité $(\Omega, \mathcal{F}, P)$ sur lequel toutes nos variables seront définies. Un exercice est de vérifier qu'un espace supportant toutes les variables aléatoires que l'on considérera dans chaque cas existe bel et bien.

3.1 Marche aléatoire sur $\mathbb{Z}$

On a déjà rencontré ce modèle plusieurs fois. On va en voir une généralisation. On travaillera sur un espace de probabilité $(\Omega, \mathcal{F}, P)$ que l'on suppose assez grand pour que l'on puisse définir toutes les variables aléatoires que l'on utilisera dessus.

3.1.1 La marche biaisée sur $\mathbb{Z}$

Soit $p \in (0,1)$. On considère $X_1', X_2', \dots$ une suite i.i.d. de variables aléatoires de Bernoulli, $X_i \sim \text{Bern}(p)$. Soit S_0 une variable aléatoire à valeurs dans $\mathbb{Z}$ telle que $P(S_0 = n) > 0$ pour tout $n \in \mathbb{Z}$, indépendante des X_is. On définit alors

$$X_k = 2X_k' - 1,\ k \geq 1, \quad S_n = S_0 + \sum_{k=1}^{n} X_k.$$

On notera P_s la loi P conditionnellement à l'événement $\{S_0 = s\}$.

Le comportement de la marche change avec le paramètre p comme sur l'image 3.1.

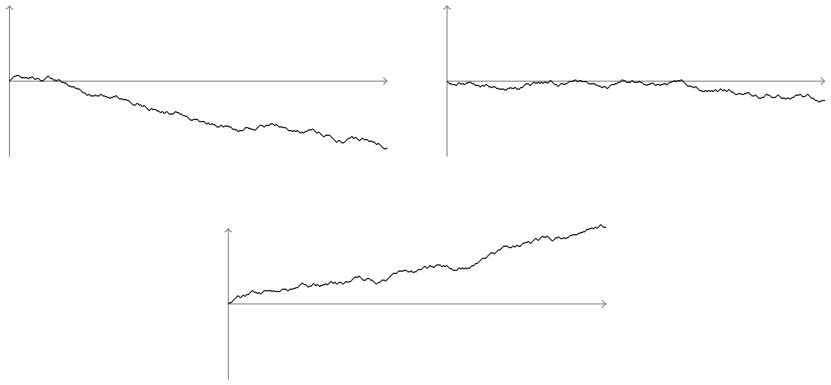

FIGURE 3.1 – Marche aléatoire biaisée jusqu'au temps 500. Haut-gauche : $p = 0.4$, Haut-droite : $p = 0.5$, bas : $p = 0.6$.

Pour $1 \leq i \leq j$, on notera

$$Z_{i,j} = \sum_{k=i}^{j} X_k.$$

Quelques observations :

1. Comme les X_i sont i.i.d., les vecteurs

   $$(X_m, X_{m+1}, \ldots, X_{m+n-1}) \text{ et } (X_1, X_2, \ldots, X_n)$$

   ont la même loi pour tout $n, m \geq 1$.
2. Par le point précédent, les vecteurs

   $$(Z_{m,m}, Z_{m,m+1}, \ldots, Z_{m,m+n-1}) \text{ et } (Z_{1,1}, Z_{1,2}, \ldots, Z_{1,n})$$

   ont la même loi pour tout $n, m \geq 1$.
3. De la même manière, si $n_1, \ldots, n_m \geq 1$, les vecteurs

   $$(X_1, \ldots, X_{n_1}),\ (X_{n_1+1}, \ldots, X_{n_1+n_2}),\ \ldots$$

   forment une famille de vecteurs aléatoires indépendants.

Des points précédents, il suit (vu en exercice) que

$$P_s\Big(\bigcap_{k=1}^{m+n}\{S_k = s_k\} \,\Big|\, \bigcap_{k=1}^{m}\{S_k = s_k\}\Big) = P_{s_m}\Big(\bigcap_{k=1}^{n}\{S_k = s_{k+m}\}\Big). \tag{3.1}$$

On définit alors

$$\tau_a = \inf\{n \geq 0 : S_n = a\}, \quad \tau_a^+ = \inf\{n \geq 1 : S_n = a\}$$

avec $\inf \varnothing = \infty$ par convention. On remarque que les événements $\{\tau_a = t\}$ sont bien des ensembles mesurables car

$$\{\tau_a = t\} = \{S_t = a\} \cap \bigcap_{k=0}^{t-1}\{S_k \neq a\}.$$

On a aussi que $\{\tau_a = \infty\}$ est mesurable car

$$\{\tau_a = \infty\} = \{\tau_a < \infty\}^c = \Big(\bigcup_{t\geq 0}\{\tau_a = t\}\Big)^c.$$

La même chose s'applique pour τ_a^+.

On va utiliser plusieurs fois une observation "élémentaire" qui suit d'un principe très important dans l'étude des marches aléatoires et, plus tard du mouvement brownien : le *principe de réflexion.*

Lemme 3.1.1. *Soit $n, b \in \mathbb{Z}$. Alors, pour tout $p \in (0,1)$,*

$$P_0(\tau_0 > n, S_n = b) = \frac{|b|}{n} P_0(S_n = b).$$

Démonstration. On remarque que l'identité est triviale si b est tel que $P_0(S_n = b) = 0$ (soit car la parité de b et n est différente soit car $|b|$ est trop grand). On regarde donc les cas pour lesquels $P_0(S_n = b) > 0$. Le résultat est alors équivalent à

$$P_0(\tau_0 > n \,|\, S_n = b) = \frac{|b|}{n}.$$

On va se ramener à un problème combinatoire : on sait que sous $P_0(\ \,|\, S_n = b)$ les trajectoires $(s_0, \dots, s_n)$ avant probabilité positives sont les trajectoires telles que

$$s_0 = 0,\ s_n = b,\ |s_k - s_{k-1}| = 1\ \forall k = 1, \dots, n,$$

notons $\mathcal{T}_{a,b}(n)$ l'ensemble des trajectoires $(s_0, \dots, s_n)$ avec $s_0 = a, s_n = b$ et $|s_k - s_{k-1}| = 1\ \forall k = 1, \dots, n$. On a alors

$$P_0(S_n = b) = \sum_{(s_0,\dots,s_n)\in\mathcal{T}_{0,b}(n)} \prod_{i=1}^{n} P(X_i = s_i - s_{i-1}) = \binom{n}{\frac{n+b}{2}} p^{\frac{n+b}{2}} (1-p)^{\frac{n-b}{2}}$$

car toutes les trajectoires dans $\mathcal{T}_{0,b}(n)$ font exactement $\frac{n+b}{2}$ pas vers le haut et $\frac{n-b}{2}$ pas vers le bas. On obtient alors que la probabilité d'une trajectoire dans $\mathcal{T}_{0,b}(n)$ sous $P_0(\ \,|\, S_n = b)$ est donnée par

$$P_0(S_0 = s_0, \dots, S_n = s_n \,|\, S_n = b) = \frac{p^{\frac{n+b}{2}}(1-p)^{\frac{n-b}{2}}}{\binom{n}{\frac{n+b}{2}} p^{\frac{n+b}{2}}(1-p)^{\frac{n-b}{2}}} = \frac{1}{|\mathcal{T}_{0,b}(n)|}.$$

On a donc que la loi de la trajectoire de la marche sous $P_0(\ \,|\, S_n = b)$ est juste la loi uniforme sur $\mathcal{T}_{0,b}(n)$. Le résultat est alors réduit à montrer que le nombre de trajectoires dans $\mathcal{T}_{0,b}(n)$ qui restent strictement au dessus de 0 du temps 1 au temps $n-1$ est égale à $\frac{|b|}{n}|\mathcal{T}_{0,b}(n)|$. Par symétrie, on traitera seulement le cas $b > 0$. On note

- $\mathcal{T}^+_{a,b}(n)$ l'ensemble des trajectoires dans $\mathcal{T}_{a,b}(n)$ qui restent strictement au dessus de 0 entre le temps 1 et le temps n ;
- $\mathcal{T}^{\pm}_{a,b}(n)$ l'ensemble des trajectoires dans $\mathcal{T}_{a,b}(n)$ qui reviennent en 0 au moins une fois entre le temps 1 et le temps n.

On remarque ensuite que les trajectoires dans $\mathcal{T}^+_{0,b}(n)$ ont toutes $s_1 = 1$ (sinon la trajectoire doit passer par 0 pour atteindre $b > 0$). En particulier,

$$|\mathcal{T}^+_{0,b}(n)| = |\mathcal{T}^+_{1,b}(n-1)|.$$

On va utiliser le fait que

$$|\mathcal{T}^+_{1,b}(n-1)| = |\mathcal{T}_{1,b}(n-1)| - |\mathcal{T}^\pm_{1,b}(n-1)|,$$

et le fait que l'on connaît

$$|\mathcal{T}_{1,b}(n-1)| = |\mathcal{T}_{0,b-1}(n-1)| = \binom{n-1}{\frac{n+b}{2}-1} = \frac{n+b}{2n}\binom{n}{\frac{n+b}{2}}.$$

On utilise alors le principe de réflexion : l'ensemble $\mathcal{T}^\pm_{1,b}(n-1)$ est en bijection avec l'ensemble $\mathcal{T}_{-1,b}(n-1)$ de la manière suivante : on réfléchit à travers la hauteur 0 la portion de trajectoire entre le temps 0 et le premier temps auquel la marche atteint 0 (voir l'image 3.2).

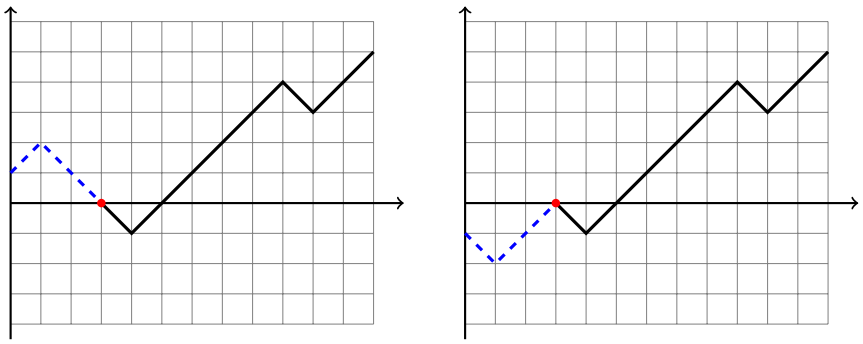

FIGURE 3.2 – Principe de réflexion : bijection entre $\mathcal{T}^\pm_{1,b}(n-1)$ et $\mathcal{T}_{-1,b}(n-1)$.

Cela nous donne

$$|\mathcal{T}^\pm_{1,b}(n-1)| = |\mathcal{T}_{-1,b}(n-1)| = |\mathcal{T}_{0,b+1}(n-1)| = \binom{n-1}{\frac{n+b}{2}} = \frac{n-b}{2n}\binom{n}{\frac{n+b}{2}}.$$

On combine alors les deux quantités pour obtenir

$$|\mathcal{T}^+_{0,b}(n)| = |\mathcal{T}^+_{1,b}(n-1)| = \frac{n+b}{2n}\binom{n}{\frac{n+b}{2}} - \frac{n-b}{2n}\binom{n}{\frac{n+b}{2}} = \frac{b}{n}|\mathcal{T}_{0,b}(n)|,$$

qui est le résultat voulu. □

Une application directe de ce résultat est l'égalité

$$P_0(\tau_0^+ > n) = \sum_{b\in\mathbb{Z}} P_0(\tau_0 > n, S_n = b) = \frac{1}{n}\sum_{b\in\mathbb{Z}} |b| P_0(S_n = b) = \frac{1}{n}E_0(|S_n|). \tag{3.2}$$

3.1.2 Récurrence de la marche aléatoire simple symétrique

On commence par montrer que la marche simple symétrique ($p = 1/2$) est *récurrente* :

$$P_a(\tau_a^+ < \infty) = 1.$$

En mots : on revient presque sûrement à son point de départ.

Théorème 3.1.2. *Si $p = 1/2$, alors*

$$P_0(\tau_0^+ < \infty) = 1.$$

Démonstration. Le point de départ est l'égalité (3.2) :

$$P_0(\tau_0 > n) = \frac{1}{n} E_0(|S_n|).$$

on va alors borner $E_0(|S_n|)$: par l'inégalité de Jensen,

$$E_0(|S_n|) \leq E_0(|S_n|^2)^{1/2} = \Big(\sum_{i,j=1}^{n} E(X_i X_j) \Big)^{1/2} = \sqrt{n}.$$

car les X_i sont indépendants et $E(X_i) = 0$, $E(X_i^2) = 1$. Donc,

$$P_0(\tau_0^+ > n) \leq \frac{1}{\sqrt{n}},$$

ce qui implique

$$P_0(\tau_0^+ = \infty) = \lim_{n \to \infty} P_0(\tau_0^+ > n) = 0.$$

□

3.1.3 Transience de la marche biaisé avec $p \neq 1/2$

On va maintenant montrer que le contraire est vrai pour la marche biaisée.

Théorème 3.1.3. *Si $p \neq 1/2$, alors*

$$P_0(\tau_0^+ = \infty) > 0.$$

Ce phénomène de "ne jamais revenir avec probabilité positive" est appelé *transience* de la marche. Une manière de voir pourquoi ce phénomène se produit est de remarquer que

- $E_0(S_n) = (2p - 1)n$ croit linéairement avec n ;
- $\mathrm{Var}_0(S_n) = n$ et donc la "déviation typique" de S_n loin de $E(S_n)$ est d'ordre $\sqrt{n}$ qui est beaucoup plus petit que $E_0(S_n)$. On a donc que la marche au temps n sera typiquement très loin de son point de départ.

Démonstration. Notez qu'il suffit de montrer le résultat pour $p > 1/2$ par symétrie. On suppose donc $p > 1/2$. On démarre encore avec l'égalité (3.2) :

$$P_0(\tau_0^+ > n) = \frac{1}{n} E_0(|S_n|).$$

On utilise ensuite la borne

$$E_0(|S_n|) \geq E_0(S_n) = \sum_{i=1}^{n} E(X_i) = n(2p-1).$$

Cela nous donne que pour tout $n \geq 1$, $P_0(\tau_0^+ > n) \geq 2p - 1$ et donc

$$P_0(\tau_0^+ = \infty) = \lim_{n\to\infty} P_0(\tau_0^+ > n) \geq 2p - 1 > 0.$$

□

3.2 Percolation

Le modèle de *percolation par arête* sur un graphe $G = (V, E)$ (avec un nombre fini ou dénombrable de sommets et d'arêtes) est une mesure de probabilité sur $\Omega = \mathcal{P}(E)$: pour chaque arête $e \in E$, on garde cette arête avec probabilité $p \in [0, 1]$ et on l'enlève avec probabilité $1 - p$, indépendamment des autres. On formalise ceci comme suit :

— On se donne $X_e, e \in E$ une famille i.i.d. de variables aléatoires de Bernoulli de paramètre p : $X_e \sim \text{Bern}(p)$.

— On pose $X = (X_e, e \in E)$ le processus stochastique indexé par E contenant tout les X_e. On regarde alors l'ensemble d'arêtes aléatoire

$$E_X = \{e \in E : \ X_e = 1\}.$$

L'étude des modèles de percolation porte sur l'étude des propriétés de connectivité du graphe aléatoire (V, E_X). On va s'intéresser à un cas en particulier : la percolation sur le graphe donné par $V = \mathbb{Z}^2$ et

$$E = \{\{i, j\} \subset \mathbb{Z}^2 : \ \|i - j\|_1 = 1\},$$

où $\|x\|_1 = |x_1| + |x_2|$, $x = (x_1, x_2) \in \mathbb{Z}^2$. Pour $F \subset E$ fini, on notera $\{x \overset{F}{\leftrightarrow} y\}$ pour l'événement

$$\{\exists n \geq 1 \text{ et } (x_0 = x, x_1, \ldots, x_n = y) : \ \{x_k, x_{k-1}\} \in E, \ X_{\{x_k, x_{k-1}\}} = 1 \ \forall k = 1, \ldots, n\}$$

"il existe un chemin d'arêtes ouvertes contenu dans F entre x et y". On définira de la même manière

$$\{A \overset{F}{\leftrightarrow} B\} = \bigcup_{x\in A, y\in B} \{x \overset{F}{\leftrightarrow} y\},$$

"l'ensemble A est connecté à l'ensemble B par un chemin ouvert dans F". Notez que ces événements ne dépendent que des X_e avec $e \in F$, comme F est fini, ces événements sont bien dans la tribu cylindrique.

On notera P_p la mesure de percolation sur $(\mathbb{Z}^2, E)$ de paramètre p (la loi du processus stochastique $(X_e)_{e\in E}$). On définit $\Lambda_n = \{-n, \dots, n\}^2$ et $\partial\Lambda_n = \{x \in \Lambda_n : \ \|x\|_\infty = n\}$ (où $\|x\|_\infty = \max(|x_1, x_2|)$). On notera aussi

$$E_n = \{\{x, y\} \in E : \ x, y \in \Lambda_n\}.$$

Le but de cette section sera de montrer l'existence d'une transition de percolation au sens suivant :

Théorème 3.2.1. *Il existe $p_c \in (0, 1)$ tel que*

— *pour $p < p_c$,*

$$P_p(0 \overset{E_n}{\longleftrightarrow} \partial\Lambda_n) \xrightarrow{n\to\infty} 0;$$

— *pour $p > p_c$,*

$$\inf_{n\geq 1} P_p(0 \overset{E_n}{\longleftrightarrow} \partial\Lambda_n) > 0.$$

On va procéder en trois étapes :

1. on montre que pour p assez petit, $P_p(0 \overset{E_n}{\longleftrightarrow} \partial\Lambda_n) \xrightarrow{n\to\infty} 0$;
2. on montre que pour p assez grand, $\inf_{n\geq 1} P_p(0 \overset{E_n}{\longleftrightarrow} \partial\Lambda_n) > 0$;
3. on montre que $P_p(0 \overset{E_n}{\longleftrightarrow} \partial\Lambda_n)$ est monotone en p.

Pour simplifier les notations, on notera

$$\theta_n(p) = P_p(0 \overset{E_n}{\longleftrightarrow} \partial\Lambda_n).$$

3.2.1 Pas de percolation quand p est petit.

L'argument central sera un usage approprié de la sous-σ-additivité des mesures de probabilité. On va prouver

Lemme 3.2.2. *Pour $p < \frac{1}{4}$,*

$$\theta_n(p) \leq \frac{(4p)^n}{1-4p}$$

pour tout $n \geq 1$. En particulier, $\theta_n(p) \xrightarrow{n\to\infty} 0$.

Démonstration. On se rappelle qu'un *chemin auto-évitant* γ de longueur n est une suite de sommets $(\gamma_0, \dots, \gamma_n)$ tel que $\{x_{k-1}, x_k\} \in E$ et $i \neq j$ implique $x_i \neq x_j$. On remarque alors que si $0 \overset{E_n}{\longleftrightarrow} \partial\Lambda_n$, il existe un chemin auto-évitant (noté SAW [1]) partant de 0 et ayant une longueur au moins n tel que toutes les arêtes parcourues par le chemin ont $X_e = 1$.

1. pour *self-avoiding walk*

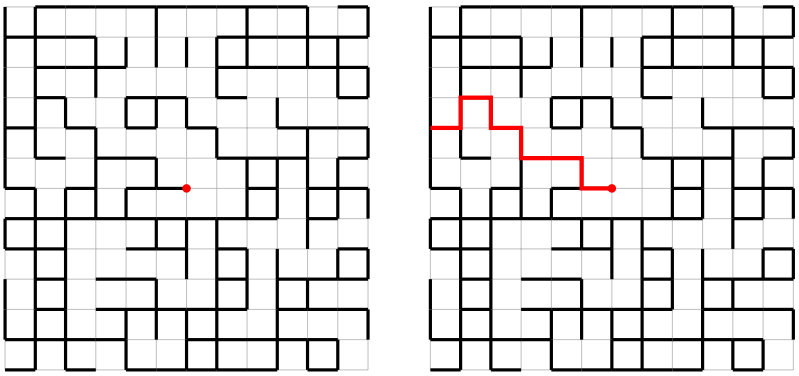

FIGURE 3.3 – Chemin auto-évitant induit par une connection.

On a alors

$$\{0 \overset{E_n}{\longleftrightarrow} \partial\Lambda_n\} \subset \bigcup_{k\geq n} \bigcup_{(\gamma_0=0,\dots,\gamma_k) \text{ SAW}} \bigcap_{k=1}^{n} \{X_{\{\gamma_k,\gamma_{k-1}\}} = 1\}.$$

Maintenant pour $(\gamma_0 = 0, \dots, \gamma_k)$ un chemin auto-évitant, on a que toutes les arêtes sont distinctes et donc par indépendance des X_es,

$$P_p\Big(\bigcap_{k=1}^{n} \{X_{\{\gamma_k,\gamma_{k-1}\}} = 1\}\Big) = \prod_{k=1}^{n} P_p(X_{\{\gamma_k,\gamma_{k-1}\}} = 1) = p^k.$$

On utilise alors la sous-σ-additivité de P_p pour obtenir

$$\begin{aligned} P_p(0 \overset{E_n}{\longleftrightarrow} \partial\Lambda_n) &\leq \sum_{k\geq n} \sum_{(\gamma_0=0,\dots,\gamma_k) \text{ SAW}} P_p\Big(\bigcap_{k=1}^{n} \{X_{\{\gamma_k,\gamma_{k-1}\}} = 1\}\Big) \\ &= \sum_{k\geq n} p^k |\big\{(\gamma_0 = 0, \dots, \gamma_k) \text{ SAW}\big\}|. \end{aligned}$$

On compte maintenant le nombres de chemins auto-évitant de longueur k partant de 0. Il y a moins de chemins auto-évitant que de chemins tout court, et il y a 4^k chemins de longueur k partant de 0 (4 choix de direction à chaque pas). On obtient

$$\theta_n(p) \leq \sum_{k\geq n} p^k 4^k = (4p)^n \sum_{k\geq 0} (4p)^k = \frac{(4p)^n}{1-4p}$$

quand $4p < 1$. □

Remarque 3.2.1. *On peut remarquer qu'il est aisé d'améliorer le* $1/4$ *en* $1/3$ *!*

3.2.2 Percolation quand p proche de 1.

On va procéder d'une manière similaire à l'argument pour p petit : au lieu de regarder quels chemins permettent de connecter loin, on va regarder quels événements empêchent la connexion. On introduit la notion de *dualité planaire* via l'image suivante :

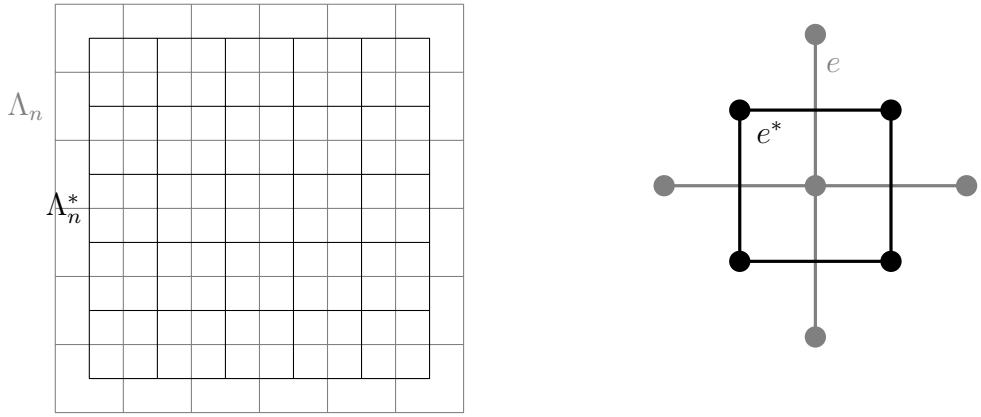

FIGURE 3.4 – Dualité planaire : on place un sommet dual au centre de chaque faces du réseau carré et on place une arête duale entre les plus proches voisins. Chaque arête duale traverse exactement une arête primale (en leurs milieus respectifs).

On notera $\Lambda_n^* = (1/2, 1/2) + \{-n, \dots, n-1\}^2$ l'ensemble des sommets duaux qui sont les centres d'une face dans Λ_n. Pour une arête $e \in E$, on notera e^* l'unique arête duale qui croise e en son milieu. On notera $(\mathbb{Z}^2)^*$ l'ensemble des sommets duaux et E^* l'ensemble des arêtes duales.

Lemme 3.2.3. *Pour tout $p > 7/8$,*

$$\theta_n(p) \geq \frac{1}{4}.$$

Démonstration. De la même manière que la connexion induit l'existence d'un chemin auto-évitant allat de 0 au bord de la boîte, l'absence de connexion entre 0 et le bord de la boîte induit l'existence d'un circuit auto-évitant γ dans Λ_n^* entourant 0, tel que toute les arêtes primales traversées par γ sont fermées.

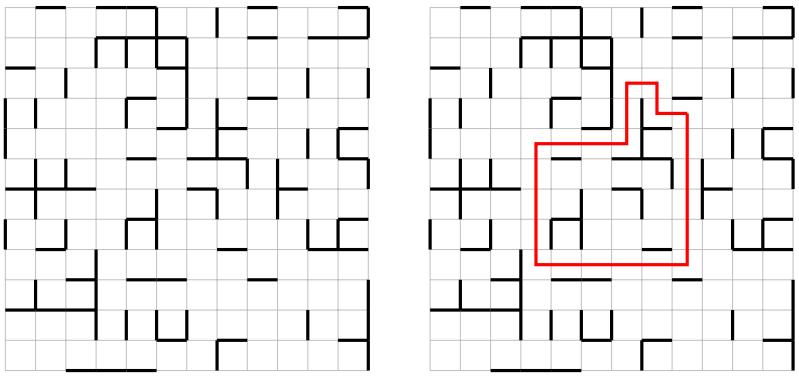

FIGURE 3.5 – Circuit dual auto-évitant bloquant les connections.

On a donc l'inclusion d'événements

$$\{0 \nleftrightarrow \partial\Lambda_n\} \subset \bigcup_{\gamma \in \mathcal{C}} \bigcap_{e : e^* \in \gamma} \{X_e = 0\},$$

où $\mathcal{C}$ est l'ensemble des circuits duaux auto-évitent qui entourent 0. De plus, par indépendance, pour $\gamma \in \mathcal{C}$ fixé,

$$P_p\Big(\bigcap_{e:e^*\in\gamma}\{X_e = 0\}\Big) = \prod_{e:e^*\in\gamma} P_p(X_e = 0) = (1-p)^{|\gamma|}$$

où $|\gamma|$ est le nombre d'arêtes duales dans γ. On peut alors utiliser la sous-σ-additivité (comme dans le Lemme 3.2.2) pour obtenir

$$P(0 \nleftrightarrow \partial\Lambda_n) \leq \sum_{k\geq 4} |\mathcal{C}_k|(1-p)^k,$$

où $\mathcal{C}_k$ est le nombre de circuits duaux auto-évitent qui entourent 0 qui contiennent k arêtes duales. La somme commence à 4 car le plus petit circuit qui entour 0 est de longueur 4. On va maintenant borner $|\mathcal{C}_k|$. On remarque que si $\gamma \in \mathcal{C}_k$, il doit avoir une arête dont le milieu est dans l'ensemble $\{0\} \times \{-1/2, -3/2, -5/2, \dots\}$. De plus, cette arête ne peut pas être plus basse que $-k/2$ (car il faut descendre jusqu'à cette arête depuis 0 puis remonter jusqu'à 0). On peut donc borner $|\mathcal{C}_k|$ par la somme sur $m = 0, \dots, k$ ($-m - \frac{1}{2}$ sera la hauteur de l'arête mentionnée précédemment) du nombre de circuits duaux de longueur k qui passent par l'arête $e_m = \{(-0.5, -m - \frac{1}{2}), (0.5, -m - \frac{1}{2})\}$. Ce nombre de circuits est alors borné par le nombre de marche auto-évitantes (sur $((\mathbb{Z}^2)^*, E^*)$) de longueur k démarrant en $(0.5, -m - \frac{1}{2})$. Ce nombre est alors plus petit que 4^k par le même argument quand dans la preuve du Lemme 3.2.2. On a alors obtenu

$$P(0 \nleftrightarrow \partial\Lambda_n) \leq \sum_{k\geq 4}(1-p)^k \sum_{m=0}^{k} 4^k = \sum_{k\geq 4}(4(1-p))^k(k+1).$$

On prend alors $p > \frac{7}{8}$ (ce qui implique $1 - p < \frac{1}{8}$) et on calcule la dernière somme (en utilisant $4(1-p) \leq \frac{1}{2}$) et on trouve [2]

$$P(0 \nleftrightarrow \partial\Lambda_n) \leq \frac{12}{16} = \frac{3}{4},$$

ce qui implique

$$\theta_n(p) = 1 - P(0 \nleftrightarrow \partial\Lambda_n) \geq \frac{1}{4}.$$

□

3.2.3 Monotonicité en p.

On se tourne maintenant vers le fait que la fonction $\theta_n(p)$ est croissante en p.

Lemme 3.2.4. *$n \mapsto \theta_n(p)$ est décroissante et $p \mapsto \theta_n(p)$ est croissante sur $[0,1]$. En particulier, il existe un unique $p_c \in [0,1]$ tel que*

2. On utilise que pour $|x| < 1$,

$$\sum_{k\geq 4} x^k(k+1) = \frac{d}{dx}\sum_{k\geq 5} x^k = \frac{d}{dx}\Big(\frac{x^5}{1-x}\Big) = \frac{x^4}{1-x}\Big(5 + \frac{x}{1-x}\Big)$$

— $p < p_c$ *implique* $\theta_n(p) \xrightarrow{n\to\infty} 0$;
— $p > p_c$ *implique* $\inf_{n\geq 1} \theta_n(p) > 0$.

Démonstration. L'observation $n \mapsto \theta_n(p)$ est décroissante suit du fait que si on connecte 0 à distance $n+1$, alors on connecte forcément 0 à distance n. Donc on a l'inclusion

$$\{0 \xleftrightarrow{E_{n+1}} \partial\Lambda_{n+1}\} \subset \{0 \xleftrightarrow{E_n} \partial\Lambda_n\}.$$

Pour la monotonicité en p, on va construire un objet que l'on appelle un *couplage* entre P_p et $P_{p'}$ pour $p < p'$: c'est une mesure de probabilité $(\Omega', \mathcal{G}, Q)$ et une paire de fonctions mesurables $\phi_p, \phi_{p'}$ telles que

$$(\phi_p)_* Q = P_p, \quad (\phi_{p'})_* Q = P_{p'}.$$

On a déjà vu un couplage : le produit de deux espaces de probabilité (muni de la mesure produit) donne un *couplage indépendant* : les projections fournissent les fonctions mesurables voulues. L'avantage de regarder des couplages plus généraux est que l'on peut en construire avec des propriétés plus particulières.

On commence par construire l'espace de probabilité voulu : on se donne n'importe quel espace $(\Omega', \mathcal{G}, Q)$ sur lequel on peut définir une famille i.i.d. $U_e, e \in E$ de variable uniformes sur $[0,1]$: $U_e \sim \mathrm{Uni}([0,1])$. On définit alors pour $p \in [0,1]$:

$$\phi_{p,e}(U) = \mathbb{1}_{[0,p]}(U_e).$$

Comme les $U_e, e \in E$ forment une famille indépendante, les $\phi_{p,e}(U), e \in E$ font de même. Leur loi est donc une loi produit, il suffit de calculer les marginales pour connaître la loi du processus stochastique $\phi_p = (\phi_{p,e}(U))_{e\in E}$: ce sont des variables de Bernoulli (car elles ne peuvent prendre que les valeurs 0 ou 1) et leur paramètre est donné par

$$Q(\phi_{p,e}(U) = 1) = Q(U_e \in [0,p]) = p.$$

On a donc que l'image par ϕ_p de Q est P_p. On remarque aussi que par construction, si $p' \geq p$,

$$\phi_{p,e}(U) \leq \phi_{p',e}(U) \ \forall e \in E.$$

En particulier, vu que rajouter des arêtes ouvertes ne fait que faciliter la connexion entre 0 et $\partial\Lambda_n$,

$$\begin{aligned} &P_{p'}(0 \xleftrightarrow{E_n} \partial\Lambda_n) - P_p(0 \xleftrightarrow{E_n} \partial\Lambda_n) = \\ &\quad = E_Q\Big(\mathbb{1}_{\{0\xleftrightarrow{E_n}\partial\Lambda_n\}}(\phi_{p'}(U))\Big) - E_Q\Big(\mathbb{1}_{\{0\xleftrightarrow{E_n}\partial\Lambda_n\}}(\phi_p(U))\Big) \\ &\quad = E_Q\Big(\underbrace{\mathbb{1}_{\{0\xleftrightarrow{E_n}\partial\Lambda_n\}}(\phi_{p'}(U)) - \mathbb{1}_{\{0\xleftrightarrow{E_n}\partial\Lambda_n\}}(\phi_p(U))}_{\geq 0}\Big) \geq 0, \end{aligned}$$

ce qui est la monotonicité requise. □

3.2.4 Preuve du Théorème 3.2.1

Le Lemme 3.2.4 nous donne le fait que p_c est une quantité bien définie. Le Lemme 3.2.2 assure que $p_c > 0$ et finalement le Lemme 3.2.3 implique que $p_c < 1$, ce qui conclu la preuve du Théorème 3.2.1.