

Introduction aux probabilités

2023 - 2024

François Simenhaus à partir des notes de Justin Salez

Table des matières

1	Espaces probabilisés	4
1.1	Univers	4
1.2	Tribus	6
1.3	Mesures de probabilité	11
1.4	Espaces probabilisés	16
2	Variables aléatoires réelles	18
2.1	Définition et propriétés	18
2.2	Variables discrètes	22
2.3	Variables à densité	25
2.4	Opérations usuelles	27
3	Espérance	31
3.1	Définition formelle	31
3.2	Propriétés fondamentales	33
3.3	Cas particuliers : variables discrètes ou à densité	36
3.4	Fluctuations autour de l'espérance	38
4	Indépendance	42
4.1	Probabilités conditionnelles	42
4.2	Événements indépendants	44
4.3	Variables aléatoires indépendantes	47
5	Sommes de variables indépendantes	51
5.1	Formules de convolution	51
5.2	Espérance et variance d'une somme	53

5.3	Loi faible des grands nombres	55
6	Fonctions caractéristiques (<i>Hors-programme</i>)	58
6.1	Définition et exemples	58
6.2	Propriétés fondamentales	60
7	Convergence en loi (<i>Hors-programme</i>)	62
A	Ensembles dénombrables	65
B	Quelques conseils de rédaction	68

Les probabilités sont au coeur du cursus à Dauphine et occupent de façon générale une place importante dans toutes les formations en mathématiques. Le contenu de ce cours vous resservira donc dans votre parcours, ici ou ailleurs. C'est aussi une première occasion de vous familiariser avec des notions difficiles, et notamment celles de théorie de la mesure que vous reverrez l'année prochaine dans le cours d'intégration.

Le cours a deux aspects aussi importants l'un que l'autre. Il comporte une partie théorique qui est l'introduction de la modélisation mathématique du hasard (espace probabilisé, variable aléatoire, indépendance,...). Il comporte également une partie plus technique avec de nombreux calculs permettant de se familiariser avec les lois les plus connues.

Les probabilités se sont développées à partir du 17^{ème} siècle mais il faut attendre le 20^{ème} pour avoir une formalisation de la théorie (ce qui ne signifie pas bien sûr que rien d'intéressant n'ait été fait avant). C'est Kolmogorov, en 1930, qui s'appuie sur la théorie de la mesure de Borel et Lebesgue pour proposer l'axiomatique que nous utilisons encore. Aujourd'hui la théorie des probabilités constitue une composante importante des mathématiques (premières médailles Fields en probabilités relativement récentes) et en particulier dans l'école mathématique française.

Je remercie Justin Salez qui m'a transmis tous les supports qu'il a conçus et utilisés lorsqu'il assurait ce cours. Le polycopié de cours comme les feuilles de TDs sont proches de ce qu'il m'a donné. Je remercie également José Trashorras pour les nombreuses améliorations faites sur les feuilles de TDs.

La formule qui détermine la note en fin de semestre est

$$\text{Note} = 0,1 \text{ CC1} + 0,2\text{P} + 0,1 \text{ CC2} + 0,6 \text{ E}$$

N'hésitez pas à m'écrire si vous avez des questions, des remarques sur le cours ou que vous avez noté des erreurs dans le poly :

mail : simenghaus@ceremade.dauphine.fr

bureau : B640

Les documents pédagogiques et toutes les informations sont sur l'équipe Teams dédiée au cours :

6jfi3t2

BON SEMESTRE À TOUTES ET TOUS !

1 Espaces probabilisés

Nous allons apprendre ici à modéliser des expériences aléatoires telles que le lancer d'une pièce de monnaie, le tirage de boules dans une urne ou le mélange d'un paquet de cartes. Par définition, les résultats de ces expériences sont imprévisibles et susceptibles de changer d'une fois sur l'autre, ce qui ne semble pas très compatible avec le langage mathématique auquel on est habitué. Il nous faudra donc un peu d'efforts pour parvenir à une description rigoureuse.

Avant de lire la suite, essayez de vous demander comment vous auriez utilisé les objets mathématiques donc vous avez l'habitude (ensembles, fonctions,...) pour représenter mathématiquement le hasard dont nous avons tous une conception intuitive. La théorie exposée ci-dessous ne vous en semblera que plus élégante !

L'objet fondamental qui permet de décrire convenablement une expérience aléatoire est l'*espace probabilisé*. Il est constitué de trois ingrédients : un *univers* Ω , une *tribu* \mathcal{F} , et une *mesure de probabilité* \mathbb{P} (parfois on dit simplement une *probabilité*). Nous allons à présent passer un peu de temps sur chacune de ces notions.

1.1 Univers

Pour décrire une expérience aléatoire, il faut commencer par en spécifier les résultats possibles. L'ensemble de tous ces résultats est appelé *univers*, et sera noté Ω . La nouveauté cette année est que nous n'imposerons aucune restriction sur la taille de cet ensemble : il pourra être fini, infini dénombrable, ou infini non-dénombrable. Notons que le choix de l'univers n'est pas unique : il y a plusieurs façons raisonnables de décrire les choses. Voici quelques exemples :

1. Lancer d'un dé : $\Omega = \{1, 2, 3, 4, 5, 6\}$.
2. Deux lancers de dé successifs : $\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$.
3. Lancer d'une pièce de monnaie : $\Omega = \{\text{Pile}, \text{Face}\}$.
4. Pour $n \in \mathbb{N}$ fixé, n lancers successifs d'une pièce de monnaie : $\Omega = \{\text{Pile}, \text{Face}\}^n$.
5. Une suite de lancers d'une pièce de monnaie : $\Omega = \{\text{Pile}, \text{Face}\}^{\mathbb{N}}$. *Question : Ω est-il dénombrable ? Lire l'Annexe A.*
6. Deux tirages successifs avec remise dans une urne contenant une boule rouge, une boule verte et une boule bleue : $\Omega = \{R, V, B\} \times \{R, V, B\}$.
7. Même exemple, sans remise : $\Omega = \{(R, V), (R, B), (V, B), (V, R), (B, R), (B, V)\}$.
8. Mélange d'un paquet de cartes : $\Omega = \mathfrak{S}_{52}$, le groupe des permutations d'ordre 52.
9. Nombre d'éruptions d'un volcan durant le prochain siècle : $\Omega = \mathbb{N}$.

10. Proportion de votes démocrates à la prochaine élection présidentielle américaine : $\Omega = [0, 1]$.
11. Durée de vie d'une ampoule neuve : $\Omega = [0, \infty[$.

Voici quelques exemples d'univers un peu plus sophistiqués :

12. *Percolation*. On fixe un entier $n \in \mathbb{N}$ et on considère un écran carré de taille (en pixels) n , $B_n = \{1, \dots, n\}^2$. Chaque point (pixel) de B_n peut être allumée ou non. L'ensemble des pixels allumés est aléatoire ce qui nous donne comme univers

$$\Omega = \{0, 1\}^{B_n},$$

où 1 signifie que le pixel est allumée et 0 qu'il est éteint.

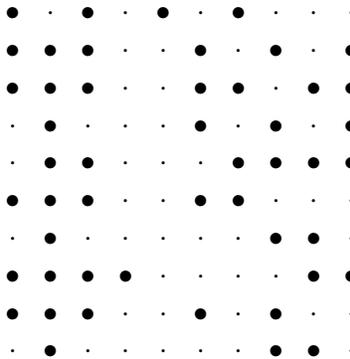


FIGURE 1: Un exemple d'élément dans Ω pour $n = 10$. Les points épais correspondent à des 1 (pixel allumé) et les petits aux 0 (pixel éteint).

Ici Ω est fini mais on peut aussi penser à un écran infini et l'univers est alors $\Omega = \{0, 1\}^{\mathbb{Z}^2}$ qui n'est plus un espace fini, ni même dénombrable.

13. On fixe à nouveau un entier $n \in \mathbb{N}$ et on considère cette fois une population de taille n dont les individus sont numérotés de 1 à n . Le fait pour un individu d'en connaître un autre est aléatoire et on souhaite modéliser ces relations de connaissance. On considère donc le graphe (penser aux graphes représentant les liens dans un réseau social) dont l'ensemble des sommets est $\mathcal{S} = \{1, \dots, n\}$ et l'ensemble des arrêtes non orientées est $\mathcal{A} = \{\{i, j\}, 1 \leq i < j \leq n\}$. On peut considérer comme univers l'ensemble des parties de l'ensemble des arrêtes,

$$\Omega = \mathcal{P}(\mathcal{A}).$$

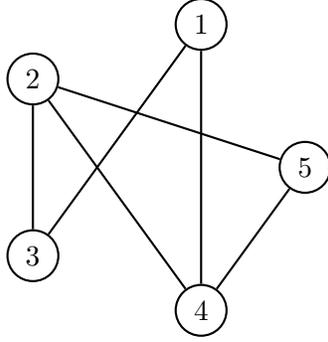


FIGURE 2: Exemple 13. Un élément $\omega \in \Omega$ pour $n = 5$.

1.2 Tribus

Notre expérience aléatoire va « produire » un résultat $\omega \in \Omega$, et nous allons chercher à savoir si ce résultat est tombé dans telle ou telle partie $A \subseteq \Omega$ de notre univers. Une fois l'univers Ω spécifié, notre tâche suivante consiste à dresser une liste de tous les événements $A \subseteq \Omega$ qui seront susceptibles de nous intéresser. L'ensemble de ces événements sera appelé *tribu*, et noté \mathcal{F} . Une tribu \mathcal{F} est donc une collection ou encore une famille de parties de Ω . On a donc

$$\mathcal{F} \subset \mathcal{P}(\Omega).$$

Remarque 1. *Ce sont les événements, i.e. les parties de Ω qui font partie de \mathcal{F} dont nous allons évaluer les probabilités et non les éléments de Ω . Une première raison simple pour cela est que pour certaines expériences aléatoires la probabilité d'un résultat fixé est nulle quelque soit le résultat considéré (par exemple un jet de flechette sur une cible : la probabilité que la flechette tombe sur un point donné de la cible est nulle)*

Remarque 2. *Il peut sembler tentant à ce stade du cours de prendre systématiquement $\mathcal{F} = \mathcal{P}(\Omega)$, quitte à ce que de nombreux événements aient probabilité 0, mais ça n'est malheureusement pas pertinent dans certains cas. Le problème est que $\mathcal{P}(\Omega)$ est dans certains cas « trop gros » pour que l'on puisse construire des probabilités « naturelles » sur tout $\mathcal{P}(\Omega)$, c'est-à-dire associer une probabilité à chaque partie de Ω en respectant les règles qu'une probabilité doit satisfaire (et que l'on verra dans la partie suivante). C'est notamment le cas quand Ω n'est pas dénombrable. Il faut alors se restreindre à une sous famille stricte de $\mathcal{P}(\Omega)$. Nous reviendrons un peu plus loin sur ce point.*

Pour définir une tribu, nous allons imposer de respecter quelques règles de stabilité.

Définition 1 (Tribu). *Une tribu sur Ω est un ensemble \mathcal{F} de parties de Ω vérifiant :*

1. (Ensemble vide) $\emptyset \in \mathcal{F}$.
2. (Complémentaire) Si $A \in \mathcal{F}$, alors $A^c \in \mathcal{F}$.
3. (Union dénombrable) Si $(A_n)_{n \geq 1}$ est une suite d'éléments de \mathcal{F} , alors $\bigcup_{n \geq 1} A_n \in \mathcal{F}$.

Remarque 3. On rappelle qu'un ensemble dénombrable est un ensemble qui est en bijection avec une partie de \mathbb{N} (ou, de façon équivalente, on peut aussi définir un ensemble dénombrable comme un ensemble qui peut être injecté dans \mathbb{N}). On notera donc que dans ce cours un ensemble dénombrable peut être fini ou dénombrable infini. Dans d'autres documents vous trouverez peut-être une autre convention où dénombrable signifie seulement infini dénombrable et cela exclut donc les ensembles finis. Cette notion de dénombrabilité joue un rôle important en théorie de probabilités (et plus généralement en maths !). Pour vous familiariser avec cette notion je vous conseille de vous reporter à l'Annexe A et surtout de faire les exercices de la première feuille de TD.

Nous réserverons désormais le terme *événement* aux seules parties $A \in \mathcal{P}(\Omega)$ qui sont dans \mathcal{F} .

Exemple 1. Par exemple, l'obtention d'un chiffre pair avec un dé correspond à la réalisation de l'événement $A = \{2, 4, 6\}$ dans l'univers $\Omega = \{1, 2, 3, 4, 5, 6\}$, tandis que l'obtention d'un double avec deux lancers de dé successifs correspond à la réalisation de l'événement $A = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$ dans l'univers $\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$.

Exemple 2. Dans l'Exemple 12, on peut considérer l'événement suivant : il existe un chemin aux plus proches voisins de sites allumés qui relie le côté gauche du carré B_n au côté droit :

$$A = \left\{ \begin{array}{l} \omega \in \Omega, \exists K \geq 0, \gamma : \{0, \dots, K\} \rightarrow B_n, \\ \text{pour tout } i \in \{0, \dots, K-1\}, \|\gamma(i+1) - \gamma(i)\|_1 = 1, \\ \gamma(1)_1 = 1, \gamma(K)_1 = n \\ \text{pour tout } i \in \{0, \dots, K\}, \omega_{\gamma(i)} = 1 \end{array} \right\}$$

Exemple 3. Sur l'univers $\Omega = \{1, 2, 3, 4, 5, 6\}$, l'ensemble $\mathcal{F} = \{\emptyset, \{2, 4, 6\}, \{1, 2, 3, 4, 5, 6\}\}$ n'est pas une tribu, car il contient $\{2, 4, 6\}$ mais pas son complémentaire. En revanche, l'ensemble $\mathcal{F} = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \{1, 2, 3, 4, 5, 6\}\}$ est bien une tribu.

Il y a une correspondance entre le vocabulaire décrivant une expérience aléatoire et les notations ensemblistes : si A et B sont des événements,

1. A : A s'est réalisé.
2. A^c : A ne se réalise pas
3. $A \cup B$: A ou B se réalisent

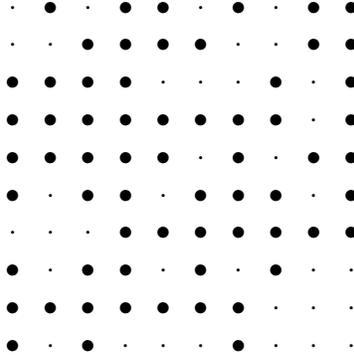


FIGURE 3: Cet élément $(\omega_x)_{x \in B_n} \in \Omega$ appartient-il à A ?

4. $A \cap B$: A et B se réalisent
5. $A \subset B$: si A se réalise alors B aussi
6. $A \cap B = \emptyset$: A et B ne peuvent se réaliser en même temps

Et la liste est bien sûr à compléter...

La définition d'une tribu est concise, mais elle implique la stabilité de \mathcal{F} par toutes sortes d'opérations ensemblistes classiques, notamment celles listées ci-dessous.

Proposition 1. *Soit \mathcal{F} une tribu sur un univers Ω . Alors on a aussi :*

1. (Ensemble plein) $\Omega \in \mathcal{F}$.
2. (Union finie) Si $A, B \in \mathcal{F}$, alors $A \cup B \in \mathcal{F}$.
3. (Intersection finie) Si $A, B \in \mathcal{F}$, alors $A \cap B \in \mathcal{F}$.
4. (Différence) Si $A, B \in \mathcal{F}$, alors $A \setminus B \in \mathcal{F}$.
5. (Différence symétrique) Si $A, B \in \mathcal{F}$, alors $A \Delta B := (A \setminus B) \cup (B \setminus A) \in \mathcal{F}$.
6. (Intersection dénombrable) Si $(A_n)_{n \geq 1}$ est une suite d'éléments de \mathcal{F} , alors $\bigcap_{n \geq 1} A_n \in \mathcal{F}$.

Démonstration. 1. On a $\Omega = \emptyset^c$. Or \emptyset est dans \mathcal{F} donc son complémentaire aussi.

2. Dans ce cours "dénombrable" = "fini ou dénombrable" donc il n'y a rien à montrer.

3. On a $A \cap B = (A^c \cup B^c)^c$ et on utilise encore la stabilité par passage au complémentaire.

4. On remarque que $A \setminus B = A \cap B^c$ et on utilise la stabilité par intersection établie au point précédent et la stabilité par passage au complémentaire.

5. On a $A \Delta B := (A \setminus B) \cup (B \setminus A) \in \mathcal{F}$ et on utilise les points précédents.

6. Enfin, $\bigcap_{n \geq 1} A_n = (\bigcup_{n \geq 1} A_n^c)^c$ et on utilise encore la stabilité du passage au complémentaire. □

L'ensemble $\mathcal{P}(\Omega)$ de toutes les parties est évidemment une tribu, et c'est celle que nous adopterons systématiquement lorsque Ω est fini ou dénombrable. Sur des espaces plus gros, on procédera plutôt comme suit : on commencera par spécifier un ensemble \mathcal{C} de parties que l'on souhaiterait absolument voir figurer dans la tribu, puis on y ajoutera seulement les parties qui manquent pour que les axiomes d'une tribu soient satisfaits. Cette idée est formalisée par la définition suivante.

Définition 2 (Tribu engendrée). *Soit \mathcal{C} un ensemble quelconque de parties de Ω . La tribu engendrée par \mathcal{C} est la plus petite (au sens de l'inclusion) tribu sur Ω qui contient toutes les parties $A \in \mathcal{C}$. On la note $\sigma(\mathcal{C})$.*

Remarque 4. *Le lecteur inquiet de l'existence de $\sigma(\mathcal{C})$ pourra vérifier que l'ensemble*

$$\sigma(\mathcal{C}) := \bigcap_{\substack{\mathcal{F}: \text{tribu sur } \Omega \\ \mathcal{F} \supseteq \mathcal{C}}} \mathcal{F},$$

est bien une tribu (car une intersection de tribu est encore une tribu) qui contient \mathcal{C} , et qui est incluse dans toute autre tribu contenant \mathcal{C} . L'intersection est par ailleurs non vide puisque $\mathcal{P}(\Omega)$ est une tribu et contient \mathcal{C} . Cette expression n'a cependant que peu d'intérêt pratique et elle ne permet pas de lister tous les éléments de $\sigma(\mathcal{C})$.

On notera bien, car cela sert souvent, que si

1. \mathcal{G} est une tribu ;
2. \mathcal{G} contient \mathcal{C} ,

alors \mathcal{G} contient $\sigma(\mathcal{C})$.

Exercice 1. *Soit \mathcal{C} l'ensemble de toutes les parties finies de \mathbb{R} . Justifier que \mathcal{C} n'est pas une tribu sur l'univers $\Omega = \mathbb{R}$, puis déterminer la tribu engendrée par \mathcal{C} .*

Les intervalles réels constituent une famille tout-à-fait raisonnable de parties de l'univers $\Omega = \mathbb{R}$, mais ne forment pas une tribu (pourquoi ?). Nous sommes donc naturellement amenés à considérer la tribu qu'ils engendrent. Cette dernière est suffisamment importante pour mériter une définition.

Définition 3 (Tribu borélienne). *On appelle tribu borélienne, et on note $\mathcal{B}(\mathbb{R})$, la tribu engendrée sur \mathbb{R} par les intervalles de la forme $]a, b[$ avec $a, b \in \mathbb{R}$ et $a < b$. Les éléments de $\mathcal{B}(\mathbb{R})$ sont appelés les boréliens.*

Une définition équivalente (mais hors-programme pour ce semestre !) que vous rencontrerez souvent et qui permet de définir les tribus boréliennes de façon plus générale (pour tous les ensembles munis d'une métrique par exemple ou d'une topologie) :

Définition 4 (Tribu borélienne). *On appelle tribu borélienne la tribu engendrée par les ouverts de \mathbb{R} .*

On rappelle qu'un ensemble $\mathcal{O} \subset \mathbb{R}$ est un ouvert si pour tout $x \in \mathcal{O}$ il existe $\varepsilon > 0$ tel que $]x - \varepsilon, x + \varepsilon[\subset \mathcal{O}$. Ainsi $]0, 2[$ est un ouvert (comme tous les intervalles ouverts) mais $[1, 3[$ ne l'est pas. La preuve de l'équivalence des deux définitions de la tribu borélienne est l'objet de l'exercice 2 ci-dessous.

La tribu borélienne est énorme : elle contient bien-sûr les intervalles, mais aussi toutes les parties qui peuvent être obtenues à partir de ceux-ci en effectuant un nombre arbitraire d'unions dénombrables et de passages au complémentaire. Dans la pratique, tous les ensembles réels auxquels on peut penser sont boréliens, et il est bien difficile de construire un contre-exemple. À titre d'exercice, nous invitons le lecteur à considérer un ensemble de réels rencontré au cours de sa scolarité – par exemple les irrationnels – et à vérifier qu'il s'agit bien d'un borélien. Dans la suite, nous munirons toujours \mathbb{R} de la tribu borélienne sauf lorsque nous préciserons explicitement que ça n'est pas le cas. L'exercice qui suit conforte ce choix, puisqu'il montre que $\mathcal{B}(\mathbb{R})$ est également engendrée par plusieurs autres familles toutes aussi naturelles.

Exercice 2. *Montrer que $\mathcal{B}(\mathbb{R})$ est aussi engendrée par chacune des familles suivantes :*

1. *Les intervalles de la forme $[a, b]$ avec $a, b \in \mathbb{R}$ et $a < b$.*
2. *Les intervalles de la forme $] - \infty, a]$ avec $a \in \mathbb{R}$.*
3. *(hors – programme) Les parties ouvertes de \mathbb{R} .*
4. *(hors – programme) Les parties fermées de \mathbb{R} .*

Il est rarement possible de décrire explicitement tous les éléments de $\sigma(\mathcal{C})$ (où \mathcal{C} est un ensemble de parties de Ω). Aussi lorsque l'on veut montrer que $\sigma(\mathcal{C})$ est incluse dans une classe \mathcal{G} on raisonne souvent de la façon suivante.

1. On commence par montrer que $\mathcal{C} \subset \mathcal{G}$.
2. On montre que \mathcal{G} est une tribu.
3. Et on conclut en notant que \mathcal{G} est une tribu et contient \mathcal{C} , donc contient la plus petite tribu qui contient \mathcal{C} , c'est -à-dire $\sigma(\mathcal{C})$.

C'est une méthode à retenir car vous l'utiliserez de nombreuses fois ! Il est au contraire souvent piégeur de vouloir raisonner directement en se donnant un événement A de la tribu $\sigma(\mathcal{C})$ et en essayant de montrer qu'il est dans la tribu \mathcal{G} . Il y a cependant quelques cas exceptionnels où on peut facilement décrire la tribu $\sigma(\mathcal{C})$, notamment lorsque \mathcal{C} est une partition de Ω (voir l'Exercice 3 du TD 2).

Illustrons cette méthode en traitant le point 1 de l'exercice ci-dessus. On note \mathcal{H} la tribu engendrée par les intervalles de la forme $[a, b]$ avec $a, b \in \mathbb{R}$ et $a < b$. Considérons $a, b \in \mathbb{R}$ avec $a < b$.

Montrons d'abord que $\mathcal{B}(\mathbb{R}) \subset \mathcal{H}$. On a pour tout $a < b$

$$]a, b[= \cup_{n \geq 1} [a + 1/n, b - 1/n]$$

donc, par stabilité de \mathcal{H} par union dénombrable, $]a, b[\in \mathcal{H}$. On a donc que \mathcal{H} est une tribu et \mathcal{H} contient tous les intervalles de la forme $]a, b[$ avec $a, b \in \mathbb{R}$ et $a < b$ donc \mathcal{H} contient $\mathcal{B}(\mathbb{R})$. Réciproquement montrons que $\mathcal{H} \subset \mathcal{B}(\mathbb{R})$. Pour tout $a < b$,

$$[a, b] = \cap_{n \geq 1}]a - 1/n, b + 1/n[$$

donc, par stabilité de $\mathcal{B}(\mathbb{R})$ par intersection dénombrable, $[a, b] \in \mathcal{B}(\mathbb{R})$. On a donc que $\mathcal{B}(\mathbb{R})$ est une tribu et $\mathcal{B}(\mathbb{R})$ contient tous les intervalles de la forme $[a, b]$ avec $a, b \in \mathbb{R}$ et $a < b$ donc $\mathcal{B}(\mathbb{R})$ contient \mathcal{H} .

Il est difficile de montrer que $\mathcal{P}(\mathbb{R}) \neq \mathcal{B}(\mathbb{R})$ et cela nécessite d'admettre l'axiome du choix. Vous verrez cependant l'année prochaine des exemples de telles parties dans le cours d'intégration. Vous pouvez aussi aller voir le paradoxe de Banach-Tarski, très bien expliqué sur wikipedia, pour un joli théorème qui montre au passage l'existence de parties non boreliennes dans \mathbb{R}^3 .

1.3 Mesures de probabilité

Maintenant que nous nous sommes mis d'accord sur la liste des événements susceptibles de nous intéresser, il ne nous reste plus qu'à préciser les chances qu'ils ont de se réaliser. À chaque événement $A \in \mathcal{F}$, nous allons associer une probabilité $\mathbb{P}(A) \geq 0$. L'application $A \mapsto \mathbb{P}(A)$ devra seulement respecter deux règles très naturelles : l'événement "certain" Ω se réalise avec probabilité 1, et les probabilités s'ajoutent lorsque l'on réunit une suite d'événements deux-à-deux disjoints.

Définition 5 (Mesure de probabilité ou, seulement, probabilité). *Soit Ω un univers muni d'une tribu \mathcal{F} . Une mesure de probabilité sur (Ω, \mathcal{F}) est une fonction $\mathbb{P}: \mathcal{F} \rightarrow [0, 1]$ qui vérifie :*

1. $\mathbb{P}(\Omega) = 1$
2. pour toute suite $(A_n)_{n \geq 1}$ d'éléments de \mathcal{F} deux-à-deux disjoints

$$\mathbb{P} \left(\bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

La seconde propriété est appelée σ -additivité (le σ réfère à la dénombrabilité).

Remarque 5. On notera bien qu'une probabilité est une fonction qui ne prend pas comme argument les éléments ω de l'univers Ω mais les événements A de la tribu \mathcal{F} .

Exemple 4 (Lois discrètes). Soit Ω un ensemble fini ou dénombrable, et $p: \Omega \rightarrow [0, 1]$ vérifiant

$$\sum_{x \in \Omega} p(x) = 1.$$

Alors l'application $\mathbb{P}: \mathcal{P}(\Omega) \rightarrow [0, 1]$ définie par

$$\mathbb{P}(A) := \sum_{x \in A} p(x)$$

est une mesure de probabilité sur $(\Omega, \mathcal{P}(\Omega))$. On pourra appeler p une "fonction de poids" même si ce terme n'est pas complètement répandu. Attention à nouveau : p n'est pas une probabilité ! Mais cette fonction permet de définir une probabilité.

On notera que toute probabilité \mathbb{P} sur un ensemble Ω dénombrable admet la fonction de poids $p: \Omega \rightarrow [0, 1]$ définie par

$$p(x) = \mathbb{P}(\{x\}), \quad x \in \Omega.$$

Les mesures de probabilités sur un univers dénombrable sont appelées lois discrètes. Elles admettent une fonction de poids qui les caractérisent.

Démonstration. La preuve est facile : $\mathbb{P}(\Omega) = \sum_{x \in \Omega} p(x) = 1$ et pour toute suite $(A_n)_{n \geq 1}$ d'éléments de \mathcal{F} deux-à-deux disjoints

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{x \in \bigcup_{n=1}^{\infty} A_n} p(x) = \sum_{n \geq 1} \sum_{x \in A_n} p(x) = \sum_{n \geq 1} \mathbb{P}(A_n).$$

□

Lorsque Ω est fini, un exemple important consiste à prendre $p(\omega) = \frac{1}{\text{card}(\Omega)}$ pour tout $\omega \in \Omega$ (tous les résultats sont équiprobables). On obtient alors la loi discrète importante suivante.

Exemple 5 (Loi uniforme sur un ensemble fini). Soit Ω un ensemble fini et non-vide, muni de la tribu $\mathcal{F} = \mathcal{P}(\Omega)$. Alors l'application $\mathbb{P}: \mathcal{F} \rightarrow [0, 1]$ définie par

$$\mathbb{P}(A) := \frac{\text{card}(A)}{\text{card}(\Omega)}, \tag{1}$$

est une mesure de probabilité appelée loi uniforme sur Ω .

Démonstration. Facile pour 1. Pour 2 : soit $(A_n)_{n \geq 1}$ une suite d'éléments de \mathcal{F} deux-à-deux dis-joints. La sigma-additivité vient facilement de la propriété suivante du cardinal $\text{card}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \text{card}(A_n)$.

On aurait aussi pu simplement noter que \mathbb{P} est la probabilité associée à la fonction de poids $p : x \in \Omega \mapsto 1/\text{Card}(\Omega)$. □

Exemple 6 (Mesure de Dirac). *On considère un univers Ω et une tribu \mathcal{F} sur cette univers. Pour tout $x \in \Omega$, on appelle masse de Dirac en x et on note δ_x la mesure de probabilité suivante :*

$$\delta_x : \mathcal{F} \rightarrow [0, 1]$$

$$A \rightarrow \mathbb{P}(A) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

Intuitivement cette probabilité correspond à un tirage déterministe de x .

De la définition générale d'une mesure de probabilité découlent diverses "règles de calcul" bien utiles en pratique. En voici une liste non-exhaustive, à retenir pour les exercices.

Proposition 2 (Règles de calcul). *Soit \mathbb{P} une mesure de probabilité quelconque. Alors,*

1. (Vide) *On a toujours*

$$\mathbb{P}(\emptyset) = 0.$$

2. (Passage au complémentaire) *Pour tout événement A ,*

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

3. (Différence) *Pour deux événements A, B quelconques,*

$$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

En particulier si $A \subset B$ alors $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$.

4. (Monotonie) *Pour deux événements A, B quelconques,*

$$A \subseteq B \implies \mathbb{P}(A) \leq \mathbb{P}(B).$$

5. (Formule du crible) *Pour deux événements A, B quelconques,*

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

6. (Sous-additivité) Pour une suite quelconque d'événements $(A_n)_{n \geq 1}$,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

7. (Continuité croissante) Pour une suite croissante d'événements $(A_n)_{n \geq 1}$,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} \uparrow A_n\right) = \lim_{n \rightarrow \infty} \uparrow \mathbb{P}(A_n).$$

8. (Continuité décroissante) Pour une suite décroissante d'événements $(A_n)_{n \geq 1}$,

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} \downarrow A_n\right) = \lim_{n \rightarrow \infty} \downarrow \mathbb{P}(A_n).$$

Démonstration. 1. La première est conséquence de la seconde !

2. On note que $\Omega = A \cup A^c$ et que ces deux ensembles sont disjoints donc

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c).$$

3. On note que $B = (B \setminus A) \cup (A \cap B)$ et que ces deux ensembles sont disjoints donc

$$\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B).$$

4. Conséquence du point précédent. Si $A \subset B$ alors $A \cap B = A$ et

$$\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A).$$

et comme $\mathbb{P}(B \setminus A) \geq 0$ on obtient le résultat.

5. Il suffit d'écrire l'union disjointe $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$. On utilise ensuite le point 3 pour conclure.

6. On définit la suite d'événements disjoints suivants : $B_1 = A_1$, $B_2 = A_2 \setminus B_1$ et pour $n \geq 1$, $B_{n+1} = A_{n+1} \setminus (\cup_{i=1}^n A_i)$. On a alors $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$...mais on a gagné que les événements de cette union sont disjoints 2 à 2 et donc

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n),$$

où, pour la dernière inégalité on a utilisé que pour tout $n \geq 1$, $B_n \subset A_n$.

7. On reprend la suite $(B_n)_{n \geq 1}$ introduite au point précédent et on obtient, en utilisant que pour tout $N \geq 1$, $\bigcup_{n=1}^N A_n = \bigcup_{n=1}^N B_n$,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) = \lim_{N \rightarrow +\infty} \uparrow \sum_{n=1}^N \mathbb{P}(B_n) = \lim_{N \rightarrow +\infty} \uparrow \mathbb{P}(A_N)$$

8. D'après le point 2,

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} \downarrow A_n\right) = 1 - \mathbb{P}\left(\bigcup_{n=1}^{\infty} \uparrow A_n^c\right)$$

et comme la suite $(A_n^c)_{n \geq 1}$ est croissante, d'après le point précédent

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} \uparrow A_n^c\right) = \lim_{n \rightarrow \infty} \uparrow \mathbb{P}(A_n^c) = 1 - \lim_{n \rightarrow \infty} \downarrow \mathbb{P}(A_n),$$

et on obtient le résultat.

Les flèches dans les notations $\bigcup \uparrow$ et $\bigcap \downarrow$ n'ont pas d'autres utilité que de rappeler aux lecteurs quand l'union est croissante ou l'intersection décroissante. Elles sont bien sûr facultative dans la rédaction. \square

Ces règles permettent de déduire la probabilité de certains événements à partir de celles d'autres événements. Au delà de l'intérêt pratique évident, une conséquence théorique importante est qu'il n'est pas nécessaire de spécifier la probabilité $\mathbb{P}(A)$ de chaque événement $A \in \mathcal{F}$ pour décrire une mesure de probabilité \mathbb{P} : si $\mathcal{C} \subseteq \mathcal{F}$ est une collection d'événements "suffisamment grosse", alors la connaissance de $\mathbb{P}(A)$ pour tout $A \in \mathcal{C}$ suffira à reconstruire entièrement l'application $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$. Le résultat suivant, admis, donne un sens rigoureux à l'expression "suffisamment grosse".

Théorème 1. *Soit $\mathcal{C} \subseteq \mathcal{F}$ une collection d'événements telle que*

1. \mathcal{C} engendre la tribu \mathcal{F} , i.e. $\sigma(\mathcal{C}) = \mathcal{F}$;
2. \mathcal{C} est stable par intersection finie : si $A \in \mathcal{C}$ et $B \in \mathcal{C}$, alors $A \cap B \in \mathcal{C}$.

Alors toute mesure de probabilité sur (Ω, \mathcal{F}) est déterminée par ses valeurs sur \mathcal{C} . Autrement dit, deux mesures de probabilités \mathbb{P} et \mathbb{Q} sur (Ω, \mathcal{F}) telles que $\mathbb{P}(A) = \mathbb{Q}(A)$ pour tout $A \in \mathcal{C}$ sont égales sur tout \mathcal{F} .

La démonstration de ce résultat déborde le programme de cette année. C'est une conséquence directe du *lemme de classe monotone*. Les lecteurs les plus curieux pourront faire le dernier exercice du TD 2 qui établit ce résultat. La condition 2 est importante dans le théorème précédent comme le prouve le contre-exemple suivant.

Exercice 3. On considère $\Omega = \{1, 2, 3, 4, 5\}$ et la collection de parties $\mathcal{C} = \{\{1, 2, 3\}, \{2, 4\}, \{3, 4, 5\}\}$.

1. Vérifier que $\sigma(\mathcal{C}) = \mathcal{P}(\Omega)$ (montrer que tous les singletons sont dans $\sigma(\mathcal{C})$ puis conclure en utilisant la stabilité par union dénombrable).
2. On note \mathbb{P} la probabilité uniforme et on définit la probabilité \mathbb{Q} par la fonction de poids $p(1) = 3/10$, $p(2) = 1/10$, $p(3) = 1/5$, $p(4) = 3/10$, $p(5) = 1/10$. Montrer que \mathbb{P} et \mathbb{Q} coïncident sur \mathcal{C} mais pas sur $\mathcal{P}(\Omega)$. Est ce que cela contredit le Théorème 1 ?

Voici une application importante de ce résultat, que nous utiliserons dans le chapitre suivant.

Corollaire 1 (Cas borélien). Une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est entièrement déterminée par ses valeurs sur les intervalles de la forme $] - \infty, t]$ avec $t \in \mathbb{R}$.

1.4 Espaces probabilisés

Nous disposons désormais de tout le vocabulaire nécessaire pour définir notre objet fondamental.

Définition 6 (Espace probabilisé). Un espace probabilisé (ou espace de probabilité) est un triplet $(\Omega, \mathcal{F}, \mathbb{P})$, où Ω est un univers, \mathcal{F} une tribu sur Ω , et \mathbb{P} une probabilité sur (Ω, \mathcal{F}) .

En résumé, le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ constitue le modèle mathématique de notre expérience aléatoire : l'univers Ω représente l'ensemble des résultats possibles de l'expérience ; la tribu \mathcal{F} précise les événements, c'est-à-dire les parties $A \subseteq \Omega$ dont nous allons déterminer la probabilité ; enfin, la mesure de probabilité \mathbb{P} spécifie les chances (les probabilités) que ces événements ont de se réaliser. Voici une illustration simple pour fixer les idées.

Exemple 7. On lance deux fois un dé, et on s'intéresse aux chances d'obtenir un double. L'univers des possibles est $\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\}$, que l'on munit de la tribu $\mathcal{F} = \mathcal{P}(\Omega)$ et de la loi uniforme (1). Dans l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, l'obtention d'un double correspond à l'événement

$$A = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\},$$

et la probabilité qu'il se réalise se calcule ainsi :

$$\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{6}{36} = \frac{1}{6} \approx 0.17.$$

Ainsi, dans notre modèle, il y a environ 17% de chances d'obtenir un double en lançant deux dés.

Exemple 8. Reprenons l'exemple de la percolation, Exemple 12. On peut donc considérer l'espace probabilisé suivant, $\Omega = \{0, 1\}^{B_n}$, $\mathcal{F} = \mathcal{P}(\Omega)$ et \mathbb{P} la probabilité uniforme sur Ω (qui correspond à

un tirage indépendants où chaque pixel est allumé ou éteint avec probabilité $1/2$, comme nous le verrons au Chapitre 4). Avec les objets simples construits dans cette section, nous pouvons déjà poser quelques questions très difficiles. Par exemple que vaut $\mathbb{P}(A)$, où on rappelle que A est l'événement "les côtés gauche et droite de l'écran sont connectés par un chemin de pixels allumés" ? Quelle est la limite de cette quantité quand $n \rightarrow +\infty$? Il est étonnant que l'on puisse déjà au terme de ce premier chapitre poser des questions qui dépassent largement ce que nous allons apprendre cette année et même tout ce que nous allons apprendre pendant le cursus dauphinois !

Nous terminons par un peu de vocabulaire concernant les espaces probabilisés.

- ▷ Comme on l'a déjà dit, les parties A qui sont dans la tribu \mathcal{F} sont appelés *événements*.
- ▷ Un événement A tel que $\mathbb{P}(A) = 1$ est dit *presque-sûr*. Il se produit *presque-sûrement* (p.s.). On notera bien que Ω est un événement qui se produit p.s. mais que ça n'est pas forcément le seul.
- ▷ Un événement A tel que $\mathbb{P}(A) = 0$ est dit *négligeable*. On notera cette fois que si \emptyset est un événement négligeable ça n'est pas forcément le seul.
- ▷ Le couple (Ω, \mathcal{F}) (sans mesure de probabilité) porte le nom d'*espace mesurable*. C'est un espace "en attente" d'une mesure de probabilité.

Pour finir ce chapitre revenons un instant sur la notion de tribu. Pourquoi avons nous besoin de cette notion alors qu'il semble que l'on pourrait prendre systématiquement l'ensemble des parties de l'univers comme tribu ?

Supposons que l'on veuille modéliser l'expérience aléatoire qui consiste à jeter une flechette uniformément sur le segment $[0, 1]$. On est alors tenté de considérer l'univers $\Omega = [0, 1]$, la tribu $\mathcal{F} = \mathcal{P}([0, 1])$ et une probabilité \mathbb{P} qui rende compte de l'uniformité du jet. En particulier, la probabilité de tomber dans un intervalle doit être proportionnelle à sa longueur et on souhaiterait donc que \mathbb{P} vérifie la propriété suivante :

$$\text{pour tout } 0 < a < b < 1, \quad \mathbb{P}(]a, b[) = b - a \quad (\star)$$

On peut cependant montrer qu'il n'existe pas d'application sur $\mathcal{P}([0, 1])$ vérifiant les deux points de la définition d'une probabilité et la propriété (\star) .

De manière plus générale, le théorème d'Ulam (que vous retrouverez sur wikipedia mais qui dépasse très largement le programme de ce cours et même de cours de niveaux bien plus élevés) nous dit qu'il n'existe pas de nombreuses probabilités qui semblent pourtant naturelles sur $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$. Il est donc dans ce cas nécessaire de renoncer à donner une probabilité à chaque partie de \mathbb{R} ; il faut se restreindre à des sous-tribus de $\mathcal{P}(\mathbb{R})$...et à introduire la tribu borélienne.

2 Variables aléatoires réelles

Nous supposons donné un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ modélisant notre expérience aléatoire. Bien souvent, ce n'est pas directement son résultat $\omega \in \Omega$ lui-même qui va nous intéresser, mais une certaine quantité $X(\omega) \in \mathbb{R}$ construite à partir de ce résultat. Dans le cas de deux lancers de dé par exemple, on pourra s'intéresser à la somme des deux chiffres obtenus, au plus grand des deux chiffres, etc. C'est à l'étude de ces "observables" que nous allons nous consacrer.

2.1 Définition et propriétés

Soit $X: \Omega \rightarrow \mathbb{R}$ une **fonction**. Puisque la réalisation ω est choisie au hasard, son image $X(\omega)$ sera aussi aléatoire, et nous allons chercher à estimer les chances pour que $X(\omega)$ tombe dans telle ou telle partie $B \in \mathcal{P}(\mathbb{R})$. De manière équivalente, cela revient à estimer les chances pour que ω lui-même tombe dans l'image réciproque de cette partie par X :

$$\{X \in B\} := X^{-1}(B) = \{\omega \in \Omega: X(\omega) \in B\}. \quad (2)$$

Remarque 6 (Notations). *Dans l'équation précédente il y a deux notations auxquelles il faut prêter attention.*

1. On prendra garde à ne pas confondre la notation X^{-1} avec la notation f^{-1} que l'on utilise pour la fonction réciproque d'une bijection f . Ici X (qui est bien une fonction même si la notation qui utilise une majuscule peut sembler peu familière!) n'a aucune raison d'être supposée bijective et $X^{-1}: \mathcal{P}(\mathbb{R}) \rightarrow \mathcal{P}(\Omega)$ associe à une partie de l'espace d'arrivée \mathbb{R} une partie de l'espace de départ Ω constituée de tous les antécédents des éléments de B .
2. Par ailleurs, pour plus de lisibilité et pour favoriser l'intuition probabiliste, on utilisera souvent la notation $\{X \in B\}$ pour $X^{-1}(B)$, et nous écrirons $\mathbb{P}(X \in B)$ plutôt que $\mathbb{P}(X^{-1}(B))$. Cette notation conforte l'idée que X représente un "nombre aléatoire", mais elle ne devra pas faire oublier qu'il s'agit en réalité d'une fonction.

Pour que notre question (2) ait un sens, il faut que $X^{-1}(B)$ soit un événement de notre tribu \mathcal{F} afin que l'on puisse lui attribuer une probabilité par \mathbb{P} , et ce pour toute partie $B \in \mathcal{P}(\mathbb{R})$ susceptible de nous intéresser. Comme expliqué plus haut, les parties "intéressantes" de \mathbb{R} seront toujours pour nous les boréliens, et nous sommes ainsi conduits à la définition importante suivante.

Définition 7 (Variable aléatoire réelle). *On appelle variable aléatoire réelle (on écrira v.a.r.) sur $(\Omega, \mathcal{F}, \mathbb{P})$ une application $X: \Omega \rightarrow \mathbb{R}$ qui vérifie la propriété de mesurabilité suivante :*

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad X^{-1}(B) \in \mathcal{F}. \quad (3)$$

Les fonctions qui satisfont (3) sont dites **mesurables** de (Ω, \mathcal{F}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Plus généralement, on dit d'une fonction f d'un espace muni d'une tribu (A, \mathcal{A}) et à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ qu'elle est mesurable, ou même si on veut préciser $\mathcal{A} - \mathcal{B}(\mathbb{R})$ -mesurable si

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad f^{-1}(B) \in \mathcal{A}.$$

Nous verrons plus loin la définition d'une fonction borélienne qui correspond au cas où $(A, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, voir la Définition 11.

Une variable aléatoire réelle n'est donc rien d'autre qu'une fonction mesurable d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

La probabilité qui figure dans l'espace de départ ne joue certes aucun rôle dans la mesurabilité mais on réserve tout de même le nom de *variable aléatoire* au cas où l'espace de départ est muni d'une probabilité.

Lorsqu'on a une variable aléatoire construite sur un espace de probabilité, on peut définir une nouvelle probabilité, cette fois sur l'espace d'arrivée $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, qui joue un rôle essentiel.

Définition 8 (Loi d'une variable aléatoire). *Soit X une variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$. L'application $\mathbb{P}_X: \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ définie pour tout $B \in \mathcal{B}(\mathbb{R})$ par*

$$\mathbb{P}_X(B) := \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B),$$

est bien définie et est une mesure de probabilité sur l'espace d'arrivée $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, appelée loi de X .

Remarque 7 (Notation encore). *Nous utiliserons la notation pratique $X \sim \mu$ pour dire que X a pour loi μ , c'est-à-dire $\mathbb{P}_X = \mu$.*

Preuve. Il faut montrer que la définition de la loi d'une variable aléatoire fait bien sens. Le fait que l'on puisse considérer l'image de $\{X \in B\}$ par la probabilité \mathbb{P} vient du fait que X est une variable aléatoire et donc $\{X \in B\} \in \mathcal{F}$. Il faut ensuite montrer que \mathbb{P}_X introduit dans la Définition 8 est bien une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ en vérifiant les deux points de la définition. Tout d'abord $\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(\Omega) = 1$. Par ailleurs pour toute suite $(A_n)_{n \geq 0}$ de boréliens, on note que $\{X \in \cup_{n \geq 0} A_n\} = \cup_{n \geq 0} \{X \in A_n\}$. Comme X est une variable aléatoire, pour tout $n \geq 0$, $\{X \in A_n\} \in \mathcal{F}$. Si on suppose de plus que les $(A_n)_{n \geq 0}$ sont disjoints 2 à 2 alors les $\{X \in A_n\} \in \mathcal{F}$ sont également 2 à 2 disjoints. On en déduit en utilisant la σ -additivité,

$$\mathbb{P}_X(\cup_{n \geq 0} A_n) = \mathbb{P}(X \in \cup_{n \geq 0} A_n) = \mathbb{P}(\cup_{n \geq 0} \{X \in A_n\}) = \sum_{n \geq 0} \mathbb{P}(X \in A_n) = \sum_{n \geq 0} \mathbb{P}_X(A_n).$$

□

Remarque 8 (Espace d'arrivée). *On peut généraliser les définitions ci-dessus en remplaçant l'espace d'arrivée \mathbb{R} par un ensemble arbitraire E , et la tribu borélienne $\mathcal{B}(\mathbb{R})$ par n'importe quelle tribu \mathcal{E} sur E . On dit alors que $X : (\Omega, \mathcal{F}, \mathbb{P}) \mapsto (E, \mathcal{E})$ est une variable aléatoire (à valeurs dans (E, \mathcal{E})) si c'est une fonction mesurable c'est-à-dire si pour $B \in \mathcal{E}$,*

$$\{X \in B\} := \{\omega \in \Omega, X(\omega) \in B\} \in \mathcal{F}.$$

On peut alors définir la loi de X comme la probabilité \mathbb{P}_X sur l'espace d'arrivée (E, \mathcal{E}) qui a tout événement $B \in \mathcal{E}$ associe

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

Cependant, c'est principalement le cas particulier des variables aléatoires réelles qui va nous occuper ce semestre. Quelques exemples pour $E : \mathbb{R}^2$, un espace de suites, de fonctions ou encore des arbres ou des cartes (et on obtient alors des arbres aléatoires comme les arbres de Galton Watson qui servent pour modéliser des généalogies ou des cartes aléatoires).

La mesure de probabilité \mathbb{P}_X résume tout ce qui pourra nous intéresser au sujet de la variable aléatoire X : elle décrit les chances pour que X “tombe” dans n'importe quel borélien $B \in \mathcal{B}(\mathbb{R})$. C'est un point important de la construction du modèle probabiliste. Les variables aléatoires sont des fonctions mais on ne s'intéresse pas réellement à la relation fonctionnelle c'est-à-dire à savoir quelle est l'image de chacun des points de l'espace de départ. On cherche plutôt à déterminer la probabilité que la fonction prenne telle ou telle valeur ou tombe dans tel ou tel ensemble de l'espace d'arrivée. La probabilité \mathbb{P}_X porte exactement cette information. On notera d'ailleurs que bien souvent l'espace de départ n'est même pas précisé. Il constitue une sorte de boîte noire ou de générateur de hasard que l'on n'explique pas. On se contente de considérer une variable et d'en préciser la loi : “Soit X une variable aléatoire de loi...”.

Spécifier $\mathbb{P}_X(B)$ pour tout $B \in \mathcal{B}(\mathbb{R})$ serait en pratique assez fastidieux (de la même façon qu'il était difficile de définir $\mathbb{P}(A)$ pour tout $A \in \mathcal{F}$), mais le Corollaire 1 nous autorise heureusement à nous restreindre aux boréliens de la forme $B =] - \infty, t]$ avec $t \in \mathbb{R}$. Nous sommes ainsi amenés à introduire l'objet suivant, qui jouera un rôle central dans l'étude de X .

Définition 9 (Fonction de répartition). *Soit X une variable aléatoire réelle. On appelle fonction de répartition de X la fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ définie comme suit : pour tout $t \in \mathbb{R}$,*

$$F_X(t) = \mathbb{P}_X(] - \infty, t]) = \mathbb{P}(X \leq t).$$

La motivation donnée ci-dessus est suffisamment importante pour mériter le nom de théorème.

Théorème 2 (La fonction de répartition caractérise la loi). *Soient X et Y deux variables aléatoires réelles. Alors les propositions suivantes sont équivalentes :*

1. X et Y ont même loi : $\mathbb{P}(X \in B) = \mathbb{P}(Y \in B)$ pour tout $B \in \mathcal{B}(\mathbb{R})$.
2. X et Y ont même fonction de répartition : $F_X(t) = F_Y(t)$ pour tout $t \in \mathbb{R}$.

Démonstration. L'implication $1 \Rightarrow 2$ est triviale, et la réciproque est assurée par le Corollaire 1. \square

Les fonctions de répartition ont des propriétés particulières, dont voici la description complète.

Théorème 3 (Caractérisation). *Une fonction $F: \mathbb{R} \rightarrow [0, 1]$ est la fonction de répartition d'une variable aléatoire réelle **si et seulement si** elle satisfait les trois propriétés suivantes :*

1. F est croissante : $s \leq t \implies F(s) \leq F(t)$.
2. F est continue à droite : pour tout $t \in \mathbb{R}$, on a $F(t+h) \rightarrow F(t)$ lorsque $h \rightarrow 0, h > 0$.
3. $F(t) \rightarrow 0$ lorsque $t \rightarrow -\infty$, et $F(t) \rightarrow 1$ lorsque $t \rightarrow +\infty$.

Démonstration. Nous nous contenterons de démontrer le sens facile (la réciproque étant par ailleurs réellement une question délicate), à savoir que toute fonction de répartition possède bien les trois propriétés ci-dessus.

1. Soit $s \leq t$. Comme $] -\infty, s] \subset] -\infty, t]$, on a $\mathbb{P}_X(] -\infty, s]) \leq \mathbb{P}_X(] -\infty, t])$ et donc $F(s) \leq F(t)$.
2. Soit $t \in \mathbb{R}$. Comme F est monotone, il suffit de montrer que la suite $(F(t+1/n))_{n \geq 1}$ converge vers $F(t)$. On remarque pour ça que $] -\infty, t] = \bigcap_{n \geq 1} \downarrow] -\infty, t+1/n]$ et on utilise la continuité décroissante pour en déduire que $(\mathbb{P}_X(\downarrow] -\infty, t+1/n])_{n \geq 1}$ converge vers $\mathbb{P}_X(] -\infty, t])$.
3. On a $\emptyset = \bigcap_{n \geq 1} \downarrow] -\infty, -n]$, donc en utilisant encore la continuité décroissante, $(F(-n))_{n \geq 1}$ converge vers 0. Cela suffit à montrer que $F(t) \rightarrow 0$ lorsque $t \rightarrow -\infty$ puisque F est croissante. Par ailleurs $\mathbb{R} = \bigcup_{n \geq 1} \uparrow] -\infty, n]$ et on en déduit, en utilisant la continuité croissante, que $(F(n))_{n \geq 1}$ converge vers 1. C'est suffisant pour montrer que $F(t) \rightarrow 1$ lorsque $t \rightarrow +\infty$ puisque F est croissante.

\square

Notons aussi que le fait que la fonction de répartition soit croissante assure qu'elle a une limite à droite et à gauche en tout point (exercice!). La fonction de répartition est donc *càdlàg* : elle est continue à droite avec une limite à gauche. Vous verrez souvent cette acronyme qui est utilisé même hors de France!

Le Théorème 2 assure que la fonction de répartition F_X caractérise $\mathbb{P}(X \in B)$ pour tout borélien $B \in \mathcal{B}(\mathbb{R})$ mais en pratique il est souvent difficile de déterminer la valeur de $\mathbb{P}(X \in B)$ à partir de F_X . Pour quelques boréliens simples, on peut cependant aisément trouver la probabilité recherchée :

Exercice 4. Soit X une variable aléatoire réelle, et $a \leq b$ deux réels. Établir les formules suivantes

1. $\mathbb{P}(X > a) = 1 - F_X(a)$.
2. $\mathbb{P}(X < a) = F_X(a-)$, où $F_X(a-)$ désigne la limite à gauche de F_X au point a .
3. $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$.
4. $\mathbb{P}(X = b) = F_X(b) - F_X(b-)$.
5. $\mathbb{P}(a < X < b) = F_X(b-) - F_X(a)$.

Un rapide corrigé :

1. Facile : passer au complémentaire!
2. On note que $] -\infty, a[= \cup_{n \geq 1} \uparrow] -\infty, a - 1/n]$ donc $\mathbb{P}_X(] -\infty, a[) = \lim_{n \rightarrow +\infty} \mathbb{P}_X(] -\infty, a - 1/n])$ et on conclut facilement puisque F_X est croissante.
3. On note que $]a, b] =] -\infty, b] \setminus] -\infty, a]$ et on conclut sans difficulté.
4. On a $\{b\} =] -\infty, b] \setminus] -\infty, b[$ et on conclut avec le point 2.
5. On a $]a, b[=] -\infty, b[\setminus] -\infty, a]$ et on conclut avec les points précédents.

On notera donc d'après le point 4. que les points de discontinuité de F_X correspondent aux « atomes » de \mathbb{P}_X c'est-à-dire aux $x \in \mathbb{R}$ tels que $\mathbb{P}(X = x) > 0$.

Nous introduisons maintenant deux classes particulières de variables aléatoires qui joueront un rôle important dans tout ce cours.

2.2 Variables discrètes

Un cas simple pour vérifier la mesurabilité est celui où l'ensemble $\text{Im}(X) := \{X(\omega) : \omega \in \Omega\}$ (rien à voir donc avec la notation d'algèbre linéaire : l'image n'a pas ici de structure algébrique particulière!) des valeurs prises par notre variable aléatoire X est dénombrable (rappelons que cela inclut en particulier le cas où cet ensemble est fini). Dans ce cas, il suffit de vérifier la mesurabilité (3) pour les singletons $B = \{x\}$, $x \in \text{Im}(X)$, qui sont bien des boréliens. En effet, une fois que l'on sait que $\{X = x\} \in \mathcal{F}$ pour tout $x \in \text{Im}(X)$, on peut en déduire que $\{X \in B\} \in \mathcal{F}$ pour tout borélien $B \in \mathcal{B}(\mathbb{R})$ en écrivant

$$\{X \in B\} = \bigcup_{x \in \text{Im}(X) \cap B} \{X = x\},$$

et en utilisant la stabilité de la tribu \mathcal{F} par réunion finie ou dénombrable. On notera bien que l'argument ne fonctionne plus si $\text{Im}(X)$ n'est pas dénombrable puisque l'union précédente n'est plus nécessairement dénombrable. Les variables aléatoires dont l'image est dénombrable sont appelées **variables discrètes**.

On s'intéresse maintenant à la caractérisation de leur loi. Comme les événements $(\{X = x\})_{x \in \text{Im}(X) \cap B}$ sont deux-à-deux disjoints, la σ -additivité de \mathbb{P} nous autorise à écrire

$$\mathbb{P}(X \in B) = \sum_{x \in \text{Im}(X) \cap B} \mathbb{P}(X = x). \quad (4)$$

Ainsi la loi d'une v.a.r. discrète X est caractérisée par :

1. l'ensemble $\text{Im}(X)$ des valeurs qu'elle peut prendre (qui est donc un ensemble dénombrable) ;
2. sa fonction de poids qui à toute valeur $x \in \text{Im}(X)$ associe la probabilité $\mathbb{P}(X = x)$ que cette valeur soit prise.

Les nombres $(\mathbb{P}(X = x))_{x \in \text{Im}(X)}$ ont une propriété importante : en prenant $B = \mathbb{R}$ dans (4), il vient

$$\sum_{x \in \text{Im}(X)} \mathbb{P}(X = x) = 1. \quad (5)$$

Réciproquement, si \mathcal{X} est un ensemble dénombrable de \mathbb{R} (mais on pourrait faire la même construction pour un ensemble dénombrable quelconque), étant donnée une famille dénombrable de nombres positifs $(p(x))_{x \in \mathcal{X}}$ vérifiant $\sum_{x \in \mathcal{X}} p(x) = 1$, on peut construire une v.a.r. discrète X vérifiant $\text{Im}(X) = \mathcal{X}$ et $\mathbb{P}(X = x) = p(x)$ pour tout $x \in \mathcal{X}$ en considérant par exemple l'espace probabilisé :

$$\Omega = \mathcal{X}, \quad \mathcal{F} = \mathcal{P}(\mathcal{X}), \quad \mathbb{P}: A \mapsto \sum_{x \in A} p(x),$$

et la variable aléatoire

$$X: (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

$$\omega \mapsto \omega.$$

On vérifie en effet bien que $\text{Im}(X) = \mathcal{X}$. La fonction X est bien une variable aléatoire puisque la tribu de départ est $\mathcal{P}(\Omega)$. Enfin pour tout $x \in \mathcal{X}$, $\mathbb{P}(X = x) = \mathbb{P}(\{x\}) = p(x)$. Nous avons donc ainsi complètement caractérisé les lois des variables aléatoires réelles discrètes.

Exemple 9 (Somme de deux dés). *On munit l'univers $\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\}$ de la tribu complète $\mathcal{F} = \mathcal{P}(\Omega)$, et de la loi uniforme \mathbb{P} définie en (1). Pour tout $\omega = (\omega_1, \omega_2) \in \Omega$, on pose*

$$X(\omega) := \omega_1 + \omega_2.$$

Alors l'application $X: \Omega \rightarrow \mathbb{R}$ ainsi définie est une variable aléatoire discrète sur Ω . L'ensemble des valeurs possibles est $X(\Omega) = \{2, 3, \dots, 12\}$ et l'on calcule aisément les probabilités associées :

x	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Voici les cinq lois discrètes les plus utilisées en pratique, à connaître sur le bout des doigts.

Exemple 10 (Lois discrètes usuelles). *Nous dirons qu'une variable aléatoire X suit la :*

1. *Loi uniforme sur \mathcal{X} (fini, non-vide), notée $\mathcal{U}(\mathcal{X})$, si $\text{Im}(X) = \mathcal{X}$ et*

$$\forall x \in \mathcal{X}, \quad \mathbb{P}_X(\{x\}) = \mathbb{P}(X = x) = \frac{1}{|\mathcal{X}|}.$$

2. *Loi de Bernoulli de paramètre $p \in [0, 1]$, notée $\mathcal{B}(p)$, si $\text{Im}(X) = \{0, 1\}$ et*

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p.$$

Autrement dit, l'application X est de la forme $X = \mathbf{1}_A$, avec $A \in \mathcal{F}$ et $\mathbb{P}(A) = p$.

3. *Loi binomiale de paramètres $n \in \mathbb{N}$ et $p \in [0, 1]$, notée $\mathcal{B}(n, p)$, si $\text{Im}(X) = \{0, 1, \dots, n\}$ et*

$$\forall k \in \{0, \dots, n\}, \quad \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Notons au passage que la somme des probabilités vaut bien 1, comme on le voit en prenant $a = p$ et $b = 1 - p$ dans la formule du binôme de Newton : pour tout $a, b \in \mathbb{C}$,

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}.$$

4. *Loi géométrique avec probabilité de succès $p \in]0, 1]$, notée $\mathcal{G}(p)$, si $\text{Im}(X) = \mathbb{N}^*$ et*

$$\forall k \in \mathbb{N}^*, \quad \mathbb{P}(X = k) = p(1-p)^{k-1}.$$

Le fait que la somme des probabilités vaut 1 s'obtient ici en prenant $x = 1 - p$ dans le développement suivant : pour tout $x \in \mathbb{C}$ tel que $|x| < 1$:

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k.$$

5. *Loi de Poisson de paramètre $\lambda \in]0, \infty[$, notée $\mathcal{P}(\lambda)$, si $\text{Im}(X) = \mathbb{N}$ et*

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}.$$

Le fait que la somme des probabilités vaut 1 s'obtient cette fois en prenant $x = \lambda$ dans le développement en série entière de la fonction exponentielle : pour tout $x \in \mathbb{C}$:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Essayons donner une intuition pour chacune de ces lois :

1. La loi uniforme modélise un tirage dans un ensemble fini où chaque élément à la même probabilité d'être tiré.
2. La loi de Bernoulli modélise le résultat d'un pile ou face avec une pièce biaisée.
3. La loi binomiale de paramètres $n \in \mathbb{N}$ et $p \in [0, 1]$ représente le nombre de pile dans une série de n lancers indépendants (nous verrons plus loin la définition de cette notion) avec une pièce biaisée de paramètre p .
4. La loi géométrique est le nombre de lancers qu'il faut dans le jeu précédent pour avoir le premier pile.
5. La loi de Poisson modélise le nombre de cas traitées dans une file d'attente ou un guichet pendant une durée fixée. Nous donnerons plus loin une explication plus précise sur ce point.

Exercice 5. *Expliciter la fonction de répartition de X lorsque $X \sim \mathcal{B}(p)$ et lorsque $X \sim \mathcal{U}(\{1, 2, \dots, n\})$.*

2.3 Variables à densité

Définition 10 (Densité). *Une densité (de probabilité) est une fonction $f: \mathbb{R} \rightarrow \mathbb{R}_+$ intégrable, avec*

$$\int_{-\infty}^{+\infty} f(x) dx = 1. \quad (6)$$

Dans ce cas, la fonction $F: \mathbb{R} \rightarrow [0, 1]$ définie par la formule

$$F(t) := \int_{-\infty}^t f(x) dx, \quad (7)$$

vérifie clairement les trois propriétés du Théorème 3, et est donc une fonction de répartition. Les v.a.r. dont la fonction de répartition est de cette forme seront dites à densité.

Ici, le terme « intégrable » doit être compris au sens du cours d'Analyse 3 (une fonction f continue par morceaux est intégrable si $\int f(x) dx < +\infty$). Vous verrez l'année prochaine une théorie plus puissante de l'intégration qui permet d'étendre cette définition.

Remarque 9 (Continuité). Si une v.a.r. X admet une densité f au sens ci-dessus, alors la représentation intégrale (7) implique (entre autres choses) que sa fonction de répartition F est continue. Au vu de l'Exercice 4, nous en déduisons en particulier que pour tout $x \in \mathbb{R}$,

$$\mathbb{P}(X = x) = F_X(x) - F_X(x-) = 0.$$

Les variables à densité sont donc radicalement différentes des variables discrètes introduites précédemment puisqu'elles ne mettent de « masse » sur aucun point. On dit qu'elles n'ont pas d'« atome ».

L'Exercice 4 implique également que pour tous réels $a \leq b$,

$$\mathbb{P}(a < X < b) = \mathbb{P}(X \in]a, b[) = \int_a^b f(x) dx,$$

et que la formule ne change pas si l'on remplace les inégalités strictes par des inégalités larges. De manière générale, vous verrez l'année prochaine une théorie de l'intégration qui permet de donner du sens à la formule suivante :

$$\mathbb{P}(X \in A) = \int_A f(x) dx = \int \mathbf{1}_A(x) f(x) dx \quad \text{pour tout } A \in \mathcal{B}(\mathbb{R}).$$

Exemple 11 (Densités usuelles). Voici les cinq lois à densité les plus utilisées en pratique.

1. Loi uniforme sur $]a, b[$ (avec $a, b \in \mathbb{R}$ et $a < b$), notée $\mathcal{U}(]a, b[)$:

$$f(x) = \frac{1}{b-a} \mathbf{1}_{]a, b[}(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in]a, b[\\ 0 & \text{sinon.} \end{cases}$$

2. Loi exponentielle de paramètre $\lambda > 0$, notée $\mathcal{E}(\lambda)$:

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{]0, \infty[}(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{sinon.} \end{cases}.$$

3. Loi gaussienne (ou normale) de moyenne $\mu \in \mathbb{R}$ et variance $\sigma^2 > 0$, notée $\mathcal{N}(\mu, \sigma^2)$:

$$f(x) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}.$$

La vérification de la condition (6) se déduit ici de la célèbre formule suivante, due à Gauss :

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}.$$

4. Loi Gamma de paramètres $r, \lambda > 0$, notée $\Gamma(r, \lambda)$.

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} \mathbf{1}_{]0, \infty[}(x).$$

La condition (6) découle ici de la définition même de la fonction $\Gamma:]0, \infty[\rightarrow]0, \infty[$:

$$\Gamma(r) := \int_0^{+\infty} x^{r-1} e^{-x} dx.$$

Notons que la loi Gamma est une généralisation de la loi exponentielle (prendre $r = 1$).

5. Loi de Cauchy de paramètre $\lambda > 0$, notée $\mathcal{C}(\lambda)$:

$$f(x) = \frac{\lambda}{\pi(\lambda^2 + x^2)}.$$

On pourra vérifier que l'intégrale de cette fonction vaut 1 en dérivant la fonction \arctan .

Assurez vous de savoir tracer les graphes de ces différentes fonctions de densité et notamment de comparer leur décroissance à l'infini.

Exercice 6. Calculer la fonction de répartition de X pour $X \sim \mathcal{U}(]a, b[)$, $X \sim \mathcal{E}(\lambda)$ et $X \sim \mathcal{C}(\lambda)$.

2.4 Opérations usuelles

Nous serons amenés à considérer diverses opérations naturelles sur les v.a.r. : somme, produit, limite, composition par certaines applications, etc. Quelques précautions s'imposent afin d'assurer la propriété de mesurabilité (3). Nous commençons par un lemme très utile.

Lemme 2 (Critère de mesurabilité). Soit $X: \Omega \rightarrow \mathbb{R}$ une fonction et \mathcal{C} un ensemble de parties de \mathbb{R} tel que $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$. Pour que X soit mesurable, il suffit que pour tout $B \in \mathcal{C}$,

$$\{X \in B\} \in \mathcal{F}.$$

Démonstration. On définit $\mathcal{G} := \{B \in \mathcal{B}(\mathbb{R}) : \{X \in B\} \in \mathcal{F}\}$. On veut donc montrer que $\mathcal{G} = \mathcal{B}(\mathbb{R})$. On commence par vérifier que \mathcal{G} est une tribu :

1. $\{X \in \emptyset\} = \emptyset \in \mathcal{F}$;
2. Soit $B \in \mathcal{G}$. On a $\{X \in B^c\} = \{X \in B\}^c \in \mathcal{F}$ par stabilité par passage au complémentaire. Donc $B^c \in \mathcal{G}$;
3. Soit $(B_n)_{n \geq 1}$ une suite de parties appartenant à \mathcal{G} . On a alors $\{X \in \cup_{n \geq 1} B_n\} = \cup_{n \geq 1} \{X \in B_n\} \in \mathcal{F}$ par stabilité par union dénombrable. Donc $\cup_{n \geq 1} B_n \in \mathcal{G}$.

On a donc prouvé que \mathcal{G} est une tribu et comme de plus $\mathcal{C} \subset \mathcal{G}$, on en déduit $\sigma(\mathcal{C}) \subset \mathcal{G}$. \square

Proposition 3 (sup, inf, lim). *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles. Alors, les applications suivantes sont des variables aléatoires réelles dès lors qu'elles sont bien définies :*

$$\begin{aligned} \omega \mapsto \sup_{n \geq 1} X_n(\omega), & \quad \omega \mapsto \inf_{n \geq 1} X_n(\omega), & \quad \omega \mapsto \lim_{n \rightarrow \infty} X_n(\omega) \\ \omega \mapsto \limsup_{n \geq 1} X_n(\omega), & \quad \omega \mapsto \liminf_{n \geq 1} X_n(\omega). \end{aligned}$$

Démonstration. Commençons par le cas $X(\omega) = \sup_{n \geq 1} X_n(\omega)$. Pour tout $t \in \mathbb{R}$, on peut écrire

$$\{X \leq t\} = \bigcap_{n \geq 1} \{X_n \leq t\}. \quad (8)$$

Pour chaque $n \geq 1$, X_n est une variable aléatoire réelle, donc $\{X_n \leq t\} \in \mathcal{F}$. Comme \mathcal{F} est stable par intersection dénombrable, nous en déduisons que $\{X \leq t\} \in \mathcal{F}$. C'est vrai pour tout $t \in \mathbb{R}$, et on conclut en appliquant le Lemme 2 à l'ensemble \mathcal{C} des intervalles de la forme $] - \infty, t], t \in \mathbb{R}$. Le cas où $X(\omega) = \inf_{n \geq 1} X_n(\omega)$ s'obtient en remplaçant \leq par \geq et $] - \infty, t]$ par $[t, +\infty[$ dans l'argument ci-dessus.

On en déduit que $\limsup_{n \geq 1} X_n = \inf_{N \geq 1} \sup_{n \geq N} X_n$ et $\liminf_{n \geq 1} X_n = \sup_{N \geq 1} \inf_{n \geq N} X_n$ sont également des variables aléatoires. Enfin, lorsque $X(\omega) = \lim_{n \rightarrow \infty} X_n(\omega)$, il suffit de noter que

$$X = \liminf_{n \geq 1} X_n, \quad (\text{ou d'ailleurs } \limsup_{n \geq 1} X_n) \quad (9)$$

(voir l'Exercice 6 de la première feuille de TD), puis d'utiliser ce que l'on vient de prouver. \square

Étant données une variable aléatoire réelle X et une fonction $h: \mathbb{R} \rightarrow \mathbb{R}$, il est naturel de chercher à définir une nouvelle variable aléatoire réelle $h(X)$ en les composant comme suit :

$$\begin{aligned} h(X): \Omega & \rightarrow \mathbb{R} \\ \omega & \mapsto h(X(\omega)). \end{aligned}$$

Cependant, pour que cette application soit bien une v.a.r., il faut imposer une condition sur h .

Définition 11 (Fonction borélienne). *Une fonction $h: \mathbb{R} \rightarrow \mathbb{R}$ est dite borélienne si elle vérifie*

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad \{h \in B\} = h^{-1}(B) \in \mathcal{B}(\mathbb{R}). \quad (10)$$

En reprenant le vocabulaire que nous avons introduit pour définir une variable aléatoire, on

peut dire qu'une fonction borélienne est une fonction *mesurable* de $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Proposition 4 (Composition). *Si X est une variable aléatoire réelle et h une fonction borélienne alors $h(X)$ est une variable aléatoire réelle.*

Démonstration. Soit $B \in \mathcal{B}(\mathbb{R})$. On a $\{h(X) \in B\} = \{X \in h^{-1}(B)\}$. Or, comme h est borélienne $h^{-1}(B) \in \mathcal{B}(\mathbb{R})$ et, comme X est une variable aléatoire, $\{X \in h^{-1}(B)\} \in \mathcal{F}$. On en déduit que $h(X)$ est bien une variable aléatoire. \square

Exercice 7 (Exemples importants). *Montrer que les fonctions suivantes sont boréliennes :*

1. Une fonction $h: \mathbb{R} \rightarrow \mathbb{R}$ continue. [indication : utiliser que $\mathcal{B}(\mathbb{R}) = \sigma(O, O \text{ ouvert de } \mathbb{R})$]
2. Une fonction $h: \mathbb{R} \rightarrow \mathbb{R}$ monotone (croissante ou décroissante).
3. Une fonction indicatrice $h = \mathbf{1}_B$, avec $B \in \mathcal{B}(\mathbb{R})$.
4. La composée de deux fonctions boréliennes.
5. Le supremum, l'infimum et la limite d'une suite de fonctions boréliennes (lorsqu'ils existent).

On pourra pour cet exercice utiliser le critère de mesurabilité établi au Lemme 2 et qui se reformule ici de la façon suivante :

Lemme 3 (Critère de mesurabilité). *Soit $h: \mathbb{R} \rightarrow \mathbb{R}$ une fonction et \mathcal{C} un ensemble de parties de \mathbb{R} tel que $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$. Pour que h soit borélienne, il suffit que pour tout $B \in \mathcal{C}$,*

$$\{h \in B\} \in \mathcal{B}(\mathbb{R}).$$

On peut enfin combiner plusieurs variables aléatoires grâce au résultat suivant (admis), qui montre en particulier que la somme et le produit de deux variables aléatoires réelles sont encore des variables aléatoires réelles (prendre $F(x, y) = x + y$ ou $F(x, y) = xy$). On rappelle auparavant la définition d'une fonction continue de \mathbb{R}^n dans \mathbb{R} :

Définition 12. *On dit d'une fonction $F: \mathbb{R}^n \rightarrow \mathbb{R}$ qu'elle est **continue** si pour tout $x \in \mathbb{R}^n$ et tout $\varepsilon > 0$ il existe $\alpha > 0$ tel que pour tout $z \in \mathbb{R}^n$:*

$$\|z - x\| < \alpha \implies |F(z) - F(x)| < \varepsilon.$$

La notation $\|z\|$ désigne la norme euclidienne de z définie par

$$\|z\| = \sqrt{\langle z, z \rangle} = \sqrt{\sum_{i=1}^n z_i^2},$$

où $\langle \cdot, \cdot \rangle$ est le produit scalaire canonique sur \mathbb{R}^n .

On ne demande pas dans ce cours de maîtriser cette notion que vous étudierez en détail dans le cours d'analyse du prochain semestre. On admet donc pour le moment que les fonctions de plusieurs variables que l'on sera amené à considérer sont bien continues. Cela nous permettra d'utiliser le résultat suivant :

Proposition 5 (Fonctions de plusieurs variables). *Soient X_1, \dots, X_n des variables aléatoires réelles sur $(\Omega, \mathcal{F}, \mathbb{P})$, et soit $F: \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue. Alors l'application*

$$\omega \mapsto F(X_1(\omega), \dots, X_n(\omega)) \quad (11)$$

est une variable aléatoire réelle, que nous noterons simplement $F(X_1, \dots, X_n)$.

Exercice 8 (Indistingabilité). *Soient X, Y des v.a.r. sur $(\Omega, \mathcal{F}, \mathbb{P})$. Justifier que $\{X = Y\}$ est un événement, puis montrer que si $X = Y$ p.s. (i.e. $\mathbb{P}(X = Y) = 1$), alors X et Y ont même loi.*

Corrigé. On note tout d'abord que $\{X = Y\} = \{X - Y = 0\}$. Comme $(x, y) \mapsto x - y$ est continue, $X - Y$ est bien une variable aléatoire et $\{X - Y = 0\} = \{X - Y \in \{0\}\}$ appartient à \mathcal{F} . Si $X = Y$ p.s. alors pour tout $B \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}(Y \in B) = \mathbb{P}(Y \in B, X = Y) + \mathbb{P}(Y \in B, X \neq Y) = \mathbb{P}(X \in B, X = Y) + \mathbb{P}(Y \in B, X \neq Y).$$

Or $\mathbb{P}(Y \in B, X = Y) = \mathbb{P}(X \in B, X = Y) = \mathbb{P}(X \in B)$ car $\mathbb{P}(X = Y) = 1$. Et, par ailleurs, $\mathbb{P}(Y \in B, X \neq Y) \leq \mathbb{P}(X \neq Y) = 0$. On en déduit bien que

$$\mathbb{P}_Y(B) = \mathbb{P}(Y \in B) = \mathbb{P}(X \in B) = \mathbb{P}_X(B).$$

3 Espérance

Ce chapitre est consacré à une quantité très importante que l'on peut associer à une variable aléatoire réelle X : son espérance mathématique $\mathbb{E}[X]$. Intuitivement, ce nombre représente la valeur théorique moyenne autour de laquelle la variable aléatoire X va fluctuer. C'est la meilleure estimation (voir l'Exercice 11) que l'on puisse donner de X . C'est la valeur que l'on peut *espérer*. La définition formelle est un cas particulier de l'intégrale d'une fonction par rapport à une mesure, théorie qui ne sera développée que l'an prochain. Nous nous contentons cette année d'une présentation allégée.

3.1 Définition formelle

L'espérance d'une v.a.r. X sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ est définie en trois étapes :

1. (Variables étagées) On commence par le cas simple où la variable X est *étagée*, c'est-à-dire qu'elle ne prend qu'un nombre fini de valeurs. Autrement dit, on a

$$X = \sum_{i=1}^n a_i \mathbf{1}_{A_i}, \quad (12)$$

avec $n \geq 0$ un entier, a_1, \dots, a_n des réels (2 à 2 distincts), et A_1, \dots, A_n des événements de la tribu \mathcal{F} (on a donc pour tout $1 \leq i \leq n$, $A_i = \{X = a_i\}$). On note que la variable X est donc une variable qui vaut a_i avec probabilité $\mathbb{P}(A_i)$ ($i = 1, \dots, n$). On rappelle que, dans ce cas la loi \mathbb{P}_X de X est caractérisée par :

- (a) l'image de X , les réels a_1, \dots, a_n ,
- (b) pour tout i , la probabilité $\mathbb{P}(X = a_i) = \mathbb{P}(A_i)$.

Dans ce cas, on définit l'espérance de X par :

$$\mathbb{E}[X] := \sum_{i=1}^n a_i \mathbb{P}(A_i). \quad (13)$$

Ce cas discret fini correspond bien à l'intuition que l'on a de l'espérance (ou de la moyenne) c'est-à-dire une somme des a_i , $i = 1 \dots, n$ pondérés par les poids $\mathbb{P}(X = a_i)$, $i = 1 \dots, n$. La construction de l'espérance mathématique permet de généraliser ce cas à toutes les variables aléatoires (pour lesquelles cette notion d'espérance a du sens). On note ici (mais c'est vrai en général) que l'espérance ne dépend de X qu'à travers sa loi \mathbb{P}_X c'est-à-dire que deux variables qui ont même loi ont même espérance.

2. (Variables positives) On considère ensuite le cas où la v.a.r. X est positive, c'est-à-dire telle que $\text{Im}(X) \subseteq [0, \infty[$. Dans ce cas, on définit $\mathbb{E}[X] \in [0, \infty]$ ainsi :

$$\mathbb{E}[X] := \sup \{ \mathbb{E}[Z] : Z \text{ étagée}, Z \leq X \}. \quad (14)$$

Cette quantité est toujours positive (prendre $Z = 0$ dans la définition ci-dessus), mais attention : elle peut très bien être égale à $+\infty$! Il faudra bien garder cela en tête. Il y a derrière cette définition l'idée qu'une variable aléatoire positive peut être bien approximée par une variable étagée : la Remarque 11 plus loin est une formulation précise de cette idée.

3. (Variables intégrables) Enfin, dans le cas d'une variable aléatoire réelle X quelconque, on écrit

$$X = X^+ - X^-, \quad (15)$$

avec $X^+ := \max(X, 0)$ et $X^- := \max(-X, 0)$. Comme X^+, X^- sont des variables aléatoires positives, les quantités $\mathbb{E}[X^+], \mathbb{E}[X^-]$ sont bien définies, et on aimerait poser

$$\mathbb{E}[X] := \mathbb{E}[X^+] - \mathbb{E}[X^-]. \quad (16)$$

Attention cependant : cette définition n'a pas de sens si $\mathbb{E}[X^+] = \mathbb{E}[X^-] = +\infty$. Nous dirons donc que la variable aléatoire réelle X est **intégrable** si $\mathbb{E}[X^+] < \infty$ et $\mathbb{E}[X^-] < \infty$ et dans ce cas seulement, nous définissons son espérance par la formule (16). On peut vérifier que la condition $(\mathbb{E}[X^+] < \infty \text{ et } \mathbb{E}[X^-] < \infty)$ est équivalente à $\mathbb{E}[|X|] < +\infty$ car $|X| = X^+ + X^-$.

Nous n'avons donc le droit d'écrire $\mathbb{E}[X]$ que dans deux cas : quand X est positive ou quand X est intégrable, c'est-à-dire $\mathbb{E}[|X|] < +\infty$.

Nous n'allons pas aller beaucoup plus loin dans la description de la construction de l'espérance. Revenons seulement pour conclure cette présentation sur l'idée qu'une variable positive peut être approximée par une suite de variables étagées. En effet **toute v.a.r. positive X est la limite simple d'une suite croissante $(X_n)_{n \geq 0}$ de v.a.r. positives et étagées**, par exemple la suite définie pour tout $n \geq 1$ par

$$X_n = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbf{1}_{X \in [\frac{k}{2^n}, \frac{k+1}{2^n}[}$$

La variable X_n est simplement la variable qui prend la valeur $k/2^n$ sur l'événement $\{X \in [\frac{k}{2^n}, \frac{k+1}{2^n}[}$ pour $k = 0, \dots, n2^n - 1$ et 0 sur l'événement $\{X \geq n\}$. On notera que pour $\omega \in \Omega$ et $n \geq 1$ fixé, un seul terme de la somme ci-dessus est non nul (mais l'indice correspondant dépend de ω bien

sûr). Nous verrons plus loin que $\mathbb{E}[X] = \lim_{n \rightarrow \infty} \uparrow \mathbb{E}[X_n]$. Cela fournit une intuition pour le point 2. de la construction de l'espérance. Quand la variable ne prend qu'un nombre fini de valeurs, l'espérance est la moyenne pondérée ; quand elle est positive on peut l'approximer par une suite de variables étagées et prendre la limite de la suite des espérances. Il y a bien sûr quelques précautions techniques à prendre pour mener à bien la construction mais les idées sont là !

3.2 Propriétés fondamentales

Avant d'énoncer quelques propositions, un petit point de vocabulaire : on dit d'une propriété relative à $\omega \in \Omega$ qu'elle est vraie **presque sûrement** (et on note souvent **p.s.**) si l'ensemble des ω pour lesquels elle est vraie est de probabilité 1. De façon équivalente on dit d'une propriété qu'elle est vraie p.s. si les ω qui ne la vérifient pas forment un ensemble de probabilité nulle. Par exemple on dit que X est positive p.s. si

$$\mathbb{P}(X \geq 0) = \mathbb{P}(\omega \in \Omega \text{ t.q. } X(\omega) \geq 0) = 1 \text{ ou, de façon équivalente } \mathbb{P}(X < 0) = 0.$$

Autre exemple : on dit que $(X_n)_{n \geq 0}$ converge p.s. vers X si

$$\mathbb{P}(\omega \in \Omega \text{ t.q. } X_n(\omega) \rightarrow X(\omega)) = 1.$$

La convergence p.s. est donc très proche de la convergence simple à ceci près qu'il peut ne pas y avoir convergence en certains points de Ω et que ces points forment un événement de probabilité nulle. Un dernier exemple : on dit que $(X_n)_{n \geq 0}$ est croissante p.s. si

$$\mathbb{P}(\omega \text{ t.q. } (X_n(\omega))_{n \geq 0} \text{ est croissante}) = 1.$$

De la définition de l'espérance que l'on a vue en 3.1, on peut déduire les propriétés suivantes (admisses), très utiles en pratique.

Proposition 6 (Propriétés fondamentales de l'espérance).

1. (Espérance d'une indicatrice) Pour tout événement $A \in \mathcal{F}$, on a

$$\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A). \tag{17}$$

Ce point est une conséquence directe de la définition de l'espérance pour des variables étagées.

2. (Linéarité) Soient X, Y des v.a.r. intégrables sur $(\Omega, \mathcal{F}, \mathbb{P})$, et λ un réel. Alors $X + \lambda Y$ est

une v.a.r. intégrable, et on a

$$\mathbb{E}[X + \lambda Y] = \mathbb{E}[X] + \lambda \mathbb{E}[Y]. \quad (18)$$

De plus, cette identité reste valable sans hypothèse d'intégrabilité lorsque $X, Y, \lambda \geq 0$.

3. (Monotonie) Soient X, Y des v.a.r. positives ou intégrables sur $(\Omega, \mathcal{F}, \mathbb{P})$. Alors,

$$X \leq Y \quad p.s. \implies \mathbb{E}[X] \leq \mathbb{E}[Y]. \quad (19)$$

On peut formuler cette propriété de façon équivalente :

$$X \geq 0 \quad p.s. \implies \mathbb{E}[X] \geq 0. \quad (20)$$

4. (Valeur absolue) Une v.a.r. X est intégrable si et seulement si $\mathbb{E}[|X|] < \infty$, auquel cas

$$|\mathbb{E}[X]| \leq \mathbb{E}[|X|]. \quad (21)$$

Les deux points suivants sont fondamentaux : ils permettent de passer de la convergence p.s. à la convergence des espérances au prix à chaque fois d'une hypothèse additionnelle :

5. (Théorème de convergence monotone) Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. convergeant p.s. vers X . On suppose de plus que

(a) toutes les variables $(X_n)_{n \geq 1}$ sont positives p.s.

(b) $(X_n)_{n \geq 1}$ est une suite croissante p.s.

Alors,

$$\mathbb{E}[X_n] \xrightarrow{n \rightarrow \infty} \mathbb{E}[X]. \quad (22)$$

On notera que l'on peut bien considérer les espérances ci-dessus puisque les variables considérées sont positives.

6. (Théorème de convergence dominée) Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. convergeant p.s. vers X . On suppose de plus qu'il existe une v.a.r. **intégrable** Y telle que

$$|X_n| \leq Y \quad p.s. \text{ pour tout } n \geq 1.$$

Alors X et les $X_n, n \geq 1$ sont intégrables et

$$\mathbb{E}[X_n] \xrightarrow{n \rightarrow \infty} \mathbb{E}[X]. \quad (23)$$

Remarque 10. Du point 3 dans la proposition précédente nous déduisons que si une variable aléatoire X est bornée p.s. c'est-à-dire s'il existe $M \in \mathbb{R}$ tel que $|X| \leq M$ p.s. alors X est intégrable (car $\mathbb{E}[|X|] \leq \mathbb{E}[M] = M < +\infty$).

Rappelons maintenant le lemme d'approximation étagée et tirons en quelques conséquences :

Remarque 11 (Approximation étagée). Toute v.a.r. positive X est la limite simple d'une suite croissante $(X_n)_{n \geq 0}$ de v.a.r. positives et étagées, par exemple

$$X_n = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbf{1}_{X \in [\frac{k}{2^n}, \frac{k+1}{2^n}[}$$

La variable X_n est simplement la variable qui prend la valeur $k/2^n$ sur l'événement $\{X \in [\frac{k}{2^n}, \frac{k+1}{2^n}[}$ pour $k = 0, \dots, n2^n - 1$ et 0 sur l'événement $\{X \geq n\}$. En utilisant la propriété 5 ci-dessus ainsi que la définition de l'espérance dans le cas des variables étagées, on en déduit que

$$\begin{aligned} \mathbb{E}[X] &= \lim_{n \rightarrow \infty} \uparrow \mathbb{E}[X_n] \\ &= \lim_{n \rightarrow \infty} \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbb{P}_X \left(\left[\frac{k}{2^n}, \frac{k+1}{2^n} \right] \right). \end{aligned} \tag{24}$$

Une conséquence importante : si X est une v.a.r. quelconque, on peut appliquer (24) à X^+ et X^- pour en déduire que **l'existence de $\mathbb{E}[X]$ et sa valeur sont entièrement déterminées par \mathbb{P}_X** . Autrement dit si X et Y sont deux v.a.r de même loi alors elles ont même espérance. En particulier, modifier une v.a.r. sur un événement négligeable ne modifie pas sa loi (Exercice 8), et ne modifie donc pas non-plus son espérance (Remarque 11) : si $X = Y$ p.s. alors $\mathbb{E}(X) = \mathbb{E}(Y)$.

Attention, la réciproque est fautive ! Deux variables qui ont même espérance n'ont pas forcément même loi : par exemple la variable qui prend la valeur 1 avec probabilité 1 a pour espérance 1. Une variable qui prend pour valeur 2 ou 0 avec probabilité 1/2 a également pour espérance 1. Pour construire une réciproque il faut donc demander plus que l'égalité des espérances. C'est l'objet du résultat suivant qui montre que la loi \mathbb{P}_X est à son tour déterminée par la donnée des espérances de toutes les fonctions "raisonnables" de X . C'est une caractérisation importante de la loi d'une variable qu'il faut maîtriser.

Proposition 7 (Espérances et loi). Pour des v.a.r. X, Y , les conditions suivantes sont équivalentes :

1. X et Y ont même loi.
2. $\mathbb{E}[h(X)] = \mathbb{E}[h(Y)]$ pour toute fonction $h: \mathbb{R} \rightarrow \mathbb{R}$ borélienne positive.
3. $\mathbb{E}[h(X)] = \mathbb{E}[h(Y)]$ pour toute fonction $h: \mathbb{R} \rightarrow \mathbb{R}$ continue bornée.

Remarque 12. *D'autres classe de fonction feraient l'affaire dans la proposition précédente. On notera que dans le point 2 l'espérance est bien définie puisque la variable $h(X)$ est positive; dans le point 3 l'espérance est bien définie car $h(X)$ est bornée donc intégrable.*

Démonstration. (1 \implies 2) Si X est une v.a.r. et h une fonction borélienne positive, alors la loi de $h(X)$ est déterminée par celle de X . En effet, pour tout $B \in \mathcal{B}(\mathbb{R})$, on peut écrire

$$\mathbb{P}_{h(X)}(B) = \mathbb{P}(h(X) \in B) = \mathbb{P}(X \in h^{-1}(B)) = \mathbb{P}_X(h^{-1}(B)).$$

Ainsi, l'égalité $\mathbb{P}_X = \mathbb{P}_Y$ implique $\mathbb{P}_{h(X)} = \mathbb{P}_{h(Y)}$ et donc $\mathbb{E}[h(X)] = \mathbb{E}[h(Y)]$ par la Remarque 11. (2 \implies 3) Soit $h: \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue et bornée. Alors h_+ et h_- sont boréliennes (car continues) et positives (par construction), donc l'hypothèse 2 nous assure que

$$\mathbb{E}[h^+(X)] = \mathbb{E}[h^+(Y)] \quad \text{et} \quad \mathbb{E}[h^-(X)] = \mathbb{E}[h^-(Y)].$$

Ces espérances étant finies (h bornée), la décomposition $h = h^+ - h^-$ entraîne $\mathbb{E}[h(X)] = \mathbb{E}[h(Y)]$. (3 \implies 1) Fixons $t \in \mathbb{R}$, et montrons que $F_X(t) = F_Y(t)$ (cela suffira, puisque la fonction de répartition caractérise la loi). Pour cela, on aimerait appliquer l'hypothèse à la fonction $h = \mathbf{1}_{]-\infty, t]}$ puisque $\mathbb{E}(h(X)) = \mathbb{P}(X \leq t)$, mais h n'est pas continue. On va donc l'approximer par une suite de fonctions continues et utiliser un théorème de convergence. Pour tout $n \geq 1$, on définit donc la fonction $h_n: \mathbb{R} \rightarrow [0, 1]$ qui vaut 1 sur $]-\infty, t]$, 0 sur $[t + \frac{1}{n}, +\infty[$, et qui décroît linéairement de 1 à 0 sur $[t, t + \frac{1}{n}]$. Comme h_n est continue et bornée, l'hypothèse assure que pour tout $n \geq 1$,

$$\mathbb{E}[h_n(X)] = \mathbb{E}[h_n(Y)]. \tag{25}$$

Il nous reste encore à passer à la limite $n \rightarrow \infty$ dans cette égalité. Comme pour tout $x \in \mathbb{R}$, $(h_n(x))_{n \geq 1}$ converge vers $h(x)$, on a aussi $(h_n(X))_{n \geq 1}$ converge vers $h(X)$ et, comme de plus pour tout $n \geq 1$, $|h_n(X)| \leq 1$ on peut utiliser le théorème de convergence dominée pour obtenir

$$\mathbb{E}[h(X)] = \mathbb{E}[h(Y)].$$

Comme $h = \mathbf{1}_{]-\infty, t]}$, cette identité n'est rien d'autre que $F_X(t) = F_Y(t)$. □

3.3 Cas particuliers : variables discrètes ou à densité

Dans le cas où la v.a.r. est discrète ou à densité, l'espérance $\mathbb{E}[X]$ s'exprime plus facilement et parfois même se calcule aisément. Les résultats de cette partie sont très importants et utiles en pratique pour les calculs. Ils doivent être parfaitement maîtrisés.

Proposition 8 (Cas discret). *Soit X une v.a.r. discrète. Alors on a toujours*

$$\mathbb{E}[|X|] = \sum_{x \in \text{Im}(X)} |x| \mathbb{P}(X = x). \quad (26)$$

De plus, X est intégrable si et seulement si cette somme est finie, auquel cas on a

$$\mathbb{E}[X] = \sum_{x \in \text{Im}(X)} x \mathbb{P}(X = x). \quad (27)$$

Plus généralement, pour toute fonction borelienne $h: \text{Im}(X) \rightarrow \mathbb{R}$, telle que $h \geq 0$ ou $h(X)$ intégrable (par exemple si h est bornée),

$$\mathbb{E}[h(X)] = \sum_{x \in \text{Im}(X)} h(x) \mathbb{P}(X = x). \quad (28)$$

On notera que pour savoir si $h(X)$ est intégrable on peut calculer $\mathbb{E}[|h(X)|] = \sum_{x \in \text{Im}(X)} |h(x)| \mathbb{P}(X = x)$ en appliquant la formule précédente à la fonction positive $|h|$.

Démonstration. On se contente de montrer la dernière égalité dans le cas où h est positive. On note $\text{Im}(X) = \{a_k, k \geq 1\}$. On a donc $X = \sum_{k=1}^{+\infty} a_k 1_{X=a_k}$ et $h(X) = \sum_{k=1}^{+\infty} h(a_k) 1_{X=a_k}$. Pour tout $n \geq 1$ on définit la variable étagée $Y_n = \sum_{k=1}^n h(a_k) 1_{X=a_k}$. Comme $(Y_n)_{n \geq 1}$ est une suite positive croissante et convergeant vers $h(X)$ on a, par le théorème de convergence monotone,

$$\mathbb{E}(Y_n) \rightarrow \mathbb{E}(h(X)).$$

Or pour tout $n \geq 1$, la variable Y_n est étagée et on a donc

$$\mathbb{E}(Y_n) = \sum_{k=1}^n h(a_k) \mathbb{P}(X = a_k),$$

et on déduit que $\mathbb{E}(Y_n)$ converge vers $\sum_{k=1}^{+\infty} h(a_k) \mathbb{P}(X = a_k)$. □

Proposition 9 (Cas à densité). *Soit X une v.a.r. de densité f . Alors*

$$\mathbb{E}[|X|] = \int_{-\infty}^{+\infty} |x| f(x) dx. \quad (29)$$

De plus, X est intégrable si et seulement si l'intégrale ci-dessus est finie, auquel cas

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x) dx. \quad (30)$$

Plus généralement, pour toute fonction borélienne $h: \mathbb{R} \rightarrow \mathbb{R}$, telle que $h \geq 0$ ou $h(X)$ intégrable (par exemple si h est bornée),

$$\mathbb{E}[h(X)] = \int_{-\infty}^{+\infty} h(x)f(x)dx \quad (31)$$

On remarque que cette formule permet de déterminer si $h(X)$ est intégrable puisque $|h|$ étant positif, $\mathbb{E}[|h(X)|] = \int_{-\infty}^{+\infty} |h(x)|f(x)dx$.

Réciproquement, si une v.a.r. X admet la représentation intégrale (31) pour toute fonction $h: \mathbb{R} \rightarrow \mathbb{R}$ continue bornée (ou pour toute fonction $h: \mathbb{R} \rightarrow \mathbb{R}$ borélienne positive), alors X admet f pour densité. On vérifie en effet que sa fonction de répartition est bien celle d'une variable de densité f en prenant h de la forme $1_{]-\infty, t]}$, $t \in \mathbb{R}$. C'est un point à retenir et une méthode très efficace pour montrer qu'une variable est à densité et pour déterminer cette densité.

Exercice 9 (Lois usuelles). Dans chacun des cas suivants, calculer $\mathbb{E}[X]$ si cette quantité existe.

1. $X \sim \mathcal{U}(\{1, \dots, n\})$ avec $n \in \mathbb{N}^*$.
2. $X \sim \mathcal{B}(p)$ avec $p \in [0, 1]$.
3. $X \sim \mathcal{B}(n, p)$ avec $n \in \mathbb{N}$ et $p \in [0, 1]$.
4. $X \sim \mathcal{G}(p)$ avec $p \in]0, 1]$.
5. $X \sim \mathcal{P}(\lambda)$ avec $\lambda \in]0, \infty[$.
6. $X \sim \mathcal{U}([a, b])$ avec $a, b \in \mathbb{R}$ et $a < b$.
7. $X \sim \mathcal{E}(\lambda)$ avec $\lambda \in]0, \infty[$.
8. $X \sim \mathcal{N}(\mu, \sigma^2)$ avec $\mu \in \mathbb{R}$ et $\sigma^2 > 0$.
9. $X \sim \Gamma(\lambda, r)$ avec $\lambda, r \in]0, \infty[$.
10. $X \sim \mathcal{C}(\lambda)$ avec $\lambda \in]0, \infty[$.

3.4 Fluctuations autour de l'espérance

L'espérance $\mathbb{E}[X]$ représente la valeur centrale autour de laquelle la v.a.r. X va fluctuer. Nous allons ici préciser cette intuition à l'aide de deux célèbres inégalités dites de *déviations* : très importantes en pratique, ces inégalités quantifient les risques pour que X tombe "loin" de $\mathbb{E}[X]$.

Proposition 10 (Inégalité de Markov). Soit X une v.a.r. positive p.s. Alors, pour tout $a > 0$,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}, \quad (32)$$

Démonstration. Pour tout $a > 0$, on a clairement

$$a\mathbf{1}_{X \geq a} \leq X. \quad (33)$$

Par monotonie de l'espérance, on en déduit que $\mathbb{E}[a\mathbf{1}_{X \geq a}] \leq \mathbb{E}[X]$, d'où le résultat. \square

L'inégalité de Markov est bien une inégalité de déviation par rapport à l'espérance. Elle permet de contrôler la probabilité que la variable prenne de « grandes valeurs ». Elle assure par exemple que n'importe quelle v.a.r. positive a , par exemple, moins de 5% de chances d'être 20 fois plus grande que sa moyenne. Notons au passage une conséquence intéressante dans le cas où $\mathbb{E}[X] = 0$.

Proposition 11. *Soit X une v.a.r. positive (p.s.). Alors $\mathbb{E}[X] = 0$ si et seulement si $X = 0$ p.s.*

Démonstration. Si $X = 0$ p.s., alors $\mathbb{E}[X] = 0$ car X a même espérance que la variable constante égale à 0. Réciproquement, si $\mathbb{E}[X] = 0$, alors l'inégalité de Markov entraîne que pour tout $n \geq 1$,

$$\mathbb{P}\left(X \geq \frac{1}{n}\right) = 0.$$

Comme

$$\{X > 0\} = \{\exists n \geq 1, X \geq \frac{1}{n}\} = \bigcup_{n=1}^{\infty} \left\{X \geq \frac{1}{n}\right\},$$

on en déduit par sous-additivité dénombrable que

$$\mathbb{P}(X > 0) \leq \sum_{n=1}^{\infty} \mathbb{P}\left(X \geq \frac{1}{n}\right). \quad (34)$$

Donc $\mathbb{P}(X > 0) = 0$. Par ailleurs $\mathbb{P}(X < 0) = 0$ par hypothèse, donc $\mathbb{P}(X \neq 0) = 0$. \square

On dit d'une variable aléatoire X qu'elle est de *carré intégrable* lorsque $\mathbb{E}(X^2) < +\infty$ (on rappelle que $\mathbb{E}(X^2)$ est toujours bien défini car $X^2 \geq 0$ mais peut être infini). On note que si X est de carré intégrable alors X est intégrable. En effet on rappelle que pour tout réels a, b ,

$$|ab| \leq \frac{1}{2}(a^2 + b^2).$$

On en déduit que $2|X| \leq 1 + X^2$ et donc, par monotonie, si $\mathbb{E}(|X|) < +\infty$ on a aussi $\mathbb{E}(X^2) < +\infty$. La réciproque en revanche est fautive : une variable peut être intégrable sans être de carré intégrable. Considérons par exemple la variable aléatoire X discrète à valeurs dans \mathbb{N}^* et dont la loi est caractérisée par

$$\mathbb{P}(X = k) = \frac{C}{k^3}, \quad k \geq 1,$$

où C est une constante adaptée (pour que la série somme bien à 1!). On a alors

$$\mathbb{E}(|X|) = \sum_{k=1}^{+\infty} k\mathbb{P}(X = k) = \sum_{k=1}^{+\infty} \frac{C}{k^2} < +\infty,$$

mais

$$\mathbb{E}(X^2) = \sum_{k=1}^{+\infty} k^2\mathbb{P}(X = k) = \sum_{k=1}^{+\infty} \frac{C}{k} = +\infty.$$

Définition 13 (Variance). *Soit X une v.a.r. de carré intégrable (i.e. $\mathbb{E}[X^2] < \infty$). On peut définir sa variance :*

$$\text{Var}(X) := \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[(X - \mathbb{E}[X])^2] < +\infty. \quad (35)$$

On note que, comme pour l'espérance, la variance d'une variable aléatoire est entièrement déterminée par sa loi.

Remarque 13 (La variance est toujours positive). *En utilisant la Proposition 11, la seconde expression montre que l'on a toujours $\text{Var}(X) \geq 0$, et qu'il y a égalité si et seulement si X est constante p.s. puisque $X = \mathbb{E}(X)$ p.s.*

Remarque 14. *Le fait qu'une variable soit intégrable n'implique pas qu'elle soit de carré intégrable et donc qu'on puisse définir sa variance. En fait on pourrait, mais ça n'est pas le choix fait dans ce cours, donner du sens à la variance de X en utilisant $\mathbb{E}[(X - \mathbb{E}[X])^2]$ qui est toujours bien défini puisque $(X - \mathbb{E}[X])^2 \geq 0$. Mais on notera bien que dans ce cas, rien n'assure que cette quantité soit finie.*

La variance d'une v.a.r. X mesure l'écart quadratique moyen entre X et $\mathbb{E}[X]$: elle est d'autant plus grande que les fluctuations de X autour de son espérance sont importantes. Cela est formalisé par l'inégalité suivante, qui est en fait un cas particulier "déguisé" de l'inégalité de Markov.

Proposition 12 (Inégalité de Bienaymé–Tchebychev). *Soit X une v.a.r. de carré intégrable. Alors,*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\text{Var}(X)}{a^2}, \quad (36)$$

pour tout $a > 0$.

Démonstration. La fonction $u \mapsto u^2$ étant strictement croissante sur $[0, \infty[$, on a

$$\{|X - \mathbb{E}[X]| \geq a\} = \{(X - \mathbb{E}[X])^2 \geq a^2\}.$$

Le résultat cherché s'obtient simplement en prenant les probabilités de ces évènements, puis en appliquant l'inégalité de Markov à la v.a.r. positive $(X - \mathbb{E}[X])^2$, qui a pour espérance $\text{Var}(X)$. \square

L'inégalité de Bienaymé-Tchebychev est la plus célèbre des inégalités de déviation. Elle garantit par exemple qu'une v.a.r. d'espérance μ et de variance σ^2 a au moins 99% de chances de tomber dans l'intervalle $]\mu - 10\sigma, \mu + 10\sigma[$. Il est notable que cette propriété soit valable sans autres hypothèses sur la loi qu'elle admette un second moment (i.e. qu'elle soit de variance finie). La quantité $\sigma = \sqrt{\text{Var}(X)}$ s'appelle l'*écart-type* de X . Elle joue donc également un rôle important en statistique descriptive dans la mesure où elle rend compte de l'« étalement » de la probabilité.

Exercice 10 (Lois usuelles). *Calculer la variance de chacune des 10 lois usuelles, lorsqu'elle existe.*

Exercice 11 (L'espérance est la meilleure prédiction). *Soit X une v.a.r. de carré intégrable. Montrer que le point $a = \mathbb{E}[X]$ est l'unique minimiseur de la fonction $a \mapsto \mathbb{E}[(X - a)^2]$.*

4 Indépendance

4.1 Probabilités conditionnelles

Considérons un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Comme nous l'avons vu, celui-ci modélise une certaine expérience au cours de laquelle un résultat est produit "au hasard". Nous cherchons à estimer les chances pour que ce résultat tombe dans une certaine partie $A \in \mathcal{F}$, c'est-à-dire pour que l'événement A se réalise. Sans information supplémentaire, la réponse à cette question est $\mathbb{P}(A)$. Mais si l'on apprend par ailleurs qu'un autre événement B s'est réalisé, alors la question de savoir si A se réalise revient en réalité à savoir si l'événement $A \cap B$ se réalise. Nous sommes donc tentés de modifier notre espace probabilisé en remplaçant la mesure $A \mapsto \mathbb{P}(A)$ par la mesure $A \mapsto \mathbb{P}(A \cap B)$. Il ne faut bien-sûr pas oublier de re-normaliser cette dernière afin d'en faire une véritable mesure de probabilité, ce qui nous amène tout naturellement à la définition suivante :

Définition 14 (Probabilité conditionnelle). Soient A, B deux événements, $\text{avec } \mathbb{P}(B) > 0$. La probabilité de A sachant B est définie par la formule

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (37)$$

L'application $A \mapsto \mathbb{P}(A|B)$ est une mesure de probabilité sur (Ω, \mathcal{F})

Vérifions ce dernier point. Pour tout $A \in \mathcal{F}$ $A \cap B \in \mathcal{F}$ par stabilité par intersection dénombrable et, de plus, $\mathbb{P}(A|B) \in [0, 1]$ car $A \cap B \subset B$ donc $\mathbb{P}(A \cap B) \leq \mathbb{P}(B)$. De plus $\mathbb{P}(\Omega|B) = \mathbb{P}(B)/\mathbb{P}(B) = 1$. Enfin pour toute suites $(A_k)_{k \geq 1}$ d'événements de \mathcal{F} disjoint 2 à 2,

$$\mathbb{P}(\cup_{k \geq 1} A_k | B) = \frac{\mathbb{P}((\cup_{k \geq 1}^{(disj.)} A_k) \cap B)}{\mathbb{P}(B)} = \frac{\sum_{k \geq 1} \mathbb{P}(A_k \cap B)}{\mathbb{P}(B)} = \sum_{k \geq 1} \mathbb{P}(A_k | B).$$

Exemple 12 (Lancer de dé). On lance un dé. Quelle est la probabilité que le chiffre obtenu soit pair, sachant qu'il est inférieur ou égal à 3 ? On se place bien-sûr sur l'univers $\Omega = \{1, 2, 3, 4, 5, 6\}$, que l'on munit de la tribu de toutes les parties, et de la loi uniforme. L'obtention d'un chiffre pair et celle d'un chiffre inférieur ou égal à 3 correspondent respectivement aux événements $A = \{2, 4, 6\}$ et $B = \{1, 2, 3\}$. Comme $A \cap B = \{2\}$, la réponse cherchée est

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/6}{3/6} = \frac{1}{3}.$$

Notons que la probabilité (inconditionnelle) de A est $\mathbb{P}(A) = 3/6 = 1/2$. Ainsi, le fait de savoir que le chiffre obtenu est inférieur ou égal à 3 rend moins probable l'obtention d'un chiffre pair.

Lemme 4 (Formule de Bayes). Soient A et B deux événements de probabilité non-nulle. Alors,

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B)\mathbb{P}(A|B)}{\mathbb{P}(A)}. \quad (38)$$

Voici une application typique de la formule de Bayes.

Exercice 12 (Application). 90% des étudiants révisent l'examen de Probabilité 1. La probabilité de validation est 0.8 pour un étudiant ayant révisé, et 0.2 pour un étudiant n'ayant pas révisé.

1. On rencontre un candidat qui a validé l'examen. Quelle est la probabilité qu'il ait révisé ?
2. Même question pour un candidat n'ayant pas validé l'examen.
3. Même question pour un candidat dont on ne sait rien.

Lemme 5 (Formules des probabilités totales). Soient $(B_n)_{n \geq 1}$ une famille d'événements formant une partition de Ω . Alors pour tout événement A , on a

$$\mathbb{P}(A) = \sum_{n=1}^{\infty} \mathbb{P}(B_n)\mathbb{P}(A|B_n), \quad (39)$$

avec la convention que $\mathbb{P}(B)\mathbb{P}(A|B) = 0$ si $\mathbb{P}(B) = 0$.

Démonstration. Il suffit de remarquer que $A = \cup_{n \geq 1} (A \cap B_n)$ et que les événements de cette union sont disjoints 2 à 2. On a donc

$$\mathbb{P}(A) = \sum_{n \geq 1} \mathbb{P}(A \cap B_n) = \sum_{n \geq 1} \mathbb{P}(A|B_n)\mathbb{P}(B_n).$$

□

Enfin les définitions que l'on vient de voir s'appliquent également lorsque la probabilité que l'on considère est en fait la loi d'une variable aléatoire :

Définition 15 (Loi conditionnelle). Soit X une v.a.r. et B un événement de probabilité non-nulle, tous deux sur le même espace probabilisé. Alors l'application

$$A \mapsto \mathbb{P}(X \in A|B) \quad (40)$$

définit une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ appelée loi conditionnelle de X sachant B .

Exercice 13 (Poisson filtré). Soient $\lambda > 0$ et $p \in]0, 1]$ des paramètres. Sur un même espace probabilisé, on considère deux v.a.r. N et X . On suppose que $N \sim \mathcal{P}(\lambda)$ et que pour tout $n \in \mathbb{N}$, la loi conditionnelle de X sachant $\{N = n\}$ est $\mathcal{B}(n, p)$. Quelle est la loi de la v.a.r. X ?

4.2 Événements indépendants

Définition 16. Deux événements A et B sont dits indépendants (noté $A \perp B$) si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \quad (41)$$

Intuitivement, deux événements sont indépendants lorsque la réalisation de l'un ne nous renseigne absolument pas quant-à celle de l'autre. En effet, la condition $A \perp B$ est trivialement vérifiée si $\mathbb{P}(A) = 0$ ou $\mathbb{P}(B) = 0$, et en dehors de ce cas dégénéré, on a clairement

$$A \perp B \iff \mathbb{P}(A|B) = \mathbb{P}(A) \iff \mathbb{P}(B|A) = \mathbb{P}(B). \quad (42)$$

Pour montrer que deux événements sont indépendants on privilégiera cependant la Définition 16 à son interprétation (42) pour laquelle il faut toujours vérifier qu'on n'est pas en train de diviser par une probabilité nulle !

Exemple 13 (Cartes et dé).

1. Lorsqu'on tire une carte dans un paquet de 52 cartes, les événements "tirer un cœur" et "tirer une dame" sont indépendants, mais pas les événements "tirer une dame" et "tirer une figure".
2. Lorsqu'on jette un dé, les événements "le résultat est pair" et "le résultat est divisible par 3" sont indépendants, mais pas les événements "le résultat est pair" et "le résultat est ≤ 3 "... En effet on a déjà vu le second point tandis que pour le premier, en notant $A = \{2, 4, 6\}$ et $B = \{3, 6\}$ les événements correspondant, on note que $\mathbb{P}(A \cap B) = \mathbb{P}(\{6\}) = 1/6$ et $\mathbb{P}(A)\mathbb{P}(B) = 3/6 \times 2/6 = 1/6$.

Lemme 6 (Indépendance et complémentaire). Pour deux événements $A, B \in \mathcal{F}$, on a toujours

$$A \perp B \iff A \perp B^c. \quad (43)$$

Démonstration. Il suffit d'observer que $A \cap B^c = A \setminus (A \cap B)$

$$\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A)\mathbb{P}(B^c).$$

□

De ce lemme, on déduit donc que pour deux événements A, B ,

$$A \perp B \iff A \perp B^c \iff A^c \perp B^c \iff A^c \perp B. \quad (44)$$

Définition 17 (Indépendance d'une famille d'événements). Soit $(A_i)_{i \in I}$ une famille quelconque d'événements. Les événements $(A_i)_{i \in I}$ sont dits indépendants si pour toute partie finie $K \subset I$,

$$\mathbb{P} \left(\bigcap_{i \in K} A_i \right) = \prod_{i \in K} \mathbb{P}(A_i) \quad (45)$$

Remarque 15 (Indépendance 2-à-2). Attention, l'indépendance de la famille $(A_i)_{i \in I}$ est beaucoup plus forte que l'indépendance $A_i \perp\!\!\!\perp A_j$ pour tout $i \neq j$ dans I , qui correspond au cas où $\text{Card}(K) = 2$. Voici un exemple très simple et à garder en tête pour lequel les deux notions diffèrent :

Exercice 14. On lance deux fois un pièce. Les trois événements suivants sont-ils indépendants ? indépendants 2 à 2 ?

A : on obtient pile au premier lancer ;

B : on obtient pile au second lancer ;

C : on obtient le même résultat aux deux lancers.

On peut modéliser cette expérience par l'espace suivant : $\Omega = \{0, 1\}^2$, $\mathcal{F} = \mathcal{P}(\Omega)$ et \mathbb{P} la probabilité uniforme. On a donc $A = \{(1, 0), (1, 1)\}$; $B = \{(0, 1), (1, 1)\}$ et $C = \{(0, 0), (1, 1)\}$. On vérifie alors facilement que ces trois événements sont indépendants 2 à 2 mais que $\mathbb{P}(A \cap B \cap C) = \mathbb{P}((1, 1)) = \frac{1}{4}$ tandis que $\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{8}$, donc les trois événements ne sont pas indépendants.

Remarque 16 (Indépendance 2-à-2). Attention, l'indépendance de la famille $(A_i)_{i \in I}$ est également beaucoup plus forte que la seule propriété

$$\mathbb{P} \left(\bigcap_{i \in I} A_i \right) = \prod_{i \in I} \mathbb{P}(A_i) \quad (46)$$

comme le prouve l'exercice suivant :

1. Proposer un espace de probabilités pour modéliser deux lancers successifs et indépendants d'un dé à 6 faces. Décrire mathématiquement les événements

$A =$ "le second jet vaut 1, 2 ou 3" ;

$B =$ "le second jet vaut 3, 4 ou 5" ;

$C =$ "la somme des deux jets vaut 9".

2. Montrer que

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$$

mais que A, B, C ne sont pas indépendants.

Remarque 17 (Stabilité par extraction et passage au complémentaire). Soient $(A_i)_{i \in I}$ un famille d'événements indépendants. Alors il en est de même de

1. toute sous-famille $(A_i)_{i \in J}$ avec $J \subseteq I$;
2. toute famille $(B_i)_{i \in I}$ obtenue en remplaçant certains des A_i par leur complémentaire, i.e.

$$\forall i \in I, \quad B_i = A_i \quad \text{ou} \quad B_i = A_i^c. \quad (47)$$

Proposition 13 (Lemme de Borel-Cantelli). Pour des événements $(A_n)_{n \geq 1}$, on a toujours

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \implies \mathbb{P} \left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \right) = 0.$$

Si les événements $(A_n)_{n \geq 1}$ sont indépendants, on a aussi

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = +\infty \implies \mathbb{P} \left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \right) = 1.$$

Un commentaire, avant de démarrer la preuve, sur l'événement que l'on étudie $\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$. Il s'agit de la limite supérieure des A_n , $n \geq 1$ notée $\limsup_n A_n$. Il s'agit de l'ensemble des éléments $\omega \in \Omega$ qui appartiennent à une infinité de A_n :

$$\limsup A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n = \{\omega \text{ t.q. } \forall k \geq 1, \omega \in \bigcup_{n=k}^{\infty} A_n\} = \{\omega \text{ t.q. } \forall k \geq 1 \exists n \geq k, \omega \in A_n\}.$$

Démonstration. Pour le premier point, on note que la suite d'événements $(\bigcup_{n=k}^{\infty} A_n)_{k \geq 1}$ est décroissante pour l'inclusion donc

$$\mathbb{P} \left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \right) = \lim_{k \rightarrow \infty} \mathbb{P} \left(\bigcup_{n=k}^{\infty} A_n \right).$$

De plus par sigma sous-additivité pour tout $k \geq 1$,

$$\mathbb{P} \left(\bigcup_{n=k}^{\infty} A_n \right) \leq \sum_{n \geq k} \mathbb{P}(A_n),$$

et obtient le résultat car le reste d'une série convergente converge vers 0.

Pour le second point (un peu plus délicat), nous allons montrer que l'événement complémentaire $\bigcup_{k \geq 1} \bigcap_{n \geq k} A_n^c$ (c'est l'ensemble des ω qui sont dans tous les A_n^c à partir d'un certain rang) est de probabilité nulle. Pour cela nous allons montrer que pour tout $k \geq 1$, $\mathbb{P}(\bigcap_{n \geq k} A_n^c) = 0$ ce qui implique bien $\mathbb{P}(\bigcup_{k \geq 1} \bigcap_{n \geq k} A_n^c) = 0$ par sigma sous-additivité (ou en utilisant le fait que l'union

est croissante). En utilisant l'indépendance de la famille $(A_n)_{n \geq 1}$ on obtient que pour tout $K \geq k$,

$$\mathbb{P}\left(\bigcap_{n \geq k} A_n^c\right) \leq \mathbb{P}\left(\bigcap_{n=k}^K A_n^c\right) = \prod_{n=k}^K \mathbb{P}(A_n^c).$$

S'il existe $n \geq k$ tel que $\mathbb{P}(A_n^c) = 0$, $\mathbb{P}(\bigcap_{n \geq k} A_n^c) = 0$. Sinon on passe à l'exponentielle et on doit donc étudier la série de terme général $\ln(1 - \mathbb{P}(A_n))$. Si la suite $(\mathbb{P}(A_n))_{n \geq k}$ ne tend pas vers 0 la série est grossièrement divergente. Sinon, comme le terme général est de signe constant et équivalent à $-\mathbb{P}(A_n)$, les deux séries sont de même nature et on obtient $\sum_{n \geq 1} \ln(1 - \mathbb{P}(A_n)) = -\infty$ puis $\lim_{K \rightarrow +\infty} \prod_{n=k}^K \mathbb{P}(A_n^c) = 0$. \square

Autrement dit, des événements trop rares ne se produisent qu'un nombre fini de fois, et des événements « non rares » (au sens où la série des probabilités diverge) et *indépendants* se produisent un nombre infini de fois.

Exercice 15. *On lance une infinité de fois une pièce équilibrée. Montrer qu'on obtient un nombre infini de pile et un nombre infini de face, presque-sûrement.*

4.3 Variables aléatoires indépendantes

Toutes les v.a.r. ci-dessous sont supposées définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Comme pour les événements, nous commençons par le cas simple de deux variables aléatoires indépendantes.

Définition 18 (Deux variables indépendantes). *Deux v.a.r. X et Y sont dites indépendantes si pour tous boréliens $A, B \in \mathcal{B}(\mathbb{R})$*

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B). \quad (48)$$

On note souvent $X \perp\!\!\!\perp Y$.

Autrement dit X et Y sont indépendantes si pour tous boréliens $A, B \in \mathcal{B}(\mathbb{R})$ les événements “ X appartient à A ” et “ Y appartient à B ” sont indépendants.

Remarque 18 (Cas discret). *Dans le cas où les v.a.r. X et Y sont discrètes, on a $X \perp\!\!\!\perp Y$ ssi*

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \quad (49)$$

pour tout $x \in \text{Im}(X)$ et tout $y \in \text{Im}(Y)$. En effet, le sens direct est facile et, réciproquement, pour

tout $A \subset \text{Im}(X)$ et $B \subset \text{Im}(Y)$,

$$\begin{aligned} \mathbb{P}(\{X \in A\} \cap \{Y \in B\}) &= \mathbb{P}\left(\bigcup_{x \in A \cap \text{Im}(X), y \in B \cap \text{Im}(Y)} \{X = x\} \cap \{Y = y\}\right) \\ &= \sum_{x \in A \cap \text{Im}(X), y \in B \cap \text{Im}(Y)} \mathbb{P}(X = x) \mathbb{P}(Y = y) \\ &= \mathbb{P}(A) \mathbb{P}(B), \end{aligned}$$

où, pour la seconde égalité, on a utilisé que les événements dans l'union sont 2 à 2 disjoints.

Exemple 14 (Lancer de deux dés). On modélise le lancer successifs de deux dés par la loi uniforme sur l'univers $\Omega = \{1, 2, \dots, 6\} \times \{1, 2, \dots, 6\}$. On note X et Y les résultats respectifs du premier et du second lancer. Montrer que les v.a.r. X et Y sont indépendantes.

Exemple 15 (Maximum de v.a.r. indépendantes). Soient X et Y deux v.a.r. indépendantes, et soit $Z = \max(X, Y)$. Exprimer F_Z en fonction de F_X et F_Y . Dans le cas particulier où X et Y suivent toutes deux la loi $\mathcal{U}(0, 1)$, montrer que Z admet une densité que l'on explicitera.

Remarque 19 (Stabilité par application de fonction). Si $X \perp\!\!\!\perp Y$ alors on a $g(X) \perp\!\!\!\perp h(Y)$ pour toutes fonctions boréliennes $h, g: \mathbb{R} \rightarrow \mathbb{R}$. En effet, pour tous boreliens A et B , $g^{-1}(A)$ et $h^{-1}(B)$ sont également des boréliens car g et h sont boréliennes donc

$$\begin{aligned} \mathbb{P}(g(X) \in A, h(Y) \in B) &= \mathbb{P}(X \in g^{-1}(A), Y \in h^{-1}(B)) \\ &= \mathbb{P}(X \in g^{-1}(A)) \mathbb{P}(Y \in h^{-1}(B)) \\ &= \mathbb{P}(g(X) \in A) \mathbb{P}(h(Y) \in B). \end{aligned}$$

Proposition 14 (Caractérisation). Les conditions suivantes sont équivalentes.

(i) $X \perp\!\!\!\perp Y$

(ii) Pour tout $s, t \in \mathbb{R}$, on a

$$\mathbb{P}(X \leq s, Y \leq t) = F_X(s) F_Y(t) \quad (50)$$

(iii) Pour toutes fonctions boréliennes $g, h: \mathbb{R} \rightarrow \mathbb{R}$, positives ou bornées,

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)] \mathbb{E}[h(Y)], \quad (51)$$

Cette formule est en fait vraie dès lors que ces espérances existent.

Démonstration. La preuve est admise. Voici cependant quelques idées pour les différentes étapes.

1. (i) \implies (iii) On vérifie que c'est vrai pour g et h indicatrices de boreliens (c'est alors la définition) puis on généralise en utilisant l'approximation par fonctions étagées.
2. (iii) \implies (ii) C'est la partie facile : on prend $g = 1_{]-\infty, s]}(\cdot)$ et $h = 1_{]-\infty, t]}(\cdot)$.
3. (ii) \implies (i) Il faut utiliser le lemme de classe monotone.

□

La définition de l'indépendance se généralise au cas de plusieurs variables aléatoires, comme suit.

Définition 19 (Indépendance de plusieurs variables). *Des v.a.r. X_1, \dots, X_n sont dites indépendantes si pour tous boreliens $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$*

$$\mathbb{P} \left(\bigcap_{i=1}^n \{X_i \in B_i\} \right) = \prod_{i=1}^n \mathbb{P}(X_i \in B_i). \quad (52)$$

Pour une famille infinie de v.a.r. $(X_i)_{i \in I}$, on parle d'indépendance lorsque toute sous-famille finie est constituée de v.a.r. indépendantes, c'est-à-dire que pour toute partie finie $K \subset I$ et tout choix des boreliens $(B_i)_{i \in K}$,

$$\mathbb{P} \left(\bigcap_{i \in K} \{X_i \in B_i\} \right) = \prod_{i \in K} \mathbb{P}(X_i \in B_i). \quad (53)$$

Remarque 20 (Indépendance 2-à-2). *Attention, comme pour la notion d'indépendance pour les événements, l'indépendance de la famille $(X_i)_{i \in I}$ est beaucoup plus forte que l'indépendance $X_i \perp\!\!\!\perp X_j$ pour tout $i \neq j$ dans I , qui correspond au cas $\text{card}(K) = 2$.*

Remarque 21 (Stabilité par extraction et application de fonctions). *Soient $(X_i)_{i \in I}$ une famille quelconque de v.a.r. indépendantes. Alors il en est de même de*

1. toute sous-famille $(X_i)_{i \in J}$ avec $J \subseteq I$;
2. toute famille $(Y_i)_{i \in I}$ obtenue en posant $Y_i = h_i(X_i)$ avec $h_i: \mathbb{R} \rightarrow \mathbb{R}$ borélienne.
3. toute famille $(Y_j)_{j \in J}$ obtenue en posant $Y_j = h_j(X_i: i \in I_j)$, où les $(I_j)_{j \in J}$ sont des parties 2-à-2 disjointes et finies de I , et où $h_j: \mathbb{R}^{|I_j|} \rightarrow \mathbb{R}$ est une fonction continue pour tout $j \in J$.

Proposition 15 (Caractérisation). *Soient X_1, \dots, X_n des v.a.r. définies sur le même espace probabilisé. Alors, les conditions suivantes sont équivalentes :*

1. Les v.a.r. X_1, \dots, X_n sont indépendantes.

2. Pour tout $(t_1, \dots, t_n) \in \mathbb{R}^n$, on a

$$\mathbb{P}(X_1 \leq t_1, \dots, X_n \leq t_n) = F_{X_1}(t_1) \times \dots \times F_{X_n}(t_n). \quad (54)$$

3. Pour toutes fonctions boréliennes h_1, \dots, h_n positives ou bornées, on a

$$\mathbb{E} \left[\prod_{i=1}^n h_i(X_i) \right] = \prod_{i=1}^n \mathbb{E} [h_i(X_i)]. \quad (55)$$

Cette formule est en fait vraie dès que les espérances sont bien définies.

Cette proposition prolonge donc la Proposition 14. Elle est également admise.

Définition 20 (Variables i.i.d.). Lorsque les v.a.r. $(X_i)_{i \in I}$ sont indépendantes et qu'elles ont toutes la même loi, on dit qu'elles sont i.i.d. (indépendantes et identiquement distribuées).

Exercice 16 (Interprétation de la loi géométrique). Soit $p \in]0, 1]$ un paramètre et $(X_n)_{n \geq 1}$ une suite de v.a.r. i.i.d. de loi $\mathcal{B}(p)$. On s'intéresse à la position du premier succès :

$$N = \inf\{n \geq 1 : X_n = 1\},$$

avec la convention $\inf \emptyset = +\infty$. Quelle est la loi de la variable aléatoire N ?

Exercice 17 (Un exemple d'application du lemme de Borel Cantelli). Soit $(X_n)_{n \geq 1}$ une suite de variable i.i.d. de loi Bernoulli de paramètre p (où p est un réel de $]0, 1[$).

1. Montrer qu'il y a p.s. une infinité de n tels que $X_n = 1$.
 2. Pour tout $n \geq 0$, on définit l'événement $A_n = \{X_{n+1} = \dots = X_{2n} = 1\}$. Montrer que p.s. il n'y a qu'un nombre fini (mais aléatoire) de A_n qui sont réalisés.
1. On note que les événements $(\{X_n = 1\})_{n \geq 1}$ sont indépendants et $\mathbb{P}(X_n = 1) = p$. Donc $\sum_{n \geq 1} \mathbb{P}(X_n = 1) = +\infty$ et du lemme de Borel Cantelli on déduit que $\mathbb{P}(\limsup_n \{X_n = 1\}) = 1$.
 2. Pour tout $n \geq 0$, comme les variables X_{n+1}, \dots, X_{2n} sont indépendantes $\mathbb{P}(A_n) = \mathbb{P}(X_{n+1} = 1) \dots \mathbb{P}(X_{2n} = 1) = p^n$. On en déduit que $\sum_{n \geq 1} \mathbb{P}(A_n) < +\infty$ et donc, d'après le lemme de Borel Cantelli, $\mathbb{P}(\limsup_n \{X_n = 1\}) = 0$.

5 Sommes de variables indépendantes

5.1 Formules de convolution

Étant données deux v.a.r. indépendantes X, Y , on s'intéresse ici à leur somme

$$Z := X + Y.$$

Dans les meilleurs cas, on peut calculer explicitement la loi de Z .

Proposition 16 (Cas discret). *Soient X, Y des v.a.r. discrètes et indépendantes. Alors Z est discrète et on peut donc caractériser sa loi par*

1. son image, $\text{Im}(Z) = \{x + y : (x, y) \in \text{Im}(X) \times \text{Im}(Y)\}$,
2. et la fonction de poids définie, pour tout $z \in \text{Im}(Z)$, par

$$\begin{aligned} \mathbb{P}(Z = z) &= \sum_{x \in \text{Im}(X)} \mathbb{P}(X = x) \mathbb{P}(Y = z - x) \\ &= \sum_{y \in \text{Im}(Y)} \mathbb{P}(X = z - y) \mathbb{P}(Y = y) \end{aligned} \tag{56}$$

Preuve. On a clairement

$$\begin{aligned} \{Z = z\} &= \bigcup_{x \in \text{Im}(X)} \{X = x, Z = z\} \\ &= \bigcup_{x \in \text{Im}(X)} \{X = x, Y = z - x\} \end{aligned}$$

Il suffit alors de remarquer que l'union est dénombrable et disjointe et d'utiliser que X et Y sont indépendantes. \square

Proposition 17 (Cas à densité). *Soit X, Y deux variables aléatoires indépendantes de densités respectives f_X et f_Y . Alors la somme $Z = X + Y$ admet une densité donnée par*

$$f_Z(z) = \int_{\mathbb{R}} f_X(w) f_Y(z - w) dw, \quad \text{pour tout } z \in \mathbb{R}. \tag{57}$$

Démonstration. Admise. \square

Les formules qui apparaissent dans (56) et (57) définissent ce qu'on appelle le *produit de convolution* (discret et continu).

Corollaire 7. *On déduit en particulier de ces formules que (on rappelle que X et Y sont indépendantes) :*

1. Si $X \sim \mathcal{B}(n, p)$ et $Y \sim \mathcal{B}(m, p)$ alors $Z \sim \mathcal{B}(n + m, p)$.
2. Si $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ alors $Z \sim \mathcal{P}(\lambda + \mu)$.
3. Si $X \sim \mathcal{N}(\mu, \sigma^2)$ et $Y \sim \mathcal{N}(\nu, \tau^2)$ alors $Z \sim \mathcal{N}(\mu + \nu, \sigma^2 + \tau^2)$.
4. Si $X \sim \Gamma(r, \lambda)$ et $Y \sim \Gamma(s, \lambda)$ alors $Z \sim \Gamma(r + s, \lambda)$.

Correction.

1. L'image de Z est bien $\{0, \dots, n + m\}$ et pour tout $z \in \text{Im}(Z)$,

$$\begin{aligned}
\mathbb{P}(Z = z) &= \sum_{x=0}^n \mathbb{P}(X = x) \mathbb{P}(Y = z - x) \\
&= \sum_{x=0}^n \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{z-x} p^{z-x} (1-p)^{m-(z-x)} \\
&= p^z (1-p)^{n+m-z} \sum_{x=0}^n \binom{n}{x} \binom{m}{z-x} \\
&= p^z (1-p)^{n+m-z} \binom{n+m}{z}.
\end{aligned}$$

Dans le calcul ci-dessus on utilise que par convention $\binom{n}{k} = 0$ dès que $k < 0$ ou $k > n$. On peut retrouver la dernière égalité en dénombrant les parties à z éléments d'un ensemble à $n + m$ éléments selon le nombre d'éléments qu'elle possède dans les n premiers éléments. On retrouve sinon ce résultat de façon plus analytique en utilisant le binôme de Newton puis en développant $(1 + u)^n (1 + u)^m$.

4. Pour tout $z \in \mathbb{R}$

$$\begin{aligned}
f_Z(z) &= \int \frac{\lambda^r}{\Gamma(r)} w^{r-1} e^{-\lambda w} 1_{[0, +\infty[}(w) \frac{\lambda^s}{\Gamma(s)} (z-w)^{s-1} e^{-\lambda(z-w)} 1_{[0, +\infty[}(z-w) dw \\
&= \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} e^{-\lambda z} 1_{[0, +\infty[}(z) \int_0^z w^{r-1} (z-w)^{s-1} dw \\
&\stackrel{(u=w/z)}{=} \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} z^{r+s-1} e^{-\lambda z} 1_{[0, +\infty[}(z) \int_0^1 u^{r-1} (1-u)^{s-1} du \\
&= \frac{\lambda^{r+s}}{\Gamma(r+s)} z^{r+s-1} e^{-\lambda z} 1_{[0, +\infty[}(z).
\end{aligned}$$

Pour la dernière égalité on a utilisé l'égalité $\mathcal{B}(r, s) := \int_0^1 u^{r-1} (1-u)^{s-1} du = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$ que l'on admettra dans le cadre de ce cours.

En remarquant que $\mathcal{B}(1, p) = \mathcal{B}(p)$ et que $\Gamma(1, \lambda) = \mathcal{E}(\lambda)$, on en déduit en particulier, par une récurrence immédiate, les deux résultats importants suivants.

Corollaire 8. (*Interprétation des lois Binômiales et Gamma*) Soit $n \geq 1$ un entier.

1. Si X_1, \dots, X_n sont des v.a.r. i.i.d. de loi $\mathcal{B}(p)$ ($p \in [0, 1]$), alors $X_1 + \dots + X_n \sim \mathcal{B}(n, p)$.
2. Si X_1, \dots, X_n sont des v.a.r. i.i.d. de loi $\mathcal{E}(\lambda)$ ($\lambda \in]0, \infty[$), alors $X_1 + \dots + X_n \sim \Gamma(n, \lambda)$

5.2 Espérance et variance d'une somme

La formule de convolution est parfois dure à exploiter. Le lecteur pourra par exemple essayer de montrer que la somme de deux v.a.r. de Cauchy indépendantes est encore une v.a.r. de Cauchy. Nous y reviendrons, munis d'outils plus sophistiqués. En général, il est difficile voire impossible de calculer la loi de Z explicitement. Néanmoins, comme on va le voir, on peut toujours facilement accéder aux deux statistiques les plus importantes concernant Z : son espérance et sa variance.

Concernant l'**espérance**, l'inégalité triangulaire $|Z| \leq |X| + |Y|$ montre que **si les variables X et Y sont intégrables, alors leur somme $Z = X + Y$ l'est aussi**. De plus, dans ce cas, la linéarité de l'espérance nous autorise à écrire

$$\mathbb{E}[Z] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Notons que **l'indépendance de X et Y ne joue aucun rôle ici**. La situation est un tout petit peu plus compliquée pour ce qui est de la variance.

Définition 21 (Covariance). Soient X et Y des v.a.r. de carré intégrable. Alors l'inégalité $2|XY| \leq X^2 + Y^2$ montre que la v.a.r. XY est intégrable, et l'on peut donc définir la covariance :

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \end{aligned}$$

Remarque 22 (Variance). Dans le cas particulier où $X = Y$, on retrouve la variance :

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

Le lien entre covariance et indépendance est résumé dans l'énoncé suivant :

Lemme 9 (Indépendance et covariance). Soient X, Y des v.a.r. de carré intégrable. Alors,

$$X \perp\!\!\!\perp Y \implies \text{Cov}(X, Y) = 0.$$

Attention : la réciproque est fautive en général, comme le montre l'exemple suivant !

Exemple 16 (Contre-exemple pour la réciproque). Soient X, U des v.a.r. indépendantes avec $X \sim \mathcal{N}(0, 1)$ et $U \sim \mathcal{U}(\{-1, 1\})$. On pose $Y := UX$. Vérifions que $Y \sim \mathcal{N}(0, 1)$ et que $\text{Cov}(X, Y) = 0$, mais que X et Y ne sont pas indépendantes.

Soit h une fonction continue bornée. On a alors

$$\mathbb{E}(h(Y)) = \mathbb{E}(h(X)1_{U=1}) + \mathbb{E}(h(-X)1_{U=-1}) \stackrel{\text{ll}}{=} \mathbb{E}(h(X))\mathbb{P}(U = 1) + \mathbb{E}(h(-X))\mathbb{P}(U = -1)$$

et comme $\mathbb{P}(U = 1) = \mathbb{P}(U = -1) = 1/2$ et $X \stackrel{\text{(loi)}}{=} -X$, on obtient $\mathbb{E}(h(Y)) = \mathbb{E}(h(X))$. On en déduit que Y suit également une $\mathcal{N}(0, 1)$. De plus

$$\text{Cov}(X, Y) = \mathbb{E}(X^2U) \stackrel{\text{ll}}{=} \mathbb{E}(X^2)\mathbb{E}(U) = 0.$$

Il nous reste à montrer que X et Y ne sont pas indépendantes (ce qui est très intuitif puisque les deux variables ont même valeur absolue!). On peut par exemple remarquer que

$$\mathbb{P}(|Y| \leq 1, |X| > 1) = 0$$

alors que

$$\mathbb{P}(|Y| \leq 1)\mathbb{P}(|X| > 1) > 0.$$

Proposition 18 (Bilinéarité). L'ensemble $L^2(\Omega, \mathcal{F}, \mathbb{P})$ des v.a.r. de carré intégrable sur $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace vectoriel, sur lequel la covariance $\text{Cov}(\cdot, \cdot)$ définit une forme bilinéaire symétrique : pour toutes v.a.r. $X, Y, Z \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ et tout $\lambda \in \mathbb{R}$

$$\begin{aligned} \text{Cov}(X, Y) &= \text{Cov}(Y, X); \\ \text{Cov}(\lambda X + Y, Z) &= \lambda \text{Cov}(X, Z) + \text{Cov}(Y, Z). \end{aligned}$$

D'après la Remarque 22, la forme quadratique associée est la variance.

Démonstration. Pour $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ et $\lambda \in \mathbb{R}$, on a toujours

$$\begin{aligned} (\lambda X + Y)^2 &\leq \lambda^2 X^2 + Y^2 + 2|\lambda| |XY| \\ &\leq (\lambda^2 + |\lambda|)X^2 + (1 + |\lambda|)Y^2, \end{aligned}$$

où l'on a utilisé l'inégalité $2|ab| \leq a^2 + b^2$. En utilisant monotonie et linéarité de l'espérance, on en déduit que $\lambda X + Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Ainsi, $L^2(\Omega, \mathcal{F}, \mathbb{P})$ est un espace vectoriel. La symétrie de la covariance est évidente et la bilinéarité est une conséquence de la linéarité de l'espérance. \square

Corollaire 10 (Variance d'une somme). *Pour $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, on a toujours*

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

En particulier, si X, Y sont indépendantes, alors

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

Plus généralement, pour $X_1, \dots, X_n \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, on a toujours

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

En particulier, si X_1, \dots, X_n sont indépendantes 2 à 2 (et donc en particulier si X_1, \dots, X_n forment une famille de variables indépendantes), alors

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

Démonstration. En effet il suffit d'écrire

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^n X_i\right) &= \text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right) \\ &= \sum_{i=1}^n \text{Cov}(X_i, X_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j), \end{aligned}$$

où l'on a utilisé la Remarque 22, puis la bilinéarité et la symétrie de $\text{Cov}(\cdot, \cdot)$. □

Exercice 18. *Retrouver la variance des loi $\mathcal{B}(n, p)$ et $\Gamma(n, \lambda)$ pour $n \in \mathbb{N}^*$, $p \in [0, 1]$ et $\lambda \in]0, \infty[$.*

5.3 Loi faible des grands nombres

Une conséquence de ce qui précède est le résultat fondamental suivant concernant la somme d'un grand nombre de variables indépendantes et identiquement distribuées. Cela valide l'intuition selon laquelle l'espérance d'une v.a.r. représente la moyenne obtenue sur un grand nombre d'expériences.

Théorème 4 (Loi faible des grands nombres). *Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. i.i.d. de carré*

intégrable et d'espérance μ . Pour tout $n \geq 1$, on définit la moyenne empirique

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

Alors pour tout $\varepsilon > 0$, on a

$$\mathbb{P}(|Z_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2 n}, \quad (58)$$

où σ^2 désigne la variance des $(X_n)_{n \geq 1}$. En particulier, $(Z_n)_{n \geq 1}$ **converge en probabilité** vers μ : pour tout $\varepsilon > 0$, on a

$$\mathbb{P}(|Z_n - \mu| \geq \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0.$$

Preuve. Par linéarité de l'espérance, on a

$$\mathbb{E}[Z_n] = \frac{\mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]}{n} = \mu.$$

Comme les variables X_1, \dots, X_n sont indépendantes, les variances s'ajoutent également :

$$\text{Var}(Z_n) = \frac{\text{Var}(X_1) + \cdots + \text{Var}(X_n)}{n^2} = \frac{\sigma^2}{n},$$

En appliquant l'inégalité de Bienaymé–Tchebychev à la v.a.r. Z_n , on obtient pour tout $\varepsilon > 0$,

$$\mathbb{P}(|Z_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2 n}.$$

La conclusion en découle immédiatement. □

Remarque 23. Pour tout $\lambda \in \mathbb{R}$, en posant $\varepsilon = \frac{\lambda\sigma}{\sqrt{n}}$, l'équation (58) nous apprend que la probabilité que Z_n tombe dans $[\mu - \frac{\lambda\sigma}{\sqrt{n}}, \mu + \frac{\lambda\sigma}{\sqrt{n}}]$ est au moins $1 - \frac{1}{\lambda^2}$, indépendamment de n . On peut montrer (grâce au théorème central limite que vous verrez au prochain trimestre) que cette probabilité tend en fait vers une constante universelle explicite lorsque $n \rightarrow \infty$. On retiendra en tout cas pour l'instant que, pour n grand, la moyenne empirique Z_n est avec grande probabilité dans une fenêtre d'ordre $1/\sqrt{n}$ centrée en μ .

Remarque 24. En inspectant soigneusement la preuve la loi faible des grands nombre, on peut vérifier que l'hypothèse *i.i.d.* peut être affaiblie : on ne s'est finalement servi que de $\mathbb{E}[X_n] = \mu$, $\sup_n \text{Var}(X_n) < \infty$, et $\text{Cov}(X_n, X_m) = 0$ lorsque $n \neq m$.

De manière générale, on dit qu'une suite de variables aléatoires $(X_n)_{n \geq 0}$ **converge en proba-**

bilité vers une variable aléatoire X si pour tout $\varepsilon > 0$,

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0.$$

On notera que la convergence presque-sûre

$$X_n \xrightarrow[n \rightarrow \infty]{} X \quad p.s.$$

i.e.

$$\mathbb{P}(\omega \text{ t.q. } X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X(\omega)) = 1,$$

implique la convergence en probabilité :

Proposition 19. Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires convergeant presque sûrement vers une variable aléatoire X . Alors $(X_n)_{n \geq 0}$ converge également en probabilité vers X .

Démonstration. En effet, pour tout $\varepsilon > 0$ la suite de variables aléatoires définies par

$$Y_n = 1_{\{|X_n - X| \geq \varepsilon\}}, \quad n \geq 0,$$

converge presque sûrement vers 0. De plus pour tout $n \geq 0$, $|Y_n| \leq 1$ et la variable aléatoire constante égale à 1 est intégrable, donc, d'après le théorème de convergence dominée, la suite $(\mathbb{E}(Y_n))_{n \geq 0}$ converge vers 0. On conclut en remarquant que pour tout $n \geq 0$, $\mathbb{E}(Y_n) = \mathbb{P}(|X_n - X| \geq \varepsilon)$. \square

La réciproque est cependant fautive comme le prouve l'exemple suivant. On considère une suite $(X_n)_{n \geq 1}$ de variables aléatoires indépendantes telles que pour tout $n \geq 1$, la variable X_n suit une loi de Bernoulli de paramètre $1/n$. On vérifie alors que pour tout $\varepsilon > 0$ et tout $n \geq 1$, $\mathbb{P}(|X_n| > \varepsilon) = 1/n$ donc la suite de variables converge bien vers 0 en probabilité. En revanche, en utilisant le second lemme de Borel-Cantelli, on obtient que $\mathbb{P}(\limsup_{n \geq 1} \{|X_n| > \varepsilon\}) = 1$ ce qui signifie que presque sûrement $|X_n| > \varepsilon$ une infinité de fois ce qui implique qu'il n'y a pas convergence p.s. vers 0.

Exercice 19. On lance $n = 1000$ fois une pièce de monnaie et l'on compte la proportion Z de piles obtenus. À l'aide de l'inégalité de Bienaymé-Tchebychev, déterminer un intervalle dans lequel Z ait au moins 90% de chances de se trouver. Même question pour $n = 25000$.

Exercice 20. On dispose d'un grand nombre d'observations réelles, supposés indépendantes et de même loi inconnue. Proposer un moyen d'estimer la fonction de répartition associée.

Fin du cours 2022 - 2023 (on verra où on arrive en 2024 !)

6 Fonctions caractéristiques (*Hors-programme*)

6.1 Définition et exemples

Commençons par un mot sur l'espérance de quantités aléatoires à valeurs complexes. Soit X une v.a.r. et $h: \mathbb{R} \rightarrow \mathbb{C}$ une fonction à valeurs complexes. Écrivons $h = f + ig$ avec $f = \Re(h)$ et $g = \Im(h)$. Si $f(X)$ et $g(X)$ sont des v.a.r. intégrables, on peut tranquillement poser

$$\mathbb{E}[h(X)] := \mathbb{E}[f(X)] + i\mathbb{E}[g(X)].$$

On vérifie immédiatement que les propriétés usuelles de l'espérance (la linéarité, notamment) sont préservées par cette extension au cas complexe. En particulier, si la v.a.r. X est discrète, on a

$$\mathbb{E}[h(X)] = \sum_{x \in \mathfrak{S}(X)} h(x)\mathbb{P}(X = x)$$

dès que la somme est absolument convergente, tandis que si X admet une densité f_X , on a

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x)f_X(x) dx,$$

dès que l'intégrale est absolument convergente.

Définition 22. La fonction caractéristique d'une v.a.r. X est la fonction $\Phi_X: \mathbb{R} \rightarrow \mathbb{C}$ définie par

$$\Phi_X(t) := \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)].$$

On calcule aisément les fonctions caractéristiques suivantes (le faire en exercice).

Proposition 20 (Lois usuelles).

1. Si $X \sim \mathcal{B}(p)$, alors $\Phi_X(t) = 1 - p + pe^{it}$;
2. Si $X \sim \mathcal{B}(n, p)$, alors $\Phi_X(t) = (1 - p + pe^{it})^n$;
3. Si $X \sim \mathcal{P}(\lambda)$, alors $\Phi_X(t) = \exp(\lambda(e^{it} - 1))$;
4. Si $X \sim \mathcal{G}(p)$, alors $\Phi_X(t) = \frac{pe^{it}}{1 - (1-p)e^{it}}$;
5. Si $X \sim \mathcal{U}(-a, a)$, alors $\Phi_X(t) = \text{sinc}(at)$;
6. Si $X \sim \mathcal{E}(\lambda)$, alors $\Phi_X(t) = \frac{\lambda}{\lambda - it}$.

Le cas $X \sim \mathcal{N}(\mu, \sigma^2)$ est plus délicat.

Proposition 21 (Fonction caractéristique de la loi normale). *Pour $X \sim \mathcal{N}(\mu, \sigma^2)$, on a*

$$\Phi_X(t) = \exp\left(i\mu t - \frac{t^2\sigma^2}{2}\right).$$

Démonstration. Supposons d'abord $\mu = 0, \sigma^2 = 1$. On a alors

$$\Phi_X(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \cos(tx) e^{-\frac{x^2}{2}} dx,$$

la partie imaginaire étant nulle par parité (c'est vrai pour n'importe quelle variable aléatoire satisfaisant $-X \stackrel{\text{loi}}{=} X$). Le terme sous l'intégrale est continument dérivable par rapport à t , et sa dérivée $x \mapsto -x \sin(tx) e^{-\frac{x^2}{2}}$ est dominée indépendamment de t par la fonction $x \mapsto |x| e^{-\frac{x^2}{2}}$, qui est intégrable sur \mathbb{R} . On admettra que cela nous donne le droit d'invertir dérivée et intégrale :

$$\begin{aligned} \Phi'_X(t) &= -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \sin(tx) x e^{-\frac{x^2}{2}} \mathcal{D}x \\ &= \left[\sin(tx) e^{-\frac{x^2}{2}} \right]_{-\infty}^{+\infty} - \frac{t}{\sqrt{2\pi}} \int_{\mathbb{R}} \cos(tx) e^{-\frac{x^2}{2}} \mathcal{D}x \\ &= -t\Phi_X(t). \end{aligned}$$

On déduit de cette équation différentielle classique que $\Phi_X(t) = \Phi_X(0) e^{-\frac{t^2}{2}} = e^{-\frac{t^2}{2}}$. Pour le cas général, on remarque que $X = \mu + \sigma X_0$ avec $X_0 \sim \mathcal{N}(0, 1)$. Ainsi,

$$\Phi_X(t) = \mathbb{E}\left[e^{it(\mu + \sigma X_0)}\right] = e^{i\mu t} \Phi_{X_0}(t\sigma),$$

et l'on conclut en utilisant la formule obtenue pour Φ_{X_0} . □

Proposition 22 (Fonction caractéristique de la loi Gamma). *Pour $X \sim \Gamma(r, \lambda)$, on a*

$$\Phi_X(t) = \left(\frac{\lambda}{\lambda - it}\right)^r.$$

Démonstration. On applique la même méthode que ci-dessus. Par définition, on a

$$\Phi_X(t) = \frac{\lambda^r}{\Gamma(r)} \int_0^\infty x^{r-1} e^{itx - \lambda x} \mathcal{D}x.$$

Le terme sous l'intégrale est dérivable par rapport à t , et sa dérivée $ix^r e^{itx - \lambda x}$ est dominée indépendamment

de t par $x^r e^{-\lambda x}$ qui est intégrable sur $]0, \infty[$. Cela nous autorise à intervertir dérivée et intégrale :

$$\begin{aligned}\Phi'_X(t) &= \frac{i\lambda^r}{\Gamma(r)} \int_0^\infty x^r e^{itx-\lambda x} \mathcal{D}x \\ &= \frac{i\lambda^r}{\Gamma(r)} \left[x^r \frac{e^{itx-\lambda x}}{it-\lambda} \right]_0^\infty + \frac{ir\lambda^r}{(\lambda-it)\Gamma(r)} \int_0^\infty x^{r-1} e^{itx-\lambda x} \mathcal{D}x \\ &= \frac{ir}{\lambda-it} \Phi_X(t).\end{aligned}$$

La solution de cette équation différentielle avec $\Phi_X(0) = 1$ est $\Phi_X(t) = \left(\frac{\lambda}{\lambda-it}\right)^r$. □

6.2 Propriétés fondamentales

Lemme 11 (Propriétés élémentaires). Φ_X est continue, avec $|\Phi_X(t)| \leq 1$ et $\Phi_X(0) = 1$.

Théorème 5 (La fonction caractéristique porte bien son nom). *Si les variables aléatoires réelles X et Y sont telles que $\Phi_X = \Phi_Y$, alors X et Y ont la même loi.*

Ainsi, en théorie, on peut retrouver à partir de Φ_X tout ce que l'on désire savoir sur la loi de X . La formule d'inversion de Levy est une belle illustration de ce principe.

Théorème 6 (Formule d'inversion de Levy). *Soit X une variable aléatoire telle que $\int_{\mathbb{R}} |\Phi_X(t)| \mathcal{D}t < \infty$. Alors X admet une densité, donnée par la formule suivante :*

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \Phi_X(t) \mathcal{D}t.$$

Exercice 21 (Exemple important). *Vérifier cette formule dans le cas $X \sim \mathcal{N}(0, 1)$.*

Exercice 22 (Fonction caractéristique d'une variable de Cauchy). *Soit X une variable aléatoire de densité $f_X(x) = \frac{\lambda}{2} e^{-\lambda|x|}$. Calculer Φ_X , puis appliquer la formule d'inversion de Levy. En déduire la fonction caractéristique d'une variable aléatoire de Cauchy.*

Nous donnons maintenant une proposition immédiate, mais qui fait toute la puissance des fonctions caractéristiques pour l'étude des sommes de variables indépendantes.

Proposition 23 (Somme de variables indépendantes). *Si X, Y sont indépendantes, alors*

$$\Phi_{X+Y} = \Phi_X \Phi_Y.$$

Cette forme produit est infiniment plus simple que la formule de convolution des densités. De plus, elle est valable pour des variables indépendantes absolument quelconques !

Corollaire 12. *En particulier, on voit immédiatement que*

1. Si $X \perp\!\!\!\perp Y$ avec $X \sim \mathcal{B}(n, p)$ et $Y \sim \mathcal{B}(m, p)$, alors $X + Y \sim \mathcal{B}(n + m, p)$.
2. Si $X \perp\!\!\!\perp Y$ avec $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$, alors $X + Y \sim \mathcal{P}(\lambda + \mu)$.
3. Si $X \perp\!\!\!\perp Y$ avec $X \sim \mathcal{N}(\mu, \sigma^2)$ et $Y \sim \mathcal{N}(\nu, \tau^2)$, alors $X + Y \sim \mathcal{N}(\mu + \nu, \sigma^2 + \tau^2)$.
4. Si $X \perp\!\!\!\perp Y$ avec $X \sim \Gamma(r, \lambda)$ et $Y \sim \Gamma(s, \lambda)$, alors $X + Y \sim \Gamma(r + s, \lambda)$.
5. Si $X \perp\!\!\!\perp Y$ avec $X \sim \mathcal{C}(\lambda)$ et $Y \sim \mathcal{C}(\mu)$, alors $X + Y \sim \mathcal{C}(\lambda + \mu)$.

Notons que la dernière assertion est nouvelle.

Nous terminons cette section en faisant le lien entre fonction caractéristique et moments.

Théorème 7 (Moments). *Soit X une v.a.r., et soit $n \in \mathbb{N}$. On suppose que $\mathbb{E}[|X|^n] < \infty$. Alors, Φ_X est n fois continûment dérivable sur \mathbb{R} , et la dérivée n -ième est donnée par*

$$\Phi_X^{(n)}(t) = i^n \mathbb{E}[X^n e^{itX}].$$

En particulier, Φ_X admet un développement de Taylor d'ordre n autour de 0 :

$$\Phi_X(t) = 1 + i\mathbb{E}[X]t - \frac{\mathbb{E}[X^2]}{2}t^2 + \dots + \frac{i^n \mathbb{E}[X^n]}{n!}t^n + o(t^n), \text{ lorsque } t \rightarrow 0.$$

Exemple 17. *Par unicité du développement de Taylor, on en déduit aussitôt que :*

1. Si $X \sim \mathcal{E}(\lambda)$, alors pour tout $n \in \mathbb{N}$, $\mathbb{E}[X^n] = \frac{n!}{\lambda^n}$.
2. Si $X \sim \mathcal{N}(0, \sigma^2)$, alors pour tout $n \in \mathbb{N}$, $\mathbb{E}[X^{2n}] = \frac{\sigma^{2n} (2n)!}{2^n n!}$ et $\mathbb{E}[X^{2n+1}] = 0$.

Ces formules peuvent aussi être établies par récurrence à l'aide d'une intégration par parties.

7 Convergence en loi (*Hors-programme*)

Définition 23 (Convergence en loi). On dit qu'une suite de v.a.r. $(X_n)_{n \geq 1}$ converge en loi vers une v.a.r. X , noté $X_n \xrightarrow[n \rightarrow \infty]{(loi)} X$, lorsque l'une des conditions équivalentes suivantes est vérifiée :

1. Fonctions test : $\mathbb{E}[h(X_n)] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[h(X)]$ pour tout $h: \mathbb{R} \rightarrow \mathbb{R}$ continue et bornée.
2. Fonctions de répartition : $F_{X_n}(t) \xrightarrow[n \rightarrow \infty]{} F_X(t)$ en tout $t \in \mathbb{R}$ où F_X est continue.
3. Fonctions caractéristiques : $\Phi_{X_n}(t) \xrightarrow[n \rightarrow \infty]{} \Phi_X(t)$ en tout point $t \in \mathbb{R}$.

Remarque 25 (Terminologie). On parle aussi de convergence en distribution, notée $X_n \xrightarrow[n \rightarrow \infty]{d} X$.

Exemple 18. Si $X_n \sim \mathcal{U}\left(\left\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\right\}\right)$, alors $X_n \xrightarrow[n \rightarrow \infty]{(loi)} X$ avec $X \sim \mathcal{U}(]0, 1[)$. En effet,

$$F_{X_n}(t) = \mathbb{P}(X_n \leq t) = \begin{cases} 0 & \text{si } t \leq 0 \\ \frac{\lfloor nt \rfloor}{n} & \text{si } 0 < t < 1 \\ 1 & \text{si } t \geq 1 \end{cases}$$

En remarquant que $nt - 1 < \lfloor nt \rfloor \leq nt$, on voit que $\frac{\lfloor nt \rfloor}{n} \rightarrow t$ lorsque $n \rightarrow \infty$, et l'on a donc bien convergence vers la fonction de répartition de la loi uniforme sur $[0, 1]$. On aurait pu également passer par les fonctions tests : pour $h: \mathbb{R} \rightarrow \mathbb{R}$ continue et bornée, on a bien

$$\mathbb{E}[h(X_n)] = \frac{1}{n} \sum_{k=1}^n h\left(\frac{k}{n}\right) \xrightarrow[n \rightarrow \infty]{} \int_0^1 h(x) \mathcal{D}x = \mathbb{E}[h(X)],$$

par le théorème d'approximation des intégrales par les sommes de Riemann. Enfin, si l'on préfère utiliser les fonctions caractéristiques, on a :

$$\Phi_{X_n}(t) = \frac{1}{n} \sum_{k=1}^n e^{\frac{ikt}{n}} = \frac{e^{\frac{it}{n}}(e^{it} - 1)}{n(e^{\frac{it}{n}} - 1)} \xrightarrow[n \rightarrow \infty]{} ace^{it} - 1it = \Phi_X(t).$$

Exemple 19. Dans chacun des cas suivants, on a $X_n \xrightarrow[n \rightarrow \infty]{(loi)} X$.

1. $X_n \sim \mathcal{B}(n, p_n)$ avec $np_n \rightarrow \lambda$, et $X \sim \mathcal{P}(\lambda)$.
2. $X_n \sim \frac{1}{n} \mathcal{G}(p_n)$ avec $np_n \rightarrow \lambda$, et $X \sim \mathcal{E}(\lambda)$
3. $X_n = n \min(U_1, \dots, U_n)$ avec $\{U_n\}_{n \geq 1}$ i.i.d. de loi $\mathcal{U}(]0, 1[)$, et $X \sim \mathcal{E}(1)$.

Proposition 24. Si $X_n \xrightarrow[n \rightarrow \infty]{(loi)} X$ et si $g: \mathbb{R} \rightarrow \mathbb{R}$ est continue alors $g(X_n) \xrightarrow[n \rightarrow \infty]{(loi)} g(X)$.

Démonstration. Immédiat en utilisant la caractérisation par les fonctions test. □

Proposition 25 (Cas des variables discrètes). Soient X, X_1, X_2, \dots des v.a.r. à valeurs dans \mathbb{Z} . Alors, $X_n \xrightarrow[n \rightarrow \infty]{(loi)} X$ si et seulement si pour tout $k \in \mathbb{Z}$,

$$\mathbb{P}(X_n = k) \xrightarrow[n \rightarrow \infty]{} \mathbb{P}(X = k).$$

Exercice 23 (Application). Étudier la convergence en loi de $X_n \sim \mathcal{B}(n, p_n)$ lorsque $np_n \rightarrow \lambda$.

Proposition 26 (Cas des variables à densité). Soient X, X_1, X_2, \dots des v.a.r. admettant des densités $f_X, f_{X_1}, f_{X_2}, \dots$. On suppose que pour (presque) tout $t \in \mathbb{R}$,

$$f_{X_n}(t) \xrightarrow[n \rightarrow \infty]{} f_X(t).$$

Alors on peut conclure que $X_n \xrightarrow[n \rightarrow \infty]{d} X$. Attention, la réciproque est fautive en général!

Exercice 24 (Application). Pour $n \geq 1$, on pose $Z_n := \frac{X_n - n}{\sqrt{n}}$ avec $X_n \sim \Gamma(n, 1)$. Déterminer la densité de Z_n , et en déduire que la suite $(Z_n)_{n \geq 1}$ converge en loi vers une limite que l'on précisera.

Nous énonçons maintenant le théorème central de ce cours, qui raffine la loi des grands nombres.

Théorème 8 (Théorème central limite, ou TCL). Soient $(X_n)_{n \geq 1}$ des v.a.r. i.i.d. de carré intégrable, d'espérance μ et de variance σ^2 . Alors,

$$\frac{(X_1 + \dots + X_n) - n\mu}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{(loi)} Z \sim \mathcal{N}(0, \sigma^2).$$

Autrement dit, pour tous $a \leq b$ réels,

$$\mathbb{P}(X_1 + \dots + X_n \in [n\mu + a\sigma\sqrt{n}, n\mu + b\sigma\sqrt{n}]) \xrightarrow[n \rightarrow \infty]{} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} \mathcal{D}x.$$

Démonstration. On pose $Z_n := \frac{(X_1 + \dots + X_n) - n\mu}{\sqrt{n}}$. Alors pour tout $n \geq 1$, $t \in \mathbb{R}$,

$$\begin{aligned} \Phi_{Z_n}(t) &= \mathbb{E} \left[\prod_{k=1}^n e^{\frac{it(X_k - \mu)}{\sqrt{n}}} \right] \\ &= \prod_{k=1}^n \mathbb{E} \left[e^{\frac{it(X_k - \mu)}{\sqrt{n}}} \right] \\ &= \left(\Phi_{X_1 - \mu} \left(\frac{t}{\sqrt{n}} \right) \right)^n \\ &= \left(1 - \frac{(\sigma t)^2}{2n} + o\left(\frac{1}{n}\right) \right)^n \xrightarrow[n \rightarrow \infty]{} e^{-\frac{(\sigma t)^2}{2}} = \Phi_Z(t). \end{aligned}$$

On a successivement utilisé l'indépendance de X_1, \dots, X_n , le fait qu'elles ont même loi, et le développement de Taylor à l'ordre 2 en zéro de la fonction caractéristique d'une v.a.r. dans L^2 . \square

A Ensembles dénombrables

La notion de *dénombrabilité* joue un rôle important en théorie de la mesure et en probabilités, notamment car une tribu est stable par union *dénombrable*. Il faut donc prendre le temps de bien comprendre ce qu'elle signifie et avoir en tête les principales propriétés et exemples d'ensembles dénombrable ou non.

Intuitivement un ensemble est dénombrable si on peut compter ses éléments. Cette intuition nous conduit à la définition suivante,

Définition 24. *On dit qu'un ensemble E est dénombrable s'il est en bijection avec une partie de \mathbb{N} .*

Dans la définition précédente la partie de \mathbb{N} avec laquelle E est en bijection peut être finie ou infinie. L'ensemble E peut donc lui-même être fini ou infini. Certaines références n'inclut pas les ensembles finis dans les ensembles dénombrables mais ça n'est pas le choix fait ici. Quand un ensemble dénombrable n'est pas fini on précise parfois qu'il est *infini dénombrable* et on peut montrer (le faire !) qu'il est alors en bijection avec \mathbb{N} .

Les ensembles suivants sont dénombrables (voir le TD 0 pour les exercices correspondants aux preuves) :

tous les ensembles finis, \mathbb{N} , \mathbb{Z} , \mathbb{Q} .

La proposition suivante est souvent utile pour établir qu'un ensemble est dénombrable. Là encore, la preuve de ces propriétés constitue un excellent exercice (voir le TD 0).

Proposition 27.

1. *Un sous-ensemble d'un ensemble dénombrable est dénombrable,*
2. *un produit fini d'ensembles dénombrables est dénombrable,*
3. *une union dénombrable d'ensembles dénombrables est dénombrable,*
4. *s'il existe une injection f d'un ensemble E dans un ensemble dénombrable alors E est dénombrable,*
5. *S'il existe une surjection g d'un ensemble dénombrable dans un ensemble E alors E est dénombrable.*

Un exemple important d'ensemble non dénombrable est l'ensemble des réels \mathbb{R} .

Proposition 28. *L'ensemble \mathbb{R} n'est pas dénombrable.*

Démonstration. La clé de la preuve, due à Cantor, est le résultat suivant :

Lemme 13. *L'ensemble E des suites à valeurs dans $\{0, 1\}$ n'est pas dénombrable.*

Remarquons avant la preuve que l'ensemble E est en bijection avec l'ensemble $\mathcal{P}(\mathbb{N})$ des parties de \mathbb{N} . On notera aussi que le résultat reste identique si on remplace l'ensemble $\{0, 1\}$ par n'importe quel ensemble fini.

Preuve du lemme 13. Supposons E dénombrable. On peut donc l'écrire

$$A = \{a^1, a^2, \dots\}$$

où pour tout $p \in \mathbb{N}$, a^p désigne une suite $(a_n^p)_{n \geq 1}$ à valeurs dans $\{0, 1\}$.

On considère maintenant la suite $b \in E$ définie par

$$\forall n \geq 1, \quad b_n = \begin{cases} 0 & \text{si } a_n^n = 1 \\ 1 & \text{si } a_n^n = 0 \end{cases}$$

Par construction, pour tout $p \in \mathbb{N}$, b diffère de a^p puisque $b_p \neq a_p^p$. Ceci est bien sûr absurde puisque la suite b est bien un élément de E . □

Avant de passer à l'étape suivante, nous allons un petit peu améliorer notre lemme. On remarque tout d'abord que le sous-ensemble E' des suites de E constantes égales à 1 à partir d'un certain rang est dénombrable. En effet

$$E' = \bigcup_{N=1}^{+\infty} \{a \in E \text{ tel que } a_n = 1, \forall n \geq N\}$$

est une union dénombrable d'ensembles dénombrables (finis même) donc E' est dénombrable. On en déduit que $E \setminus E'$ n'est pas dénombrable (sinon E serait dénombrable comme union de deux ensembles dénombrables).

Ce lemme amélioré va nous permettre de montrer que $[0, 1[$ n'est pas dénombrable (ce qui implique que \mathbb{R} n'est pas dénombrable). Il faut pour cela prouver que $E \setminus E'$ est en bijection avec $[0, 1[$. On considère le développement dyadique c'est-à-dire l'application qui à $x \in [0, 1[$ associe la suite $a \in E \setminus E'$ définie par récurrence par $a_1 = [2x]$ et

$$\forall n > 1, \quad a_n = [2^n x] - \sum_{k=1}^{n-1} a_k 2^{n-k}.$$

Nous n'allons pas prouver que cette application est bien la bijection recherchée. Il est en revanche important de comprendre géométriquement comment un nombre réel est codé par cette application. Cette dernière propriété (admise donc) nous permet de conclure la preuve : puisque $E \setminus E'$ n'est

pas dénombrable, $[0, 1[$ ne l'est pas non plus et donc \mathbb{R} non plus. □

Notons donc aussi que l'ensemble $\mathbb{R} \setminus \mathbb{Q}$ n'est pas dénombrable sinon \mathbb{R} le serait également comme union de deux ensembles dénombrables.

B Quelques conseils de rédaction

En mathématiques, la rédaction joue un rôle essentiel qui dépasse largement la question du style. Il s'agit de donner forme à votre raisonnement et de le communiquer à votre lecteur. La façon d'écrire un raisonnement est codifiée et la rédaction désigne donc la maîtrise de ce code et non une question d'élégance. Une faute de rédaction est donc similaire à une faute de logique. Souvent les rédactions peu compréhensibles dans les copies sont synonymes de raisonnements faux. Voici quelques conseils pour vous aider à rédiger correctement. Cette liste n'est pas exhaustive !

1. Il faut **séparer la partie mathématique du texte écrite sous forme de proposition de la partie écrite en français**. Par exemple, \forall, \exists sont des quantificateurs logiques que vous ne devez pas utiliser comme abréviations mais uniquement comme symboles dans les phrases logiques.
2. Il faut **introduire toutes les notations que vous utilisez**.
3. Toutes les variables n'ont pas la même **durée de vie**. Par exemple :
 - Les phrases "Soit $\varepsilon > 0$." ou "Il existe donc $\varepsilon > 0$." introduisent un réel que l'on appelle ε . Pour toute la durée de la preuve la lettre ε désignera ce réel.
 - Dans la proposition " $\forall \varepsilon > 0, \varepsilon/2 > 0$ " ou encore dans " $\exists \varepsilon > 0$ " la variable ε est quantifiée par \forall ou \exists . La lettre ε n'est alors définie que le temps de la phrase logique et non plus pour toute la durée de la preuve. Sortie de la phrase logique, elle n'a plus de sens !
 - Quand on écrit $\sum_{i=1}^{10} i^2$, la lettre i désigne un indice muet de sommation qui n'a de sens que le temps de la somme. En dehors de cette formule, la lettre i ne désigne plus rien.Il faut respecter ces durées de vie. Changer la valeur d'une variable déjà affectée est une faute logique.
4. Voici un **exemple important** de rédaction d'une proposition du type :

$$\forall \varepsilon > 0 \exists \alpha > 0 \text{ tel que } P(\varepsilon, \alpha)$$

où P désigne une proposition logique.

La rédaction de la preuve d'une telle proposition peut (c'est en fait très souvent le cas) s'organiser de la façon suivante :

- (a) *Première étape*. On doit montrer que quelque chose est vraie pour tout $\varepsilon > 0$ (la proposition : $\exists \alpha > 0$ tel que $P(\varepsilon, \alpha)$). On fixe donc arbitrairement un ε et on va prouver la proposition pour cet ε arbitraire. La phrase type permettant de fixer ε est :
Soit $\varepsilon > 0$.
On a désormais fixé un ε pour toute la preuve et on va pouvoir travailler avec.

(b) *Deuxième étape.* L'étape suivante est de proposer un $\alpha > 0$ tel que $P(\varepsilon, \alpha)$ est vrai (où ε est celui introduit à l'étape précédente). On va donc travailler pour cela en utilisant des théorèmes, des calculs ou des raisonnements variés. A l'issue de ce travail on doit avoir construit ou exhiber un certain α . Le résultat de ce travail doit être clairement indiquée, par exemple :

On pose $\alpha = \dots$

ou

Il existe donc α tel que ...

(c) *Troisième étape.* La dernière étape consiste à prouver que $P(\alpha, \varepsilon)$ est vrai où ε et α sont ceux obtenus à l'issue des deux étapes précédentes. Ce qui peut de nouveau nécessiter raisonnements et autres calculs... On marque souvent la conclusion par :

On a donc bien $P(\varepsilon, \alpha)$.

Entraînez vous à souligner dans vos démonstrations ces étapes clés (en gras dans l'exemple au-dessus) qui structurent le raisonnement !

5. **Bien utiliser les théorèmes :**

(a) Vérifier soigneusement toutes les hypothèses du théorème utilisé.

(b) Nommer explicitement le théorème utilisé.

6. Familiarisez vous avec les **différents types de raisonnements** et leur rédaction type (absurde, contraposée, récurrence,...)