

# Probabilités continues<sup>1</sup>

Pascal Maillard & Olivier Hénard

29 août 2024

1. Notes du cours « Probabilités » (MDD303) de la troisième année de licence double diplôme » à l'Université Paris Saclay



# Table des matières

<b>0</b>	<b>Préliminaires</b>	<b>7</b>
0.1	Bibliographie conseillée . . . . .	7
0.2	Bagage mathématique . . . . .	7
0.3	Rappels du cours de probas en L2 . . . . .	8
0.3.1	Probabilités . . . . .	8
0.3.2	Variables aléatoires . . . . .	8
0.3.3	Par la suite. . . . .	9
<b>1</b>	<b>Variables aléatoires réelles</b>	<b>11</b>
1.1	Définition et axiomes . . . . .	11
1.1.1	Axiomes des probabilités . . . . .	12
1.2	Événements . . . . .	14
1.3	Fonction de répartition . . . . .	15
1.4	Classes de variables aléatoires . . . . .	17
1.5	Quelques variables aléatoires discrètes importantes . . . . .	22
1.5.1	Loi de Dirac . . . . .	23
1.5.2	Loi uniforme . . . . .	23
1.5.3	Loi de Bernoulli . . . . .	23
1.5.4	Loi binomiale . . . . .	24
1.5.5	Loi de Poisson . . . . .	24
1.5.6	Loi géométrique . . . . .	25
1.5.7	Autres lois discrètes . . . . .	27
1.6	Quelques variables aléatoires à densité importantes . . . . .	28
1.6.1	Loi uniforme . . . . .	28
1.6.2	Loi normale (ou gaussienne) . . . . .	29
1.6.3	Loi exponentielle . . . . .	31
1.6.4	Loi gamma. . . . .	32
1.6.5	Loi Beta. . . . .	33
1.6.6	Loi de Cauchy. . . . .	34
1.6.7	Autres lois à densité . . . . .	34
1.7	Loi d'une fonction d'une variable aléatoire . . . . .	35
1.7.1	Fonction discrète d'une v.a.r. . . . .	35
1.7.2	Fonction régulière d'une variable aléatoire à densité . . . . .	36
1.8	Simulation de lois . . . . .	37

<b>2</b>	<b>Espérance et moments</b>	<b>41</b>
2.1	Espérance d'une variable aléatoire . . . . .	41
2.1.1	Variable aléatoire positive ou nulle . . . . .	41
2.1.2	Variable aléatoire signée . . . . .	43
2.2	Espérance d'une fonction d'une variable aléatoire . . . . .	43
2.2.1	Caractérisation des lois de fonctions de variables aléatoires . . . . .	47
2.3	Moments et quantités associées . . . . .	49
2.3.1	Espérance . . . . .	50
2.3.2	Variance . . . . .	50
2.4	Fonction génératrice des moments . . . . .	52
2.5	Inégalités . . . . .	55
2.5.1	Inégalité de Markov . . . . .	55
2.5.2	Inégalité de Tchebychev . . . . .	55
2.5.3	Inégalité de Jensen . . . . .	56
2.6	Autres définitions de l'espérance . . . . .	57
2.6.1	À la Lebesgue. . . . .	57
2.6.2	À la Riemann-Stieltjes . . . . .	59
<b>3</b>	<b>Variables aléatoires indépendantes</b>	<b>61</b>
3.1	Variables aléatoires simultanées . . . . .	61
3.1.1	Variables aléatoires discrètes . . . . .	63
3.1.2	Vecteur aléatoire à densité . . . . .	63
3.1.3	Fonction d'un vecteur aléatoire . . . . .	65
3.2	Variables aléatoires indépendantes . . . . .	66
3.3	Fonctions de variables aléatoires indépendantes . . . . .	70
3.3.1	Maximum et minimum . . . . .	71
3.3.2	Somme . . . . .	72
3.3.3	Cas général . . . . .	75
<b>4</b>	<b>Suites de variables aléatoires et théorèmes limites</b>	<b>83</b>
4.1	Convergence de suites de variables aléatoires . . . . .	83
4.1.1	Convergence en loi . . . . .	83
4.1.2	Convergence en probabilité . . . . .	86
4.1.3	Convergence dans $L^p$ . . . . .	88
4.1.4	Convergence presque sûre . . . . .	88
4.2	Convergence de couples aléatoires . . . . .	89
4.3	Loi des grands nombres . . . . .	91
4.4	Théorème central limite . . . . .	92
<b>5</b>	<b>Variables aléatoires dépendantes</b>	<b>95</b>
5.1	Lois marginales . . . . .	95
5.2	Lois conditionnelles . . . . .	97
5.3	Covariance et corrélation . . . . .	98

<b>6</b>	<b>Appendice : Toolbox.</b>	<b>103</b>
6.1	Ensemble dénombrable . . . . .	103
6.2	Familles sommables . . . . .	103
6.3	Séries de fonctions et intégrales à paramètre . . . . .	105
6.4	Théorèmes de Fubini et Tonelli . . . . .	108



# Chapitre 0

## Préliminaires

### 0.1 Bibliographie conseillée

Pour le contenu du cours :

- APPEL, W., *Probabilités pour les non probabilistes*, HK (une somme assez impressionnante, dans un style très prépa qui pourra plaire à certains lecteurs)
- ROSS, S., *Initiation aux probabilités*, Presses polytechniques et universitaires romandes
- BARBE, P. ET LEDOUX, M., *Probabilité L3/M1*, EDP Sciences (niveau plus élevé, inclut notamment la théorie de la mesure)

Pour les préliminaires en mathématiques et en probabilités en particulier, voir ci-dessous.

### 0.2 Bagage mathématique

Les probabilités mettent en œuvre un grand nombre d'outils mathématiques, surtout du domaine de l'analyse. Pour profiter pleinement du cours, il est donc impératif de rafraîchir ses connaissances en analyse et d'être à l'aise en calcul. En particulier, il est important de revoir les sujets suivants qui ont été traités en L1-L2 (liste non exhaustive) :

- *suites et limites*
- *sommes et séries* : convergence (absolue ou non), somme télescopique, changement d'indice, double sommes, séries entières, valeurs des séries classiques : somme des entiers, série géométrique, ...
- *fonctions* : continuité, définition de la dérivée, dérivation des produits, des composées, des fonctions réciproques, DLs, développement de Taylor, propriétés des fonctions classiques (polynômes, exponentielle, logarithme, fonctions trigonométriques, et fonctions trigonométriques inverses), comportement asymptotique
- *intégration* : intégrale sur un intervalle borné, intégrale impropre, intégration par parties, changement de variables
- *séries de fonctions et intégrales à paramètre* : convergence simple/uniforme/normale, échange limite  $\longleftrightarrow$  somme, limite  $\longleftrightarrow$  intégrale, dérivation sous les signes somme et intégrale

Ces sujets sont bien exposés dans un grand nombre de livres et de notes de cours. Le dernier sujet (séries de fonctions et intégrales à paramètre) est peut-être moins bien traité, c'est pourquoi nous donnons ci-dessous les résultats à retenir. Nous aurons également besoin de résultats sur les sujets suivants traités dans le cours d'Intégration en L3 :

- *théorèmes de Fubini et Tonelli* : interversion série  $\longleftrightarrow$  série, série  $\longleftrightarrow$  intégrale, intégrale  $\longleftrightarrow$  intégrale
  - *intégrales multiples* : intégration sur un domaine de  $\mathbb{R}^d$ , changement de coordonnées
- Nous renvoyons à l'appendice de ce cours pour ces sujets.

### 0.3 Rappels du cours de probas en L2

On rappelle quelques éléments de probabilités vus en L2. Référence : notes de cours d'Anne Broise (disponible sur Dokeos).

#### 0.3.1 Probabilités

On considère une expérience aléatoire ayant un nombre fini ou dénombrable de *résultats* possibles. Exemple : jet d'un ou de plusieurs dés, réalisation d'un sondage avec un nombre fini de réponses, mesure du nombre de particules émis par une matière radioactive pendant un certain temps, observation du nombre de personnes dans une file d'attente à un moment donné, nombre d'enfants d'une personne prise au hasard dans une population... On formalise une telle expérience souvent par un ensemble fini ou *dénombrable*<sup>1</sup>  $\Omega$  qui contient tous les résultats possibles. Un élément de  $\Omega$  est typiquement noté  $\omega$  et on suppose qu'on peut lui associer une *probabilité*  $p(\omega)$  qui est un nombre entre 0 et 1. On suppose de plus que la somme des probabilités de tous les résultats vaut 1 :

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

La fonction  $p$  est alors parfois appelée *fonction de masse* où *germe de probabilité*. Ce dernier nom s'explique par le fait que cette fonction permet de donner une probabilité à n'importe quel *événement*, un événement étant un ensemble de résultats possédant une certaine caractéristique commune. On penserait par exemple à l'événement « nombre pair » lors du lancer d'un dé de six faces, qui correspond à l'ensemble  $\{2, 4, 6\}$ . Formellement, un *événement* n'est alors rien d'autre qu'une *partie* de  $\Omega$ . On les note typiquement par des lettres majuscules du début de l'alphabet :  $A, B, C, \dots$ . La *probabilité* d'un événement  $A \subset \Omega$ , notée  $P(A)$  dans ce cours, est alors la somme des probabilités des résultats que comprend cet événement :

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

La fonction  $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$  est également appelée *probabilité* (ou *mesure de probabilité*) sur  $\Omega$  et la fonction  $p$  est alors appelée sa *fonction de masse* ou son *germe*. Par exemple, dans le cas du dé ci-dessus, on prendrait typiquement  $\Omega = \{1, \dots, 6\}$  et  $p(\omega) = 1/6$  pour tout  $\omega \in \Omega$ .

#### 0.3.2 Variables aléatoires

Il arrive que l'espace  $\Omega$  soit trop gros pour travailler avec directement ou qu'on ne s'intéresse qu'à certaines propriétés du résultat de l'expérience. Typiquement, on associe alors un nombre réel à un résultat qui contient l'information qu'on cherche. Par exemple, lors d'un sondage on pourrait définir  $X$  comme étant le numéro du candidat ayant remporté le plus de

1. On renvoie le lecteur à l'annexe pour la définition d'un ensemble dénombrable



voix.  $X$  est alors une variable qui peut prendre n'importe quelle valeur dans  $\{1, \dots, n\}$ , où  $n$  est le nombre de candidats, et sa valeur est aléatoire, c'est-à-dire elle dépend du résultat de l'expérience aléatoire. On l'appelle alors une *variable aléatoire réelle* ou tout simplement une *variable aléatoire*<sup>2</sup>. On note les variables aléatoires typiquement par des lettres majuscules de la fin de l'alphabet :  $X, Y, Z, \dots$ . Remarquons aussi que formellement, une variable aléatoire  $X$  n'est rien d'autre qu'une fonction  $X : \Omega \rightarrow \mathbb{R}$ .

### 0.3.3 Par la suite...

*Le moment est venu pour une remarque importante qui va simplifier grandement la suite des opérations.*

Il s'est avéré au fil du temps qu'il est souvent plus commode de travailler entièrement avec des variables aléatoires et d'ignorer complètement l'espace  $\Omega$ . Les avantages sont multiples :

- On peut modifier  $\Omega$  (par exemple en ajoutant des informations supplémentaires sur le résultat de l'expérience) sans avoir à réécrire des calculs, énoncés etc. concernant des variables aléatoires.
- On n'est pas obligé de préciser  $\Omega$ , seulement des éléments qui nous intéressent.
- On peut plus facilement formuler des énoncés généraux ne dépendant pas de la définition précise d'un espace  $\Omega$ .
- Dans le cas où  $\Omega$  ne peut pas être fini ou dénombrable, on n'est pas obligé de considérer des espaces généraux et abstraits, mais seulement des variables aléatoires à valeurs dans  $\mathbb{R}$ , donc des probabilités sur  $\mathbb{R}$ .
- Philosophiquement, cette approche peut être considérée plus satisfaisante dans le sens où dans une situation réelle on ne connaît jamais entièrement « l'univers » (donc  $\Omega$ ) mais plutôt partiellement à partir d'observations (les variables aléatoires).
- Enfin, techniquement, si effectivement la théorie des probabilités repose sur la théorie de la mesure (abstraite) et de l'intégration, les objets d'étude sont distincts dans les deux domaines ; en probabilité, l'espace  $\Omega$  est souvent relégué au second plan, les quantités d'intérêt étant souvent définies en fonction des seules *lois* des variables aléatoires en jeu. Si une compréhension fine des probabilités passe par l'espace  $\Omega$ , on gagnera en première approche à ignorer  $\Omega$ .

*Par conséquent, dans la suite du cours, nous n'allons plus entendre parler d'espace  $\Omega$ , mais très souvent de variables aléatoires.*

---

2. On peut considérer des variables aléatoires qui ne prennent pas valeurs dans l'ensemble des nombres réelles mais dans des espaces plus généraux, mais nous n'allons pas considérer ces variables aléatoires dans ce cours.



# Chapitre 1

## Variables aléatoires réelles

### 1.1 Définition et axiomes

**Une fonction de masse pour une variable aléatoire continue ?** On cherche à modéliser une expérience ayant un résultat aléatoire. On suppose qu'on observe le résultat à partir d'une *variable aléatoire (v.a.)*  $X$  qui prend ses valeurs dans les nombres réels. Précédemment, nous avons toujours supposé qu'une variable aléatoire pouvait prendre seulement un nombre fini ou dénombrable de valeurs, on dit dans ce cas que c'est une variable aléatoire *discrète*. Maintenant nous souhaitons nous affranchir de cette contrainte. Mais nous allons nous faciliter la vie : nous ne chercherons pas à définir formellement ce qu'est cette variable aléatoire mais seulement ce qu'on peut faire avec. Concrètement, on souhaite donner un sens à la probabilité que la valeur de cette v.a.  $X$  tombe dans un ensemble  $A \subset \mathbb{R}$  donné, c'est-à-dire, à l'expression  $P(X \in A)$ .

On se trouve ici confronté à un nouveau problème : précédemment, quand une variable aléatoire ne pouvait prendre qu'un nombre fini ou dénombrable de valeurs (notons l'ensemble de valeurs  $S_X$ ), on pouvait toujours écrire la probabilité  $P(X \in A)$  comme la somme  $\sum_{x \in A} P(X = x)$ . Donc il suffisait de donner la valeur de  $P(X = x)$  pour tout  $x \in S_X := \{x : P(X = x) \neq 0\}$  (ce qu'on appelle *la fonction de masse*) pour définir les probabilités  $P(X \in A)$  pour tout  $A \subset S_X$ , c'est-à-dire la loi de la variable aléatoire  $X$ . Il s'avère que cela n'est plus le cas en général. Par exemple, supposons qu'on souhaite donner un sens à un nombre choisi uniformément dans l'intervalle  $[0, 1]$ . On va modéliser cela par une variable aléatoire  $X$  qui prend ses valeurs dans  $[0, 1]$ . Quelle doit être la probabilité que la variable aléatoire tombe dans un ensemble  $A \subset [0, 1]$  donné ? Considérons pour l'instant seulement le cas où  $A$  consiste en le singleton  $\{x\}$  pour  $x \in [0, 1]$  réel donné. Puisque  $X$  correspond à un nombre choisi uniformément dans l'intervalle  $[0, 1]$ , on s'attend à ce que la probabilité que  $X$  soit égal à un nombre donné ne dépende pas de ce nombre, autrement dit que :

$$P(X = x) = P(X = y) \text{ pour tout } x, y \in [0, 1].$$

Mais puisque l'intervalle  $[0, 1]$  contient un nombre infini de points, cela entraîne que  $P(X = x) = 0$  pour tout  $x \in [0, 1]$  (dans le cas inverse, on pourrait déduire que  $P(X \in [0, 1]) = \infty$ ). La connaissance de  $P(X = x)$  pour tout  $x$  ne permet donc pas de définir  $P(X \in A)$  pour tout ensemble  $A$ .

### 1.1.1 Axiomes des probabilités

On doit alors trouver un autre moyen pour définir les probabilités  $P(X \in A)$  pour des ensembles  $A \subset \mathbb{R}$ , c'est-à-dire pour donner la loi de  $X$ . L'observation importante est la suivante : ces probabilités doivent obéir à certaines règles, par exemple, si  $A, B \subset \mathbb{R}$  sont disjoints (c'est-à-dire d'intersection vide), alors nous souhaitons que

$$P(X \in A \cup B) = P(X \in A) + P(X \in B).$$

Pour définir toutes les lois possibles d'une variable aléatoire  $X$  à valeurs dans  $\mathbb{R}$ , on commence alors par énoncer un certain nombre de règles, appelés *axiomes*. Pour simplifier le développement par la suite, nous introduisons à ce stade une convention :

*Par la suite, chaque ensemble  $A \subset \mathbb{R}$  est supposé être de la forme suivante : soit un intervalle, soit la réunion d'un nombre fini d'intervalles disjoints. La notion d'intervalle inclut les intervalles ouverts, semi-ouverts ou fermés, qu'ils soient finis ou infinis, et en particulier l'ensemble vide  $\emptyset$  ainsi que l'ensemble  $\mathbb{R}$  tout entier.*

On note  $A^c = \mathbb{R} \setminus A$  le *complémentaire* de  $A$ . On remarque que si  $A \subset \mathbb{R}$  est de la forme ci-dessus, alors  $A^c$  l'est aussi (preuve laissée en exercice). Également, si  $A$  et  $B$  sont de la forme ci-dessus, alors  $A \cup B$  et  $A \cap B$  le sont aussi (exercice). Par contre, si  $A_0, A_1, A_2, \dots$  est une suite d'ensembles de la forme ci-dessus, alors  $\bigcup_{n \in \mathbb{N}} A_n$  et  $\bigcap_{n \in \mathbb{N}} A_n$  ne sont plus nécessairement de cette forme.

Nous allons préciser maintenant les règles de calcul des quantités  $P(X \in A)$ , c'est-à-dire les *axiomes* qu'on suppose être valables : on suppose que

1.  $P(X \in A) \geq 0$  pour tout  $A \in \mathbb{R}$  (*positivité*)
2.  $P(X \in \emptyset) = 0$  et  $P(X \in \mathbb{R}) = 1$ ,
3. Si  $A, B \subset \mathbb{R}$  sont disjoints, alors

$$P(X \in A \cup B) = P(X \in A) + P(X \in B) \quad (\text{additivité}).$$

Nous allons ensuite introduire un quatrième axiome, mais tout d'abord, remarquons que ces axiomes entraînent les propriétés suivantes :

- Si  $A \subset B$ , alors

$$P(X \in A) \leq P(X \in B) \quad (\text{monotonie de } A \mapsto P(X \in A)).$$

- $P(X \in A) \in [0, 1]$  pour tout  $A \in \mathbb{R}$ .
- Si  $A \subset B$ , alors

$$P(X \in B \setminus A) = P(X \in B) - P(X \in A).$$

En particulier ( $B = \mathbb{R}$ ),

$$P(X \in A^c) = 1 - P(X \in A).$$

- Si  $A_0, A_1, \dots, A_n \subset \mathbb{R}$  sont deux à deux disjoints, alors

$$P\left(X \in \bigcup_{k=0}^n A_k\right) = \sum_{k=0}^n P(X \in A_k) \quad (\text{additivité finie}).$$

La propriété de monotonie et la formule pour  $P(X \in B \setminus A)$  se montrent ainsi : pour  $A \subset B$ , on écrit  $B$  sous la forme de la réunion disjointe  $B = A \cup (B \setminus A)$ . L'axiome 3 donne alors :

$$P(X \in B) = P(X \in A) + P(X \in B \setminus A).$$

La propriété de monotonie en découle puisque  $P(X \in B \setminus A) \geq 0$  par l'axiome 1. La formule  $P(X \in B \setminus A)$  s'obtient en réarrangeant. Finalement, le fait que  $P(X \in A) \in [0, 1]$  pour tout  $A \in \mathcal{R}$  est une simple conséquence de la monotonie appliquée à  $\emptyset \subset A \subset \mathbb{R}$ . La propriété d'additivité finie découle de la propriété d'additivité par une récurrence finie.

On peut finalement présenter le quatrième axiome :

4. Si  $A_0, A_1, A_2, \dots$  est une suite croissante d'ensembles (c'est-à-dire  $A_n \subset A_{n+1}$  pour tout  $n \in \mathbb{N}$ ) et telle que  $\bigcup_{n \in \mathbb{N}} A_n$  est encore un ensemble de la forme ci-dessus, alors

$$P\left(X \in \bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} P(X \in A_n).$$

Notons que cette limite existe puisque  $P(X \in A_n)$  est une suite croissante par la monotonie de  $A \mapsto P(X \in A)$  énoncée ci-dessus.

Cet axiome nécessite éventuellement une explication : avec les axiomes 1 à 3, on avait déjà l'inégalité «  $\geq$  » (exercice). Le vrai énoncé de l'axiome 4 est qu'on a aussi l'inégalité inverse, à savoir «  $\leq$  ». Intuitivement, cela signifie que lors un passage à la limite, on ne peut pas créer de la probabilité *ex nihilo*.

Remarquons aussi que l'axiome 4 est équivalent à l'axiome suivant, comme le démontre un passage au complémentaire :

- 4'. Si  $A_0, A_1, A_2, \dots$  est une suite décroissante d'ensembles (c'est-à-dire  $A_n \supset A_{n+1}$  pour tout  $n \in \mathbb{N}$ ) et telle que  $\bigcap_{n \in \mathbb{N}} A_n$  est encore un ensemble de la forme ci-dessus, alors

$$P\left(X \in \bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} P(X \in A_n).$$

Les axiomes mis en place, on peut passer à la définition d'une variable aléatoire :

**Définition 1.1.** On dit que  $X$  est une *variable aléatoire* s'il existe une famille de nombres  $(P(X \in A))_{A \subset \mathbb{R}}$  (ou, autrement dit, une *fonction*  $A \mapsto P(X \in A)$  à valeurs dans  $\mathbb{R}$ ) qui satisfait aux axiomes 1, 2, 3 et 4 ci-dessus. Cette famille/fonction est appelée la *loi* de  $X$ .

Une conséquence importante de l'axiome 4 et de l'axiome d'additivité est la propriété de *sigma-additivité*<sup>1</sup> :

- 4". Si  $A_0, A_1, A_2, \dots$  est une suite d'ensembles deux à deux *disjoints* (c'est-à-dire que  $m \neq n$  implique  $A_m \cap A_n = \emptyset$ ), alors

$$P\left(X \in \bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(X \in A_n) \quad (\text{sigma-additivité}).$$

---

1. "sigma" fait en général référence à une opération dénombrable

Posons en effet  $B_n = \cup_{k=0}^n A_k$  (réunion disjointe), alors  $B_0, B_1, B_2, \dots$  est une suite croissante d'ensembles qui vérifie  $\cup_{n \in \mathbb{N}} B_n = \cup_{n \in \mathbb{N}} A_n$ , appliquant alors successivement l'axiome 4 (deuxième égalité) et la propriété d'additivité (troisième égalité), on obtient :

$$P\left(X \in \bigcup_{n \in \mathbb{N}} A_n\right) = P\left(X \in \bigcup_{n \in \mathbb{N}} B_n\right) = \lim_{n \rightarrow \infty} P(X \in B_n) = \lim_{n \rightarrow \infty} \sum_{k=0}^n P(X \in A_k),$$

et cette dernière expression correspond bien à  $\sum_{n \in \mathbb{N}} P(X \in A_n)$  par définition.

**Proposition 1.2.** *Les jeux d'axiome suivants sont équivalents :*

1. 1, 2, 3, 4
2. 1, 2, 3, 4'
3. 1, 2, 4''

*Parmi ces jeux d'axiomes équivalents, il est courant de retenir le jeu d'axiomes suivants : 1 2 et 4''. C'est ces 3 axiomes qui figurent usuellement dans définition d'une mesure de probabilité sur  $\mathbb{R}$ .*

Notons que nous aurons également parfois besoin de définir  $P(X \in A)$  pour  $A$  un ensemble qui n'est pas de la forme ci-dessus, par exemple  $A = \mathbb{N}$  ou n'importe quel ensemble dénombrable. On se bornera aux ensembles  $A$  qui sont la *limite monotone* d'une suite d'ensembles  $(A_n)_{n \geq 0}$  ayant la forme ci-dessus (réunion d'un nombre fini d'intervalles). Par limite monotone on entend les ensembles  $\bigcup_n A_n$  (ou  $\bigcap_n A_n$ ) pour une suite croissante (ou décroissante)  $(A_n)_{n \geq 0}$ , comme dans les axiomes 4 et 4'. On *définit* alors  $P(X \in A) = \lim_{n \rightarrow \infty} P(X \in A_n)$ , où on note que la limite est toujours bien définie par monotonie de la suite  $P(X \in A_n)$ . Un exemple d'un tel ensemble est l'ensemble des rationnels  $\mathbb{Q}$ , ou n'importe quel ensemble dénombrable  $S \subset \mathbb{R}$ . Pour un tel ensemble, notons qu'avec la définition ci-dessus, on a  $P(X \in S) = \sum_{x \in S} P(X = x)$ .

## 1.2 Événements

Nous rappelons qu'un *événement* est simplement un sous-ensemble de l'ensemble des résultats d'une expérience aléatoire, et l'événement est dit "réalisé" si le résultat de l'expérience fait partie de cet ensemble. Par exemple, lors d'un lancer de dé, on pourrait s'intéresser à l'événement  $E$  "le dé donne un nombre pair". Si  $X$  est la v.a. qui représente le nombre donné par le dé, cet événement s'écrit alors symboliquement  $E = \{X \in \{2, 4, 6\}\}$ , ou encore  $E = \{X \text{ pair}\}$ . L'événement  $E$  se réalise alors avec probabilité  $P(E) = P(X \in \{2, 4, 6\})$  (notez qu'on n'écrit plus les accolades dans cette dernière expression).

Dans ce cours, nous considérons seulement les événements qui sont définis à partir de variables aléatoires, donc des événements de la forme  $E = \{X \in A\}$  pour  $X$  une variable aléatoire et  $A \subset \mathbb{R}$ . La probabilité d'un tel événement est alors définie comme  $P(E) = P(X \in A)$ . On peut également appliquer les opérations habituelles pour les ensembles à ces événements. Par exemple :

$$\begin{aligned} \{X \in A\}^c &= \{X \in A^c\} \\ \{X \in A\} \cup \{X \in B\} &= \{X \in A \cup B\} \\ \{X \in A\} \cap \{X \in B\} &= \{X \in A \cap B\} \end{aligned}$$

**Définition 1.3.** Un événement de probabilité 1 est dit presque sûr. Un événement de probabilité nulle est dit négligeable.

Ainsi, un événement négligeable est le complémentaire d'un événement presque sûr. On rencontre quelquefois la notation :

$$X \in A, \text{ p.s.}^2 \quad \text{pour un événement } A \text{ presque sûr}$$

et qui se lit : « presque sûrement,  $X$  appartient à  $A$  ».

### 1.3 Fonction de répartition

Nous avons vu dans la Section 1.1 la définition de la loi d'une variable aléatoire à partir des axiomes 1 à 4. La loi d'une variable aléatoire  $X$  est donnée par une famille de nombres réels indicée par les ensembles  $A \subset \mathbb{R}$ . Cette définition semble assez lourde si on veut définir une loi particulière, car il faudrait prendre en compte *tous* les ensembles  $A \subset \mathbb{R}$  qui peuvent avoir des formes assez complexes. Dans le cas d'une v.a. discrète  $X$  (c'est-à-dire pour laquelle  $X$  ne prend qu'un nombre fini ou dénombrable de valeurs, étude à venir), on peut se ramener aux valeurs  $P(X = x)$  pour tout  $x$  (c'est-à-dire à la *fonction de masse*), ce qui représente une grande simplification. Nous avons vu dans la section précédente que ceci n'est plus possible dans le cadre général, car ces valeurs peuvent être égales à zéro pour tout  $x \in \mathbb{R}$ . L'astuce consiste alors à considérer les valeurs de  $P(X \leq x)$  pour tout  $x \in \mathbb{R}$ . (Remarquons qu'on aurait également pu prendre  $P(X < x)$ ,  $P(X \geq x)$  ou  $P(X > x)$ , mais la convention a retenu  $P(X \leq x)$ ). Ceci donne lieu à une fonction  $F_X : \mathbb{R} \rightarrow [0, 1]$  définie par

$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R}.$$

La fonction  $F_X$  s'appelle la *fonction de répartition de  $X$* .

Il nous reste à montrer que la fonction de répartition d'une variable aléatoire définit sa loi et à déterminer quelles fonctions sont des fonctions de répartition d'une variable aléatoire.

**Théorème 1.4.** 1. La fonction de répartition  $F_X$  d'une v.a.r. a les propriétés suivantes :

- elle est croissante,
- $F_X(-\infty) := \lim_{x \rightarrow -\infty} F_X(x) = 0$ ,
- $F_X(+\infty) := \lim_{x \rightarrow +\infty} F_X(x) = 1$ ,
- elle est continue à droite.

2. La loi d'une v.a.r.  $X$  est déterminée de manière unique par sa fonction de répartition  $F_X$ .

3. Chaque fonction sur  $\mathbb{R}$  ayant les propriétés données dans 1. est la fonction de répartition d'une v.a. réelle.

Ainsi avec cette fonction on a atteint le même degré de simplicité qu'avec la fonction de masse dans le cas d'une v.a. discrète, car dans les deux cas nous avons réduit la notion de loi à la notion d'une fonction sur  $\mathbb{R}$ , un objet que nous connaissons depuis le collège ! Notons que la simplicité de la fonction de répartition est conceptuelle, car bien souvent, elle n'admettra pas d'expression numérique particulièrement simple.

---

2. et dans la littérature anglo-saxonne, a.s., pour almost surely, remplace p.s.

*Démonstration.*

1. Montrons les propriétés énoncées :

- Croissance : Ceci est une conséquence immédiate de la monotonie de  $P(X \in A)$  en  $A$ . Plus précisément, soient  $x, y \in \mathbb{R}$  avec  $x < y$ . Alors  $] -\infty, x] \subset ] -\infty, y]$ , si bien que

$$F_X(x) = P(X \in ] -\infty, x]) \leq P(X \in ] -\infty, y]) = F_X(y).$$

- On remarque d'abord que la fonction  $F_X$  est croissante et bornée inférieurement (par 0), la limite  $F_X(-\infty) = \lim_{x \rightarrow -\infty} F_X(x)$  existe donc bien. En particulier,

$$F_X(-\infty) = \lim_{n \rightarrow \infty} F_X(-n) = \lim_{n \rightarrow \infty} P(X \in ] -\infty, -n]).$$

La suite d'ensembles  $A_n = ] -\infty, -n]$  est décroissante et d'intersection vide. L'axiome 4' donne donc :

$$F_X(-\infty) = P\left(X \in \bigcap_{n=1}^{\infty} ] -\infty, -n]\right) = P(X \in \emptyset) = 0.$$

- Pareil que  $F_X(-\infty)$ , on utilise le fait que  $F_X$  est bornée par 1 pour montrer l'existence de la limite, puis l'axiome 4 pour l'identifier.
- Continuité à droite. Soit  $x \in \mathbb{R}$ . Puisque  $F_X$  est croissante et bornée inférieurement, la limite  $F_X(x+) := \lim_{y \downarrow x} F_X(y)$  existe. En particulier,

$$F_X(x+) = \lim_{n \rightarrow \infty} F_X(x + \frac{1}{n}) = P(X \in \bigcap_{n=1}^{\infty} ] -\infty, x + \frac{1}{n}]),$$

où on a encore utilisé l'axiome 4'. Le point clé est alors l'égalité suivante :

$$\bigcap_{n=1}^{\infty} ] -\infty, x + \frac{1}{n}] = ] -\infty, x].$$

Cela donne que  $F_X(x+) = P(X \in ] -\infty, x]) = F_X(x)$  et donc  $F_X$  est continue à droite en  $x$ . Puisque  $x \in \mathbb{R}$  était arbitraire,  $F_X$  est continue à droite sur  $\mathbb{R}$ .

2. On veut montrer que pour tout ensemble  $A$  qui est une union finie d'intervalles disjoints,  $P(X \in A)$  s'écrit à partir de la fonction de répartition  $F_X$ . On procède cas par cas :

- $A = ] -\infty, x]$ ,  $x \in \mathbb{R}$  : Par définition,  $P(X \in ] -\infty, x]) = P(X \leq x) = F_X(x)$ .
- $A = ] -\infty, x[$ ,  $x \in \mathbb{R}$  : L'intervalle ouvert  $] -\infty, x[$  s'écrit comme l'union dénombrable d'intervalles fermés :

$$] -\infty, x[ = \bigcup_{n=1}^{\infty} ] -\infty, x - \frac{1}{n}].$$

L'axiome 4 donne alors

$$P(X < x) = \lim_{n \rightarrow \infty} P(X \leq x - \frac{1}{n}) = \lim_{n \rightarrow \infty} F_X(x - \frac{1}{n}) = F_X(x-),$$

où  $F_X(x-) := \lim_{y \uparrow x} F_X(y)$  qui existe car la fonction  $F_X$  est croissante.

- $A = ]x, \infty[$  ou  $A = [x, \infty[$ ,  $x \in \mathbb{R}$  : On utilise le fait que  $F(A) = 1 - F(A^c)$  et on se ramène aux deux premiers cas.



—  $A = ]x, y]$ ,  $x < y$  : On obtient

$$P(x < X \leq y) = P(X \leq y) - P(X \leq x) = F_X(y) - F_X(x).$$

—  $A = [x, y]$ ,  $A = [x, y[$  ou  $A = ]x, y[$  : Similaire à  $A = ]x, y]$ , avec éventuellement  $F(y-)$  et  $F(x-)$  à la place de  $F(y)$  ou  $F(x)$ . Par exemple

$$\mathbb{P}(X \in [x, y]) = \mathbb{P}(X \leq x) - \mathbb{P}(X < y) = F_X(x) - F_X(y-)$$

—  $A = I_1 \cup \dots \cup I_n$ , avec  $I_1, \dots, I_n$  des intervalles disjoints, alors par l'axiome 3, on a

$$P(X \in A) = P(X \in I_1) + \dots + P(X \in I_n).$$

On se ramène alors au cas d'un intervalle traité précédemment.

3. Si  $F$  est une fonction avec les propriétés du théorème, on définit la loi d'une variable aléatoire  $X$  à partir de  $F$  en utilisant les formules pour  $P(X \in A)$  ci-dessus. Il suffit alors de montrer que cette loi satisfait aux axiomes 1 à 4. On omet la preuve de ce fait qui est laissée en exercice. La fonction  $F$  est alors par définition la fonction de répartition de cette variable aléatoire. □

Posons pour tout réel  $x$  :  $F_X(x-) = \lim_{t \rightarrow x, t < x} F_X(t)$  la limite à gauche de  $F_X$  en  $x$  (qui existe par croissance de la fonction  $F_X$ ). Soulignons le fait important suivant rencontré au cours de la preuve : la fonction de répartition  $F_X$  admet (en tant que fonction croissante) une limite à gauche en tout point, égale à :

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

**Exercice 1.5.** Soit  $F$  une fonction de répartition. On suppose qu'il existe  $x_0 < x_1$  deux réels tels que  $\lim_{x \rightarrow x_0, x > x_0} F(x) = 0$ ,  $\lim_{x \rightarrow x_1, x < x_1} F(x) = 1$ .

1. Calculer  $F(x_0)$  et  $F(x_1)$ .
2. En déduire que  $F$  est caractérisée par sa restriction à l'intervalle  $]x_0, x_1[$ .

## 1.4 Classes de variables aléatoires

Ayant établi la forme générale de la loi d'une v.a. réelle au travers de sa fonction de répartition, il convient de distinguer des classes de v.a. à l'aide de leurs fonction de répartition. Pour ce faire, il est utile d'introduire la notion *d'atome*.

**Définition 1.6.** On dit que  $x \in \mathbb{R}$  est un atome de masse  $m > 0$  de la loi de  $X$  si  $P(X = x) = m$  ou, de façon équivalente,  $F_X(x) - F_X(x-) = m$

La dernière propriété vient de :  $P(X = x) = P(X \leq x) - P(X < x) = F_X(x) - F_X(x-)$ . Ainsi un nombre réel  $x \in \mathbb{R}$  est un atome de masse  $m > 0$  pour la loi de  $X$  si et seulement si  $x$  est un point de discontinuité de la fonction  $F_X$ . Dans ce cas,  $m$  est la hauteur du saut en  $x$ , c'est-à-dire  $m = F_X(x) - F_X(x-)$ . En particulier, une v.a.r.  $X$  n'a pas d'atome si et seulement si sa fonction de répartition est *continue*.

On renvoie le lecteur à l'annexe pour la définition d'un ensemble *dénombrable*, et d'un ensemble *au plus dénombrable*.

**Corollaire 1.7.** *La loi d'une variable aléatoire  $X$  possède un nombre au plus dénombrable, c'est-à-dire fini ou dénombrable, d'atomes.*

*Démonstration.* Il suffit de noter que l'ensemble  $S_X$  des atomes de  $X$  vérifie :

$$S_X := \{x \in \mathbb{R} : P(X = x) > 0\} = \bigcup_{n \geq 1} \{x \in \mathbb{R} : P(X = x) > 1/n\}$$

et l'ensemble  $\{x \in \mathbb{R} : P(X = x) > 1/n\}$  est de cardinal  $< n$ , donc  $S_X$  est une réunion dénombrable d'ensembles finis, donc est en particulier dénombrable.  $\square$

**Variables aléatoires discrètes.** Les variables aléatoires *discrètes* sont celles dont toute la masse se trouve concentrée dans les atomes.

**Définition 1.8.** Soit  $X$  v.a.r. On dit que  $X$  est discrète s'il existe un sous-ensemble  $S \subset \mathbb{R}$  au plus dénombrable tel que  $P(X \in S) = 1$ .

La définition ci-dessus serait vide si  $\mathbb{R}$  était lui-même dénombrable, ce qui n'est bien entendu pas le cas.

**Définition 1.9.** On appelle ensemble des atomes de la loi de  $X$  et on note  $S_X$  l'ensemble

$$S_X := \{x : P(X = x) > 0\}$$

$S_X$  est toujours *au plus* dénombrable et donc peut s'écrire sous la forme :  $S_X = \{s_k, k \in \mathbb{N}\}$  ou  $S_X = \{s_k, 0 \leq k \leq n\}$  pour un certain  $n \in \mathbb{N}$  (de sorte que le premier cas peut être vu comme le cas  $n = +\infty$ ) et des réels  $s_k$  deux à deux distincts. Par abus de notation, on posera simplement  $S_X = \{s_k\}$ .

**Définition 1.10.** Soit  $X$  v.a.r. discrète, de support  $S_X = \{s_k\}_k$ . On appelle *fonction de masse* de  $X$  la fonction :

$$\begin{aligned} S_X &\longrightarrow \mathbb{R} \\ s_k &\longmapsto p_k := \mathbb{P}(X = s_k) \end{aligned}$$

Les éléments  $s_k$  s'appellent alors les *atomes* de la loi, et  $p_k$  est la *masse* de l'atome  $s_k$

On peut maintenant énoncer une CNS sur le caractère discret de  $X$  en fonction des masses des atomes :

**Proposition 1.11.** *Soit  $X$  v.a.r. Alors  $X$  est discrète ssi l'ensemble  $S_X$  des atomes de  $X$  vérifie  $\mathbb{P}(X \in S_X) = 1$  ssi  $\sum_k p_k = 1$ .*

Intuitivement, cela signifie que toute la masse de la loi de  $X$  est concentrée dans ses seuls atomes : on lit aussi que la loi de  $X$  est "purement atomique".

*Démonstration.* Puisque l'ensemble des atomes est au plus dénombrable du Corollaire 1.7,  $\mathbb{P}(X \in S_X) = 1$  implique directement que  $X$  est discrète. Si réciproquement  $X$  est discrète, soit  $S$  ensemble au plus dénombrable qui vérifie  $\mathbb{P}(X \in S) = 1$ . D'abord, on doit avoir  $S \subset S_X$  car pour  $x \in S_X \cap S^c$ , on aurait  $0 < \mathbb{P}(X = x) \leq \mathbb{P}(X \in S^c) = 1 - \mathbb{P}(X \in S) = 0$ ,

absurde. Dès lors on peut écrire  $S$  sous forme de la réunion disjointe :  $S = S_X \cup (S \setminus S_X)$  donc  $\mathbb{P}(X \in S \setminus S_X) = \sum_{s \in S \setminus S_X} \mathbb{P}(X = s) = \sum_{s \in S \setminus S_X} 0 = 0$  donc :

$$1 = \mathbb{P}(X \in S) = \mathbb{P}(X \in S \setminus S_X) + \mathbb{P}(X \in S_X) = \mathbb{P}(X \in S_X)$$

L'équivalence concernant la somme des masses des atomes découle simplement du fait que  $S_X$  est une réunion au plus dénombrable de singletons (les atomes) :

$$\mathbb{P}(X \in S_X) = \mathbb{P}(X \in \bigcup_k \{x_k\}) = \sum_k \mathbb{P}(X = x_k) = \sum_k p_k$$

□

Un corollaire de la démonstration précédente est que  $S_X$  est, dans le cas des v.a.r. discrètes, le plus petit ensemble  $S \subset \mathbb{R}$  tel que  $\mathbb{P}(X \in S) = 1$  : on l'appelle aussi le *support* de la loi de  $X$ .

Il est utile à ce moment de définir la fonction indicatrice de l'ensemble  $A$  :

$$\mathbf{1}_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A. \end{cases}$$

**Proposition 1.12.** *Soit  $X$  une v.a.r. discrète, d'ensemble d'atomes  $S = \{s_k\}_k$  de masses respectives  $p_k = P(X = s_k)$ . La fonction de masse caractérise la loi : pour tout ensemble  $A$  :*

$$P(X \in A) = \sum_k p_k \mathbf{1}_A(s_k).$$

En particulier, la fonction de répartition s'écrit :

$$F_X(x) = \sum_k p_k \mathbf{1}_{\{s_k \leq x\}}.$$

*Démonstration.* Les deux cas  $S$  fini et dénombrable sont quasiment identiques, traitons le cas  $S$  dénombrable  $S = \{s_k, k \in \mathbb{N}\}$ . On décompose comme suit l'événement  $\{X \in A\}$  :

$$P(X \in A) = P(X \in A \cap S) + P(X \in A \cap S^c) = P(X \in \bigcup_{k \in \mathbb{N}} (A \cap \{s_k\})) + 0 = \sum_{k \in \mathbb{N}} P(X \in A \cap \{s_k\})$$

ayant utilisé à la seconde égalité que  $0 \leq P(X \in A \cap S^c) \leq P(X \in S^c) = 1 - P(X \in S) = 0$ , et à la troisième égalité la propriété de sigma-additivité (additivité suffit dans le cas fini). Maintenant,

$$\sum_{k \in \mathbb{N}} P(X \in A \cap \{s_k\}) = \sum_{k \in \mathbb{N}} \mathbf{1}_A(s_k) P(X = s_k) = \sum_{k \in \mathbb{N}} \mathbf{1}_A(s_k) p_k.$$

□

Intuitivement, une v.a. est discrète si sa fonction de répartition n'évolue que par sauts. Il est difficile de formaliser précisément cette notion en général mais on peut formuler la condition suffisante.

**Proposition 1.13.** Soit  $X$  v.a.r. S'il existe une suite  $(x_n)_{n \in \mathbb{Z}} \in [-\infty, +\infty]$  bi-infinie, croissante au sens large, de nombres réels  $\dots \leq x_{-1} \leq x_0 \leq x_1 \leq \dots$ , avec  $\lim_{n \rightarrow -\infty} x_n = -\infty$  et  $\lim_{n \rightarrow \infty} x_n = \infty$  telle que pour tout  $k \in \mathbb{Z}$  pour lequel  $x_k < x_{k+1}$ , la restriction  $(F_X)|_{]x_k, x_{k+1}[}$  est constante, alors  $X$  est discrète, d'ensembles d'atomes  $S_X \subset \{(x_n)_{n \in \mathbb{Z}}\}$ , précisément :  $S_X = \{x_n : F_X(x_n) > F_X(x_{n-1})\}$ .

*Démonstration.* Soit  $k \in \mathbb{Z}$  tel que  $x_k < x_{k+1}$ . On a  $P(X \in ]x_k, x_{k+1}[) = F_X(x_{k+1}-) - F_X(x_k) = F_X(x_{k+1}-) - F_X(x_{k+}) = 0$  puisque  $(F_X)|_{]x_k, x_{k+1}[}$  est constante. Maintenant,

$$\begin{aligned} P(X \in \mathbb{R} \setminus \bigcup_k \{x_k\}) &= P(X \in \bigcup_{k: x_k < x_{k+1}} ]x_k, x_{k+1}[) \\ &= \sum_{k: x_k < x_{k+1}} F_X(x_{k+1}-) - F_X(x_{k+}) = 0, \end{aligned}$$

ayant utilisé à la deuxième égalité que l'union est au plus dénombrable. Ceci implique  $P(X \in \bigcup_k \{x_k\}) = 1$  et donc  $X$  est donc bien discrète par définition.  $\square$

La condition énoncée à la proposition 1.13 n'est pas nécessaire car il existe des ensembles (infinis) dénombrables qu'on ne peut pas écrire sous la forme  $\{x_k\}_k$  pour une suite croissante en  $k$  (par exemple l'ensemble des rationnels).

### Variables aléatoires à densité.

**Définition 1.14.** Soit  $X$  v.a.r. On dit que  $X$  est à densité<sup>3</sup> si la fonction de répartition  $F_X$  s'écrit sous la forme

$$F_X(x) = \int_{-\infty}^x f_X(y) dy, \quad x \in \mathbb{R}, \quad (1.1)$$

pour une fonction intégrable  $f_X : \mathbb{R} \rightarrow \mathbb{R}_+$ . Dans ce cas, on dit que la fonction  $f_X$  est une densité de  $X$  (ou de la loi de  $X$ ), et que  $X$  possède la densité  $f_X$ .

La densité  $f_X$  ainsi définie n'est pas unique (on peut la changer sur un nombre fini ou même dénombrable de points sans changer l'intégrale et donc la fonction de répartition), mais si  $X$  admet une densité continue, alors celle-ci est unique<sup>4</sup>. En ce sens, parler de la densité d'une variable aléatoire est un abus de notation.

Notons que la densité intègre à 1, puisque  $\int_{-\infty}^{+\infty} f_X(y) dy = \lim_{+\infty} F_X = 1$ .

La densité joue un rôle analogue à la fonction de masse dans le cas discret ; l'intuition est que, pour un "élément infinitésimal"  $dx$  autour de  $x$ , on a :

$${}''\mathbb{P}(X \in dx) = f_X(x)dx''$$

et l'exercice suivant donne un sens plus formel à cette égalité :

**Exercice 1.15.** Soit  $X$  une variable aléatoire à densité, dont la densité  $f_X$  est continue au point  $x$ .

3. Dans la littérature, on rencontre aussi le vocabulaire v.a. *absolument continues* pour désigner ces v.a., ou même plus improprement, v.a. *continues*; cette dernière appellation en particulier est trompeuse car elle semble impliquer qu'une fonction de répartition continue est celle d'une variable aléatoire à densité.

4. plus généralement, si  $X$  admet une densité continue par morceaux, alors la densité est unique en dehors des points de discontinuité de la densité

1. Montrer (à l'aide de l'égalité des accroissements finis par exemple) que

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathbb{P}(X \in [x - \varepsilon/2, x + \varepsilon/2])}{\varepsilon} = f_X(x)$$

En particulier, on note le

**Lemme 1.16.** *Soit  $X$  v.a.r. à densité. Alors l'ensemble des atomes  $S_X$  de la loi de  $X$  vérifie :  $S_X = \emptyset$ , c'est-à-dire que loi de  $X$  est sans atome.*

Ce lemme n'admet pas de réciproque. Nous reviendrons plus tard sur ce point subtil : il existe des variables aléatoires sans atomes qui ne sont pas à densité; du point de vue des fonctions de répartition, cela revient à dire qu'il existe des fonctions continues croissantes qui ne s'écrivent pas comme primitive de leur dérivée (partout où celle-ci est définie).

*Démonstration.* La fonction  $F_X$  est continue en tant que primitive, voir (1.1), en particulier, pour tout  $x \in \mathbb{R}$  :

$$P(X = x) = F_X(x) - F_X(x-) = 0$$

c'est-à dire que  $x$  n'est pas un atome. □

Ainsi pour une variable aléatoire à densité  $\mathbb{P}(X \in S_X) = 0$ , à comparer avec le fait que  $\mathbb{P}(X \in S_X) = 1$  pour une variable aléatoire discrète : on voit donc que ces deux types de variables aléatoires correspondent aux deux valeurs extrêmes pour  $\mathbb{P}(X \in S_X)$ . En particulier, si  $X$  est une v.a. à densité, pour tout ensemble au plus dénombrable  $T \subset \mathbb{R}$ , on a  $\mathbb{P}(X \in T) = 0$ , ce qui montre la différence de nature entre v.a.r. discrète et v.a.r. à densité. La densité caractérise la fonction de répartition de  $X$  et donc la loi de  $X$  (du théorème 1.4). On peut en outre facilement exprimer cette loi à l'aide de la densité. Voici donc l'analogue pour les v.a.r. à densité de la Proposition 1.12.

**Proposition 1.17.** *Si  $X$  est une v.a. à densité de densité  $f_X$ , alors on a pour tout ensemble  $A$ <sup>5</sup> :*

$$P(X \in A) = \int_{-\infty}^{\infty} f_X(x) \mathbb{1}_A(x) dx.$$

*Démonstration.* On se donne  $a, b \in [-\infty, +\infty]$ . On commence par vérifier la résultat pour  $A = ]a, b]$  :  $P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx = \int_{-\infty}^{+\infty} f_X(x) \mathbb{1}_A(x) dx$ . Les cas  $A = [a, b]$ ,  $A = ]a, b[$ ,  $A = ]a, b[$ ,  $A = ]a, b]$  ne posent pas de difficulté, la fonction  $F_X$  étant continue,  $F_X(x) = F_X(x-)$  pour tout réel  $x$  et donc, on a :  $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b) = P(a < X \leq b) = \int_a^b f_X(x) dx = \int_{-\infty}^{+\infty} f_X(x) \mathbb{1}_A(x) dx$ . Noter que ces relations valent encore pour  $a = -\infty$  et  $b = +\infty$ . Par linéarité de l'intégrale, pour  $A$  réunion d'un nombre fini d'intervalles disjoints de  $\mathbb{R}$ , on a encore :  $P(X \in A) = \int_A f_X(x) dx = \int_{-\infty}^{+\infty} \mathbb{1}_I(x) f_X(x) dx$  □

Le résultat suivant est utile pour identifier la densité à partir de la fonction de répartition.

**Proposition 1.18.** *Soit  $X$  v.a.r. dont la fonction de répartition  $F_X$  est  $\mathcal{C}^0(\mathbb{R})$  (continue), et  $\mathcal{C}^1$  par morceaux (c'est-à-dire qu'il existe un nombre fini de points  $x_1 < x_2 < \dots < x_n$  t.q.  $F_X$  est  $\mathcal{C}^1$  sur  $] - \infty, x_1[$ ,  $]x_1, x_2[$ ,  $\dots$ ,  $]x_{n-1}, x_n[$ ,  $]x_n, +\infty[$ ) alors  $X$  est une v.a. à densité et une densité de  $X$  est  $F'_X$ .*

---

5. Le résultat s'étend aux ensembles mesurables au sens de Lebesgue; en revanche à la différence du résultat pour les variables aléatoires discrètes on ne peut cette fois-ci considérer des ensemble  $A$  quelconques; en pratique cependant, il est très difficile de construire des ensembles non mesurables

*Démonstration.* Les propriétés de la fonction  $F_X$  garantissent qu'on peut écrire pour tout  $x, y \in [-\infty, \infty]$ ,  $F_X(y) - F_X(x) = \int_x^y F'_X(z) dz$  et donc, avec  $x = -\infty$ , en particulier, on a pour tout  $y \in \mathbb{R}$ ,  $F_X(y) = \int_{-\infty}^y F'_X(z) dz$ , donc  $F'_X$  est une densité de  $X$ .  $\square$

On utilise souvent cette proposition pour vérifier qu'une variable aléatoire est à densité. Cependant, une alternative à cette proposition (et à la vérification de ses hypothèses ...) est d'utiliser tout simplement la définition 1.14 et de mettre  $F$  directement sous la forme (1.1) : du point de vue de la rédaction, c'est souvent plus simple, même si en pratique, pour trouver l'intégrande, on dérive bien la fonction de répartition, comme dans la proposition précédente...). Voyez plutôt :

**Exercice 1.19.** On pose  $F(x) = x^2 \mathbb{1}_{[0,1/2[}(x) + (1 - 3(1 - x)^2) \mathbb{1}_{[1/2,1[}(x) + \mathbb{1}_{[1,+\infty[}(x)$ .

1. Appliquer la proposition précédente pour montrer que  $F$  est la fonction de répartition d'une variable aléatoire la densité dont on précisera la densité.
2. Vérifier que  $F(x) = \int_{-\infty}^x [2t \mathbb{1}_{[0,1/2[}(t) + 6(1-t) \mathbb{1}_{[1/2,1[}(t)] dt$ , et conclure comme précédemment.

Toutes les v.a.r. à densité que nous rencontrerons dans ce cours auront une fonction de répartition du type précédent ( $\mathcal{C}^0(\mathbb{R})$  et  $\mathcal{C}^1$  par morceaux). Cependant, la forme générale autorise des fonctions de répartition de v.a.r. à densité qui ne sont pas de ce type. L'ensemble des fonctions qui s'écrivent comme primitive d'une fonction intégrable ne se réduit pas à l'ensemble des fonctions  $\mathcal{C}^0(\mathbb{R})$  et  $\mathcal{C}^1$  par morceaux : il s'agit de l'ensemble des fonctions absolument continues, que le lecteur curieux est invité à découvrir par ses propres moyens (diverses caractérisations de cet ensemble existe, googlez pour les trouver).

### Les autres variables aléatoires.

- Quasiment toutes les variables aléatoires que nous allons rencontrer dans ce cours font partie d'une des deux classes ci-dessus. Parmi les autres, on peut mentionner celles qui sont une combinaison des deux types de variables aléatoires, voir l'exercice ci-dessous.
- Il est aussi légitime de se demander s'il existe des fonctions de répartition qui sont continues mais ne s'écrivent pas comme la primitive d'une autre fonction. La réponse est oui, mais la construction d'une telle fonction dépasse le contenu de ce cours. Les curieux sont invités à googler « l'escalier de Cantor, » également appelé « l'escalier du diable. . . ».

Voici un exemple de combinaison des deux types de variables aléatoires tel que mentionné ci-dessus :

**Exercice 1.20.** Soit  $X$  v.a.r. de fonction de répartition  $F_X(t) = (1+t)/2 \cdot \mathbb{1}_{[0,1[}(t) + \mathbb{1}_{[1,+\infty[}(t)$ .

1. Vérifier qu'il s'agit d'une fonction de répartition.
2. Vérifier que  $\mathbb{P}(X \in S_X) = 1/2$ .
3.  $X$  est-elle discrète, est-elle à densité ?
4. Soit  $y \in [0, 1]$ . Construire une variable aléatoire  $Y$  telle que  $\mathbb{P}(Y \in S_Y) = y$  (on pourra considérer une généralisation simple de  $F_X$ ).

## 1.5 Quelques variables aléatoires discrètes importantes

NB : Pour des détails au sujet de cette section, se référer aux chapitres 4 et 5 du livre de Ross.

### 1.5.1 Loi de Dirac

La loi de Dirac est la loi la plus simple qui existe : Si  $x \in \mathbb{R}$ , on dit que  $X$  suit la loi de Dirac *en*  $x$  (noté parfois symboliquement  $X \sim \delta_x$ ), si

$$P(X = x) = 1.$$

De la proposition 1.12, on peut aussi écrire pour toute partie  $A \subset \mathbb{R}$  :

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

En particulier ( $A = ]-\infty, y]$ ), la fonction de répartition est donnée par

$$F_X(y) = \mathbb{1}_{]-\infty, y]}(x) = \begin{cases} 1, & \text{si } y \geq x \\ 0, & \text{sinon.} \end{cases}$$

On note parfois symboliquement :  $X \sim \delta_x$ , ce qu'on lit «  $X$  suit la loi de Dirac en  $x$ . »

Si  $X$  suit la loi de Dirac en  $x \in \mathbb{R}$ , on dit aussi que  $X$  est presque sûrement (p.s.) égale à  $x$ .

Noter que la loi de Dirac en  $x$  est *la* loi discrète d'ensemble d'atomes  $S_X = \{x\}$ . La valeur de l'atome en  $x$  vaut en effet nécessairement 1 dans ce cas.

### 1.5.2 Loi uniforme

La loi uniforme est la loi la plus simple d'une variable aléatoire prenant un nombre fini de valeurs. Soit  $E$  une partie finie de  $\mathbb{R}$ , par exemple  $E = \{1, \dots, n\}$  pour  $n \in \mathbb{N}$ . Une v.a.  $X$  suit la *loi uniforme* sur  $S$  (symboliquement :  $X \sim \text{Unif}(S)$ ) si pour tout  $x \in A$

$$P(X = x) = \frac{1}{\text{Card}(S)}.$$

En conséquence on peut écrire comme suit la fonction de répartition :

$$P(X \in A) = \frac{\text{Card}(A \cap S)}{\text{Card}(S)}, \quad F_X(x) = \frac{\text{Card}(]-\infty, x] \cap S)}{\text{Card}(S)}.$$

Noter que  $\text{Unif}(S)$  est *la* loi discrète d'ensembles d'atomes  $S$  dont la fonction de masse est constante sur  $S$ . La loi uniforme intervient dans de nombreuses applications, par exemple dans les jeux de hasard (lancer de dés, tirage de cartes, ...).

### 1.5.3 Loi de Bernoulli

On dit que  $X$  suit la loi de Bernoulli de paramètre  $p \in [0, 1]$  (symboliquement :  $X \sim \text{Ber}(p)$ ), si

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

La fonction de répartition s'écrit :

$$F_X(x) = \begin{cases} 0, & \text{si } x < 0 \\ 1 - p, & \text{si } x \in [0, 1[ \\ 1, & \text{si } x \geq 1. \end{cases}$$

Le nom de la loi de Bernoulli vient de *l'expérience de Bernoulli* : c'est une expérience aléatoire ayant deux issues possibles : *succès* ou *échec*. On pose alors  $X = 1$  en cas de succès et  $X = 0$  en cas d'échec, et si  $p$  est la probabilité de succès, alors  $X$  suit la loi de Bernoulli de paramètre  $p$ . Notons que dans le cas  $p = 1/2$ ,  $\text{Ber}(p)$  coïncide avec  $\text{Unif}(\{0, 1\})$ .

### 1.5.4 Loi binomiale

La loi binomiale est une généralisation importante de la loi de Bernoulli. On dit que  $X$  suit la loi binomiale de paramètres  $n \in \mathbb{N}$  et  $p \in [0, 1]$  (symboliquement :  $X \sim \text{Bin}(n, p)$ ), si

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

où on rappelle la définition du coefficient binomial :

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

La fonction de répartition n'admet pas d'expression plus simple que son expression générale à partir de la fonction de masse. Noter que  $\text{Bin}(1, p)$  coïncide avec  $\text{Ber}(p)$ .

Interprétation : Il s'agit du nombre de succès lors de  $n$  répétitions indépendantes d'une expérience de Bernoulli de probabilité de succès  $p$ . Il nous faudra encore du temps pour justifier cette interprétation : cela ne viendra que bien plus tard, à l'exercice 3.33, une fois la notion d'indépendance introduite.

Le fait que cette fonction de masse définit une probabilité correspond au théorème binomial de Newton : développer  $(a+b)^n$  revient à choisir les indices où l'on choisit  $a$ , soit une partie  $I$  de  $\{1, \dots, n\}$ , puis on note qu'une telle partie s'écrit  $I = \{1 \leq i_1 < \dots < i_k \leq n\}$  pour un certain  $k \in \{0, \dots, n\}$ ,

$$(a+b)^n = \sum_{I \subset \{1, \dots, n\}} a^{|I|} b^{n-|I|} = \sum_{k=0}^n \sum_{1 \leq i_1 < \dots < i_k \leq n} a^k b^{n-k} = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k},$$

À la dernière ligne on a noté que

$$\text{Card}\{1 \leq i_1 < \dots < i_k \leq n\} = \frac{n \times (n-1) \times \dots \times (n-k+1)}{k!} = \binom{n}{k},$$

(choisir une suite de  $k$  éléments distincts de  $\{1, \dots, n\}$  sans remplacement, soit  $n \times (n-1) \times \dots \times n-k+1$  choix, puis réordonner par ordre croissant :  $k!$  suites donneront le même réordonnement croissant). Il suffit maintenant de poser  $a = p$  et  $b = 1-p$ .

**Exercice 1.21.** Soit  $X = (X_1, \dots, X_n)$  une variable uniforme sur l'ensemble fini  $\{0, 1\}^n$  des mots binaires de longueur  $n$ . Pour  $x \in \{0, 1\}^n$ , on pose  $\varphi(x) = \sum_{i=1}^n x_i$ . Montrer que  $\varphi(X)$  suit la loi  $\text{Bin}(n, 1/2)$ .

### 1.5.5 Loi de Poisson

On dit que  $X$  suit la loi de Poisson de paramètre  $\lambda \geq 0$  (symboliquement,  $X \sim \text{Po}(\lambda)$ ), si pour tout  $k \in \mathbb{N}$ ,

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



Sa fonction de répartition n'admet pas d'expression plus simple que son expression générale à partir de la fonction de masse. Le fait que la fonction de masse définisse une probabilité correspond au DSE de la fonction  $\exp$  : pour  $z$  complexe,  $e^z = \sum_{k \geq 0} \frac{z^k}{k!}$ , obtenu en calculant les dérivées successives de  $\exp$  en 0.

La loi de Poisson de paramètre  $\lambda \geq 0$  peut-être vue comme une approximation de la loi binomiale de paramètres  $n$  et  $p = \lambda/n$ , dès lors que  $p$  est petit. En effet, on peut vérifier (exercice) que pour  $\lambda \geq 0$  et  $k \in \mathbb{N}$  fixés,

$$\binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}, \quad n \rightarrow \infty.$$

Plus précisément, on a le théorème suivant :

**Theorème 1.22** (Approximation Poisson de la loi Binomiale; Prokhorov, Stein,...). *Pour tout  $n \in \mathbb{N}$  et  $p \in [0, 1]$ ,*

$$\sum_{k \in \mathbb{N}} \left| \binom{n}{k} p^k (1-p)^{n-k} - e^{-np} \frac{(np)^k}{k!} \right| \leq \min\{2np^2, 3p\}.$$

En particulier, si  $X_n \sim \text{Bin}(n, p)$  et  $X \sim \text{Po}(np)$ , on a, pour tout sous-ensemble  $J \subset \mathbb{N}$ ,

$$|P(X_n \in J) - P(X \in J)| \leq \min\{2np^2, 3p\}.$$

Notons que la conséquence énoncée découle comme suit de l'inégalité triangulaire et du premier énoncé :

$$\begin{aligned} |P(X_n \in J) - P(X \in J)| &= \left| \sum_{k \in J} P(X_n = k) - \sum_{j \in J} P(X = k) \right| \\ &\leq \sum_{k \in J} |P(X_n = k) - P(X = k)| \\ &\leq \sum_{k \in \mathbb{N}} |P(X_n = k) - P(X = k)| \\ &= \sum_{k \in \mathbb{N}} \left| \binom{n}{k} p^k (1-p)^{n-k} - e^{-np} \frac{(np)^k}{k!} \right| \end{aligned}$$

Ce résultat montre donc que la loi de Poisson de paramètre  $\lambda = np$  est une bonne approximation de la loi binomiale de paramètres  $n$  et  $p$  dès lors que  $p$  est petit (et cela indépendamment de  $n$ , un résultat souvent méconnu).

### 1.5.6 Loi géométrique

On dit que  $X$  suit la loi géométrique de paramètre  $p \in ]0, 1]$  (symboliquement :  $X \sim \text{Geom}(p)$ ), si

$$P(X = k) = p \times (1-p)^{k-1}, \quad k \in \mathbb{N}^* = \{1, 2, \dots\}.$$

Sa fonction de répartition admet une expression simple (exercice) ; on rappelle que pour  $x \in \mathbb{R}$ , la *partie entière*  $[x]$  est le plus grand entier plus petit ou égal à  $x$ . On obtient alors :

$$F_X(x) = \begin{cases} 1 - (1-p)^{[x]}, & \text{si } x \geq 0 \\ 0, & \text{si } x < 0. \end{cases}$$

En particulier, on peut retenir la formule suivante :

$$P(X > k) = (1 - p)^k, \quad \text{pour tout } k \in \mathbb{N}.$$

Une v.a.  $X$  de loi géométrique de paramètre  $p$  représente le nombre de d'expériences de Bernoulli indépendantes de probabilité de succès  $p$  jusqu'à la première expérience fructueuse. En effet, la probabilité d'avoir eu des échecs pendant les  $k - 1$  premiers essais et un succès au  $k$ -ième essai est exactement égal à  $(1 - p)^{k-1} \times p$ , si les expériences sont indépendantes.

Le fait que cette fonction de masse définit une probabilité correspond au DSE de la fonction  $1/(1 - z)$ . Rappelons le raisonnement : pour tout complexe  $z$ ,

$$\sum_{k=0}^n z^k \cdot (1 - z) = \sum_{k=0}^n z^k - \sum_{k=0}^n z^{k+1} = 1 - z^{n+1}$$

donc pour  $z \neq 1$ ,  $\sum_{k=0}^n z^k = \frac{1-z^{n+1}}{1-z}$  puis par passage à la limite pour  $|z| < 1$ ,

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

Ceci entraîne, pour  $p \in ]0, 1]$ ,  $\sum_{k=0}^{\infty} (1-p)^{k-1} = \frac{1}{1-(1-p)} = \frac{1}{p}$ , c'est-à-dire  $\sum_{k=0}^{\infty} p(1-p)^{k-1} = 1$ .

**Proposition 1.23.** *Soit  $X$  v.a.r. Alors  $X$  suit la loi géométrique ssi  $X$  vérifie la propriété suivante*

$$\mathbb{P}(X \in \mathbb{N}^*) = 1 \text{ et } P(X - k > \ell | X > k) = P(X > \ell), \quad \text{pour tout } k, \ell \in \mathbb{N},$$

dite propriété de perte de mémoire.

Autrement dit, pour tout  $k \in \mathbb{N}$ , la loi de  $X - k$  conditionnellement à l'événement  $\{X > k\}$  est la même que celle de  $X$  : si  $X$  modélise un temps de vie, la loi du temps de vie restant ne dépend pas de l'âge ; il n'y a pas de vieillissement.

*Démonstration.* Si la v.a.r satisfait la propriété de perte de mémoire, alors on montre par récurrence sur l'entier  $\ell \in \mathbb{N}^*$  (en prenant  $k = 1$  dans l'énoncé précédent) :

$$P(X > \ell) = P(X > 1)^\ell,$$

ce qui entraîne

$$\begin{aligned} P(X = \ell) &= P(X > \ell - 1) - P(X > \ell) && \text{car } P(X \in \mathbb{N}^*) = 1 \\ &= P(X > 1)^{\ell-1} - P(X > 1)^\ell \\ &= P(X > 1)^{\ell-1}(1 - P(X > 1)) \\ &= (1 - P(X = 1))^{\ell-1}P(X = 1), \end{aligned}$$

c'est-à-dire que  $X$  suit une loi géométrique (dire qu'elle est de paramètre  $\mathbb{P}(X = 1)$  n'est pas informatif car ceci vaut toujours.) La réciproque consiste en un simple calcul : Si  $X$  suit une loi  $\text{Geom}(p)$ , alors  $P(X > \ell) = P(X > 1)^\ell$  et ceci entraîne

$$P(X - k > \ell | X > k) = \frac{P(\{X - k > \ell\} \cap \{X > k\})}{P(X > k)} = \frac{P(\{X > k + \ell\})}{P(X > k)} = \frac{P(X > 1)^{k+\ell}}{P(X > 1)^k} = P(X > 1)^\ell$$

□

*Remarque 1.24.* Il est également fréquent de rencontrer la loi géométrique sur  $\mathbb{N}$ ; il s'agit simplement de la loi géométrique décalée de 1 : précisément  $Y \sim \text{Geom}_{\mathbb{N}}(p)$  si  $Y+1 \sim \text{Geom}(p)$  si

$$P(Y = k) = p \times (1 - p)^k, \quad k \in \mathbb{N} = \{0, 1, 2, \dots\}.$$

On fera bien attention à ajouter l'indice  $\mathbb{N}$  si on manipule cette loi; les deux lois se rencontrent aussi fréquemment l'une que l'autre en pratique.

### 1.5.7 Autres lois discrètes

Voici quelques autres lois discrètes que nous mentionnons brièvement; à la différence des précédentes, elles ne sont pas à connaître par coeur.

**Loi binomiale négative** On dit que  $X$  suit la loi binomiale négative de paramètres  $r \in \mathbb{N}$  et  $p \in ]0, 1]$  (symboliquement :  $X \sim \text{NegBin}(r, p)$ ), si

$$P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n \in \{r, r+1, \dots\}$$

Interprétation : il s'agit du nombre de répétitions indépendantes dans une suite d'expériences de Bernoulli jusqu'à cumuler  $r$  succès. Noter que la dernière expérience (incluse dans le décompte) est nécessairement un succès; ainsi les  $r-1$  autres succès doivent être répartis parmi les  $n-1$  expériences précédentes, ce qui donne le facteur combinatoire  $\binom{n-1}{r-1}$  tandis que chaque configuration a pour probabilité  $p^r (1-p)^{n-r}$ . En particulier,  $\text{Geom}(p) = \text{BN}(1, p)$ . Le fait que cette fonction de masse définit une probabilité est un calcul intéressant :

$$(1-q)^{-r} = ((1-q)^{-1})^r = \left(\sum_{j \geq 0} q^j\right)^r = \left(\sum_{j_1 \geq 0} q^{j_1}\right) \times \dots \times \left(\sum_{j_r \geq 0} q^{j_r}\right) = \sum_{j_1, \dots, j_r \geq 0} q^{j_1 + \dots + j_r}$$

Maintenant, pour trouver le coefficient de  $q^k$  dans cette expression, il faut se demander,  $k \in \mathbb{N}$  étant fixé combien de  $r$ -uplets d'entier  $j_1, \dots, j_r$  réalisent la contrainte  $1_{\{j_1 + \dots + j_r = k\}}$ . Cet ensemble est en bijection avec les chemins Nord-Est du plan de  $(0, 0)$  à  $(k, r-1)$ , dénombré par le coefficient binomial  $\binom{r-1+k}{r-1}$ , aussi :

$$\begin{aligned} (1-q)^{-r} &= \sum_{k \geq 0} q^k \left( \sum_{j_1, \dots, j_r \geq 0} 1_{\{j_1 + \dots + j_r = k\}} \right) \\ &= \sum_{k \geq 0} \binom{r-1+k}{r-1} q^k \end{aligned}$$

Ainsi, en réindiquant et en posant  $q = 1 - p$ , on tombe bien sur l'identité cherchée :

$$1 = \sum_{k \geq 0} \binom{r-1+k}{r-1} q^k (1-q)^r = \sum_{n \geq r} \binom{n-1}{r-1} (1-p)^{n-r} p^r$$

*Remarque 1.25.* Il existe une autre convention pour la loi la loi Binomiale négative qui consiste à compter le nombre d'échecs avant de cumuler  $r$  succès : c'est tout simplement la loi de  $X-r$  :

$$P(X - r = j) = \binom{r+j-1}{r-1} p^r (1-p)^j, \quad j \in \mathbb{N} = \{0, 1, \dots\}$$

**Loi hypergéométrique** On dit que  $X$  suit la loi hypergéométrique de paramètres  $n, N, m \in \mathbb{N}$  avec  $n \leq N$  et  $m \leq N$  (symboliquement :  $X \sim \text{Hypergeo}(n, N, m)$ ), si

$$P(X = k) = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}}, \quad k = 0, \dots, \min\{m, n\}.$$

Interprétation : on tire sans remise un échantillon de  $n$  boules d'une urne en contenant  $N$ , dont  $m$  blanches et  $N - m$  noires. Si  $X$  désigne le nombre de boules blanches tirées, alors  $X \sim \text{Hypergeo}(n, N, m)$ .

**Exercice 1.26.** Soit  $X_N$  de loi  $\text{Hypergeo}(n, N, m)$ . On suppose que  $m = m_N$  est une suite telle que  $m_N/N \rightarrow p$  quand  $N \rightarrow \infty$ . Montrer que

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

**Loi zéta (ou de Zipf)** On dit que  $X$  suit une loi zéta (ou de Zipf) de paramètre  $\alpha$  si

$$P(X = k) = C/k^{\alpha+1}, \quad k \in \mathbb{N}^*,$$

pour  $C = 1/\sum_{k=1}^{\infty} k^{-(\alpha+1)} = 1/\zeta(\alpha+1)$ , avec  $\zeta$  la fonction zéta de Riemann.

## 1.6 Quelques variables aléatoires à densité importantes

### 1.6.1 Loi uniforme

Soit  $I \subset \mathbb{R}$  un intervalle borné et notons  $|I|$  sa longueur. On dit que  $X$  suit la loi uniforme sur  $I$  (symboliquement,  $X \sim \text{Unif}(I)$ ), si  $X$  possède la densité

$$f_X(x) = \frac{1}{|I|} \mathbb{1}_I(x).$$

Si on note  $a = \inf I$  et  $b = \sup I$ , de sorte que  $|I| = b - a$ , la fonction de répartition s'écrit

$$F_X(t) = \begin{cases} 0, & \text{si } t < a \\ \frac{t-a}{b-a}, & \text{si } t \in [a, b] \\ 1, & \text{si } t > b. \end{cases}$$

On pose donc alors la notation  $\text{Unif}(a, b)$  pour cette loi (ce qui a l'avantage de ne pas préciser si l'intervalle est ouvert ou fermé, ou ouvert en un point et fermé en l'autre). Cette loi est l'analogie continu de la loi uniforme discrète. En effet, alors que la fonction de masse de cette dernière est constante sur l'ensemble des valeurs que prend la variable aléatoire, c'est cette fois-ci la densité qui est constante sur  $I$ . Insistons sur le fait que l'intervalle est nécessairement borné :  $a$  ne peut prendre la valeur  $-\infty$ ,  $b$  ne peut prendre la valeur  $+\infty$ . En particulier, il n'existe pas de mesure uniforme sur  $\mathbb{R}$  (tout comme il n'existe pas de mesure uniforme sur  $\mathbb{N}$  dans le cas discret).

**Exercice 1.27.** Soit  $X \sim \text{Unif}(0, 1)$ , et  $a, b \in \mathbb{R}$ . Quelle est la loi de  $a \cdot X + b$  ?

### 1.6.2 Loi normale (ou gaussienne)

On dit que  $X$  suit la loi normale (ou gaussienne) de moyenne  $\mu \in \mathbb{R}$  et variance  $\sigma^2 > 0$  (symboliquement,  $X \sim \mathcal{N}(\mu, \sigma^2)$ ), si  $X$  possède la densité

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Pour calculer l'intégrale  $\int_{-\infty}^{\infty} f_X(x) dx$ , on se ramène par changement de variables à l'intégrale de Gauss  $\int_{-\infty}^{\infty} e^{-t^2} dt$  et on montre ainsi que  $f_X$  est bien une densité de probabilité, c'est-à-dire son intégrale vaut 1.

Si  $\mu = 0$ , on dit que  $X$  est *centrée*, et si  $\sigma^2 = 1$  on dit que aussi que  $X$  est *réduite*. La fonction de répartition de la loi normale centrée réduite  $\mathcal{N}(0, 1)$  est notée  $\Phi$  :

$$\Phi(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

Cette intégrale n'admet pas d'expression simple, mais elle est tabulée et disponible dans tout logiciel de calcul<sup>6</sup> : des tables de valeurs de  $\Phi$  sont facilement accessibles. Notons aussi l'encadrement suivant de la queue de distribution, de très bonne qualité pour  $x$  grand :

**Exercice 1.28.** *Montrer l'encadrement suivant (on pourra considérer des intégration par parties successives de  $1 - \Phi(t)$ ) :*

$$\left(\frac{1}{t} - \frac{1}{t^3}\right) e^{-\frac{t^2}{2}} \leq \mathbb{P}(X \geq t) \leq \frac{1}{t} e^{-\frac{t^2}{2}}, \quad x > 0$$

La forme de la densité  $f_X$  est la fameuse « courbe en forme de cloche » de Gauss. Elle est symétrique autour de  $\mu$ , croissante à gauche de  $\mu$  et décroissante à droite. Plus  $\sigma^2$  est petit, plus la densité est resserrée autour de  $\mu$ . On peut facilement relier les lois normales pour les différentes valeurs des paramètres par la formule suivante :

**Lemme 1.29.** *Si  $N \sim \mathcal{N}(0, 1)$  et  $\mu, \sigma \in \mathbb{R}$ ,  $\sigma \neq 0$ , alors  $\mu + \sigma N \sim \mathcal{N}(\mu, \sigma^2)$ .*

*Démonstration.* On traite d'abord le cas  $\sigma > 0$ . Le lemme se vérifie directement à l'aide des définitions comme suit :

$$\begin{aligned} P(\mu + \sigma N \leq t) &= P\left(N \leq \frac{t - \mu}{\sigma}\right) \\ &= \int_{-\infty}^{\frac{t - \mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\ &= \int_{-\infty}^t \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z - \mu)^2}{2\sigma^2}} dz \end{aligned}$$

où l'on a posé le changement de variable affine  $y = \frac{z - \mu}{\sigma}$  à la dernière ligne. Pour le cas  $\sigma < 0$ , ou bien on utilise la même méthode (mais il faut faire être attentif dans la manipulation des inégalités), ou bien on note que,  $N$  et  $-N$  ayant même loi,  $\mu + \sigma N$  et  $\mu - \sigma N$  ont même loi.  $\square$

6. plus précisément, c'est la fonction erreur, aussi notée "erf",  $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy$  qu'on trouve souvent tabulée

Pour cette raison, on *définit* la loi  $\mathcal{N}(\mu, 0)$  comme étant la loi de Dirac en  $\mu$ , ce n'est donc plus une loi à densité mais une loi discrète.

La loi normale est omniprésente en sciences : elle donne en effet une bonne approximation de nombre de lois qui décrivent des sommes de variables aléatoires. Par exemple, le théorème de de Moivre-Laplace dit que la loi binomiale est bien approchée quand  $n$  est grand par la loi gaussienne :

**Théorème 1.30** (de Moivre, Laplace). *Soit  $p \in ]0, 1[$ . Pour tout  $n \in \mathbb{N}$ , soit  $X_n$  une variable aléatoire de loi binomiale de paramètres  $n$  et  $p$ . Alors, pour tout  $-\infty \leq a < b \leq +\infty$ ,*

$$P\left(np + a \times \sqrt{p(1-p)n} \leq X_n \leq np + b \times \sqrt{p(1-p)n}\right) \rightarrow \Phi(b) - \Phi(a), \quad n \rightarrow \infty$$

*Heuristiquement, ce théorème dit que si  $X_n \sim \text{Bin}(n, p)$ , alors quand  $n$  est grand et  $p$  fixé,*

$$X_n \approx np + \sqrt{p(1-p)n} \times N, \quad \text{avec } N \sim \mathcal{N}(0, 1)$$

*c'est-à-dire qu'on approche la loi  $\text{Bin}(n, p)$  par  $\mathcal{N}(np, np(1-p))$ .*

de Moivre fut le premier à établir ce théorème en 1733 dans le cas particulier  $p = \frac{1}{2}$ ; et Laplace a pu le généraliser en 1812 pour toute valeur de  $p \in ]0, 1[$ . Il s'agit d'un cas particulier du théorème central limite, qui sera montré en fin de cours (nous ne ferons donc pas de démonstration dès à présent). On verra aussi quel est le sens des paramètres  $np$  et  $np(1-p)$  et en quoi ils sont naturels. En pratique, on juge que l'approximation est bonne dès lors que  $np(1-p) \geq 10$ , critère que l'on vérifiera avant toute application lors d'un exercice.

**Correction de continuité** Lorsqu'on cherche à approcher une loi discrète par une loi à densité, l'évaluation numérique des probabilités peut poser problème. Précisément, en pratique, on utilise souvent le théorème dans le contexte suivant :  $a < b \in \mathbb{Z}$  étant deux entiers donnés, on cherche à estimer la probabilité de l'événement :

$$\{a \leq X_n \leq b\}.$$

Une première possibilité consiste à écrire :

$$\begin{aligned} \mathbb{P}(a \leq X_n \leq b) &= \mathbb{P}\left(\frac{a - np}{\sqrt{np(1-p)}} \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq \frac{b - np}{\sqrt{np(1-p)}}\right) \\ &\approx \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

où l'approximation est justifiée par le théorème de de Moivre-Laplace. Maintenant, observons que  $a$  et  $b$  étant des entiers, on aurait aussi bien pu noter l'égalité d'événements suivante :  $\{a - 1 < X_n < b + 1\} = \{a \leq X_n \leq b\}$  et dans ce cas, on est tenté d'écrire :

$$\begin{aligned} \mathbb{P}(a - 1 < X_n < b + 1) &= \mathbb{P}\left(\frac{a - 1 - np}{\sqrt{np(1-p)}} \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq \frac{b + 1 - np}{\sqrt{np(1-p)}}\right) \\ &\approx \Phi\left(\frac{b + 1 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - 1 - np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

toujours d'après le théorème. Quantitativement, la différence entre ces deux approximations peut être significative; la *correction dite de continuité* consiste à couper la poire en deux en retenant l'intervalle d'extrémités  $a - 1/2$  et  $b + 1/2$ , et donc à retenir la valeur

$$\Phi\left(\frac{b + 1/2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - 1/2 - np}{\sqrt{np(1-p)}}\right)$$

comme approximation. On observe en pratique que cette correction peut donner une approximation significativement meilleure.

Une autre façon de justifier cette approximation est la suivante : si on dispose d'une variable discrète  $X_N$  et d'une variable à densité  $X$  qui approche celle-ci, il est naturel de retenir comme approximation :

$$\mathbb{P}(X_n = k) \approx \mathbb{P}(k - 1/2 < X \leq k + 1/2).$$

On somme ensuite sur les valeurs de  $k \in \{a, \dots, b\}$ . Ici,  $X_n \sim \text{Bin}(n, p)$  et  $X \sim \mathcal{N}_1(np, np(1-p))$  et si l'on pose  $N \sim \mathcal{N}_1(0, 1)$ , on peut écrire :

$$\begin{aligned} \mathbb{P}(a \leq X_n \leq b) &= \sum_{j=a}^b \mathbb{P}(X_n = j) \\ &\approx \sum_{j=a}^b \mathbb{P}(j - 1/2 < X \leq j + 1/2) \\ &= \mathbb{P}(a - 1/2 < X \leq b + 1/2) \\ &= \mathbb{P}(a - 1/2 < np + \sqrt{np(1-p)}N \leq b + 1/2) \\ &= \mathbb{P}\left(\frac{a - 1/2 - np}{\sqrt{np(1-p)}} < N \leq \frac{b + 1/2 - np}{\sqrt{np(1-p)}}\right) \\ &= \Phi\left(\frac{b + 1/2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - 1/2 - np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

### 1.6.3 Loi exponentielle

On dit que  $X$  suit la loi exponentielle de paramètre  $\lambda > 0$  (symboliquement,  $X \sim \text{Exp}(\lambda)$ ), si  $X$  possède la densité

$$f_X(x) = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}_+}(x)$$

Sa fonction de répartition admet une expression simple (exercice) :

$$F_X(t) = (1 - e^{-\lambda t}) \mathbb{1}_{\mathbb{R}_+}(t)$$

En particulier, on peut retenir la formule suivante :

$$P(X > t) = e^{-\lambda t}, \quad t \geq 0.$$

La loi exponentielle est l'analogie continu de la loi géométrique. Comme elle, elle satisfait et se trouve caractérisée par la propriété d'absence de mémoire.

**Proposition 1.31.** *Soit  $X$  v.a.r. Alors  $X$  suit la loi exponentielle ssi  $X$  vérifie la propriété suivante*

$$P(X \in \mathbb{R}^+) = 1 \text{ et } P(X > t + s | X > t) = P(X > s), \quad \text{pour tout } t, s \in \mathbb{R}^+$$

dite propriété d'absence de mémoire.

Le lecteur est invité à comparer cette propriété à son analogue discret pour la loi géométrique.

*Démonstration.* Si la v.a.r satisfait la propriété d'absence de mémoire, alors on note tout d'abord que la propriété d'absence de mémoire s'écrit : pour tout  $t, s \in \mathbb{R}^+$ ,  $P(X > t + s) = P(X > t)P(X > s)$ , ce qui permet de montrer pour  $p, q$  entiers naturels non nuls,

$$P(X > \frac{p}{q}) = P(X > \frac{1}{q})^p = P(X > 1)^{\frac{p}{q}},$$

puis par décroissance en  $t$  de  $t \in \mathbb{R}^+ \mapsto P(X > t)$ , il suit que pour  $t \in \mathbb{R}^+$  cette fois,

$$P(X > t) = P(X > 1)^t = e^{-\lambda t}, \quad \lambda := -\log(P(X > 1)),$$

et on reconnaît la fonction de répartition de la loi Exp (de paramètre  $-\log(\mathbb{P}(X > 1))$ ), mais ceci est toujours vrai ... ) La réciproque est immédiate par un simple calcul.  $\square$

**Lemme 1.32.** *Si  $X \sim \text{Exp}(\lambda)$  pour  $\lambda > 0$ , et  $\mu > 0$ , alors  $\mu \cdot X \sim \text{Exp}(\lambda/\mu)$ .*

*Démonstration.* Soit  $x \in \mathbb{R}$ .  $F_{\mu \cdot X}(x) = F_X(x/\mu) = (1 - e^{-\frac{\lambda}{\mu}x})\mathbb{1}_{\mathbb{R}^+}(x)$  et on reconnaît la fonction de répartition de la loi  $\text{Exp}(\lambda/\mu)$ .  $\square$

#### 1.6.4 Loi gamma.

On dit qu'une variable aléatoire  $X$  suit la loi gamma de paramètres  $\alpha, \beta > 0$  (symboliquement,  $X \sim \Gamma(\alpha, \beta)$ ) si elle admet la densité

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{]0, +\infty[}(x),$$

où  $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$  est appelée la fonction gamma d'Euler ; elle vérifie  $\Gamma(n) = (n-1)!$  pour tout entier naturel  $n$ , et interpole donc la fonction factorielle. La fonction de répartition de la loi  $\Gamma(\alpha, \beta)$  n'admet pas d'expression simple.

La loi gamma est une généralisation de la loi exponentielle : on retrouve  $\text{Exp}(\beta)$  quand  $\alpha = 1$ , et on verra que lorsque le paramètre  $\alpha$  est un entier, la loi gamma de paramètres  $\alpha, \beta > 0$  peut être obtenue comme la loi de la somme de  $\alpha$  variables aléatoires *indépendantes* de loi  $\text{Exp}(\beta)$  (la notion d'indépendance de variables aléatoires sera introduite plus tard d'ailleurs le cours). Les valeurs non entières de  $\alpha$  donnent alors un sens à des sommes *non entières* de variables aléatoires  $\text{Exp}(\beta)$ , tout comme la fonction Gamma d'Euler, notée  $\Gamma$ , peut être vue comme une interpolation de la fonction factorielle.

Elle est aussi reliée à la loi de Poisson, et intervient par ce biais dans de nombreuses applications :

**Exercice 1.33.** *Soit  $t > 0$  un réel et  $n \in \mathbb{N}$  un entier, soit  $X \sim \text{Po}(t)$  et  $Y \sim \Gamma(n, 1)$ . Montrer que*

$$\mathbb{P}(X < n) = \mathbb{P}(Y > t).$$



Il s'agit de faire subir une IPP à la quantité :

$$\begin{aligned} P(Y > t) &= \frac{1}{(n-1)!} \int_t^\infty x^{n-1} e^{-x} dx \\ &= \frac{1}{(n-1)!} \left( \left[ - \int_t^\infty x^{n-1} e^{-x} \right]_t^\infty + (n-1) \int_t^\infty x^{n-2} e^{-x} dx \right) \\ &= \frac{1}{(n-1)!} t^{n-1} e^{-t} + \frac{1}{(n-2)!} \int_t^\infty x^{n-2} e^{-x} dx \end{aligned}$$

et donc par récurrence finie,

$$P(Y > t) = \sum_{j=0}^{n-1} \frac{1}{j!} t^j e^{-t} = \sum_{j=0}^{n-1} P(X = j) = P(X < n)$$

### 1.6.5 Loi Beta.

On dit qu'une variable aléatoire  $X$  suit la loi beta de paramètres  $a, b > 0$  (symboliquement,  $X \sim \beta(a, b)$ ) si elle admet la densité

$$f_X(x) = \frac{1}{\beta(a, b)} x^{a-1} (1-x)^{b-1} \mathbb{1}_{]0,1[}(x),$$

où  $\beta(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$  est la fonction beta, reliée comme suit à la fonction  $\Gamma$ ,

$$\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad (1.2)$$

en particulier, si  $a$  et  $b$  sont des entiers non nuls,  $\beta(a, b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}$ , ce qui peut se prouver par IPP. La fonction de répartition de la loi  $\beta(a, b)$  n'admet pas d'expression simple. Nous verrons plus tard dans le cours (une fois la notion d'indépendance introduite) que les lois  $\beta$  et  $\Gamma$  sont liées par des relations étroites, tous comme les fonctions  $\beta$  et  $\Gamma$ ; au passage, la relation (1.2) se justifie comme suit :

$$\begin{aligned} \Gamma(a)\Gamma(b) &= \int_0^\infty x^{a-1} e^{-x} dx \int_0^\infty y^{b-1} e^{-y} dy \\ &= \int_{\mathbb{R}^2} (x+y)^{a+b-2} \left(\frac{x}{x+y}\right)^{a-1} \left(\frac{y}{x+y}\right)^{b-1} e^{-(x+y)} \mathbb{1}_{0 < x, 0 < y} dx dy \\ &= \int_0^\infty \left( \int_0^\infty (x+y)^{a+b-2} \left(\frac{x}{x+y}\right)^{a-1} \left(\frac{y}{x+y}\right)^{b-1} e^{-(x+y)} dy \right) dx \\ &= \int_0^\infty s^{a+b-2} e^{-s} \left( \int_0^s \left(\frac{x}{s}\right)^{a-1} \left(1 - \frac{x}{s}\right)^{b-1} dx \right) ds \\ &= \left( \int_0^\infty s^{a+b-1} e^{-s} ds \right) \left( \int_0^1 r^{a-1} (1-r)^{b-1} dr \right) = \Gamma(a+b)\beta(a, b). \end{aligned}$$

ayant utilisé Fubini positif (aussi appelé Fubini-Tonelli) à la première ligne, posé le changement de variables affine  $x/s = r$  la dernière ligne, ce qui donne  $dr = dx/s$ . Dans le cas particulier où  $a$  et  $b$  sont des entiers, ce calcul s'exprime simplement à l'aide du coefficient binomial (on a un peu transformé l'expression en prenant les variables  $a$  et  $b$  plutôt que  $a-1$  et  $b-1$  :

$$\binom{a+b}{a} \int_0^1 x^a (1-x)^b dx = \frac{1}{a+b+1},$$

on peut en fait donner un sens probabiliste à cette jolie identité (exercice, une fois vue la notion d'indépendance!).

### 1.6.6 Loi de Cauchy.

On dit qu'une variable aléatoire  $X$  suit la loi de Cauchy, et on note  $X \sim \mathcal{C}(a)$  si  $X$  possède la densité :

$$f_X(x) = \frac{a}{\pi} \frac{1}{a^2 + x^2}.$$

Sa fonction de répartition vaut (exercice)

$$F_X(t) = \int_{-\infty}^t \frac{a}{\pi} \frac{1}{a^2 + x^2} dx = \int_{-\infty}^t \frac{1}{\pi} \frac{1}{1 + (x/a)^2} \frac{dx}{a} = \int_{-\infty}^{t/a} \frac{1}{\pi} \frac{1}{1 + y^2} dy = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x}{a}\right).$$

La densité de cette loi est encore en « forme de courbe en cloche », mais elle décroît beaucoup moins vite que la densité de la loi normale. Elle possède des propriétés remarquables, à commencer par l'égalité le lien suivant entre les lois de  $X$  et  $1/X$  (variable aléatoire presque sûrement bien définie puisque  $P(X \neq 0) = 1$ ).

**Exercice 1.34.** 1. Montrer l'identité  $\arctan(t) + \arctan(1/t) = \frac{\pi}{2}(\mathbf{1}_{t>0} - \mathbf{1}_{t<0})$  valable pour tout  $t \neq 0$ .

2. Soit  $X \sim \mathcal{C}(a)$ . Déduire de l'identité précédente la loi de  $1/X$ .

Si  $\Theta \sim \text{Unif}(-\pi/2, \pi/2)$  est un angle uniforme entre  $-\pi/2$  et  $\pi/2$ , la variable aléatoire  $\tan(\Theta) \sim \mathcal{C}(1)$ , puisque, pour  $t \in \mathbb{R}$ ,

$$P(\tan(\Theta) \leq t) = P(\Theta \leq \arctan(t)) = \frac{1}{2} + \frac{\arctan(t)}{\pi}.$$

Interprétation : si un rayon lumineux issu de l'origine forme un angle  $\Theta$  avec l'axe des abscisses, l'intersection de ce rayon avec la droite d'équation  $x = 1$  est  $\tan(\Theta)$ .

### 1.6.7 Autres lois à densité

À nouveau, à la différence des lois précédentes, les lois de cette section ne sont pas à connaître par coeur.

**Loi de Weibull.** On dit qu'une variable aléatoire  $X$  suit la loi Weibull de paramètres  $\alpha, \beta > 0$  si sa fonction de répartition s'écrit

$$F_X(x) = \begin{cases} 1 - \exp(-(x/\alpha)^\beta), & \text{si } x \geq 0 \\ 0, & \text{si } x < 0. \end{cases}$$

En dérivant, on obtient la densité

$$f_X(x) = \left(\frac{\beta}{\alpha}\right) \left(\frac{x}{\alpha}\right)^{\beta-1} \exp(-(x/\alpha)^\beta) \mathbf{1}_{x \geq 0}.$$

Cette loi est une autre généralisation de la loi exponentielle. Elle est utilisée entre autre pour modéliser des lois d'usures de pièces.

## 1.7 Loi d'une fonction d'une variable aléatoire

Si  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  est une fonction et  $X$  une variable aléatoire, alors  $\varphi(X)$  est encore un nombre qui décrit le résultat d'une expérience aléatoire, il est donc naturel de considérer  $\varphi(X)$  comme une nouvelle variable aléatoire. Plus généralement, si  $\varphi$  est seulement définie sur un ensemble  $E \subset \mathbb{R}$  et si la variable  $X$  est à valeurs dans  $E$ , donc si  $P(X \in E^c) = 0$ , il y a un sens à parler de  $\varphi(X)$ . Pour définir  $\varphi(X)$  comme une variable aléatoire, il faut alors définir  $P(\varphi(X) \in A)$  pour tout ensemble  $A$ . La seule définition naturelle est la suivante : pour tout ensemble  $A \subset \mathbb{R}$ , réunion finie d'intervalles disjoints,

$$P(\varphi(X) \in A) := P(X \in \varphi^{-1}(A)),$$

où  $\varphi^{-1}(A)$  est la préimage de  $A$  par l'application  $\varphi$ . On suppose ici que l'ensemble  $\varphi^{-1}(A)$  est encore une réunion finie d'intervalles disjoints. Ceci est une hypothèse sur la fonction  $\varphi$  qui est satisfaite par toute fonction « raisonnable ». On peut vérifier (exercice), que la fonction  $A \mapsto P(\varphi(X) \in A)$  ainsi définie satisfait encore aux axiomes 1 à 4. En pratique il suffira pour calculer la fonction de répartition de  $\varphi(X)$  de déterminer les ensembles  $\varphi^{-1}(]-\infty, x])$ .

### 1.7.1 Fonction discrète d'une v.a.r.

Dans le cas où la fonction  $\varphi$  envoie le support de la v.a.r.  $X$  (quelconque) sur un ensemble au plus dénombrable,  $\varphi(X)$  est une variable aléatoire discrète (et ceci même si la v.a.r.  $X$  ne l'était pas). Dans ce cas, la loi de  $\varphi(X)$  est caractérisée par sa fonction de masse. Plutôt que d'énoncer un théorème général, nous détaillons la méthode à suivre sur quelques exemples.

**Exemple 1.35.**  $\varphi : \mathbb{R} \rightarrow \{0, 1\}$ ,  $\varphi(x) = \mathbb{1}_{x \geq 0}$ ,  $X$  une v.a. quelconque. La fonction  $\varphi$  ne prend alors que deux valeurs, 0 ou 1. Par conséquent, la v.a.  $\varphi(X) = \mathbb{1}_{X \geq 0}$  est discrète. Sa fonction de masse est donnée par :

$$P(\mathbb{1}_{X \geq 0} = 1) = P(X \geq 0), \quad P(\mathbb{1}_{X \geq 0} = 0) = P(X < 0).$$

La v.a.  $\mathbb{1}_{X \geq 0}$  suit alors la loi de Bernoulli de paramètre  $p = P(X \geq 0)$ .

**Exemple 1.36.**  $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ ,  $\varphi(x) = \min(x, N)$  (avec  $N \in \mathbb{N}$ ),  $X$  une v.a. discrète à valeurs dans  $\mathbb{N}$ . Puisque  $X$  est une v.a. discrète,  $\varphi(X)$  l'est aussi. Sa fonction de masse est donnée par

$$P(\min(X, N) = n) = \begin{cases} P(X = n), & n \in \{0, \dots, N-1\} \\ P(X \geq N) = \sum_{k=N}^{\infty} P(X = k), & n = N. \end{cases}$$

**Exemple 1.37.**  $\varphi : ]0, 1[ \rightarrow \mathbb{N}$ ,  $\varphi(x) = \lfloor Nx \rfloor$  (avec  $N \in \mathbb{N}$ ),  $X \sim \text{Unif}(0, 1)$ . La fonction  $\varphi$  est à valeurs dans  $\{0, \dots, N-1\}$ , la v.a.  $\varphi(X)$  est donc discrète. Sa fonction de masse est donnée pour tout  $n \in \{0, \dots, N-1\}$  par :

$$P(\lfloor NX \rfloor = n) = P(n/N \leq X < (n+1)/N) = \int_{n/N}^{(n+1)/N} 1 \, dx = \frac{1}{N}.$$

La v.a.  $\lfloor NX \rfloor$  suit donc la loi uniforme sur  $\{0, \dots, N-1\}$ .

**Exemple 1.38.**  $\varphi : ]0, 1[ \rightarrow \mathbb{N}$ ,  $\varphi(x) = \lfloor x \rfloor$ ,  $X \sim \text{Exp}(\lambda)$ . La fonction  $\varphi$  est à valeurs dans  $\mathbb{N}$ , la v.a.  $\varphi(X)$  est donc discrète. Sa fonction de masse est donnée pour tout  $n \in \mathbb{N}$  par :

$$P(\lfloor X \rfloor = n) = P(n \leq X < n+1) = \int_n^{n+1} \lambda e^{-\lambda x} dx = e^{-\lambda n} - e^{-\lambda(n+1)} = e^{-\lambda n}(1 - e^{-\lambda})$$

La v.a.  $\lfloor X \rfloor$  suit donc la loi  $\text{Geom}_{\mathbb{N}}(1 - e^{-\lambda})$ .

**Exercice 1.39.** Retrouver ce résultat sur la partie entière d'une variable aléatoire de loi Exp en comparant les propriétés d'absence de mémoire 1.23 et 1.31 satisfaites par les lois Geom et Exp respectivement.

### 1.7.2 Fonction régulière d'une variable aléatoire à densité

Dans le cas où  $X$  est une v.a.r. à densité et la fonction  $\varphi$  est suffisamment régulière, alors  $\varphi(X)$  est encore une variable aléatoire à densité; la méthode consiste alors à calculer la fonction de répartition pour en déduire la densité à l'aide la Proposition 1.18. Nous détaillons plusieurs exemples, tout d'abord un exemple générique, c'est-à-dire avec une v.a.  $X$  quelconque

**Exemple 1.40.**  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ ,  $\varphi(x) = x^2$ ,  $X$  une v.a.r. On commence par calculer la fonction de répartition de  $\varphi(X)$  : pour  $x \geq 0$ ,

$$F_{X^2}(x) = P(X^2 \leq x) = P(-\sqrt{x} \leq X \leq \sqrt{x}) = F_X(\sqrt{x}) - F_X((-\sqrt{x})-).$$

C'est-à-dire :

$$F_{X^2}(x) = \begin{cases} F_X(\sqrt{x}) - F_X((-\sqrt{x})-), & x \geq 0 \\ 0, & x < 0. \end{cases}$$

Si maintenant  $X$  est à densité de densité  $f_X$ , alors on peut écrire que :

$$F_X(\sqrt{x}) - F_X((-\sqrt{x})-) = \int_{-\infty}^{\sqrt{x}} f_X(y) dy - \int_{-\infty}^{-\sqrt{x}} f_X(y) dy = \int_{-\sqrt{x}}^{\sqrt{x}} f_X(y) dy.$$

Il existe alors deux méthodes pour conclure :

- ou bien on observe que  $x \mapsto F_{X^2}(x)$  est de classe  $\mathcal{C}^1$  sur  $] -\infty, 0[ \cup ]0, +\infty[$  : par la Proposition 1.18, il suffit de dériver cette fonction (dérivation d'une fonction composée) pour obtenir que  $X^2$  est encore à densité, de densité :

$$f_{X^2}(x) = \begin{cases} \frac{1}{2\sqrt{x}}(f_X(\sqrt{x}) + f_X(-\sqrt{x})) & x > 0 \\ 0 & \text{sinon.} \end{cases}$$

- ou bien on se met directement sous la forme de la définition d'une variable aléatoire à densité en utilisant deux changements de variables comme suit :

$$\begin{aligned} \int_{-\sqrt{x}}^{\sqrt{x}} f_X(y) dy &= \int_{-\sqrt{x}}^0 f_X(y) dy + \int_0^{\sqrt{x}} f_X(y) dy \\ &= \int_0^{\sqrt{x}} (f_X(-y) + f_X(y)) dy \\ &= \int_0^x [f_X(-\sqrt{z}) + f_X(\sqrt{z})] \frac{1}{2\sqrt{z}} dz \end{aligned}$$

Et maintenant un exemple pratique où  $X$  suit une loi donnée (et où les ensembles de niveaux sont des réunions infinies d'intervalles).

**Exemple 1.41.**  $\varphi : \mathbb{R} \rightarrow [0, 1[$ ,  $\varphi(x) = x - \lfloor x \rfloor$ ,  $X \sim \text{Exp}(\lambda)$ . On commence par calculer la fonction de répartition de  $\varphi(X)$  : pour tout  $x \in [0, 1]$ ,

$$\begin{aligned}
 F_{\varphi(X)}(x) &= P(\varphi(X) \leq x) \\
 &= P(0 \leq X - \lfloor X \rfloor \leq x) \\
 &= P(\lfloor X \rfloor \leq X \leq \lfloor X \rfloor + x) \\
 &= P(\{\lfloor X \rfloor \leq X \leq \lfloor X \rfloor + x\} \cap \bigcup_{k \in \mathbb{N}} \{\lfloor X \rfloor = k\}) \quad \text{en distinguant selon la valeur de la partie entière} \\
 &= \sum_{k \in \mathbb{N}} P(\{\lfloor X \rfloor \leq X \leq \lfloor X \rfloor + x\} \cap \{\lfloor X \rfloor = k\}) \\
 &= \sum_{k \in \mathbb{N}} P(\{k \leq X \leq k + x\} \cap \{\lfloor X \rfloor = k\}) \\
 &= \sum_{k \in \mathbb{N}} P(k \leq X \leq k + x) \\
 &= \sum_{k \in \mathbb{N}} \int_k^{k+x} \lambda e^{-\lambda x} dx \\
 &= \sum_{k \in \mathbb{N}} e^{-\lambda k} (1 - e^{-\lambda x}) \\
 &= \frac{1 - e^{-\lambda x}}{1 - e^{-\lambda}}
 \end{aligned}$$

C'est-à-dire :

$$F_{\varphi(X)}(x) = \frac{1 - e^{-\lambda x}}{1 - e^{-\lambda}} \mathbb{1}_{[0,1[}(x) + \mathbb{1}_{[1,\infty[}(x)$$

et on identifie la densité via l'écriture :

$$F_{\varphi(X)}(x) = \int_{-\infty}^x \frac{\lambda e^{-\lambda y}}{1 - e^{-\lambda}} \mathbb{1}_{[0,1]}(y) dy$$

Une dernière méthode possible est la méthode dite de la fonction muette, elle aussi basée sur le changement de variable, qui sera étudiée au chapitre suivant.

## 1.8 Simulation de lois

On conclut ce (long!) chapitre par une méthode de simulation basée sur la notion de fonction de répartition. On appelle fonction quantile l'inverse *généralisé* de la fonction de répartition défini par :

$$F_X^{-1}(y) = \inf \{x : F_X(x) \geq y\}, \quad y \in ]0, 1[$$

On énonce d'abord quelques propriétés de la fonction quantile (non nécessaires pour la méthode de simulation qui suit, et qu'on pourra sauter en première lecture).

**Lemme 1.42.** *La fonction quantile  $F_X^{-1}$  est croissante, continue à gauche et admet des limites à droite.*

On note que les propriétés de continuité sont échangées par rapport à celle de la fonction de répartition  $F_X$ . Notons aussi que la preuve (purement analytique) de ces propriétés se base sur les seules hypothèses de croissance et de continuité à droite de  $F_X$ .

*Démonstration.* Si  $y \leq z$  sont deux réels dans  $]0, 1[$ , on a l'inclusion d'ensembles  $\{x : F(x) \geq y\} \supset \{x : F(x) \geq z\}$  qui implique bien  $F^{-1}(y) \leq F^{-1}(z)$  : c'est la croissance. Si maintenant  $(y_n)_n$  est une suite croissante de  $]0, 1[$ , avec  $y = \lim y_n \in ]0, 1[$ , considérons  $z = \lim_n F^{-1}(y_n)$ . Alors  $F(z + \epsilon) \geq y_n$  pour tout  $n$  donc  $F(z + \epsilon) \geq y$  d'où  $z + \epsilon \geq F^{-1}(y)$  puis  $z \geq F^{-1}(y)$ . Par croissance de la suite  $(y_n)_n$  et de  $F^{-1}$ , on a par ailleurs toujours l'inégalité dans l'autre sens :  $z \leq F^{-1}(y)$ , d'où l'égalité annoncée, et la continuité à gauche. L'existence des limites à droite découle de la propriété de croissance.  $\square$

Donnons un exemple pour lequel  $F^{-1}$  ne sera pas continue à droite : supposons qu'on puisse trouver  $y \in ]0, 1[$  et  $x_1 < x_2$  tels que  $F(x_1) = F(x_2) = y$  (c'est à dire que  $[x_1, x_2]$  est un intervalle de constance de la fonction  $F$ ) alors, si  $(y_n)_n$  est une suite strictement décroissante de limite  $y$ , on a  $F^{-1}(y_n) \geq x_2$  tandis que  $F^{-1}(y) \leq x_1$ , en conséquence de quoi :  $F^{-1}(y_n) - F^{-1}(y) \geq x_2 - x_1 > 0$ .

L'intérêt de la fonction quantile est de permettre la simulation très simple de la variable aléatoire associée :

**Proposition 1.43.** *Soit  $U$  variable aléatoire de loi uniforme  $\text{Unif}(0, 1)$  et  $X$  une var. Alors  $F_X^{-1}(U)$  suit la loi de  $X$ .*

L'intérêt de la méthode réside dans le fait qu'elle vaut quelque soit la variable aléatoire réelle  $X$  (on utilise seulement la notion de fonction de répartition).

*Démonstration.* Il suffit de vérifier par double inclusion que quelque soit  $u, x \in ]0, 1[$ ,

$$F_X^{-1}(u) \leq x \text{ ssi } u \leq F_X(x),$$

et on conclura alors en prenant les probabilités des deux événements  $\{F_X^{-1}(U) \leq x\} = \{U \leq F_X(x)\}$  :  $\mathbb{P}(F_X^{-1}(U) \leq x) = \mathbb{P}(U \leq F_X(x)) = F_X(x)$  et donc  $F_X^{-1}(U)$  a pour fonction de répartition  $F_X$ , ce qui suffit à conclure.

Supposons  $F_X^{-1}(u) \leq x$ . Posons  $y = F_X^{-1}(u)$ . Alors quelque soit  $\epsilon > 0$ ,  $F_X(y + \epsilon) \geq u$  et par continuité à droite de  $F_X$ , on tire  $F_X(y) = F_X(y+) \geq u$ . Comme par ailleurs  $F_X(y) \leq F_X(x)$  suit de la croissance de  $F_X$ , on a bien  $u \leq F_X(x)$  comme attendu. Réciproquement, si l'on suppose  $u \leq F_X(x)$ , alors  $x \in \{z : F_X(z) \geq u\}$  donc  $F_X^{-1}(u) \leq x$  suit de la définition de  $F_X$  comme l'infimum de l'ensemble  $\{z : F_X(z) \geq u\}$ .  $\square$

Attention, on n'a pas toujours  $F_X \circ F_X^{-1}(u) = u$  : considérer le cas où  $F_X(x-) < u < F_X(x)$  par exemple ; de même on n'a pas toujours  $F_X^{-1} \circ F_X(x) = x$ , considérer le cas où il existe  $w < x$  tel que  $F_X(w) = F_X(x)$  : nous invitons le lecteur à faire des dessins dans chacun de ces cas.

De façon intéressante, le calcul de  $F_X^{-1}$  permet souvent de comprendre la dépendance d'une loi en ses paramètres. Considérer le cas des deux exemples suivants :

**Exemple 1.44.**  $X \sim \text{Exp}(\lambda)$ . Pour  $y \in ]0, 1[$ ,  $F_X^{-1}(y) = \frac{-\log(1-y)}{\lambda}$  : ainsi  $\frac{-\log(1-U)}{\lambda}$  (ou encore  $\frac{-\log(U)}{\lambda}$  qui a même loi) suit la loi  $\text{Exp}(\lambda)$  et donc  $\lambda X \sim \text{Exp}(1)$ .

**Exemple 1.45.**  $X \sim \mathcal{C}(a)$ , Pour  $y \in ]0, 1[$ ,  $F_X^{-1}(y) = a \tan(\pi(y - \frac{1}{2}))$ , ainsi  $a \tan(\pi(U - \frac{1}{2}))$  (ou encore  $a \tan(\pi(U))$  qui a même loi) suit la loi  $\mathcal{C}(a)$  et donc  $X/a \sim \mathcal{C}(1)$ .

En pratique le calcul numérique de  $F^{-1}$  peut être long (comment calcule-t-on un "log" sur une machine?) et d'autres méthodes (comme la méthode du rejet) sont utilisées dans ce cas.





## Chapitre 2

# Espérance et moments

Supposons qu'on lance une pièce équilibrée  $n$  fois et qu'on s'intéresse à la proportion de lancers « pile » parmi ces  $n$  lancers. Puisque la pièce est non biaisée, chaque face a probabilité  $1/2$  d'apparaître, et on s'attend alors à ce que la proportion de piles approche  $1/2$  quand  $n$  tend vers l'infini : c'est l'interprétation dite fréquentiste, dans laquelle les probabilités s'interprètent comme une fréquence d'apparition limite. Plus généralement, supposons qu'on ait une variable aléatoire  $X$  décrivant le résultat d'une expérience aléatoire et qu'on répète  $n$  fois cette expérience en notant  $x_1, \dots, x_n$  les valeurs de cette variable aléatoire observées lors des  $n$  répétitions (le cas précédent correspond à  $X$  qui suit la loi Bernoulli de paramètre  $1/2$ ). On s'attend alors à ce que la moyenne  $(X_1 + \dots + X_n)/n$  approche une constante quand  $n$  tend vers l'infini, égale à la somme sur les valeurs possibles de la variable aléatoire  $X$  des valeurs possiblement prises par la variable fois leur probabilité ; cette somme pondérée, qui serait en un sens la moyenne de la variable aléatoire  $X$ , est appelée l'*espérance* de la variable aléatoire  $X$ , quelquefois l'espérance mathématique, et notée  $E[X]$ . La définition de l'espérance de  $X$  (et plus généralement, celle de  $f(X)$  où  $f$  est une fonction) est l'un des concepts les plus importants en théorie des probabilités.

L'espérance permet de définir les *moments* d'une variable aléatoire ; la suite des moments fournit en retour une description de la loi qui peut être utile : dans certains cas, elle caractérise la loi ; surtout, elle fournit des bornes intéressantes sur les probabilités d'événements qu'on ne peut calculer directement. Elle est aussi utilisée pour définir la *fonction génératrice des moments*, qui fournit encore une autre façon de décrire la loi de la variable aléatoire, tout comme la fonction de répartition, mais avec un résultat souvent plus explicite (l'expression de la fonction de répartition ne se simplifie que rarement).

## 2.1 Espérance d'une variable aléatoire

### 2.1.1 Variable aléatoire positive ou nulle

Dans le cas où  $X$  est une variable aléatoire positive ou nulle on peut toujours définir son espérance :

**Définition 2.1.** Soit  $X$  une variable aléatoire *positive ou nulle*, discrète ou à densité . Alors

l'espérance de  $X$ , notée  $E[X] \in [0, \infty]$ , est définie par

$$E[X] = \begin{cases} \sum xP(X = x) & \text{si } X \text{ est discrète} \\ \int_0^{+\infty} xf_X(x) dx & \text{si } X \text{ admet la densité } f_X, \end{cases}$$

et on dit que  $X$  est intégrable si la somme ou l'intégrale dans l'expression précédente est convergente.

L'espérance  $E[X]$  est intuitivement la moyenne pondérée des valeurs que  $X$  peut prendre, les poids étant les probabilités que ces valeurs soient prises. On peut la rapprocher à la notion de *centre de gravité* au sens de la mécanique. Pour illustrer cette analogie, supposons d'abord que  $X$  est discrète et notons  $x_i$  les valeurs que peut prendre  $X$ . Imaginons-nous alors des masses situées sur l'axe réelle aux abscisses  $x_i$  et de poids  $P(X = x_i)$ . L'espérance  $E[X]$  correspond alors à l'abscisse du centre de gravité de ce groupe de masses. Si  $X$  est à densité, alors on s'imagine une masse d'une unité répartie continuellement sur axe réelle selon la densité  $f_X$ . L'espérance  $E[X]$  correspond alors encore à l'abscisse du centre de gravité de cette masse.

Insistons en particulier sur le fait que l'espérance d'une variable aléatoire positive est positive (les termes de la somme, comme l'intégrande dans l'intégrale sont positifs) : c'est la positivité de l'intégrale.

**Exemple 2.2.** Soit  $X \sim \text{Ber}(p)$ ,  $p \in [0, 1]$ . Alors

$$E[X] = 0 \cdot (1 - p) + 1 \cdot p = p.$$

Interprétation : en répétant  $n$  fois une expérience de Bernoulli de probabilité de succès  $p$ , on s'attend à ce que la proportion de succès approche  $p$  quand  $n$  tend vers l'infini.

**Exemple 2.3.** Soit  $X \sim \text{Geom}(p)$ ,  $p \in ]0, 1]$ . Alors

$$E[X] = \sum_{n=1}^{\infty} np(1-p)^{n-1} = \frac{1}{p},$$

ou la dernière série se calcule par dérivation de la série entière  $\sum_{n=1}^{\infty} (1-p)^n = \frac{1}{p}$  à l'intérieur du disque de convergence, c'est-à-dire pour  $|1-p| < 1$ . Interprétation : Quand on répète une expérience de Bernoulli de probabilité de succès  $p$  jusqu'à ce qu'elle réussisse, on doit la répéter en moyenne  $1/p$  fois. Par exemple, si une expérience a 5% de chance de réussir, il faut la répéter en moyenne  $1/0.05 = 20$  fois jusqu'à ce qu'elle réussisse.

**Exemple 2.4.** Soit  $X \sim \text{Exp}(\lambda)$ ,  $\lambda > 0$ . Alors

$$E[X] = \int_0^{\infty} x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}.$$

Interprétation : Une pièce ayant un taux de panne constant égal à  $\lambda$  a une durée de vie de  $1/\lambda$  en moyenne.

Toutes les variables aléatoires précédentes sont intégrables.

### 2.1.2 Variable aléatoire signée

**Définition 2.5.** Soit  $X$  une variable aléatoire discrète ou à densité. On dit que  $X$  est intégrable si

$$\begin{cases} \sum |x|P(X = x) < \infty & \text{si } X \text{ est discrète,} \\ \int_{-\infty}^{+\infty} |x|f_X(x) dx < \infty & \text{si } X \text{ admet la densité } f_X. \end{cases}$$

Dans ce cas on définit l'espérance de  $X$  par la formule suivante :

$$E[X] = \begin{cases} \sum xP(X = x) & \text{si } X \text{ est discrète,} \\ \int_{-\infty}^{+\infty} xf_X(x) dx & \text{si } X \text{ admet la densité } f_X. \end{cases}$$

On connaît déjà une loi non intégrable, c'est la *loi de Cauchy*. En effet, si  $X \sim \mathcal{C}(a)$ , alors

$$\int_{\mathbb{R}} \frac{a}{\pi} \cdot \frac{|x|}{a^2 + x^2} dx = \infty,$$

car l'intégrande est équivalent à  $C/x$  en  $+\infty$  par exemple.

**Exemple 2.6.** Soit  $X \sim \text{Unif}(a, b)$ ,  $a < b$ . Alors

$$E[X] = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \left[ \frac{1}{2}x^2 \right]_a^b = \frac{a+b}{2},$$

car  $b^2 - a^2 = (a+b)(b-a)$ . Interprétation : Le centre de gravité d'une masse répartie uniformément dans l'intervalle  $[a, b]$  se trouve au point d'abscisse  $(a+b)/2$ .

## 2.2 Espérance d'une fonction d'une variable aléatoire

Il sera utile d'étendre la définition de l'espérance ci-dessus à l'espérance d'une fonction d'une variable aléatoire.

**Proposition 2.7** (Formule de transfert). *Soit  $X$  une variable aléatoire discrète ou à densité et  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  une fonction.  $\varphi(X)$  est une variable aléatoire intégrable ssi*

$$\begin{cases} \sum_x |\varphi(x)|P(X = x) < \infty & \text{si } X \text{ est discrète} \\ \int_{-\infty}^{+\infty} |\varphi(x)|f_X(x) dx < \infty & \text{si } X \text{ admet la densité } f_X. \end{cases}$$

Dans ce cas, son espérance est donnée par :

$$E[\varphi(X)] = \begin{cases} \sum \varphi(x)P(X = x) & \text{si } X \text{ est discrète} \\ \int_{-\infty}^{+\infty} \varphi(x)f_X(x) dx & \text{si } X \text{ admet la densité } f_X. \end{cases} \quad (2.1)$$

Notons que l'intégrabilité garantit l'absolue convergence de la série/intégrale qui donne l'espérance. Cette définition généralise celle de  $E[X]$  qu'on retrouve en prenant pour  $\varphi$  l'application identité  $\varphi(x) = x$ . Appliquant cette définition avec  $\varphi = \mathbf{1}_A$ , on retrouve bien :

$$E[\mathbf{1}_A(X)] = 1 \cdot P(X \in A) + 0 \cdot P(X \notin A) = P(X \in A),$$

ce qui est cohérent avec l'exemple 2.2 puisque  $\mathbf{1}_A(X)$  suit une loi de Bernoulli de paramètre  $P(X \in A)$ .

*Démonstration.* Considérons d'abord le cas où  $X$  est discrète. Alors  $\varphi(X)$  l'est aussi et :

$$P(\varphi(X) = y) = \sum_x P(X = x) \mathbf{1}_{\varphi(x)=y}$$

On commence par traiter l'énoncé au sujet de l'intégrabilité, en écrivant la suite d'égalités suivantes, dans  $[0, +\infty]$  à l'aide de Fubini positif :

$$\begin{aligned} \sum_y |y| P(\varphi(X) = y) &= \sum_y |y| \sum_x P(X = x) \mathbf{1}_{\varphi(x)=y} \\ &= \sum_y \sum_x |\varphi(x)| P(X = x) \mathbf{1}_{\varphi(x)=y} \\ &= \sum_x \varphi(x) P(X = x) \sum_y \mathbf{1}_{|\varphi(x)|=y} \\ &= \sum_x |\varphi(x)| P(X = x), \end{aligned}$$

On en déduit bien que

$$\sum_y |y| P(\varphi(X) = y) < \infty \text{ ssi } \sum_x |\varphi(x)| P(X = x) < \infty$$

Ainsi :

$$\begin{aligned} E[\varphi(X)] &= \sum_y y P(\varphi(X) = y) = \sum_y y \sum_x P(X = x) \mathbf{1}_{\varphi(x)=y} \\ &= \sum_y \sum_x \varphi(x) P(X = x) \mathbf{1}_{\varphi(x)=y} \\ &= \sum_x \varphi(x) P(X = x) \sum_y \mathbf{1}_{\varphi(x)=y} \\ &= \sum_x \varphi(x) P(X = x), \end{aligned}$$

On considère maintenant le cas où  $X$  est à densité. On traitera seulement ici le cas où  $\varphi$  est  $\mathcal{C}^1$  et de dérivée  $\varphi'(x) > 0$  pour tout  $x \in \mathbb{R}$ . Alors il existe  $-\infty \leq a < b \leq +\infty$  tels que  $\text{Im}(\varphi) = (a, b)$ . Soit  $x \in \text{Im}(\varphi)$ . On a alors :

$$\begin{aligned} P(\varphi(X) \leq x) &= P(X \leq \varphi^{-1}(x)) \\ &= \int_{-\infty}^{\varphi^{-1}(x)} f_X(y) dy \\ &= \int_a^x f_X(\varphi^{-1}(y)) (\varphi^{-1})'(y) dy \\ &= \int_{-\infty}^x f_X(\varphi^{-1}(y)) (\varphi^{-1})'(y) \mathbf{1}_{\text{Im}\varphi}(y) dy \end{aligned}$$

puis on étend facilement l'énoncé à  $x \in \mathbb{R}$ , ce qui permet d'identifier  $f_X(\varphi^{-1}(y))(\varphi^{-1})'(y)\mathbf{1}_{\text{Im}\varphi}(y)$  comme une densité de  $\varphi(X)$ . Maintenant,

$$\begin{aligned} E[\varphi(X)] &= \int_{-\infty}^{+\infty} y f_{\varphi(X)}(y) dy = \int_{-\infty}^{+\infty} y f_X(\varphi^{-1}(y))(\varphi^{-1})'(y) dy \\ &= \int_{-\infty}^{+\infty} \varphi(x) f_X(x) dx, \end{aligned}$$

par un changement de variable  $x = \varphi^{-1}(y)$ ,  $dx = (\varphi^{-1})'(y) dy$ . Ceci conclut la preuve du théorème.  $\square$

Une conséquence importante qui découle très facilement de la formule de transfert est la suivante :

**Corollaire 2.8** (Linéarité de l'intégrale). *Si  $f_1$  et  $f_2$  sont des fonctions telles que  $f_1(X)$  et  $f_2(X)$  sont intégrables, alors  $(f_1 + f_2)(X)$  est intégrable et*

$$E[(f_1 + f_2)(X)] = E[f_1(X)] + E[f_2(X)],$$

*En particulier, si  $X$  est intégrable,*

$$E[aX + b] = aE[X] + b,$$

*Démonstration.* Et dans le cas où  $X$  est une variable aléatoire discrète. Noter qu'on a la droit de découper la somme en 2 car  $f_1(X)$  et  $f_2(X)$  sont intégrables :

$$\begin{aligned} E[(f_1 + f_2)(X)] &= \sum_x (f_1(x) + f_2(x))P(X = x) \\ &= \sum_x f_1(x)P(X = x) + \sum_x f_2(x)P(X = x) \\ &= E[f_1(X)] + E[f_2(X)] \end{aligned}$$

De même, si  $X$  est à densité, par (2.1),

$$\begin{aligned} E[(f_1 + f_2)(X)] &= \int_{-\infty}^{+\infty} (f_1(x) + f_2(x))f_X(x) dx \\ &= \int_{-\infty}^{+\infty} f_1(x)f_X(x) dx + \int_{-\infty}^{+\infty} f_2(x)f_X(x) dx \\ &= E[f_1(X)] + E[f_2(X)]. \end{aligned}$$

Pour le deuxième item, on prend seulement  $f_1(x) = a.x$  et  $f_2(x) = b$ .  $\square$

Une conséquence directe du théorème de transfert, de la linéarité et de la positivité est le résultat suivant de comparaison, dont la preuve est laissée au lecteur :

**Lemme 2.9.** *Soit  $X$  une variable aléatoire à valeurs dans  $I$ . Si  $\varphi_1, \varphi_2 : I \rightarrow \mathbb{R}^+$  vérifient  $\varphi_1 \leq \varphi_2$ , alors  $E[\varphi_1(X)] \leq E[\varphi_2(X)]$ . En particulier, si  $\varphi_2(X)$  est intégrable, alors  $\varphi_1(X)$  est intégrable.*

Pour les variables aléatoires de loi symétrique, certains calculs d'espérance peuvent se simplifier.

**Définition 2.10.** On dit que la loi de la variable aléatoire  $X$  est symétrique si  $X$  et  $-X$  sont égales en loi, c'est-à-dire que pour tout  $A \subset \mathbb{R}$ ,

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

On peut encore écrire cette dernière égalité :  $\mathbb{P}(X \in A) = \mathbb{P}(X \in -A)$  si  $-A = \{-x, x \in A\}$ . Notons la condition suffisante suivante :

**Lemme 2.11.** *La loi de  $X$  est symétrique dans les deux cas suivants :*

1.  $X$  est une variable aléatoire discrète dont la fonction de masse est paire.
2.  $X$  est une variable aléatoire à densité dont une densité est paire.

*Démonstration.* Si  $X$  est une variable aléatoire discrète dont la fonction de masse est paire :

$$\mathbb{P}(-X \in A) = \mathbb{P}(X \in -A) = \sum_{x_k \in -A} \mathbb{P}(X = x_k) = \sum_{-x_k \in A} \mathbb{P}(X = -x_k) = \sum_{x_k \in A} \mathbb{P}(X = x_k) = \mathbb{P}(X \in A)$$

Si  $X$  est une variable aléatoire à densité dont une densité est paire :

$$\mathbb{P}(-X \in A) = \mathbb{P}(X \in -A) = \int_{\mathbb{R}} \mathbf{1}_{-A}(x) f_X(x) dx = \int_{\mathbb{R}} \mathbf{1}_{-A}(-x) f_X(-x) dx = \int_{\mathbb{R}} \mathbf{1}_A(x) f_X(x) dx$$

□

**Lemme 2.12.** *Soit  $X$  une variable aléatoire symétrique et  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  une fonction impaire, c'est-à-dire  $\varphi(-x) = -\varphi(x)$  pour tout  $x \in \mathbb{R}$ , telle que  $\varphi(X)$  est intégrable. Alors*

$$E[\varphi(X)] = 0$$

Il est clef ici que la variable considérée soit intégrable.

*Démonstration.* On note simplement que :

$$\begin{aligned} E[\varphi(X)] &= E[\varphi(-X)] && (X \stackrel{\text{loi}}{=} -X) \\ &= E[-\varphi(X)] && (\varphi \text{ impaire}) \\ &= -E[\varphi(X)] && (\text{linéarité de l'espérance}). \end{aligned}$$

Par conséquent,  $E[\varphi(X)] = 0$ .

□

Par exemple, si  $X$  est symétrique et intégrable, alors  $E[X] = 0$  (utiliser  $\varphi(x) = x$ ). si l'on note  $x_+ = \max\{x, 0\} = x\mathbf{1}_{x>0}$  et  $x_- = \max\{-x, 0\} = -x\mathbf{1}_{x<0}$  les parties positives et négatives du réel  $x$ , qui vérifient  $x = x_+ - x_-$  et  $|x| = x_+ + x_-$ . On peut relier l'intégrabilité d'une variable aléatoire à celle de ses parties positives et négatives comme suit :

**Lemme 2.13.** *Soit  $X$  une variable aléatoire.  $X$  est intégrable ssi  $|X|$  est intégrable ssi  $X_+$  et  $X_-$  sont intégrables. Dans ce cas, on peut alors écrire :*

$$E[|X|] = E[X_+] + E[X_-] \text{ et } E[X] = E[X_+] - E[X_-].$$

### 2.2.1 Caractérisation des lois de fonctions de variables aléatoires

La formule dite de transfert permet de caractériser la loi d'une fonction d'une variable aléatoire ; on comparera cette méthode avec celle exposée à la section 1.7 qui utilise la seule notion de fonction de répartition. Les deux méthodes sont également recevables, et vous pouvez choisir celle que vous préférez. La proposition suivante permet de caractériser la loi d'une variable aléatoire  $g(X)$ .

**Proposition 2.14.** *Soit  $X$  une v.a.r, et  $g : \mathbb{R} \rightarrow \mathbb{R}$  une fonction. S'il existe une fonction  $f : \mathbb{R} \rightarrow \mathbb{R}$  pour toute fonction  $\varphi$  bornée,*

$$\mathbb{E}[\varphi(g(X))] = \int_{\mathbb{R}} \varphi(y)f(y)dy$$

alors  $f$  est une densité de probabilité, et  $g(X)$  est une v.a.r. à densité de densité  $f$ .

La fonction  $\varphi$  porte le nom de "fonction test" ou encore "fonction muette" : aussi la méthode est quelquefois appelée méthode de la fonction muette. Pour que le résultat se mette sous la forme demandée il faut en fait partir d'une v.a.r.  $X$  à densité.

*Démonstration.* Il suffit de choisir  $\varphi = \mathbb{1}_{]-\infty, x]}$  : on alors

$$\mathbb{P}(g(X) \leq x) = \mathbb{E}[\mathbb{1}_{]-\infty, x]}(g(X))] = \int_{-\infty}^{+\infty} \mathbb{1}_{]-\infty, x]}(y)f(y)dy = \int_{-\infty}^x f(y)dy$$

et puisque la fonction de répartition caractérise la loi, la v.a.  $g(X)$  admet bien  $f$  pour densité.  $\square$

On note qu'il suffit en fait que la classe de fonctions  $\varphi$  pour laquelle la propriété vaut contienne les fonctions  $(\mathbb{1}_{]-\infty, x]}, x \in \mathbb{R})$ , ce qui revient à appliquer la caractérisation des lois par les fonctions de répartition. Conceptuellement il n'y a pas de différence entre les deux méthodes, et on pourrait même argumenter que l'emploi d'une fonction test obscurcit le propos par rapport à l'usage de la seule fonction de répartition. Les exemples suivants pourront aider à faire son choix entre les deux méthodes.

**Exemple 2.15.**  $g : \mathbb{R} \rightarrow \mathbb{R}$ ,  $g(x) = x^2$ ,  $X$  une v.a.r. à densité, de densité  $f_X$ ,  $\varphi$  bornée.

$$\begin{aligned} \mathbb{E}[\varphi(X^2)] &= \int_{\mathbb{R}} \varphi(x^2)f_X(x)dx \\ &= \int_{-\infty}^0 \varphi(x^2)f_X(x)dx + \int_0^{\infty} \varphi(x^2)f_X(x)dx \\ &= \int_0^{\infty} \varphi(y)f_X(-\sqrt{y})\frac{1}{2\sqrt{y}}dy + \int_0^{\infty} \varphi(y)f_X(\sqrt{y})\frac{1}{2\sqrt{y}}dy \\ &= \int_{\mathbb{R}} \varphi(y)(f_X(-\sqrt{y}) + f_X(\sqrt{y}))\mathbb{1}_{\mathbb{R}^+}(y)dy \end{aligned}$$

$g : \mathbb{R} \rightarrow \mathbb{R}$ ,  $g(x) = |x|$ ,  $X$  une v.a.r. à densité, de densité  $f_X$ ,  $\varphi$  bornée.

$$\begin{aligned} \mathbb{E}[\varphi(|X|)] &= \int_{\mathbb{R}} \varphi(|x|) f_X(x) dx \\ &= \int_{-\infty}^0 \varphi(-x) f_X(x) dx + \int_0^{\infty} \varphi(x) f_X(x) dx \\ &= \int_0^{\infty} \varphi(y) f_X(-y) dy + \int_0^{\infty} \varphi(x) f_X(x) dy \\ &= \int_{\mathbb{R}} \varphi(y) (f_X(-y) + f_X(y)) \mathbb{1}_{\mathbb{R}^+}(y) dy \end{aligned}$$

$g : \mathbb{R} \rightarrow \mathbb{R}$ ,  $g(x) = x - \lfloor x \rfloor$ ,  $X \sim \text{Exp}(\lambda)$ ,

$$\begin{aligned} \mathbb{E}[\varphi(X - \lfloor X \rfloor)] &= \int_{\mathbb{R}} \varphi(x - \lfloor x \rfloor) \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}^+}(x) dx \\ &= \int_0^{\infty} \varphi(x - \lfloor x \rfloor) \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}^+}(x) dx \\ &= \sum_{n=0}^{\infty} \int_n^{n+1} \varphi(x - \lfloor x \rfloor) \lambda e^{-\lambda x} dx \\ &= \sum_{n=0}^{\infty} \int_0^1 \varphi(n + x - \lfloor n + x \rfloor) \lambda e^{-\lambda(n+x)} dx \\ &= \left( \sum_{n=0}^{\infty} e^{-\lambda n} \right) \left( \int_0^1 \varphi(x) \lambda e^{-\lambda x} dx \right) \\ &= \int_{\mathbb{R}} \varphi(x) \left( \frac{\lambda e^{-\lambda x}}{1 - e^{-\lambda}} \mathbb{1}_{[0,1]}(x) \right) dx \end{aligned}$$

$g : \mathbb{R} \rightarrow \mathbb{R}$ ,  $g(x) = 1/x$ ,  $X \sim \mathcal{C}(a)$ ,

$$\begin{aligned} \mathbb{E}[\varphi(1/X)] &= \int_{\mathbb{R}} \varphi\left(\frac{1}{x}\right) \frac{a}{\pi} \frac{1}{a^2 + x^2} dx \\ &= \int_{-\infty}^0 \varphi\left(\frac{1}{x}\right) \frac{1}{\pi} \frac{1}{a^2 + x^2} dx + \int_0^{+\infty} \varphi\left(\frac{1}{x}\right) \frac{a}{\pi} \frac{1}{a^2 + x^2} dx \\ &= \int_{-\infty}^0 \varphi(y) \frac{a}{\pi} \frac{1}{a^2 + (\frac{1}{y})^2} \frac{dy}{y^2} + \int_0^{+\infty} \varphi(y) \frac{a}{\pi} \frac{1}{a^2 + (\frac{1}{y})^2} \frac{dy}{y^2} \\ &= \int_{\mathbb{R}} \varphi(y) \frac{1/a}{\pi} \frac{1}{(1/a)^2 + y^2} dy \end{aligned}$$

donc  $g(X) \sim \mathcal{C}(1/a)$ .

Cette proposition a évidemment son pendant discret, dont la preuve est similaire.

**Proposition 2.16.** *Soit  $X$  une v.a.r., et  $g : \mathbb{R} \rightarrow \mathbb{R}$  une fonction. S'il existe un ensemble  $S$  au plus dénombrable et une fonction  $p : S \rightarrow ]0, +\infty[$  telle que pour toute fonction  $\varphi$  bornée,*

$$\mathbb{E}[\varphi(g(X))] = \sum_{y \in S} \varphi(y) p(y)$$

alors  $g(X)$  est une variable aléatoire discrète, de fonction de masse  $p$ .



Un nouvel exemple dans ce dernier cadre.

**Exemple 2.17.**  $g : \mathbb{R} \rightarrow \mathbb{R}$ ,  $g(x) = \lfloor x \rfloor$ ,  $X \sim \text{Exp}(\lambda)$  une v.a.r. à densité, de densité  $f_X$ ,

$$\begin{aligned} \mathbb{E}[\varphi(\lfloor X \rfloor)] &= \int_{\mathbb{R}} \varphi(\lfloor x \rfloor) \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}^+}(x) dx \\ &= \sum_{n=0}^{\infty} \int_n^{n+1} \varphi(\lfloor x \rfloor) \lambda e^{-\lambda x} dx \\ &= \sum_{n=0}^{\infty} \varphi(n) \int_n^{n+1} \lambda e^{-\lambda x} dx \\ &= \sum_{n=0}^{\infty} \varphi(n) (e^{-\lambda n} - e^{-\lambda(n+1)}) \\ &= \sum_{n=0}^{\infty} \varphi(n) e^{-\lambda n} (1 - e^{-\lambda}) \end{aligned}$$

Donc  $\lfloor X \rfloor$  est un v.a.r. discrète de fonction de masse  $n \in \mathbb{N} \mapsto \mathbb{P}(\lfloor X \rfloor = n) = e^{-\lambda n} (1 - e^{-\lambda})$ .

## 2.3 Moments et quantités associées

**Définition 2.18.** Si  $X$  est une variable aléatoire telle que  $X^k$  est intégrable, alors on dit que  $X$  admet un moment d'ordre  $k$  et on appelle la quantité  $E[X^k]$ ,  $k \in \mathbb{N}^*$  son  $k$ -ième moment (ou le  $k$ -ième moment de sa loi).

On dit encore que  $X$  est de puissance  $k$  intégrable lorsque  $X^k$  est intégrable. Les moments d'une variable aléatoire décrivent différents aspects de sa loi. Aussi et surtout ils permettent de donner des bornes sur des probabilités d'événements, comme nous le verrons plus loin dans ce chapitre. Nous allons nous concentrer dans ce cours sur les deux premiers moments (c'est-à-dire  $k = 1, 2$ ), qui sont les plus importants. Tout d'abord quelques généralités. Formellement, on peut toujours définir les moments pairs, puisque la fonction  $x \mapsto x^k$  est positive pour  $k \in \mathbb{N}$  pair, mais ces moments peuvent alors être infinis.

**Lemme 2.19.** Soit  $k < m$  deux entiers. Si  $X$  admet un moment d'ordre  $m$ , alors  $X$  admet un moment d'ordre  $k$  et

$$E[|X|^k] \leq E[|X|^m] + 1$$

*Démonstration.* On note que  $|x|^k \leq 1 + |x|^m$  pour tout  $x \in \mathbb{R}$  et on applique le lemme de comparaison 2.9.  $\square$

On note que les moments impairs sont simples à calculer quand la loi est paire :

**Lemme 2.20.** Si la loi de  $X$  est symétrique, et  $X$  admet un moment d'ordre  $k \in \mathbb{N}$  impair, alors  $E[X^k] = 0$ .

*Démonstration.* C'est une application du lemme 2.12 avec la fonction  $\varphi(x) = x^k$  impaire puisque  $k$  est impair.  $\square$

### 2.3.1 Espérance

Le premier moment  $E[X]$  est défini pour toute v.a.r. intégrable, ou positive ou nulle. Il est appelé *l'espérance* ou la *moyenne* de la variable aléatoire  $X$ , et dans le cas où  $\mathbb{E}[X] = 0$ , on dit que la v.a.  $X$  est *centrée*. Notons qu'il est toujours possible de centrer la v.a.  $X$  en lui retranchant son espérance : on appelle  $X - E[X]$  la *variable centrée* ; par linéarité de l'espérance, on a bien  $E[X - E[X]] = E[X] - E[X] = \mu - \mu = 0$ .

### 2.3.2 Variance

Le second moment est  $E[X^2]$ , il est défini pour toute variable aléatoire, puisque  $X^2 \geq 0$ , mais il peut être infini.

**Définition 2.21.** Soit  $X$  variable aléatoire. On dit que  $X$  est de carré intégrable si  $E[X^2] < +\infty$ . Dans ce cas, la *variance* de  $X$  est définie par

$$\text{Var}(X) = E[X^2] - E[X]^2.$$

Si  $X$  est de carré intégrable,  $X$  admet un premier moment, donc la définition utilisant  $E[X^2]$  et  $E[X]$  est légitime. Une définition équivalente plus intuitive est la suivante :

**Lemme 2.22.** Soit  $X$  variable aléatoire de carré intégrable. On a aussi :

$$\text{Var}(X) = E[(X - E[X])^2]$$

De la définition comme du lemme, on a que  $\text{Var}(X) \geq 0$  (utiliser Jensen pour la définition ou la positivité de la fonction carré pour le lemme). La variance est un *coefficient de dispersion* : il mesure comment les valeurs de  $X$  sont étalées autour de l'espérance. Si la variance est grande, les valeurs sont très étalées, alors que si la variance est petite, les valeurs sont concentrées autour de l'espérance. (Un cas extrême est donnée par une v.a.r. de variance nulle, voir 2.24)

*Démonstration.* On écrit :

$$(X - E[X])^2 = X^2 - 2XE[X] + E[X]^2$$

relation dont on prend l'espérance (noter que tous les termes à droite sont bien intégrables) ; grâce à la linéarité de l'espérance, on obtient :

$$E[(X - E[X])^2] = E[X^2 - 2XE[X] + E[X]^2] = E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2$$

□

Si  $X$  a une « dimension », c'est à dire si  $X$  mesure une quantité qui possède une dimension physique (par exemple une longueur en mètres), alors la variance aura comme dimension le *carré* de la dimension de  $X$  (par exemple mètres carrés). Pour obtenir une quantité de la même dimension que  $X$  on prend alors la racine carrée de la variance qu'on nomme *écart-type*, noté souvent  $\sigma \geq 0$ , et définit par

$$\sigma = \sqrt{\text{Var}(X)}.$$

Cela conduit à noter la variance par  $\sigma^2$ . L'écart-type donne l'ordre de grandeur de la distance entre les valeurs typiques de  $X$  et son espérance  $\mu$ .

**Définition 2.23.** Si  $X$  est une v.a. de carré intégrable, d'espérance  $\mu$  et d'écart-type  $\sigma > 0$ , on appelle

$$\frac{X - \mu}{\sigma}$$

la *variable centrée réduite* (ou simplement *variable réduite*) associée à  $X$ .

Elle est d'espérance nulle et de variance égale à 1 (et écart-type égal à 1), car

$$\begin{aligned} E\left[\frac{X - \mu}{\sigma}\right] &= \frac{1}{\sigma} \mathbb{E}[X - \mu] = \frac{1}{\sigma} \times 0 = 0 \\ \text{Var}\left(\frac{X - \mu}{\sigma}\right) &= \mathbb{E}\left[\left(\frac{X - \mu}{\sigma}\right)^2\right] = \frac{1}{\sigma^2} \mathbb{E}[(X - \mu)^2] = \frac{1}{\sigma^2} \text{Var}(X) = 1. \end{aligned}$$

En poursuivant le raisonnement « dimensionnel » ci-dessus, la variable centrée réduite est une quantité *adimensionnelle*, car c'est le quotient de deux quantités de même dimension. Passer à la variable centrée réduite est le moyen « naturel » de normaliser une variable aléatoire de telle façon à obtenir une variable d'espérance nulle et de variance 1.

**Propriétés de la variance.** Nous résumons ici quelques propriétés importantes de la variance :

**Proposition 2.24.** Soit  $X$  une variable aléatoire de carré intégrable.

1. Si  $a, b \in \mathbb{R}$ ,

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

(avec  $0 \times \infty = 0$ ). En particulier,  $\text{Var}(X + b) = \text{Var}(X)$ .

2.  $\text{Var}(X) = 0$  si et seulement si  $X$  est p.s. constante, c'est-à-dire que  $X \sim \delta_\mu$ .

*Démonstration.* 1. Par linéarité de l'espérance,  $E[aX + b] = aE[X] + b = a\mu + b$ , si bien que

$$\begin{aligned} \text{Var}(aX + b) &= E[(aX + b - E[aX + b])^2] \\ &= E[(aX + b - aE[X] + b)^2] \\ &= E[a^2(X - E[X])^2] \\ &= a^2 \text{Var}(X). \end{aligned}$$

2. On définit la variable aléatoire  $Y = (X - \mu)^2$ . Alors  $Y \geq 0$  et  $Y = 0$  si et seulement si  $X = \mu$ . Le Lemme 2.25 ci-dessous appliqué à  $Y$  donne alors

$$X \sim \delta_\mu \iff E[Y] = 0 \iff \text{Var}(X) = 0.$$

□

**Lemme 2.25.** Soit  $Y$  une v.a. positive. Alors  $Y \sim \delta_0$  si et seulement si  $E[Y] = 0$ .

La preuve de la réciproque (l'implication non évidente) sera faite à la fin de la Section 2.5.1.

**L'espérance est la constante qui approche le mieux la variable aléatoire** Dans le cadre des variables aléatoires de carré intégrable, une façon naturelle d'introduire espérance et variance est de les voir comme solution du problème d'optimisation suivant :

**Lemme 2.26.** *Soit  $X$  une variable aléatoire de carré intégrable. On a :*

$$\text{Var}(X) = \min_{a \in \mathbb{R}} E[(X - a)^2],$$

et le minimum est atteint en un unique élément  $a$  égal à  $E[X]$ .

Cet énoncé dit que la *constante* qui approche le mieux la variable aléatoire  $X$  au sens de la distance  $(X, Y) \mapsto \mathbb{E}[(X - Y)^2]$  (dite des "moindres carrés" ou encore du "risque quadratique", un terme plus particulièrement employé en économie) est la constante égale à  $E[X]$ .

*Démonstration.* Par linéarité de l'intégrale,

$$\begin{aligned} E[(X - a)^2] &= E[a^2 - 2aX + X^2] \\ &= a^2 - 2aE[X] + E[X^2] \\ &= (a - E[X])^2 + (E[X^2] - E[X]^2) \\ &= (a - E[X])^2 + \text{Var}(X) \\ &\geq \text{Var}(X), \end{aligned}$$

avec égalité si et seulement si  $a = E[X]$ . □

Un corollaire simple de ce résultat est alors que

**Corollaire 2.27.** *Soit  $X$  variable aléatoire à valeurs dans  $[0, 1]$ . Alors  $X$  est de carré intégrable et*

$$\text{Var}(X) \leq 1/4$$

*Démonstration.* D'après le lemme précédent, pour tout  $y \in \mathbb{R}$ ,  $\text{Var}(X) \leq E[(X - y)^2]$ . On choisit alors  $y = 1/2$  et on observe que  $|X - 1/2| \leq 1/2$  implique alors :  $\text{Var}(X) \leq E[(X - 1/2)^2] \leq E[1/4] = 1/4$ . □

On peut se demander quelle constante approche le mieux une variable aléatoire au sens de la norme  $L^1$  cette fois, c'est-à-dire en quel point  $a \mapsto \mathbb{E}[|X - a|]$  atteint son infimum (en fait un minimum). C'est un exercice plus ardu, qui conduit à l'introduction de la notion de médiane.

## 2.4 Fonction génératrice des moments

On définit pour tout réel  $t$ , la *fonction génératrice des moments*  $\varphi_X$  de la variable aléatoire  $X$  par

$$\varphi_X(t) = E[e^{tX}] \in ]0, +\infty[.$$

Comme son nom l'indique, c'est un outil adapté au calcul des moments au moyen de développements en série entière :

**Proposition 2.28.** *Supposons qu'il existe  $t_0 > 0$  tel que  $\varphi_X(t) < \infty$  pour tout  $|t| < t_0$ . Alors pour tout  $|t| < t_0$ ,*

$$\varphi_X(t) = \sum_{k=0}^{\infty} E[X^k] \frac{t^k}{k!},$$

*et le rayon de convergence la série entière du membre de droite est supérieur ou égal à  $t_0$ . Dans ce cas,  $\varphi_X$  admet des dérivées à tout ordre sur  $] -t_0; t_0[$  et on a, pour tout  $k \in \mathbb{N}^*$ ,*

$$\mathbb{E}[X^k] = \varphi_X^{(k)}(0),$$

où  $\varphi_X^{(k)}$  est la  $k$ -ième dérivée de la fonction  $\varphi_X$ .

*Démonstration.* On suppose  $X$  discrète, la preuve pour  $X$  à densité étant très similaire. On montre d'abord que  $E[e^{|tX|}] < \infty$  pour tout  $|t| < t_0$ . En effet,  $e^{|tx|} \leq e^{tx} + e^{-tx}$  pour tout  $x \in \mathbb{R}$  et donc

$$E[e^{|tX|}] \leq E[e^{tX}] + E[e^{-tX}] < \infty,$$

par hypothèse. Ceci garantit l'intégrabilité de  $e^{tX}$  pour  $|t| < t_0$ , et on calcule alors pour ces valeurs de  $t$ ,

$$\begin{aligned} \varphi_X(t) &= E[e^{tX}] \\ &= \sum_x e^{tx} P(X = x) \\ &= \sum_x \left( \sum_{k=0}^{\infty} \frac{(tx)^k}{k!} P(X = x) \right). \end{aligned}$$

Puisque

$$\sum_x \left( \sum_{k=0}^{\infty} \frac{|tx|^k}{k!} P(X = x) \right) = E[e^{|tX|}] < \infty$$

le théorème de Fubini s'applique et donne :

$$\varphi_X(t) = \sum_{k=0}^{\infty} \left( \sum_x \frac{(tx)^k}{k!} P(X = x) \right),$$

c'est-à-dire que cette série (en  $k$ ) converge et a comme limite  $\varphi_X(t)$ . En réarrangeant, on obtient

$$\begin{aligned} \varphi_X(t) &= \sum_{k=0}^{\infty} \left( \sum_x x^k P(X = x) \right) \frac{t^k}{k!} \\ &= \sum_{k=0}^{\infty} E[X^k] \frac{t^k}{k!}. \end{aligned}$$

En particulier, cette dernière série entière converge pour tout  $|t| < t_0$ , si bien que son rayon de convergence est supérieur ou égal à  $t_0$ . Le résultat sur les dérivées successives en 0 en découle.  $\square$

La fonction génératrice permet non seulement de retrouver les moments, mais également toute la loi. Ceci est le contenu du théorème suivant (admis) :

**Theorème 2.29.** *Supposons  $\varphi_X$  finie dans un voisinage de 0. Alors la loi de  $X$  est l'unique loi de fonction génératrice  $\varphi_X$ .*

Nous donnons maintenant plusieurs exemples :

**Exemple 2.30.** Soit  $a, b \in \mathbb{R}$ . La fonction génératrice des moments d'une combinaison linéaire s'exprime simplement :

$$\varphi_{aX+b}(t) = E[e^{t(aX+b)}] = E[e^{(at)X}]e^{tb} = \varphi_X(at)e^{tb}.$$

**Exemple 2.31.**  $X \sim \text{Po}(\lambda)$ ,  $\lambda \geq 0$ . Alors pour tout  $t \in \mathbb{R}$ ,

$$\varphi_X(t) = E[e^{tX}] = \sum_{n=0}^{\infty} e^{tn} e^{-\lambda} \frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} = e^{\lambda(e^t-1)}.$$

**Exemple 2.32.**  $X \sim \mathcal{N}(0, 1)$ . Alors pour tout  $t \in \mathbb{R}$ ,

$$\begin{aligned} \varphi_X(t) &= \int_{-\infty}^{+\infty} e^{-x^2/2+tx} \frac{dx}{\sqrt{2\pi}} \\ &= \int_{-\infty}^{+\infty} e^{-(x-t)^2/2+t^2/2} \frac{dx}{\sqrt{2\pi}} \\ &= e^{t^2/2} \int_{-\infty}^{+\infty} e^{-y^2/2} \frac{dy}{\sqrt{2\pi}} \quad (y = x - t) \\ &= e^{t^2/2}. \end{aligned}$$

Si  $Y \sim \mathcal{N}(\mu, \sigma^2)$ , alors  $Y \stackrel{\text{loi}}{=} \mu + \sigma X$ , donc

$$\varphi_Y(t) = e^{t\mu} \varphi_X(\sigma t) = e^{t\mu + \sigma^2 t^2/2}.$$

En particulier, si  $X \sim \mathcal{N}(0, 1)$ , on retrouve les moments de  $X$  par unicité du développement en série entière (DSE) :

$$\varphi_X(t) = \sum_{k=0}^{\infty} \frac{(t^2/2)^k}{k!} = \sum_{k=0}^{\infty} \frac{(2k)!}{2^k k!} \frac{t^{2k}}{(2k)!},$$

si bien que  $E[X^{2k+1}] = 0$  pour tout  $k \in \mathbb{N}$ , et

$$E[X^{2k}] = \frac{(2k)!}{2^k k!} = \prod_{i=1}^k (2i - 1),$$

cette dernière quantité étant également notée  $(2n - 1)!!$ . En particulier,  $E[X^4] = 3$ .

**Exemple 2.33.**  $X \sim \text{Exp}(1)$ . Alors

$$\varphi_X(t) = \int_0^{\infty} e^{tx} e^{-x} dx = \int_0^{\infty} e^{-(1-t)x} dx = \begin{cases} \frac{1}{1-t}, & t < 1 \\ +\infty, & t \geq 1. \end{cases}$$

On reconnaît une série géométrique :

$$\varphi_X(t) = \sum_{k=0}^{\infty} t^k = \sum_{k=0}^{\infty} k! \frac{t^k}{k!},$$

et on en déduit par unicité du DSE que  $\mathbb{E}[X^k] = k!$  pour tout  $k \in \mathbb{N}$ .

## 2.5 Inégalités

Nous présentons trois inégalités faisant intervenir l'espérance d'une variable aléatoire.

### 2.5.1 Inégalité de Markov

L'inégalité de Markov est l'inégalité fondamentale permettant de borner des probabilités par des espérances. Cette inégalité repose sur une inégalité simple de fonctions élémentaires.

**Lemme 2.34.** *Soit  $y \geq 0$  et  $x > 0$ . Alors*

$$\mathbf{1}_{y \geq x} \leq \frac{y}{x}.$$

*Démonstration.* On distingue les deux cas : si  $y \geq x$ , alors  $\mathbf{1}_{y \geq x} = 1 \leq y/x$ , et si  $y < x$ , alors  $0 \leq y/x$ , car  $y$  et  $x$  sont positifs.  $\square$

**Lemme 2.35.** *Soit  $X$  une v.a. et  $B \subset \mathbb{R}$ . Alors*

$$P(X \in B) = E[\mathbf{1}_{X \in B}].$$

*Démonstration.* La v.a.  $\mathbf{1}_{X \in B}$  suit la loi de Bernoulli de paramètre  $P(X \in B)$  car elle est à valeurs dans  $\{0, 1\}$  et elle vaut 1 avec probabilité  $P(X \in B)$ . Son espérance vaut alors  $P(X \in B)$ .  $\square$

**Theorème 2.36** (Inégalité de Markov). *Soit  $X$  une v.a. positive. Alors pour tout  $x > 0$ ,*

$$P(X \geq x) \leq \frac{E[X]}{x}.$$

*Démonstration.* Soit  $x > 0$ . Puisque  $X \geq 0$ , le Lemme 2.34 montre que  $\mathbf{1}_{X \geq x} \leq X/x$ . Par le Lemme 2.35,

$$P(X \geq x) = E[\mathbf{1}_{X \geq x}] \leq E\left[\frac{X}{x}\right] = \frac{E[X]}{x},$$

par linéarité de l'espérance.  $\square$

*Corollaire : preuve du Lemme 2.25.* Si  $Y \sim \delta_0$ , alors par définition de l'espérance,  $E[Y] = 0 \times 1 = 0$ . Si  $Y \not\sim \delta_0$ , alors il existe  $x > 0$  tel que  $P(Y \geq x) > 0$ . Par l'inégalité de Markov,

$$E[Y] \geq P(Y \geq x) \times x > 0.$$

Ceci conclut la preuve.  $\square$

### 2.5.2 Inégalité de Tchebychev

**Theorème 2.37** (Inégalité de Tchebychev). *Soit  $X$  une v.a. d'espérance finie. Alors pour tout  $x > 0$ ,*

$$P(|X - E[X]| > x) \leq \frac{\text{Var}(X)}{x^2}.$$

*Démonstration.* On remarque que

$$P(|X - E[X]| > x) = P((X - E[X])^2 > x^2).$$

La variable  $(X - E[X])^2$  étant positive, on peut appliquer l'inégalité de Markov (Theorem 2.36) pour obtenir

$$P(|X - E[X]| > x) \leq \frac{E[(X - E[X])^2]}{x^2} = \frac{\text{Var}(X)}{x^2}.$$

□

### 2.5.3 Inégalité de Jensen

Soit  $I \subset \mathbb{R}$  intervalle. On rappelle qu'une fonction  $\varphi : I \rightarrow \mathbb{R}$  est *convexe* si pour tout  $x, y \in I$ ,  $t \in [0, 1]$ ,

$$\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y).$$

*Remarque 2.38.* On note que si  $X$  est une variable aléatoire telle que  $\mathbb{P}(X = x) = p$ ,  $\mathbb{P}(X = y) = 1 - p$ , alors l'inégalité précédente s'écrit encore très simplement :

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)].$$

L'inégalité de Jensen consiste en la généralisation de cette inégalité aux variables aléatoires réelles  $X$  intégrables.

Si  $\varphi \in C^2(I)$ , alors  $\varphi$  est convexe si et seulement si  $\varphi'' \geq 0$ . Des exemples sont les fonctions

$$(x \mapsto e^x), (x \mapsto e^{-x}) \text{ et } (x \mapsto |x|^k) \text{ pour } k \geq 1.$$

Une fonction  $\varphi$  telle que  $-\varphi$  est convexe est dite *concave*. Exemples :

$$(x > 0 \mapsto \log x) \text{ et } (x \geq 0 \mapsto x^k) \text{ pour } k \leq 1$$

On admet qu'une fonction convexe admet des *minorants affines* en tout point : si  $x_0 \in I$ , alors il existe  $a \in \mathbb{R}$  tel que pour tout  $x \in I$ ,

$$\varphi(x_0) + a(x - x_0) \leq \varphi(x).$$

(Dans le cas où  $f$  est dérivable en  $x_0$ , on peut simplement choisir  $a = f'(x_0)$ ; sinon, de la convexité,  $f$  admet des dérivées à gauche et à droite distinctes, avec nécessairement  $f'((x_0)-) < f'((x_0)+)$ ; tout nombre  $a \in [f'((x_0)-), f'((x_0)+)]$  convient alors.)

**Theorème 2.39.** *Soit  $X$  une v.a. d'espérance finie, à valeurs dans un intervalle  $I \subset \mathbb{R}$  et  $\varphi : I \rightarrow \mathbb{R}$  convexe. Alors*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)],$$

*inégalité dans lequel le membre de droite peut être infini.*

Noter les deux cas particuliers suivants (variables aléatoires discrètes/à densité) : si les  $(p_k)$  sont des réels qui somment à 1, et les réels  $(x_k)_k$  sont tels que  $\sum p_k |x_k| < \infty$

$$\varphi\left(\sum_k p_k x_k\right) \leq \sum_k \varphi(x_k) p_k,$$

et si  $f : \mathbb{R} \rightarrow \mathbb{R}^+$  est une densité telle que  $\int |g(x)| f(x) dx < \infty$ , alors

$$\varphi\left(\int g(x) f(x) dx\right) \leq \int \varphi(g(x)) f(x) dx.$$



*Démonstration.* Posons  $x_0 = E[X]$ . On peut alors montrer qu'il existe  $a \in \mathbb{R}$ , tel que pour tout  $x \in \mathbb{R}$ ,

$$\varphi(x_0) + a(x - x_0) \leq \varphi(x).$$

Par conséquent,

$$E[\varphi(X)] \geq E[\varphi(x_0) + a(X - x_0)] = \varphi(x_0) + a(E[X - x_0]) = \varphi(x_0) + a(E[X] - x_0) = \varphi(x_0),$$

car  $x_0 = E[X]$ . □

**Exemple 2.40.** Soit  $t \in \mathbb{R}$ . La fonction  $x \mapsto e^{tx}$  est convexe sur  $\mathbb{R}$  donc, pour  $X$  variable aléatoire réelle intégrable,

$$e^{tE[X]} \leq E[e^{tX}],$$

ce dernier terme pouvant être infini.

**Exemple 2.41.** On peut préciser l'inégalité dans le lemme 2.19 comme suit. Soit  $k < m$  deux entiers. Si  $X$  admet un moment d'ordre  $k$ , alors, par convexité de  $x \mapsto x^{\frac{m}{k}}$  sur  $]0, +\infty[$ ,

$$E[|X|^k]^{\frac{m}{k}} \leq E[(|X|^k)^{\frac{m}{k}}] = E[|X|^m],$$

ce dernier moment pouvant être infini.

## 2.6 Autres définitions de l'espérance

On se demande ici s'il est possible d'offrir une version unifiée de la définition de l'espérance, qui pour le moment diffère selon que la variable sous-jacente est à densité ou discrète. Une telle définition doit : - faire intervenir la seule fonction de répartition (c'est la seule façon que l'on a de définir une loi générique), - coïncider avec les définitions précédentes pour les variables discrètes ou à densité. Une idée commune aux deux méthodes ci-dessous est l'intégration par parties. Cette fin de chapitre s'adresse aux étudiants désireux d'aller plus loin.

### 2.6.1 À la Lebesgue.

**Lemme 2.42.** Soit  $X$  variable aléatoire positive ou nulle de fonction de répartition  $F_X$ . On a l'égalité suivante dans  $[0, +\infty[$  :

$$E[X] = \int_0^\infty P(X > t) dt = \int_0^\infty (1 - F_X(t)) dt.$$

Cette égalité signifie que les quantités de part et d'autre du symbole d'égalité sont soit simultanément finies et égales, soit simultanément infinies.

On notera en particulier que l'intégrabilité de  $X$  est équivalente à l'intégrabilité de la fonction  $t \mapsto P(X > t) = 1 - F_X(t)$  au voisinage de  $+\infty$ .

Bien sûr, on ne peut *démontrer* ce résultat que dans le cas des variables discrètes ou à densité pour lesquelles l'espérance a été préalablement définie ; cependant, la formule a un sens dès lors que la fonction de répartition est définie, et on pourrait donc la prendre comme une *définition* de l'espérance pour toute variable aléatoire positive ou nulle. Nous avons choisi de ne pas procéder de la sorte pour des raisons pédagogiques, cette définition n'étant pas très naturelle de prime abord (on ne voit pas en quoi  $E[X]$  représente une moyenne pondérée).

*Démonstration.* Si  $X$  est à densité, de densité  $f_X$ , alors :

$$\begin{aligned}
 E[X] &= E \left[ \int_0^\infty \mathbb{1}_{X>t} dt \right] \\
 &= \int_0^\infty \left( \int_0^\infty \mathbb{1}_{x>t} dt \right) f_X(x) dx \\
 &= \int_0^\infty \left( \int_0^\infty \mathbb{1}_{x>t} f_X(x) dx \right) dt && \text{de Fubini positif} \\
 &= \int_0^\infty E[\mathbb{1}_{X>t}] dt \\
 &= \int_0^\infty P(X > t) dt && \text{du lemme 2.35}
 \end{aligned}$$

Le cas où  $X$  est discrète est analogue et laissé au lecteur.  $\square$

Dans le cas de variables aléatoires à valeurs entières, le résultat précédent prend une forme plus simple, très utilisée dans la pratique :

**Corollaire 2.43.** *Soit  $X$  variable aléatoire à valeurs dans  $\mathbb{N}$ , on a l'égalité suivante dans  $[0, +\infty[$  :*

$$E[X] = \sum_{n \in \mathbb{N}} P(X > n) = \sum_{n \in \mathbb{N}^*} P(X \geq n),$$

*Démonstration.* Deux possibilités : ou bien on reprend le raisonnement précédent en notant que pour tout entier  $m$ ,  $m = \sum_{n \in \mathbb{N}} \mathbb{1}_{m>n} = \sum_{n \in \mathbb{N}^*} \mathbb{1}_{m>n}$ , ou bien on note que pour tout  $t \in [n, n+1[$ ,  $P(X > t) = P(X \geq n+1)$ , d'où l'on déduit :

$$\int_n^{n+1} P(X > t) dt = \int_n^{n+1} P(X \geq n+1) dt = P(X \geq n+1)$$

qui implique bien le résultat cherché puisque :

$$\begin{aligned}
 \int_0^\infty P(X > t) dt &= \sum_{n \in \mathbb{N}} \int_n^{n+1} P(X > t) dt \\
 &= \sum_{n \in \mathbb{N}} P(X \geq n+1) \\
 &= \sum_{n \in \mathbb{N}^*} P(X \geq n) \\
 &= \sum_{n \in \mathbb{N}} P(X > n)
 \end{aligned}$$

$\square$

Aussi, notons le lien entre ces formules et l'intégrale de Riemann-Lebesgue, dont l'idée fondatrice est de considérer les contributions à l'aire sous la courbe en procédant à un découpage de l'axe des ordonnées plutôt que de l'axe des abscisses.

**Exemple 2.44.** On sait que si  $X \sim \text{Geom}(p)$ ,  $0 < p \leq 1$ , alors, pour  $n \in \mathbb{N}$ ,  $\mathbb{P}(X > n) = (1-p)^n$  (les  $n$  premières expériences de Bernoulli sont des échecs), et ainsi  $\mathbb{E}[X] = \sum_{n \in \mathbb{N}} (1-p)^n = 1/p$ .

Nous détaillons maintenant une autre méthode pour définir l'espérance d'une variable aléatoire à partir de sa seule fonction de répartition, fondée sur une autre notion d'intégrale.

### 2.6.2 À la Riemann-Stieltjes

L'intégrale de Stieltjes (aussi appelée Riemann-Stieltjes) est une intégrale très adaptée au calcul des probabilités : elle permet notamment d'unifier les deux cadres théoriques ici présentés séparément ; celui du calcul d'espérance de variables discrètes ou continues.

Soit  $a < b$  deux réels,  $f, g : [a, b] \rightarrow \mathbb{R}$ . On appelle *partition* de  $[a, b]$  une suite finie  $\pi = (x_0, \dots, x_n)$  de nombres réels avec  $a = x_0 < x_1 < \dots < x_n = b$  : l'entier  $n$  s'appelle alors la longueur de la partition, et le *pas de la partition* est la quantité  $\max_{i=0 \dots n-1} (x_{i+1} - x_i)$  ; considérons la quantité

$$I(f, g, \pi) = \sum_{i=0}^{n-1} f(c_i)[g(x_{i+1}) - g(x_i)]$$

pour  $c_i \in [x_i, x_{i+1}]$ . Si ces quantités admettent une limite lorsque le pas de la subdivision tend vers 0, alors la limite est appelée l'intégrale de Stieltjes de  $f$  par rapport à  $g$ , et notée :

$$\int_a^b f(x) dg(x).$$

Précisément, on dit que  $A \in \mathbb{R}$  est l'intégrale de Stieltjes de  $f$  par rapport à  $g$  si quelque soit  $\varepsilon > 0$ , il existe  $\delta > 0$  tel que pour toute partition  $\pi$  de  $[a, b]$  de pas  $\leq \delta$ , et pour tout choix de  $c_0, \dots, c_{n-1}$  où  $n$  désigne la longueur de la partition  $\pi$ ,  $|I(f, g, \pi) - A| \leq \varepsilon$ . Cette définition coïncide avec celle de l'intégrale de Riemann usuelle lorsque  $g(x) = x$ . Aussi, si  $g$  est dérivable alors on peut interpréter l'intégrale obtenue comme une intégrale de Riemann :

$$\int_a^b f(x) dg(x) = \int_a^b f(x) g'(x) dx.$$

Cette définition étant posée, l'enjeu principal est alors de comprendre quelles fonctions  $f$  (les *intégrandes*) et quelles fonctions  $g$  (les *intégrateurs*) donnent lieu à une intégrale de Stieltjes bien définie : un résultat important est l'existence de l'intégrale lorsqu'on prend  $f$  continue et  $g$  à variations bornée ( $\sum_{i=0}^{n-1} |g(x_{i+1}) - g(x_i)|$  admet une borne uniforme sur l'ensemble des partitions  $\pi$  de  $[a, b]$ ). Dans notre cadre probabiliste, soit  $X$  une variable aléatoire à valeurs dans  $[a, b]$  p.s., alors prenant pour  $g$  la fonction de répartition  $F_X$  de  $X$  ( $g$  étant croissante, elle est évidemment à variation bornée) on obtient, pour toute fonction  $\varphi$  continue :

$$\mathbb{E}[\varphi(X)] = \int_a^b \varphi(x) dF_X(x)$$

et cette formule vaut pour toute v.a.r.  $X$ , qu'elle soit discrète, à densité, ou ni l'une ni l'autre. Bien entendu, tout comme pour l'introduction de l'intégrale de Lebesgue, donner un sens à une intégrale ne permet pas plus de la calculer...

Dans le cas où la v.a.r.  $X$  ne prend pas ses valeurs dans un intervalle borné, alors on peut montrer que, pour  $\varphi$  continue bornée,

$$\int_a^b \varphi(x) dF_X(x) \rightarrow \int_{-\infty}^{+\infty} \varphi(x) dF_X(x) \in \mathbb{R}, \text{ quand } a \rightarrow -\infty, b \rightarrow \infty$$

et cette dernière quantité est prise comme définition de  $\mathbb{E}[\varphi(X)]$ .



## Chapitre 3

# Variables aléatoires indépendantes

### 3.1 Variables aléatoires simultanées

Nous avons vu qu'une variable aléatoire  $X$  décrit un aspect d'une expérience aléatoire. Par exemple, lors de  $n$  lancers d'une pièce équilibrée, on peut s'intéresser au nombre de lancers *pile* qu'on modélise par une variable aléatoire  $X$  de loi  $\text{Bin}(n, 1/2)$ . On pourrait également s'intéresser au nombre de lancers *face* qu'on pourrait également modéliser par une variable aléatoire  $Y$  de même loi. Ces deux variables aléatoires décrivent donc deux aspects de la même expérience aléatoire. Ces deux variables aléatoires sont bien sûr de même loi, mais elles retournent des résultats distincts (sauf si  $n$  est pair et  $X = Y = n/2$ ).

On peut aussi considérer  $n$  lancers de deux pièces équilibrées, et compter le nombre de piles  $X$  de la première pièce, et le nombre de piles  $Y$  de la seconde pièce. À nouveau,  $X$  et  $Y$  ont même loi, mais cette fois-ci, moins que les deux pièces ne partagent un mystérieux lien,  $X$  ne donne aucune information sur  $Y$  : c'est l'indépendance (stochastique).

Bien entendu, on peut ne pas se limiter à deux variables aléatoires, et définir  $n$  variables aléatoires  $X_1, \dots, X_n$  pour rendre compte des différents aspects d'une même expérience aléatoire. De manière générale, on peut considérer une infinité de variables aléatoires  $(X_i)_{i \in I}$  indicés par un ensemble  $I$  au plus dénombrable (dans ce cours).

D'un point de vue abstrait, il est utile de voir la donnée de  $n$  variables aléatoires  $X_1, \dots, X_n$  comme la donnée d'*un seul* vecteur aléatoire  $X = (X_1, \dots, X_n)$  à valeurs dans  $\mathbb{R}^n$ . On définit alors la loi du vecteur aléatoire  $X = (X_1, \dots, X_n)$  par la fonction

$$A \mapsto P(X \in A) = P((X_1, \dots, X_n) \in A), \quad A \subset \mathbb{R}^n$$

On dit encore qu'il s'agit de la *loi jointe* des variables aléatoires  $X_1, \dots, X_n$ . On doit encore se restreindre à des ensembles  $A$  pas complètement arbitraires, mais nous allons passer cela sous le silence. Cette fonction devra encore satisfaire à des axiomes tout à fait analogues aux axiomes 1.-4. vus au début du cours pour des variables aléatoires réelles : il suffit de remplacer dans ces axiomes  $X$  par  $(X_1, \dots, X_n)$  et  $A \subset \mathbb{R}$  par  $A \subset \mathbb{R}^n$ .

On voit dans cette définition qu'on peut également considérer  $(X_1, \dots, X_n)$  comme un *vecteur aléatoire*, la loi jointe des  $X_1, \dots, X_n$  est également appelée la *loi du vecteur aléatoire*  $(X_1, \dots, X_n)$ .

**Définition 3.1.** Pour tout  $i \in \{1, \dots, n\}$ , la loi de  $X_i$  donnée, pour tout  $A \subset \mathbb{R}$ , par :

$$P(X_i \in A) = P((X_1, \dots, X_n) \in \mathbb{R}^{i-1} \times A \times \mathbb{R}^{n-i}).$$

est appelée la *loi marginale* de la loi du vecteur  $X$ .

La *fonction de répartition* d'un vecteur aléatoire  $X = (X_1, \dots, X_n)$  est définie par

$$\mathbb{R}^n \mapsto [0, 1], x = (x_1, \dots, x_n) \mapsto F_X(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

Elle a des propriétés tout à fait analogues à celle d'une v.a.r. comme le montre le théorème suivant :

**Theorème 3.2.**

1. La fonction de répartition  $F_{X_1, \dots, X_n}$  d'un vecteur aléatoire  $X = (X_1, \dots, X_n)$  a les propriétés suivantes :
  - elle est croissante en chaque coordonnée,
  - $F_X(-\infty) := \lim F_X(x_1, \dots, x_n) = 0$ , quand  $\min\{x_1, \dots, x_n\} \rightarrow -\infty$
  - $F_X(+\infty) := \lim F_X(x_1, \dots, x_n) = 1$ , quand  $\min\{x_1, \dots, x_n\} \rightarrow +\infty$
  - elle est continue à droite en chaque coordonnée.
2. La loi jointe d'un vecteur aléatoire  $X = (X_1, \dots, X_n)$  est déterminée de manière unique par sa fonction de répartition conjointe  $F_{X_1, \dots, X_n}$ .

Rappelons nous qu'en dimension 1, ce théorème était complété par un troisième point, à savoir "3. Chaque fonction sur  $\mathbb{R}$  ayant les propriétés données dans 1. est la fonction de répartition d'une variable aléatoire  $X$ ". Ce point ne vaut pas ici, comme le montre le (contre)-exemple ci-dessous.

**Exemple 3.3.** Soit la fonction  $F$  suivante, pour  $n = 2$  :

$$F(x, y) = \mathbb{1}_{x+y \geq 0}$$

$F$  vérifie les 4 tirets du point 1 : elle est en chaque coordonnée croissante et continue à droite, et les valeurs des limites conviennent. Supposons maintenant qu'il existe un vecteur  $X = (X_1, X_2)$  de fonction de répartition  $F$ , alors, écrivant  $] - \infty, 1] \times ] - \infty, 1]$  comme la réunion de quatre rectangle disjoints basés sur le point  $(-1, -1)$ , on peut écrire :

$$\begin{aligned} P((X_1, X_2) \in ] - 1, 1] \times ] - 1, 1]) &= F(1, 1) - F(1, -1) - F(-1, 1) + F(-1, -1) \\ &= \mathbb{1}_{1+1 \geq 0} - \mathbb{1}_{1-1 \geq 0} - \mathbb{1}_{-1+1 \geq 0} + \mathbb{1}_{-1-1 \geq 0} \\ &= 1 - 2 + 0 = -1 < 0 \end{aligned}$$

ce qui est absurde. Donc  $F$  n'est pas la fonction de répartition d'un vecteur aléatoire de  $\mathbb{R}^2$ .

*Preuve (esquisse).* Nous n'allons pas démontrer ce théorème mais seulement en donner quelques idées clés pour nous en convaincre. Le point central est le fait que la fonction de répartition conjointe définit la loi jointe. En fait, en suivant la preuve du théorème dans le cas univarié, on voit facilement que la fonction de répartition définit de manière unique les probabilités de la forme

$$P(X_1 \in A_1, \dots, X_n \in A_n), \quad A_1, \dots, A_n \subset \mathbb{R},$$

donc les probabilités  $P((X_1, \dots, X_n) \in A)$  pour des ensembles  $A$  de la forme  $A_1 \times \dots \times A_n$ . On appelle ces ensembles des *pavés*. Pour passer de là à des ensembles généraux, on approche ces ensembles par des réunions ou intersections d'un nombre dénombrable de pavés. Il est laissé au lecteur de se convaincre à travers quelques exemples que cela est bien possible pour des ensembles « raisonnables » (par exemple des boules, des hyperplans...)  $\square$

### 3.1.1 Variables aléatoires discrètes

**Définition 3.4.** On dit que  $X = (X_1, \dots, X_n)$  est un vecteur aléatoire *discret* à valeurs dans  $\mathbb{R}^n$  ssi il existe un sous-ensemble au plus dénombrable  $S \subset \mathbb{R}^n$  tel que  $\mathbb{P}(X \in S) = 1$ .

**Proposition 3.5.**  $X$  est un vecteur aléatoire discret ssi les  $X_i$  sont des variables aléatoires réelles discrètes.

*Démonstration.* Si les v.a.  $X_1, \dots, X_n$  sont discrètes, c'est-à-dire si elles prennent leurs valeurs dans des ensembles au plus dénombrables  $E_1, \dots, E_n \subset \mathbb{R}$ , alors le vecteur  $(X_1, \dots, X_n)$  prend ses valeurs dans l'ensemble  $E_1 \times \dots \times E_n$ , qui est également au plus dénombrable. La réciproque est claire puisque la projection sur toute coordonnée d'un ensemble au plus dénombrable reste au plus dénombrable.  $\square$

La fonction de masse du vecteur aléatoire discret  $X = (X_1, \dots, X_n)$  est donnée par :

$$x = (x_1, \dots, x_n) \mapsto P(X_1 = x_1, \dots, X_n = x_n), \quad x_1, \dots, x_n \in \mathbb{R},$$

elle détermine la loi de ce vecteur au travers de la formule :

$$P((X_1, \dots, X_n) \in A) = \sum_{(x_1, \dots, x_n) \in A} P(X_1 = x_1, \dots, X_n = x_n), \quad (3.1)$$

(on rappelle que la somme porte sur tous les  $(x_1, \dots, x_n)$  tels que la probabilité correspondante est non-nulle, donc il s'agit d'une somme portant sur un ensemble au plus dénombrable). En particulier, on a le calcul suivant des lois marginales.

**Proposition 3.6.** Soit  $(X_1, \dots, X_n)$  vecteur aléatoire discret. Alors pour tout  $k \in \{1, \dots, n\}$ ,  $X_k$  est une variable aléatoire discrète de fonction de masse donnée par

$$P(X_k = x) = \sum_{x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n} P(X_1 = x_1, \dots, X_{k-1} = x_{k-1}, X_k = x, X_{k+1} = x_{k+1}, \dots, X_n = x_n).$$

*Démonstration.* (3.1) donne, avec  $x \in \mathbb{R}$  et  $A = \mathbb{R}^{k-1} \times \{x\} \times \mathbb{R}^{n-k}$  :

$$\begin{aligned} P(X_k = x) &= \sum_{(x_1, \dots, x_n) \in \mathbb{R}^{k-1} \times \{x\} \times \mathbb{R}^{n-k}} P(X_1 = x_1, \dots, X_n = x_n) \\ &= \sum_{x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n} P(X_1 = x_1, \dots, X_{k-1} = x_{k-1}, X_k = x, X_{k+1} = x_{k+1}, \dots, X_n = x_n). \end{aligned}$$

$\square$

### 3.1.2 Vecteur aléatoire à densité

**Définition 3.7.** Un vecteur aléatoire  $X = (X_1, \dots, X_n)$  est à *densité* s'il existe une fonction  $f_{X_1, \dots, X_n} : \mathbb{R}^n \rightarrow \mathbb{R}_+$  telle que, pour tout  $A \subset \mathbb{R}^n$ ,

$$P((X_1, \dots, X_n) \in A) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathbf{1}_A(y_1, \dots, y_n) f_{X_1, \dots, X_n}(y_1, \dots, y_n) dy_1 \dots dy_n.$$

La fonction  $f_{X_1, \dots, X_n}$  est alors appelée *la densité conjointe des v.a.  $X_1, \dots, X_n$* .

Notons que l'ordre des intégrales ci-dessus n'importe pas grâce au théorème de Fubini-Tonelli. On peut vérifier que cela définit bien une loi satisfaisant aux axiomes 1.-4.

**Proposition 3.8.** *Un vecteur aléatoire  $X = (X_1, \dots, X_n)$  est à densité ssi il existe une fonction  $f_{X_1, \dots, X_n} : \mathbb{R}^n \rightarrow \mathbb{R}_+$  telle que*

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{X_1, \dots, X_n}(y_1, \dots, y_n) dy_1 \cdots dy_n.$$

*Démonstration.* Il suffit de prendre  $A = \prod_{1 \leq i \leq n} ]-\infty, x_i]$  pour voir que la fonction de répartition prend cette forme. Réciproquement, si un vecteur aléatoire possède cette fonction de répartition, alors elle coïncide avec la fonction de répartition de  $X_1, \dots, X_n$ . Puisque la fonction de répartition détermine la loi, elle est alors bien donnée par cette formule.  $\square$

**Proposition 3.9.** *Une fonction  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  est la densité d'un vecteur aléatoire de  $\mathbb{R}^n$  si et seulement si*

- *$f$  est positive et*
- $\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \dots, x_n) dx_1 \cdots dx_n = 1.$

La proposition suivante détaille le calcul des lois marginales.

**Proposition 3.10.** *Supposons que le vecteur aléatoire  $(X_1, \dots, X_n)$  admette la densité  $f_{X_1, \dots, X_n}$ . Alors pour tout  $k \in \{1, \dots, n\}$ , la v.a.  $X_k$  admet la densité*

$$f_{X_k}(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(x_1, \dots, x_{k-1}, x, x_{k+1}, \dots, x_n) dx_1 \cdots dx_{k-1} dx_{k+1} \cdots dx_n.$$

*Démonstration.* Cas à densité. On rappelle que la fonction de répartition du vecteur  $(X_1, \dots, X_n)$  est donnée par

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{X_1, \dots, X_n}(y_1, \dots, y_n) dy_1 \cdots dy_n.$$

En faisant tendre  $x_j \rightarrow +\infty$  pour  $j \neq k$  et en changeant l'ordre d'intégration, on obtient

$$F_{X_k}(x) = \int_{-\infty}^x \left( \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{X_1, \dots, X_n}(y_1, \dots, y_{k-1}, y, y_{k+1}, \dots, y_n) dy_1 \cdots dy_{k-1} dy_{k+1} \cdots dy_n \right) dy.$$

Ceci montre que  $X_k$  est à densité avec la densité écrite dans l'énoncé.  $\square$

En revanche, contrairement à ce qui se passait dans le cas discret, (voir proposition 3.5), il *ne suffit pas* que chacune des variables  $X_1, \dots, X_n$  soit à densité pour que le vecteur  $(X_1, \dots, X_n)$  soit à densité. Par exemple, soit  $X$  une v.a. à densité; on pose  $Y = -X$ , alors  $(X, Y)$  ne définit pas un vecteur aléatoire à densité. Voici un contre-exemple : le couple  $(X, Y)$  prend ses valeurs dans la droite  $D = \{(x, -x) : x \in \mathbb{R}\}$ , or, pour toute fonction positive  $f$ , d'après le théorème de Fubini-Tonelli :

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{1}_D(x, y) f(x, y) dx dy = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} \mathbb{1}_{\{-y\}}(x) f(x, y) dx \right) dy = 0,$$

car la fonction  $\mathbb{1}_{\{-y\}}(x) f(x, y)$  est nulle sauf en un point, donc son intégrale est nulle.



### 3.1.3 Fonction d'un vecteur aléatoire

Si  $X_1, \dots, X_n$  sont des v.a. et  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $m \in \mathbb{N}^*$  une fonction, alors  $Y = \varphi(X_1, \dots, X_n)$  est encore un vecteur aléatoire, c'est-à-dire que ses coordonnées  $Y_1, \dots, Y_m$  satisfont

$$P(f(X_1, \dots, X_n) \in A) = P((X_1, \dots, X_n) \in \varphi^{-1}(A)), \quad A \subset \mathbb{R}^m.$$

#### Espérance

Si  $X = (X_1, \dots, X_n)$  est un vecteur aléatoire discret et  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  (il est important que  $\varphi$  soit à valeurs réelle désormais) est telle que  $\varphi$  est positive, ou

$$\sum_{(x_1, \dots, x_n) \in A} |\varphi(x_1, \dots, x_n)| P(X_1 = x_1, \dots, X_n = x_n) < \infty,$$

alors on définit l'espérance de  $\varphi(X)$  dans le cas discret par :

$$E[\varphi(X)] = E[\varphi(X_1, \dots, X_n)] = \sum_{(x_1, \dots, x_n) \in A} \varphi(x_1, \dots, x_n) P(X_1 = x_1, \dots, X_n = x_n),$$

De même, si  $X = (X_1, \dots, X_n)$  est un vecteur aléatoire à densité et  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  est telle que  $\varphi$  est positive, ou

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi(y_1, \dots, y_n) |f_{X_1, \dots, X_n}(y_1, \dots, y_n)| dy_1 \cdots dy_n < \infty,$$

alors :

$$E[\varphi(X)] = E[\varphi(X_1, \dots, X_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi(y_1, \dots, y_n) f_{X_1, \dots, X_n}(y_1, \dots, y_n) dy_1 \cdots dy_n.$$

Dans le cas d'un vecteur aléatoire  $X$  discret,  $Y = \varphi(X)$  est une variable aléatoire réelle discrète, et on disposait *déjà* d'une définition de l'espérance : on doit alors vérifier que les deux définitions coïncident, et la preuve est rigoureusement la même que dans le cas où  $X$  est une variable aléatoire réelle. De même, dans le cas où  $Y = \varphi(X)$  est à densité, on disposait déjà d'une définition de l'espérance pour  $Y$ , et il faudrait vérifier que les deux définitions coïncident. Finalement, dans le cas d'un vecteur aléatoire  $X$  à densité, il est possible que  $\varphi(X)$  ne soit pas à densité, mais l'expression ci-dessus fait encore sens : il s'agit alors d'une définition.

L'espérance satisfait encore les propriétés vues précédemment :

**Proposition 3.11.** (*Linéarité de l'espérance*) Si  $X = (X_1, \dots, X_n)$  est un vecteur aléatoire et  $\varphi_1, \dots, \varphi_m : \mathbb{R}^n \rightarrow \mathbb{R}$  sont des fonctions positives ou telles que  $\varphi_1(X), \dots, \varphi_m(X)$  sont intégrables, alors

$$E\left[\sum_{i=1}^m \varphi_i(X_1, \dots, X_n)\right] = \sum_{i=1}^m E[\varphi_i(X_1, \dots, X_n)].$$

On peut encore mettre le résultat précédent sous la forme plus simple suivante (en fait équivalente) :

**Corollaire 3.12.** Si les  $X_i$  sont intégrables pour tout  $i = 1, \dots, n$ , alors :

$$E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n].$$

Il est crucial de noter que cette relation ne suppose *aucune forme d'indépendance* des coordonnées du vecteur aléatoire (l'indépendance est définie dans la section à venir). Cette relation sera comparée avec profit avec la relation similaire sur la variance, à venir également, Proposition 3.23, qui elle nécessite l'indépendance. Enfin, on note que la formule de sommation permet de *définir*  $E[X_1 + \dots + X_n]$ , pour un vecteur aléatoire  $X = (X_1, \dots, X_n)$  qui n'est pas forcément discret ou à densité (mais dont les marginales appartiennent à l'une de ces deux classes).

## 3.2 Variables aléatoires indépendantes

Les liens entre variables aléatoires ne sont pas toujours aussi simples que pour le premier exemple ci-dessus des nombres des lancers pile ou face, où il y avait un lien déterministe entre les deux variables aléatoires ( $X + Y = n$ ). Par exemple, si on note encore  $X_1, \dots, X_n$  les résultats des  $n$  lancers pile ou face, alors ces variables ne semblent avoir aucun lien entre elles, car le résultat d'un lancer n'influence pas les résultats des autres lancers. C'est ce qu'on appelle *l'indépendance (stochastique)*, elle se traduit mathématiquement par la définition suivante :

**Définition 3.13.** On dit qu'un vecteur aléatoire  $X = (X_1, \dots, X_n)$  est à coordonnées indépendantes, ou simplement que les variables aléatoires  $X_1, \dots, X_n$  sont *indépendantes*, si pour tout  $A_1, \dots, A_n \subset \mathbb{R}$ ,

$$P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1) \times \dots \times P(X_n \in A_n). \quad (3.2)$$

*Remarque 3.14.* Pour comprendre l'intuition de la définition, considérons le cas  $n = 2$ ,  $X_1$  et  $X_2$  indépendantes, et calculons la probabilité conditionnelle suivante, pour  $A_2$  tel que  $P(X_2 \in A_2) \neq 0$  :

$$P(X_1 \in A_1 | X_2 \in A_2) = \frac{P(X_1 \in A_1, X_2 \in A_2)}{P(X_2 \in A_2)} = P(X_1 \in A_1).$$

Intuitivement, la donnée d'une information portant sur  $X_2$  n'influence pas la loi de  $X_1$ .

**Définition 3.15.** On dit qu'un vecteur aléatoire  $X = (X_1, \dots, X_n)$  est à coordonnées deux à deux indépendantes, ou simplement que les variables aléatoires  $X_1, \dots, X_n$  sont deux à deux *emphindépendantes*, si pour tout  $1 \leq i \neq j \leq n$ ,  $X_i$  et  $X_j$  sont indépendantes.

Des variables aléatoires  $X_1, \dots, X_n$  indépendantes sont bien sûr deux à deux indépendantes, mais la réciproque est fautive :

**Exercice 3.16.** Soit  $X_1, X_2, X_3$  trois variables aléatoires indépendantes de loi de Rademacher :  $P(X_i = \pm 1) = 1/2$ . Montrer que les variables aléatoires

$$Y_1 = X_2 X_3, Y_2 = X_1 X_3, Y_3 = X_1 X_2$$

sont deux à deux indépendantes mais pas indépendantes.

Ainsi la relation "être indépendant" n'est pas transitive :  $X_1$  peut être indépendant de  $X_2$  et  $X_2$  de  $X_3$ , sans que  $X_1$  ne soit indépendant de  $X_3$ .

**Proposition 3.17.** *Si  $X = (X_1, \dots, X_n)$  est à coordonnées indépendantes alors on a la formule suivante pour la fonction de répartition conjointe :*

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \times \dots \times F_{X_n}(x_n), \quad x_1, \dots, x_n \in \mathbb{R}. \quad (3.3)$$

*Inversement, si la fonction de répartition conjointe d'un vecteur aléatoire  $X = (X_1, \dots, X_n)$  s'écrit sous la forme (3.3), alors ses coordonnées sont indépendantes.*

*Démonstration.* En appliquant (3.2) à des ensembles de la forme  $A_i = ]-\infty, x_i]$ ,  $x_i \in \mathbb{R}$ , on obtient la formule pour la fonction de répartition conjointe. La réciproque tient au fait que la fonction de répartition conjointe détermine la loi.  $\square$

**Proposition 3.18.** *Soit  $X_1, \dots, X_n$  des variables aléatoires indépendantes (resp. deux à deux indépendantes), et  $f_1, \dots, f_n : \mathbb{R} \rightarrow \mathbb{R}$  des fonctions. Alors  $f_1(X_1), \dots, f_n(X_n)$  sont encore indépendantes (resp. deux à deux indépendantes).*

*Démonstration.* Il suffit de prouver le résultat pour l'indépendance globale. On se donne  $A_1, \dots, A_n$  des parties de  $\mathbb{R}$  et on note que

$$\{f_1(X_1) \in A_1, \dots, f_n(X_n) \in A_n\} = \{X_1 \in f_1^{-1}(A_1), \dots, X_n \in f_n^{-1}(A_n)\}$$

puis on applique la définition de l'indépendance de  $X_1, \dots, X_n$  pour calculer la probabilité de cet événement.  $\square$

On peut se demander si des variables aléatoires indépendantes  $X_1, \dots, X_n$  avec des lois marginales prescrites *existent*. C'est le cas car si on prend (3.3) comme *définition* d'une fonction de répartition conjointe. En revanche, s'il y a existence, il n'y a pas en revanche unicité en général : on dit d'un vecteur aléatoire dont les lois marginales sont prescrites qu'il est un couplage de ces lois.

**Proposition 3.19.** *Soit  $(X_1, \dots, X_n)$  un vecteur aléatoire.*

- *Si  $(X_1, \dots, X_n)$  est un vecteur aléatoire discret, alors ses coordonnées sont indépendantes si et seulement si la fonction de masse se factorise sous la forme :*

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \times \dots \times P(X_n = x_n), \quad x_1, \dots, x_n \in \mathbb{R}.$$

- *Si  $(X_1, \dots, X_n)$  est un v.a. à densité, alors ses coordonnées sont indépendantes si et seulement si le vecteur admet la densité :*

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \times \dots \times f_{X_n}(x_n), \quad x_1, \dots, x_n \in \mathbb{R}. \quad (3.4)$$

*Démonstration.* Nous montrons seulement le cas à densité, le cas discret, plus facile, est laissé en exercice. Posons

$$g(x_1, \dots, x_n) = f_{X_1}(x_1) \times \dots \times f_{X_n}(x_n), \quad x_1, \dots, x_n \in \mathbb{R}.$$

Alors  $g$  est positive et

$$\begin{aligned} & \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{X_1}(x_1) \times \dots \times f_{X_n}(x_n) dx_1 \dots dx_n \\ &= \left( \int_{-\infty}^{\infty} f_{X_1}(x_1) dx_1 \right) \times \dots \times \left( \int_{-\infty}^{\infty} f_{X_n}(x_n) dx_n \right) \\ &= 1, \end{aligned}$$

donc  $g$  est bien la densité conjointe de  $n$  variables aléatoires, notons-les  $Y_1, \dots, Y_n$ . On a pour tous  $A_1, \dots, A_n \subset \mathbb{R}$ ,

$$\begin{aligned} P(Y_1 \in A_1, \dots, Y_n \in A_n) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbb{1}_{A_1 \times \cdots \times A_n}(x_1, \dots, x_n) g(x_1, \dots, x_n) dx_1 \cdots dx_n \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbb{1}_{A_1}(x_1) f_{X_1}(x_1) \times \cdots \times \mathbb{1}_{A_n}(x_n) f_{X_n}(x_n) dx_1 \cdots dx_n \\ &= \left( \int_{-\infty}^{\infty} \mathbb{1}_{A_1}(x_1) f_{X_1}(x_1) dx_1 \right) \times \cdots \times \left( \int_{-\infty}^{\infty} \mathbb{1}_{A_n}(x_n) f_{X_n}(x_n) dx_n \right) \\ &= P(X_1 \in A_1) \times \cdots \times P(X_n \in A_n). \end{aligned}$$

Par conséquent,  $X_1, \dots, X_n$  sont indépendantes si et seulement si leur loi jointe est égale à la loi jointe de  $Y_1, \dots, Y_n$ , donc si et seulement si  $X_1, \dots, X_n$  admettent la densité conjointe  $g$ . Ceci permet de conclure.  $\square$

Au regard des formules vues précédemment sur les espérances de fonctions de vecteurs aléatoires, (3.2) peut encore s'écrire :

$$E\left[\prod_{i=1}^n \mathbb{1}_{A_i}(X_i)\right] = \prod_{i=1}^n E[\mathbb{1}_{A_i}(X_i)]$$

La proposition suivante généralise cette identité :

**Corollaire 3.20.** Soient  $X_1, \dots, X_n$  des v.a. indépendantes discrètes ou à densité et  $g_1, \dots, g_n : \mathbb{R} \rightarrow \mathbb{R}$  des fonctions telles que  $g_1(X_1), \dots, g_n(X_n)$  soient intégrables. On a alors :

$$E\left[\prod_{i=1}^n g_i(X_i)\right] = \prod_{i=1}^n E[g_i(X_i)]$$

*Démonstration.* En effet, dans le cas à densité, on a

$$\begin{aligned} E\left[\prod_{i=1}^n g_i(X_i)\right] &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^n g_i(x_i) f_{X_1}(x_1) \cdots f_{X_n}(x_n) dx_1 \cdots dx_n \\ &= \prod_{i=1}^n \left( \int_{-\infty}^{\infty} g_i(x_i) f_{X_i}(x_i) dx_i \right) \text{ de Fubini} \\ &= \prod_{i=1}^n E[g_i(X_i)]. \end{aligned}$$

Le cas discret se démontre de manière analogue.  $\square$

En particulier, en prenant  $g_i(x) = x$  pour tout  $i$ , on obtient la formule, pour  $X_1, \dots, X_n$  des variables aléatoires intégrables,

$$E[X_1 \cdots X_n] = E[X_1] \times \cdots \times E[X_n]. \quad (3.5)$$

*Remarque 3.21.* On notera donc que pour des variables aléatoires indépendantes, il suffit que chaque variable soit intégrable pour que leur produit le soit : cela ne vaut pas pour des variables qui ne sont pas indépendantes (prendre  $n = 2$  et  $X_1 = X_2$  pour s'en convaincre : on a alors inégalité stricte entre les deux membres de l'égalité (3.5) sauf si la variable aléatoire  $X_1$  est presque sûrement constante d'après la définition de la variance.)

Le critère suivant s'avère souvent très utile pour démontrer l'indépendance de v.a. :

**Proposition 3.22.** Soit  $X = (X_1, \dots, X_n) \in \mathbb{R}^n$  un vecteur aléatoire discret ou à densité. Supposons qu'il existe des fonctions  $f_1, \dots, f_n : \mathbb{R} \rightarrow \mathbb{R}_+$  telles que :

- $P(X_1 = x_1, \dots, X_n = x_n) = f_1(x_1) \times \dots \times f_n(x_n)$  (cas discret)
- $f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_1(x_1) \times \dots \times f_n(x_n)$  (cas à densité).

Alors les coordonnées  $X_1, \dots, X_n$  du vecteur aléatoire  $X$  sont indépendantes.

Les fonctions  $f_i$  de l'énoncé précédent sont en fait proportionnelles aux marginales du vecteur aléatoire; l'avantage d'appliquer cette proposition est qu'on n'a pas à calculer ces constantes; d'un point de vue théorique, on peut toujours expliciter ces constantes :

- $f_i(x_i) / \sum_x f_i(x)$  est la fonction de masse de  $X_i$ .
- $f_i(x_i) / \int_{\mathbb{R}} f_i(x_i) dx_i$  est la densité marginale de  $X_i$ .

*Démonstration.* On traite seulement le cas à densité, le cas discret est analogue. On pose

$$C_i = \int_{-\infty}^{\infty} f_i(x) dx, \quad i = 1, \dots, n.$$

Aussi, de Fubini positif,

$$\begin{aligned} C_1 \times \dots \times C_n &= \left( \int_{-\infty}^{\infty} f_1(x_1) dx_1 \right) \times \dots \times \left( \int_{-\infty}^{\infty} f_n(x_n) dx_n \right) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_1(x_1) \times \dots \times f_n(x_n) dx_1 \dots dx_n \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= 1. \end{aligned}$$

En particulier,  $0 < C_i < \infty$  pour tout  $i$ . On définit alors des fonctions  $g_i = f_i/C_i$  qui sont toutes positives et d'intégrale 1, donc des densités de v.a.. Notons  $Y_1, \dots, Y_n$  des v.a. indépendantes de densités respectives  $g_1, \dots, g_n$ . Alors, puisque  $C_1 \times \dots \times C_n = 1$ ,

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_1(x_1) \times \dots \times f_n(x_n) = g_1(x_1) \times \dots \times g_n(x_n) = f_{Y_1, \dots, Y_n}(x_1, \dots, x_n).$$

Ceci montre que la loi jointe des v.a.  $X_1, \dots, X_n$  est égale à la loi jointe de  $Y_1, \dots, Y_n$ . En particulier, les v.a.  $X_1, \dots, X_n$  sont indépendantes.  $\square$

### Variance d'une somme de v.a. indépendantes

On a vu que la variance était lié au carré de la distance entre la variable aléatoire et son espérance. La notion de couple de variables indépendantes permet de donner une nouvelle interprétation. Si  $X_1$  et  $X_2$  sont deux variables aléatoires indépendantes et identiquement distribuées, de carré intégrables, par linéarité de l'espérance puis par indépendance,

$$E[(X_1 - X_2)^2] = E[X_1^2] - 2E[X_1 X_2] + E[X_2^2] = 2(E[X_1^2] - E[X_1]^2) = 2 \text{Var}(X_1),$$

c'est à dire que la variance mesure aussi (à un facteur multiplicatif 2 près) l'espérance du carré de la distance entre deux réalisations indépendantes.

**Proposition 3.23.** Soient  $X_1, \dots, X_n$  des v.a. indépendantes tel que  $E[X_i^2] < \infty$  pour tout  $i \in \{1, \dots, n\}$  (donc  $E[|X_i|] < \infty$  pour tout  $i$ ). Alors

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n).$$

*Démonstration.* Posons  $Y_i = X_i - E[X_i]$  pour tout  $i$ . Alors  $E[Y_i] = 0$  pour tout  $i$ .

$$\begin{aligned} \text{Var}(X_1 + \dots + X_n) &= \text{Var}(Y_1 + \dots + Y_n) \\ &= E[(Y_1 + \dots + Y_n)^2] \\ &= E \left[ \sum_{i=1}^n Y_i^2 + 2 \sum_{1 \leq i < j \leq n} Y_i Y_j \right] \\ &= \sum_{i=1}^n E[Y_i^2] + 2 \sum_{1 \leq i < j \leq n} E[Y_i Y_j], \end{aligned}$$

par linéarité de l'espérance. Par indépendance,  $E[Y_i Y_j] = E[Y_i]E[Y_j] = 0$  pour tous  $i < j$ . De plus,  $E[Y_i^2] = \text{Var}(Y_i)$  pour tout  $i$ . Ceci permet de conclure.  $\square$

Notons une conséquence remarquable : de l'inégalité de Tchebychev, on a pour des variables  $(X_i)_{1 \leq i \leq n}$  indépendantes centrées, et de même variance :

$$\mathbb{P}(X_1 + \dots + X_n \geq \alpha\sqrt{n}) \leq \frac{\text{Var}(X_1)}{\alpha^2}$$

Par exemple, pour des variables i.i.d. de Rademacher de paramètre 1/2, pour lesquelles  $P(X_1 = \pm 1) = 1/2$ , on obtient :

$$\mathbb{P}(X_1 + \dots + X_n \geq \alpha\sqrt{n}) \leq \frac{1}{\alpha^2}.$$

Alors que l'ensemble des valeurs possibles de la variable aléatoire  $X_1 + \dots + X_n$  est

$$\{-n, \dots, \dots, n\},$$

elle se concentre en fait sur un sous-ensemble bien plus petit,

$$\{-\lfloor \alpha\sqrt{n} \rfloor, \dots, \lfloor \alpha\sqrt{n} \rfloor\},$$

avec probabilité  $\geq 1 - \frac{1}{\alpha^2}$ . Cet énoncé profond de la théorie des probabilité se trouvera précisé dans le chapitre consacré au théorème central limite. En dépit de sa simplicité, il n'est pas exagéré de dire que c'est l'énoncé le plus important de ce cours.

### 3.3 Fonctions de variables aléatoires indépendantes

Dans cette section, on s'intéresse à la loi de certaines fonctions particulières d'un vecteur aléatoire dont les  $n$  coordonnées sont des v.a. indépendantes  $X_1, \dots, X_n$ . On considère les cas du maximum ou minimum ainsi que de la somme.

### 3.3.1 Maximum et minimum

On s'intéresse à la loi de  $Y = \max(X_1, \dots, X_n)$  pour des variables aléatoires  $X_1, \dots, X_n$  indépendantes. L'outil le plus adapté à la caractérisation de cette loi est la fonction de répartition. En effet,

$$F_Y(x) = P(\max(X_1, \dots, X_n) \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = F_{X_1, \dots, X_n}(x, \dots, x),$$

et l'indépendance des  $X_1, \dots, X_n$  donne alors

$$F_Y(x) = F_{X_1}(x) \cdots F_{X_n}(x).$$

La loi du minimum  $Z = \min(X_1, \dots, X_n)$  s'obtient de manière similaire :

$$F_Z(x) = 1 - P(X_1 > x, \dots, X_n > x) = 1 - \prod_{i=1}^n (1 - F_{X_i}(x)).$$

Si de plus,  $X_1, \dots, X_n$  sont conjointement à densité, alors la densité du vecteur se met sous la forme  $f_{X_1}(x_1) \cdots f_{X_n}(x_n)$ , et

$$F_Y(x) = \int_{-\infty}^x \sum_{i=1}^n f_{X_i}(t) \cdot \prod_{j \neq i} F_{X_j}(t) dt,$$

donc  $Y$  est à densité, de densité :

$$f_Y(x) = \sum_{i=1}^n f_{X_i}(x) \cdot \prod_{j \neq i} F_{X_j}(x). \quad (3.6)$$

De même, et sous les mêmes hypothèses,  $Z$  est à densité, de densité :

$$f_Z(x) = \sum_{i=1}^n f_{X_i}(x) \cdot \prod_{j \neq i} (1 - F_{X_j}(x)). \quad (3.7)$$

L'interprétation des relations est comme suit : dans l'expression de la densité de  $Y$ ,  $f_{X_i}(x)$  quantifie la probabilité que  $X_i$  soit "égal" à  $x$  puis  $\prod_{j \neq i} F_{X_j}(x)$  la probabilité que les autres variables soient inférieures à cette valeur, auquel cas le maximum est égal à  $x$ , réalise par la variable  $X_i$ ; de même, dans l'expression de la densité de  $Z$ ,  $f_{X_i}(x)$  quantifie la probabilité que  $X_i$  soit "égal" à  $x$  puis  $\prod_{j \neq i} (1 - F_{X_j}(x))$  la probabilité que les autres variables soient supérieures à cette valeur, auquel cas le minimum est égal à  $x$ , réalise par la variable  $X_i$ .

**Exemple 3.24.** Soient  $X_1, \dots, X_n$  i.i.d. (indépendantes et identiquement distribuées) de loi Unif(0, 1). Alors,

$$F_{\max(X_1, \dots, X_n)}(x) = \begin{cases} 0, & x \leq 0 \\ x^n, & x \in [0, 1] \\ 1, & x \geq 1. \end{cases}$$

**Exercice 3.25.** Retrouver les expressions (3.3.1) et (3.3.1) des densités du minimum et du maximum de variables aléatoires indépendantes conjointement à densité à l'aide de la méthode de la fonction muette.

### 3.3.2 Somme

Dans ce chapitre, on s'intéresse à la loi de  $S = X_1 + \dots + X_n$  à  $n$  fixé. On présente deux méthodes pour la déterminer.

#### Première méthode : par la fonction génératrice des moments

Dans le cas précédent du maximum, il était commode de calculer la fonction de répartition puisque celle-ci s'écrivait comme le produit des fonctions de répartition respectives des v.a.  $X_1, \dots, X_n$ . Qu'en est-il pour la somme  $S = \sum_{i=1}^n X_i$ ? Existe-t-il encore une quantité fonction de  $S$  qui s'écrit comme un produit de mêmes quantités fonction des  $X_i$ ? La réponse est oui, et l'outil en question est la fonction génératrice des moments. En effet, pour tout  $t \in \mathbb{R}$ ,

$$\varphi_S(t) = E[e^{tS}] = E[e^{t(X_1 + \dots + X_n)}] = E[e^{tX_1} \dots e^{tX_n}] = E[e^{tX_1}] \dots E[e^{tX_n}] = \prod_{i=1}^n \varphi_{X_i}(t),$$

par indépendance.

**Exemple 3.26.** On suppose que  $X_i \sim \text{Po}(\lambda_i)$ ,  $\lambda_i \geq 0$ , pour tout  $i$ . On rappelle que  $\varphi_{X_i}(t) = e^{\lambda_i(e^t - 1)}$ ,  $t \in \mathbb{R}$ . Par conséquent,

$$\varphi_{X_1 + \dots + X_n}(t) = e^{\lambda_1(e^t - 1)} \dots e^{\lambda_n(e^t - 1)} = e^{(\lambda_1 + \dots + \lambda_n)(e^t - 1)},$$

si bien que  $X_1 + \dots + X_n \sim \text{Po}(\lambda_1 + \dots + \lambda_n)$ .

**Exemple 3.27.** On suppose que  $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ ,  $\mu_i \in \mathbb{R}$ ,  $\sigma_i^2 > 0$ , pour  $i = 1 \dots n$ . On rappelle, voir 2.32, que  $\varphi_{X_i}(t) = e^{t\mu_i + \sigma_i^2 t^2 / 2}$ ,  $t \in \mathbb{R}$ . Par conséquent,

$$\varphi_{X_1 + \dots + X_n}(t) = e^{t\mu_1 + \sigma_1^2 t^2 / 2} \dots e^{t\mu_n + \sigma_n^2 t^2 / 2} = e^{t(\mu_1 + \dots + \mu_n) + (\sigma_1^2 + \dots + \sigma_n^2)t^2 / 2},$$

si bien que  $X_1 + \dots + X_n \sim \mathcal{N}(\mu_1 + \dots + \mu_n, \sigma_1^2 + \dots + \sigma_n^2)$ . Noter que les valeurs des paramètres ne sont pas surprenants : Admettons que l'on sache *a priori* que  $X_1 + \dots + X_n$  suit une loi normale  $\mathcal{N}(\mu, \sigma^2)$  pour un certain  $\mu$  et un certain  $\sigma^2$  à déterminer, alors, puisque *l'on sait* que  $\mu_i$  est l'espérance de  $X_i$  et  $\sigma_i^2$  la variance de  $X_i$ , d'après les règles d'addition des espérances et des variances (dans le cas indépendant), on a que  $\mu = \mu_1 + \dots + \mu_n$  et  $\sigma^2 = \sigma_1^2 + \dots + \sigma_n^2$ .

#### Deuxième méthode : par la convolution des densités

Supposons maintenant que les v.a.  $X_1, \dots, X_n$  sont soit discrètes et à valeurs dans  $\mathbb{Z}$ , soit à densité. Au lieu de calculer la fonction génératrice des moments de leur somme, on peut aussi calculer directement la fonction de masse ou densité. Pour cela, nous introduisons la notation suivante.

**Définition 3.28.** Soient  $f, g : \mathbb{R} \rightarrow \mathbb{R}_+$ . On définit leur *convoluée*  $f \star g : \mathbb{R} \rightarrow \mathbb{R}_+ \cup \{\infty\}$  par

$$(f \star g)(x) = \int_{-\infty}^{\infty} f(t)g(x-t) dt \stackrel{t \rightarrow x-t}{=} \int_{-\infty}^{\infty} f(x-t)g(t) dt.$$

Si  $p, q : \mathbb{Z} \rightarrow \mathbb{R}_+$ , on définit leur *convoluée discrète*  $p \star_{\mathbb{Z}} q : \mathbb{Z} \rightarrow \mathbb{R}_+ \cup \{\infty\}$  par

$$(p \star_{\mathbb{Z}} q)(n) = \sum_{m \in \mathbb{Z}} p(m)q(n-m) \stackrel{m \rightarrow n-m}{=} \sum_{m \in \mathbb{Z}} p(n-m)q(m).$$

On peut vérifier (exercice) que les opérations  $\star$  et  $\star_{\mathbb{Z}}$ , appelées *convolution/convolution discrète* sont commutatives et associatives.



**Proposition 3.29.** Soient  $X_1, \dots, X_n$  des variables aléatoires indépendantes, ou bien discrètes et à valeurs dans  $\mathbb{Z}$ , ou bien à densité.

- Cas discret : Notons  $p_{X_i}(n) = P(X_i = n)$  la fonction de masse de  $X_i$ . Alors la fonction de masse de  $S = X_1 + \dots + X_n$  est égale à

$$p_S = p_{X_1} \star_{\mathbb{Z}} \dots \star_{\mathbb{Z}} p_{X_n}.$$

- Cas à densité : la v.a.  $S = X_1 + \dots + X_n$  admet la densité

$$f_S = f_{X_1} \star \dots \star f_{X_n}.$$

*Démonstration.* Il suffit de considérer le cas  $n = 2$ , le cas général se montre par récurrence. On note alors  $X = X_1$  et  $Y = X_2$ . On commence par le cas discret. Il suffit alors de décomposer :

$$\begin{aligned} p_S(n) &= P(X + Y = n) \\ &= \sum_{m \in \mathbb{Z}} P(X = m, X + Y = n) \\ &= \sum_{m \in \mathbb{Z}} P(X = m, Y = n - m) \\ &= \sum_{m \in \mathbb{Z}} P(X = m)P(Y = n - m) && \text{par indépendance} \\ &= (p_X \star_{\mathbb{Z}} p_Y)(n) \end{aligned}$$

Dans le cas à densité, c'est moins simple et on calcule la fonction de répartition de  $S$  :

$$\begin{aligned} P(S \leq z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{1}_{x+y \leq z} f_X(x) f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{1}_{x \leq z} f_X(x - y) f_Y(y) dx dy && (x \mapsto x - y) \\ &= \int_{-\infty}^z \left( \int_{-\infty}^{\infty} f_X(x - y) f_Y(y) dy \right) dx \\ &= \int_{-\infty}^z (f_X \star f_Y)(x) dx. \end{aligned}$$

Ceci montre que  $S$  admet la densité  $f_S = f_X \star f_Y$  (par définition de la densité d'une variable aléatoire réelle).  $\square$

**Exemple 3.30.** Soient  $X, Y \sim \text{Unif}(0, 1)$  indépendantes, donc de densité  $f = \mathbf{1}_{[0,1]}$ . Alors  $X + Y$  admet la densité

$$\begin{aligned} f_{X+Y}(x) &= (f \star f)(x) = \int_{-\infty}^{\infty} \mathbf{1}_{[0,1]}(x - t) \mathbf{1}_{[0,1]}(t) dt = \int_0^1 \mathbf{1}_{x-1 \leq t \leq x} dt \\ &= x \mathbf{1}_{[0,1]}(x) + (2 - x) \mathbf{1}_{[1,2]}(x). \end{aligned}$$

**Exemple 3.31.** Soient  $\alpha_1, \alpha_2, \beta > 0$ , et  $X \sim \Gamma(\alpha_1, \beta)$  et  $Y \sim \Gamma(\alpha_2, \beta)$  indépendantes. Soit  $y > 0$ . Alors  $X + Y$  admet la densité suivante en  $y > 0$  :

$$\begin{aligned}
 f_{X+Y}(y) &= (f_X \star f_Y)(y) \\
 &= \int_{\mathbb{R}} f_X(x) f_Y(y-x) dx \\
 &= \int_0^y \frac{\beta^{\alpha_1}}{\Gamma(\alpha_1)} x^{\alpha_1-1} e^{-\beta x} \frac{\beta^{\alpha_2}}{\Gamma(\alpha_2)} (y-x)^{\alpha_2-1} e^{-\beta(y-x)} dx \\
 &= \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^y x^{\alpha_1-1} (y-x)^{\alpha_2-1} e^{-\beta y} dx \\
 &= \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y^{\alpha_1+\alpha_2-1} e^{-\beta y} \int_0^y \left(\frac{x}{y}\right)^{\alpha_1-1} \left(1-\frac{x}{y}\right)^{\alpha_2-1} \frac{dx}{y} \\
 &= \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y^{\alpha_1+\alpha_2-1} e^{-\beta y} \int_0^1 z^{\alpha_1-1} (1-z)^{\alpha_2-1} dz
 \end{aligned}$$

Maintenant deux densités de probabilité proportionnelles sont nécessairement égales, on en tire que

$$f_{\alpha_1, \beta} \star f_{\alpha_2, \beta}(y) = \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1 + \alpha_2)} y^{\alpha_1+\alpha_2-1} e^{-\beta y} \mathbf{1}_{y>0} = f_{\alpha_1+\alpha_2, \beta}(y)$$

avec en prime l'expression de l'intégrale suivante en terme de fonction  $\Gamma$  :

$$\int_0^1 z^{\alpha_1-1} (1-z)^{\alpha_2-1} dz = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}$$

De ce résultat on tire (par une récurrence finie) que la somme de  $n$  variables i.i.d de loi  $\text{Exp}(\beta)$  suit une loi  $\Gamma(n, \beta)$ .

**Exercice 3.32.** Retrouver ce résultat à l'aide de la méthode portant sur la fonction génératrice des moments.

### Comparaison des deux méthodes

Chacune des deux méthodes a ses avantages et ses inconvénients. La méthode par la fonction génératrice des moments est souvent plus simple et rapide, car une simple multiplication est plus facile à calculer qu'une convolution. En revanche, même si on réussit à calculer la fonction génératrice des moments de  $S$ , rien ne dit qu'on sera capable d'inverse cette fonction génératrice, c'est à dire de trouver la loi dont cette fonction est la fonction génératrice. Cette méthode peut donc ne pas aboutir, auquel cas on pourra passer à la méthode de la convolution des densités.

Dans le cas particulier de la loi Binomiale, on peut aussi se dispenser de convolution et calculer directement la loi de la somme.

**Exercice 3.33.** Soit  $n \geq 1$  entier, et  $(B_i)_{1 \leq i \leq n}$  une suite de variables aléatoires i.i.d. selon la loi  $\text{Ber}(p)$ , pour  $p \in [0, 1]$ . On s'intéresse à la loi de  $S_n = \sum_{1 \leq i \leq n} B_i$ .

1. Soit  $x \in \{0, 1\}$ . Observer que  $\mathbb{P}(B_1 = x) = p^x (1-p)^{1-x}$ .

2. En déduire que si  $(x_i)_{1 \leq i \leq n} \in \{0, 1\}^n$ ,  $\mathbb{P}(\bigcap_{i \leq n} \{B_i = x_i\}) = p^{\sum x_i} (1-p)^{n-\sum x_i}$ .
3. Quel est le cardinal de l'ensemble  $\{(x_i)_{1 \leq i \leq n} \in \{0, 1\}^n : \sum_{1 \leq i \leq n} x_i = k\}$
4. En déduire que, pour  $0 \leq k \leq n$ ,  $\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$
5. Calculer la fonction génératrice des moments de  $S_n$  et de la loi  $\text{Bin}(n, p)$  et retrouver ce résultat.

**Exercice 3.34.** Soit  $n \geq 1$  entier, et  $(B_i)_{i \geq 1}$  une suite de variables aléatoires i.i.d. selon la loi  $\text{Ber}(p)$ , pour  $p \in ]0, 1[$ .

1. On pose  $L = \min\{n \geq 1, B_n = 1\} \in \mathbb{N} \cup \{\infty\}$ . Quelle est la loi de  $L$  ?
2. On pose  $N_r = \min\{n \geq 1, \sum_{i=1}^n B_i = r\}$ . Observer que

$$\{N = n\} = \left\{ \sum_{i=1}^{n-1} B_i = r-1, B_n = 1 \right\}.$$

Quelle est la loi de  $N_r$  ?

3. Montrer que  $N_r$  est distribué comme  $L_1 + \dots + L_r$  avec les  $L_i$  i.i.d. selon la loi de  $L$ .

### 3.3.3 Cas général

Soient  $X_1, \dots, X_n$  des v.a. conjointement à densité<sup>1</sup> et  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  une fonction. On s'intéresse à la loi du vecteur aléatoire  $\varphi(X_1, \dots, X_n)$ . Sous des conditions convenables sur la fonction  $\varphi$ , ce vecteur aléatoire admettra encore une densité jointe qu'on peut exprimer en fonction de la densité conjointe des v.a.  $X_1, \dots, X_n$ . Pour cela, nous rappelons d'abord le changement de variable multidimensionnel.

#### Changement de variable dans $\mathbb{R}^n$

Si  $\varphi(x) = (\varphi_1(x), \dots, \varphi_n(x))$  pour  $x \in \mathbb{R}^n$ , on définit la *matrice jacobienne* de la fonction  $\varphi$  par

$$\mathcal{J}_\varphi : x \mapsto \begin{pmatrix} \frac{\partial \varphi_1}{\partial x_1}(x) & \cdots & \frac{\partial \varphi_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial \varphi_n}{\partial x_1}(x) & \cdots & \frac{\partial \varphi_n}{\partial x_n}(x) \end{pmatrix},$$

partout où toutes les dérivées partielles sont bien définies. Notons que la matrice jacobienne est une *fonction* à valeurs dans les matrices.

Le déterminant de la matrice jacobienne joue un rôle important, il est appelé *le jacobien* :

$$\begin{cases} \mathbb{R}^n \rightarrow \mathbb{R} \\ x \mapsto J_\varphi(x) = \det \mathcal{J}_\varphi(x) \end{cases}$$

La matrice jacobienne et le jacobien étant des notions locales, ils sont également définis pour des fonctions sur un ouvert  $O \subset \mathbb{R}^n$ . On dira qu'une fonction  $\varphi$  est *continument dérivable* sur  $O$ , et on notera  $\varphi \in C^1(O)$ , si la matrice jacobienne  $\mathcal{J}_\varphi(x)$  est définie et continue sur  $O$ .

Nous avons besoin de savoir comment le volume  $d$ -dimensionnel d'un petit parallélépipède  $I_x$  de centre  $x$  se transforme par l'application  $\varphi$ . La réponse est donnée par le jacobien :

$$\text{vol}(\varphi(I_x)) \approx |J_\varphi(x)| \times \text{vol}(I_x).$$

---

1. que nous ne supposons pas forcément indépendantes même si ce sera le cas dans la plupart des exemples

**Theorème 3.35** (Changement de variables dans  $\mathbb{R}^n$ ). Soit  $\varphi$  une bijection continûment différentiable ainsi que son inverse  $\varphi^{-1}$  d'un ouvert  $O \subset \mathbb{R}^n$  sur un ouvert  $O'$  de  $\mathbb{R}^n$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  bornée et  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  intégrable. Alors :

$$\int_O g(\varphi(x))f(x)dx = \int_{O'} g(y)f(\varphi^{-1}(y)) |J_{\varphi^{-1}}(y)| dy, \quad (3.8)$$

avec de plus :

$$J_{\varphi^{-1}}(y) = \frac{1}{J_{\varphi}(\varphi^{-1}(y))} \quad (3.9)$$

*Remarque 3.36* (Au sujet de la valeur absolue du jacobien). En dimension  $d = 1$ , lorsque  $O = ]a, b[$  avec  $a < b$  et  $\varphi$  est strictement décroissante alors  $O' = ]\varphi(b), \varphi(a)[$  et  $\varphi^{-1}$  est décroissante si bien que  $|J_{\varphi^{-1}}(y)| = -(\varphi^{-1})'(y)$ . Ainsi le second membre de (3.8) s'écrit :

$$\int_{\varphi(b)}^{\varphi(a)} g(y)f(\varphi^{-1}(y))(-\varphi^{-1})'(y)dy = \int_{\varphi(a)}^{\varphi(b)} g(y)f(\varphi^{-1}(y))(\varphi^{-1})'(y)dy$$

et on retrouve bien le changement de variables usuel dans  $\mathbb{R}$ .

*Remarque 3.37*. La relation (3.9) entre  $J_{\varphi^{-1}}$  et  $J_{\varphi}$  est l'analogie multidimensionnel de la relation  $(\varphi^{-1})'(y) = 1/(\varphi' \circ \varphi^{-1})(y)$ ; elle permet quelquefois de se dispenser du calcul du jacobien de la fonction réciproque, voire de la fonction réciproque elle-même.

### Application : fonction d'un vecteur à densité

Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire à densité à valeurs dans un ouvert  $O \subset \mathbb{R}^n$ , de densité conjointe  $f_{X_1, \dots, X_n}$ . On suppose que  $\varphi : O \rightarrow O'$  est une bijection continûment différentiable ainsi que son inverse  $\varphi^{-1}$ . On s'intéresse au vecteur aléatoire

$$Y = \varphi(X) = \varphi(X_1, \dots, X_n).$$

Pour caractériser la loi de ce vecteur, on combine la méthode de la fonction muette avec le changement de variables multidimensionnel : dans la formule (3.8) ci-dessus, on prends pour  $g$  une fonction bornée qui est la fonction "test" ou fonction "muette", et pour  $f$  la densité du vecteur aléatoire  $X$ , qui est donc intégrable. Le changement de variable multidimensionnel donne :

$$\begin{aligned} E[g(\varphi(X))] &= \int_O g(\varphi(x))f(x)dx \\ &= \int_{O'} g(y)f(\varphi^{-1}(y)) |J_{\varphi^{-1}}(y)| dy \\ &= \int_{\mathbb{R}^d} g(y)f(\varphi^{-1}(y)) |J_{\varphi^{-1}}(y)| \mathbf{1}_{O'}(y) dy, \end{aligned}$$

c'est-à-dire que  $\varphi(X)$  est à densité, de densité  $f(\varphi^{-1}(y))|J_{\varphi^{-1}}(y)|\mathbf{1}_{O'}(y)$ . Dans le cas où  $\varphi$  n'est pas bijective sur  $O$  entier, il faut partitionner  $O$  en "bouts" sur lesquels  $\varphi$  est bijective. La meilleure façon de comprendre ce résultat est par la pratique. Nous développons dans la suite de nombreux exemples.

**Exemple 3.38.** Soit  $X, Y$  indépendantes de loi  $\text{Unif}(0, 1)$ . Soit  $\varphi : O = ]0, 1[^2 \rightarrow \mathbb{R}^2$  donnée par

$$(z, w) = \varphi(x, y) = (x + y, x - y).$$

Déterminons  $O' = \varphi(O)$ .  $(x, y) = \varphi^{-1}(z, w) = (\frac{z+w}{2}, \frac{z-w}{2})$  donne

$$(z, w) \in O' \text{ ssi } (\frac{z+w}{2}, \frac{z-w}{2}) \in O \text{ ssi } (z+w \in ]0, 2[ \text{ et } z-w \in ]0, 2[).$$

En d'autres termes, l'ensemble  $O'$  est le carré d'extrémités  $(0, 0)$ ,  $(1, 1)$ ,  $(2, 0)$  et  $(1, -1)$  (notons que son aire vaut 2).  $\varphi$  est une bijection continûment différentiable, ainsi que son inverse, de  $O$  sur  $O'$ , et  $J_{\varphi^{-1}}(z, w) = 1/2$  pour tout  $(z, w) \in O'$ . On peut donc calculer, pour  $g$  bornée,

$$\begin{aligned} E[g(Z, W)] &= E[g(X + Y, X - Y)] \\ &= \int_O g(x + y, x - y) dx dy \\ &= \int_{O'} g(z, w) \frac{dz dw}{2} \\ &= \int_{O'} g(z, w) \frac{\mathbf{1}_{O'}(z, w)}{2} dz dw \end{aligned}$$

Ainsi  $(Z, W)$  possède la densité uniforme sur  $O'$ ; ses marginales sont données par :

$$\begin{aligned} f_Z(z) &= \int \mathbf{1}_{\{-z < w < 2-z, z-2 < w < z\}} dw / 2 \\ &= \mathbf{1}_{\{0 < z < 1\}} \int \mathbf{1}_{\{-z < w < z\}} dw / 2 + \mathbf{1}_{\{1 \leq z < 2\}} \int \mathbf{1}_{\{z-2 < w < 2-z\}} dw / 2 \\ &= \mathbf{1}_{\{0 < z < 2\}} (z \wedge (2 - z)) \end{aligned}$$

(on retrouve bien la densité calculée en 3.30) et

$$\begin{aligned} f_W(w) &= \int \mathbf{1}_{\{-w < z < 2-w, w < z < 2+w\}} dz / 2 \\ &= \mathbf{1}_{\{-1 < w < 0\}} \int \mathbf{1}_{\{-w < z < 2+w\}} dz / 2 + \mathbf{1}_{\{0 \leq w < 1\}} \int \mathbf{1}_{\{w < z < 2-w\}} dz / 2 \\ &= \mathbf{1}_{\{-1 < w < 1\}} \frac{1 - |w|}{2} \end{aligned}$$

Enfin,  $Z$  et  $W$  ne sont pas indépendantes car  $(Z, W)$  n'admet *pas* la densité produit  $f_Z(z)f_W(w)$ .

**Exemple 3.39.** Soit  $X, Y$  indépendantes de loi  $\text{Exp}(\lambda)$ . On reprend la même fonction  $\varphi$  mais cette fois-ci définie sur  $O = ]0, +\infty[^2 \rightarrow \mathbb{R}^2$ . Alors

$$(z, w) \in O' = \varphi(O) \text{ ssi } (z + w > 0 \text{ et } z - w > 0) \text{ ssi } z > |w|.$$

Ainsi  $O' = \{(z, w) \in \mathbb{R}^2, |w| < z\}$ . On a encore que  $\varphi$  est une bijection continûment différentiable ainsi que son inverse, et, pour  $g$  bornée :

$$\begin{aligned} E[g(Z, W)] &= E[g(X + Y, X - Y)] \\ &= \int_O g(x + y, x - y) \lambda^2 e^{-\lambda(x+y)} dx dy \\ &= \int_{O'} g(z, w) \lambda^2 e^{-\lambda z} \frac{dz dw}{2} \\ &= \int_{\mathbb{R}^2} g(z, w) \lambda^2 e^{-\lambda z} \mathbf{1}_{|w| < z} \frac{dz dw}{2} \end{aligned}$$

donc le couple  $(Z, W)$  admet la densité jointe  $f_{Z,W}(z, w) = \frac{\lambda^2}{2} e^{-\lambda z} \mathbf{1}_{|w| < z}$  et on peut calculer les densités marginales :

$$\begin{aligned} f_Z(Z) &= \lambda^2/2 e^{-\lambda z} \left( \int_{-z}^z dw \right) \\ &= \lambda^2 z e^{-\lambda z} \mathbf{1}_{z>0}, \end{aligned}$$

on reconnaît donc que  $Z$  suit une loi  $\Gamma(2, \lambda)$ , ainsi que :

$$\begin{aligned} f_W(w) &= \lambda/2 \int_{-z}^z \mathbf{1}_{z>|w|} \lambda e^{-\lambda z} dz \\ &= \frac{\lambda}{2} e^{-\lambda|w|} \end{aligned}$$

encore appelée loi de Laplace de paramètre  $\lambda$ , et qui peut être réalisée comme la valeur absolue d'une variable aléatoire de loi  $\text{Exp}(\lambda)$ .

**Exemple 3.40.** Enfin, soit  $X, Y$  indépendantes de loi normale centrée réduite,  $\mathcal{N}(0, 1)$ . On considère toujours la même fonction  $\varphi$  mais cette fois-ci définie sur  $O = \mathbb{R}^2$  entier ;  $\varphi$  définit une bijection sur  $O' = \mathbb{R}^2$  cette fois. On obtient donc, pour  $g$  bornée, que

$$\begin{aligned} E[g(Z, W)] &= E[g(X + Y, X - Y)] \\ &= \int_O g(x + y, x - y) \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dx dy \\ &= \int_{O'} g(z, w) \frac{1}{2\pi} e^{-\frac{((z+w)/2)^2 + ((z-w)/2)^2}{2}} \frac{dz dw}{2} \\ &= \int_{\mathbb{R}^2} g(z, w) \frac{1}{4\pi} e^{-\frac{z^2+w^2}{4}} dz dw \end{aligned}$$

Ainsi  $(Z, W)$  admet la densité  $f_{Z,W}(z, w) = \frac{1}{4\pi} e^{-\frac{z^2+w^2}{4}}$  de lois marginales  $f_Z(z) = f_W(z) = \frac{1}{\sqrt{4\pi}} e^{-\frac{z^2}{4}}$  et donc la relation  $f_{Z,W}(z, w) = f_Z(z) f_W(w)$  implique (cette fois-ci !) que  $Z$  et  $W$  sont indépendantes de loi  $\mathcal{N}(0, 2)$ .

**Exemple 3.41.** Considérons la somme de variables aléatoires  $X_1, \dots, X_n$  i.i.d. selon la loi exponentielle de paramètre  $\beta$  (on pourra faire le lien avec la somme de v.a. de loi Gamma documenté dans l'exemple 3.31). Puisque  $\beta e^{-\beta x} \mathbf{1}_{x>0}$  est la densité de la loi exponentielle, le vecteur aléatoire  $(X_1, \dots, X_n)$  possède la densité

$$\beta^n e^{-\beta(x_1 + \dots + x_n)} \mathbf{1}_{x_1>0, \dots, x_n>0}.$$

On pose  $\varphi(x_1, \dots, x_n) = (y_1, \dots, y_n)$  avec  $y_i = \sum_{j=1}^i x_j$  ;  $\varphi$  est une bijection de  $O = ]0, +\infty[^n$  sur  $O' = \{0 < y_1 < \dots < y_n\}$ , elle est continûment différentiable ainsi que son inverse ; pour cette transformation,  $J_\varphi$  vaut identiquement 1 sur  $O$  et donc, de la formule (3.9), pour tout  $y \in O'$ ,  $J_{\varphi^{-1}}(y) = 1$  (sans même avoir besoin d'explicitier l'inverse  $\varphi^{-1}$ ). Ainsi, pour  $g : \mathbb{R} \rightarrow \mathbb{R}$

bornée, on obtient

$$\begin{aligned}
E[g(X_1 + \dots + X_n)] &= \int_O g(x_1 + \dots + x_n) \beta^n e^{-\beta(x_1 + \dots + x_n)} \mathbb{1}_{x_1 > 0, \dots, x_n > 0} dx_1 \dots dx_n \\
&= \int_{O'} g(y_n) \beta^n e^{-\beta y_n} \mathbb{1}_{0 < y_1 < y_2 < \dots < y_n} dy_1 \dots dy_n \\
&= \int_{\mathbb{R}} g(y_n) \beta^n e^{-\beta y_n} \left( \int_{\mathbb{R}^{n-1}} \mathbb{1}_{0 < y_1 < y_2 < \dots < y_n} dy_1 \dots dy_{n-1} \right) dy_n \\
&= \int_{\mathbb{R}} g(y_n) \frac{\beta^n}{(n-1)!} (y_n)^{n-1} e^{-\beta y_n} \mathbb{1}_{y_n > 0} dy_n
\end{aligned}$$

ayant utilisé que

$$\begin{aligned}
\int_{\mathbb{R}^{n-1}} \mathbb{1}_{0 < y_1 < y_2 < \dots < y_n} dy_1 \dots dy_{n-1} &= \int_{\mathbb{R}^{n-2}} y_2 \mathbb{1}_{0 < y_2 < \dots < y_n} dy_2 \dots dy_{n-1} \\
&= \int_{\mathbb{R}^{n-3}} \frac{y_3}{2} \mathbb{1}_{0 < y_3 < \dots < y_n} dy_3 \dots dy_{n-1} \\
&= \frac{(y_n)^{n-1}}{(n-1)!}
\end{aligned}$$

Ainsi  $X_1 + \dots + X_n$  suit une loi  $\Gamma(n, \beta)$

Les exemples suivants sont plus avancés et sont donnés à destination du lecteur curieux.

**Exemple 3.42.** Soient  $X_1, \dots, X_n$  des v.a. i.i.d. de loi  $\text{Unif}(0, 1)$  et soit  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  la fonction

$$\varphi(x_1, \dots, x_n) = x_{(1)}, \dots, x_{(n)},$$

où  $x_{(1)}, \dots, x_{(n)}$  sont les valeurs  $x_1, \dots, x_n$  rangées par ordre croissant. On s'intéresse au vecteur aléatoire  $Y = (X_{(1)}, \dots, X_{(n)}) = \varphi(X_1, \dots, X_n)$ . La fonction  $\varphi$  n'est évidemment pas injective, ni de classe  $C^1$ , mais on peut se restreindre à l'ouvert

$$O = \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_i \neq x_j, i \neq j\}.$$

Alors le vecteur aléatoire  $(X_1, \dots, X_n)$  est p.s. à valeurs dans  $O$ , car  $P(X_i = X_j) = 0$  pour tout  $i \neq j$ , par continuité de la loi  $\text{Unif}(0, 1)$ .

La matrice jacobienne : Soit  $x = (x_1, \dots, x_n) \in O$  et pour  $i \in \{1, \dots, n\}$ , notons  $\pi(i)$  l'indice tel que  $x_i = x_{(\pi(i))}$ , soit le rang de  $x_i$  dans l'arrangement croissant. Alors pour tout  $i$  et pour  $\varepsilon > 0$  assez petit,

$$\varphi(x_1, \dots, x_{i-1}, x_i + \varepsilon, x_{i+1}, \dots, x_n) = (x_{\pi(1)}, \dots, x_{\pi(i-1)}, x_{\pi(i)} + \varepsilon, x_{\pi(i+1)}, \dots, x_{\pi(n)}),$$

puisque l'ordre des  $x_1, \dots, x_n$  ne change pas quand on perturbe un peu les valeurs. Par conséquent, la matrice jacobienne est donnée par

$$\mathcal{J}_\varphi(x_1, \dots, x_n) = (\mathbb{1}_{j=\pi(i)})_{i,j=1,\dots,n}.$$

En particulier, le jacobien est donné par le signe  $\sigma(\pi)$  de la permutation  $\pi$ .

$$J_\varphi(x_1, \dots, x_n) = \sigma(\pi) \in \{-1, 1\},$$

Il reste à déterminer la préimage de  $\varphi$ . Clairement, l'image de  $O$  par  $\varphi$  est incluse dans l'ensemble

$$E = \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_1 < \dots < x_n\}.$$

Inversement, étant donné  $x \in E$ , sa préimage  $\varphi^{-1}(x)$  consiste en tous les points dans  $D$  obtenus en permutant les coordonnées de  $x$ . Il existe exactement  $n!$  telles permutations et chacune donne un point différent, si bien que  $\text{Card}(\varphi^{-1}(x)) = n!$ . Puisque  $f_{X_1, \dots, X_n} = \mathbb{1}_{[0,1]^n}$  on obtient pour tout  $x \in E$ ,

$$f_{X_{(1)}, \dots, X_{(n)}}(x) = n! \times \mathbb{1}_{S(x)}, \quad \text{avec } S = \{(x_1, \dots, x_n) \in \mathbb{R}^n : 0 \leq x_1 < \dots < x_n \leq 1\}.$$

Autrement dit, le vecteur aléatoire  $(X_{(1)}, \dots, X_{(n)})$  est de loi uniforme dans l'ensemble  $S$ .

**Exemple 3.43.** Si  $X$  et  $Y$  sont deux VARs indépendantes de lois respectives  $\Gamma(\alpha_1, \beta)$  et  $\Gamma(\alpha_2, \beta)$  (notons qu'on choisit le même second paramètre  $\beta$ ), alors

$$(S, R) = \left( X + Y, \frac{X}{X + Y} \right)$$

sont deux VARs indépendantes également, de lois respectives  $\Gamma(\alpha_1 + \alpha_2, \beta)$  et  $\Gamma(\alpha_2, \beta)$ . On mène alors le calcul suivant, qu'on mène par deux changements de variable affines successifs :

$$\begin{aligned} & \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\beta^{\alpha_1+\alpha_2}} E \left[ \varphi \left( X + Y, \frac{X}{X + Y} \right) \right] \\ &= \int_{\mathbb{R}^2} \varphi \left( x + y, \frac{x}{x + y} \right) x^{\alpha_1-1} y^{\alpha_2-1} e^{-\beta(x+y)} \mathbb{1}_{x>0, y>0} dx dy \\ &= \int_{\mathbb{R}^2} \varphi \left( x + y, \frac{x}{x + y} \right) (x + y)^{\alpha_1+\alpha_2-2} \left( \frac{x}{x + y} \right)^{\alpha_1-1} \left( \frac{y}{x + y} \right)^{\alpha_2-1} e^{-\beta(x+y)} \mathbb{1}_{x>0, y>0} dx dy \\ &= \int_0^\infty \left( \int_0^\infty \varphi \left( x + y, \frac{x}{x + y} \right) (x + y)^{\alpha_1+\alpha_2-2} \left( \frac{x}{x + y} \right)^{\alpha_1-1} \left( \frac{y}{x + y} \right)^{\alpha_2-1} e^{-\beta(x+y)} dy \right) dx \\ &= \int_0^\infty \left( \int_0^\infty \varphi \left( s, \frac{x}{s} \right) s^{\alpha_1+\alpha_2-2} \left( \frac{x}{s} \right)^{\alpha_1-1} \left( 1 - \frac{x}{s} \right)^{\alpha_2-1} e^{-\beta s} ds \right) dx \\ &= \int_0^\infty \varphi(s, r) \left( s^{\alpha_1+\alpha_2-1} e^{-\beta s} \mathbb{1}_{s>0} \right) \left( r^{\alpha_1-1} (1-r)^{\alpha_2-1} \mathbb{1}_{r>0} \right) dr ds \end{aligned}$$

Ainsi la densité du couple  $(S, R)$  est la densité produit :

$$\frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} s^{\alpha_1+\alpha_2-1} e^{-\beta s} \mathbb{1}_{s>0} r^{\alpha_1-1} (1-r)^{\alpha_2-1} \mathbb{1}_{r>0}$$

La première marginale est donc la loi  $\Gamma(\alpha_1 + \alpha_2, \beta)$  et la seconde la loi  $\beta(\alpha_1, \alpha_2)$  et on en tire également par identification, le lien déjà établi entre fonctions  $\Gamma$  et  $\beta$  :

$$\frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} = \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1 + \alpha_2)} \frac{1}{\beta(a, b)} \quad \text{d'où } \beta(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}.$$

**Exemple 3.44.** Soit  $\varphi(x) = Ax + b$  avec  $A$  une matrice carrée et  $b \in \mathbb{R}^n$ . Alors

$$\mathcal{J}_\varphi \equiv A,$$



en particulier,  $J_\varphi \equiv \det A$ . Pour appliquer les théorèmes, il faut alors que  $\det A \neq 0$ , c'est-à-dire que  $A$  soit inversible. La fonction  $\varphi$  est alors bijective d'inverse

$$\varphi^{-1}(x) = A^{-1}(x - b).$$

On obtient donc :

$$f_{AX+b}(x) = |\det A^{-1}| \times f_X(A^{-1}(x - b)).$$

**Exemple 3.45.** Soient  $X_1, \dots, X_n$  des v.a. i.i.d. de loi  $\mathcal{N}(0, 1)$ . On appelle le vecteur  $X = (X_1, \dots, X_n)$  un *vecteur gaussien standard* (*n-dimensionnel*). Sa densité conjointe est donnée par

$$f_X(x_1, \dots, x_n) = \prod_{i=1}^n \frac{e^{-x_i^2/2}}{\sqrt{2\pi}} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_{i=1}^n x_i^2},$$

ce qu'on peut encore écrire à l'aide de la norme euclidienne  $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$  de  $x = (x_1, \dots, x_n)$  :

$$f_X(x) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{\|x\|_2^2}{2}}.$$

Si maintenant  $A$  est une matrice de dimension  $n \times n$  et  $\mu \in \mathbb{R}^n$ , alors on dit que  $AX + \mu$  est un *vecteur gaussien de moyenne  $\mu$  et de matrice de covariance  $\Sigma = AA^t$*  et on note  $AX + \mu \sim \mathcal{N}(\mu, \Sigma)$ . Dans le cas où  $A$  est inversible, par l'exemple précédent,  $AX + \mu$  admet la densité conjointe

$$\begin{aligned} f_{AX+\mu}(x) &= |\det A^{-1}| \times \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \|A^{-1}(x-\mu)\|_2^2} \\ &= \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \times e^{-\frac{1}{2} \langle \Sigma^{-1}(x-\mu), (x-\mu) \rangle}, \end{aligned}$$

où la seconde identité vient du fait que  $\det \Sigma = \det A \times \det A^t = (\det A)^2$  et pour tout  $y \in \mathbb{R}^n$ ,

$$\|A^{-1}y\|_2^2 = \langle A^{-1}y, A^{-1}y \rangle = \langle (A^{-1})^t A^{-1}y, y \rangle = \langle (AA^t)^{-1}y, y \rangle = \langle \Sigma^{-1}y, y \rangle.$$

En particulier, la loi de  $AX + \mu$  ne dépend que de  $\mu$  et  $\Sigma$  : si  $B$  est une autre matrice telle que  $BB^t = \Sigma$ , alors  $AX + \mu$  et  $BX + \mu$  ont même loi et on peut montrer que cela reste vrai même si  $A$  n'est pas inversible (ce qui justifie la notation  $AX + \mu \sim \mathcal{N}(\mu, \Sigma)$  qui ne fait pas mention de  $A$ ).

**Exemple 3.46.** Pour un couple  $(x, y) \in \mathbb{R}^2$ , notons  $(r, \theta)$  ses coordonnées polaires, i.e.  $(x, y) = (r \cos \theta, r \sin \theta)$ . Ceci induit un difféomorphisme, avec  $D := \mathbb{R}^2 \setminus (]-\infty, 0] \times \{0\})$ ,

$$\varphi : D \rightarrow \mathbb{R}_+^* \times ]-\pi, \pi[, \quad (x, y) \mapsto (r, \theta).$$

Son inverse est donné par

$$\varphi^{-1}(r, \theta) = (r \cos \theta, r \sin \theta),$$

dont le jacobien est donné par

$$J_{\varphi^{-1}}(r, \theta) = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r(\cos^2 \theta + \sin^2 \theta) = r.$$

On applique la fonction  $\varphi$  à un vecteur gaussien standard  $(X, Y)$ , i.e.  $X, Y \sim \mathcal{N}(0, 1)$  et sont indépendantes. Celui-ci est à valeurs dans  $D$ , car  $P((X, Y) \notin D) \leq P(Y = 0) = 0$ . Notons  $(R, \Theta) = \varphi(X, Y)$ , i.e.  $(X, Y) = (R \cos \Theta, R \sin \Theta)$ . On obtient donc :

$$\begin{aligned} f_{R, \Theta}(r, \theta) &= |J_{\varphi^{-1}}(r, \theta)| \times f_{X, Y}(\varphi^{-1}(r, \theta)) \\ &= r \frac{1}{2\pi} e^{-r^2(\cos^2 \theta + \sin^2 \theta)} \\ &= r e^{-r^2/2} \times \frac{1}{2\pi}. \end{aligned}$$

Ceci est le produit des deux fonctions  $f_R(r) = r e^{-r^2/2}$  et  $f_\Theta(\theta) = \frac{1}{2\pi}$ . Par conséquent, les v.a.  $R$  et  $\Theta$  sont indépendantes et de densités proportionnelles à  $f_R$  et  $f_\Theta$ , respectivement. Mais puisque  $f_\Theta$  est d'intégrale 1,  $f_R$  doit l'être aussi et donc  $f_R$  et  $f_\Theta$  sont en fait les densités. On résume :  *$R$  et  $\Theta$  sont des v.a. indépendantes, avec  $\Theta$  de loi Unif $(-\pi, \pi)$  et  $R$  de loi de densité*

$$f_R(r) = r e^{-r^2/2} \mathbf{1}_{r>0}.$$

## Chapitre 4

# Suites de variables aléatoires et théorèmes limites

Dans ce chapitre, nous allons énoncer deux théorèmes concernant le comportement limite quand  $n \rightarrow \infty$  de la somme  $X_1 + \dots + X_n$ , où  $X_1, X_2, \dots$  sont des variables i.i.d. (indépendantes et identiquement distribuées). Pour cela, nous allons d'abord introduire des notions de convergence pour des suites de variables aléatoires.

### 4.1 Convergence de suites de variables aléatoires

Dans cette section,  $X_1, X_2, \dots$  désigne une suite de variables aléatoires, c'est-à-dire une famille de v.a. indexée par  $\mathbb{N}^*$ . On introduit plusieurs notions de convergence.

#### 4.1.1 Convergence en loi

On dit que la suite  $X_1, X_2, \dots$  *converge en loi* s'il existe une variable aléatoire  $X$  telle que *pour tout*  $x \in \mathbb{R}$  *tel que*  $F_X$  *est continue en*  $x$  (donc pour tout  $x \in \mathbb{R}$  qui n'est pas un atome de  $X$ ),

$$F_{X_n}(x) \rightarrow F_X(x), \quad n \rightarrow \infty.$$

On dit alors que  $X_n$  converge en loi vers  $X$  quand  $n \rightarrow \infty$  et on note  $X_n \xrightarrow{\text{loi}} X$ .

*Remarque.* La convergence en loi de variables aléatoires porte seulement sur les lois marginales des v.a.  $X, X_1, X_2, \dots$  et non pas sur leur loi jointe. Le lien entre les v.a. ne joue donc aucun rôle pour la convergence en loi, elles peuvent même décrire des expériences aléatoires différentes. Il serait même plus naturel de parler de la convergence *des lois* au lieu de la convergence des variables aléatoires, mais cette dernière terminologie est bien établie.

De la définition de la convergence en loi découlent des convergences de probabilités plus générales : si  $X_n$  converge en loi vers  $X$  quand  $n \rightarrow \infty$  et si  $I$  est un intervalle tel que  $\inf I$  et  $\sup I$  ne sont pas des atomes de  $X$ , alors

$$P(X_n \in I) \rightarrow P(X \in I), \quad n \rightarrow \infty.$$

**Exemple 4.1.** Soit  $X_n$  la v.a. constante égale à  $1/n$  (donc  $X_n \sim \delta_{1/n}$ ). Vérifions que  $X_n \rightarrow 0$  quand  $n \rightarrow \infty$ . La fonction de répartition de la loi  $\delta_0$  est  $F_0 = \mathbf{1}_{[0, \infty[}$ . Elle est discontinue en

0 (un atome de la loi  $\delta_0$ ) et continue ailleurs. Pour  $x < 0$ , on a

$$F_{X_n}(x) = \mathbb{1}_{[1/n, \infty[}(x) = 0 = F_0(x), \quad \text{pour tout } n,$$

et pour  $x > 0$ ,

$$F_{X_n}(x) = \mathbb{1}_{[1/n, \infty[}(x) = 1 = F_0(x) \quad \text{pour tout } n \text{ tel que } 1/n < x.$$

Par conséquent,  $F_{X_n}(x) \rightarrow F_0(x)$  pour tout  $x \in \mathbb{R}^*$ , et donc  $X_n \xrightarrow{\text{loi}} 0$ . Notons que  $F_{X_n}(0) = 0$  pour tout  $n$ , alors que  $F_0(0) = 1$ , donc  $F_{X_n}(0) \not\rightarrow F_0(0)$ .

**Exemple 4.2.** Soit  $\lambda > 0$ ,  $X_n \sim \text{Geom}_{\mathbb{N}^*}(\lambda/n)$ , et  $X \sim \text{Exp}(\lambda)$ . Alors, pour  $t \geq 0$  réel,  $P(X_n \leq t) = 1 - (1 - \lambda/n)^{\lfloor t \rfloor}$ , et partant,  $P(X_n/n \leq t) = 1 - (1 - \lambda/n)^{\lfloor tn \rfloor}$ . Les bornes  $tn - 1 \leq \lfloor tn \rfloor \leq tn$  permettent alors de conclure que :

$$P\left(\frac{X_n}{n} \leq t\right) \rightarrow 1 - e^{-\lambda t}$$

et la convergence vaut encore pour  $t \leq 0$  avec limite 0 ; au final, on a donc convergence en tout point  $t$ , et donc  $X_n/n \xrightarrow{\text{loi}} X$ .

**Lemme 4.3.** Si les v.a.  $(X_n)_{n \geq 1}$  et  $X$  sont à valeurs dans  $\mathbb{N}$ , on a l'équivalence

$$X_n \xrightarrow{\text{loi}} X \iff P(X_n = k) \rightarrow P(X = k) \forall k \in \mathbb{N},$$

*Démonstration.* Il suffit de noter que  $F_{X_n}(x) = \sum_{k=0}^{\lfloor x \rfloor} P(X_n = k)$  pour tout  $x$ . □

L'exemple suivant est fondamental.

**Exemple 4.4.** Soit  $p_n \in [0, 1]$  une suite telle que  $np_n \rightarrow \lambda \geq 0$ ,  $X_n \sim \text{Bin}(n, p_n)$ , et  $X \sim \text{Po}(\lambda)$ . Alors

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} (p_n)^k (1 - p_n)^{n-k} \\ &= \frac{n(n-1) \dots (n-k+1)}{k!} (p_n)^k (1 - p_n)^{n-k} \\ &= \frac{n(n-1) \dots (n-k+1)}{n^k} \cdot \frac{1}{k!} \cdot (np_n)^k \cdot e^{(n-k) \ln(1-p_n)} \\ &\rightarrow 1 \cdot \frac{1}{k!} \cdot \lambda^k \cdot e^{-k} = P(X = k) \end{aligned}$$

Ainsi

$$\text{Bin}(n, p_n) \rightarrow \text{Po}(\lambda), \quad \text{si } p_n \rightarrow 0 \text{ et } np_n \rightarrow \lambda.$$

D'un point de vue théorique il est bon de noter la caractérisation suivante de la convergence en loi :

**Proposition 4.5.**  $X_n$  converge en loi vers  $X$  si et seulement s'il existe un sous-ensemble dense  $D \subset \mathbb{R}$  tel que pour tout  $x \in D$

$$F_{X_n}(x) \rightarrow F_X(x), \quad n \rightarrow \infty. \tag{4.1}$$

*Démonstration.* Pour le sens direct, il suffit de noter que l'ensemble  $S_X$  des atomes de  $X$  étant au plus dénombrable, son complémentaire (sur lequel on a convergence de  $F_{X_n}$  vers  $F$ ) est dense dans  $\mathbb{R}$  (pour voir ceci noter qu'un ensemble de mesure de Lebesgue non nulle est non vide). Pour le sens réciproque il faut travailler un peu plus : on se donne  $D \subset \mathbb{R}$  sous-ensemble dense sur lequel (4.1) a lieu et  $x \in \mathbb{R} \setminus S_X$ . Soit  $\varepsilon > 0$ , par densité de  $D$ , il existe  $y, z \in D$  tels que  $x - \varepsilon < y < x < z < x + \varepsilon$  et l'inégalité  $F_{X_n}(y) \leq F_{X_n}(x) \leq F_{X_n}(z)$  donne alors, en prenant la limite en  $n$  :

$$F_X(y) \leq \liminf F_{X_n}(x) \leq \limsup F_{X_n}(x) \leq F_X(z)$$

En prenant maintenant la limite en  $\varepsilon \rightarrow 0$  et en utilisant que  $x \in \mathbb{R} \setminus S_X$ , on a :

$$0 \leq F_X(z) - F_X(y) \leq F_X(x + \varepsilon) - F_X(x - \varepsilon) = F_X(x) - F_X(x-) \rightarrow 0 \text{ quand } \varepsilon \rightarrow 0$$

de sorte que  $\liminf F_{X_n}(x) = \limsup F_{X_n}(x)$  pour  $x \in \mathbb{R} \setminus S_X$  et on conclut bien que  $\lim_n F_{X_n}(x) = F_X(x)$ .  $\square$

Une seconde caractérisation très utile est la suivante (elle se généralise mieux à des espaces d'état plus généraux, et fait pour cette raison souvent office de définition dans les cours plus avancés).

**Proposition 4.6.**  $X_n$  converge en loi vers  $X$  si et seulement si pour toute fonction continue bornée  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  :

$$\mathbb{E}[\varphi(X_n)] \rightarrow \mathbb{E}[\varphi(X)]$$

*Démonstration.* Pour l'implication on se contente de prouver la propriété pour une fonction  $\varphi$  continue positive strictement croissante. La fonction  $\varphi$  induit une bijection de  $\mathbb{R}$  sur un intervalle  $I$  de  $\mathbb{R}^+$ , et  $\{\varphi(X_n) > x\} = \{X_n > \varphi^{-1}(x)\}$ . Si  $S_X$  est au plus dénombrable c'est encore le cas de  $\varphi^{-1}(S_X)$ , et donc pour tout  $x$  en dehors de cet ensemble,  $P(\varphi(X_n) > x) = P(X_n > \varphi^{-1}(x)) \rightarrow P(X > \varphi^{-1}(x)) = P(\varphi(X) > x)$ , on déduit du théorème de convergence dominée que :  $E[\varphi(X_n)] = \int P(\varphi(X_n) > x) dx \rightarrow \int P(\varphi(X) > x) dx = E[\varphi(X)]$ .

On considère maintenant la réciproque. Soit  $y \in \mathbb{R}$ . On introduit les deux fonctions  $\varphi_{\varepsilon,-}(x) = 1 \wedge (-\frac{1}{\varepsilon}(x - y)) \vee 0$  et  $\varphi_{\varepsilon,+}(x) = 1 \wedge (-\frac{1}{\varepsilon}(x - y - \varepsilon)) \vee 0^1$ , et on note :

$$\mathbb{E}[\varphi_{\varepsilon,-}(X_n)] \leq \mathbb{P}(X_n \leq y) \leq \mathbb{E}[\varphi_{\varepsilon,+}(X_n)]$$

Puisque  $\varphi_{\varepsilon,-}$  et  $\varphi_{\varepsilon,+}$  sont continues et bornées, il suit de l'hypothèse que :

$$\mathbb{E}[\varphi_{\varepsilon,-}(X)] \leq \liminf \mathbb{P}(X_n \leq y) \leq \limsup \mathbb{P}(X_n \leq y) \leq \mathbb{E}[\varphi_{\varepsilon,+}(X)].$$

Des inégalités  $\mathbf{1}_{x \leq y - \varepsilon} \leq \varphi_{\varepsilon,-} \leq \varphi_{\varepsilon,+} \leq \mathbf{1}_{x \leq y + \varepsilon}$ , on déduit alors :

$$\mathbb{P}(X \leq y - \varepsilon) \leq \liminf \mathbb{P}(X_n \leq y) \leq \limsup \mathbb{P}(X_n \leq y) \leq \mathbb{P}(X \leq y + \varepsilon),$$

ce qui implique :

$$\mathbb{P}(X < y) \leq \liminf \mathbb{P}(X_n \leq y) \leq \limsup \mathbb{P}(X_n \leq y) \leq \mathbb{P}(X \leq y).$$

Si maintenant  $y$  n'est pas un atome de  $X$ , les deux membres extrêmes de l'inégalité coïncident et on conclut bien que :  $\lim_{n \rightarrow \infty} \mathbb{P}(X_n \leq y) = \mathbb{P}(X \leq y)$ .  $\square$

1. on rappelle la notation  $x \wedge y = \min\{x, y\}$ ,  $x \vee y = \max\{x, y\}$

### 4.1.2 Convergence en probabilité

On dit que la suite  $X_1, X_2, \dots$  converge en probabilité s'il existe une variable aléatoire  $X$  telle que

$$\forall \varepsilon > 0 : P(|X_n - X| > \varepsilon) \rightarrow 0, \quad n \rightarrow \infty.$$

On dit alors que  $X_n$  converge en probabilité vers  $X$  quand  $n \rightarrow \infty$  et on note  $X_n \xrightarrow{P} X$ .

**Exemple 4.7.** Soit  $X$  une v.a. et  $X_n = \min(X, n)$ . Alors pour tout  $\varepsilon > 0$ ,

$$P(|X_n - X| > \varepsilon) \leq P(X_n \neq X) = P(X > n) \rightarrow 0, \quad n \rightarrow \infty.$$

Par conséquent,  $X_n \xrightarrow{P} X, n \rightarrow \infty$ .

**Proposition 4.8.** 1. La convergence en probabilité entraîne la convergence en loi.

2. On a équivalence entre

$$X_n \xrightarrow{P} X \iff |X_n - X| \xrightarrow{\text{loi}} 0 \iff X_n - X \xrightarrow{\text{loi}} 0.$$

3. Si la limite est une constante, les deux notions de convergence sont équivalentes. Autrement dit, si  $c \in \mathbb{R}$ , alors

$$X_n \xrightarrow{\text{loi}} c \iff X_n \xrightarrow{P} c.$$

*Démonstration.* 1. Soit  $x \in \mathbb{R}$  tel que  $F_X$  est continue en  $x$ . Soit  $\varepsilon > 0$ . Alors,

$$\begin{aligned} F_{X_n}(x) &= P(X_n \leq x) \\ &= P(X_n \leq x, |X - X_n| \leq \varepsilon) + P(X_n \leq x, |X_n - X| > \varepsilon) \\ &\leq P(X \leq x + \varepsilon, |X - X_n| \leq \varepsilon) + P(|X_n - X| > \varepsilon) \\ &\leq F_X(x + \varepsilon) + P(|X_n - X| > \varepsilon). \end{aligned}$$

Puisque  $X_n \xrightarrow{P} X$ , le second terme tend vers 0, si bien que

$$\limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x + \varepsilon).$$

Par ailleurs,

$$\begin{aligned} P(X \leq x - \varepsilon) &\leq P(X \leq x - \varepsilon, |X - X_n| \leq \varepsilon) + P(X \leq x - \varepsilon, |X - X_n| > \varepsilon) \\ &\leq P(X_n \leq x) + P(|X - X_n| > \varepsilon) \end{aligned}$$

Donc :

$$\begin{aligned} F_{X_n}(x) &= P(X_n \leq x) \\ &\geq P(X \leq x - \varepsilon) - P(|X_n - X| > \varepsilon) \\ &= F_X(x - \varepsilon) - P(|X_n - X| > \varepsilon). \end{aligned}$$

Puis

$$\liminf_{n \rightarrow \infty} F_{X_n}(x) \geq F_X(x - \varepsilon).$$

Laissant  $\varepsilon > 0$ , puisque  $F_X$  est continue en  $x$ , on obtient :

$$F_{X_n}(x) \rightarrow F_X(x), \quad n \rightarrow \infty.$$

2. Montrons la première équivalence : puisque  $F_{|X_n - X|}(x) = 0$  pour tout  $x < 0$ , on a

$$\begin{aligned} |X_n - X| \xrightarrow{\text{loi}} 0 &\iff \forall x > 0 : F_{|X_n - X|}(x) \rightarrow 1 \\ &\iff \forall x > 0 : P(|X_n - X| > x) \rightarrow 0 \\ &\iff X_n \xrightarrow{P} X. \end{aligned}$$

Pour la deuxième équivalence, on a

$$\begin{aligned} X_n - X \xrightarrow{\text{loi}} 0 &\iff \forall x > 0 : F_{X_n - X}(x) \rightarrow 1 \text{ et } F_{X_n - X}(-x) \rightarrow 0 \\ &\iff \forall x > 0 : P(X_n - X > x) \rightarrow 0 \text{ et } P(X_n - X < -x) \rightarrow 0 \\ &\iff \forall x > 0 : P(X_n - X > x) + P(X_n - X < -x) \rightarrow 0 \\ &\iff \forall x > 0 : P(|X_n - X| > x) \rightarrow 0 \\ &\iff X_n \xrightarrow{P} X. \end{aligned}$$

3. Il suffit de montrer que la convergence en loi entraîne la convergence en probabilité si la limite est une constante. Si  $X_n \xrightarrow{\text{loi}} c$ , alors pour tout  $\varepsilon > 0$ ,

$$P(|X_n - c| > \varepsilon) = P(X_n < c - \varepsilon) + P(X_n > c + \varepsilon) \leq F_{X_n}(c - \varepsilon) + (1 - F_{X_n}(c + \varepsilon)) \rightarrow 0,$$

car les deux termes tendent vers 0. □

Si  $(x_n)_{n \geq 1}$  est une suite de nombres qui converge vers  $x \in \mathbb{R}$  et  $g : \mathbb{R} \rightarrow \mathbb{R}$  une fonction, alors  $g(x_n) \rightarrow g(x)$  si  $g$  est continue en  $x$ . Pour les v.a., il existe un résultat analogue :

**Théorème 4.9** (Lemme de l'application continue). *Supposons que  $X$  est à valeurs dans un ensemble  $E \subset \mathbb{R}$ . Soit  $g : \mathbb{R} \rightarrow \mathbb{R}$  une fonction qu'on suppose continue en tout point  $x \in E$ . Alors*

1.  $X_n \xrightarrow{P} X$  implique  $g(X_n) \xrightarrow{P} g(X)$ .
2.  $X_n \xrightarrow{\text{loi}} X$  implique  $g(X_n) \xrightarrow{\text{loi}} g(X)$ .

*Démonstration.* 1. On suppose  $X_n \xrightarrow{P} X$ . Pour simplifier, on suppose que les v.a.  $(X_n)_{n \geq 1}$  et  $X$  sont à valeurs dans un compact  $K$  et que  $g$  est continue sur  $K$  (il s'avère qu'on peut toujours se ramener à ce cas). Alors  $g$  est uniformément continue, c'est-à-dire, pour tout  $\varepsilon > 0$ , il existe  $\delta > 0$  tel que pour tout  $x, y \in K$ ,  $|g(x) - g(y)| > \varepsilon$  implique  $|x - y| \geq \delta$ . Par conséquent,

$$P(|g(X_n) - g(X)| > \varepsilon) \leq P(|X_n - X| > \delta) \rightarrow 0, \quad n \rightarrow \infty,$$

si bien que  $g(X_n) \xrightarrow{P} g(X)$ ,  $n \rightarrow \infty$ .

2. Si  $\varphi$  est continue bornée,  $g$  étant continue, la composée  $\varphi \circ g$  est encore continue bornée et il suit que  $\mathbb{E}[\varphi \circ g(X_n)] \rightarrow \mathbb{E}[\varphi \circ g(X)]$ , ce qui prouve la convergence en loi. □

**Exemple 4.10.** Comme ci-dessus, soit  $p \in ]0, 1[$  et pour tout  $n \in \mathbb{N}^*$ , soit  $X_n \sim \text{Bin}(n, p)$ , et soit  $X \sim \mathcal{N}(0, 1)$ . Posons  $X_n = (Y_n - np)/\sqrt{p(1-p)n}$ . On définit

$$g(x) = \begin{cases} 1/x, & \text{if } x \neq 0 \\ 0, & \text{if } x = 0. \end{cases}$$

Alors  $g$  est continue sur  $\mathbb{R}^*$  et  $X$  est à valeurs dans  $\mathbb{R}^*$ , car  $P(X = 0) = 0$ . Par le lemme de l'application continue,

$$g(X_n) \xrightarrow{\text{loi}} g(X), \quad n \rightarrow \infty,$$

ce qu'on peut écrire avec un très léger abus de notation,

$$1/X_n \xrightarrow{\text{loi}} 1/X, \quad n \rightarrow \infty.$$

### 4.1.3 Convergence dans $L^p$

Soit  $p > 0$ . On dit que la suite  $X_1, X_2, \dots$  converge dans  $L^p$  s'il existe une variable aléatoire  $X$  avec  $E[|X|^p] < \infty$  et telle que

$$E[|X_n - X|^p] \rightarrow 0, \quad n \rightarrow \infty.$$

et on note alors  $X_n \xrightarrow{L^p} X$ . Le cas où  $p = 2$  est le plus courant et on dit alors que  $X_n$  converge en moyenne quadratique vers  $X$ .

**Proposition 4.11.** Soit  $0 < q < p$ . Si  $X_n \xrightarrow{L^p} X$  alors  $X_n \xrightarrow{L^q} X$ .

*Démonstration.* Cela découle de l'inégalité de Jensen. Supposons la convergence  $L^p$ . Puisque  $p/q > 1$ , la fonction  $x \geq 0 \mapsto x^{p/q}$  est convexe et il suit que

$$E[|X_n - X|^{p/q}] \leq E[|X_n - X|^p] \rightarrow 0.$$

□

**Proposition 4.12.** Soit  $p > 0$ . La convergence dans  $L^p$  implique la convergence en probabilité.

*Démonstration.* Supposons que  $X_n \xrightarrow{L^p} X$ . On note que  $\{|X_n - X| > \varepsilon\} = \{|X_n - X|^p > \varepsilon^p\}$  puis, d'où on déduit par l'inégalité de Markov que, pour tout  $\varepsilon > 0$ ,

$$P(|X_n - X| > \varepsilon) = P(|X_n - X|^p > \varepsilon^p) \leq \frac{E[|X_n - X|^p]}{\varepsilon^p} \rightarrow 0.$$

□

### 4.1.4 Convergence presque sûre

On dit que la suite  $X_1, X_2, \dots$  converge presque sûrement s'il existe une variable aléatoire  $X$  telle que

$$P\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1.$$

On dit alors que  $X_n$  converge presque sûrement vers  $X$  quand  $n \rightarrow \infty$  et on note  $X_n \xrightarrow{\text{p.s.}} X$ .

Nous n'allons pas utiliser cette notion de convergence par la suite. On peut montrer que la convergence presque sûre implique la convergence en probabilité (mais pas nécessairement la convergence dans  $L^p$ ).



## 4.2 Convergence de couples aléatoires

On introduit dans cette section les notions de convergence en loi et en probabilité de couples aléatoires. Elles se généralisent de manière évidente à des vecteurs aléatoires, mais nous nous restreignons aux couples pour rendre la notation plus simple.

Soit  $(X_n, Y_n)_{n \geq 1}$  une suite de couples aléatoires et  $(X, Y)$  un autre couple aléatoire.

**Convergence en loi.** On dit que  $(X_n, Y_n)$  converge en loi vers  $(X, Y)$  et on écrit  $(X_n, Y_n) \xrightarrow{\text{loi}} (X, Y)$  si pour tout  $x, y \in \mathbb{R}$  tels que  $F_{X,Y}$  est continue en  $(x, y)$ ,

$$F_{X_n, Y_n}(x, y) \rightarrow F_{X, Y}(x, y), \quad n \rightarrow \infty.$$

On utilisera souvent le résultat suivant (qui est valable aussi dans le cas uni-dimensionnel) :

**Lemme 4.13.**  $(X_n, Y_n) \xrightarrow{\text{loi}} (X, Y)$  si et seulement s'il existe un ensemble dense  $E \subset \mathbb{R}^2$  tel que pour tout  $(x, y) \in E$ ,

$$F_{X_n, Y_n}(x, y) \rightarrow F_{X, Y}(x, y), \quad n \rightarrow \infty.$$

*Démonstration.* Exercice, utilise la croissance des fonctions de répartition. □

La définition qui se généralise le mieux est en fait celle qui utilise les fonctions continues bornées :

**Proposition 4.14.**  $(X_n, Y_n)$  converge en loi vers  $(X, Y)$  si et seulement si pour toute fonction continue bornée  $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$  :

$$\mathbb{E}[\varphi(X_n, Y_n)] \rightarrow \mathbb{E}[\varphi(X, Y)]$$

**Convergence en probabilité.** On dit que  $(X_n, Y_n)$  converge en probabilité vers  $(X, Y)$  et on écrit  $(X_n, Y_n) \xrightarrow{P} (X, Y)$  si pour tout  $\varepsilon > 0$ ,

$$P(|X_n - X| + |Y_n - Y| > \varepsilon) \rightarrow 0, \quad n \rightarrow \infty.$$

Remarque : on peut remplacer la norme  $|X_n - X| + |Y_n - Y|$  par n'importe quelle autre norme, par exemple la norme euclidienne.

**Proposition 4.15.** 1. La convergence en probabilité entraîne la convergence en loi.

2. On a équivalence entre

$$(X_n, Y_n) \xrightarrow{P} (X, Y) \iff (X_n - X, Y_n - Y) \xrightarrow{\text{loi}} (0, 0) \iff |X_n - X| + |Y_n - Y| \xrightarrow{\text{loi}} 0.$$

3. Si la limite est une constante, les deux notions de convergence sont équivalentes. Autrement dit, si  $c, c' \in \mathbb{R}$ , alors

$$(X_n, Y_n) \xrightarrow{\text{loi}} (c, c') \iff (X_n, Y_n) \xrightarrow{P} (c, c').$$

*Démonstration.* Preuves similaires au cas univarié. □

**Theorème 4.16** (Lemme de l'application continue ; version couple). *Supposons que le couple  $(X, Y)$  est à valeurs dans un ensemble  $E \subset \mathbb{R}^2$ . Soit  $g : \mathbb{R}^2 \rightarrow \mathbb{R}$  ou  $g : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  une fonction qu'on suppose continue en tout point  $x \in E$ . Alors*

1.  $(X_n, Y_n) \xrightarrow{P} (X, Y)$  implique  $g(X_n, Y_n) \xrightarrow{P} g(X, Y)$ .
2.  $(X_n, Y_n) \xrightarrow{\text{loi}} (X, Y)$  implique  $g(X_n, Y_n) \xrightarrow{\text{loi}} g(X, Y)$ .

*Démonstration.* La première partie se démontre comme dans le cas univarié. Pour la seconde, on admet qu'on peut encore se ramener à une convergence en probabilité. La preuve est alors similaire.  $\square$

- Lemme 4.17.**
1.  $(X_n, Y_n) \xrightarrow{P} (X, Y)$  si et seulement si  $X_n \xrightarrow{P} X$  et  $Y_n \xrightarrow{P} Y$ .
  2. Si  $(X_n, Y_n) \xrightarrow{\text{loi}} (X, Y)$ , alors  $X_n \xrightarrow{\text{loi}} X$  et  $Y_n \xrightarrow{\text{loi}} Y$  (la réciproque est fautive en général)
  3. Si  $X_n \xrightarrow{\text{loi}} X$  et  $Y_n \xrightarrow{\text{loi}} c$ , où  $c \in \mathbb{R}$  est une constante, alors  $(X_n, Y_n) \xrightarrow{\text{loi}} (X, c)$ .
  4. Si  $X_n \xrightarrow{\text{loi}} X$  et  $Y_n \xrightarrow{\text{loi}} Y$  et si  $X_n$  et  $Y_n$  sont indépendantes pour tout  $n \in \mathbb{N}$ , alors  $(X_n, Y_n) \xrightarrow{\text{loi}} (X, Y)$ , où  $X$  et  $Y$  sont indépendantes.

*Démonstration.* 1. Si  $(X_n, Y_n) \xrightarrow{P} (X, Y)$ , alors le lemme de l'application continue appliqué aux fonctions  $(x, y) \mapsto x$  et  $(x, y) \mapsto y$  donne  $X_n \xrightarrow{P} X$  et  $Y_n \xrightarrow{P} Y$ . Pour la réciproque, soit  $\varepsilon > 0$ . Puisque  $|X_n - X| + |Y_n - Y| > \varepsilon$  implique  $|X_n - X| > \varepsilon/2$  ou  $|Y_n - Y| > \varepsilon/2$ ,

$$P(|X_n - X| + |Y_n - Y| > \varepsilon) \leq P(|X_n - X| > \varepsilon/2) + P(|Y_n - Y| > \varepsilon/2).$$

Par conséquent, quand  $n \rightarrow \infty$ ,  $P(|X_n - X| + |Y_n - Y| > \varepsilon) \rightarrow 0$  pour tout  $\varepsilon > 0$  si et seulement si  $P(|X_n - X| > \varepsilon) \rightarrow 0$  et  $P(|Y_n - Y| > \varepsilon) \rightarrow 0$  pour tout  $\varepsilon > 0$ .

2. Encore lemme de l'application continue comme ci-dessus.
3. La fonction de répartition de  $(X, c)$  s'écrit

$$F_{X,c}(x, y) = F_X(x) \mathbf{1}_{y \geq c}.$$

Soit  $A \subset \mathbb{R}$  un ensemble dense de points  $x \in \mathbb{R}$  tels que  $F_{X_n}(x) \rightarrow F_X(x)$ . Alors pour tout  $y < c$ ,

$$F_{X_n, Y_n}(x, y) \leq F_{Y_n}(y) \rightarrow 0,$$

et pour tout  $y > c$ ,

$$F_{X_n, Y_n}(x, y) \geq F_{X_n}(x) - P(Y_n > y) \rightarrow F_X(x),$$

donc

$$F_{X_n, Y_n}(x, y) \rightarrow F_{X,c}(x, y) \quad \forall (x, y) \in A \times (\mathbb{R} \setminus \{c\}).$$

L'ensemble  $A \times (\mathbb{R} \setminus \{c\})$  est dense dans  $\mathbb{R}^2$ , ce qui permet de conclure que  $(X_n, Y_n) \xrightarrow{\text{loi}} (X, c)$ .

4. Par indépendance,  $F_{X,Y}(x, y) = F_X(x)F_Y(y)$  pour tout  $(x, y) \in \mathbb{R}$ . Par conséquent, pour tout  $x, y \in \mathbb{R}$  points de continuité respectifs de  $F_X$  et  $F_Y$ ,

$$F_{X_n, Y_n}(x, y) = F_{X_n}(x)F_{Y_n}(y) \rightarrow F_X(x)F_Y(y) = F_{X,Y}(x, y).$$

Cet ensemble de points  $(x, y)$  étant dense dans  $\mathbb{R}^2$ , on conclut que  $(X_n, Y_n) \xrightarrow{\text{loi}} (X, c)$ .  $\square$

**Corollaire 4.18** (Lemme de Slutsky). *Supposons que  $X_n \xrightarrow{loi} X$  et  $Y_n \xrightarrow{loi} c$ , où  $c \in \mathbb{R}$  est une constante. Alors*

- $X_n + Y_n \xrightarrow{loi} X + c$
- $X_n Y_n \xrightarrow{loi} cX$
- Si  $c \neq 0$ ,  $X_n/Y_n \xrightarrow{loi} X/c$ .

### 4.3 Loi des grands nombres

**Theorème 4.19** (Loi des grands nombres). *Soient  $X_1, X_2, \dots$  i.i.d. avec  $E[|X_1|] < \infty$ . Alors*

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{P} E[X_1].$$

*Remarque 4.20.* Comme expliqué lors de l'introduction de l'espérance, la loi des grands nombres justifie l'interprétation de celle-ci comme « moyenne » de la v.a. : la *moyenne empirique* de  $n$  réalisations de cette v.a. approche son espérance quand  $n$  est grand.

*Remarque 4.21.* Dans l'énoncé de la loi des grands nombres, on peut en fait remplacer la convergence en probabilité par de la convergence presque sûre. C'est ce qu'on appelle aussi la *loi forte des grands nombres*.

*Démonstration.* Nous commençons par fournir une preuve très simple sous l'hypothèse  $E[X_1^2] < \infty$ . On montre alors la convergence en moyenne quadratique qui on le rappelle implique la convergence en probabilité. Le cas général où on a seulement  $E[|X_1|] < \infty$  est plus évolué. Remarquons d'abord que  $E[(X_1 + \dots + X_n)/n] = nE[X_1]/n = E[X_1]$ , par linéarité de l'espérance. Par conséquent,

$$\begin{aligned} E \left[ \left( \frac{X_1 + \dots + X_n}{n} - E[X_1] \right)^2 \right] &= \text{Var} \left( \frac{X_1 + \dots + X_n}{n} \right) \\ &= \frac{1}{n^2} (\text{Var}(X_1) + \dots + \text{Var}(X_n)) \quad \text{par indépendance} \\ &= \frac{\text{Var}(X_1)}{n} \rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

Nous présentons maintenant l'extension au cas intégrable, c'est-à-dire qu'on suppose seulement  $E[|X_1|] < \infty$ <sup>2</sup>. Posons  $S_n = X_1 + \dots + X_n$ . Quitte à considérer  $X_1 - E[X_1]$  on peut supposer que les  $(X_i)_{1 \leq i \leq N}$  sont des variables aléatoires centrées, ce qu'on fera désormais  $E[X_1] = 0$ . Maintenant, pour  $N$  entier, on pose :

$$X_i^N = X_i \mathbb{1}_{|X_i| \leq N} \text{ et } S_n^N = \sum_{1 \leq i \leq n} X_i^N,$$

c'est-à-dire qu'on considère les variables tronquées ainsi que leur somme. On se fixe  $\varepsilon > 0$ . Il nous suffit de montrer que pour tout  $n$  suffisamment grand,  $P(|S_n| > \varepsilon n) \leq \varepsilon$ . Pour ce faire, on note que par convergence dominée :  $E[X_1^N] \rightarrow E[X_1] = 0$  et  $E[|X_1 - X_1^N|] \rightarrow 0$ , et donc on peut choisir  $N$  suffisamment grand tel que

$$|E[X_1^N]| \leq \varepsilon/4 \text{ et } E[|X_1 - X_1^N|] = E[|X_1| \mathbb{1}_{|X_1^N| > N}] \leq \varepsilon^2/4.$$

2. à ne lire qu'en seconde lecture, cette preuve est d'un niveau significativement plus avancé que le reste du cours

(Le choix de ce dernier terme sera expliqué dans quelques lignes). Puisque les variables bornées sont de carré intégrable, la première partie de la preuve implique :  $P(|S_n^N - nE[|X_1^N|]| > \varepsilon n/4) \rightarrow 0$ , et donc il existe un rang  $n_0$  à partir duquel on a  $P(|S_n^N - nE[|X_1^N|]| > \varepsilon n/4) \leq \varepsilon/2$ . Notons également que  $S_n - S_n^N = \sum_{1 \leq i \leq N} X_i \mathbb{1}_{|X_i| > N}$  donc

$$P(|S_n - S_n^N| > \varepsilon n/2) \leq 2 \frac{E[|S_n - S_n^N|]}{\varepsilon n} \leq 2n \frac{E[|X_1 - X_1^N|]}{\varepsilon n} \leq \frac{\varepsilon}{2}$$

grâce à la majoration précédente de  $E[|X_1| \mathbb{1}_{|X_1^N| > N}]$ . Finalement pour  $n \geq n_0$  :

$$\begin{aligned} P(|S_n| > \varepsilon n) &= P(|S_n^N| > \varepsilon n/2) + P(|S_n - S_n^N| > \varepsilon n/2) \\ &\leq P(|S_n^N - nE[|X_1^N|]| > \varepsilon n/4) + P(|S_n - S_n^N| > \varepsilon n/2) \\ &\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \leq \varepsilon. \end{aligned}$$

□

## 4.4 Théorème central limite

La loi des grands nombres nous dit que si  $X_1, \dots, X_n$  sont i.i.d. d'espérance finie, alors  $X_1 + \dots + X_n \approx n \times E[X_1]$ . Il est naturel de se demander quelles sont les *fluctuations* autour de cette moyenne, c'est-à-dire, quelle est l'ordre de grandeur et la loi de  $X_1 + \dots + X_n - n \times E[X_1]$ ? Les réponses diffèrent selon que  $\sigma^2 = \text{Var}(X_1)$  est finie ou non, et on suppose dorénavant que  $\sigma^2 < \infty$ . Alors  $\text{Var}(X_1 + \dots + X_n) = n\sigma^2$ , ce qui laisse penser que  $X_1 + \dots + X_n - n \times E[X_1]$  est d'ordre  $\sigma\sqrt{n}$ . Le théorème central limite donne une réponse positive et très précise à cette question.

**Théorème 4.22** (Théorème central limite). *Soient  $X_1, X_2, \dots$  i.i.d. de variance  $\sigma^2 = \text{Var}(X_1) < \infty$ . Alors*

$$\frac{X_1 + \dots + X_n - nE[X_1]}{\sigma\sqrt{n}} \xrightarrow{\text{loi}} \mathcal{N}(0, 1).$$

Les propriétés de stabilité de la loi normale/gaussienne montrent qu'on avait guère le choix pour la loi limite, comme le justifie l'exercice suivant :

**Exercice 4.23.** *On suppose dans cet exercice connaître l'énoncé du théorème central limite, à l'exception de la forme de la loi limite : précisément, on sait qu'on a convergence, mais on ne sait pas vers quelle loi limite.*

1. *Si l'on choisit  $X_1, X_2, \dots$  v.a.r. i.i.d. selon la loi normale, quelle loi suit  $\frac{X_1 + \dots + X_n - nE[X_1]}{\sigma\sqrt{n}}$  ?*
2. *En déduire la loi limite.*

Nous allons démontrer ce théorème à l'aide des fonctions génératrices des moments. Nous aurons besoin du lemme suivant que nous admettons.

**Lemme 4.24.** *Soit  $(X_n)_{n \geq 1}$  une suite de v.a. et  $X$  une v.a. telle que  $\varphi_X$  est finie dans un voisinage de 0. Supposons que  $\varphi_{X_n}(t) \rightarrow \varphi_X(t)$  pour tout  $t \in \mathbb{R}$ . Alors  $X_n \xrightarrow{\text{loi}} X$ .*

*Démonstration du théorème central limite.* On travaillera sous l'hypothèse supplémentaire que la fonction génératrice des moments  $\varphi(t) = E[e^{tX_1}]$  est finie dans un voisinage de 0 (dans le cas général il faudrait passer par la *fonction caractéristique*  $t \mapsto E[e^{itX}]$  que nous n'avons pas introduite). Sans perte de généralité, quitte à remplacer  $X_1$  par sa version centrée réduite, on peut supposer  $E[X_1] = 0$  et  $\sigma^2 := E[X_1^2] = 1$ . On rappelle que

$$\varphi(0) = 1, \quad \varphi'(0) = E[X_1] = 0, \quad \varphi''(0) = E[X_1^2] = \text{Var}(X_1) = 1.$$

Par conséquent, le développement en 0 de  $\varphi$  est

$$\varphi(t) = 1 + \frac{1}{2}t^2 + o(t^2) = e^{\frac{1}{2}t^2 + o(t^2)}.$$

Par indépendance, on a pour  $t$  fixé et  $n \rightarrow \infty$ ,

$$\varphi_{\frac{X_1 + \dots + X_n}{\sqrt{n}}}(t) = \varphi\left(\frac{t}{\sqrt{n}}\right)^n = e^{n\left(\frac{1}{2}\frac{t^2}{n} + o(1/n)\right)} \rightarrow e^{\frac{t^2}{2}}.$$

On reconnaît la fonction génératrice des moments de la loi  $\mathcal{N}(0, 1)$ . Le lemme 4.24 montre alors que

$$\frac{X_1 + \dots + X_n}{\sqrt{n}} \xrightarrow{\text{loi}} \mathcal{N}(0, 1).$$

□

Un corollaire immédiat est le théorème d'approximation 1.30 de de Moivre - Laplace.



# Chapitre 5

## Variables aléatoires dépendantes

### 5.1 Lois marginales

On rappelle que la loi jointe de  $n$  variables aléatoires  $X_1, \dots, X_n$  détermine la loi de chacune des variables par la formule

$$P(X_k \in A) = P((X_1, \dots, X_n) \in \mathbb{R}^{k-1} \times A \times \mathbb{R}^{n-k}), \quad k \in \{1, \dots, n\}.$$

On appelle les lois des v.a.  $X_1, \dots, X_n$  dans ce contexte les *lois marginales* du vecteur aléatoire  $X_1, \dots, X_n$ . On peut obtenir la fonction de répartition de  $X_k$  en fonction de la *fonction de répartition conjointe* par la formule

$$\begin{aligned} F_{X_k}(x) &= \lim_{x_j \rightarrow +\infty \forall j \neq k} F_{X_1, \dots, X_n}(x_1, \dots, x_{k-1}, x, x_{k+1}, \dots, x_n) \\ &=: F_{X_1, \dots, X_n}(+\infty, \dots, +\infty, x, +\infty, \dots, +\infty), \quad k \in \{1, \dots, n\}. \end{aligned}$$

Dans le cas où les v.a. sont discrètes ou conjointement à densité, on a les formules suivantes pour les fonctions de masse ou les densités des lois marginales :

**Proposition 5.1.** — *Supposons que les v.a.  $X_1, \dots, X_n$  sont discrètes. Alors pour tout  $k \in \{1, \dots, n\}$  la fonction de masse de  $X_k$  est donnée par*

$$P(X_k = x) = \sum_{x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n} P(X_1 = x_1, \dots, X_{k-1} = x_{k-1}, X_k = x, X_{k+1} = x_{k+1}, \dots, X_n = x_n).$$

— *Supposons que les v.a.  $X_1, \dots, X_n$  admettent la densité conjointe  $f_{X_1, \dots, X_n}$ . Alors pour tout  $k \in \{1, \dots, n\}$ , la v.a.  $X_k$  admet la densité*

$$f_{X_k}(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(x_1, \dots, x_{k-1}, x, x_{k+1}, \dots, x_n) dx_1 \cdots dx_{k-1} dx_{k+1} \cdots dx_n.$$

*Démonstration.* Cas discret : On rappelle que pour tout  $A \subset \mathbb{R}^n$ ,

$$P((X_1, \dots, X_n) \in A) = \sum_{(x_1, \dots, x_n) \in A} P(X_1 = x_1, \dots, X_n = x_n).$$

Avec  $x \in \mathbb{R}$  et  $A = \mathbb{R}^{k-1} \times \{x\} \times \mathbb{R}^{n-k}$ , cela donne

$$\begin{aligned} P(X_k = x) &= \sum_{(x_1, \dots, x_n) \in \mathbb{R}^{k-1} \times \{x\} \times \mathbb{R}^{n-k}} P(X_1 = x_1, \dots, X_n = x_n) \\ &= \sum_{x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n} P(X_1 = x_1, \dots, X_{k-1} = x_{k-1}, X_k = x, X_{k+1} = x_{k+1}, \dots, X_n = x_n). \end{aligned}$$

Cas à densité. On rappelle que la fonction de répartition conjointe de  $X_1, \dots, X_n$  est donnée par

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{X_1, \dots, X_n}(y_1, \dots, y_n) dy_1 \cdots dy_n.$$

En faisant tendre  $x_j \rightarrow +\infty$  pour  $j \neq k$  et en changeant l'ordre d'intégration, on obtient

$$F_{X_k}(x) = \int_{-\infty}^x \left( \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{X_1, \dots, X_n}(y_1, \dots, y_{k-1}, y, y_{k+1}, \dots, y_n) dy_1 \cdots dy_{k-1} dy_{k+1} \cdots dy_n \right) dy.$$

Ceci montre que  $X_k$  est à densité avec la densité écrite dans l'énoncé.  $\square$

**Exemple 5.2.** On teste des individus d'une population pour l'infection à un virus. Pour un individu pris au hasard, on note  $X = 1$  si l'individu est infecté,  $X = 0$  sinon ainsi que  $Y = 1$  si le test est positif,  $Y = 0$  sinon. Après un grand nombre de tests, on a pu établir empiriquement la loi jointe de  $X$  et  $Y$ , dont la fonction de masse est représentée dans le *tableau croisé* suivant :

	$X = 0$	$X = 1$
$Y = 0$	0.7	0
$Y = 1$	0.2	0.1

On calcule les lois marginales :

$$\begin{aligned} P(X = 0) &= P(X = 0, Y = 0) + P(X = 0, Y = 1) = 0.9 \\ P(X = 1) &= 1 - P(X = 0) = 0.1 \\ P(Y = 0) &= P(X = 0, Y = 0) + P(X = 1, Y = 0) = 0.7 \\ P(Y = 1) &= 1 - P(Y = 0) = 0.3. \end{aligned}$$

En mots, 10% de la population sont infectés et le test est positif chez 30% de la population.

**Exemple 5.3.** Soit  $(X, Y)$  un couple aléatoire de loi uniforme sur le triangle  $T = \{(x, y) \in [0, 1]^2 : x + y \geq 1\}$ , i.e. de loi jointe  $f_{X,Y} = \mathbb{1}_T / \text{aire}(T) = 2\mathbb{1}_T$ . Alors  $X$  est de densité

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy \\ &= \int_{-\infty}^{+\infty} 2\mathbb{1}_{[0,1]}(x) \mathbb{1}_{(y \in [0,1], x+y \geq 1)} dy \\ &= 2\mathbb{1}_{[0,1]}(x) \mathbb{1}_{1-x}^x 1 dy \\ &= 2x \mathbb{1}_{[0,1]}(x). \end{aligned}$$

Par symétrie ( $(X, Y) \stackrel{\text{loi}}{=} (Y, X)$ ), on a également  $f_Y(x) = f_X(x) = 2x \mathbb{1}_{[0,1]}(x)$ .



## 5.2 Lois conditionnelles

Soient  $X$  et  $Y$  deux v.a. discrètes. Alors pour tout  $y$  tel que  $P(Y = y) > 0$ , la *loi de  $X$  conditionnellement à  $Y = y$*  est la loi discrète de fonction de masse

$$p_{X|Y}(x | y) = P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

Dans le cas où les v.a.  $X$  et  $Y$  ont une densité conjointe  $f_{X,Y}$ , et  $y \in \mathbb{R}$  est tel que  $f_Y(y) > 0$ , on définit de manière analogue la *loi de  $X$  conditionnellement à  $Y = y$*  comme étant la loi de densité

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

On peut traiter la loi conditionnelle comme n'importe quelle loi. En particulier, on peut définir *l'espérance conditionnelle* :

$$E[f(X) | Y = y] = \begin{cases} \sum_x f(x) \times p_{X|Y}(x | y) & \text{(cas discret)} \\ \int_{-\infty}^{+\infty} f(x) f_{X|Y}(x | y) dx & \text{(cas à densité)}. \end{cases}$$

On vérifie aisément les formules suivantes (dites *formules de la probabilité totale*) :

— Si  $X, Y$  discrètes de fonctions de masse  $p_X$  et  $p_Y$ , alors

$$p_X(x) = \sum_y p_{X|Y}(x | y) p_Y(y)$$

$$E[f(X)] = \sum_y E[f(X) | Y = y] p_Y(y) \quad \text{pour tout } f.$$

— Si  $X, Y$  conjointement à densité,

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X|Y}(x | y) f_Y(y) dy$$

$$E[f(X)] = \int_{-\infty}^{+\infty} E[f(X) | Y = y] f_Y(y) dy \quad \text{pour tout } f.$$

**Exemple 5.4.** On reprend l'exemple du test d'infection de la section précédente. On calcule les lois conditionnelles :

$$\begin{array}{ll} P(Y = 1 | X = 0) = 0.2/0.9 \approx 22\% & P(Y = 0 | X = 0) \approx 78\% \\ P(Y = 1 | X = 1) = 0.1/0.1 = 1 & P(Y = 0 | X = 1) = 0 \\ P(X = 1 | Y = 0) = 0/0.7 = 0 & P(X = 0 | Y = 0) = 1 \\ P(X = 1 | Y = 1) = 0.1/0.3 \approx 33\% & P(X = 0 | Y = 1) \approx 67\%. \end{array}$$

En mots, on retient

- Chez une personne infectée, le test est positif dans 100% des cas, mais aussi dans 22% des cas chez une personne non infectée.
- Une personne dont le test est négatif peut être 100% sûre de ne pas être infectée.
- Une personne dont le test est positif est réellement infectée dans un tiers des cas seulement.

**Exemple 5.5.** On reprend l'exemple de la loi uniforme sur le triangle. On calcule pour tout  $y \in [0, 1]$  la densité de  $X$  conditionnellement à  $Y = y$  :

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{2\mathbb{1}_{T(x,y)}}{2y\mathbb{1}_{[0,1]}(y)} = \frac{1}{y}\mathbb{1}_{(x \in [0,1], x+y \geq 1)} = \frac{1}{y}\mathbb{1}_{[1-y,1]}(x).$$

Autrement dit, conditionnellement à  $Y = y$ ,  $X$  suit la loi uniforme sur l'intervalle  $[1 - y, 1]$ . En particulier, l'espérance de  $X$  conditionnellement à  $Y = y$  vaut

$$E[X | Y = y] = \frac{1 - y + 1}{2} = 1 - \frac{y}{2}.$$

### 5.3 Covariance et corrélation

Soient  $X$  et  $Y$  des v.a. de carrés intégrables, c'est-à-dire  $E[X^2] < \infty$  et  $E[Y^2] < \infty$ . On définit la *covariance* de  $X$  et  $Y$  par

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

*NB : A priori, il n'est pas clair que cette covariance est bien définie et finie, mais nous allons voir ci-dessous que c'est toujours le cas.*

En développant le produit et en utilisant la linéarité de l'espérance, on obtient

$$\text{Cov}(X, Y) = E[XY - XE[Y] - E[X]Y + E[X]E[Y]] = E[XY] - 2E[X]E[Y] + E[X]E[Y],$$

si bien que

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y],$$

une formule qui rappelle la formule  $\text{Var}(X) = E[X^2] - E[X]^2$ . D'ailleurs, par définition,

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

donc la variance d'une v.a. est égale à sa covariance avec elle-même.

**Proposition 5.6.** Soient  $X, Y, Z$  des v.a. de carrés intégrables et  $a, b \in \mathbb{R}$ ,

1.  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
2.  $\text{Cov}(aX + bY, Z) = a \text{Cov}(X, Z) + b \text{Cov}(Y, Z)$ .
3.  $\text{Cov}(a, X) = 0$ .
4.  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$ .

*NB : Les deux premières propriétés de la proposition ci-dessus disent que la covariance est une forme bilinéaire symétrique.*

*Démonstration.* 1. Evident par la définition.

2. Par linéarité de l'espérance,

$$\begin{aligned} \text{Cov}(aX + bY, Z) &= E[(aX + bY - E[aX + bY])(Z - E[Z])] \\ &= aE[(X - E[X])(Z - E[Z])] + bE[(Y - E[Y])(Z - E[Z])] \\ &= a \text{Cov}(X, Z) + b \text{Cov}(Y, Z). \end{aligned}$$

3. Puisque  $E[a] = a$ ,

$$\text{Cov}(a, X) = E[(a - a)X] = E[0X] = 0.$$

4. On suppose que  $E[X] = E[Y] = 0$  quitte à remplacer les v.a. par leurs variables centrées. Alors,

$$\begin{aligned} \text{Var}(X + Y) &= E[(X + Y)^2] \\ &= E[X^2 + Y^2 + 2XY] \\ &= E[X^2] + E[Y^2] + 2E[XY] \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y). \end{aligned}$$

□

La covariance est une mesure de dépendance entre les variables  $X$  et  $Y$ . Pour illustrer cela, on considère un exemple :

**Exemple 5.7.** Soient  $A$  et  $B$  des événements (par exemple de la forme  $\{X \in S\}$  pour une v.a.  $X$  et  $S \subset \mathbb{R}$ ) et soient  $\mathbf{1}_A$  et  $\mathbf{1}_B$  leurs indicatrices. On calcule

$$\begin{aligned} E[\mathbf{1}_A] &= P(A), & E[\mathbf{1}_B] &= P(B) \\ E[\mathbf{1}_A \mathbf{1}_B] &= P(A \cap B) \\ \text{Cov}(\mathbf{1}_A, \mathbf{1}_B) &= E[\mathbf{1}_A \mathbf{1}_B] - E[\mathbf{1}_A]E[\mathbf{1}_B] = P(A \cap B) - P(A)P(B). \end{aligned}$$

Par conséquent,

$$\text{Cov}(\mathbf{1}_A, \mathbf{1}_B) = 0 \iff P(A \cap B) = P(A)P(B) \iff A \text{ et } B \text{ sont indépendants.}$$

Ceci conforte l'idée de la covariance comme une mesure de dépendance entre les v.a.

On peut même interpréter le signe de la covariance. Supposons que  $P(A) \neq 0$  et  $P(B) \neq 0$ , alors on peut également écrire

$$\text{Cov}(\mathbf{1}_A, \mathbf{1}_B) = P(B)(P(A | B) - P(A)) = P(A)(P(B | A) - P(B))$$

avec  $P(A | B) = P(A \cap B)/P(B)$  la probabilité de  $A$  conditionnellement à  $B$ . Cette écriture montre que

$$\text{Cov}(\mathbf{1}_A, \mathbf{1}_B) > 0 \iff P(A | B) > P(A) \iff P(B | A) > P(B).$$

Les deux dernières inégalités signifient que la réalisation d'un des deux événements augmente la chance que l'autre événement se réalise. On dit dans ce cas que les deux événements sont *positivement corrélés*. Dans le cas d'une inégalité dans l'autre sens, on dit qu'ils sont *négativement corrélés*.

On résume : Soient  $A$  et  $B$  deux événements. Alors,  $A$  et  $B$  sont

- *positivement corrélés* si  $P(A \cap B) > P(A)P(B)$  ( $\iff \text{Cov}(\mathbf{1}_A, \mathbf{1}_B) > 0$ )
- *négativement corrélés* si  $P(A \cap B) < P(A)P(B)$  ( $\iff \text{Cov}(\mathbf{1}_A, \mathbf{1}_B) < 0$ )
- *indépendantes* si  $P(A \cap B) = P(A)P(B)$  ( $\iff \text{Cov}(\mathbf{1}_A, \mathbf{1}_B) = 0$ ).

En vue du dernier exemple, on dit que deux v.a.  $X$  et  $Y$  sont

- *positivement corrélées* si  $\text{Cov}(X, Y) > 0$

- *négativement corrélées* si  $\text{Cov}(X, Y) < 0$
- *non corrélées* si  $\text{Cov}(X, Y) = 0$

On remarque que l'absence de corrélation est équivalente à  $E[XY] = E[X]E[Y]$ . On a alors l'implication

$$X \text{ et } Y \text{ indépendantes} \implies X \text{ et } Y \text{ non corrélées.}$$

La réciproque est fautive en général, mais elle est vraie dans certains cas particuliers (par exemple quand  $X$  et  $Y$  sont des v.a. de Bernoulli, cf exemple ci-dessus).

**Exemple 5.8.** Soit  $X$  une v.a. de loi symétrique avec  $E[X^4] < \infty$ , par exemple  $X \sim \mathcal{N}(0, 1)$ . On pose  $Y = X^2$  qui est de carré intégrable. Evidemment,  $X$  et  $Y$  ne sont en général pas indépendantes. En effet, on vérifie facilement que  $X$  et  $Y$  sont indépendantes si et seulement si  $X \sim \delta_0$  (calculer  $P(Y > \varepsilon^2 \mid |X| > \varepsilon)$  pour tout  $\varepsilon > 0$ ). Par contre,

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = E[X^3] - E[X]E[X^2] = 0,$$

car  $E[X^3] = E[X] = 0$  par la symétrie de  $X$ . Ceci donne un exemple où les v.a.  $X$  sont non corrélées mais dépendantes.

Pour quantifier la corrélation entre deux variables aléatoires, on introduit le *coefficient de corrélation*  $\rho(X, Y)$  comme suit :

$$\rho(X, Y) = \text{Cov}\left(\frac{X}{\sqrt{\text{Var}(X)}}, \frac{Y}{\sqrt{\text{Var}(Y)}}\right) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}.$$

**Proposition 5.9.** *Le coefficient de corrélation satisfait aux inégalités suivantes :*

$$-1 \leq \rho(X, Y) \leq 1.$$

Cette proposition sera une conséquence du résultat suivant :

**Lemme 5.10** (Inégalité de Cauchy–Schwarz). *Soient  $X, Y$  des v.a. de carrés intégrables. Alors,*

$$E[|XY|] \leq \sqrt{E[X^2]E[Y^2]}.$$

*Démonstration.* On peut supposer que  $X \geq 0$  et  $Y \geq 0$  et que  $E[X^2] = E[Y^2] = 1$ , sinon on remplace  $X$  et  $Y$  par  $X' = |X|/\sqrt{E[X^2]}$  et  $Y' = |Y|/\sqrt{E[Y^2]}$  (notons que les deux cotés de l'inégalité sont homogènes en  $X$  et  $Y$ ). On a alors,

$$0 \leq E[(X - Y)^2] = E[X^2 + Y^2 - 2XY] = E[X^2] + E[Y^2] - 2E[XY] = 2(1 - E[XY]).$$

Par conséquent,  $E[|XY|] = E[XY] \leq 1$ . □

*Démonstration de la proposition.* Par l'inégalité de Jensen,

$$|\text{Cov}(X, Y)| = |E[(X - E[X])(Y - E[Y])]| \leq E[|(X - E[X])(Y - E[Y])|].$$

Par l'inégalité de Cauchy–Schwarz,

$$E[|(X - E[X])(Y - E[Y])|] \leq \sqrt{E[(X - E[X])^2]E[(Y - E[Y])^2]} = \sqrt{\text{Var}(X) \text{Var}(Y)}.$$

Ces deux inégalités donnent  $|\rho(X, Y)| \leq 1$ . □

*Remarque 5.11.* La preuve ci-dessus montre aussi ce que nous avons admis en haut : que la covariance est toujours bien définie et finie pour des v.a. de carrés intégrables.

**Exemple 5.12.** Soient  $U, V, W$  des v.a. de carrés intégrables, indépendantes, et de variance égale à 1. On pose

$$X = U + aV, \quad Y = U + bW,$$

pour  $a, b \in \mathbb{R}$ . On peut voir  $X$  et  $Y$  comme des observations « bruitées » d'une même quantité  $U$ , les v.a.  $aV$  et  $bW$  modélisant le « bruit ». Par indépendance,

$$\text{Cov}(X, Y) = \text{Cov}(U, U) + \text{Cov}(U, bW) + \text{Cov}(aV, U) + \text{Cov}(aV, bW) = \text{Var}(U) = 1.$$

et  $\text{Var}(X) = \text{Var}(U) + \text{Var}(aV) = 1 + a^2$  et  $\text{Var}(Y) = 1 + b^2$ . Par conséquent,

$$\rho(X, Y) = \frac{1}{\sqrt{(1+a^2)(1+b^2)}}.$$

On observe :

- Les v.a.  $X$  et  $Y$  sont positivement corrélées.
- La corrélation est parfaite (c'est-à-dire  $\rho = 1$ ) si  $a = b = 0$ .
- La corrélation tend vers 0 quand  $|a| \rightarrow \infty$  ou  $|b| \rightarrow \infty$ , ce qui correspond à un bruit qui devient de plus en plus fort.

Il ressort des deux derniers exemples que le coefficient de corrélation mesure bien des relations *linéaires* entre les v.a., mais pas nécessairement des relations non-linéaires.

**Exemple 5.13.** On reprend l'exemple du test d'infection. On calcule

$$\text{Var}(X) = E[X^2] - E[X]^2 = P(X = 1) - P(X = 1)^2 = 0.1 - 0.1^2 = 0.09$$

$$\text{Var}(Y) = P(Y = 1) - P(Y = 1)^2 = 0.3 - 0.3^2 = 0.21$$

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = P(X = 1, Y = 1) - P(X = 1)P(Y = 1) = 0.1 - 0.03 = 0.07$$

$$\rho(X, Y) = \frac{0.07}{\sqrt{0.09 \times 0.21}} \approx 0.51.$$

Ce calcul montre que les v.a.  $X$  et  $Y$  sont sensiblement, mais pas parfaitement, corrélées.

**Exemple 5.14.** On reprend l'exemple de la loi uniforme sur le triangle. On calcule

$$E[X] = \int_0^1 x \times 2x \, dx = 2/3$$

$$E[X^2] = \int_0^1 x^2 \times 2x \, dx = 2/4 = 1/2$$

$$\text{Var}(X) = E[X^2] - E[X]^2 = 1/2 - 4/9 = 1/18.$$

Et pareil pour  $Y$ , car  $Y \stackrel{\text{loi}}{=} X$ . De plus,

$$\begin{aligned}
 E[XY] &= \int_0^1 \int_0^1 2xy \mathbf{1}_{x+y \geq 1} dx dy \\
 &= \int_0^1 x [y^2]_{1-x}^1 dx \\
 &= \int_0^1 x(1 - (1-x)^2) dx \\
 &= \int_0^1 (2x^2 - x^3) dx \\
 &= 2/3 - 1/4 \\
 &= 5/12.
 \end{aligned}$$

Par conséquent,

$$\begin{aligned}
 \text{Cov}(X, Y) &= E[XY] - E[X]E[Y] = 5/12 - (2/3)^2 = 15/36 - 16/36 = -1/36 \\
 \rho(X, Y) &= \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} = \frac{-1/36}{1/18} = -1/2.
 \end{aligned}$$

Les v.a.  $X$  et  $Y$  sont alors sensiblement mais pas parfaitement négativement corrélées. Ceci s'explique par le fait que si l'une des v.a. est petite, l'autre doit être grande pour que la somme soit supérieure à 1, il s'agit donc d'un biais « dans l'autre sens ».

# Chapitre 6

## Appendice : Toolbox.

### 6.1 Ensemble dénombrable

Rappelons d'abord la définition d'ensemble fini :

**Définition 6.1.** Un ensemble  $S$  est dit *fini* s'il est en bijection avec une partie finie de l'ensemble  $\mathbb{N}$  des entiers naturels. On peut alors sans perte de généralité prendre la partie de  $\mathbb{N}$  sous la forme  $\{1, \dots, n\}$ , et  $n$  s'appelle alors le cardinal de  $S$  ;

**Définition 6.2.** Une ensemble  $S$  est dit *dénombrable* s'il est en bijection avec l'ensemble  $\mathbb{N}$  des entiers naturels.

**Exemple 6.3.** Voici des exemples d'ensembles dénombrables : l'ensemble des entiers pairs  $2\mathbb{N}$ , ou celui des entiers impairs  $2\mathbb{N}+1$  ; un peu plus subtil, l'ensemble  $\mathbb{N}^2$  des couples d'entiers, puis celui  $\mathbb{Q} = \{p/q, p \in \mathbb{Z}, q \in \mathbb{N}^*\}$  des nombres rationnels ; en revanche l'ensemble  $\mathbb{R}$  des nombres réels n'est pas dénombrable<sup>1</sup>, ni partant, celui  $\mathbb{R} \setminus \mathbb{Q}$  des nombres irrationnels.

On appellera ensemble au plus dénombrable un ensemble fini ou dénombrable.

### 6.2 Familles sommables

**Définition 6.4.** Soit  $I$  un ensemble arbitraire. Une famille de nombre réels *positifs*  $\{x_i, i \in I\}$  indicée par  $I$  est dite sommable si

$$\sup_{K \subset I} \left\{ \sum_{i \in K} x_i \right\} < \infty$$

où  $K$  parcourt les sous-ensembles finis de  $I$ . Dans ce cas, on note  $\sum_{i \in I} x_i$  le supremum, et on appelle "somme de la famille des  $x_i$ ", ce supremum.

Dans le cas où  $I$  est dénombrable, il est naturel de se ramener à une somme indicée par les entiers naturels  $\mathbb{N}$  :

---

1. Hors programme : pour le voir, on commence par vérifier à l'aide de l'argument dit diagonal de Cantor<sup>2</sup> que l'ensemble  $\{0, 1\}^{\mathbb{N}}$  des suites à valeurs dans  $\{0, 1\}$  n'est pas dénombrable, puis on note que cet ensemble peut être mis en correspondance<sup>3</sup> avec l'ensemble  $[0, 1]$  via le développement dyadique  $x = \sum_{i \geq 1} x_i 2^{-i}$  d'un nombre réel  $x \in [0, 1]$ .

**Proposition 6.5.** Soit  $\phi : \mathbb{N} \rightarrow I$  une bijection, la famille  $\{x_i, i \in I\}$  est sommable ssi la suite des sommes partielles  $(\sum_{k=0}^n x_{\phi(k)})_n$  est bornée, et la somme de la famille (telle que précédemment définie) coïncide avec la limite de la suite des sommes partielles :

$$\sum_{i \in I} x_i = \lim_{n \rightarrow \infty} \sum_{k=0}^n x_{\phi(k)}.$$

Le point remarquable dans cette proposition est que la limite dans le membre de droite ne dépend pas de l'ordre des termes choisi au travers de  $\phi$ . On propose de démontrer ce fait dans l'exercice suivant :

**Exercice 6.6.** Soit  $(a_k)_{k \in \mathbb{N}}$  une suite de nombres réels positifs qui forme une famille sommable et  $\sigma, \nu : \mathbb{N} \rightarrow \mathbb{N}$  deux bijections.

1. Soit  $n \in \mathbb{N}$ . Montrer qu'il existe  $m = m(n) \in \mathbb{N}$  tel que  $\sum_{k=0}^m a_{\sigma_k} \geq \sum_{k=0}^n a_{\nu_k}$ .
2. En déduire que  $\sum_{k \geq 0} a_{\sigma_k} \geq \sum_{k \geq 0} a_{\nu_k}$ , et conclure à l'égalité de ces deux nombres.

On peut, toujours dans le cas  $I$  dénombrable, relâcher l'hypothèse de positivité au prix d'une hypothèse d'absolue convergence. On définit ainsi :

**Définition 6.7.** Soit  $I$  un ensemble arbitraire. Une famille de nombre réels (non nécessairement positifs)  $\{x_i, i \in I\}$  indicée par  $I$  est dite sommable si la famille de nombre réels positifs  $\{|x_i|, i \in I\}$  est sommable.

Bien entendu, il revient au même de demander à ce que les deux familles de nombre réels positifs  $\{(x_i)_+, i \in I\}$  (les parties positives) et  $\{(x_i)_-, i \in I\}$  (les parties négatives) soient sommables. On rappelle les définitions :  $x_+ := \max x, 0$ , et  $x_- := \max -x, 0$ .

**Définition 6.8.** Soit une famille  $\{x_i, i \in I\}$  de nombre réels sommable. On appelle somme de la famille  $\{x_i, i \in I\}$ , et on note  $\sum_{i \in I} x_i$  la différence :

$$\sum_{i \in I} x_i := \sum_{i \in I} (x_i)_+ - \sum_{i \in I} (x_i)_-$$

On a alors l'analogie de la proposition précédente (à vérifier) :

**Proposition 6.9.** Soit  $I$  un ensemble dénombrable, et  $\phi : \mathbb{N} \rightarrow I$  une bijection. La suite des sommes partielles  $(\sum_{k=0}^n x_{\phi(k)})_n$  converge, de limite indépendante de  $\phi$  : on l'appelle encore la somme de la famille.

**Exercice 6.10.** Soit  $(a_k)_{k \in \mathbb{N}}$  suite de nombres réels telle que la suite  $(|a_k|)_{k \in \mathbb{N}}$  forme une famille sommable.

1. Montrer que  $\sum_{k=0}^n a_k$  converge quand  $n \rightarrow \infty$
2. Soit  $\phi : \mathbb{N} \rightarrow \mathbb{N}$  une bijection. Montrer que  $\sum_{k=0}^n a_k$  converge quand  $n \rightarrow \infty$  vers la même limite.

L'exemple suivant montre qu'on ne peut se passer de l'hypothèse d'absolue convergence lors de la manipulation de nombres signés : soit  $(x_k)_{k \in \mathbb{N}}$  une famille de réels telle que

1.  $\lim_k x_k = 0$ ,
2.  $\sum (x_k)_+ = \sum (x_k)_- = +\infty$ .

Alors pour tout  $\ell \in \mathbb{R}$ , on peut trouver une bijection  $\phi : \mathbb{N} \rightarrow \mathbb{N}$  (qui dépend bien sûr de  $\ell$ ) telle que la suite des sommes partielles  $(\sum_{k=0}^n x_{\phi(k)})_n$  converge vers  $\ell$  (exercice!).



### 6.3 Séries de fonctions et intégrales à paramètre

Référence pour cette section : Notes de Thierry Ramond du cours « Analyse et Convergence 2, » disponible sur Dokeos.

Nous nous intéressons à des quantités ayant une des formes suivantes :

$$\begin{array}{c|c} \sum_{k \in \mathbb{N}} f_{n,k} & \int_{-\infty}^{\infty} f_n(y) dy \\ \hline \sum_{k \in \mathbb{N}} f_k(x) & \int_{-\infty}^{\infty} f(x, y) dy \end{array}$$

Dans ce tableau,  $n, k \in \mathbb{N}$ ,  $y \in \mathbb{R}$  et  $x \in I$  avec  $I$  un intervalle. On peut aussi considérer d'autres intégrales telles que  $\int_{-\infty}^a$  ou  $\int_a^{\infty}$  pour  $a, b \in \mathbb{R}$  mais on peut se ramener au cas ci-dessus en posant  $f_n(y) = 0$  ou  $f(x, y) = 0$  en dehors du domaine d'intégration. Pour les deux quantités de la première ligne du tableau, on souhaite alors avoir des conditions sous lesquelles on a le droit d'échanger le signe somme ou intégrale et le signe  $\lim_{n \rightarrow \infty}$ , c'est-à-dire :

$$\text{Q : a-t-on } \lim_{n \rightarrow \infty} \sum_{k \in \mathbb{N}} f_{n,k} = \sum_{k \in \mathbb{N}} \lim_{n \rightarrow \infty} f_{n,k} \quad \text{ou} \quad \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f_n(y) dy = \int_{-\infty}^{\infty} \lim_{n \rightarrow \infty} f_n(y) dy \quad ?$$

De manière analogue, pour les deux quantités de la deuxième ligne du tableau, on souhaite avoir des conditions sous lesquelles on a le droit d'échanger le signe somme ou intégrale et le signe  $\lim_{x \rightarrow x_0}$ , ou, autrement dit,

$$\text{Q' : les quantités } \sum_{k \in \mathbb{N}} f_k(x) \quad \text{ou} \quad \int_{-\infty}^{\infty} f(x, y) dy \quad \text{sont-elles continues en } x ?$$

Pour ces deux dernières quantités, on souhaite également savoir si on a le droit de dériver (par rapport à  $x$ ) sous les signes somme ou intégrale, c'est-à-dire :

$$\text{Q'' : a-t-on } \frac{d}{dx} \sum_{k \in \mathbb{N}} f_k(x) = \sum_{k \in \mathbb{N}} \frac{d}{dx} f_k(x) \quad \text{ou} \quad \frac{d}{dx} \int_{-\infty}^{\infty} f(x, y) dy = \int_{-\infty}^{\infty} \frac{\partial}{\partial x} f(x, y) dy \quad ?$$

La clé pour répondre à ces questions est la notion de *convergence normale*. On l'énonce séparément pour les quantités de la première colonne et celles de la deuxième colonne du tableau ci-dessus.

**Définition 6.11** (Convergence normale). On dit que les séries  $\sum_{k \in \mathbb{N}} f_{n,k}$  ou  $\sum_{k \in \mathbb{N}} f_k(x)$  convergent normalement s'il existe une suite  $(g_k)_{k \in \mathbb{N}}$  de nombres positifs telle que

1.  $\sum_{k \in \mathbb{N}} g_k < \infty$  et
2.  $|f_{n,k}| < g_k$  pour tout  $n \in \mathbb{N}$  (ou  $|f_k(x)| < g_k$  pour tout  $x \in I$ ).

De manière analogue, on dit que les intégrales  $\int_{-\infty}^{\infty} f_n(y) dy$  ou  $\int_{-\infty}^{\infty} f(x, y) dy$  convergent normalement s'il existe une fonction (positive)  $g : \mathbb{R} \rightarrow \mathbb{R}_+$  telle que

1.  $\int_{-\infty}^{\infty} g(y) dy < \infty$  et
2.  $|f_n(y)| < g(y)$  pour tout  $n \in \mathbb{N}$  (ou  $|f(x, y)| < g(y)$  pour tout  $x \in I$ ).

Les théorèmes suivants donnent alors les réponses aux questions précédentes :

**Theorème 6.12.** *Supposons qu'une des quantités du tableau ci-dessus converge normalement. Alors la réponse à la question  $Q$  ou  $Q'$  concernant cette quantité est « oui ».*

**Theorème 6.13.** 1. *Supposons que  $f_k : I \rightarrow \mathbb{R}$  est dérivable pour tout  $k$ . Supposons de plus que*

(a)  $\sum_{n \in \mathbb{N}} f_k(x)$  converge simplement (i.e. converge pour tout  $x \in I$ ), et

(b)  $\sum_{n \in \mathbb{N}} \frac{d}{dx} f_k(x)$  converge normalement.

Alors la réponse à la question  $Q'$  concernant  $\sum_{n \in \mathbb{N}} f_k$  est « oui ».

2. *Supposons que  $f(x, y) : I \times \mathbb{R} \rightarrow \mathbb{R}$  admet une dérivée partielle par rapport à  $x$  pour tout  $y$ . Supposons de plus que*

(a)  $\int_{-\infty}^{\infty} f(x, y) dy$  converge pour tout  $x \in I$ , et

(b)  $\int_{-\infty}^{\infty} \frac{\partial}{\partial x} f(x, y) dy$  converge normalement.

Alors la réponse à la question  $Q'$  concernant  $\int_{-\infty}^{\infty} f(x, y) dy$  est « oui ».

En probabilités, on rencontre assez souvent des séries entières, et les résultats précédents prennent alors une forme très simple. Rappelons notamment le résultat de dérivation terme à terme des séries entières.

Étant donné une suite  $(a_n)_{n \in \mathbb{N}} \in \mathbb{C}$  de nombres complexes<sup>4</sup>, la série entière associée est formellement notée  $\sum_{n \geq 0} a_n z^n$ . La somme de la série entière est alors la fonction qui à tout complexe  $z$  tel que  $\sum a_n z^n$  converge, associe

$$f(z) = \sum_n a_n z^n.$$

La première question est celle de la convergence de la somme. À cet effet, on introduit :

**Définition 6.14** (Rayon et disque de convergence). On appelle rayon de convergence de la série entière le réel  $R$  défini par

$$R = \sup\{r \geq 0 : \sum_n |a_n| r^n \text{ est bornée}\} \in [0, +\infty].$$

et on appelle disque de convergence de la série entière  $\sum a_n z^n$  le disque ouvert  $D_R = \{z \in \mathbb{C} : |z| < R\}$  de rayon  $R$  est de centre 0.

On peut alors fournir une réponse à la question précédente : une série entière converge *absolument* sur son disque de convergence. De plus la convergence est *normale* sur tout disque fermé inclus dans le disque de convergence, en particulier sur  $\overline{D}_r = \{z \in \mathbb{C} : |z| \leq r\}$ , avec  $r < R$ . Par définition, la série diverge grossièrement sur  $\mathbb{C} \setminus \overline{D}_r$ , puisque son terme général ne tend pas vers 0; enfin, la situation sur le cercle  $\overline{D}_r \setminus D_r$  est bien plus délicate et aucune ne possibilité ne peut être exclue<sup>5</sup>

Concernant la dérivation terme à terme, le cadre des séries entières est particulièrement commode comme le montre le théorème suivant :

**Theorème 6.15.** *La somme  $f(z) = \sum_n a_n z^n$  de la série entière est indéfiniment dérivable sur son disque de convergence  $D_R$ . Sa dérivée est la somme de la série dérivée terme à terme, qui a même rayon de convergence :*

$$\forall z \in D_R, f'(z) = \sum_{n=1}^{\infty} n a_n z^{n-1} = a_1 + 2a_2 z + \dots + n a_n z^{n-1} + \dots$$

4. nous aurons surtout besoin dans ce cours de coefficients réels, mais les preuves sont identiques

5. comparer le comportement de  $\sum_n z^n$  et  $\sum_n z^n/n^2$  au bord du disque de convergence.

Si  $R > 0$ , on a, pour tout  $n \in \mathbb{N}$ ,

$$a_n = \frac{f^{(n)}(0)}{n!}.$$

En particulier deux séries entières qui coïncident sur un intervalle ouvert contenant 0 sont égales, puisqu'on peut identifier la suite de leurs coefficients à l'aide de la formule précédente. On rappelle maintenant les développements en série entière les plus classiques :

$$\forall z \in \mathbb{C}, \quad |z| < 1, \quad \frac{1}{1-z} = \sum_{n \in \mathbb{N}} z^n \quad \text{et} \quad \forall z \in \mathbb{C}, \quad \exp(z) = \sum_{n \in \mathbb{N}} \frac{z^n}{n!}$$

Ayant mentionné le développement de l'exponentielle, on note celui du logarithme (de rayon de convergence 1 cette fois)

$$\forall z \in \mathbb{C}, \quad |z| < 1, \quad \ln(1+z) = \sum_{n=1}^{+\infty} (-1)^{n+1} \frac{z^n}{n} = z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} + \dots,$$

on pourra d'ailleurs retenir plutôt le DSE de  $z \mapsto -\ln(1-z)$  dont l'expression est particulièrement simple. Bien entendu, on peut aussi former cos et sin à partir des parties réelles et imaginaires de  $e^{ix}$  et on obtient :

$$\forall z \in \mathbb{C}, \quad \cos z = \sum_{n=0}^{+\infty} (-1)^n \frac{z^{2n}}{(2n)!} = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \dots$$

$$\forall z \in \mathbb{C}, \quad \sin z = \sum_{n=0}^{+\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!} = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \dots$$

On rappelle la formule du binôme de Newton

$$\forall z \in \mathbb{C}, \quad (1+z)^n = \sum_{k=0}^n \binom{n}{k} z^k = 1 + \sum_{k=1}^n \frac{n(n-1)\dots(n-k+1)}{k!} z^k$$

dont la dernière expression se généralise en fait pour tout  $\alpha \in \mathbb{R}$  (!) :

$$\forall z \in \mathbb{C}, \quad |z| < 1, \quad (1+z)^\alpha = 1 + \sum_{n=1}^{+\infty} \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!} z^n = 1 + \alpha z + \frac{\alpha(\alpha-1)}{2} z^2 + \dots,$$

de rayon de convergence 1. Cette formule est très puissante : noter que si l'on prend  $\alpha = -1$ , et  $-z$  à la place de  $z$  on retrouve le développement de  $(1-z)^{-1}$  donné précédemment. et plus généralement, si l'on prend pour  $\alpha$  un entier relatif négatif et que l'on substitue  $-z$  à  $z$  on obtient :

$$\forall z \in \mathbb{C}, \quad |z| < 1, \quad k \in \mathbb{N}^* \quad (1-z)^{-n} = 1 + \sum_{k=1}^{+\infty} \binom{n+k-1}{k} z^k$$

formule qu'on peut aussi retrouver par dérivation successive du développement de  $(1-z)^{-1}$ . Un dernier développement remarquable est celui d'arctan, qu'on retrouve en considérant par exemple sa dérivée :

$$\forall z \in \mathbb{C}, \quad |z| < 1, \quad \arctan z = \sum_{n=0}^{+\infty} (-1)^n \frac{z^{2n+1}}{2n+1} = z - \frac{z^3}{3} + \frac{z^5}{5} - \frac{z^7}{7} + \dots$$

## 6.4 Théorèmes de Fubini et Tonelli

La démonstration de ce théorème sera vue en cours d'intégration. Il est cependant fort probable qu'on ait besoin de l'utiliser en probabilités avant que vous ne l'ayez vu en intégration. On vous demandera donc un petit acte de foi. On s'intéresse à des quantités de la forme

$$\sum_{n \in \mathbb{N}} \left( \sum_{m \in \mathbb{N}} f_{n,m} \right), \quad \sum_{n \in \mathbb{N}} \left( \int_{-\infty}^{\infty} f_n(y) dy \right), \quad \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} f(x, y) dy \right) dx.$$

On souhaite savoir si on a le droit d'invertir les signes somme et/ou intégrale, c'est-à-dire, on souhaite répondre à la question suivante : *Les quantités ci-dessus sont-elles égales à (respectivement) :*

$$\sum_{m \in \mathbb{N}} \left( \sum_{n \in \mathbb{N}} f_{n,m} \right), \quad \int_{-\infty}^{\infty} \left( \sum_{n \in \mathbb{N}} f_n(y) dy \right), \quad \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} f(x, y) dx \right) dy ?$$

Une première réponse est apportée par le *théorème de Tonelli* :

**Théorème 6.16** (Tonelli, parfois appelé Fubini–Tonelli ou Fubini cas positif). *Supposons les nombres  $f_{n,m}$  sont positifs :  $f_{n,m} \geq 0$  pour tout  $n, m \in \mathbb{N}$ . Alors on a l'égalité dans  $[0, +\infty]$*

$$\sum_{n \in \mathbb{N}} \left( \sum_{m \in \mathbb{N}} f_{n,m} \right) = \sum_{m \in \mathbb{N}} \left( \sum_{n \in \mathbb{N}} f_{n,m} \right),$$

*ce qui signifie que les deux membres sont soit simultanément finis et égaux soit simultanément infinis.*

Ce théorème est très utile pour nous car les probabilités sont justement des quantités positives ! *Notons aussi que le théorème est vrai même si les séries/intégrales divergent, il n'y a donc pas besoin de vérifier au préalable qu'elles convergent.*

Si les termes ne sont pas positifs, on peut souvent appliquer le théorème suivant :

**Théorème 6.17** (Fubini). *Supposons qu'une des conditions équivalentes sont satisfaites :*

- $\sum_{n \in \mathbb{N}} (\sum_{m \in \mathbb{N}} |f_{n,m}|) < \infty$  ou
- $\sum_{m \in \mathbb{N}} (\sum_{n \in \mathbb{N}} |f_{n,m}|) < \infty$

*Alors on a l'égalité suivante :*

$$\sum_{n \in \mathbb{N}} \left( \sum_{m \in \mathbb{N}} f_{n,m} \right) = \sum_{m \in \mathbb{N}} \left( \sum_{n \in \mathbb{N}} f_{n,m} \right),$$

*c'est-à-dire que les deux séries convergent et ont la même limite.*

*Les résultats analogues pour les autres quantités sont également vrais.*

Notons que l'équivalence des deux conditions du théorème découle de Fubini cas positif.