

Probabilités continues¹

Pascal Maillard

15 décembre 2017

1. Notes du cours « Probabilités » (M316) de la troisième année de licence « Mathématiques et Interactions » à l'Université Paris-Sud

Table des matières

0	Préliminaires	5
0.1	Bibliographie conseillée	5
0.2	Bagage mathématique	5
0.2.1	Séries de fonctions et intégrales à paramètre	6
0.2.2	Théorèmes de Fubini et Tonelli	7
0.3	Rappels du cours de probas en L2	8
0.3.1	Probabilités	8
0.3.2	Variables aléatoires	9
0.3.3	Par la suite...	9
1	Variables aléatoires réelles	11
1.1	Définition et axiomes	11
1.2	Événements	13
1.2.1	Indépendance d'événements et probabilité conditionnelle.	14
1.3	Fonction de répartition	14
1.4	Classes de variables aléatoires	16
1.5	Quelques variables aléatoires discrètes importantes	18
1.5.1	Loi de Dirac	18
1.5.2	Loi uniforme	19
1.5.3	Loi de Bernoulli	19
1.5.4	Loi binomiale	20
1.5.5	Loi de Poisson	20
1.5.6	Loi géométrique	21
1.5.7	Autres lois discrètes	21
1.6	Quelques variables aléatoires continues importantes	22
1.6.1	Loi uniforme	22
1.6.2	Loi normale (ou gaussienne)	22
1.6.3	Loi exponentielle	23
1.6.4	Autres lois à densité	24
1.7	Loi d'une fonction d'une variable aléatoire	24
2	Espérance et moments	31
2.1	Espérance d'une variable aléatoire	31
2.2	Espérance d'une fonction d'une variable aléatoire	32
2.3	Moments et quantités associées	35
2.3.1	Espérance ou moyenne	35

2.3.2	Variance	35
2.3.3	Asymétrie	37
2.3.4	Kurtosis	37
2.4	Fonction génératrice des moments	38
2.5	Inégalités	40
2.5.1	Inégalité de Markov	40
2.5.2	Inégalité de Tchebychev	41
2.5.3	Inégalité de Jensen	41
3	Variables aléatoires indépendantes	43
3.1	Variables aléatoires simultanées	43
3.1.1	Variables aléatoires discrètes	44
3.1.2	Variables aléatoires conjointement à densité	45
3.1.3	Fonction de variables aléatoires simultanées	45
3.1.4	Espérance	45
3.2	Variables aléatoires indépendantes	46
3.3	Fonctions de variables aléatoires indépendantes	48
3.3.1	Maximum et minimum	48
3.3.2	Somme	49
4	Suites de variables aléatoires et théorèmes limites	53
4.1	Convergence de suites de variables aléatoires	53
4.1.1	Convergence en loi	53
4.1.2	Convergence en probabilité	54
4.1.3	Convergence en moyenne quadratique	56
4.1.4	Convergence presque sûre	57
4.2	Convergence de couples aléatoires	57
4.3	Loi des grands nombres	59
4.4	Théorème central limite	60
5	Variables aléatoires dépendantes	61
5.1	Lois marginales	61
5.2	Lois conditionnelles	63
5.3	Covariance et corrélation	64
5.4	Fonction d'un vecteur aléatoire	68

Chapitre 0

Préliminaires

0.1 Bibliographie conseillée

Pour le contenu du cours :

- ROSS, S., *Initiation aux probabilités*, Presses polytechniques et universitaires romandes
- BOGAERT, P., *Probabilités pour scientifiques et ingénieurs*, de boeck

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

0.2 Bagage mathématique

Les probabilités mettent à l'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), télescopage, changement d'indice, double sommes, séries entières, les sommes et séries fondamentales : somme sur les entiers, série géométrique, ...
- *fonctions* : continuité, définition de la dérivée, dérivation de produits et de composées, DLs, développement de Taylor, propriétés des fonctions classiques (polynomes, exponentielle, logarithme, fonctions trigonométriques), 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 anticipons déjà les théorèmes de Fubini et Tonelli que nous donnons ci-dessous. En ce qui concerne les intégrales multiples, nous allons patienter.

0.2.1 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^b 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 :

$$Q : \text{a-t-on } \lim_{n \rightarrow \infty} \sum_{k \in \mathbb{N}} f_{n,k} = \sum_{k \in \mathbb{N}} \lim_{n \rightarrow \infty} f_{n,k} \quad \text{ou} \quad \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f_n(y) dy = \int_{-\infty}^{\infty} \lim_{n \rightarrow \infty} f_n(y) dy \quad ?$$

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,

$$Q' : \text{les quantités } \sum_{k \in \mathbb{N}} f_k(x) \quad \text{ou} \quad \int_{-\infty}^{\infty} f(x, y) dy \quad \text{sont-elles continues en } x ?$$

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 :

$$Q'' : \text{a-t-on } \frac{d}{dx} \sum_{k \in \mathbb{N}} f_k(x) = \sum_{k \in \mathbb{N}} \frac{d}{dx} f_k(x) \quad \text{ou} \quad \frac{d}{dx} \int_{-\infty}^{\infty} f(x, y) dy = \int_{-\infty}^{\infty} \frac{\partial}{\partial x} f(x, y) dy \quad ?$$

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 0.1 (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 0.2. *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 0.3. 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 ».

0.2.2 Théorèmes de Fubini et Tonelli

On s'intéresse maintenant 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'intervertir 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* :

Theorème 0.4 (Tonelli, parfois appelé Fubini–Tonelli ou Fubini cas positif). *Supposons que les nombres $f_{n,m}$ sont positifs : $f_{n,m} \geq 0$ pour tout $n, m \in \mathbb{N}$. Alors*

$$\sum_{n \in \mathbb{N}} \left(\sum_{m \in \mathbb{N}} f_{n,m} \right) = \sum_{m \in \mathbb{N}} \left(\sum_{n \in \mathbb{N}} f_{n,m} \right).$$

Les résultats analogues pour les autres quantités sont également vrais.

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 :

Theorème 0.5 (Fubini). *Supposons qu'une des conditions équivalentes sont satisfaites :*

- $\sum_{n \in \mathbb{N}} \left(\sum_{m \in \mathbb{N}} |f_{n,m}| \right) < \infty$ ou
- $\sum_{m \in \mathbb{N}} \left(\sum_{n \in \mathbb{N}} |f_{n,m}| \right) < \infty$

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.

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 Ω qui contient tous les résultats possibles. Un élément de Ω est typiquement noté ω 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 Ω . 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 Ω 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 Ω est 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 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*¹. 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 changer notre façon de penser et de formaliser les expériences probabilistes.

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 Ω . Les avantages sont multiples :

- On peut modifier Ω (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 Ω , seulement des éléments qui nous intéressent.
- On peut plus facilement formuler des énoncés généraux ne dépendant pas de la définition précise d'un espace Ω .
- Dans le cas où Ω 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 Ω) mais plutôt partiellement à partir d'observations (les variables aléatoires).

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

1. 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

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 valeurs dans les nombres réels. Précédemment, nous avons toujours supposé qu'une variable aléatoire peut 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. Par contre, nous allons nous faciliter la vie : nous ne cherchons 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 devant 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 $x \in S_X$ 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 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 exactement un point. Puisque X correspond à un nombre choisi uniformément dans l'intervalle $[0, 1]$, on s'attend à ce que la probabilité que X est égal à un nombre donné ne dépend pas de ce nombre, autrement dit :

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

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 .

On doit alors trouver un autre moyen pour définir les probabilités $P(X \in A)$ pour des ensembles $A \subset \mathbb{R}$. L'observation importante est la suivante : ces probabilités ne peuvent être complètement arbitraires mais doivent obéir à certaines règles, par exemple, si $A, B \subset \mathbb{R}$ sont

disjoints, c'est-à-dire que leur intersection est 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, les intervalles finis, semi-infinis 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

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

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

— Si $A \subset B$, alors

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

— $P(X \in A) \in [0, 1]$ 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 ($B = \mathbb{R}$),

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

En effet, la propriété de monotonie et la formule pour $P(X \in B \setminus A)$ se montrent ainsi : pour $A \subset B$, on a $B = A \cup (B \setminus A)$, les deux ensembles étant disjoints. L'axiome 3 donne alors :

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

La propriété de monotonie 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 $A \in \mathbb{R}$ est une simple conséquence de la monotonie.

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 $P(X \in A)$ en A énoncée ci-dessus.

Cet axiome nécessite éventuellement une explication : avec seulement les axiomes 1 à 3, on aurait déjà une partie de l'égalité, à savoir l'inégalité $\ll \geq \gg$ (exercice). Le vrai énoncé de l'axiome 4 est alors qu'on a aussi l'inégalité inverse, à savoir $\ll \leq \gg$. Ceci correspond à dire 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 : 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 à 4 ci-dessus. Cette famille/fonction est appelée la *loi* de X . On dit également que X *suit* cette loi.

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 $P(X \in A) = \lim_{n \rightarrow \infty} \mathbb{P}(X \in A_n)$, 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 rationnelles \mathbb{Q} , ou n'importe quel ensemble dénombrable $D \subset \mathbb{R}$. Pour un tel ensemble, notons qu'avec la définition ci-dessus, on a $\mathbb{P}(X \in D) = \sum_{x \in D} \mathbb{P}(X = x)$.

1.2 Événements

Nous rappelons qu'un *événement* est un ensemble de résultats d'une expérience aléatoire ayant une caractéristique commune. Si le résultat de l'expérience fait partie de cette ensemble on dit alors que cet événement s'est réalisé. Par exemple, lors d'un lancer de dé, on pourrait s'intéresser à l'événement que le nombre tombé est un nombre pair. Si X est alors la v.a. qui représente le nombre tombé, 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 une certaine probabilité, à savoir $P(E) = P(X \in \{2, 4, 6\})$ (notez qu'on n'écrit plus les accolades dans cette dernière expression).

Dans ce cours, nous considérons seulement les événements qui sont définies 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, avec les notations évidentes. Par exemple :

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

1.2.1 Indépendance d'événements et probabilité conditionnelle.

On dit que deux événements E et F sont *indépendants* si

$$P(E \cap F) = P(E) \times P(F).$$

Ceci s'interprète comme quoi les deux événements n'ont aucune influence l'un sur l'autre, c'est-à-dire, sachant que l'un s'est produit, la probabilité que l'autre s'est produit reste la même. Ceci conduit à la notion de *probabilité conditionnelle* : si E est un événement de probabilité non nulle, alors pour tout autre événement F on définit la probabilité de F sachant E par :

$$P(F | E) = \frac{P(F \cap E)}{P(E)}.$$

Il découle alors immédiatement de la définition que E et F sont indépendants si et seulement si $P(F | E) = P(F)$.

On peut également introduire la notion de *loi conditionnelle* : Si X est une variable aléatoire et E un événement de probabilité non nulle, alors on note $P(X \in A | E)$ la probabilité de l'événement $\{X \in A\}$ sachant E , pour tout $A \subset \mathbb{R}$. On vérifie alors aisément (exercice) que l'application $A \mapsto P(X \in A | E)$ satisfait toujours aux axiomes 1 à 4 de la Section 1.1. On appelle cette application la *loi conditionnelle de X sachant E* .

La notion d'indépendance se généralise aussi à plusieurs événements, mais c'est un peu plus compliqué et ne sera pas traité ici.

1.3 Fonction de répartition

Nous avons vu dans la Section 1.1 la définition d'une variable aléatoire et de sa loi à partir des axiomes 1 à 4. La loi d'une variable aléatoire X était donnée par une famille de nombres réels indexées par les ensembles $A \subset \mathbb{R}$. Cette définition semble être assez encombrante 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 où X ne prend qu'un nombre fini ou dénombrable de valeurs), 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

$$F_X(x) = P(X \leq x) \text{ pour tout } x \in \mathbb{R}.$$

La fonction F_X s'appelle la *fonction de répartition de X* . 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 bien depuis le collège ! Evidemment, dans le cas d'une variable aléatoire discrète, la fonction de masse de sa loi peut avoir une expression plus simple que sa fonction de répartition, nous allons en voir des exemples plus tard.

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

Theorème 1.1.

1. La fonction de répartition F_X d'une variable aléatoire réelle 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 variable aléatoire réelle 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.

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 puisque la fonction F_X est croissante et bornée inférieurement (par 0), la limite $F_X(-\infty) = \lim_{x \rightarrow -\infty} F_X(x)$ existe. 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,

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

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 infinie 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 $F(x-) := \lim_{y \uparrow x} F(x)$.

- $A =]x, \infty[$ 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

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

- $A = [x, y]$, $A = [x, y[$ ou $A =]x, y[$: Similaire à $A =]x, y]$, avec éventuellement $F(y-)$ et $F(x-)$ à la place de $F(y)$ ou $F(x)$.
- $A = I_1 \cup \dots \cup I_n$, 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. □

1.4 Classes de variables aléatoires

Ayant établi la forme générale de la loi d'une v.a. réelle à travers sa fonction de répartition, il convient de distinguer plusieurs classes de v.a. à l'aide de leurs fonction de répartition. Pour ce faire, il convient d'introduire la notion *d'atome*. On dit que $x \in \mathbb{R}$ est un atome de masse

$m > 0$ de la loi de X si $P(X = x) = m$. Ceci peut être exprimé avec la fonction de répartition car pour tout $x \in \mathbb{R}$, on a l'égalité

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

Par conséquent, x est un atome de masse $m > 0$ de la loi de X si et seulement si x est un point de discontinuité de F_X et m est la hauteur du saut en x , c'est-à-dire $m = F_X(x) - F_X(x-)$. Il découle de la définition qu'une v.a. X n'a pas d'atomes si et seulement si sa fonction de répartition est continue. Notons que puisque la fonction F_X est croissante d'après le théorème 1.1, elle a au plus un nombre dénombrable de sauts (résultat classique en analyse) et donc au plus un nombre dénombrable d'atomes. Cela mérite d'être mis en avant :

Corollaire 1.2. *Une variable aléatoire X a au plus un nombre dénombrable d'atomes.*

Concentrons-nous maintenant sur deux classes de variables aléatoires, sans doute les plus importantes : les variables aléatoires *discrètes* et les v.a. *à densité*.

Variables aléatoires discrètes. On dit qu'une variable aléatoire réelle X est dite *discrète* si sa fonction de répartition est une fonction en escalier, c'est-à-dire si elle est constante entre deux sauts successifs. Cette notion coïncide avec la notion de variable aléatoire vue en L2. En effet, soit S l'ensemble des atomes de X , alors S est un ensemble fini ou dénombrable d'après le Corollaire 1.2. Puisque F_X est constante entre deux sauts successifs, la somme des hauteurs des sauts est égale à $F(+\infty) - F(-\infty) = 1$, autrement dit, $P(X \in S) = 1$. On dit à ce moment-là que X est *à valeurs dans S* ou que la loi de X est *portée par S* . X est alors une variable aléatoire discrète dans le sens du cours de L2. En particulier, sa loi est déterminée par la fonction de masse $x \mapsto P(X = x)$, car

$$P(X \in A) = \sum_{x \in A} P(X = x), \quad \text{pour } A \subset \mathbb{R},$$

où la somme se fait sur tous les $x \in A$ tels que $P(X = x) > 0$, c'est-à-dire sur tous les atomes qui sont contenus dans A . En particulier,

$$F_X(x) = \sum_{y \leq x} P(X = y),$$

avec la même convention pour la somme.

Variables aléatoires à densité. On dit qu'une variable aléatoire réelle X est *à densité* ou *continue*¹ si la fonction de répartition F_X admet la forme

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

pour une fonction intégrable $f_X : \mathbb{R} \rightarrow \mathbb{R}_+$. On appelle la fonction f_X la *densité* de X (ou de la loi de X). Elle joue le rôle analogue de la fonction de masse du cas discret. La loi de

1. Dans la littérature, ces v.a. sont aussi parfois appelées *absolument continues*, le terme « continue » ayant dans ce cas un sens plus général.

X s'exprime facilement à l'aide de la densité, par exemple, si $a, b \in \mathbb{R}$ avec $a \leq b$, alors par continuité de la fonction de répartition,

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

Aussi, la fonction F_X étant la primitive d'une fonction elle est continue et la loi est sans atomes. Autrement dit,

$$P(X = x) = 0 \text{ pour tout } x \in \mathbb{R},$$

et donc, pour tout $a, b \in \mathbb{R}$ avec $a \leq b$,

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

En général, on peut écrire pour un ensemble $A \subset \mathbb{R}$:

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

où $\mathbf{1}_A$ est la fonction indicatrice de l'ensemble A , c'est-à-dire

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

Notons que la densité f_X n'est pas unique : par exemple, en modifiant sa valeur en un point, la valeur de son intégrale est inchangée et donc la loi reste la même.

Les autres variables aléatoires. Quasiment toutes les variables aléatoires que nous allons rencontrer dans ce cours font partie des deux classes ci-dessus. Parmi les autres, on peut mentionner celles qui sont une combinaison de ces deux, cf TD. 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... »

1.5 Quelques variables aléatoires discrètes importantes

NB : Pour des détails pour 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

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

Ceci est équivalent (voir preuve ci-dessous) à

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

En particulier, la fonction de répartition est donnée par

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

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 , ou encore que X est presque sûrement constante (égale à x). Inversement, on peut voir n'importe quel nombre $x \in \mathbb{R}$ comme une variable aléatoire de loi de Dirac en x .

Voyons maintenant une preuve complète de la formule pour $P(X \in A)$ ci-dessus : on commence par décomposer

$$P(X \in A) = P(X \in A \cap \{x\}) + P(X \in A \setminus \{x\}).$$

Le premier terme est égal à $P(X \in \emptyset) = 0$ si $x \notin A$ et à $P(X = x) = 1$ si $x \in A$, donc il est égal à $\mathbf{1}_A(x)$. Le second terme est majoré par $P(X \in \mathbb{R} \setminus \{x\})$ par la monotonie de $P(X \in A)$ en A . Mais $P(X \in \mathbb{R} \setminus \{x\}) = 1 - P(X = x) = 1 - 1 = 0$, et donc $P(X \in A \setminus \{x\}) = 0$. Ceci montre bien que $P(X \in A) = \mathbf{1}_A(x)$.

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 $x \in S$,

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

ce qui implique pour tout $A \subset \mathbb{R}$,

$$P(X \in A) = \frac{\text{Card}(A \cap S)}{\text{Card}(S)}.$$

La fonction de répartition est

$$F_X(x) = \frac{\text{Card}(]-\infty, x] \cap S)}{\text{Card}(S)}.$$

La loi uniforme intervient dans de nombre 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 Bernoulli* : c'est une expérience aléatoire ayant deux issues possibles : *succès* ou *échec*. Si on pose alors $X = 1$ en cas de succès et $X = 0$ en cas d'échec, et si p est la probabilité de succès de l'expérience, alors X suit la loi de Bernoulli de paramètre p .

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

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

où on rappelle la définition du coefficient binomial :

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad \text{si } k \in \{0, \dots, n\}, \quad = 0 \text{ sinon.}$$

Notons que cela coïncide avec la loi de Bernoulli de paramètre p quand $n = 1$. La fonction de répartition n'admet pas d'expression plus simple que son expression générale à partir de la fonction de masse.

Une variable aléatoire X de loi binomiale de paramètres n et p représente le nombre de succès dans n répétitions indépendantes d'une expérience de Bernoulli ayant une probabilité de succès p .

1.5.5 Loi de Poisson

On dit que X suit la loi de Poisson de paramètre $\lambda \geq 0$ (symboliquement, $X \sim \text{Poi}(\lambda)$), 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.

La loi de Poisson de paramètre $\lambda \geq 0$ peut-être vu 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 :

Théorème 1.3 (Prokhorov, Stein, ...). *Pour tout $n \in \mathbb{N}$ et $p \in [0, 1]$,*

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

Ce théorème montre donc de manière très précise 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 petit².

2. On voit souvent dans la littérature la contrainte supplémentaire que $\lambda = np$ ne soit pas trop grand pour

1.5.6 Loi géométrique

On dit que X suit la loi géométrique de paramètre $p \in]0, 1]$ (symboliquement : $X \sim \text{Geo}(p)$), 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 fois que l'on doit répéter une expérience de Bernoulli de probabilité de succès p jusqu'au premier succès. 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 $(1 - p)^{k-1} \times p$.

La loi géométrique est aussi la seule loi d'une variable aléatoire à valeurs dans \mathbb{N}^* ayant la propriété suivante, dite la *propriété de perte de mémoire* :

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

Ceci peut se reformuler de la façon suivante : pour tout $k \in \mathbb{N}$, la loi de $X - k$ conditionnellement à l'événement $\{X > k\}$ est la même que celle de X .

1.5.7 Autres lois discrètes

Voici quelques autres lois discrètes que nous mentionnons brièvement :

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

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

Ceci correspond au nombre de répétitions d'une expérience de Bernoulli jusqu'au r -ième succès. En particulier, $\text{Geo}(p) = \text{BN}(1, p)$.

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

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

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 $X \sim \text{Hypergeo}(n, N, m)$.

que l'approximation soit bonne. Même si la preuve est plus facile avec cette contrainte, celle-ci n'est en fait pas nécessaire, comme le montre le théorème 1.3. Ceci dit, quand λ est grand, la loi de Poisson elle-même est souvent approchée par la loi normale (voir Section 1.6.2).

Loi zéta (ou de Zipf) On dit que X suit une loi zéta (ou de Zipf) de paramètre α 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 ζ la fonction zéta de Riemann.

1.6 Quelques variables aléatoires continues 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 admet la densité

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

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

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

Notons que la loi ne dépend que de a et b et non de la forme exacte de I , ainsi, on voit également la notation $\text{Unif}(a, b)$ pour cette loi.

Cette loi est l'analogie continue 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 .

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 admet 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 $\int_{-\infty}^{\infty} e^{-t^2} dt$ (cf exo 9 du TD1) 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 de plus $\sigma^2 = 1$ on dit que X est centrée réduite. La fonction de répartition de la loi normale centrée réduite $\mathcal{N}(0, 1)$ est notée Φ :

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

Des tables de valeurs de Φ seront distribuées en TD.

La forme de la densité f_X est la fameuse « courbe en forme de cloche » de Gauss. Elle est symétrique autour de μ , croissante à gauche de μ et décroissante à droite. Plus σ^2 est petit, plus la densité est resserrée autour de μ .

On peut facilement relier les lois normales pour les différentes valeurs des paramètres par la formule suivante (plus tard) :

$$\text{Si } N \sim \mathcal{N}(0, 1) \text{ et } \mu, \sigma \in \mathbb{R}, \sigma \neq 0, \text{ alors } \mu + \sigma N \sim \mathcal{N}(\mu, \sigma).$$

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

La loi normale est omniprésente en sciences. Ceci est dû au fait qu'elle sert comme bonne approximation pour nombre de lois dès lors qu'un paramètre devient grand. 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 1.4 (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 $X \sim \text{Bin}(n, p)$, alors quand n est grand et p fixé,

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

Une généralisation significative de ce théorème sera apportée par le *théorème central limite* que nous verrons plus tard.

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 admet la densité

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

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

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

En particulier, on peut retenir la formule suivante :

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

La loi exponentielle est l'analogie continu de la loi géométrique. Comme elle, elle possède une propriété de perte de mémoire :

$$P(X - x > y \mid X > x) = P(X > y), \quad \text{pour tout } x, y \geq 0.$$

Ceci peut se reformuler de la façon suivante : pour tout $x \geq 0$, la loi de $X - x$ conditionnellement à l'événement $\{X > x\}$ est la même que celle de X . Pour cette propriété, X peut représenter la durée de vie d'une pièce qui ne présente pas de phénomène de vieillissement : Le temps qu'il lui reste à vivre sachant qu'il vit encore au temps t ne dépend pas de t (en loi).

1.6.4 Autres lois à densité

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

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

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x},$$

où $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ est appelée la fonction gamma d'Euler. Sa fonction de répartition n'admet pas d'expression simple.

La loi gamma est une autre généralisation de la loi exponentielle. On retrouve cette dernière quand $\alpha = 1$. La loi gamma intervient dans de nombreuses applications.

Loi de Cauchy. On dit qu'une variable aléatoire X suit la loi de Cauchy, si elle admet la densité

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

Sa fonction de répartition vaut (exercice)

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x.$$

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.

1.7 Loi d'une fonction d'une variable aléatoire

Si $f : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction et X une variable aléatoire, alors $f(X)$ est encore un nombre qui décrit le résultat d'une expérience aléatoire, il est donc naturel de considérer $f(X)$ comme une nouvelle variable aléatoire. Plus généralement, si f 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 $f(X)$. Pour définir $f(X)$ comme une variable aléatoire dans le sens de la Section 1.1, il faut alors définir $P(f(X) \in A)$ pour tout ensemble A . La seule définition naturelle est la suivante : pour tout ensemble $A \subset \mathbb{R}$, réunion finie d'intervalles disjoints,

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

où $f^{-1}(A)$ est la préimage de A par l'application f . On suppose ici que l'ensemble $f^{-1}(A)$ est encore une réunion finie d'intervalles disjoints. Ceci est une hypothèse sur la fonction f qui est satisfaite par toute fonction « raisonnable ». On peut vérifier (exercice), que la fonction $A \mapsto P(f(X) \in A)$ ainsi définie satisfait encore aux axiomes 1 à 4.

Ayant défini la variable aléatoire $f(X)$ et sa loi nous souhaitons maintenant connaître sa fonction de répartition et/ou sa fonction de masse (quand il s'agit d'une variable aléatoire discrète) ou sa fonction de densité (quand il s'agit d'une variable aléatoire à densité).

Calcul de la fonction de répartition. Il est souvent commode d'essayer de calculer la fonction de répartition de $f(X)$. Par définition, on a

$$F_{f(X)}(x) = P(f(X) \in]-\infty, x]) = P(X \in f^{-1}(]-\infty, x])).$$

Puisque $f^{-1}(]-\infty, x])$ est par hypothèse réunion d'un nombre fini d'intervalles disjoints, cela permet d'exprimer la fonction de répartition de $f(X)$ en fonction de celle de X . Le cas le plus simple est celui où f est une fonction strictement croissante ou décroissante, donc injective, et il convient de le formuler comme un énoncé :

Corollaire 1.5. Soient $I, J \subset \mathbb{R}$ des intervalles, $f : I \rightarrow J$ une fonction continue, strictement croissante ou décroissante et surjective (donc bijective) et X une v.a. à valeurs dans I . Alors la fonction de répartition de $f(X)$ s'écrit :

$$F_{f(X)}(x) = \begin{cases} F_X(f^{-1}(x)), & \text{si } f \text{ est (strictement) croissante et } x \in J \\ 1 - F_X(f^{-1}(x)-), & \text{si } f \text{ est (strictement) décroissante } x \in J \\ 0, & \text{si } x \notin J \text{ et } x \leq \inf J \\ 1, & \text{si } x \notin J \text{ et } x \geq \sup J. \end{cases}$$

Démonstration. — f (strictement) croissante : Dans ce cas, $f(X) \leq x$ si et seulement si $X \leq f^{-1}(x)$, et donc

$$F_{f(X)}(x) = P(X \leq f^{-1}(x)) = F_X(f^{-1}(x)).$$

— f (strictement) décroissante : Dans ce cas, $f(X) \leq x$ si et seulement si $X \geq f^{-1}(x)$, et donc

$$F_{f(X)}(x) = P(X \geq f^{-1}(x)) = 1 - P(X < f^{-1}(x)) = 1 - F_X(f^{-1}(x)-).$$

Exemple 1.6. Soit X une v.a. positive, c'est-à-dire à valeurs dans \mathbb{R}_+ . On considère la fonction

$$f : \mathbb{R}_+ \rightarrow \mathbb{R}_+, \quad f(x) = x^2.$$

Alors f est une fonction continue, strictement croissante et surjective avec comme inverse $f^{-1}(x) = \sqrt{x}$. Par le Corollaire 1.5, on a alors

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

Si X est une v.a. générale, alors le Corollaire 1.5 ne s'applique pas. On reprend alors la définition : pour tout $x \geq 0$,

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

On résume :

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

Calcul de la fonction de masse pour des v.a. discrètes. Rappelons qu'une variable aléatoire est dite discrète si elle ne prend qu'un nombre au plus dénombrable de valeurs. Sous quelles conditions, $f(X)$ est-elle discrète? Deux conditions sont les plus courantes :

- Si f ne prend qu'un nombre fini ou dénombrable de valeurs et X quelconque.
- Si X est une v.a. discrète et f quelconque.

Si une de ces deux conditions est vérifiée, on peut alors essayer de calculer directement la fonction de masse de $f(X)$. Ceci est souvent plus commode que de calculer la fonction de répartition. On a par définition, pour tout $x \in \mathbb{R}$,

$$P(f(X) = x) = P(X \in f^{-1}(x)).$$

On en déduit les formules suivantes selon que X est discrète ou à densité :

Corollaire 1.7. *Supposons que $f(X)$ est une v.a. discrète (par exemple si une des deux conditions suivantes sont vérifiées). Alors la fonction de masse de $f(X)$ s'exprime comme suit :*

- Si X est discrète,

$$P(f(X) = x) = \sum_{y: f(y)=x} P(X = y).$$

- Si X admet une densité f_X ,

$$P(f(X) = x) = \int_{f^{-1}(\{x\})} f_X(y) dy.$$

Exemple 1.8. $f: \mathbb{R} \rightarrow \{0, 1\}$, $f(x) = \mathbf{1}_{x \geq 0}$, X une v.a. quelconque. La fonction f ne prend alors que deux valeurs, 0 ou 1. Par conséquent, la v.a. $f(X) = \mathbf{1}_{X \geq 0}$ est discrète. Sa fonction de masse est donnée par :

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

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

Exemple 1.9. $f: \mathbb{R}_+ \rightarrow \mathbb{R}_+$, $f(x) = \min(x, N)$ (avec $N \in \mathbb{N}$), X une v.a. à valeurs dans \mathbb{N} . Puisque X est une v.a. discrète, $f(X)$ 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 1.10. $f:]0, 1[\rightarrow \mathbb{N}$, $f(x) = \lfloor Nx \rfloor$ (avec $N \in \mathbb{N}$), $X \sim \text{Unif}(0, 1)$. La fonction f est à valeurs dans $\{0, \dots, N-1\}$, la v.a. $f(X)$ 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 $\{0, \dots, N-1\}$.

Calcul de la densité pour des v.a. à densité. Supposons que X soit une v.a. à densité. Alors sous certaines conditions sur f , $f(X)$ sera également une v.a. à densité et on peut alors calculer facilement la densité de $f(X)$ en fonction de la densité de X et de (la dérivée de) la fonction f . Nous allons d'abord énoncer et démontrer un théorème avec des hypothèses un peu restrictives, puis un théorème avec des hypothèses plus générales (sans preuve) :

Théorème 1.11. *Supposons que X admet une densité f_X et est à valeurs dans un intervalle $I \subset \mathbb{R}$. Supposons en plus que f satisfait aux hypothèses suivantes :*

- $f \in C^1(I)$
- f est strictement croissante ou strictement décroissante (en particulier, f est injective)
- $f'(x) = 0$ pour un nombre fini de $x \in I$.

Alors $f(X)$ est encore une v.a. à densité donnée par

$$f_{f(X)}(x) = \begin{cases} f_X(f^{-1}(x)) |(f^{-1})'(x)|, & \text{si } x \in f(I) \text{ et } |(f^{-1})'(x)| \text{ existe.} \\ 0, & \text{sinon.} \end{cases},$$

Remarque 1.12. On rappelle que $(f^{-1})'(x) = \frac{1}{f'(f^{-1}(x))}$, dès lors que $f'(f^{-1}(x)) \neq 0$. La condition « $|(f^{-1})'(x)|$ existe » dans le théorème ci-dessus signifie alors que $f'(f^{-1}(x)) \neq 0$.

Démonstration. On peut supposer que I est un intervalle ouvert, car la probabilité que X prenne valeur en un des deux bords de l'intervalle est nulle puisque X est à densité. De plus, pour simplifier on suppose que $f'(x) \neq 0$ pour tout $x \in I$, sinon on découpe l'intervalle I en les points où f' s'annule et on traite chaque morceau séparément. Par le théorème des fonctions implicites, f^{-1} est alors de classe C^1 sur $f(I)$.

On considère d'abord le cas où f est croissante, donc $f'(x) > 0$ pour tout $x \in I$. Pour tout $x \in f(I)$, on calcule alors la fonction de répartition comme suit :

$$\begin{aligned} F_{f(X)}(x) &= F_X(f^{-1}(x)) && \text{par le Corollaire 1.5} \\ &= \int_{\inf I}^{f^{-1}(x)} f_X(y) dy \\ &= \int_{\inf f(I)}^x f_X(f^{-1}(z)) (f^{-1})'(z) dz && \text{chgt de var. } z = f(y), dy = (f^{-1})'(z) dz. \end{aligned}$$

Ceci montre que $f(X)$ admet la densité valant $f_X(f^{-1}(x)) |(f^{-1})'(x)|$ sur $f(I)$ et 0 sinon.

Le cas où f est décroissante est similaire. \square

Règle mnémotechnique. Pour mémoriser la formule du théorème 1.11, il convient d'utiliser la règle mnémotechnique suivante. On part de l'expression

$$f_X(x) dx$$

On effectue alors un changement de variable $z = f(x)$, $dx = (f^{-1})'(z) dz$, comme si l'expression ci-dessus apparaissait dans une intégrale. Sauf qu'on prend la valeur absolue de la dérivée pour contrer le fait que les bornes de l'intégrale s'inversent quand la fonction f est décroissante. Ceci donne alors

$$f_X(f^{-1}(z)) |(f^{-1})'(z)| dz$$

Ceci est la densité recherchée. On peut bien sûr aussi l'écrire

$$f_X(f^{-1}(z)) \frac{1}{|f'(f^{-1}(z))|} dz.$$

Heuristique. Considérons encore l'expression

$$f_X(x) dx$$

On peut s'imaginer cela comme une répartition de masse : une masse de 1 kg est répartie sur \mathbb{R} de telle façon que l'intervalle infinitésimal de longueur dx autour de x porte $f_X(x) \times dx$ kg. On déplace alors la masse selon une fonction f : la masse qui se trouvait en x se trouve désormais en $z = f(x)$. Combien de masse se trouve-t-il dans un intervalle I_z de longueur infinitésimale dz autour de z ? Considérons l'application inverse f^{-1} . Celle-ci dilate l'espace autour de z par un facteur $|(f^{-1})'(z)|$, si bien que la masse dans l'intervalle I_z est la masse qui était auparavant dans un intervalle de longueur $|(f^{-1})'(z)| \times dz$ autour de $x = f^{-1}(z)$. L'intervalle I_z contient donc

$$f_X(f^{-1}(z)) \times |(f^{-1})'(z)| \times dz$$

kg de masse.

Exemple 1.13. $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = ax + b$, $a \neq 0$, $b \in \mathbb{R}$. Alors f est strictement monotone (croissante quand $a > 0$ et décroissante quand $a < 0$). De plus, $f^{-1}(x) = (x - b)/a$, donc $(f^{-1})'(x) = 1/a$ pour tout $x \in \mathbb{R}$. Le théorème 1.11 donne alors

$$f_{aX+b}(x) = \frac{1}{|a|} f_X\left(\frac{x-b}{a}\right).$$

L'effet des valeurs a et b sur la densité est alors le suivant : Si $a > 0$ mais $a < 1$, la densité est resserrée vers l'origine, alors que si $a > 1$, elle est dilatée. Si $a < 0$, alors la densité est réfléchie à l'axe verticale, puis resserrée ou dilatée selon que $|a| < 1$ ou > 1 . Puis, la densité est translatée de b , donc vers la droite si $b > 0$ et vers la gauche si $b < 1$.

Exemples importants :

- Si $I \subset \mathbb{R}$ intervalle, $U \sim \text{Unif}(I)$ et $a, b \in \mathbb{R}$, alors $aU + b \sim \text{Unif}(aI + b)$, où $aI + b = \{ax + b : x \in I\}$.
- Si $X \sim \mathcal{N}(0, 1)$, $\mu, \sigma \in \mathbb{R}$, $\sigma \neq 0$, alors $\mu + \sigma X \sim \mathcal{N}(\mu, \sigma^2)$. En particulier, $-X \sim \mathcal{N}(0, 1)$. Inversement, si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors $(X - \mu)/\sigma \sim \mathcal{N}(0, 1)$.
- Si $X \sim \text{Exp}(\lambda)$, $t > 0$, alors $tX \sim \text{Exp}(\lambda/t)$.
- Si $X \sim \Gamma(\alpha, \beta)$, $t > 0$, alors $tX \sim \Gamma(\alpha, \beta/t)$.

Exemple 1.14. On reprend l'exemple $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, $f(x) = x^2$. L'inverse est alors $f^{-1}(x) = \sqrt{x}$ et $(f^{-1})'(x) = \frac{1}{2\sqrt{x}}$ pour $x > 0$. Par conséquent, une densité de la v.a. X^2 est

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

Généralisation du théorème. On peut facilement assouplir l'hypothèse que f est monotone dans le théorème 1.11. Cela donne le théorème suivant, dont la preuve est laissée en exercice :

Théorème 1.15. *Supposons que X admet une densité f_X et est à valeurs dans $E \subset \mathbb{R}$, réunion finie d'intervalles disjoints. Supposons en plus que f satisfait aux hypothèses suivantes :*

- $f \in C^1(E)$
- il existe $N \in \mathbb{N}$ tel que $\text{Card}(f^{-1}(\{x\})) \leq N$ pour tout x ,
- $f'(x) = 0$ pour un nombre fini de $x \in E$.

Alors $f(X)$ est encore une v.a. à densité donnée par

$$f_{f(X)}(x) = \sum_{y \in E: f(y)=x, f'(y) \neq 0} f_X(y) \frac{1}{|f'(y)|}.$$

Exemple 1.16. On reprend l'exemple $f(x) = x^2$, avec cette fois-ci X une v.a. à densité sur \mathbb{R} . Alors

$$f^{-1}(\{x\}) = \begin{cases} \{\pm\sqrt{x}\}, & x > 0 \\ \{0\}, & x = 0 \\ \emptyset, & x < 0. \end{cases}$$

De plus, $|f'(y)| = 2|y|$ pour tout $y \in \mathbb{R}$. Le théorème 1.15 donne alors

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

□

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. On s'attend alors à ce que cette proportion approche $1/2$ quand n tend vers l'infini. 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 suivant 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, cette constante serait d'un sens la moyenne de la variable aléatoire X . Ceci est en effet vrai et on appelle cette moyenne *l'espérance* de la variable aléatoire X , notée $E[X]$. La définition de cette espérance (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 entre autres de définir les *moments* d'une variable aléatoire qui permettent entre autres de décrire certains aspects de sa loi. Elle est aussi utilisée pour définir la *fonction génératrice des moments*, qui fournit une autre façon de décrire la loi de la variable aléatoire, similaire à la fonction de répartition.

Nous verrons tous ces concepts dans ce chapitre.

2.1 Espérance d'une variable aléatoire

Soit X une variable aléatoire discrète ou à densité. Alors *l'espérance* de X , notée $E[X]$, est définie par

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

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.

Exemple 2.1. 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 2.2. Soit $X \sim \text{Geo}(p)$, $p \in]0, 1]$. Alors

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

cette série ayant été calculée dans l'Exercice 2 de la première feuille de TD.

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 2.3. 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 $(a+b)/2$.

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

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

cette intégrale ayant été calculée dans l'Exercice 1 de la première feuille de TD.

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

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. Soit X une variable aléatoire discrète ou à densité et f une fonction. Alors l'espérance de $f(X)$, notée $E[f(X)]$, est définie par

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

Cette définition généralise celle de $E[X]$ qu'on retrouve en prenant f l'application identité $f(x) = x$. L'espérance $E[f(X)]$ est encore intuitivement la moyenne pondérée des valeurs que $f(X)$ peut prendre, les poids étant les probabilités que ces valeurs soient prises.

La définition de $E[f(X)]$ ci-dessus devrait susciter une question. Car quand $f(X)$ est-elle même discrète ou continue, en appliquant la première définition de l'espérance à la variable aléatoire $Y = f(X)$, on devrait obtenir

$$E[f(X)] = \begin{cases} \sum yP(f(X) = y) & \text{si } f(X) \text{ est discrète} \\ \int_{-\infty}^{+\infty} yf_{f(X)}(y) dy & \text{si } f(X) \text{ admet la densité } f_{f(X)}. \end{cases} \quad (2.2)$$

La question est donc : Les deux expressions de $E[f(X)]$ dans (2.1) et (2.2) sont-elles équivalentes ? Autrement dit, l'espérance $E[f(X)]$ est-elle *bien définie* ? Le théorème suivant dit que c'est effectivement le cas :

Théorème 2.5. *Les deux expressions de l'espérance dans (2.1) et (2.2) sont équivalentes.*

Démonstration. Considérons d'abord le cas où X est discrète. Alors $f(X)$ l'est aussi et le Corollaire 1.7 donne

$$P(f(X) = y) = \sum_{x:f(x)=y} P(X = x).$$

Par conséquent,

$$\begin{aligned} \sum_y yP(f(X) = y) &= \sum_y y \sum_{x:f(x)=y} P(X = x) \\ &= \sum_y \sum_{x:f(x)=y} f(x)P(X = x) \\ &= \sum_x f(x)P(X = x), \end{aligned}$$

car dans la dernière double somme, chaque x apparaît exactement une fois.

On considère maintenant le cas où X est à densité. On suppose de plus que f est C^1 et de dérivée $f'(x) > 0$ pour tout $x \in \mathbb{R}$ et on admet le théorème dans le cas général. D'après le Théorème 1.11, on a alors

$$\begin{aligned} \int_{-\infty}^{+\infty} yf_{f(X)}(y) dy &= \int_{-\infty}^{+\infty} yf_X(f^{-1}(y))(f^{-1})'(y) dy \\ &= \int_{-\infty}^{+\infty} f(x)f_X(x) dx, \end{aligned}$$

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

Exemple 2.6. Soit f une fonction affine, $f(x) = ax + b$, $a, b \in \mathbb{R}$. Soit X une variable aléatoire discrète ou à densité. Si X est discrète, par (2.1),

$$E[aX + b] = \sum_x (ax + b)P(X = x) = a \sum_x xP(X = x) + b \sum_x P(X = x) = aE[X] + b,$$

car $\sum_x P(X = x) = 1$. De même, si X est à densité, par (2.1),

$$E[aX + b] = \int_{-\infty}^{+\infty} (ax + b)f_X(x) dx = a \int_{-\infty}^{+\infty} xf_X(x) dx + b \int_{-\infty}^{+\infty} f_X(x) dx = aE[X] + b,$$

car $\int_{-\infty}^{+\infty} f_X(x) dx = 1$. L'espérance est alors une application linéaire : pour toute variable aléatoire X et tout $a, b \in \mathbb{R}$,

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

on appelle cette propriété plus simplement *la linéarité de l'espérance*. De manière analogue on montre que si f_1, \dots, f_n sont des fonctions, alors

$$E[f_1(X) + \dots + f_n(X)] = E[f_1(X)] + \dots + E[f_n(X)],$$

ce qu'on appelle encore la *linéarité de l'espérance*.

Exemple 2.7. Une variable aléatoire X est dite symétrique, si X et $-X$ sont égales en loi. Soit X une telle variable aléatoire et $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction impaire, c'est-à-dire $f(-x) = -f(x)$ pour tout $x \in \mathbb{R}$. Alors

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

Par conséquent,

$$E[f(X)] = 0.$$

En particulier, $E[X] = 0$ (utiliser $f(x) = x$).

Exemple 2.8. Soit X une variable aléatoire. On définit $X_+ = \max(X, 0)$ et $X_- = \max(-X, 0)$, si bien que $X = X_+ - X_-$. Par linéarité de l'espérance,

$$E[X] = E[X_+] - E[X_-].$$

On appelle X_+ (X_-) la *partie positive (négative) de X* . X_+ et X_- sont des variables aléatoires positives, c'est-à-dire à valeurs dans \mathbb{R}_+ .

Considérations techniques sur l'existence de l'espérance La définition de l'espérance $E[X]$ fait intervenir des séries et des intégrales indéfinies, elle n'existe donc pas toujours *a priori*. Si X est une v.a. positive, alors elle est toujours bien définie (par exemple comme limite croissante des sommes partielles ou des intégrales de $-L$ à L), mais peut être égale à $+\infty$. En revanche, si X peut prendre des valeurs négatives, alors on écrit $X = X_+ - X_-$, où $X_+ = \max(X, 0)$ et $X_- = \max(-X, 0)$ sont les parties positives et négatives de X définies ci-dessus. On définit alors séparément $E[X_+]$ et $E[X_-]$, puis $E[X]$ par la différence de ces deux quantités pourvu que les deux ne sont pas infinies, dans quel cas l'espérance n'est pas définie. On résume : **l'espérance est bien définie pourvu qu'au moins une des deux quantités $E[X_+]$ ou $E[X_-]$ est finie, dans quel cas on définit $E[X] = E[X_+] - E[X_-]$.**

Ceci se généralise bien sûr à l'espérance d'une fonction d'une variable aléatoire : Si on note $f_+ = \max(f, 0)$ et $f_- = \max(-f, 0)$, alors $E[f(X)]$ est définie si $E[f_+(X)]$ ou $E[f_-(X)]$ est finie, dans quel cas on pose $E[f(X)] = E[f_+(X)] - E[f_-(X)]$.

On connaît déjà une loi dont l'espérance n'est pas bien définie, c'est la *loi de Cauchy*. En effet, si X suit la loi de Cauchy, alors

$$\mathbb{E}[X_+] = \int_0^{\infty} \frac{x}{\pi(1+x^2)} dx = \infty,$$

et puisque X est symétrique, $E[X_-] = E[X_+] = \infty$. Donc malgré le fait que la loi est symétrique, on n'a pas le droit d'écrire $E[X] = 0$, cette quantité n'étant pas définie (et on verra plus tard que cela a des conséquences importantes!).

2.3 Moments et quantités associées

Si X est une variable aléatoire, la quantité $E[X^k]$, $k \in \mathbb{N}^*$, est appelée son *k-ième moment* (ou le *k-ième moment* de sa loi). Les moments d'une variable aléatoire décrivent différents aspects de sa loi. Nous allons nous concentrer dans ce cours sur les quatre premiers moments (donc pour $k = 1, 2, 3, 4$), qui sont les plus importants. Tout d'abord quelques généralités :

- Les moments pairs sont toujours définis et positifs, puisque la fonction $x \mapsto x^k$ est positive pour k un nombre naturel pair.
- Pour $k \in \mathbb{N}$ impair, le k -ième moment est défini et fini si et seulement si $E[|X|^k] < \infty$. Si X est symétrique, alors dans ce cas $E[X^k] = 0$, puisque la fonction $x \mapsto x^k$ est impaire.
- Soit $k < m$. Alors $E[|X|^m] < \infty$ implique $E[|X|^k] < \infty$ puisque $|x|^k \leq 1 + |x|^m$ pour tout $x \in \mathbb{R}$.

2.3.1 Espérance ou moyenne

Le premier moment $E[X]$ est appelé *l'espérance* ou *moyenne* de la variable aléatoire X . On le note souvent par μ . Nous avons vu que c'est la moyenne pondérée des valeurs que peut prendre la variable aléatoire X , avec comme poids les probabilités que ces valeurs sont prises. Si $\mu = 0$, alors on dit que X est *centrée*. Si μ est quelconque, alors $X - \mu$ est appelée la *variable centrée*. En effet, par linéarité de l'espérance, on a bien $E[X - \mu] = E[X] - \mu = \mu - \mu = 0$.

On remarque que la moyenne μ d'une variable aléatoire X n'est pas toujours définie, par exemple si X suit la loi de Cauchy, ou peut être définie mais égale à $+\infty$ ou $-\infty$. Dans la suite, si on parle d'une v.a. X de moyenne μ , on suppose toujours que μ est bien définie et finie.

2.3.2 Variance

La *variance* d'une variable aléatoire X de moyenne μ est définie par

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

C'est donc le second moment de la variable centrée $X - \mu$. Puisque $(X - \mu)^2$ est positive, la variance est toujours définie (pourvu que la moyenne est finie et bien définie) mais peut être infinie. La variance est un *coefficient de dispersion* : il mesure comment les valeurs de X sont dispersées autour de la moyenne. Si la variance est grande, les valeurs sont très dispersées, alors que si la variance est petite, les valeurs sont concentrées autour de la moyenne.

Si X a une « dimension », c'est à dire si que X mesure une quantité dans 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 σ^2 . L'écart-type donne l'ordre de grandeur de la distance entre les valeurs typiques de X et sa moyenne μ .

Si X est une v.a. de moyenne μ et écart-type σ (fini), on appelle $(X - \mu)/\sigma$ la *variable centrée réduite* (ou simplement *variable réduite*). Elle est de moyenne 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 2.9. *Soit X une variable aléatoire de moyenne μ (finie et bien définie).*

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.

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

3. $\text{Var}(X) = 0$ si et seulement si $X \sim \delta_\mu$.

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

$$\text{Var}(aX + b) = E[(aX + b - (a\mu + b))^2] = E[a^2(X - \mu)^2] = a^2 E[(X - \mu)^2] = a^2 \text{Var}(X).$$

2. On développe :

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

Par linéarité de l'espérance, on obtient,

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

3. On définit la variable aléatoire $Y = (X - \mu)^2$. Alors $Y \geq 0$ et $Y = 0$ si et seulement si $X = \mu$. Le Lemme 2.10 ci-dessous appliqué à Y donne alors

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

□

Lemme 2.10. *Soit Y une v.a. positive. Alors $Y \sim \delta_0$ si et seulement si $E[Y] = 0$.*

La preuve sera faite à la fin de la Section 2.5.1.

2.3.3 Asymétrie

Le *coefficient d'asymétrie* (ou *asymétrie*) d'une variable aléatoire X est le troisième moment de la variable centrée réduite, donc la quantité

$$\gamma_1 = E \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right]$$

Le coefficient d'asymétrie ne change pas quand on applique une transformation linéaire à la variable aléatoire. Il décrit alors un aspect de la *forme* de la densité ou la fonction de masse de la variable aléatoire et est invariant par translation ou changement d'échelle. En effet, comme l'indique le nom, le coefficient d'asymétrie quantifie le degré d'asymétrie de la densité ou la fonction de masse. Si celle-ci est « penchée vers la droite », l'asymétrie sera positive et inversement, si elle est « penchée vers la gauche », l'asymétrie sera négative. Le coefficient d'une loi symétrique a un coefficient d'asymétrie nul mais la réciproque est fautive.

Le coefficient d'asymétrie peut se calculer à l'aide des moments de X : en développant et en utilisant la linéarité de l'espérance,

$$\begin{aligned} \gamma_1 &= E \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right] \\ &= \frac{1}{\sigma^3} E[X^3 - 3\mu X^2 + 3\mu^2 X - \mu^3] \\ &= \frac{1}{\sigma^3} (E[X^3] - 3\mu E[X^2] + 2\mu^3). \end{aligned}$$

2.3.4 Kurtosis

Le *kurtosis* ou *coefficient d'aplatissement* d'une variable aléatoire X est le quatrième moment de la variable centrée réduite, donc la quantité

$$\beta_2 = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right]$$

Comme l'asymétrie, le kurtosis décrit un aspect de la forme de la densité ou la fonction de masse de la variable aléatoire. Pour le comprendre, il convient de l'écrire sous une autre forme : Soit $Y = (X - \mu)/\sigma$ la variable centrée réduite. Alors $E[Y^2] = \text{Var}(Y) = 1$, si bien que le kurtosis vaut

$$\beta_2 = E[Y^4] = E[(Y^2)^2] - E[Y^2]^2 + E[Y^2]^2 = \text{Var}(Y^2) + 1.$$

Le kurtosis est alors à une constante près égal à la variance de Y^2 . Il sera alors grand si la loi de Y^2 donne beaucoup de masse aux grandes ainsi qu'aux petites valeurs de Y^2 , et petit si elle donne beaucoup de masse aux valeurs proches de 1 et peu de masse aux grandes et petites valeurs. Typiquement, un grand kurtosis se traduit par une forme plutôt « pointue » de la densité/fonction de masse et un petit kurtosis par une forme plutôt « aplatie ». Cependant, quand la densité/fonction de masse possède plusieurs maxima, cette description peut être erronée.

Propriétés du kurtosis.

- La formule ci-dessus montre que le kurtosis est toujours plus grand ou égal à 1. Cette valeur minimale est atteinte pour X suivant la loi uniforme sur $\{-1, 1\}$, car $Y^2 = X^2 = 1$ et donc $\text{Var}(Y^2) = 0$.
- Le kurtosis de la loi normale est égal à 3 (cf TD). Il est commode de comparer le kurtosis d'une variable aléatoire X à cette valeur. On appelle alors une loi *mésokurtique*, si $\beta_2 = 3$, *leptokurtique* si $\beta_2 > 3$ et *platykurtique* si $\beta_2 < 3$.
- Comme le coefficient d'asymétrie, le kurtosis peut se calculer à l'aide des moments de X : en développant et en utilisant la linéarité de l'espérance,

$$\begin{aligned}\beta_2 &= E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] \\ &= \frac{1}{\sigma^4} E[X^4 - 4\mu X^3 + 6\mu^2 X^2 - 4\mu^3 X + \mu^4] \\ &= \frac{1}{\sigma^4} (E[X^4] - 4\mu E[X^3] + 6\mu^2 E[X^2] - 3\mu^4).\end{aligned}$$

2.4 Fonction génératrice des moments

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

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

Elle porte son nom du fait qu'elle permet d'obtenir tous les moments de la v.a. X :

Proposition 2.11. *Supposons qu'il existe $t_0 > 0$ tel que $\varphi_X(t) < \infty$ pour tout $|t| < t_0$. Alors pour tout $|t| < t_0$,*

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

où le rayon de convergence de cette série entière est plus grand ou égal à t_0 . En particulier, pour tout $k \in \mathbb{N}^*$,

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

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

Démonstration. On suppose que X est 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.

Pour $|t| < t_0$ on calcule alors,

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

et pareil,

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

Le théorème de Fubini donne alors,

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

c'est-à-dire que cette série 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 plus grand ou égal à t_0 . \square

La fonction génératrice permet non seulement de retrouver les moments, mais également toute la loi. Ceci est le contenu du théorème suivant dont nous omettons la preuve :

Theorème 2.12. *Supposons que φ_X est finie dans un voisinage de 0. Alors la loi de X est l'unique loi de fonction génératrice φ_X .*

Exemple 2.13. $a, b \in \mathbb{R}$.

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

Exemple 2.14. $X \sim \text{Poi}(\lambda)$, $\lambda \geq 0$. 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 2.15. $X \sim \mathcal{N}(0, 1)$. 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

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

Si $X \sim \mathcal{N}(0, 1)$, on retrouve les moments de X en développant l'exponentielle :

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

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 2.16. $X \sim \text{Exp}(1)$. 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 :

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

On déduit $\mathbb{E}[X^k] = k!$ 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. Au fond de cette inégalité il y a deux observations simples mais puissantes.

Lemme 2.17. *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 2.18. *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 2.19 (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 2.17 montre que $\mathbf{1}_{X \geq x} \leq X/x$. Par le Lemme 2.18,

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

par linéarité de l'espérance. □

Corollaire : preuve du Lemme 2.10. 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. □

2.5.2 Inégalité de Tchebychev

Theorème 2.20 (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 2.19) 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 $I \subset \mathbb{R}$ un intervalle. On rappelle qu'une fonction $f : I \rightarrow \mathbb{R}$ est *convexe* si pour tout $x, y \in I$, $t \in [0, 1]$,

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

Si $f \in C^2(I)$, alors f est convexe si et seulement si $f'' \geq 0$. Des exemples sont les fonctions e^x , e^{-x} et $|x|^k$ pour $k \geq 1$. Une fonction f telle que $-f$ est convexe est dite *concave*. Exemple : $\log x$ pour $x > 0$.

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

$$f(x) \geq f(x_0) + a(x - x_0).$$

Theorème 2.21. *Soit X une v.a. d'espérance finie, à valeurs dans un intervalle $I \subset \mathbb{R}$ et $f : I \rightarrow \mathbb{R}$ convexe. Alors,*

$$E[f(X)] \geq f(E[X]).$$

Démonstration. Posons $x_0 = E[X]$. Alors il existe $a \in \mathbb{R}$, tel que pour tout $x \in \mathbb{R}$,

$$f(x) \geq f(x_0) + a(x - x_0).$$

Par conséquent,

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

car $x_0 = E[X]$. □

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 l'expérience aléatoire. Même si elles sont de même loi, on ne les considère pas comme les mêmes variables aléatoires. Il y a d'ailleurs un lien fort entre elles : $X + Y = n$. En particulier, pour une réalisation donnée, $X = Y$ si et seulement si $X = Y = n/2$, donc dans la plupart des cas, $X \neq Y$.

On ne doit pas se limiter à deux variables aléatoires. Pour le même exemple de n lancers d'une pièce, on peut noter X_i le résultat du i -ième lancer avec la convention $X_i = 1$ pour *pile* et $X_i = 0$ pour *face*. On a donc défini n variables aléatoires X_1, \dots, X_n . De manière générale, on peut avoir une infinité de variables aléatoires $(X_i)_{i \in I}$ décrivant différents aspects d'une expérience aléatoire, pour un ensemble I quelconque.

Pour définir mathématiquement ce que représentent n variables aléatoires X_1, \dots, X_n , nous devons définir leur *loi jointe*, c'est-à-dire la fonction

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

(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) .

Les lois des variables aléatoires X_1, \dots, X_n sont appelés dans ce contexte les *lois marginales* (de la loi jointe). Elle s'obtiennent à partir de la loi jointe, car pour tout $i \in \{1, \dots, n\}$, $A \subset \mathbb{R}$,

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

La *fonction de répartition conjointe* de n variables aléatoires X_1, \dots, X_n est définie par

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

Elle a des propriétés tout à fait analogues à celle d'une variable aléatoire réelle comme le montre le théorème suivant :

Theorème 3.1.

1. La fonction de répartition F_{X_1, \dots, X_n} de n variables aléatoires X_1, \dots, X_n a les propriétés suivantes :
 - elle est croissante en chaque coordonnée,
 - $F_X(-\infty) := \lim_{x_1 \rightarrow -\infty, \dots, x_n \rightarrow -\infty} F_X(x_1, \dots, x_n) = 0$,
 - $F_X(+\infty) := \lim_{x_1 \rightarrow +\infty, \dots, x_n \rightarrow +\infty} F_X(x_1, \dots, x_n) = 1$,
 - elle est continue à droite en chaque coordonnée.
2. La loi jointe de n variables aléatoires X_1, \dots, X_n est déterminée de manière unique par la fonction de répartition conjointe F_{X_1, \dots, X_n} .
3. Chaque fonction sur \mathbb{R}^n ayant les propriétés données dans 1. est la fonction de répartition conjointe de n variables aléatoires X_1, \dots, X_n .

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

Si les v.a. X_1, \dots, X_n sont discrètes, c'est-à-dire s'ils prennent valeurs dans des ensembles au plus dénombrables $E_1, \dots, E_n \subset \mathbb{R}$, alors le vecteur (X_1, \dots, X_n) prend valeurs dans l'ensemble $E_1 \times \dots \times E_n$ qui est également au plus dénombrable. Par conséquent, la loi jointe de X_1, \dots, X_n est déterminée par la fonction de masse

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

à travers la formule

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

où on garde encore la convention que la somme porte sur tous les (x_1, \dots, x_n) tels que la probabilité correspondante est non-nulle.

3.1.2 Variables aléatoires conjointement à densité

On dit que n variables aléatoires sont *conjointement à densité* s'il existe une fonction $f_{X_1, \dots, X_n} : \mathbb{R}^n \rightarrow \mathbb{R}_+$ telle que

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

Notons que l'ordre des intégrales ci-dessus n'importe pas grâce au théorème de Fubini–Tonelli.

Plus généralement, pour tout $A \subset \mathbb{R}^n$,

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

En effet, on peut vérifier que cela définit bien une loi satisfaisant aux axiomes 1.-4. et sa fonction de répartition 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.

La fonction f_{X_1, \dots, X_n} est appelée *densité conjointe des v.a. X_1, \dots, X_n* . Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est la densité conjointe de n v.a. si et seulement si

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

On remarque qu'il ne suffit pas que chacune des variables X_1, \dots, X_n soit à densité pour que les v.a. X_1, \dots, X_n soient conjointement à densité. Par exemple, si X est une v.a. à densité et $Y = -X$, alors X et Y sont tous les deux à densité mais pas conjointement. En effet, le couple (X, Y) prend ses valeurs dans la droite $D = \{(x, -x) : x \in \mathbb{R}\}$, or, pour toute fonction f ,

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

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

3.1.3 Fonction de variables aléatoires simultanées

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

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

3.1.4 Espérance

Si X_1, \dots, X_n sont des v.a. discrètes ou conjointement à densité et $f : \mathbb{R}^n \rightarrow \mathbb{R}$, alors on définit l'espérance de $f(X_1, \dots, X_n)$ dans le cas discret par

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

et dans le cas continu/à densité par

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

On vérifiera plus tard que cette définition est sans ambiguïté, c'est-à-dire qu'elle équivaut à la définition de l'espérance de la *variable aléatoire (réelle)* $Y = f(X_1, \dots, X_n)$. Elle hérite alors les propriétés de l'espérance vues précédemment. En particulier, les conditions d'existence de l'espérance restent les mêmes : l'espérance $E[f(X_1, \dots, X_n)]$ est bien définie si $E[f_+(X_1, \dots, X_n)] < \infty$ ou $E[f_-(X_1, \dots, X_n)] < \infty$. Aussi, la *linéarité de l'espérance* : Si $f_1, \dots, f_m : \mathbb{R}^n \rightarrow \mathbb{R}$ sont des fonctions, alors

$$E[f_1(X_1, \dots, X_n) + \dots + f_m(X_1, \dots, X_n)] = E[f_1(X_1, \dots, X_n)] + \dots + E[f_m(X_1, \dots, X_n)].$$

En particulier si $m = n$ et $f_i(x_1, \dots, x_n) = f_i(x_i)$ pour tout $i = 1, \dots, n$

$$E[f_1(X_1) + \dots + f_n(X_n)] = E[f_1(X_1)] + \dots + E[f_n(X_n)].$$

Ceci permet de *définir* $E[f_1(X_1) + \dots + f_n(X_n)]$, même si les X_1, \dots, X_n ne sont pas tous discrets ou tous conjointement à densité.

3.2 Variables aléatoires indépendantes

On dit que des variables aléatoires X_1, \dots, X_n sont *indépendantes* si pour tous $A_1, \dots, A_n \subset \mathbb{R}$,

$$P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1) \times \dots \times P(X_n \in A_n). \quad (3.1)$$

En appliquant (3.1) à des ensembles de la forme $A_i =]-\infty, x_i]$, $x_i \in \mathbb{R}$, on obtient la formule suivante pour la fonction de répartition conjointe :

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \times \dots \times F_{X_n}(x_n), \quad x_1, \dots, x_n \in \mathbb{R}. \quad (3.2)$$

Inversement, puisque la fonction de répartition conjointe détermine la loi, si X_1, \dots, X_n sont des v.a. dont la fonction de répartition conjointe s'écrit par (3.2), alors les variables aléatoires sont indépendantes.

On peut se demander si des variables aléatoires indépendantes X_1, \dots, X_n avec des lois marginales prescrites *existent*. En effet, si on prend (3.2) comme *définition* d'une fonction de répartition conjointe, alors on montre aisément que cette fonction satisfait aux propriétés d'une fonction de répartition conjointe, ce qui implique l'existence par le Théorème 3.1.

Proposition 3.2. *Soient X_1, \dots, X_n des v.a.*

— *Si X_1, \dots, X_n sont des v.a. discrètes, alors elles sont indépendantes si et seulement si*

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

— *Si X_1, \dots, X_n sont des v.a. à densité, alors elles sont indépendantes si et seulement si elles admettent la densité conjointe*

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

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

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

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

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

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

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. Une conséquence immédiate de la Proposition 3.2 est la formule suivante :

$$E[g_1(X_1) \times \cdots \times g_n(X_n)] = E[g_1(X_1)] \times \cdots \times E[g_n(X_n)].$$

En effet, dans le cas à densité, on a

$$\begin{aligned} E[g_1(X_1) \times \cdots \times g_n(X_n)] &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g_1(x_1) \cdots g_n(x_n) f_{X_1}(x_1) \cdots f_{X_n}(x_n) dx_1 \cdots dx_n \\ &= \left(\int_{-\infty}^{\infty} g_1(x_1) f_{X_1}(x_1) dx_1 \right) \times \cdots \times \left(\int_{-\infty}^{\infty} g_n(x_n) f_{X_n}(x_n) dx_n \right) \\ &= E[g_1(X_1)] \times \cdots \times E[g_n(X_n)]. \end{aligned}$$

Le cas discret se démontre de manière analogue. En fait, la formule ci-dessus est vraie en toute généralité, par exemple quand certaines des v.a. sont discrètes et d'autres à densité.

En particulier, en prenant $g_i(x) = x$ pour tout i , on obtient la formule

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

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

Proposition 3.3. Soient X_1, \dots, X_n des v.a. discrètes ou conjointement 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 v.a. X_1, \dots, X_n sont indépendantes.

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

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

Notons que $C_i > 0$ pour tout i , sinon on aurait pour tout $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n \in \mathbb{R}$,

$$\int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_i = f_1(x_1) \cdots f_{i-1}(x_{i-1}) C_i f_{i+1}(x_{i+1}) \cdots f_n(x_n) = 0,$$

et donc $\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \cdots dx_n = 0$. Aussi,

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

En particulier, $C_i < \infty$ 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 $C_1 \times \cdots \times C_n = 1$,

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

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

3.3 Fonctions de variables aléatoires indépendantes

Dans cette section, on s'intéresse à la loi d'une fonction de n v.a. indépendantes X_1, \dots, X_n . On considère en particulier les cas du maximum ou minimum ainsi que de la somme.

3.3.1 Maximum et minimum

On s'intéresse à la loi de $Y = \max(X_1, \dots, X_n)$. Elle se détermine le plus simplement par 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 :

$$F_Z(x) = 1 - P(X_1 > x, \dots, X_n > x) = 1 - \prod_{i=1}^n (1 - F_{X_i}(x)).$$

Exemple 3.4. Soient X_1, \dots, X_n iid (indépendantes et identiquement distribuées) de loi Unif(0, 1). Alors,

$$F_{\max(X_1, \dots, X_n)}(x) = \begin{cases} 0, & x \leq 0 \\ x^n, & x \in [0, 1] \\ 1, & x \geq 1. \end{cases}$$

3.3.2 Somme

Dans ce chapitre, on s'intéresse à la loi de $S = X_1 + \dots + X_n$. 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 répartitions respectives des v.a. X_1, \dots, X_n . Qu'en est-il pour la somme ? Existe-t-il encore une quantité en fonction de Y qui s'écrit comme un produit des quantités respectives en fonction des X_1, \dots, X_n ? La réponse est oui : la fonction génératrice des moments. En effet, pour tout $t \in \mathbb{R}$,

$$E[e^{tS}] = E[e^{t(X_1 + \dots + X_n)}] = E[e^{tX_1} \dots e^{tX_n}] = E[e^{tX_1}] \dots E[e^{tX_n}],$$

par indépendance. Par conséquent,

$$\varphi_S(t) = \varphi_{X_1}(t) \dots \varphi_{X_n}(t).$$

Exemple 3.5. On suppose que $X_i \sim \text{Poi}(\lambda_i)$, $\lambda_i \geq 0$, 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 $X_1 + \dots + X_n \sim \text{Poi}(\lambda_1 + \dots + \lambda_n)$.

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 3.6. Soient $f, g : \mathbb{R} \rightarrow \mathbb{R}_+$. On définit leur *convolée* $f \star g : \mathbb{R} \rightarrow \mathbb{R}_+ \cup \{\infty\}$ 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 *convolée 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 3.7. Soient X_1, \dots, X_n soit discrètes et à valeurs dans \mathbb{Z} , soit 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 $f_S = f_X \star f_Y$. □

Exemple 3.8. Soient $X, Y \sim \text{Unif}(0, 1)$ indépendantes, donc de densité $f = \mathbf{1}_{[0,1]}$. Alors $X + Y$ admet la densité

$$\begin{aligned} f_{X+Y}(x) &= (f \star f)(x) = \int_{-\infty}^{\infty} \mathbf{1}_{[0,1]}(x - t) \mathbf{1}_{[0,1]}(t) dt = \int_0^1 \mathbf{1}_{x-1 \leq t \leq x} dt \\ &= x \mathbf{1}_{[0,1]}(x) + (2 - x) \mathbf{1}_{[1,2]}(x). \end{aligned}$$

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 a réussi à calculer la fonction génératrice des moments de S , il n'existe pas de formule simple pour retrouver

sa densité ou fonction de répartition, qui sont souvent les quantités qu'on souhaite calculer. Cette méthode peut donc ne pas aboutir, dans quel cas on pourra passer à la méthode de la convolution des densités.

Pour résumer, on privilégiera la méthode par la fonction génératrice des moments en premier lieu, surtout si on connaît déjà la fonction génératrice des moments des X_1, \dots, X_n ou si on connaît ou devine déjà la loi de S . Si la méthode n'aboutit pas (c'est-à-dire si on ne réussit pas à reconnaître la loi à partir de la fonction génératrice des moments), on utilisera la méthode de la convolution des densités.

Variance d'une somme de v.a. indépendantes

Proposition 3.9. *Soient X_1, \dots, X_n des v.a. indépendantes, $E[(X_i)^2] < \infty$ 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

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 iid (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 $p \in]0, 1[$. Pour tout $n \in \mathbb{N}^*$, soit $Y_n \sim \text{Bin}(n, p)$, et soit $X \sim \mathcal{N}(0, 1)$. Posons

$$X_n = \frac{Y_n - np}{\sqrt{p(1-p)n}}.$$

Alors le théorème de de Moivre–Laplace dit que $F_{X_n}(x) \rightarrow F_X(x)$ pour tout $x \in \mathbb{R}$, si bien que

$$X_n \xrightarrow{\text{loi}} X.$$

On note aussi

$$X_n \xrightarrow{\text{loi}} \mathcal{N}(0, 1),$$

ce qui évite d'introduire la v.a. X .

Exemple 4.2. 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 δ_0 est $F_0 = \mathbf{1}_{[0, \infty[}$. Elle est discontinue en 0 (un atome de la loi δ_0) et continue ailleurs. Pour $x < 0$, on a

$$F_{X_n}(x) = \mathbf{1}_{[1/n, \infty[}(x) = 0 = F_0(x), \quad \text{pour tout } n,$$

et pour $x > 0$,

$$F_{X_n}(x) = \mathbf{1}_{[1/n, \infty[}(x) = 1 = F_0(x) \quad \text{pour tout } n \text{ tel que } 1/n < x.$$

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 4.3. Si les v.a. $(X_n)_{n \geq 1}$ et X sont à valeurs dans \mathbb{N} , on a l'équivalence

$$X_n \xrightarrow{\text{loi}} X \iff P(X_n = k) \rightarrow P(X = k) \forall k \in \mathbb{N},$$

car $F_{X_n}(x) = \sum_{k=0}^{[x]} P(X_n = k)$ pour tout x . Comme exemple, nous avons montré au début du cours que

$$\text{Bin}(n, p_n) \rightarrow \text{Poi}(\lambda), \quad \text{si } p_n \rightarrow 0 \text{ et } np_n \rightarrow \lambda \geq 0.$$

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 4.4. 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 4.5. 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

$$\limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x + \varepsilon).$$

Des arguments similaires donnent

$$\liminf_{n \rightarrow \infty} F_{X_n}(x) \geq F_X(x - \varepsilon).$$

Puisque $\varepsilon > 0$ était arbitraire et F_X est continue en x , cela donne

$$F_{X_n}(x) \rightarrow F_X(x), \quad n \rightarrow \infty.$$

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. \square

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 4.6 (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. On suppose $X_n \xrightarrow{\text{loi}} X$. Soit $U \sim \text{Unif}(0, 1)$. On pose $X'_n = F_{X_n}^{-1}(U)$ et $X' = F_X^{-1}(U)$, où l'inverse généralisée d'une fonction F est définie par

$$F^{-1}(x) = \inf\{y \in \mathbb{R} : F(y) \geq x\}.$$

On peut alors vérifier que $X'_n \stackrel{\text{loi}}{=} X_n$ pour tout n et $X' \stackrel{\text{loi}}{=} X$. De plus, on peut vérifier que $X'_n \xrightarrow{P} X'$, $n \rightarrow \infty$. Par la première partie du lemme, $g(X'_n) \xrightarrow{P} g(X')$ et donc $g(X'_n) \xrightarrow{\text{loi}} g(X')$, et finalement $g(X_n) \xrightarrow{\text{loi}} g(X)$, car $g(X'_n) \stackrel{\text{loi}}{=} g(X_n)$ pour tout n et $g(X') \stackrel{\text{loi}}{=} g(X)$ \square

Exemple 4.7. Comme ci-dessus, soit $p \in]0, 1[$ et pour tout $n \in \mathbb{N}^*$, soit $Y_n \sim \text{Bin}(n, p)$, 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 peu d'abus de notation,

$$1/X_n \xrightarrow{\text{loi}} 1/X, \quad n \rightarrow \infty.$$

4.1.3 Convergence en moyenne quadratique

On dit que la suite X_1, X_2, \dots converge en moyenne quadratique (ou dans L^2) s'il existe une variable aléatoire X avec $E[X^2] < \infty$ et telle que

$$E[(X_n - X)^2] \rightarrow 0, \quad n \rightarrow \infty.$$

On dit alors que X_n converge en moyenne quadratique vers X quand $n \rightarrow \infty$ et on note $X_n \xrightarrow{L^2} X$.

Proposition 4.8. *La convergence en moyenne quadratique implique la convergence en probabilité.*

Démonstration. Supposons que $X_n \xrightarrow{L^2} X$. Par l'inégalité de Markov, pour tout $\varepsilon > 0$,

$$P(|X_n - X| > \varepsilon) = P((X_n - X)^2 > \varepsilon^2) \leq \frac{E[(X_n - X)^2]}{\varepsilon^2} \rightarrow 0.$$

Ceci montre que $X_n \xrightarrow{P} X$. □

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 la convergence en moyenne quadratique).

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 4.9. $(X_n, Y_n) \xrightarrow{\text{loi}} (X, Y)$ 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. □

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 4.10. 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. \square

Theorème 4.11 (Lemme de l'application continue ; version couples). *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 pareil. \square

Lemme 4.12. 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 en général fausse)

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 4.13 (Lemme de Slutsky). *Supposons que $X_n \xrightarrow{\text{loi}} X$ et $Y_n \xrightarrow{\text{loi}} c$, où $c \in \mathbb{R}$ est une constante. Alors*

- $X_n + Y_n \xrightarrow{\text{loi}} X + c$
- $X_n Y_n \xrightarrow{\text{loi}} cX$
- Si $c \neq 0$, $X_n/Y_n \xrightarrow{\text{loi}} X/c$.

4.3 Loi des grands nombres

Theorème 4.14 (Loi des grands nombres). *Soient X_1, X_2, \dots iid avec $E[|X_1|] < \infty$. Alors*

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{P} E[X_1].$$

Démonstration. Nous allons démontrer ce résultat seulement sous l'hypothèse $E[X_1^2] < \infty$. Le cas général 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,

$$\begin{aligned} E \left[\left(\frac{X_1 + \dots + X_n}{n} - E[X_1] \right)^2 \right] &= \text{Var} \left(\frac{X_1 + \dots + X_n}{n} \right) \\ &= \frac{1}{n^2} (\text{Var}(X_1) + \dots + \text{Var}(X_n)) \quad \text{par indépendance} \\ &= \text{Var}(X_1)/n \\ &\rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

On a alors convergence en moyenne quadratique ce qui implique convergence en probabilité. \square

Remarque 4.15. 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 4.16. 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*. Elle n'est pas facile à démontrer.

4.4 Théorème central limite

La loi des grands nombres nous dit que si X_1, \dots, X_n iid 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 4.17 (Théorème central limite). *Soient X_1, X_2, \dots iid de variance $\sigma^2 = \text{Var}(X_1) < \infty$. Alors*

$$\frac{X_1 + \dots + X_n - nE[X_1]}{\sigma\sqrt{n}} \xrightarrow{\text{loi}} \mathcal{N}(0, 1).$$

Nous allons démontrer ce théorème à l'aide des fonctions génératrices des moments. Nous aurons besoin du lemme suivant que nous admettons.

Lemme 4.18. *Soit $(X_n)_{n \geq 1}$ une suite de v.a. et X une v.a. telle que φ_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 peut supposer que $E[X_1] = 0$ et $\sigma^2 = 1$, quitte à remplacer X_i par sa variable centrée réduite. On suppose de plus 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). 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 φ 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 précédent montre alors que

$$\frac{X_1 + \dots + X_n}{\sqrt{n}} \xrightarrow{\text{loi}} \mathcal{N}(0, 1).$$

□

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 $f_{X,Y} = \mathbf{1}_T / \text{aire}(T) = 2\mathbf{1}_T$. Alors X est de densité

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy \\ &= \int_{-\infty}^{+\infty} 2\mathbf{1}_{[0,1]}(x) \mathbf{1}_{(y \in [0,1], x+y \geq 1)} dy \\ &= 2\mathbf{1}_{[0,1]}(x) \mathbf{1}_{1-x}^x \mathbf{1} dy \\ &= 2x \mathbf{1}_{[0,1]}(x). \end{aligned}$$

Par symétrie ($(X, Y) \stackrel{\text{loi}}{=} (Y, X)$), on a également $f_Y(x) = f_X(x) = 2x \mathbf{1}_{[0,1]}(x)$.

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 $Y = y$:

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{2\mathbf{1}_T(x, y)}{2y\mathbf{1}_{[0,1]}(y)} = \frac{1}{y}\mathbf{1}_{(x \in [0,1], x+y \geq 1)} = \frac{1}{y}\mathbf{1}_{[1-y,1]}(x).$$

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 transperce 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 corrélées, mais loin d'être 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 ».

5.4 Fonction d'un vecteur aléatoire

Soient X_1, \dots, X_n des v.a. conjointement à densité et $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ une fonction. On s'intéresse à la loi du vecteur aléatoire $f(X_1, \dots, X_n)$. Sous des conditions convenables à la fonction f , ce vecteur aléatoire admettra encore une densité conjointe qu'on peut exprimer en fonction de la densité conjointe des v.a. X_1, \dots, X_n . Pour cela, nous rappelons d'abord des résultats de l'analyse multidimensionnelle.

Soient f_1, \dots, f_n les marginales de la fonction f , i.e., $f(x) = (f_1(x), \dots, f_n(x))$ pour $x \in \mathbb{R}^d$. On définit la *matrice jacobienne* de la fonction f par

$$\mathcal{J}_f : x \mapsto \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(x) & \cdots & \frac{\partial f_n}{\partial x_n}(x) \end{pmatrix},$$

partout où toutes les dérivées partielles sont bien définies. Notons que la matrice jacobienne est en réalité une *fonction* à valeurs dans les matrices.

Le déterminant de la matrice jacobienne joue un rôle important et est appelé *le jacobien* :

$$J_f(x) := \det \mathcal{J}_f(x).$$

Remarque 5.15. La matrice jacobienne et le jacobien étant des notions locales, ils sont également définis pour des fonctions sur un ouvert $D \subset \mathbb{R}^n$. On dira qu'une fonction f est *continument dérivable* sur D , et on notera $f \in C^1(D)$, si la matrice jacobienne $\mathcal{J}_f(x)$ est définie et continue sur D .

Revenons à la question du début de la section : quelle est la densité conjointe du vecteur aléatoire $f(X_1, \dots, X_n)$ (si elle existe) ? Pour cela, nous avons besoin de savoir comment le volume d -dimensionnel d'un petit parallélépipède I_x de centre x se transforme par l'application f . La réponse est donnée par le jacobien :

$$\text{vol}(f(I_x)) \approx |J_f(x)| \times \text{vol}(I_x).$$

Les mêmes considérations que dans le cas unidimensionnel (Section 1.7) donnent alors les résultats suivants. Pour les deux résultats, on suppose que le vecteur aléatoire X_1, \dots, X_n admet une densité conjointe f_{X_1, \dots, X_n} , est à valeurs dans un ouvert $D \subset \mathbb{R}^n$ et que $f : D \rightarrow \mathbb{R}^n$ est une fonction.

Theorème 5.16. *Supposons que f satisfait aux hypothèses suivantes :*

- $f \in C^1(D)$
- f est injective
- $J_f(x) = 0$ pour un nombre fini de $x \in D$.

Alors le vecteur aléatoire $f(X_1, \dots, X_n)$ admet la densité conjointe donnée par

$$f_{f(X_1, \dots, X_n)}(x) = \begin{cases} |J_{f^{-1}}(x)| \times f_{X_1, \dots, X_n}(f^{-1}(x)), & \text{si } x \in f(D) \text{ et } |J_{f^{-1}}(x)| \text{ existe,} \\ 0, & \text{sinon.} \end{cases}$$

En général (pour des fonctions non injectives), on a le théorème suivant :

Theorème 5.17. *Supposons que f satisfait aux hypothèses suivantes :*

- $f \in C^1(D)$
- il existe $N \in \mathbb{N}$ tel que $\text{Card}(f^{-1}(\{x\})) \leq N$ pour tout $x \in \mathbb{R}^d$,
- $J_f(x) = 0$ pour un nombre fini de $x \in D$.

Alors le vecteur aléatoire $f(X_1, \dots, X_n)$ admet la densité conjointe donnée par

$$f_{f(X_1, \dots, X_n)}(x) = \sum_{y \in D: f(y)=x, J_f(y) \neq 0} f_{X_1, \dots, X_n}(y) \frac{1}{|J_f(y)|}.$$

Exemple 5.18. Soient X_1, \dots, X_n des v.a. iid de loi $\text{Unif}(0, 1)$ et soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ la fonction

$$f(x_1, \dots, x_n) = x_{(1)}, \dots, x_{(n)},$$

où $x_{(1)}, \dots, x_{(n)}$ sont les valeurs x_1, \dots, x_n arrangées dans l'ordre croissant. Alors on s'intéresse au vecteur aléatoire $(X_{(1)}, \dots, X_{(n)}) = f(X_1, \dots, X_n)$. La fonction f est évidemment pas injective, ni de classe C^1 , mais on peut se restreindre à l'ouvert

$$D = \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_i \neq x_j, i \neq j\}.$$

Alors le vecteur aléatoire (X_1, \dots, X_n) est à valeurs dans D , 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 $x = (x_1, \dots, x_n) \in D$ et pour $i \in \{1, \dots, n\}$, notons $\pi(i)$ l'indice tel que $x_i = x_{(\pi(i))}$, donc le rang de x_i dans l'arrangement croissant. Alors pour tout i et pour $\varepsilon > 0$ assez petit,

$$f(x_1, \dots, x_{i-1}, x_i + \varepsilon, x_{i+1}, \dots, x_n) = (x_{\pi(1)}, \dots, x_{\pi(i-1)}, x_{\pi(i)} + \varepsilon, x_{\pi(i+1)}, \dots, x_{\pi(n)}),$$

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

$$\mathcal{J}_f(x_1, \dots, x_n) = (\mathbf{1}_{j=\pi(i)})_{i,j=1,\dots,n}.$$

En particulier, le jacobien est donné par

$$J_f(x_1, \dots, x_n) = \sigma(\pi(i)) \in \{-1, 1\},$$

le signe de la permutation π .

Il reste à déterminer la préimage de f . Clairement, l'image de D par f 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 $f^{-1}(x)$ consiste en tous les points dans D obtenus en permutant les coordonnées de x . Il existe exactement $n!$ de telles permutations et chacune donne un point différent, si bien que $\text{Card}(f^{-1}(x)) = n!$. On peut finalement appliquer le Théorème 5.17 (rappel : $f_{X_1, \dots, X_n} = \mathbf{1}_{[0,1]^n}$) et obtenir pour tout $x \in E$,

$$f_{X_{(1)}, \dots, X_{(n)}}(x) = n! \times \mathbf{1}_S(x), \quad \text{avec } S = \{(x_1, \dots, x_n) \in \mathbb{R}^n : 0 \leq x_1 < \dots < x_n \leq 1\}.$$

Autrement dit, le vecteur aléatoire $(X_{(1)}, \dots, X_{(n)})$ est de loi uniforme dans l'ensemble S .

Exemple 5.19. Soit $f(x) = Ax + b$ avec A une matrice carrée et $b \in \mathbb{R}^n$. Alors

$$\mathcal{J}_f \equiv A,$$

en particulier, $J_f \equiv \det A$. Pour appliquer les théorèmes, il faut alors que $\det A \neq 0$, donc A inversible. La fonction f est alors bijective d'inverse

$$f^{-1}(x) = A^{-1}(x - b).$$

Par le Théorème 5.16, avec $X = (X_1, \dots, X_n)^t$,

$$f_{AX+b}(x) = |\det A^{-1}| \times f_X(A^{-1}(x - b)).$$

Exemple 5.20. Soient X_1, \dots, X_n des v.a. iid 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 écrire avec $x = (x_1, \dots, x_n)$ et $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$ la norme euclidienne,

$$f_X(x) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{\|x\|_2^2}{2}}.$$

Si 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 μ et de matrice de covariance $\Sigma = AA^t$* et on note $AX + \mu \sim \mathcal{N}(\mu, \Sigma)$. Si A est inversible, alors 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 $y \in \mathbb{R}^n$,

$$\|A^{-1}y\|_2 = \langle A^{-1}y, A^{-1}y \rangle = \langle (A^{-1})^t A^{-1}y, y \rangle = \langle (AA^t)^{-1}y, y \rangle = \langle \Sigma^{-1}y, y \rangle.$$

En particulier, la loi de $AX + \mu$ ne dépend que de μ et Σ : 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 5.21. Pour un couple $(x, y) \in \mathbb{R}^2$, notons (r, θ) ses coordonnées polaires, i.e. $(x, y) = (r \cos \theta, r \sin \theta)$. Ceci induit un difféomorphisme, avec $D := \mathbb{R}^2 \setminus (]-\infty, 0] \times \mathbb{R})$,

$$f : D \rightarrow \mathbb{R}_+^* \times]-\pi, \pi[, \quad (x, y) \mapsto (r, \theta).$$

Son inverse est donné par

$$f^{-1}(r, \theta) = (r \cos \theta, r \sin \theta),$$

dont le jacobien est donné par

$$J_{f^{-1}}(r, \theta) = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r(\cos^2 \theta + \sin^2 \theta) = r.$$

On applique la fonction f à un vecteur gaussien standard (X, Y) , i.e. $X, Y \sim \mathcal{N}(0, 1)$ iid. Celui-ci est à valeurs dans D , car $P((X, Y) \notin D) \leq P(Y = 0) = 0$. Notons $(R, \Theta) = f(X, Y)$, i.e. $(X, Y) = (R \cos \Theta, R \sin \Theta)$. Par le Théorème 5.16,

$$\begin{aligned} f_{R, \Theta}(r, \theta) &= |J_{f^{-1}}(r, \theta)| \times f_{X, Y}(f^{-1}(r, \theta)) \\ &= \frac{1}{r} \frac{1}{\sqrt{2\pi}} e^{-r^2(\cos^2 \theta + \sin^2 \theta)} \\ &= \sqrt{2\pi} \frac{1}{r} e^{-r^2/2} \times \frac{1}{2\pi}. \end{aligned}$$

Ceci est le produit des deux fonctions $f_R(r) = \sqrt{2\pi} \frac{1}{r} e^{-r^2/2}$ et $f_\Theta(\theta) = \frac{1}{2\pi}$. Par conséquent, les v.a. R et Θ sont indépendantes et de densités proportionnelles à f_R et f_Θ , respectivement. Mais puisque f_Θ est d'intégrale 1, f_R doit l'être aussi et donc f_R et f_Θ sont en fait les densités. On résume : R et Θ sont des v.a. indépendantes, Θ suivant la loi $\text{Unif}(-\pi, \pi)$ et R admettant la densité

$$f_R(r) = \sqrt{2\pi} \frac{1}{r} e^{-r^2/2} \mathbf{1}_{r>0}.$$