

Probabilités (INCOMPLET)

Pascal Maillard¹

9 février 2019

1. avec des modifications et ajouts d'Olivier Hénard

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	Séries entières	7
0.2.3	Théorèmes de Fubini et Tonelli	8
0.3	Rappels du cours de probas en L2	9
0.3.1	Probabilités	9
0.3.2	Variables aléatoires	10
0.3.3	Par la suite...	10
1	Variables aléatoires réelles	13
1.1	Définition et axiomes	13
1.1.1	Axiomes des probabilités	14
1.2	Événements	16
1.3	Fonction de répartition	17
1.4	Classes de variables aléatoires	19
1.5	Quelques variables aléatoires discrètes importantes	22
1.5.1	Loi de Dirac	22
1.5.2	Loi uniforme	23
1.5.3	Loi de Bernoulli	23
1.5.4	Loi binomiale	24
1.5.5	Loi de Poisson	24
1.5.6	Loi géométrique	25
1.5.7	Autres lois discrètes	25
1.6	Quelques variables aléatoires continues importantes	26
1.6.1	Loi uniforme	26
1.6.2	Loi normale (ou gaussienne)	26
1.6.3	Loi exponentielle	28
1.6.4	Loi gamma	28
1.6.5	Loi de Cauchy	28
1.6.6	Autres lois à densité	29
1.7	Loi d'une fonction d'une variable aléatoire	29

2	Espérance et moments	35
2.1	Espérance d'une variable aléatoire	35
2.1.1	Variable aléatoire positive ou nulle	35
2.1.2	Variable aléatoire signée	36
2.2	Espérance d'une fonction d'une variable aléatoire	37
2.3	Moments et quantités associées	39
2.3.1	Espérance	39
2.3.2	Variance	40
2.3.3	Asymétrie	42
2.3.4	Kurtosis	43
2.4	Fonction génératrice des moments	43
2.5	Inégalités	45
2.5.1	Inégalité de Markov	46
2.5.2	Inégalité de Tchebychev	46
2.5.3	Inégalité de Jensen	47
2.5.4	Une formule générale pour l'espérance d'une variable aléatoire positive	47

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

Ces sujets sont bien exposés dans un grand nombre de livres et de notes de cours. Le dernier sujet (séries de fonctions et intégrales à paramètre) est peut-être moins bien traité, c'est pourquoi nous donnons ci-dessous les résultats à retenir.

Nous aurons également besoin de résultats sur les sujets suivants traités dans le cours d'Intégration en L3 :

- *théorèmes de Fubini et Tonelli* : interversion série \longleftrightarrow série, série \longleftrightarrow intégrale, intégrale \longleftrightarrow intégrale

— *intégrales multiples* : intégration sur un domaine de \mathbb{R}^d , changement de coordonnées
 Nous 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

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^{∞} 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 Séries entières

En probabilités, on rencontre assez souvent des séries entières, et les résultats précédents s'appliquent alors très simplement. Une série entière est formellement notée $\sum_{n \geq 0} a_n z^n$, où $(a_n)_{n \in \mathbb{N}}$ est une suite de nombres réels (ou complexes). La somme de la série entière est la fonction qui à tout complexe z tel que $\sum a_n z^n$ converge, associe

$$f(z) = \sum_n a_n z^n.$$

Quelques exemples importants de séries entières figurent dans le Tableau 1.

La première question à se poser est de savoir où converge cette somme. À cet effet, on introduit :

Définition 0.4 (Rayon et disque de convergence). On appelle rayon de convergence de la série entière le réel R défini par

$$R = \sup\{r \geq 0 : \sum_n |a_n| r^n \text{ est bornée}\}$$

et on appelle disque de convergence de la série entière $\sum a_n r^n$ le disque ouvert $D_R = \{z \in \mathbb{C} : |z| < R\}$ de rayon R est de centre 0.

On peut alors fournir une réponse à la question précédente : une série entière converge *absolument* sur son disque de convergence. De plus la convergence est *normale* sur tout disque fermé inclus dans le disque de convergence, en particulier sur $\overline{D}_r = \{z \in \mathbb{C} : |z| \leq r\}$, avec $r < R$. Par définition, la série diverge grossièrement en dehors de \overline{D}_r ; sur le cercle de convergence $\overline{D}_R \setminus D_R$ enfin, la situation est bien plus délicate et aucune possibilité ne peut être exclue.

Concernant la dérivation terme à terme, le cadre des séries entières est particulièrement commode comme le montre le théorème suivant :

$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n = 1 + z + z^2 + z^3 + \dots$	$R = 1$
$\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} + \dots$	$R = \infty$
$\ln(1+z) = \sum_{n=1}^{+\infty} (-1)^{n+1} \frac{z^n}{n} = z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} + \dots$	$R = 1$
$n \in \mathbb{N}, (1+x)^n = \sum_{k=0}^n \binom{n}{k} x^k = 1 + nz + \frac{n(n-1)}{2} z^2 + \dots + nz^{n-1} + z^n$	$R = \infty$
$\alpha \in \mathbb{R} \setminus \mathbb{N}, (1+z)^\alpha = \sum_{n=0}^{\infty} \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!} z^n = 1 + \alpha z + \frac{\alpha(\alpha-1)}{2} z^2 + \dots$	$R = 1$
$\cos z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!} = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \dots$	$R = \infty$
$\sin z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!} = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \dots$	$R = \infty$
$\arctan z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{2n+1} = z - \frac{z^3}{3} + \frac{z^5}{5} - \frac{z^7}{7} + \dots$	$R = 1$

TABLE 1 – Quelques séries entières importantes et leur rayon de convergence R .

Theorème 0.5. *La somme $f(z) = \sum_n a_n z^n$ de la série entière est indéfiniment dérivable sur son disque de convergence D_R . Sa dérivée est la somme de la série dérivée terme à terme, qui a même rayon de convergence :*

$$\forall z \in D_R, f'(z) = \sum_{n=1}^{\infty} n a_n z^{n-1} = a_1 + 2a_2 z + \dots + n a_n z^{n-1} + \dots$$

Si donc $R > 0$, on a, pour tout $n \in \mathbb{N}$,

$$a_n = \frac{f^{(n)}(0)}{n!}.$$

En particulier, deux séries entières qui coïncident sur un intervalle ouvert contenant 0 sont égales, c'est-à-dire qu'on peut identifier la suite de leurs coefficients.

0.2.3 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'invertir les signes somme et/ou intégrale, c'est-à-dire, on souhaite répondre à la question suivante : *Les quantités ci-dessus sont-elles égales à*

(respectivement) :

$$\sum_{m \in \mathbb{N}} \left(\sum_{n \in \mathbb{N}} f_{n,m} \right), \quad \int_{-\infty}^{\infty} \left(\sum_{n \in \mathbb{N}} f_n(y) dy \right), \quad \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x, y) dx \right) dy ?$$

Une première réponse est apportée par le *théorème de Tonelli* :

Theorème 0.6 (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 0.7 (Fubini). *Supposons qu'une des conditions équivalentes sont satisfaites :*

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

Alors on a l'égalité suivante :

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

c'est-à-dire que les deux séries convergent et ont la même limite.

Les résultats analogues pour les autres quantités sont également vrais.

Notons que l'équivalence des deux conditions du théorème découle de Fubini cas positif.

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 soit trop gros pour travailler avec directement ou qu'on ne s'intéresse qu'à certaines propriétés du résultat de l'expérience. Typiquement, on associe alors un nombre réel à un résultat qui contient l'information qu'on cherche. Par exemple, lors d'un sondage on pourrait définir X comme étant le numéro du candidat ayant remporté le plus de 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 simplifier grandement la suite des opérations.

Il s'est avéré au fil du temps qu'il est souvent plus commode de travailler entièrement avec des variables aléatoires et d'ignorer complètement l'espace 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.

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.

- 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).
- Enfin, techniquement, si effectivement la théorie des probabilités use des outils de la théorie de la mesure (abstraite) et de l'intégration, les objets d'étude sont bien différents dans les deux domaines ; en probabilité, l'espace Ω est souvent relégué au second plan, les quantités d'intérêt étant souvent définies en fonction des seules lois des variables aléatoires en jeu. Si une compréhension fine des probabilités passe par l'espace Ω , on gagnera en première approche à ignorer simplement cet espace Ω .

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

Chapitre 1

Variables aléatoires réelles

1.1 Définition et axiomes

Une fonction de masse pour une variable aléatoire continue ? On cherche à modéliser une expérience ayant un résultat aléatoire. On suppose qu'on observe le résultat à partir d'une *variable aléatoire (v.a.)* X qui prend ses valeurs dans les nombres réels. Précédemment, nous avons toujours supposé qu'une variable aléatoire 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 dépende 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 .

1.1.1 Axiomes des probabilités

On doit alors trouver un autre moyen pour définir les probabilités $P(X \in A)$ pour des ensembles $A \subset \mathbb{R}$. 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 $A \subset \mathbb{R}$ (*positivité*)
2. $P(X \in \emptyset) = 0$ et $P(X \in \mathbb{R}) = 1$,
3. Si $A, B \subset \mathbb{R}$ sont disjoints, alors

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

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

- Si $A \subset B$, alors

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

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

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

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

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

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 $A \in \mathcal{R}$ est une simple conséquence de la monotonie. La propriété d'additivité finie découle de la propriété d'additivité par une récurrence finie.

On peut finalement présenter le quatrième axiome :

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

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

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

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

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

- 4'. Si A_0, A_1, A_2, \dots est une suite décroissante d'ensembles (c'est-à-dire $A_n \subset A_{n+1}$ 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.

Une conséquence importante de l'axiome 4 et de l'axiome d'additivité est la propriété de sigma-additivité :

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

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

Posons en effet $B_n = \bigcup_{k=0}^n A_k$ (réunion disjointe), alors B_0, B_1, B_2, \dots est une suite croissante d'ensembles qui vérifie $\bigcup_{n \in \mathbb{N}} B_n = \bigcup_{n \in \mathbb{N}} A_n$, appliquant alors successivement

l'axiome 4 et la propriété d'additivité, on obtient :

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

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

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

1.2 Événements

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

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

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

Définition 1.1. Un événement de probabilité 1 est dit presque sûr. Un événement de probabilité null est dit négligeable.

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

$$X \in A, \text{ p.s. }^1 \quad \text{qui signifie} \quad \mathbb{P}(X \in A) = 1.$$

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

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

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 1.2. 1. La fonction de répartition F_X d'une v.a.r. a les propriétés suivantes :

- elle est croissante,
 - $F_X(-\infty) := \lim_{x \rightarrow -\infty} F_X(x) = 0$,
 - $F_X(+\infty) := \lim_{x \rightarrow +\infty} F_X(x) = 1$,
 - elle est continue à droite.
2. La loi d'une v.a.r. X est déterminée de manière unique par sa fonction de répartition F_X .
3. Chaque fonction sur \mathbb{R} ayant les propriétés données dans 1. est la fonction de répartition d'une v.a. réelle.

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 dénombrable d'intervalles fermés :

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

L'axiome 4 donne alors

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

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

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

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

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

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

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

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

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

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

Soulignons le fait important suivant rencontré au cours de la preuve : la fonction de répartition F_X admet (en tant que fonction croissante) une limite à gauche en tout point, égale à :

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

pour tout réel x .

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*.

Définition 1.3. 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 $x \in \mathbb{R}$,

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

Lemme 1.4. Soit $x \in \mathbb{R}$. 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 alors la hauteur du saut en x , c'est-à-dire $m = F_X(x) - F_X(x-)$.

En particulier, une v.a. X n'a pas d'atomes si et seulement si sa fonction de répartition est continue.

Corollaire 1.5. La loi d'une variable aléatoire X a un nombre dénombrable d'atomes.

Démonstration. Puisque la fonction F_X est croissante d'après le théorème 1.2, elle a un nombre dénombrable de sauts (résultat classique d'analyse) \square

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 rappelle qu'une variable aléatoire X est discrète s'il existe un ensemble S dénombrable tel que $P(X \in S) = 1$. On dit à ce moment-là que X est la loi de X est *portée par* S , ou que S est le *support* de la loi de X . On rappelle que $S \subset \mathbb{R}$ est dit dénombrable si il existe une injection de S dans \mathbb{N} , ou de façon équivalente, s'il existe une surjection de \mathbb{N} dans S . On rappelle que $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ sont dénombrables mais que \mathbb{R} ne l'est pas². Noter aussi que notre définition d'un ensemble dénombrable inclut les ensembles

2. On montre par un argument de diagonalisation de Cantor que l'ensemble $\{0, 1\}^{\mathbb{N}}$ des suites à valeurs dans $\{0, 1\}$ n'est pas dénombrable, ce qui permet de voir que $[0, 1]$ ne l'est pas en considérant le développement dyadique d'un réel

finis ; pour préciser qu'un ensemble dénombrable est de plus infini, on dira simplement qu'il est infini dénombrable. S dénombrable implique que S est de la forme : $S = \{s_k, k \in \mathbb{N}\}$ ou $S = \{s_k, k \leq n\}$ pour un certain $n \in \mathbb{N}$ (cas infini et fini respectivement), pour des réels s_k deux à deux distincts. On notera simplement $S = \{s_k\}$ pour ne pas avoir à faire la distinction. On rappelle avant d'énoncer la prochaine proposition la définition de $\mathbb{1}_A$ la fonction indicatrice de l'ensemble A :

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

Proposition 1.6. *Si X est une v.a.r. discrète, d'ensemble d'atomes $S = \{s_k\}_k$ alors, notant $p_k = P(X = s_k)$ pour tout k , on a pour tout ensemble A*

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

Démonstration. Les deux cas S fini et infini sont quasiment identiques, on traite le cas infini $S = \{s_k, k \in \mathbb{N}\}$, et on mentionne les différences entre les deux cas si besoin. On décompose comme suit l'événement $\{X \in A\}$:

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

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

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

□

Définition 1.7. Soit X v.a. discrète de support $S = \{s_k\}_k$. On appelle fonction de masse de X la fonction :

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

La fonction de répartition d'une v.a. discrète X et de support $S = \{s_1, s_2, \dots\}$ prend la forme

$$F_X(x) = \sum_k p_k \mathbb{1}_{]-\infty, x]}(s_k) = \sum_k p_k \mathbb{1}_{s_k \leq x}.$$

On vérifie que F_X est constante sur chaque intervalle qui ne contient pas de saut. Réciproquement, on a la proposition suivante, dont on omet la preuve :

Proposition 1.8. *Supposons que X est une v.a. et que sa fonction de répartition F_X satisfait aux deux propriétés suivantes :*

1. *L'ensemble des sauts de F_X n'admet pas de point d'accumulation.*
2. *F_X est constante sur chaque intervalle qui ne contient pas de saut.*

Alors X est une v.a. discrète et son support est égal à l'ensemble des sauts de F_X .

On appelle une fonction croissante qui satisfait aux deux hypothèses de la proposition précédente une fonction « en escalier ». Notons que la première hypothèse ne peut pas être enlevée de la proposition, par exemple, si l'ensemble des sauts de F_X est dense dans \mathbb{R} , alors il n'existe aucun intervalle de longueur positive qui ne contient pas de saut, donc une telle fonction satisfait trivialement à la deuxième hypothèse. Il n'y a alors aucune raison que les hauteurs des sauts somment à 1 (ce qui serait nécessaire pour que X soit discrète) : on peut construire des contre-exemples en prenant une fonction de répartition continue G et en considérant la fonction $(F_X + G)/2$.

Variables aléatoires à densité.

Définition 1.9. On dit qu'une v.a.r. X est à densité³ 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 $f_X : \mathbb{R} \rightarrow \mathbb{R}_+$. Dans ce cas, on dit que la fonction f_X est une densité de X (ou de la loi de X), et que X possède la densité f_X .

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. Aussi on parlera en général d'une densité et non pas de la densité.

On note le résultat suivant :

Lemme 1.10. *La loi de X v.a. à densité est sans atome.*

Démonstration. En effet, la fonction F_X étant la primitive d'une fonction, elle est continue, donc en particulier, pour tout $x \in \mathbb{R}$:

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

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

En particulier, si X est une v.a. à densité, pour tout ensemble dénombrable $T \subset \mathbb{R}$, on a

$$\mathbb{P}(X \in T) = 0,$$

ce qui montre la différence de nature entre v.a.r. discrète et v.a.r. à densité.

La densité caractérise la fonction de répartition de X et donc la loi de X (du théorème 1.2) qui s'exprime facilement à l'aide de la densité. On énonce la proposition suivante, analogue pour les v.a. à densité de la Proposition 1.6. Elle montre que la densité joue le rôle analogue de la fonction de masse dans le cas discret ; l'intuition est que, pour un « élément infinitésimal » dx , on a : $\mathbb{P}(X \in dx) = f_X(x)dx$.

Proposition 1.11. *Si X est une v.a. de densité f_X , alors on a pour tout ensemble A :*

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

3. Dans la littérature, on rencontre aussi le vocabulaire v.a. *absolument continues* pour désigner ces v.a., ou même plus improprement, v.a. *continues*

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

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

Proposition 1.12. *Soit X une v.a. dont la fonction de répartition F_X est globalement continue et \mathcal{C}^1 par morceaux (c'est-à-dire qu'il existe un nombre fini de points $x_1 < x_2 < \dots < x_n$ t.q. F_X est \mathcal{C}^1 sur $]-\infty, x_1[,]x_1, x_2[, \dots,]x_{n-1}, x_n[,]x_n, +\infty[$) alors X est une v.a. à densité et une densité de X est F_X' .*

Toutes les v.a.r. à densité que nous rencontrerons dans ce cours auront une fonction de répartition du type précédent (continue et \mathcal{C}^1 par morceaux). Cependant, la forme générale autorise des fonctions de répartition qui ne sont pas de classe \mathcal{C}^1 .

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

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... ». Pour comprendre les propriétés spécifiques des fonctions qui s'écrivent comme primitive, on pourra aussi chercher « fonction absolument continue ... ».

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

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

En particulier, la fonction de répartition est donnée par

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

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

Si X suit la loi de Dirac en $x \in \mathbb{R}$, on dit aussi que X est presque sûrement (p.s.) égale à x , 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 $\mathbb{1}_A(x)$. 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 $P(X \in A) = \mathbb{1}_A(x)$.

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 S une partie finie de \mathbb{R} , par exemple $S = \{1, \dots, n\}$ 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

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

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

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

Plus précisément, on a le théorème suivant :

Théorème 1.13 (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 petit⁴.

4. 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* $\lfloor x \rfloor$ est le plus grand entier plus petit ou égal à x . On obtient alors :

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

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 ; à la différence des précédentes, elles ne sont pas à connaître par coeur.

Loi binomiale négative On dit que X suit la loi binomiale négative de paramètres $r \in \mathbb{N}$ et $p \in]0, 1]$ (symboliquement : $X \sim \text{BN}(r, p)$), 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 1.13. 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 possède la densité

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

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 . Insistons sur le fait que l'intervalle est nécessairement borné : a ne peut prendre la valeur $-\infty$, b ne peut prendre la valeur $+\infty$. En particulier, il n'existe pas de mesure uniforme sur \mathbb{R} (tout comme il n'existe pas de mesure uniforme sur \mathbb{N} dans le cas discret). À titre d'exercice on pourra déterminer la loi de $a \cdot X + b$ lorsque $X \sim \text{Unif}([0, 1])$, et $a, b \in \mathbb{R}$.

1.6.2 Loi normale (ou gaussienne)

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

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

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

Si $\mu = 0$, on dit que X est *centrée* et si 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 (qu'on verra plus tard) :

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

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 1.14 (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.

Correction de continuité En pratique on utilise cette approximation à n fixé, et souvent pour de petits intervalles. L'exemple typique est celui où on veut approcher la probabilité que la variable binomiale prenne une valeur donnée k . Mais alors on *ne peut pas* directement écrire :

$$\mathbb{P}(X = k) \approx \mathbb{P}(np + \sqrt{p(1-p)n} \times N = k)$$

car le membre de droite vaut 0, ce qui est une approximation peu intéressante... On utilisera donc *la correction de continuité* : cette méthode, qui est utilisée à chaque fois qu'on approche la loi d'une variable aléatoire discrète (ici, X à valeurs dans \mathbb{N}) par la loi d'une variable aléatoire à densité (ici, N ou sa transformée linéaire), consiste à écrire :

$$\mathbb{P}(X = k) \approx \mathbb{P}\left(k - \frac{1}{2} \leq np + \sqrt{p(1-p)n} \times N \leq k + \frac{1}{2}\right)$$

En particulier, on approchera

$$\mathbb{P}(X < k + 1) = \mathbb{P}(X \leq k) \approx \mathbb{P}\left(np + \sqrt{p(1-p)n} \times N \leq k + \frac{1}{2}\right)$$

Cette correction peut représenter des différences numériques significatives. On notera pour retenir cette formule que $k + 1/2$ est la moyenne entre k et $k + 1$.

1.6.3 Loi exponentielle

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

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

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

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

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 satisfait la propriété de perte de mémoire

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

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 . Du fait de cette propriété, la loi exponentielle est choisie pour modéliser 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) en effet. La propriété d'absence de mémoire caractérise en fait la loi exponentielle (il s'agit d'un exercice difficile).

1.6.4 Loi gamma

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

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

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

La loi gamma est une généralisation de la loi exponentielle : d'abord, on retrouve $\text{Exp}(\beta)$ quand $\alpha = 1$. Ensuite, on verra que lorsque le paramètre α est un entier, la loi gamma de paramètres $\alpha, \beta > 0$ peut être obtenue comme la loi de la somme de α variables aléatoires indépendantes de loi $\text{Exp}(\beta)$. Les valeurs non entières de α donnent un sens à des sommes non entières de variables aléatoires $\text{Exp}(\beta)$, tout comme la fonction Γ peut être vue comme une interpolation de la fonction factorielle.

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

1.6.5 Loi de Cauchy

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

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

Sa fonction de répartition vaut (exercice)

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

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

1.6.6 Autres lois à densité

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

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

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

En dérivant, on obtient la densité

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

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

1.7 Loi d'une fonction d'une variable aléatoire

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

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

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

Ayant défini la variable aléatoire $\varphi(X)$ 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 $\varphi(X)$. Par définition, on a

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

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

Corollaire 1.15. Soient $I, J \subset \mathbb{R}$ des intervalles, $\varphi : I \rightarrow J$ 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 $\varphi(X)$ s'écrit :

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

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

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

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

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

Exemple 1.16. Soit X une v.a. positive, c'est-à-dire à valeurs dans \mathbb{R}_+ . On considère la fonction

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

Alors φ est une fonction continue, strictement croissante et surjective avec comme inverse $\varphi^{-1}(x) = \sqrt{x}$. Par le Corollaire 1.15, 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 1.15 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, $\varphi(X)$ est-elle discrète? Deux conditions sont les plus courantes :

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

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

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

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

Corollaire 1.17. *Supposons que $\varphi(X)$ est une v.a. discrète (par exemple si une des deux conditions suivantes sont vérifiées). Alors la fonction de masse de $\varphi(X)$ s'exprime comme suit :*

- Si X est discrète,

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

- Si X admet une densité f_X ,

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

Exemple 1.18. $\varphi : \mathbb{R} \rightarrow \{0, 1\}$, $\varphi(x) = \mathbb{1}_{x \geq 0}$, X une v.a. quelconque. La fonction φ ne prend alors que deux valeurs, 0 ou 1. Par conséquent, la v.a. $\varphi(X) = \mathbb{1}_{X \geq 0}$ est discrète. Sa fonction de masse est donnée par :

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

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

Exemple 1.19. $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, $\varphi(x) = \min(x, N)$ (avec $N \in \mathbb{N}$), X une v.a. à valeurs dans \mathbb{N} . Puisque X est une v.a. discrète, $\varphi(X)$ l'est aussi. Sa fonction de masse est donnée par

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

Exemple 1.20. $\varphi :]0, 1[\rightarrow \mathbb{N}$, $\varphi(x) = \lfloor Nx \rfloor$ (avec $N \in \mathbb{N}$), $X \sim \text{Unif}(0, 1)$. La fonction φ est à valeurs dans $\{0, \dots, N-1\}$, la v.a. $\varphi(X)$ est donc discrète. Sa fonction de masse est donnée pour tout $n \in \{0, \dots, N-1\}$ par :

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

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

Calcul de la densité pour des v.a. à densité. Supposons que X soit une v.a. à densité. Alors sous certaines conditions sur φ , $\varphi(X)$ sera également une v.a. à densité et on peut alors calculer facilement la densité de $\varphi(X)$ en fonction de la densité de X et de (la dérivée de) la fonction φ . 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) :

Theorème 1.21. *Supposons que X admet une densité f_X et est à valeurs dans un intervalle $I \subset \mathbb{R}$. Supposons en plus que φ satisfait aux hypothèses suivantes :*

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

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

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

Remarque 1.22. On rappelle que $(\varphi^{-1})'(x) = \frac{1}{\varphi'(\varphi^{-1}(x))}$, dès lors que $\varphi'(\varphi^{-1}(x)) \neq 0$. La condition « $|(\varphi^{-1})'(x)|$ existe » dans le théorème ci-dessus signifie alors que $\varphi'(\varphi^{-1}(x)) \neq 0$.

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

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

$$\begin{aligned} F_{\varphi(X)}(x) &= F_X(\varphi^{-1}(x)) && \text{par le Corollaire 1.15} \\ &= \int_{\inf I}^{\varphi^{-1}(x)} f_X(y) dy \\ &= \int_{\inf \varphi(I)}^x f_X(\varphi^{-1}(z)) (\varphi^{-1})'(z) dz && \text{chgt de var. } z = \varphi(y), dy = (\varphi^{-1})'(z) dz. \end{aligned}$$

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

Le cas où φ est décroissante est similaire. □

Règle mnémotechnique. Pour mémoriser la formule du théorème 1.21, il convient d'utiliser la règle mnémotechnique suivante. On part de l'expression

$$f_X(x) dx$$

On effectue alors un changement de variable $z = \varphi(x)$, $dx = (\varphi^{-1})'(z) dz$, 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 φ est décroissante. Ceci donne alors

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

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

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

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 φ : la masse qui se trouvait en x se trouve désormais en $z = \varphi(x)$. Combien de masse se trouve-t-il dans un intervalle I_z de longueur infinitésimale dz autour de z ? Considérons l'application inverse φ^{-1} . Celle-ci dilate l'espace autour de z par un facteur $|(\varphi^{-1})'(z)|$, si bien que la masse dans l'intervalle I_z est la masse qui était auparavant dans un intervalle de longueur $|(\varphi^{-1})'(z)| \times dz$ autour de $x = \varphi^{-1}(z)$. L'intervalle I_z contient donc

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

kg de masse.

Exemple 1.23. $\varphi : \mathbb{R} \rightarrow \mathbb{R}$, $\varphi(x) = ax + b$, $a \neq 0$, $b \in \mathbb{R}$. Alors φ est strictement monotone (croissante quand $a > 0$ et décroissante quand $a < 0$). De plus, $\varphi^{-1}(x) = (x - b)/a$, donc $(\varphi^{-1})'(x) = 1/a$ pour tout $x \in \mathbb{R}$. Le théorème 1.21 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 par rapport à l'axe vertical, puis resserrée ou dilatée selon que $|a| < 1$ ou > 1 . Enfin, la densité est translatée de b , donc vers la droite si $b > 0$ et vers la gauche si $b < 0$.

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 1.24. On reprend l'exemple $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, $\varphi(x) = x^2$. L'inverse est alors $\varphi^{-1}(x) = \sqrt{x}$ et $(\varphi^{-1})'(x) = \frac{1}{2\sqrt{x}}$ 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 φ est monotone dans le théorème 1.21. Cela donne le théorème suivant, dont la preuve est laissée en exercice :

Théorème 1.25. *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 φ satisfait aux hypothèses suivantes :*

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

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

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

Exemple 1.26. On reprend l'exemple $\varphi(x) = x^2$, avec cette fois-ci X une v.a. à densité sur \mathbb{R} . Alors

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

De plus, $|\varphi'(y)| = 2|y|$ pour tout $y \in \mathbb{R}$. Le théorème 1.25 donne alors

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

□

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 précédant 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

2.1.1 Variable aléatoire positive ou nulle

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

Définition 2.1. Soit X une variable aléatoire *positive ou nulle*, discrète ou à densité. Alors *l'espérance* de X , notée $E[X] \in [0, \infty]$, est définie par

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

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

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

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

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

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

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

cette série ayant été calculée dans 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 2.4. 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 en TD. Interprétation : Une pièce ayant un taux de panne constant égal à λ a une durée de vie de $1/\lambda$ en moyenne.

Toutes les variables aléatoires précédentes sont intégrables.

2.1.2 Variable aléatoire signée

Définition 2.5. Soit X une variable aléatoire discrète ou à densité. On dit que X est intégrable si $\sum_x |x|P(X = x) < +\infty$ si X est discrète, et si $\int_0^{+\infty} |x|f_X(x) dx < +\infty$ si X admet la densité f_X . Dans ce cas on définit l'espérance de X par la formule suivante :

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

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

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

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

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

car $b^2 - a^2 = (a+b)(b-a)$.

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

2.2 Espérance d'une fonction d'une variable aléatoire

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

Proposition 2.7 (Formule de transfert). *Soit X une variable aléatoire discrète ou à densité et $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction. $\varphi(X)$ est une variable aléatoire intégrable ssi $\sum_x |\varphi(x)|P(X = x) < \infty$ si X est discrète et $\int_{-\infty}^{+\infty} |\varphi(x)|f_X(x) dx < \infty$ si X admet la densité f_X . Dans ce cas, son espérance est donnée par :*

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

Cette définition généralise celle de $E[X]$ qu'on retrouve en prenant f l'application identité $\varphi(x) = x$. L'espérance $E[\varphi(X)]$ est encore intuitivement la moyenne pondérée des valeurs que $\varphi(X)$ peut prendre, les poids étant les probabilités que ces valeurs soient prises. Appliquant cette définition avec $\varphi = \mathbb{1}_A$, on obtient bien :

$$E(\mathbb{1}_A(X)) = 1 \cdot P(X \in A) + 0 \cdot P(X \notin A)$$

Démonstration. Considérons d'abord le cas où X est discrète. Alors $\varphi(X)$ l'est aussi et le Corollaire 1.17 donne

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

Par conséquent,

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

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 φ est \mathcal{C}^1 et de dérivée $\varphi'(x) > 0$ pour tout $x \in \mathbb{R}$ et on admet le théorème dans le cas général. D'après le Théorème 1.21, on a alors

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

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

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

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

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

En particulier, si X est intégrable,

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

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

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

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

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

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

On introduit maintenant une conséquence intéressante pour les variables aléatoires de loi symétrique.

Définition 2.9. On dit que la loi de la variable aléatoire X est symétrique si X et $-X$ sont égales en loi, c'est-à-dire que pour tout $A \subset \mathbb{R}$,

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

Les calculs de moments peuvent être simplifiés dans ce contexte.

Lemme 2.10. *Soit X une variable aléatoire symétrique et $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction impaire, c'est-à-dire $\varphi(-x) = -\varphi(x)$ pour tout $x \in \mathbb{R}$, telle que $\varphi(X)$ est intégrable. Alors*

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

Démonstration. On note simplement que :

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

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

Par exemple, si X est symétrique et intégrable, alors $E[X] = 0$ (utiliser $\varphi(x) = x$). si l'on note $x_+ = \max\{x, 0\} = x\mathbb{1}_{x>0}$ et $x_- = \max\{-x, 0\} = -x\mathbb{1}_{x<0}$ les parties positives et négatives du réel x , qui vérifient $x = x_+ - x_-$ et $|x| = x_+ + x_-$. On a alors le lemme suivant :

Lemme 2.11. *Soit X une variable aléatoire. X est intégrable ssi $|X|$ est intégrable ssi X_+ et X_- sont intégrables. Dans ce cas, on peut alors écrire :*

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

On peut aussi énoncer le lemme suivant de comparaison, où l'on considère des fonctions positives ou nulles de variables aléatoires.

Lemme 2.12. *Soit X une variable aléatoire à valeurs dans I . Si $\varphi_1, \varphi_2 : I \rightarrow \mathbb{R}^+$ vérifient $\varphi_1 \leq \varphi_2$, alors $E[\varphi_1(X)] \leq E[\varphi_2(X)]$. En particulier, si $\varphi_2(X)$ est intégrable, alors $\varphi_1(X)$ est intégrable.*

2.3 Moments et quantités associées

Définition 2.13. Si X est une variable aléatoire telle que X^k est intégrable, alors on dit que X admet un moment d'ordre k et on appelle la quantité $E[X^k]$, $k \in \mathbb{N}^*$ son k -ième moment (ou le k -ième moment de sa loi).

On dit encore que X est de puissance k intégrable lorsque X^k est intégrable. Les moments d'une variable aléatoire décrivent différents aspects de sa loi. Nous allons nous concentrer dans ce cours sur les quatre premiers moments (c'est-à-dire $k = 1, 2, 3, 4$), qui sont les plus importants. Tout d'abord quelques généralités. On peut toujours définir les moments pairs, puisque la fonction $x \mapsto x^k$ est positive pour $k \in \mathbb{N}$ pair, mais ces moments peuvent alors être infinis.

Lemme 2.14. *Soit $k < m$ deux entiers. Si X admet un moment d'ordre m , alors X admet un moment d'ordre k et*

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

Démonstration. On note que $|x|^k \leq 1 + |x|^m$ pour tout $x \in \mathbb{R}$ et on applique le lemme de comparaison 2.12. \square

Les calculs peuvent être très simples dans des cas particuliers :

Lemme 2.15. *Si la loi de X est symétrique, et X admet un moment d'ordre $k \in \mathbb{N}$ impair, alors $E[X^k] = 0$.*

Démonstration. C'est une application du lemme 2.10 avec la fonction $\varphi(x) = x^k$ impaire puisque k est impair. \square

2.3.1 Espérance

Le premier moment $E[X]$ est défini pour toute variable intégrable, ou positive ou nulle. Il est appelé *l'espérance* ou la *moyenne* de la variable aléatoire X . Si $E[X] = 0$, alors on dit que X est *centrée*. On appelle $X - E[X]$ la *variable centrée*. En effet, par linéarité de l'espérance, on a bien $E[X - E[X]] = E[X] - E[X] = \mu - \mu = 0$.

2.3.2 Variance

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

Définition 2.16. Soit X variable aléatoire. On dit que X est de carré intégrable si $E[X^2] < +\infty$. Dans ce cas, la *variance* de X est définie par

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

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

Lemme 2.17. Soit X variable aléatoire de carré intégrable. On a aussi :

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

C'est donc le second moment de la variable centrée $X - E[X]$. La variance est un *coefficient de dispersion* : il mesure comment les valeurs de X sont dispersées autour de l'espérance. Si la variance est grande, les valeurs sont très dispersées, alors que si la variance est petite, les valeurs sont concentrées autour de l'espérance.

Démonstration. On écrit :

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

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

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

□

Si X a une « dimension », c'est à dire si 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 son espérance μ .

Définition 2.18. Si X est une v.a. de carré intégrable, d'espérance μ et d'écart-type σ , on appelle $(X - \mu)/\sigma$ la *variable centrée réduite* (ou simplement *variable réduite*) associée à X .

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

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

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

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

Proposition 2.19. *Soit X une variable aléatoire de carré intégrable.*

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

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

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

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

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

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

2. On définit la variable aléatoire $Y = (X - \mu)^2$. Alors $Y \geq 0$ et $Y = 0$ si et seulement si $X = \mu$. Le Lemme 2.20 ci-dessous appliqué à Y donne alors

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

□

Lemme 2.20. *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.

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

Lemme 2.21. *Soit X une variable aléatoire de carré intégrable. On a :*

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

et le minimum est atteint en un unique point égal à $E[X]$

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

Démonstration. Par linéarité de l'intégrale,

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

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

Un corollaire simple de ce résultat est alors que

Corollaire 2.22. *Soit X variable aléatoire à valeurs dans $[0, 1]$. Alors X est de carré intégrable et*

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

Démonstration. D'après le lemme précédent, pour tout $y \in \mathbb{R}$, $\text{Var}(X) \leq E[(X - y)^2] \leq 1/4$ et donc pour $y = 1/2$ en particulier, en utilisant que car $|X - 1/2| \leq 1/2$ par hypothèse, on obtient : $\text{Var}(X) \leq E[(X - 1/2)^2] \leq E[1/4] = 1/4$. □

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'au petites valeurs de Y^2 , et petit si elle donne beaucoup de masse au 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

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

Comme son nom l'indique, il s'agit d'un outil puissant pour calculer des moments au moyen de développements en série entière :

Proposition 2.23. *Supposons qu'il existe $t_0 > 0$ tel que $\varphi_X(t) < \infty$ pour tout $|t| < t_0$. Alors pour tout $|t| < t_0$,*

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

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

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

où $\varphi_X^{(k)}$ est la k -ième dérivée de la fonction 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 2.24. *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 2.25. $a, b \in \mathbb{R}$.

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

Exemple 2.26. $X \sim \text{Poi}(\lambda)$, $\lambda \geq 0$. 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 2.27. $X \sim \mathcal{N}(0, 1)$. 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 2.28. $X \sim \text{Exp}(1)$. 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 2.29. *Soit $y \geq 0$ et $x > 0$. Alors*

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

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

Lemme 2.30. *Soit X une v.a. et $B \subset \mathbb{R}$. Alors*

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

Démonstration. La v.a. $\mathbb{1}_{X \in B}$ 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 2.31 (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 2.29 montre que $\mathbb{1}_{X \geq x} \leq X/x$. Par le Lemme 2.30,

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

par linéarité de l'espérance. \square

Corollaire : preuve du Lemme 2.20. Si $Y \sim \delta_0$, alors par définition de l'espérance, $E[Y] = 0 \times 1 = 0$. Si $Y \not\sim \delta_0$, alors il existe $x > 0$ tel que $P(Y \geq x) > 0$. Par l'inégalité de Markov,

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

Ceci conclut la preuve. \square

2.5.2 Inégalité de Tchebychev

Theorème 2.32 (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 2.31) pour obtenir

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

\square

2.5.3 Inégalité de Jensen

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

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

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

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

Theorème 2.33. *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,*

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

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

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

Par conséquent,

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

car $x_0 = E[X]$. □

2.5.4 Une formule générale pour l'espérance d'une variable aléatoire positive

Lemme 2.34. *Soit X variable aléatoire positive ou nulle. On a :*

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

On notera en particulier que l'intégrabilité de X implique l'intégrabilité de la fonction $t \mapsto 1 - F_X(t)$ au voisinage de $+\infty$, alors qu'une application de l'inégalité de Markov à une variable aléatoire X intégrable donnait seulement

$$1 - F_X(t) = P(X > t) \leq \frac{E[X]}{t}$$

ce qui ne suffit pas pour établir cette intégrabilité (la fonction inverse $\mathbb{R}^* \rightarrow \mathbb{R}^*$, $t \mapsto 1/t$ n'est en effet pas intégrable au voisinage de $+\infty$).

On ne peut bien sûr démontrer ce résultat que dans le cas des variables discrètes ou à densité pour lesquelles l'espérance a été définie ; cependant, la formule a un sens dès lors que la fonction de répartition est définie, et on pourrait de fait la prendre comme une **définition** de l'espérance pour toute variable aléatoire positive ou nulle ! Cependant pour des raisons pédagogiques, il nous a paru plus prudent de nous en tenir à l'exposition précédente.

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

$$\begin{aligned}
 E[X] &= E \left[\int_0^\infty \mathbb{1}_{X>t} dt \right] \\
 &= \int_0^\infty \left(\int_0^\infty \mathbb{1}_{x>t} dt \right) f_X(x) dx \\
 &= \int_0^\infty \left(\int_0^\infty \mathbb{1}_{x>t} f_X(x) dx \right) dt && \text{de Fubini positif} \\
 &= \int_0^\infty E[\mathbb{1}_{X>t}] dt \\
 &= \int_0^\infty P(X > t) dt && \text{du lemme 2.30}
 \end{aligned}$$

Le cas où X est discrète est tout à fait analogue, nous ne le répétons pas (il s'agit de remplacer l'intégrale par rapport à $f_X(x)dx$ par une somme pondérée par $P(X = x)$. \square

Le résultat suivant sur les variables aléatoires à valeurs entières est très utile en pratique :

Corollaire 2.35. *Si de plus X est une variable aléatoire discrète à valeurs dans \mathbb{N} , alors*

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

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

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

qui implique bien le résultat cherché puisque :

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

\square