

Cours de Probabilités

Joseph Lehec

Table des matières

1	Théorie de la mesure	3
1.1	Définitions	3
1.2	La mesure de Lebesgue	4
1.3	Convergence monotone	5
1.4	Exercices	6
2	Espaces de probabilité	8
2.1	Définition	8
2.2	Conditionnement	9
2.3	Indépendance	10
2.4	Exercices	11
3	Variables aléatoires	13
3.1	Définitions	13
3.2	Variables aléatoires, fonctions de répartition	14
3.3	Loi d'une variable aléatoire	15
3.4	Variables aléatoires discrètes	16
3.5	Variables aléatoires continues	17
3.6	Exercices	18
4	Intégration	19
4.1	Intégrale d'une fonction étagée positive	19
4.2	Intégrale d'une fonction positive	21
4.3	Intégrale d'une fonction réelle	22
4.4	Exemples	23
4.5	Exercices	25
5	Espérance	27
5.1	Définition, propriétés	27
5.2	Moments, variance	28
5.3	Fonction génératrice des moments	29
5.4	Formule fondamentale	30
5.5	Retour sur les variables continues	31
5.6	Lois classiques	32
5.7	Changement de variable I	35
5.8	Un exemple ni discret ni continu	37
5.9	Exercices	37
6	Variables indépendantes	40
6.1	Définition	40
6.2	Covariance	41
6.3	Exercices	41
7	Marche aléatoire simple	43
7.1	Dénombrement de trajectoires	43
7.2	Temps de retour en 0 pour la marche symétrique	44
7.3	Loi de l'arc sinus	46
7.4	Loi des grands nombres	47
7.5	Théorème de la limite centrale	48

8	Intégrales doubles	50
8.1	Théorème de Fubini	50
8.2	Lois jointes	51
8.3	Couples continus	51
8.4	Retour sur l'indépendance	52
8.5	Changement de variable II	53
8.6	Exercices	55
9	Conditionnement	57
9.1	Cas discret	57
9.2	Cas continu	59
9.3	Exercices	60

1 Théorie de la mesure

1.1 Définitions

Définition 1.1. Soit E un ensemble. On appelle *tribu* sur E un sous-ensemble \mathcal{A} des parties de E vérifiant

- (i) $\emptyset \in \mathcal{A}$
- (ii) si $A \in \mathcal{A}$ alors $A^c \in \mathcal{A}$
- (iii) si $(A_n)_{n \in \mathbb{N}}$ est une suite d'éléments de \mathcal{A} , alors $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$

Exemple 1.2. Soit E un ensemble

- $\{\emptyset, E\}$ est une tribu.
- Soit $A \subset E$, alors $\{\emptyset, A, A^c, E\}$ est une tribu.
- $\mathcal{P}(E)$ est une tribu.

Lemme 1.3. Soit E un ensemble et \mathcal{A} une tribu sur E . Alors

- $E \in \mathcal{A}$
- \mathcal{A} est stable par union finie
- \mathcal{A} est stable par intersection dénombrable (ou finie).

Démonstration. Comme $\emptyset \in \mathcal{A}$ et comme \mathcal{A} est stable par passage au complémentaire $E = \emptyset^c \in \mathcal{A}$. Pour le deuxième point il suffit de remarquer que

$$A \cup B = A \cup B \cup \emptyset \cup \emptyset \dots$$

Pour le troisième on écrit

$$\bigcap_n A_n = \left(\bigcup_n A_n^c \right)^c. \quad \square$$

Définition 1.4. Étant donné un ensemble E et une tribu \mathcal{A} sur E , on appelle *mesure* sur (E, \mathcal{A}) une application

$$\mu: \mathcal{A} \rightarrow [0, +\infty]$$

vérifiant

$$\mu(\emptyset) = 0$$

et telle que pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} deux à deux disjoints ($A_n \cap A_m = \emptyset$ pour tous $n \neq m$) on ait

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} \mu(A_n)$$

Cette propriété est appelée σ -*additivité* ou additivité dénombrable.

Remarque. Le membre de gauche dans l'égalité précédente reste inchangé si l'on permute les A_n (l'union ne dépend pas de l'ordre). Pour que la définition précédente ait un sens il faut donc que le membre de droite soit invariant par permutation des A_n . C'est le cas puisque la somme d'une série à termes **positifs** ne dépend pas de l'ordre des termes.

Définition 1.5. On appelle *espace mesuré* tout triplet (E, \mathcal{A}, μ) , où E est un ensemble, \mathcal{A} une tribu sur E , et μ une mesure de probabilité sur (E, \mathcal{A}) .

Lemme 1.6. Un espace mesuré (E, \mathcal{A}, μ) vérifie les propriétés suivantes :

- si $A \cap B = \emptyset$ alors $\mu(A \cup B) = \mu(A) + \mu(B)$ (*additivité*),
- si $A \subset B$ alors $\mu(A) \leq \mu(B)$.

Démonstration. On écrit

$$\begin{aligned} \mu(A \cup B) &= \mu(A \cup B \cup \emptyset \cup \emptyset \cup \dots) \\ &= \mu(A) + \mu(B) + \mu(\emptyset) + \mu(\emptyset) + \dots = \mu(A) + \mu(B). \end{aligned}$$

Si $A \subset B$ on a $B = A \cup (B \setminus A)$ et l'union est disjointe. Donc $\mu(B) = \mu(A) + \mu(B \setminus A)$ et donc $\mu(B) \geq \mu(A)$. \square

Un peu de terminologie : Si $\mu(E) < +\infty$ on dit que μ est une mesure *finie*. Si $\mu(E) = 1$ on dit que μ est une mesure de probabilité. Les éléments de la tribu \mathcal{A} sont appelés *ensembles mesurables*. Un ensemble A vérifiant $\mu(A) = 0$ est dit *négligeable*. Si une propriété a lieu en dehors d'un ensemble négligeable, on dit qu'elle est vraie *presque partout*.

Exemple 1.7. Soit $a \in E$, on définit une mesure de probabilité sur $(E, \mathcal{P}(E))$ en posant $\delta_a(A) = 1$ si $a \in A$ et $\delta_a(A) = 0$ sinon. Cette mesure est appelée *masse de Dirac* en a .

Exemple 1.8. Pour tout sous-ensemble A de \mathbb{N} posons $\mu(A) = \text{card}(A)$ si A est fini et $\mu(A) = +\infty$ si A est infini. Alors μ définit une mesure sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ appelée mesure de comptage.

1.2 La mesure de Lebesgue

Dans cette partie on se place sur l'ensemble \mathbb{R} des nombres réels. Les principaux théorèmes de cette partie seront admis, leur démonstration dépasse le cadre d'une introduction à la théorie de la mesure.

Remarquons qu'une intersection de tribus est encore une tribu, ce qui justifie la définition suivante.

Définition 1.9. On appelle tribu des *Boréliens* ou tribu *Borélienne*, notée $\mathcal{B}(\mathbb{R})$, la tribu engendrée par les intervalles de \mathbb{R} ; c'est-à-dire la tribu obtenue en prenant l'intersection de toutes les tribus contenant les intervalles.

Il faut retenir cette définition ainsi :

- $\mathcal{B}(\mathbb{R})$ est une tribu.
- Tout intervalle de \mathbb{R} appartient $\mathcal{B}(\mathbb{R})$.
- Si \mathcal{A} est une tribu contenant les intervalles, alors $\mathcal{B}(\mathbb{R}) \subset \mathcal{A}$.

On ne change rien à la définition précédente en ne considérant que les intervalles ouverts, ou que les intervalles fermés : un intervalle ouvert peut toujours s'écrire comme une union dénombrables d'intervalle fermés. De même, la tribu $\mathcal{B}(\mathbb{R})$ contient les singletons. En effet, soit on considère que $\{x\} = [x, x]$ est un intervalle et donc appartient à $\mathcal{B}(\mathbb{R})$ par définition, soit on écrit

$$\{x\} = \bigcap_{n \geq 1} [x, x + 1/n].$$

Par stabilité par union dénombrable la tribu $\mathcal{B}(\mathbb{R})$ contient tous les ensembles dénombrables, en particulier l'ensemble \mathbb{Q} des nombres rationnels. En fait, il est difficile de construire un sous-ensemble de \mathbb{R} qui ne soit pas Borélien.

Théorème 1.10. Soient μ, ν deux mesures définies sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, si μ et ν coïncident sur les intervalles alors μ et ν sont égales. Autrement dit si $\mu(I) = \nu(I)$ pour tout intervalle I alors $\mu(B) = \nu(B)$ pour tout Borélien B .

Théorème 1.11. Il existe une unique mesure \mathcal{L} sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ vérifiant

$$\mathcal{L}([a, b]) = b - a$$

pour tout intervalle $[a, b]$. Cette mesure est appelée mesure de Lebesgue.

On admettra ces deux théorèmes. Remarquons quand même que l'unicité de la mesure de Lebesgue est une conséquence du Théorème 1.10.

Proposition 1.12. La mesure de Lebesgue est invariante par translation : pour tout $x \in \mathbb{R}$ et pour tout $B \in \mathcal{B}(\mathbb{R})$ l'ensemble $x + B$ est aussi Borélien et

$$\mathcal{L}(x + B) = \mathcal{L}(B).$$

Démonstration. Commençons par montrer que si $B \in \mathcal{B}(\mathbb{R})$ alors $x + B \in \mathcal{B}(\mathbb{R})$. On fixe $x \in \mathbb{R}$ et on pose

$$\mathcal{A} = \{B \subset \mathbb{R}, x + B \in \mathcal{B}(\mathbb{R})\}.$$

On vérifie aisément que \mathcal{A} est une tribu qui contient les intervalles. Par conséquent $\mathcal{B}(\mathbb{R}) \subset \mathcal{A}$, ce qu'il fallait démontrer.

Pour $B \in \mathcal{B}(\mathbb{R})$ on définit

$$\mathcal{M}(B) = \mathcal{L}(x + B).$$

Alors \mathcal{M} est une mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (exercice) et pour tout intervalle $[a, b]$ on a

$$\mathcal{M}([a, b]) = \mathcal{L}([x + a, x + b]) = (b + x) - (a + x) = b - a = \mathcal{L}([a, b]).$$

Les mesures \mathcal{M} et \mathcal{L} coïncident donc sur les intervalles. D'après le Théorème 1.10 on obtient $\mathcal{L} = \mathcal{M}$, ce qui est le résultat. \square

1.3 Convergence monotone

Dans toute la suite du cours, étant donné une suite réelle $(u_n)_{n \geq 0}$ et $l \in \mathbb{R} \cup \{+\infty\}$, on écrit $u_n \nearrow l$ lorsque la suite u_n est croissante et tend vers l . De même, si $l \in \mathbb{R} \cup \{-\infty\}$ on écrit $u_n \searrow l$ si la suite est décroissante et tend vers l . Étant donnée une suite $(A_n)_{n \geq 0}$ d'ensembles, la notation $A_n \nearrow A$ signifie que la suite est croissante pour l'inclusion ($A_n \subset A_{n+1}$ pour tout $n \in \mathbb{N}$) et que

$$\bigcup_{n \geq 0} A_n = A.$$

Enfin $A_n \searrow A$ signifie que la suite est décroissante pour l'inclusion et que $\bigcap_n A_n = A$. Les résultats de cette sous-partie **très importants**.

Proposition 1.13. Soit (E, \mathcal{A}, μ) un espace mesuré et soient $(A_n)_{n \in \mathbb{N}}$, A des éléments de \mathcal{A} . Si $A_n \nearrow A$ alors $\mu(A_n) \nearrow \mu(A)$.

Remarque. L'hypothèse de monotonie est fondamentale. Par exemple, considérons la mesure de comptage μ sur \mathbb{N} et les ensembles A_n suivants : $A_0 = \{0\}$ et $A_n = \{1\}$ pour $n \geq 1$. On a $\mu(A_n) = 1$ pour tout n mais $\mu(\bigcup_n A_n) = \mu(\{0, 1\}) = 2$.

Démonstration. On pose $A'_0 = A_0$ et $A'_n = A_n \setminus A_{n-1}$ pour tout $n \geq 1$. Les propriétés suivantes sont laissées en exercice (faire un dessin) :

- Les A'_n sont deux à deux disjoints.
- $A_n = A'_0 \cup A'_1 \cup \dots \cup A'_n$ pour tout n .
- $\bigcup_{n=0}^{\infty} A'_n = \bigcup_{n=0}^{\infty} A_n$.

Par σ -additivité, on obtient

$$\mu(A_n) = \mu(A'_0) + \dots + \mu(A'_n) \nearrow \sum_{k=0}^{\infty} \mu(A'_k) = \mu\left(\bigcup_{k=0}^{+\infty} A'_k\right) = \mu\left(\bigcup_{k=0}^{+\infty} A_k\right). \quad \square$$

Corollaire 1.14. Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'éléments de \mathcal{A} . On a

$$\mu\left(\bigcup_{n=0}^{\infty} A_n\right) \leq \sum_{n=0}^{\infty} \mu(A_n).$$

Démonstration. On commence par montrer par récurrence que pour tout $k \in \mathbb{N}$

$$\mu\left(\bigcup_{n=0}^k A_n\right) \leq \sum_{n=0}^k \mu(A_n).$$

Laissons cette partie de la preuve en exercice. Ensuite on pose $B_k = \bigcup_{n=0}^k A_n$. D'après ce qui précède, pour tout $k \in \mathbb{N}$

$$\mu(B_k) \leq \sum_{n=0}^k \mu(A_n) \leq \sum_{n=0}^{\infty} \mu(A_n). \quad (1)$$

Comme $B_k \nearrow \bigcup_n A_n$ on a $\mu(B_k) \nearrow \mu(\bigcup_n A_n)$. On obtient donc le résultat en passant à la limite dans l'inégalité (1). \square

Proposition 1.15. Soit $(B_n)_{n \geq 0}, B$ des éléments de \mathcal{A} . Si $B_n \searrow B$ et s'il existe m tel que $\mu(B_m) < +\infty$ (en particulier si la mesure μ est finie) alors $\mu(B_n) \searrow \mu(B)$.

Démonstration. On pose $A_n = B_m \setminus B_n$ pour $n \geq m$ et $A = B_m \setminus B$. Alors $A_n \nearrow A$ et donc $\mu(A_n) \nearrow \mu(A)$. Comme $\mu(A_n) = \mu(B_m) - \mu(B_n)$ et comme $\mu(B_m) < +\infty$ ceci revient à $\mu(B_n) \searrow \mu(B)$. \square

Remarque. L'hypothèse "il existe m tel que $\mu(B_m) < +\infty$ " est nécessaire. En effet, si B_n est l'intervalle $[n, +\infty[$, alors $B_n \searrow \emptyset$, pourtant $\mathcal{L}(B_n) = +\infty$ pour tout n .

1.4 Exercices

Exercice 1.1. Soit $(u_k)_{k \geq 0}$ une suite de réels positifs, Montrer que la somme de la série $\sum_{k \geq 0} u_k$ ne dépend pas de l'ordre des termes : si $\sigma: \mathbb{N} \rightarrow \mathbb{N}$ est une fonction bijective alors

$$\sum_{k \geq 0} u_k = \sum_{k \geq 0} u_{\sigma(k)}.$$

Indication : Montrer que

$$\sum_{k \geq 0} u_k = \sup \left\{ \sum_{k \in I} u_k, I \subset \mathbb{N}, I \text{ fini} \right\}.$$

Exercice 1.2. Montrer qu'une intersection quelconque de tribus est une tribu. Est-ce que cela marche aussi avec l'union ?

Exercice 1.3. Montrer que la mesure de comptage sur \mathbb{N} est bien une mesure.

Exercice 1.4. Soit (E, \mathcal{A}, μ) un espace mesuré et soient A_1, \dots, A_n des ensembles mesurables. Montrer que

1. Montrer que $\mu(A_1 \cup A_2) \leq \mu(A_1) + \mu(A_2)$.
2. Montrer par récurrence sur n que $\mu(A_1 \cup \dots \cup A_n) \leq \mu(A_1) + \dots + \mu(A_n)$.

Exercice 1.5. Soit (E, \mathcal{A}, μ) un espace mesuré et soient A_1, \dots, A_n des ensembles mesurables. Montrer que

1. Montrer que $\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2) - \mu(A_1 \cap A_2)$.
2. Montrer que

$$\begin{aligned} \mu(A_1 \cup A_2 \cup A_3) &= \mu(A_1) + \mu(A_2) + \mu(A_3) \\ &\quad - \mu(A_1 \cap A_2) - \mu(A_1 \cap A_3) - \mu(A_2 \cap A_3) + \mu(A_1 \cap A_2 \cap A_3). \end{aligned}$$

3. Donner une formule pour $\mu(A_1 \cup \dots \cup A_n)$.

Exercice 1.6. Montrer qu'un intervalle ouvert est réunion dénombrable d'intervalles fermés. Montrer qu'un intervalle fermé est intersection dénombrable d'intervalles ouverts.

Exercice 1.7. Montrer que presque tous les nombres réels sont irrationnels.

Exercice 1.8. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, montrer qu'une union dénombrable d'ensembles négligeables est négligeable.

Exercice 1.9. Soit $f: \mathbb{R} \rightarrow \mathbb{R}_+$ telle que $\mathcal{L}(\{x: f(x) \geq t\}) = 0$ pour tout réel $t > 0$. Montrer que $f(x) = 0$ presque partout.

Exercice 1.10. Soit U un sous-ensemble ouvert de \mathbb{R} . Pour $x \in U$ on pose

$$\begin{aligned} a(x) &= \inf \{a < x,]a, x[\subset U\} \\ b(x) &= \sup \{b > x,]x, b[\subset U\} \end{aligned}$$

1. Montrer que

$$U = \bigcup_{r \in \mathbb{Q} \cap U}]a(r), b(r)[$$

2. En déduire que les ouverts sont Boréliens.

Exercice 1.11. Pour $x, y \in [-1, 1]$, on pose $x \sim y$ si $x - y \in \mathbb{Q}$.

1. Montrer que \sim est une relation d'équivalence.

Soit A un sous-ensemble de $[-1, 1]$ contenant exactement un élément de chaque classe d'équivalence. On pose

$$B = \bigcup_{r \in \mathbb{Q} \cap [-2, 2]} (r + A).$$

2. Montrer que les $r + A$ sont deux à deux disjoints.

3. Montrer que $[-1, 1] \subset B \subset [-3, 3]$.

4. Montrer par l'absurde que A n'est pas Borélien.

2 Espaces de probabilité

2.1 Définition

Définition 2.1. On appelle *espace de probabilité* tout triplet (Ω, \mathcal{A}, P) , où Ω est un ensemble, \mathcal{A} une tribu sur Ω , et P une mesure de probabilité sur (Ω, \mathcal{A}) .

Cette notion permet de modéliser n'importe quelle expérience aléatoire.

Exemple 2.2. Le pile ou face est modélisé par l'espace (Ω, \mathcal{A}, P) avec $\Omega = \{P, F\}$

$$\mathcal{A} = \mathcal{P}(\Omega) = \{\emptyset, \{P\}, \{F\}, \{P, F\}\}$$

et P la probabilité donnée par $P(\{P\}) = P(\{F\}) = 1/2$.
Pour le lancer de dé, on a $\Omega = \{1, 2, 3, 4, 6\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ et

$$P(\{1\}) = P(\{2\}) = \dots = P(\{6\}) = 1/6.$$

De manière générale si Ω un ensemble fini, on définit une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ en posant

$$P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}.$$

Cette probabilité est appelée probabilité uniforme sur Ω .

Exemple 2.3. Une main de poker entre aussi dans ce cadre : l'espace des observables est l'ensemble des sous-ensembles de 5 cartes (ou mains). La probabilité d'obtenir un as (par exemple) est $p = n/N$ où n est le nombre de mains contenant un as et N est le nombre total de mains. On a $N = \binom{52}{5}$. Le nombre de mains ne contenant pas d'as est $N - n = \binom{48}{5}$. On obtient donc

$$p = 1 - \frac{\binom{48}{5}}{\binom{52}{5}} \approx 0,44.$$

Exemple 2.4 (aiguille de Buffon). On lance une aiguille et on regarde l'angle que fait l'aiguille avec une direction donnée (disons le nord). Supposons (ce qui semble naturel) que la probabilité que l'aiguille tombe dans un intervalle d'angle donné soit proportionnelle à la longueur de cet intervalle. On peut modéliser cette expérience ainsi. On prend $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ et on pose pour tout $B \in \mathcal{B}(\mathbb{R})$

$$P(B) = \frac{\mathcal{L}(B \cap [0, 2\pi])}{\mathcal{L}([0, 2\pi])} = \frac{\mathcal{L}(B \cap [0, 2\pi])}{2\pi}.$$

Remarque. De manière générale, étant donnée $a < b$ on définit une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ en posant

$$P(B) = \frac{\mathcal{L}(B \cap [a, b])}{b - a}.$$

Cette mesure de probabilité est appelée *probabilité uniforme* sur l'intervalle $[a, b]$.

Un peu de terminologie : l'espace Ω est appelé *espace des observables*, c'est l'ensemble des issues possibles de l'expérience aléatoire. Les éléments de la tribu \mathcal{A} sont appelés *événements*. Un événement A vérifiant $P(A) = 1$ est dit *quasi-certain*. On dit aussi que A a lieu *presque sûrement*. Un événement vérifiant $P(A) = 0$ est dit *négligeable*.

Rappelons les propriétés de convergence monotone vues précédemment : si $A_n \nearrow A$ alors $P(A_n) \nearrow P(A)$ et si $B_n \searrow B$ alors $P(B_n) \searrow P(B)$.

Exemple 2.5. Admettons qu'on puisse construire un espace de probabilité permettant de modéliser une suite infinie de jeu de pile ou face. Soit A_n l'événement "les n premiers lancers donnent face" et A l'événement "tous les lancers tombent sur face". D'une part $P(A_n) = 2^{-n}$ et d'autre part $A_n \searrow A$. Donc $2^{-n} = P(A_n) \searrow P(A)$. Donc $P(A) = 0$. L'événement A est négligeable, on finit presque sûrement par faire pile.

2.2 Conditionnement

Un énoncé de probabilité conditionnelle est un énoncé du type « Si B se produit alors la probabilité que A se produise est p ». On peut penser par exemple que B est l'événement « il pleut » et A « le bus est en retard ». Mathématiquement, la définition est la suivante.

Définition 2.6. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Soit B un événement vérifiant $\mathbb{P}(B) > 0$. Pour tout événement A on pose

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

Cette quantité est appelée probabilité de B sachant A ou conditionnellement à A .

Exemple 2.7. Une famille a deux enfants. Quelle est la probabilité que les deux soient des garçons conditionnellement au fait qu'au moins l'un des deux est un garçon ? Avec des notations évidentes, l'espace de probabilité est $\Omega = \{GG, GF, FG, FF\}$ muni de la probabilité uniforme. La probabilité cherchée est

$$\mathbb{P}(\{GG\} | \{GG, GF, FG\}) = \frac{\mathbb{P}(\{GG\})}{\mathbb{P}(\{GG, GF, FG\})} = \frac{1}{3}.$$

Une réponse fautive courante à ce dernier exemple est de dire : « On sait que l'un des deux enfants est un garçon, la probabilité cherchée est la probabilité que l'autre soit aussi un garçon, à savoir $1/2$. » L'erreur est la suivante : l'événement « l'autre enfant est un garçon » n'a en fait aucun sens.

Lemme 2.8 (Formule de Bayes). Soit A et B des événements non négligeables

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

Démonstration. $\mathbb{P}(A | B)\mathbb{P}(B) = \mathbb{P}(A \cap B) = \mathbb{P}(B | A)\mathbb{P}(A)$. □

Définition 2.9. Soit Ω un ensemble. Soit I un ensemble (fini ou infini) et soit $(A_i)_{i \in I}$ une famille de parties de Ω . On dit que $(A_i)_{i \in I}$ est une *partition* de Ω si les deux propriétés suivantes sont vérifiées.

- $\bigcup_{i \in I} A_i = \Omega$
- Les A_i sont deux à deux disjoints

Autrement dit, la famille $(A_i)_{i \in I}$ est une partition de Ω si pour tout $\omega \in \Omega$ il existe un et un seul $i \in I$ tel que $\omega \in A_i$.

Exemple 2.10. Pour tout $A \subset \Omega$, l'ensemble $\{A, A^c\}$ est une partition de Ω .

Proposition 2.11 (Formule des probabilités totales). Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Soit I un ensemble **fini ou dénombrable** — penser $I = \{1, \dots, n\}$ pour un certain n ou $I = \mathbb{N}$ — et soit $(B_i)_{i \in I}$ une partition de Ω en événements. Alors pour tout événement A

$$\begin{aligned} \mathbb{P}(A) &= \sum_{i \in I} \mathbb{P}(A \cap B_i) \\ &= \sum_{i \in I} \mathbb{P}(A | B_i)\mathbb{P}(B_i), \quad \text{si } \mathbb{P}(B_i) > 0 \text{ pour tout } i. \end{aligned}$$

Démonstration. Comme les B_i recouvrent Ω on a $A = \bigcup_{i \in I} (A \cap B_i)$. Et comme les B_i sont deux à deux disjoints, les $A \cap B_i$ aussi. De plus I est fini ou dénombrable donc

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \cap B_i). \quad \square$$

Un cas particulier qui revient souvent est la formule

$$\mathbb{P}(A) = \mathbb{P}(A | B)\mathbb{P}(B) + \mathbb{P}(A | B^c)\mathbb{P}(B^c)$$

valable dès que $0 < \mathbb{P}(B) < 1$.

Exemple 2.12 (Faux positifs). Une maladie affecte une personne sur 1000. Le test de dépistage n'est pas parfait : le résultat est toujours positif pour une personne malade et pour une personne saine il est positif (donc erroné) 2 fois sur 100. Quelle est la probabilité qu'une personne ayant un résultat positif au test soit effectivement malade ?

Soit T l'événement « le test est positif » et M l'événement « la personne est malade ». On cherche $P(M | T)$. On écrit

$$P(M | T) = P(T | M) \frac{P(M)}{P(T)}.$$

D'après les données du problème $P(T | M) = 1$ et $P(M) = 0,001$. De plus

$$P(T) = P(T | M)P(M) + P(T | M^c)P(M^c) = 1 \times 0,001 + 0,02 \times 0,999.$$

En regroupant tout on trouve que $P(M | T)$ est de l'ordre de 5%. Le test est probablement erroné.

2.3 Indépendance

De manière intuitive on dit que A est indépendant de B si savoir B ne change pas la probabilité de A . C'est-à-dire si

$$P(A | B) = P(A).$$

Pour que cette formule ait un sens on est obligé de supposer que $P(B) > 0$, ce qui n'est pas le cas dans la définition suivante.

Définition 2.13. Soit (Ω, \mathcal{A}, P) un espace de probabilité. Deux événements A, B sont dits *indépendants* si

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

Plus généralement soit $(A_i)_{i \in I}$ une famille d'événements (I est un ensemble quelconque). On dit que les A_i sont indépendants si

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j)$$

pour tout J fini inclus dans I .

Exemple 2.14. On tire une carte dans un paquet de 52. L'événement A : « tirer un roi » est indépendant de B : « tirer un pique ». En effet $P(A \cap B)$ est la probabilité de tirer le roi de pique, soit $1/52$, qui est bien égal à $P(A)P(B) = (1/13) \times (1/4)$.

Remarque. On dit donc que les événements A, B et C sont indépendants si les quatre propriétés suivantes sont vérifiées

$$\begin{aligned} P(A \cap B) &= P(A)P(B) \\ P(A \cap C) &= P(A)P(C) \\ P(B \cap C) &= P(B)P(C) \\ P(A \cap B \cap C) &= P(A)P(B)P(C). \end{aligned}$$

Lorsque seules les trois premières sont vérifiées on dit que A, B et C sont *deux à deux* indépendants. Quand les quatre sont vérifiées on dit parfois que les ensembles sont globalement (ou mutuellement) indépendants pour bien marquer la différence.

Exemple 2.15. On jette deux pièces, les événements suivants

- la première pièce tombe sur pile
- la deuxième tombe sur pile
- les deux pièces donnent le même résultat

sont deux à deux indépendants, mais pas mutuellement indépendants.

2.4 Exercices

Exercice 2.1. On considère une classe de N étudiants, tous nés en 1990. Quelle est la probabilité (en fonction de N) d'avoir deux étudiants nés le même jour ? Déterminer numériquement à partir de quelle valeur de N cette probabilité devient supérieure à $1/2$.

Exercice 2.2. Montrer qu'une intersection dénombrable d'événements quasi-certains est un événement quasi-certain.

Exercice 2.3. On effectue une suite infinie de pile ou face.

1. Montrer que presque sûrement un pile finit par sortir.
2. Montrer que presque sûrement n'importe séquence finie de piles et de faces (par exemple $PFPPFFPP$) finit par sortir.

Exercice 2.4. Montrer qu'un événement A est indépendant de lui-même si et seulement si $P(A) = 0$ ou $P(A) = 1$. Si $P(A) = 0$, montrer que A est indépendant de tout événement B . Ce résultat reste-t-il vrai si $P(A) = 1$?

Exercice 2.5. On suppose que A et B sont indépendants. Montrer que A est indépendant de B^c , que A^c est indépendant de B^c .

Exercice 2.6. On lance deux dés, montrer que l'événement « la somme fait 7 » est indépendant du score du premier dé. Est-ce que cela marche encore si on remplace 7 par 6 ?

Exercice 2.7. On lance un dé n fois. On appelle A_{ij} l'événement : « les scores du i -ème et du j -ème lancers sont égaux ». Montrer que les A_{ij} sont indépendants deux à deux mais pas mutuellement.

Exercice 2.8. Roulette russe : un revolver à 6 coups contient une seule balle mais on ne sait pas à quel endroit du barillet. Le premier joueur place le revolver sur sa tempe et presse la gâchette. S'il survit le deuxième joueur fait de même. Vaut-il mieux jouer en premier ou en second ? La taille du barillet importe-t-elle ?

Exercice 2.9 (Paradoxe du prisonnier). Trois prisonniers sont condamnés à mort mais le tyran décide d'en libérer un. Le prisonnier A apprend par le garde que le prisonnier B sera bien exécuté. Le survivant étant A ou C , on peut dire que A a une chance sur deux d'être libéré. D'un autre côté, entre B et C , il y a au moins une personne qui doit être exécutée, ça ne change rien pour A de savoir si c'est B ou C , la probabilité qu'il soit libéré est toujours de $1/3$. Qu'en pensez vous ? Indication : il faut préciser ce qu'aurait dit le garde si B avait été gracié. On propose deux hypothèses :

1. Le garde dit « B sera exécuté » ou « B sera gracié ».
2. Le garde dit « B sera exécuté » ou « C sera exécuté », s'ils le sont tous les deux, il tire au sort.

Quelle est, dans chacun des cas, la probabilité que A soit libéré sachant que le garde dit que B sera exécuté ?

Montrer que pour tout $p \in [0, 1/2]$, on peut donner au garde un algorithme qui fasse que la probabilité conditionnelle précédente soit p .

Exercice 2.10. Soit A_1, A_2, \dots une suite d'événements indépendants. On suppose que la série $\sum_{n=1}^{+\infty} P(A_k)$ diverge.

1. Déterminer la probabilité p_n qu'aucun des n premiers événements ne se réalise.
2. Montrer que p_n tend vers 0 (on pourra utiliser l'inégalité $1 - t \leq e^{-t}$).
3. En déduire que presque sûrement au moins un des événements A_n se produit.

Exercice 2.11 (Ruine du joueur). On considère une suite de pile ou face avec une pièce truquée tombant sur pile avec probabilité p . Au temps 0 le joueur possède une somme n . Il gagne 1 à chaque pile et perd 1 à chaque face. Le jeu s'arrête lorsque le joueur est ruiné ou lorsqu'il a atteint un montant N donné. On note r_n la probabilité que le joueur soit ruiné à la fin du jeu.

1. En conditionnant par rapport au résultat du premier pile ou face, montrer que

$$r_n = pr_{n+1} + (1-p)r_{n-1}, \quad n \in \{1, \dots, N-1\}.$$

2. Que valent r_0 et r_N ?
3. Déterminer r_n (attention au cas $p = 1/2$).

Exercice 2.12. n passagers montent dans un avion l'un après l'autre. Chacun a un numéro de siège mais le premier passager s'assied au hasard. Les passagers suivants s'asseyent à leur place lorsqu'elle est libre et choisissent un siège libre au hasard sinon. On appelle p_n la probabilité que le dernier passager s'asseye à sa place.

1. En conditionnant par rapport au siège choisi par le premier passager, exprimer p_n en fonction de p_2, \dots, p_{n-1} .
2. Déterminer p_n .

3 Variables aléatoires

3.1 Définitions

On ne sera pas toujours intéressé par le résultat complet d'une expérience aléatoire mais plutôt par une conséquence de ce résultat, c'est-à-dire une fonction de ce résultat. Une telle fonction est appelée *variable aléatoire*.

Exemple 3.1. On lance deux pièces. On appelle X le nombre de piles obtenus. De manière formelle, on a $\Omega = \{PP, PF, FP, FF\}$ muni de la probabilité uniforme, et X est définie par

$$X(PP) = 2, X(PF) = 1, X(FP) = 1, X(FF) = 0.$$

Étant donnée une fonction X de Ω dans \mathbb{R} , on s'intéresse à la *répartition* de X dans \mathbb{R} . On voudrait par exemple estimer la probabilité

$$P(\{\omega \in \Omega: X(\omega) \leq x\}),$$

pour tout réel x . Celle-ci ne sera définie que si l'ensemble

$$\{\omega \in \Omega: X(\omega) \leq x\} \in \mathcal{A}.$$

Cette remarque motive les définitions suivantes.

Définition 3.2. Soit E, F des ensembles et soient \mathcal{A}, \mathcal{B} des tribus de E et F respectivement. Une fonction $f: E \rightarrow F$ est dite mesurable si l'image réciproque (par f) de tout élément de \mathcal{B} est un élément de \mathcal{A} :

$$f^{-1}(B) \in \mathcal{A}, \quad \text{for all } B \in \mathcal{B}.$$

Remarque. Si $F = \mathbb{R}$ on prend toujours $\mathcal{B} = \mathcal{B}(\mathbb{R})$, on dit donc qu'une fonction $f: E \rightarrow \mathbb{R}$ est mesurable si pour $f^{-1}(B) \in \mathcal{A}$ pour tout Borélien B .

Exemple 3.3. Soit $A \in \mathcal{A}$, alors la fonction indicatrice de A , définie par $\mathbf{1}_A(x) = 1$ si $x \in A$ et $\mathbf{1}_A(x) = 0$ sinon, est mesurable. En effet on a

$$(\mathbf{1}_A)^{-1}(B) = \{x \in E: \mathbf{1}_A(x) \in B\} = \begin{cases} \emptyset & 0 \notin B, 1 \notin B \\ A^c & 0 \in B, 1 \notin B \\ A & 0 \notin B, 1 \in B \\ E & , 0 \in B, 1 \in B. \end{cases}$$

Dans tous les cas $\{\mathbf{1}_A \in B\}$ est mesurable.

Lemme 3.4. Soit $f: E \rightarrow \mathbb{R}$, les assertions suivantes sont équivalentes.

1. f est mesurable,
2. $f^{-1}(I) \in \mathcal{A}$ pour tout intervalle I ouvert,
3. $f^{-1}(I) \in \mathcal{A}$ pour tout intervalle I de la forme $] - \infty, x[$,
4. $f^{-1}(I) \in \mathcal{A}$ pour tout intervalle I fermé,
5. $f^{-1}(I) \in \mathcal{A}$ pour tout intervalle I de la forme $] - \infty, x]$.

Démonstration. Il est évident que $1 \Rightarrow 2$. Remarquons que l'ensemble $\mathcal{B} = \{B \subset \mathbb{R}: f^{-1}(B) \in \mathcal{A}\}$ est une tribu. Par conséquent, si \mathcal{B} contient tous les intervalles ouverts, alors \mathcal{B} contient $\mathcal{B}(\mathbb{R})$, ce qui montre que $2 \Rightarrow 1$. Les autres équivalences se démontrent de la même manière. \square

Exemple 3.5. Soit $f: \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue. Alors pour tout intervalle I ouvert $f^{-1}(I)$ est un ouvert. Or on a vu au chapitre précédent (en exercice) que les ouverts sont Boréliens. D'après le lemme précédent on en déduit que f est mesurable. On peut montrer de même qu'une fonction continue par morceaux est mesurable.

Proposition 3.6. Soit (E, \mathcal{A}) muni d'une tribu, $f, g: E \rightarrow \mathbb{R}$ des fonctions mesurables. Alors les fonctions suivantes sont mesurables.

- (i) λf
- (ii) $f + g$
- (iii) fg
- (iv) $\max(f, g)$ et $\min(f, g)$

De plus si $f(x) \neq 0$ pour tout x alors $1/f$ est mesurable.

Démonstration. Pour (ii), on montre en utilisant la densité de \mathbb{Q} dans \mathbb{R} que

$$(f + g)^{(-1)}(] - \infty, x]) = \bigcup_{r \in \mathbb{Q}} f^{-1}(] - \infty, r]) \cap g^{-1}(] - \infty, x - r]).$$

Il en résulte que $(f + g)^{(-1)}(] - \infty, x])$ est mesurable (comme réunion dénombrable d'ensembles mesurables) pour tout x . D'après ce qui précède cela suffit pour dire que $f + g$ est mesurable.

Pour (iv) on remarque que

$$\max(f, g)^{(-1)}(] - \infty, x]) = f^{-1}(] - \infty, x]) \cap g^{-1}(] - \infty, x])$$

et que $\min(f, g)^{(-1)}(] - \infty, x]) = f^{-1}(] - \infty, x]) \cup g^{-1}(] - \infty, x])$. Les autres points se démontrent de manière analogue. \square

3.2 Variables aléatoires, fonctions de répartition

On se donne un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

Définition 3.7. Une *variable aléatoire* est une fonction mesurable $X: \Omega \rightarrow \mathbb{R}$.

Notations et terminologie spécifiques aux probabilités : les variables aléatoires seront toujours désignées par des lettres majuscules X, Y, \dots . Étant donné un Borélien B , l'image réciproque de B par X :

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

sera noté simplement $\{X \in B\}$ et avec un léger abus de notation, on écrira $\mathbb{P}(X \in B)$ au lieu de $\mathbb{P}(\{X \in B\})$. Par exemple on a

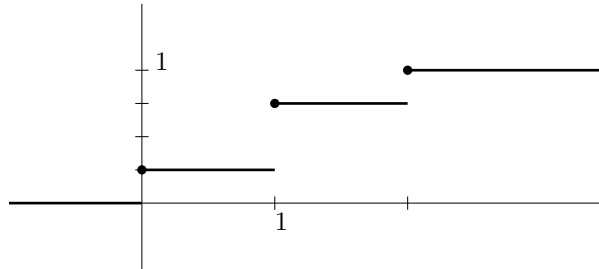
$$\mathbb{P}(X \leq x) = \mathbb{P}(\{\omega \in \Omega: X(\omega) \leq x\}).$$

Définition 3.8. Soit X une variable aléatoire. La fonction

$$F: x \in \mathbb{R} \mapsto \mathbb{P}(X \leq x)$$

est appelée fonction de répartition de X .

Exemple 3.9. La fonction de répartition de la variable X de l'exemple 3.1 est



Exemple 3.10. Soit $c \in \mathbb{R}$ et $X: \omega \in \Omega \mapsto c$ la fonction constante égale à c . Alors X est une variable aléatoire. En effet $\{X \in B\}$, qui vaut \emptyset ou Ω selon que c appartienne à B ou non, est forcément dans la tribu \mathcal{A} . La fonction de répartition de X est $F = \mathbf{1}_{[c, +\infty)}$.

Proposition 3.11. Soit X une variable aléatoire et F sa fonction de répartition. Alors

- (i) F est croissante
- (ii) F est continue à droite

(iii) $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.

Démonstration. Soit $x \leq y$, alors $\{X \leq x\} \subset \{X \leq y\}$ donc $P(X \leq x) \leq P(X \leq y)$, ce qui montre que F est croissante. Soit $x \in \mathbb{R}$, remarquons que $\{X \leq x + 1/n\} \searrow \{X \leq x\}$. Donc par convergence monotone

$$F(x + \frac{1}{n}) = P(X \leq x + \frac{1}{n}) \searrow P(X \leq x) = F(x).$$

Comme F est croissante ceci montre que F est continue à droite en x . De même $\{X \leq -n\} \searrow \{X \leq -\infty\} = \emptyset$ donc

$$F(-n) = P(X \leq -n) \searrow 0.$$

Enfin $\{X \leq n\} \nearrow \{X \leq +\infty\} = \Omega$ donc $F(n) = P(X \leq n) \nearrow 1$. □

Ces propriétés caractérisent la notion de fonction de répartition au sens suivant : si F est une fonction vérifiant les trois propriétés précédentes alors on peut trouver un espace de probabilité (Ω, \mathcal{A}, P) et une variable aléatoire $X: \Omega \rightarrow \mathbb{R}$ tels que $P(X \leq x) = F(x)$ pour tout $x \in \mathbb{R}$.

La fonction de répartition est continue à droite en x . Comme elle est croissante, elle possède forcément une limite à gauche en x , on notera cette limite $F(x-)$:

$$F(x-) = \lim_{y \rightarrow x, y < x} F(y).$$

Lemme 3.12. Soit X une variable aléatoire et F sa fonction de répartition. Soient $x \leq y$ des réels. On a

- (i) $P(X > x) = 1 - F(x)$
- (ii) $P(x < X \leq y) = F(y) - F(x)$
- (iii) $P(X < x) = F(x-)$
- (iv) $P(X = x) = F(x) - F(x-)$

En particulier F est continue en x si et seulement si $P(X = x) = 0$.

Démonstration. Les propriétés (i) et (ii) sont faciles. Pour (iii) on remarque que $\{X \leq x - 1/n\} \nearrow \{X < x\}$ donc par convergence monotone

$$F(x - 1/n) = P(X \leq x - 1/n) \nearrow P(X < x).$$

On obtient (iv) en écrivant $P(X = x) = P(X \leq x) - P(X < x)$. □

Exemple 3.13. On reprend l'exemple 3.1. On a $F(1-) = 1/4$ et $F(1) = 3/4$ ce qui montre que $P(X = 1) = 3/4 - 1/4 = 1/2$.

3.3 Loi d'une variable aléatoire

Proposition 3.14. Soit (Ω, \mathcal{A}, P) un espace de probabilité et soit $X: \Omega \rightarrow \mathbb{R}$ une variable aléatoire. On définit une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ en posant

$$P_X(B) = P(X \in B)$$

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

Démonstration. Soit $(B_n)_{n \in \mathbb{N}}$ une suite de Boréliens deux à deux disjoints. On a

$$P_X \left(\bigcup_{n \in \mathbb{N}} B_n \right) = P \left(\bigcup_{n \in \mathbb{N}} \{X \in B_n\} \right) = \sum_{n \in \mathbb{N}} P(X \in B_n) = \sum_{n \in \mathbb{N}} P_X(B_n).$$

ce qui montre que P_X est σ -additive. □

Définition 3.15. La probabilité P_X est appelée *loi* de X .

Exemple 3.16. Si X une variable constante égale à c on a pour tout Borélien $P_X(B) = 0$ si $c \notin B$ et $P_X(B) = 1$ si $c \in B$. C'est ce qu'on appelle la mesure de Dirac en c .

Exemple 3.17. Si X est le résultat d'un lancer de dé. On a

$$P_X(B) = \frac{\text{card}(B \cap \{1, 2, 3, 4, 5, 6\})}{6}$$

pour tout Borélien B .

Théorème 3.18. La fonction de répartition caractérise la loi : si X et Y ont la même fonction de répartition alors $P_X = P_Y$.

Autrement dit, si X et Y vérifient $P(X \leq x) = P(Y \leq x)$ pour tout $x \in \mathbb{R}$, alors $P(X \in B) = P(Y \in B)$ pour tout Borélien B .

Démonstration. Si $P(X \leq x) = P(Y \leq x)$ pour tout x alors $P(X \in I) = P(Y \in I)$ pour tout intervalle I . Autrement dit les lois de X et Y coïncident sur les intervalles. D'après le Théorème 1.10 elles sont égales. \square

Remarque. Attention le fait que X et Y aient la même loi ne dit rien sur $P(X = Y)$. En fait X et Y peuvent avoir la même loi tout en étant définies sur des espaces de probabilités différents, auquel cas la quantité $P(X = Y)$ n'a aucun sens.

3.4 Variables aléatoires discrètes

Définition 3.19. Une variable aléatoire X est dite discrète s'il existe $D \subset \mathbb{R}$ tel que D soit fini ou dénombrable et tel que

$$P(X \in D) = 1.$$

Remarque. Pour que cette définition ait un sens il faut que $\{X \in D\}$ soit un événement. C'est bien le cas puisque qu'on peut l'écrire comme une réunion finie ou dénombrable d'événements :

$$\{X \in D\} = \bigcup_{x \in D} \{X = x\}.$$

Définition 3.20. Soit X une variable discrète, on appelle fonction de masse la fonction

$$f: x \in \mathbb{R} \mapsto P(X = x).$$

La fonction de masse f vérifie donc les propriétés suivantes :

- (i) $\{x \in \mathbb{R} : f(x) \neq 0\}$ est fini ou dénombrable.
- (ii) En appelant x_1, x_2, \dots les points où f est non nulle, on a $\sum_i f(x_i) = 1$.

Exemple 3.21. La variable de l'exemple 3.1 est discrète puisqu'elle ne prend que les valeurs 0, 1, 2. Sa fonction de masse prend les valeurs 1/4, 1/2, 1/4 en 0, 1, 2 respectivement (et la valeur 0 partout ailleurs).

La fonction de répartition F d'une variable aléatoire discrète est une fonction en escalier avec éventuellement une infinité de « marches ». La connaissance de F détermine f et réciproquement : les points x_1, x_2, \dots où f est non nulle sont les points où F fait un saut et $f(x_i)$ est la taille du saut que fait F en x_i . Tout ceci se résume en une équation :

$$F = \sum_i f(x_i) \mathbf{1}_{[x_i, +\infty)}.$$

3.5 Variables aléatoires continues

Définition 3.22. Soit X une variable aléatoire et F sa fonction de répartition. On dira que la variable X est *continue* s'il existe une fonction f continue par morceaux (disons), positive et vérifiant $\int_{\mathbb{R}} f(t) dt = 1$ telle que

$$F(x) = \int_{-\infty}^x f(t) dt \quad (2)$$

pour tout $x \in \mathbb{R}$. La fonction f est appelée *densité* de X .

Remarque. L'hypothèse « f continue par morceaux » sera vérifiée dans tous les exemples qu'on rencontrera mais elle n'est pas vraiment nécessaire. L'important est de pouvoir donner un sens à l'intégrale de f . On verra au chapitre suivant la bonne notion d'intégrale à considérer.

Lemme 3.23. Soit X une variable continue de densité f . Alors

- (i) $P(X = x) = 0$ pour tout réel x .
- (ii) $P(a \leq X \leq b) = \int_a^b f(t) dt$ pour tous réels $a \leq b$.

Démonstration. L'équation (2) montre que F est continue ce qui implique (i). On a donc

$$P(a \leq X \leq b) = P(a < X \leq b) = F(b) - F(a) = \int_a^b f(t) dt. \quad \square$$

Si X est une variable continue on a $P(X = x) = 0$ pour tout $x \in \mathbb{R}$. Par σ -additivité, si $D = \{x_1, x_2, \dots\}$ est un ensemble dénombrable alors $P(X \in D) = \sum_i P(X = x_i) = 0$. Par conséquent une variable aléatoire ne peut pas être à la fois discrète et continue.

On a vu que la fonction de répartition d'une variable continue était continue. La réciproque n'est pas vraie. On a néanmoins le résultat suivant.

Proposition 3.24. Soit F une fonction de répartition. Si F est continue sur \mathbb{R} et de classe \mathcal{C}^1 sur \mathbb{R} privé éventuellement d'un nombre fini de points, alors F est la fonction de répartition d'une variable continue. Une densité associée est donnée par la fonction $f(x) = F'(x)$ si F est dérivable en x et $f(x) = 0$ sinon.

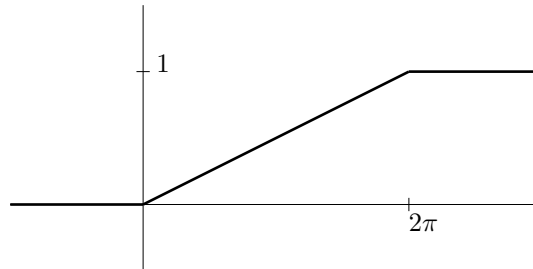
Démonstration. Sous ces hypothèses on a

$$F(x) = \int_{-\infty}^x f(t) dt,$$

pour tout $x \in \mathbb{R}$. □

Remarque. La valeur 0 attribuée à f là où F' n'est pas définie n'a pas d'importance. De manière générale, changer la valeur de la densité en un nombre fini de points ne change rien à la répartition de X .

Exemple 3.25. Soit θ l'angle que fait l'aiguille de Buffon avec le nord et F sa fonction de répartition. Le graphe de F est



On remarque que F est continue sur \mathbb{R} et \mathcal{C}^1 sur $\mathbb{R} \setminus \{0, 2\pi\}$. On a $F'(x) = 0$ si $x < 0$ ou si $x > 2\pi$ et $F'(x) = 1/(2\pi)$ si $x \in]0, 2\pi[$. La variable θ est donc continue et admet $\mathbf{1}_{]0, 2\pi[}/2\pi$ comme densité.

3.6 Exercices

Exercice 3.1. Montrer que si F et G sont des fonctions de répartition et $\lambda \in [0, 1]$, alors $\lambda F + (1 - \lambda)G$ est une fonction de répartition. La fonction FG est-elle une fonction de répartition ?

Exercice 3.2 (Loi de Poisson). Soit $\lambda > 0$. Soit X une variable aléatoire à valeurs entières ayant comme fonction de masse $P(X = n) = C\lambda^n/n!$ pour tout entier n .

1. Déterminer C .
2. Calculer $P(X > 1)$.
3. Déterminer la probabilité que X soit paire.

Exercice 3.3. Un bus passe toutes les dix minutes, à partir de midi. Un homme arrive X minutes après midi où X est une variable aléatoire de fonction de répartition :

$$P(X \leq x) = \begin{cases} 0 & x < 0 \\ x/60 & 0 \leq x < 60 \\ 1 & 60 \leq x \end{cases}$$

Quelle est la probabilité qu'il attende le bus moins de deux minutes ?

Exercice 3.4. Les personnes ayant réservé un billet d'avion manquent de se présenter à l'embarquement avec probabilité $1/10$ indépendamment les unes des autres. La compagnie A possède un avion de 9 places mais elle vend 10 tickets. La compagnie B vend 20 tickets pour son avion de 18 places.

1. Déterminer, dans chacun des cas, la fonction de masse du nombre X de passagers se présentant à l'embarquement.
2. Avec quelle compagnie risque-t-on le plus d'être en sur-réservation ?

Exercice 3.5. On considère une suite infinie de lancers de la même pièce truquée tombant sur pile avec probabilité p .

1. Déterminer la fonction de masse du temps d'attente du premier pile.
2. Déterminer la fonction de masse du temps d'attente du r -ième pile.
3. Déterminer la fonction de masse du nombre de piles obtenus au bout de n lancers.

Exercice 3.6 (Loi exponentielle). Soit f la fonction

$$f: x \mapsto e^{-x} \mathbf{1}_{\{x > 0\}}.$$

1. Montrer que f est une densité de probabilité.
2. Soit X de densité f , déterminer la fonction de répartition de X .
3. Montrer que pour tous $s, t > 0$ on a

$$P(X \geq s + t \mid X \geq s) = P(X \geq t).$$

Cette propriété est appelée absence de mémoire.

Exercice 3.7. Soit X uniforme sur $[0, 1]$ (i.e. admettant $\mathbf{1}_{[0,1]}$ comme densité). Montrer que $Y = -\ln(1 - X)$ est continue et déterminer sa densité.

Exercice 3.8 (loi normale, loi du Chi-2). On admet que

$$\int_{\mathbb{R}} e^{-x^2/2} dx = \sqrt{2\pi}.$$

Soit X une variable continue de densité

$$x \mapsto \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

1. Montrer que $P(X \leq x) = P(X \geq -x)$ pour tout $x \in \mathbb{R}$.
2. Exprimer la fonction de répartition de X^2 en fonction de celle de X .
3. En déduite que X^2 est une variable continue, et déterminer sa densité.
4. En déduire la valeur de

$$\int_0^{+\infty} x^{-1/2} e^{-x} dx.$$

4 Intégration

On se donne dans toute cette partie un espace mesuré (E, \mathcal{A}, μ) . Dans cette partie nous allons définir l'intégrale de Lebesgue des fonctions mesurables $f: E \rightarrow \mathbb{R}$ par rapport à μ . L'intégrale de f par rapport à μ sera notée

$$\int_E f d\mu \quad \text{ou} \quad \int_E f(x) \mu(dx).$$

4.1 Intégrale d'une fonction étagée positive

Définition 4.1. Une fonction $h: E \rightarrow \mathbb{R}$ est dite étagée positive si elle est mesurable et si elle ne prend qu'un nombre fini de valeurs, toutes positives.

Si h est une fonction étagée positive, il existe une suite finie $(a_1, A_1), \dots, (a_n, A_n)$ où les a_i sont des réels positifs et les A_i des ensembles mesurables telle que

$$h = \sum_{i=1}^n a_i \mathbf{1}_{A_i}. \quad (3)$$

En effet, on peut par exemple écrire

$$h = \sum_{y \in h(E)} y \mathbf{1}_{\{h=y\}}.$$

Attention la représentation (3) n'est pas unique. On a par exemple

$$\mathbf{1}_{[0,1]} + 2 \times \mathbf{1}_{]1,2]} = \mathbf{1}_{[0,2]} + \mathbf{1}_{]1,2]}.$$

Définition 4.2. Soit h une fonction étagée positive. On définit l'intégrale de h par

$$\int_E h d\mu = \sum_{y \in h(E)} y \mu(h=y) \in [0, +\infty]$$

Remarque. Il y a un léger abus de notation dans la définition précédente, la notation

$$\mu(h=y)$$

est un raccourci pour

$$\mu(\{x \in E, h(x) = y\}).$$

Cette notation sera utilisée dans tout ce chapitre.

Remarque. Nous ne supposons pas que la mesure μ est finie, il est donc possible que $\mu(A_1) = +\infty$ par exemple. La somme précédente est à comprendre avec les conventions suivantes :

$$\begin{aligned} a + \infty &= +\infty & \forall a \geq 0 \\ \infty + \infty &= +\infty \\ a \times (+\infty) &= +\infty & \forall a > 0. \\ 0 \times (+\infty) &= 0. \end{aligned}$$

Exemple 4.3. Si h est l'indicatrice d'un événement mesurable A , alors h ne prend que deux valeurs, la valeur 1 sur A et la valeur 0 sur A^c . On a donc

$$\int_E \mathbf{1}_A d\mu = 1 \times \mu(A) + 0 \times \mu(A^c) = \mu(A).$$

Proposition 4.4. Soient h_1, h_2 des fonctions étagées positives et $\lambda > 0$. On a les propriétés suivantes

- Linéarité : $\int_E (\lambda h_1) d\mu = \lambda \left(\int_E h_1 d\mu \right)$ et

$$\int_E (h_1 + h_2) d\mu = \int_E h_1 d\mu + \int_E h_2 d\mu.$$

- Monotonie : si $h_1 \leq h_2$ alors $\int_E h_1 d\mu \leq \int_E h_2 d\mu$.

Démonstration. La propriété $\int(\lambda h) = \lambda \int h d\mu$ est évidente. Montrons que $\int(h_1+h_2) = \int h_1 + \int h_2$. Soit $h = h_1 + h_2$, on a

$$h(E) = \{y_1 + y_2, y_1 \in h_1(E), y_2 \in h_2(E)\}.$$

De plus pour tout $y \in h(E)$

$$\{h = y\} = \bigcup_{\substack{y_1 \in h_1(E) \\ y_2 \in h_2(E) \\ y_1 + y_2 = y}} \{h_1 = y_1, h_2 = y_2\}$$

et cette union est disjointe. Par conséquent

$$\mu(h = y) = \sum_{\substack{y_1 \in h_1(E) \\ y_2 \in h_2(E) \\ y_1 + y_2 = y}} \mu(h_1 = y_1, h_2 = y_2).$$

On peut donc écrire

$$\begin{aligned} \int_E h d\mu &= \sum_{y \in h(E)} y \mu(h = y) \\ &= \sum_{y \in h(E)} \sum_{\substack{y_1 \in h_1(E) \\ y_2 \in h_2(E) \\ y_1 + y_2 = y}} (y_1 + y_2) \mu(h_1 = y_1, h_2 = y_2) \\ &= \sum_{y_1 \in h_1(E)} \sum_{y_2 \in h_2(E)} (y_1 + y_2) \mu(h_1 = y_1, h_2 = y_2) \\ &= \sum_{y_1 \in h_1(E)} \sum_{y_2 \in h_2(E)} y_1 \mu(h_1 = y_1, h_2 = y_2) + \sum_{y_1 \in h_1(E)} \sum_{y_2 \in h_2(E)} y_2 \mu(h_1 = y_1, h_2 = y_2). \end{aligned}$$

Or, d'après la formule des probabilités totales

$$\begin{aligned} \sum_{y_1 \in h_1(E)} \sum_{y_2 \in h_2(E)} y_1 \mu(h_1 = y_1, h_2 = y_2) &= \sum_{y_1 \in h_1(E)} y_1 \sum_{y_2 \in h_2(E)} \mu(h_1 = y_1, h_2 = y_2) \\ &= \sum_{y_1 \in h_1(E)} y_1 \mu(h_1 = y_1) = \int_E h_1 d\mu. \end{aligned}$$

Et de manière analogue

$$\sum_{y_1 \in h_1(E)} \sum_{y_2 \in h_2(E)} y_2 \mu(h_1 = y_1, h_2 = y_2) = \int_E h_2 d\mu,$$

ce qui termine la preuve de la linéarité.

La monotonie se déduit de la linéarité ainsi : si $h_1 \leq h_2$ alors $h_3 = h_2 - h_1$ est une fonction étagée positive. Donc

$$\int_E h_2 d\mu = \int_E (h_1 + h_3) d\mu = \int_E h_1 d\mu + \int_E h_3 d\mu.$$

Et comme $\int_E h_3 d\mu \geq 0$ le résultat suit. □

Exemple 4.5. On a

$$\int_E \sum_{j=1}^m b_j \mathbf{1}_{B_j} d\mu = \sum_{j=1}^m b_j \mu(B_j),$$

pour toutes suites finies b_j de réels positifs et B_j d'ensembles mesurables.

4.2 Intégrale d'une fonction positive

Définition 4.6. Soit f une fonction mesurable positive. On pose

$$\int_E f d\mu = \sup \left\{ \int_E h d\mu, h \text{ étagée positive } h \leq f \right\} \in [0, +\infty].$$

Proposition 4.7. Soient f_1, f_2 des fonctions mesurables positives et $\lambda > 0$. On a les propriétés suivantes

– *Linéarité* : $\int_E (\lambda f_1) d\mu = \lambda \left(\int_E f_1 d\mu \right)$ et

$$\int_E (f_1 + f_2) d\mu = \int_E f_1 d\mu + \int_E f_2 d\mu.$$

– *Monotonie* : si $f_1 \leq f_2$ alors $\int_E f_1 d\mu \leq \int_E f_2 d\mu$.

Démonstration de la monotonie. C'est évident : si $f_1 \leq f_2$ alors l'ensemble des fonctions étagées positives plus petite que f_1 est contenu dans l'ensemble des fonctions étagées positives plus petite que f_2 . \square

La linéarité sera démontrée à la fin de cette sous partie.

Théorème 4.8 (Convergence monotone). Soit $(f_n)_{n \geq 0}, f$ des fonctions mesurables positives telles que $f_n(x) \nearrow f(x)$ pour tout $x \in E$. Alors

$$\int_E f_n d\mu \nearrow \int_E f d\mu.$$

Remarque. Soit $(A_n)_{n \geq 0}, A$ des ensembles mesurables et supposons que $A_n \nearrow A$. On peut alors appliquer le théorème aux fonctions $\mathbf{1}_{A_n}, \mathbf{1}_A$ et on obtient $\mu(A_n) \nearrow \mu(A)$. Autrement dit on retrouve la Proposition 1.13.

Démonstration. D'après la propriété de monotonie de l'intégrale on sait déjà que la suite $(\int_E f_n d\mu)_{n \geq 0}$ est croissante et majorée par $\int_E f d\mu$. Donc il existe L tel que

$$\int_E f_n d\mu \nearrow L \leq \int_E f d\mu.$$

Soit h une fonction étagée positive telle que $h \leq f$ et soit $\epsilon > 0$. On pose

$$A_n = \{f_n \geq (1 - \epsilon)h\}$$

pour tout $n \geq 0$. Alors par définition de A_n

$$f_n \geq f_n \mathbf{1}_{A_n} \geq (1 - \epsilon)h \mathbf{1}_{A_n},$$

Posons $h_n = h \mathbf{1}_{A_n}$, par monotonie de l'intégrale on a

$$\int_E f_n d\mu \geq (1 - \epsilon) \int_E h_n d\mu. \quad (4)$$

De plus

$$h_n = \sum_{y \in h(E)} y \mathbf{1}_{\{h=y\} \cap A_n}$$

et donc

$$\int_E h_n d\mu = \sum_{y \in h(E)} y \mu(\{h=y\} \cap A_n)$$

L'hypothèse " $f_n(x) \nearrow f(x)$ pour tout x " implique facilement $A_n \nearrow E$, et donc $\{h=y\} \cap A_n \nearrow \{h=y\}$ pour tout y . D'après la convergence monotone pour les ensembles on en déduit

$$\mu(\{h=y\} \cap A_n) \nearrow \mu(\{h=y\})$$

pour tout y . On multiplie par y et on somme, on obtient

$$\int_E h_n d\mu \nearrow \int_E h d\mu.$$

En passant à la limite dans l'inégalité (4) on obtient donc

$$L \geq (1 - \epsilon) \int_E h d\mu.$$

Comme ϵ est arbitraire on en déduit $L \geq \int_E h d\mu$. Comme ceci est valable pour toute fonction h étagée plus petite que f , il vient bien

$$L \geq \int_E f d\mu,$$

ce qu'il fallait démontrer. □

Lemme 4.9. *Soit f une fonction mesurable positive, il existe une suite $(h_n)_{n \geq 0}$ de fonctions étagées positives vérifiant $h_n(x) \nearrow f(x)$ pour tout x .*

Démonstration. On pose

$$h_n = \sum_{k=0}^{4^n - 1} k \times 2^{-n} \mathbf{1}_{\{k2^{-n} \leq f < (k+1)2^{-n}\}}.$$

Clairement h_n est une fonction étagée. Si $x < 2^n$ alors $h_n(x) \leq f(x) < h_n(x) + 2^{-n}$ ce qui montre que $h_n(x) \rightarrow f(x)$. On vérifie de même que $h_n(x) \leq h_{n+1}(x)$ pour tout x et pour tout n . □

Démonstration de la linéarité de l'intégrale. L'égalité $\int_E (\lambda f_1) d\mu = \lambda \int_E f_1 d\mu$ est immédiate. Pour l'additivité on utilise le lemme précédent : il existe $(h_{n,1})_{n \geq 0}$ et $(h_{n,2})_{n \geq 0}$ des suites de fonctions étagées positives telles que $h_{n,1} \nearrow f_1$ et $h_{n,2} \nearrow f_2$. Alors

$$h_{n,1} + h_{n,2} \nearrow f_1 + f_2.$$

D'après la linéarité de l'intégrale pour les fonctions étagées on a

$$\int_E (h_{n,1} + h_{n,2}) d\mu = \int_E h_{n,1} d\mu + \int_E h_{n,2} d\mu$$

pour tout n . En passant à la limite dans cette égalité (en utilisant le théorème de convergence monotone trois fois) on obtient le résultat. □

L'inégalité suivante est **très importante**.

Proposition 4.10 (Inégalité de Markov). *Soit f une fonction positive et $a > 0$*

$$\mu(\{x \in E : f(x) \geq a\}) \leq \frac{1}{a} \int_E f d\mu.$$

Démonstration. Comme $f \geq 0$ on a $f \geq a \mathbf{1}_{\{f \geq a\}}$. Donc par monotonie

$$\int_E f d\mu \geq \int_E a \mathbf{1}_{\{f \geq a\}} d\mu = a \mu(\{f \geq a\}).$$
 □

4.3 Intégrale d'une fonction réelle

Rappelons que si x est un nombre réel on appelle partie positive de x la quantité $x_+ = \max(x, 0)$ et partie négative $x_- = -\min(x, 0)$. Les nombres x_+ et x_- sont positifs et on a

$$\begin{aligned} x &= x_+ - x_- \\ |x| &= x_+ + x_-. \end{aligned}$$

Définition 4.11. Soit $f : E \rightarrow \mathbb{R}$ une fonction mesurable. On dit que f est intégrable si

$$\int_E |f| d\mu < +\infty.$$

Comme $f_+ \leq |f|$ et $f_- \leq |f|$, les quantités $\int_E f_+ d\mu$ et $\int_E f_- d\mu$ sont également finies. On pose alors

$$\int_E f d\mu = \int_E f_+ d\mu - \int_E f_- d\mu.$$

Proposition 4.12. Soient f, g des fonctions intégrables et $\lambda > 0$. On a les propriétés suivantes

- *Linéarité* : $\int_E (\lambda f) d\mu = \lambda \left(\int_E f d\mu \right)$ et

$$\int_E (f + g) d\mu = \int_E f d\mu + \int_E g d\mu.$$

- *Monotonie* : si $f \leq g$ alors $\int_E f d\mu \leq \int_E g d\mu$.

Démonstration. L'égalité $\int_E (\lambda f) d\mu = \lambda \left(\int_E f d\mu \right)$ est immédiate. Pour l'additivité on écrit

$$f + g = (f + g)_+ - (f + g)_- = f_+ - f_- + g_+ - g_-,$$

et donc

$$(f + g)_+ + f_- + g_- = (f + g)_- + f_+ + g_+.$$

En utilisant la linéarité pour les fonctions positives on obtient

$$\int_E (f + g)_+ d\mu + \int_E f_- d\mu + \int_E g_- d\mu = \int_E (f + g)_- d\mu + \int_E f_+ d\mu + \int_E g_+ d\mu.$$

En réarrangeant les termes on obtient le résultat.

La monotonie est facile une fois qu'on a la linéarité : on écrit

$$\int_E g d\mu = \int_E f d\mu + \int_E (g - f) d\mu \geq \int_E f d\mu. \quad \square$$

4.4 Exemples

Mesures sur \mathbb{N} On considère une mesure μ sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$.

Proposition 4.13. Pour toute fonction positive $f : \mathbb{N} \rightarrow \mathbb{R}$, on a

$$\int_{\mathbb{N}} f d\mu = \lim_{n \rightarrow +\infty} \sum_{k=0}^n f(k) \mu(\{k\}).$$

L'intégrale par rapport à une mesure sur les entiers est donc la somme d'une série.

Démonstration. Posons

$$f_n(k) = \begin{cases} f(k) & \text{si } k \leq n \\ 0 & \text{sinon.} \end{cases}$$

Alors la fonction f_n est une fonction étagée

$$f_n = \sum_{k=0}^n f(k) \mathbf{1}_{\{k\}},$$

et donc

$$\int_{\mathbb{N}} f_n d\mu = \sum_{k=0}^n f(k) \mu(\{k\}).$$

D'autre part, on a clairement $f_n(k) \nearrow f(k)$ pour tout k et donc par convergence monotone

$$\int_{\mathbb{N}} f_n d\mu \nearrow \int_{\mathbb{N}} f d\mu,$$

ce qui est le résultat annoncé. □

Le résultat est également vrai si la fonction f change de signe, pourvu qu'elle soit intégrable.

Proposition 4.14. *Si $f: \mathbb{N} \rightarrow \mathbb{R}$ est intégrable (pour μ) alors la série $\sum f(k)\mu(\{k\})$ est absolument convergente et*

$$\int_{\mathbb{N}} f d\mu = \lim_{n \rightarrow +\infty} \sum_{k=0}^n f(k)\mu(\{k\}).$$

Démonstration. Il suffit de décomposer $f = f_+ - f_-$ et d'utiliser le résultat précédent. \square

La mesure de Lebesgue Supposons maintenant que $E = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ et que $\mu = \mathcal{L}$, la mesure de Lebesgue sur \mathbb{R} . Dans ce cas il est d'usage d'écrire

$$\int_{\mathbb{R}} f(x) dx$$

au lieu de

$$\int_{\mathbb{R}} f(x) \mathcal{L}(dx).$$

Soit $I = [a, b]$ un intervalle de \mathbb{R} et $f: I \rightarrow \mathbb{R}$ une fonction telle que $f \times \mathbf{1}_I$ soit intégrable. On pose

$$\int_a^b f(x) dx = \int_{[a,b]} f(x) dx = \int_{\mathbb{R}} f(x)\mathbf{1}_I(x) dx.$$

La notation $\int_a^b f(x) dx$ peut paraître ambiguë, faut-il intégrer sur $[a, b]$ ou $]a, b[$? En fait ça ne change rien au résultat. On a par exemple

$$\int_{\mathbb{R}} f(x)\mathbf{1}_{[a,b]} dx - \int_{\mathbb{R}} f(x)\mathbf{1}_{[a,b[}(x) dx = \int_{\mathbb{R}} f(b)\mathbf{1}_{\{b\}}(x) dx = f(b)\mathcal{L}(\{b\}) = 0.$$

Proposition 4.15. *Soit $f: [a, b] \rightarrow \mathbb{R}$ une fonction continue et posons pour $x \in]a, b[$*

$$F(x) = \int_a^x f(t) dt.$$

Alors F est de classe \mathcal{C}^1 sur $]a, b[$ et $F'(x) = f(x)$ pour tout $x \in]a, b[$.

Autrement dit F est la primitive de f nulle en a .

Démonstration. Remarquons que f est bornée sur $[a, b]$: il existe $M > 0$ tel que $|f| \leq M$ sur $[a, b]$, et donc

$$\int_a^b |f(x)| dx \leq M(b-a) < +\infty.$$

ce qui montre que f est intégrable sur $[a, b]$. Comme f est continue on a $f(x) - \epsilon \leq f(y) \leq f(x) + \epsilon$ pour y suffisamment proche de x . Donc par croissance de l'intégrale

$$\int_x^{x+h} (f(x) - \epsilon) dy \leq \int_x^{x+h} f(y) dy \leq \int_x^{x+h} (f(x) + \epsilon) dy$$

pour $h > 0$ suffisamment petit, ce qui donne

$$(f(x) - \epsilon)h \leq F(x+h) - F(x) \leq (f(x) + \epsilon)h.$$

On montre de même cette inégalité pour $h < 0$ suffisamment petit et on obtient

$$\lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h} = f(x),$$

ce qui est le résultat. \square

Donc si f est une fonction de classe \mathcal{C}^1 sur $[a, b]$ on a

$$f(b) - f(a) = \int_a^b f'(t) dt.$$

En appliquant ceci à un produit de fonctions on obtient :

Proposition 4.16 (Intégration par parties). Soient f et g des fonction de classe \mathcal{C}^1 sur $[a, b]$

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

Pour calculer une intégrale sur \mathbb{R} on utilise le résultat suivant.

Proposition 4.17. Soit $f: \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable positive ou intégrable. On a

$$\int_{\mathbb{R}} f(x) dx = \lim_n \int_{-n}^n f(x) dx$$

Démonstration. Si $f \geq 0$, il suffit d'appliquer le théorème de convergence monotone, si f change de signe, on écrit $f = f_+ - f_-$ et on se ramène au cas positif. \square

4.5 Exercices

Exercice 4.1. Soit (E, \mathcal{A}, μ) un espace mesuré, à quelle condition les fonctions constantes sont-elles intégrables ?

Exercice 4.2 (Masse de Dirac). Soit $a \in \mathbb{R}$, on rappelle que la masse de Dirac en a est la mesure définie par $\delta_a(A) = 1$ si $a \in A$ et $\delta_a(A) = 0$ sinon. Soit $f: \mathbb{R} \rightarrow \mathbb{R}$, montrer (en revenant à la définition de l'intégrale) que

$$\int_{\mathbb{R}} f(x) \delta_a(dx) = f(a).$$

Exercice 4.3. Soit (E, \mathcal{A}, μ) un espace mesuré, et soit $f: E \rightarrow \mathbb{R}$ une fonction mesurable et positive.

1. On suppose que f est nulle presque partout. Montrer (en revenant à la définition de l'intégrale) que

$$\int_E f d\mu = 0.$$

2. Réciproquement, on suppose que f est d'intégrale nulle. Montrer que $\mu(\{f \geq \epsilon\}) = 0$ pour tout $\epsilon > 0$. En déduire que f est nulle presque partout.

Exercice 4.4. Soit (E, \mathcal{A}, μ) un espace mesuré, soient $f, g: E \rightarrow \mathbb{R}$ des fonctions intégrables. Montrer que $\max(f, g)$ est intégrable et que

$$\max\left(\int_E f d\mu, \int_E g d\mu\right) \leq \int_E \max(f, g) d\mu.$$

Exercice 4.5. À quelle condition (sur r) la fonction $x \mapsto x^r$ est-elle intégrable sur $[1, +\infty)$? sur $[0, +\infty)$?

Exercice 4.6. Soit (E, \mathcal{A}, μ) un espace mesuré.

1. Soit $(f_k)_{k \geq 0}$ une suite de fonctions mesurables **positives**. Montrer que

$$\int_E \sum_{k=0}^{+\infty} f_k d\mu = \sum_{k=0}^{+\infty} \int_E f_k d\mu.$$

Indication : On pourra poser $g_n = \sum_{k=0}^n f_k$ et appliquer un théorème du cours.

2. Montrer que pour tout entier n on a $n = \sum_{k=0}^{+\infty} \mathbf{1}_{\{n > k\}}$.

3. En déduire que si f est une fonction mesurable à valeurs entières :

$$\int_E f d\mu = \sum_{k=0}^{+\infty} \mu(f > k).$$

Exercice 4.7 (Inégalité de Cauchy–Schwarz). Soit (E, \mathcal{A}, μ) un espace mesuré, soient f et g des fonctions de carré intégrable (c'est-à-dire que f^2 et g^2 sont intégrables). On supposera dans un premier temps $\int_E g^2 d\mu > 0$.

1. Montrer que fg est intégrable.
2. On pose $P(\lambda) = \int_E (f + \lambda g)^2 d\mu$. Montrer que P est un polynôme du second degré en λ .
3. Calculer le discriminant de P .
4. Montrer que

$$\left| \int_E fg d\mu \right| \leq \sqrt{\left(\int_E f^2 d\mu \right) \left(\int_E g^2 d\mu \right)} :$$

5. Montrer que l'inégalité précédente reste vraie si $\int_E g^2 d\mu = 0$.
6. Écrire ce que devient cette inégalité dans le cas particulier où μ est la mesure de comptage sur \mathbb{N} .

Exercice 4.8 (Inégalité de Hölder). Soient f, g des fonctions positives et intégrables et soit $\alpha \in [0, 1]$.

1. Montrer que pour tous $a, b > 0$ on a $a^\alpha b^{1-\alpha} \leq \alpha a + (1 - \alpha)b$.
2. En déduire que

$$\int_E f^\alpha g^{1-\alpha} d\mu \leq \alpha \int_E f d\mu + (1 - \alpha) \int_E g d\mu.$$

3. En appliquant l'inégalité précédente à λf et en optimisant en λ montrer que

$$\int_E f^\alpha g^{1-\alpha} d\mu \leq \left(\int_E f d\mu \right)^\alpha \left(\int_E g d\mu \right)^{1-\alpha}.$$

4. Montrer que l'inégalité de Cauchy–Schwarz est un cas particulier de l'inégalité de Hölder.

5 Espérance

Dans toute cette partie on se donne un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

5.1 Définition, propriétés

Définition 5.1. Soit $X: \Omega \rightarrow \mathbb{R}$ une variable aléatoire (i.e. une fonction mesurable). Si X est positive ou intégrable pour la mesure \mathbb{P} , l'intégrale

$$\int_{\Omega} X(\omega) \mathbb{P}(d\omega)$$

est bien définie. On appelle cette intégrale *espérance* de X et on la note $\mathbb{E}(X)$.

Si $\mathbb{E}|X| < +\infty$ on dit que X est intégrable.

Exemple 5.2. Si $X = c$ (variable constante) alors $X = c \times \mathbf{1}_{\Omega}$, donc

$$\mathbb{E}X = c \int_{\Omega} \mathbf{1}_{\Omega} d\mathbb{P} = c\mathbb{P}(\Omega) = c.$$

Dans le même ordre d'idée, pour un tout événement A

$$\mathbb{E}\mathbf{1}_A = \mathbb{P}(A).$$

Exemple 5.3. On lance n pièces, soit X le nombre de pile obtenus. Alors est une fonction étagée positive prenant les valeurs $0, 1, \dots, n$. Autrement dit,

$$X = \sum_{k=0}^n k \times \mathbf{1}_{\{X=k\}}$$

Donc

$$\mathbb{E}(X) = \sum_{k=0}^n k\mathbb{P}(X = k)$$

On a

$$\mathbb{P}(X = k) = \binom{n}{k} 2^{-n}, \quad k = 0, \dots, n.$$

Donc

$$\mathbb{E}(X) = \sum_{k=1}^n k \binom{n}{k} 2^{-n} = \sum_{k=1}^n n \binom{n-1}{k-1} 2^{-n} = n 2^{n-1} 2^{-n} = \frac{n}{2}.$$

L'espérance est une intégrale, elle vérifie donc les mêmes propriétés que l'intégrale :

- Linéarité : $\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y)$.
- Monotonie : si $X \leq Y$ alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

Exemple 5.4. On peut calculer l'espérance du nombre X de piles au bout de n lancers ainsi : on pose $Y_i = 1$ si le i -ème résultat est pile et $Y_i = 0$ si c'est face. Alors

$$\mathbb{E}(Y_i) = \mathbb{E}\mathbf{1}_{\{Y_i=1\}} = \mathbb{P}(Y_i = 1) = 1/2.$$

De plus $X = Y_1 + \dots + Y_n$. Par conséquent

$$\mathbb{E}(X) = \mathbb{E}(Y_1) + \dots + \mathbb{E}(Y_n) = \frac{n}{2}.$$

Rappelons également l'inégalité de Markov : si $X \geq 0$ alors pour tout $a > 0$

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

Remarque. L'inégalité de Markov est valable pour tout $a > 0$ mais elle n'est intéressante que pour $a > \mathbb{E}(X)$.

5.2 Moments, variance

Définition 5.5. Soit k un entier naturel, la quantité $E(X^k)$ est appelée moment d'ordre k . Le moment d'ordre k de X est défini si $X \geq 0$ (dans ce cas il peut prendre la valeur $+\infty$) ou si $|X|^k$ est intégrable (dans ce cas il est forcément fini). La variance de X est la quantité

$$\text{var}(X) = E((X - EX)^2) = E(X^2) - (EX)^2$$

Remarque. Soit X ayant un moment d'ordre 2. Alors la variable $|X|^2 + 1$ est intégrable. Comme $|X| \leq |X|^2 + 1$, la variable X est aussi intégrable par comparaison. De même on a

$$|X|^p \leq |X|^q + 1,$$

dès que $p \leq q$. Donc si le moment d'ordre q est défini, alors le moment d'ordre p est défini pour tout $p \leq q$.

Exemple 5.6. On considère toujours le nombre X de piles au bout de n lancers. Calculons $\text{var}(X)$. Commençons par remarquer que pour toute fonction f , on a

$$f(X) = \sum_{k=0}^n f(k) \mathbf{1}_{\{X=k\}},$$

et donc

$$Ef(X) = \sum_{k=0}^n f(k) P(X = k).$$

On applique ceci à $f: x \mapsto x(x-1)$, on obtient

$$E[X(X-1)] = \sum_{k=2}^n k(k-1) \binom{n}{k} 2^{-n} = n(n-1) 2^{-n} \sum_{k=2}^n \binom{n-2}{k-2} = \frac{n(n-1)}{4}.$$

On en déduit

$$\text{var}(X) = E[X(X-1)] + E(X) - (EX)^2 = \frac{n(n-1)}{4} + \frac{n}{2} - \frac{n^2}{4} = \frac{n}{4}.$$

Proposition 5.7. On a les propriétés suivantes

- $\text{var}(\lambda X) = \lambda^2 \text{var}(X)$
- $\text{var}(X + c) = \text{var}(X)$

La démonstration est laissée en exercice. La variance mesure la déviation de X à sa moyenne. Plus la variance est grande plus X a de chances d'être loin de sa moyenne. Plus précisément on a le résultat suivant :

Proposition 5.8 (Inégalité de Bienaymé-Tchebichev). Soit $a > 0$, alors

$$P(|X - EX| \geq a) \leq \frac{\text{var}(X)}{a^2}$$

Démonstration. La variable $Y = (X - EX)^2$ est positive. Par Markov

$$P(|X - EX| \geq a) = P(Y \geq a^2) \leq \frac{E(Y)}{a^2} = \frac{\text{var}(X)}{a^2}. \quad \square$$

Exemple 5.9. Soit X de carré intégrable vérifiant $\text{var}(X) = 0$. Par Bienaymé-Tchebychev

$$P\left(|X - E(X)| \geq \frac{1}{n}\right) \leq n^2 \text{var}(X) = 0$$

pour tout $n \geq 1$. De plus clairement

$$\left\{|X - E(X)| \geq \frac{1}{n}\right\} \nearrow \{|X - E(X)| > 0\}.$$

Par convergence monotone on obtient donc

$$P(|X - E(X)| > 0) = 0.$$

Donc $X = E(X)$ presque sûrement. Autrement dit X est une variable aléatoire constante.

Exemple 5.10. On appelle désormais X_n le nombre de piles obtenus au bout de n lancers. On s'intéresse à la fréquence des piles X_n/n . On a $\mathbf{E}(X_n/n) = \mathbf{E}(X_n)/n = 1/2$ et $\text{var}(X_n/n) = \text{var}(X_n)/n^2 = 1/(4n)$. Étant donné $\epsilon > 0$, on a par Bienaymé–Tchebychev

$$\mathbf{P}\left(\left|\frac{X_n}{n} - \frac{1}{2}\right| \geq \epsilon\right) \leq \frac{1}{4\epsilon^2 n}$$

En particulier, la probabilité que X_n/n dévie de sa moyenne d'au moins ϵ tend vers 0 quand n tend vers l'infini. On dit que X_n/n converge vers $1/2$ en probabilité.

5.3 Fonction génératrice des moments

Définition 5.11. Soit X une variable discrète ou continue

$$M_X : t \in \mathbb{R} \mapsto \mathbf{E}(e^{tX}) \in [0, +\infty].$$

est appelée fonction génératrice des moments de X

On a toujours $M_X(0) = 1$ mais il est possible d'avoir $M_X(t) = +\infty$ pour tout $t \neq 0$. On a le résultat suivant.

Proposition 5.12. *L'ensemble des réels t pour lesquels $M_X(t) < +\infty$ est un intervalle contenant 0.*

Démonstration. Soient $a < b < c$ tels que M_X soit finie en a et c . Remarquons qu'on a toujours $e^{bX} \leq e^{aX} + e^{cX}$. Par monotonie et linéarité de l'espérance $\mathbf{E}e^{cX} \leq \mathbf{E}e^{aX} + \mathbf{E}e^{cX} < +\infty$. \square

La dénomination “fonction génératrice des moments” vient du résultat suivant.

Théorème 5.13. *On suppose que ϕ_X est finie au voisinage de 0, c'est-à-dire : il existe $\epsilon > 0$ tel que $\phi_X(t) < +\infty$ pour $t \in [-\epsilon, +\epsilon]$. Alors $\mathbf{E}(|X|^k) < +\infty$ pour tout k et*

$$\mathbf{E}(X^k) = \phi_X^{(k)}(0),$$

$\phi_X^{(k)}$ désignant la dérivée k -ième de ϕ_X .

Démonstration. On donne seulement une démonstration formelle. On sait que

$$\phi_X(t) = \mathbf{E}(e^{tX})$$

est définie au voisinage de 0. Donc, pour t dans un voisinage de 0

$$\begin{aligned} \phi_X^{(k)}(t) &= \frac{d^k}{dt^k} \mathbf{E}(e^{tX}) \\ &= \mathbf{E}\left(\frac{d^k}{dt^k} e^{tX}\right) \\ &= \mathbf{E}(X^k e^{tX}). \end{aligned}$$

En prenant $t = 0$ on obtient le résultat. Pour avoir une vraie preuve il faudrait justifier que la dérivation sous le signe \mathbf{E} est licite. \square

Théorème 5.14. *Soit X, Y deux variables aléatoires. On suppose qu'il existe $\epsilon > 0$ tel que ϕ_X et ϕ_Y soient finies et égales sur l'intervalle $]-\epsilon, +\epsilon[$. Alors X et Y ont la même loi.*

Autrement dit, la fonction génératrice des moments caractérise la loi. Nous admettrons ce résultat et nous en verrons des applications un peu plus loin.

5.4 Formule fondamentale

On a vu au chapitre précédent que si X est une variable est à valeurs dans $\{0, \dots, n\}$ alors on peut écrire

$$X = \sum_{k=0}^n k \mathbf{1}_{\{X=k\}}$$

et donc

$$\mathbb{E}(X) = \sum_{k=0}^n k \mathbb{P}(X = k).$$

Plus généralement pour toute fonction f , on a

$$\mathbb{E}(f(X)) = \sum_{k=0}^n f(k) \mathbb{P}(X = k) = \sum_{k=0}^n f(k) \mathbb{P}_X(\{k\}),$$

où \mathbb{P}_X est la loi de X . On a donc l'égalité

$$\mathbb{E}f(X) = \int_{\mathbb{R}} f(x) \mathbb{P}_X(dx).$$

Cette égalité est en fait générale.

Proposition 5.15. *Soit X une variable aléatoire et $f: \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable. On suppose que $f(X) \geq 0$ ou que $f(X)$ est intégrable. Alors*

$$\mathbb{E}f(X) = \int_{\mathbb{R}} f(x) \mathbb{P}_X(dx).$$

Démonstration. Supposons $f = \mathbf{1}_B$ pour un Borélien B . Alors

$$\mathbb{E}\mathbf{1}_B(X) = \mathbb{E}\mathbf{1}_{\{X \in B\}} = \mathbb{P}(X \in B) = \mathbb{P}_X(B) = \int_{\mathbb{R}} \mathbf{1}_B d\mathbb{P}_X,$$

donc l'égalité est vraie.

Si f est une fonction étagée positive, alors f est combinaison linéaire d'indicatrices, et l'égalité reste vraie par linéarité de l'intégrale.

Si f est mesurable positive, il existe une suite $(h_n)_{n \geq 0}$ de fonctions étagée positives telle que $h_n \nearrow f$. On a alors

$$\mathbb{E}h_n(X) = \int_{\mathbb{R}} h_n d\mathbb{P}_X$$

pour tout n . Par convergence monotone, on obtient l'égalité cherchée en passant à la limite.

Enfin si f est un fonction qui change de signe on obtient le résultat en écrivant $f = f_+ - f_-$. \square

On retrouvera ce schéma de démonstration à plusieurs reprise dans la suite du cours.

Exemple 5.16. Si X est à valeurs dans \mathbb{N} alors \mathbb{P}_X est une mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ qui ne charge que les entiers. D'après ce qu'on a vu précédemment on a donc

$$\mathbb{E}f(X) = \sum_{n=0}^{\infty} f(n) \mathbb{P}_X(\{n\}) = \sum_{n=0}^{\infty} f(n) \mathbb{P}(X = n).$$

Calculer l'espérance d'une fonction d'une variable discrète revient donc à calculer la somme d'une série.

5.5 Retour sur les variables continues

Nous sommes maintenant en mesure de donner la bonne définition de la notion de variable continue.

Définition 5.17. On appelle *densité de probabilité* une fonction $f: \mathbb{R} \rightarrow \mathbb{R}$ mesurable, positive et vérifiant

$$\int_{\mathbb{R}} f(x) dx = 1.$$

Étant donnée une densité f on définit une mesure de probabilité μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ en posant

$$\mu(B) = \int_B f(x) dx,$$

pour tout Borélien B . On dit que μ est une mesure à densité, ou que μ est absolument continue. On dit également que f est la densité de μ .

Remarque. La densité d'une mesure est définie à un ensemble de mesure nulle près : si f est une densité de μ et que $g = f$ presque partout, alors g est aussi une densité de μ .

Exemple 5.18. Soient $a < b$ des réels. La fonction $f = \mathbf{1}_{[a,b]}/(b-a)$ est une densité de probabilité. La mesure ayant pour densité f est appelée probabilité uniforme sur $[a, b]$.

Proposition 5.19. Si μ possède une densité f , on a pour toute fonction g positive ou μ -intégrable

$$\int_{\mathbb{R}} g(x) \mu(dx) = \int_{\mathbb{R}} g(x) f(x) dx.$$

Démonstration. Par définition l'égalité est vraie si g est une indicatrice. Elle est donc vraie si g est étagée positive par linéarité. Par convergence monotone on obtient l'égalité pour g positive. Enfin, si g change de signe on a le résultat en écrivant $g = g_+ - g_-$. \square

Définition 5.20. Une variable aléatoire X est dite continue si sa loi P_X est absolument continue. Autrement dit s'il existe une fonction f mesurable, positive, d'intégrale 1 vérifiant

$$P(X \in B) = \int_B f(x) dx$$

pour tout Borélien B . On dit alors que f est une densité de X .

De manière informelle $f(x)dx$ est la probabilité que la variable X tombe dans l'intervalle $[x, x + dx]$. Si on combine la Proposition 5.15 avec la proposition précédente, on obtient : pour X variable continue de densité f on a

$$Eg(X) = \int_{\mathbb{R}} g(t) f(t) dt$$

pour toute fonction g telle que $g(X)$ soit positive ou intégrable.

Exemple 5.21. Soit X une variable uniforme sur $[a, b]$. Alors X admet comme densité la fonction $\mathbf{1}_{[a,b]}/(b-a)$. En particulier

$$EX = \frac{1}{b-a} \int_a^b x dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

Remarque. Une variable ne peut pas être à la fois discrète et continue. Si X est une variable continue (de densité f) alors

$$P(X = x) = \int_{\{x\}} f(t) dt = f(x)\mathcal{L}(\{x\}) = 0$$

pour tout $x \in \mathbb{R}$. Par conséquent $P(X \in D) = 0$ pour tout ensemble dénombrable. Donc X ne peut pas être une variable discrète.

En revanche il existe des variables qui ne soient ni discrètes ni continues.

Si X est continue de densité f , sa fonction de répartition F vérifie

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f(t) dt$$

pour tout $x \in \mathbb{R}$. En particulier F est continue sur \mathbb{R} . Attention la réciproque n'est pas vraie, ce n'est pas parce que la fonction de répartition de X est continue que la variable X l'est. On a cependant le résultat suivant.

Proposition 5.22. *Soit X une variable aléatoire et F sa fonction de répartition. Si F est continue sur \mathbb{R} et de classe \mathcal{C}^1 sur \mathbb{R} privé éventuellement d'un nombre fini de points, alors X est une variable continue. Une densité associée est donnée par la fonction f donnée par $f(x) = F'(x)$ si F est dérivable en x et $f(x) = 0$ sinon.*

Démonstration. Sous ces hypothèses on a

$$\mathbb{P}(X \in]a, b]) = F(b) - F(a) = \int_a^b f(t) dt,$$

pour tout $a < b$. En utilisant le Théorème 1.10 on en déduit que

$$\mathbb{P}(X \in B) = \int_B f(t) dt$$

pour tout Borélien B . □

5.6 Lois classiques

Loi de Bernoulli Soit X à valeurs dans $\{0, 1\}$ vérifiant

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

pour un paramètre $p \in [0, 1]$. On dit que X suit une loi de Bernoulli de paramètre p . On calcule facilement

$$\mathbb{E}(X) = p, \quad \text{var}(X) = p(1 - p),$$

ainsi que la fonction génératrice des moments

$$M_X(t) = \mathbb{P}(X = 0) + e^t \mathbb{P}(X = 1) = (1 - p) + pe^t, \quad t \in \mathbb{R}.$$

Loi binomiale On jette n fois une pièce truquée qui retombe sur pile avec probabilité p . Soit X le nombre de pile obtenus. Alors $X \in \{0, \dots, n\}$ et

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

On dit que X suit la loi binomiale de paramètres n et p . Des calculs similaires à ceux effectués dans la section précédente donnent

$$\mathbb{E}(X) = np, \quad \text{var}(X) = np(1 - p).$$

Pour la fonction génératrice des moments on a

$$M_X(t) = \mathbb{E}(e^{tX}) = \sum_{k=0}^n \binom{n}{k} e^{tk} p^k (1 - p)^{n-k} = (pe^t + 1 - p)^n,$$

pour tout $t \in \mathbb{R}$.

Loi géométrique On jette la même pièce truquée de manière répétée. Le temps d'attente X du premier pile est à valeur dans les entiers non nuls et

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

On dit que X suit la loi géométrique de paramètre p . Calculons d'abord la fonction génératrice des moments

$$M_X(t) = \mathbb{E}e^{tX} = pe^t \sum_{k=1}^{+\infty} (e^t(1 - p))^{k-1}.$$

La série converge si et seulement si $e^t(1 - p) < 1$. On obtient donc

$$M_X(t) = \frac{pe^t}{1 - (1 - p)e^t}, \quad t \in]-\infty, -\ln(1 - p)[.$$

En calculant les dérivées on trouve

$$\mathbb{E}(X) = \frac{1}{p}, \quad \text{var}(X) = \frac{1 - p}{p^2}.$$

Loi de Poisson Soit $\lambda > 0$, on dit que X suit une loi de Poisson de paramètre λ si X est à valeurs dans \mathbb{N} et

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

On obtient facilement

$$\mathbb{E}(X) = \sum_{n=1}^{+\infty} n \frac{\lambda^n}{n!} e^{-\lambda} = \lambda \sum_{n=1}^{+\infty} \frac{\lambda^{n-1}}{(n-1)!} e^{-\lambda} = \lambda.$$

Un calcul analogue montre que $\mathbb{E}[X(X - 1)] = \lambda^2$ et donc $\text{var}(X) = \lambda$. Enfin

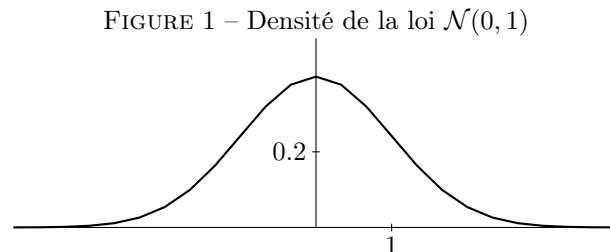
$$M_X(t) = \sum_{n=0}^{+\infty} e^{nt} \frac{\lambda^n}{n!} e^{-\lambda} = \exp(\lambda(e^t - 1)), \quad t \in \mathbb{R}.$$

Passons maintenant aux exemples de lois continues.

Gaussienne Probablement la loi la plus importante. On dit que X est une Gaussienne si X est une variable aléatoire continue de densité

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/(2\sigma^2)}, \quad x \in \mathbb{R}.$$

On démontrera que ceci définit bien une densité plus tard. On dit aussi que X suit une loi normale



de paramètres m et σ (souvent abrégée en $\mathcal{N}(m, \sigma)$). La loi $\mathcal{N}(0, 1)$ est appelée loi normale centrée réduite. Faisons les calculs d'espérance et de variance dans ce cas particulier. On a

$$\mathbb{E}(X) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} te^{-t^2/2} dt.$$

Cette intégrale est convergente : on peut par exemple dire que $te^{-t^2} \leq e^{-t}$ pour t assez grand. De plus comme on intègre une fonction impaire sur \mathbb{R} le résultat est nul. Une autre manière de faire est de remarquer que $te^{-t^2/2}$ est la dérivée de $-e^{-t^2}$. En intégrant par partie, vérifie que

$$\mathbb{E}(X^2) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} t^2 e^{-t^2/2} dt = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-t^2/2} dt = 1.$$

On a donc $\mathbb{E}(X) = 0$ et $\text{var}(X) = 1$. Pour la fonction génératrice :

$$M_X(s) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{st-t^2/2} dt = \frac{1}{\sqrt{2\pi}} e^{s^2/2} \int_{\mathbb{R}} e^{-(t-s)^2/2} dt = e^{s^2/2},$$

pour tout $s \in \mathbb{R}$. On verra plus loin que si Y suit la loi $\mathcal{N}(m, \sigma)$ alors

$$\mathbb{E}[Y] = m, \quad \text{var}(Y) = \sigma^2 \quad \text{et} \quad \mathbb{E}[e^{xY}] = e^{xm + x^2\sigma^2/2}.$$

Il n'existe pas de formule explicite pour la fonction de répartition Gaussienne

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

Néanmoins, on a les propriétés suivantes.

- $\Phi(-x) = 1 - \Phi(x)$ pour tout $x \in \mathbb{R}$.
- Pour $a > 0$ on a $1 - \Phi(a) \leq e^{-a^2/2}/2$. En effet

$$\int_a^{+\infty} e^{-x^2/2} dx \leq \int_0^{+\infty} e^{-(a+x)^2/2} dx \leq e^{-a^2/2} \int_0^{+\infty} e^{-x^2/2} dx.$$

Loi exponentielle Soit $\lambda > 0$. On dit que X suit une loi exponentielle de paramètre λ si X a pour fonction de répartition $F(x) = (1 - e^{-\lambda x})\mathbf{1}_{\mathbb{R}_+}(x)$. La variable X est alors continue de densité $f(x) = \lambda e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x)$. C'est une loi qui sert souvent à modéliser des temps d'attente (temps d'attente à un guichet, durée de vie d'une ampoule...). En intégrant par partie on trouve

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{+\infty} \lambda t e^{-\lambda t} dt = \int_0^{+\infty} e^{-\lambda t} dt = \frac{1}{\lambda}, \\ \mathbb{E}(X^2) &= \int_0^{+\infty} \lambda t^2 e^{-\lambda t} dt = \int_0^{+\infty} 2t e^{-\lambda t} dt = \frac{2}{\lambda^2}. \end{aligned}$$

Donc $\text{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \frac{1}{\lambda^2}$. Pour la fonction génératrice on trouve facilement

$$M_X(s) = \frac{\lambda}{\lambda - s}, \quad s < \lambda.$$

Loi gamma Rappelons la définition de la fonction Γ d'Euler

$$\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt$$

définie pour tout $\alpha > 0$. On a clairement $\Gamma(1) = 1$ et, en intégrant par partie $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$. On en déduit que $\Gamma(n) = (n-1)!$ pour tout entier n non nul.

Soit $\alpha > 0$ et $\lambda > 0$ on dit que X suit la loi gamma de paramètres α et λ (souvent abrégée en $\Gamma(\alpha, \lambda)$) si X a pour densité

$$f(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t} \mathbf{1}_{\{t>0\}}, \quad t \in \mathbb{R}.$$

On a

$$\mathbb{E}(X) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} t \times t^{\alpha-1} e^{-\lambda t} dt = \frac{\Gamma(\alpha+1)}{\lambda \Gamma(\alpha)} = \frac{\alpha}{\lambda}$$

De même on trouve $E(X^2) = \frac{\alpha(\alpha+1)}{\lambda^2}$ puis $\text{var}(X) = \frac{\alpha}{\lambda^2}$. Pour la fonction génératrice on calcule

$$E(e^{sX}) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} t^{\alpha-1} e^{-(\lambda-s)t} dt.$$

Cette intégrale converge si $s < \lambda$ et vaut $\lambda^\alpha / (\lambda - s)^\alpha$. On a donc

$$M_X(s) = \left(\frac{\lambda}{\lambda - s} \right)^\alpha, \quad s < \lambda.$$

Remarquons enfin que la loi $\Gamma(1, \lambda)$ n'est autre que la loi exponentielle de paramètre λ .

Loi uniforme On dit que X suit la loi uniforme sur le segment $[a, b]$ si X a pour densité la fonction $\mathbf{1}_{[a,b]}/(b-a)$. On calcule facilement

$$E(X) = \frac{a+b}{2}, \quad \text{var}(X) = \frac{(a-b)^2}{12}.$$

Pour la fonction génératrice, on a $M_X(0) = 1$ et

$$M_X(s) = \frac{e^{sb} - e^{sa}}{s(b-a)}, \quad s \neq 0.$$

Loi de Cauchy On dit que X suit une loi de Cauchy si X admet $f(x) = 1/\pi(1+x^2)$ comme densité (on montre que f est une densité en remarquant que $1/(1+x^2)$ est la dérivée de $\arctan(x)$). La variable X ne possède aucun moment puisque l'intégrale

$$\int_{\mathbb{R}} \frac{t}{1+t^2} dt$$

est divergente.

5.7 Changement de variable I

On a vu précédemment que quand X est une variable discrète et ϕ une fonction quelconque $\phi(X)$ est une variable discrète. Ceci ne marche pas pour les variables continues. En effet si ϕ est la fonction nulle, alors $\phi(X)$ est la variable constante égale à 0 qui n'est évidemment pas une variable continue (sa fonction de répartition fait un saut en 0). Pour que $\phi(X)$ soit une variable continue il faut imposer des conditions à ϕ .

Rappelons que si X est une variable continue de densité f et h une fonction positive on a

$$Eh(X) = \int_{\mathbb{R}} h(t)f(t) dt.$$

Réciproquement, supposons cette égalité vraie pour toute fonction positive h . En l'appliquant à $h = \mathbf{1}_B$ on obtient

$$P(X \in B) = \int_B f(t) dt.$$

On peut faire ceci pour tout Borélien B , ce qui montre que X est continue de densité f . Nous admettrons le résultat suivant (formule de changement de variable).

Théorème 5.23. Soit U un ouvert de \mathbb{R} et soit $\phi: U \rightarrow V$ une fonction bijective de classe \mathcal{C}^1 . Alors

$$\int_{\phi(U)} f(t) dt = \int_U f \circ \phi(t) |\phi'(t)| dt$$

pour toute fonction f telle que l'intégrale de gauche a un sens.

Soient U et V des ouverts de \mathbb{R} et $\phi: U \rightarrow V$ une fonction, on dit que ϕ est un \mathcal{C}^1 -difféomorphisme si ϕ est de classe \mathcal{C}^1 , bijective, et si sa réciproque est de classe \mathcal{C}^1 .

Proposition 5.24. Soit X une variable continue à valeurs dans un ouvert U et soit $\phi: U \rightarrow V$ un \mathcal{C}^1 -difféomorphisme. Alors $Y = \phi(X)$ est une variable continue de densité

$$f_Y(y) = f_X \circ \phi^{-1}(y) |(\phi^{-1})'(y)| \mathbf{1}_{\phi(U)}(y),$$

où f_X désigne une densité de X .

Démonstration. Soit f la densité de X . Par hypothèse f est nulle en dehors de U . Soit h une fonction positive. En appliquant la formule du changement de variable à ϕ^{-1} on a

$$\begin{aligned} \mathbb{E}h(\phi(X)) &= \int_U h \circ \phi(x) f(x) dx \\ &= \int_{\phi(U)} h(y) f \circ \phi^{-1}(y) |(\phi^{-1})'(y)| dy. \end{aligned}$$

D'où le résultat. □

La fonction h utilisée dans la preuve précédente est souvent appelée *fonction test* ; on parle de méthode de la fonction test.

Exemple 5.25. Un cas simple est quand la fonction ϕ est affine : $\phi(x) = ax + b$. Si X est une variable aléatoire de densité f , on a

$$\mathbb{E}h(aX + b) = \int_{\mathbb{R}} h(y) f\left(\frac{y-b}{a}\right) \frac{1}{a} dy$$

pour toute fonction test h , ce qui montre que la variable $aX + b$ admet $t \mapsto a^{-1}f(a^{-1}(t - b))$ comme densité.

Par exemple si X suit la loi $\mathcal{N}(0, 1)$ on trouve que $aX + b$ suit la loi $\mathcal{N}(b, a^2)$. On peut déduire facilement l'espérance et la variance de la loi $\mathcal{N}(b, a^2)$:

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b = b, \quad \text{var}(aX + b) = a^2 \text{var}(X) = a^2.$$

Ainsi que la fonction génératrice des moments

$$\mathbb{E}(e^{t(aX+b)}) = \mathbb{E}(e^{(ta)X})e^b = e^{t^2 a^2 / 2} e^b.$$

Dans le même ordre d'idée, si Y suit une loi $\Gamma(\alpha, \lambda)$ et si t est un réel strictement positif. Alors tY suit une loi $\Gamma(\alpha, \lambda/t)$.

Exemple 5.26 (loi log-normale). Soit X suivant une loi normale centrée réduite. La loi de $Y = e^X$ est appelé log-normale. On cherche sa densité. La fonction exp réalise une bijection entre \mathbb{R} et $]0, +\infty[$. Sa réciproque $\log(x)$ admet comme dérivée $1/x$. Soit h une fonction test, en utilisant la formule de changement de variable on a

$$\begin{aligned} \mathbb{E}h(Y) = \mathbb{E}h(e^X) &= \int_{\mathbb{R}} h(e^x) e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} h(y) e^{-\ln(y)^2/2} \frac{1}{y} dy. \end{aligned}$$

On en déduit que Y admet comme densité la fonction

$$g(y) = \mathbf{1}_{]0, +\infty[}(y) e^{-\ln(y)^2/2} \frac{1}{y\sqrt{2\pi}}.$$

Il n'est pas obligatoire de suivre cette méthode. On sait que $Y \geq 0$ et que pour $x \geq 0$

$$\mathbb{P}(Y \leq x) = \mathbb{P}(X \leq \ln(x)) = \Phi(\ln(x))$$

où Φ est la répartition Gaussienne :

$$\Phi(x) = \mathbb{P}(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

La fonction de répartition de Y est donc la fonction

$$F(x) = \Phi(\log(x))\mathbf{1}_{[0,+\infty[}.$$

Cette fonction est clairement \mathcal{C}^1 sur \mathbb{R}^* et on vérifie aisément qu'elle est continue en 0. Par conséquent Y est une variable continue admettant comme densité

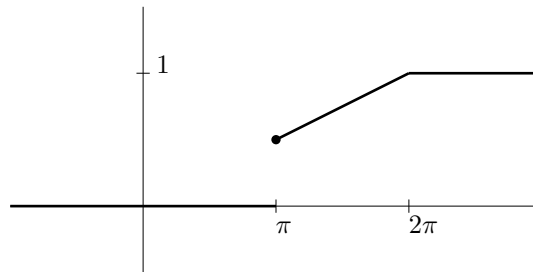
$$F'(x) = \Phi'(\log(x))\frac{1}{x} = e^{-\ln(x)^2/2} \frac{1}{x\sqrt{2\pi}}.$$

si $x \geq 0$ et $F'(x) = 0$ si $x < 0$.

5.8 Un exemple ni discret ni continu

On a vu qu'une variable aléatoire ne pouvait être à la fois discrète et continue. En revanche il existe des variables qui ne soient ni l'un ni l'autre.

Exemple 5.27. Soit θ l'angle donné par l'aiguille de Buffon et soit $X = \max(\theta, \pi)$. On a $P(X \leq x) = 0$ pour tout $x < \pi$. Pour $x \geq \pi$ on a $P(X \leq x) = P(\theta \leq x)$. La fonction de répartition de X a donc le graphe suivant



La variable X n'est pas continue puisque $P(X = \pi) = 1/2$. La variable n'est pas non plus discrète puisque sa fonction de répartition n'est pas en escalier. Plus précisément on voit sur le graphe que $P(X = x) = 0$ pour tout $x \neq \pi$. Par conséquent, si $D = \{x_1, x_2, \dots\}$ est un ensemble dénombrable $P(X \in D) = \sum_i P(X = x_i)$ vaut 0 ou $1/2$ selon que π appartienne ou pas à D . Dans tous les cas $P(X \in D) < 1$.

On peut néanmoins calculer l'espérance de X : remarquons que $\max(\theta, \pi) = \pi\mathbf{1}_{\{\theta < \pi\}} + \theta\mathbf{1}_{\{\theta \geq \pi\}}$. Par linéarité de l'espérance

$$E(X) = \pi E(\mathbf{1}_{\{\theta < \pi\}}) + E(\theta\mathbf{1}_{\{\theta \geq \pi\}}).$$

Comme θ est une variable uniforme sur $[0, 2\pi]$

$$E(\mathbf{1}_{\{\theta < \pi\}}) = P(\theta \leq \pi) = \frac{1}{2}.$$

$$E(\theta\mathbf{1}_{\{\theta \geq \pi\}}) = \frac{1}{2\pi} \int_{\pi}^{2\pi} t dt = \frac{3\pi}{4}.$$

Donc $E(X) = \pi/2 + 3\pi/4 = 5\pi/4$.

5.9 Exercices

Exercice 5.1. On lance un dé de manière répétée, et on s'arrête dès qu'il donne 1 ou un nombre supérieur ou égal à 3.

1. Quel est le score moyen du dernier lancer ?
2. Quelle est la moyenne du carré du dernier lancer ?

Exercice 5.2. Soit X une variable aléatoire intégrable à valeurs dans \mathbb{N} . Montrer que

$$E(X) = \sum_{n=0}^{\infty} P(X > n).$$

Retrouver en utilisant cette formule l'espérance de la loi géométrique.

Exercice 5.3. Chaque paquet de céréales contient une figurine. Il y a n figurines différentes.

1. Si on a déjà k figurines distinctes, combien de paquets supplémentaires faut-il acheter en moyenne pour en avoir $k + 1$.
2. Déterminer le nombre moyen de paquets à acheter pour obtenir les n figurines.
3. Donner un équivalent de ce nombre quand n tend vers l'infini.

Exercice 5.4 (loi binomiale négative). Une urne contient des boules blanches et des boules rouges, les boules blanches étant en proportion p . On tire les boules une à une avec remise et on note X_n le rang d'apparition de la n -ième boule blanche.

1. Donner la loi de X_1 et calculer son espérance.
2. Déterminer la loi de X_n . En déduire que

$$\sum_{k=n}^{+\infty} p^n (1-p)^{k-n} \binom{k-1}{n-1} = 1.$$

3. Calculer $\mathbf{E}X_n$. Indication : utiliser $k \binom{k-1}{n-1} = n \binom{k}{n}$.
4. On note Y_n le nombre de boules rouges apparues avant la n -ième boule blanche. Écrire Y_n en fonction de X_n .
5. En déduire la loi de Y_n , puis son espérance.

Exercice 5.5. On considère une urne contenant b boules bleues et r boules rouges. On effectue des tirages successifs sans remise et on s'intéresse au rang X d'apparition de la première boule bleue.

1. Calculer $\mathbf{P}(X > n)$ pour tout entier n .
2. Montrer que $\mathbf{E}(X) = (b + r + 1)/(b + 1)$.
Indication : Montrer par récurrence sur q que pour tout entier p

$$\sum_{i=0}^q \binom{p+i}{p} = \binom{p+q+1}{p+1}.$$

Exercice 5.6. On suppose que $\text{var}(X) = 0$. Que peut-on dire de X ?

Exercice 5.7. Soit X une variable aléatoire. On dit qu'un nombre m est une médiane pour X si $\mathbf{P}(X \leq m) \geq 1/2$ et $\mathbf{P}(X \geq m) \geq 1/2$.

1. Soit F la fonction de répartition de X . On pose

$$m = \inf \{x \in \mathbb{R} : F(x) \geq 1/2\}.$$

Montrer que m est bien défini puis que m est une médiane.

2. Montrer par un exemple que X peut avoir plusieurs médianes.
3. On suppose que X est de carré intégrable, montrer que toute médiane m vérifie

$$|m - \mathbf{E}(X)| \leq \sqrt{2\text{var}(X)}.$$

Exercice 5.8. Soit X une variable aléatoire et soit $Y = aX + b$.

1. Montrer que $M_Y(t) = e^{tb} M_X(at)$.
2. Calculer la fonction génératrice des moments de la loi $\mathcal{N}(0, 1)$ et en déduire celle de la loi $\mathcal{N}(m, \sigma^2)$.

Exercice 5.9. Soit X_n une variable de loi Gamma de paramètres n, λ .

1. Calculer la fonction génératrice des moments de X_n . Préciser son domaine de définition.
2. En déduire la fonction génératrice des moments M_{Y_n} de $Y_n = X_n/n$. Calculer la limite de $M_{Y_n}(t)$ en tout point $t \in \mathbb{R}$ où cela a un sens.

3. La fonction limite obtenue est la fonction génératrice des moments d'une variable réelle. Laquelle ?

Exercice 5.10. Soit α un réel strictement positif. On dit qu'un variable aléatoire X suit une loi de Pareto de paramètre α si elle est continue de densité

$$f: x \in \mathbb{R} \mapsto \frac{\alpha}{x^{\alpha+1}} \mathbf{1}_{[1,+\infty[}(x).$$

1. Déterminer la fonction de réartition de X .
2. Déterminer l'espérance de X .
3. Déterminer les lois des variables aléatoires suivantes ainsi que leurs espérances :
 - (a) $U = \ln(X)$,
 - (b) $V = \sqrt{X}$.

Exercice 5.11 (Inégalité de Jensen). Soit X une variable aléatoire et φ une fonction convexe de classe \mathcal{C}^1 . On suppose que X et $\varphi(X)$ sont intégrables. On pose $m = \mathbf{E}(X)$. En utilisant le fait que φ est au-dessus de sa tangente en m , montrer que

$$\varphi(\mathbf{E}X) \leq \mathbf{E}\varphi(X).$$

Exercice 5.12. Soit X une variable de densité $f(x) = k(x - x^2)\mathbf{1}_{[0,1]}(x)$.

1. Montrer que f est une bien une densité de probabilité.
2. Calculer la fonction de répartition de X , son espérance et sa variance.
3. On pose $Y = 1 - 2X$. Déterminer la loi de Y et calculer son espérance et sa variance.

Exercice 5.13. Soit X de loi $\mathcal{N}(0, 1)$ et Y une variable définie par

$$Y(\omega) = \begin{cases} X(\omega) & \text{si } X(\omega) > 0, \\ 2X(\omega) & \text{si } X(\omega) \leq 0. \end{cases}$$

Déterminer la loi de Y , son espérance et sa variance.

Exercice 5.14. Soit X une v.a. suivant la loi exponentielle de paramètre $\lambda > 0$. On pose $Y = [X]$, où $[X]$ désigne la partie entière de X .

1. Trouver la loi de Y , calculer son espérance et sa variance.
2. Montrer que $Z = X - [X]$ est une v.a. continue et calculer sa densité et son espérance.

Exercice 5.15. Dans un repère orthonormé, soient $O = (0, 0)$, $I = (-1, 0)$ et $J = (1, 0)$. On considère un demi-cercle posé sur le segment $[I, J]$ et un point M aléatoire sur le demi-cercle. Quelle est l'ordonnée moyenne de M si

1. L'angle $(\overrightarrow{OI}, \overrightarrow{OM})$ est uniforme sur $[0, \pi]$?
2. L'abscisse de M est uniforme sur $[-1, 1]$?

6 Variables indépendantes

6.1 Définition

Définition 6.1. Soient X, Y des variables aléatoires. On dit que X et Y sont indépendantes si pour tous $A, B \in \mathcal{B}(\mathbb{R})$

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

Proposition 6.2. Soient X et Y des variables discrètes, X et Y sont indépendantes si et seulement si

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y), \quad \forall x, y \in \mathbb{R}. \quad (5)$$

Remarque. Attention ceci ne marche que pour les variables discrètes. Par exemple si X et Y sont continues, on a toujours

$$\mathbb{P}(X = x, Y = y) = 0 = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

Démonstration. On suppose qu'on a (5) et que X et Y sont à valeurs dans \mathbb{N} (par exemple). Soient A, B des Boréliens de \mathbb{R} , on a

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B) &= \sum_{i \in A, j \in B} \mathbb{P}(X = i, Y = j) \\ &= \sum_{i \in A, j \in B} \mathbb{P}(X = i)\mathbb{P}(Y = j) \\ &= \sum_{i \in A} \mathbb{P}(X = i) \sum_{j \in B} \mathbb{P}(Y = j) \\ &= \mathbb{P}(X \in A)\mathbb{P}(Y \in B), \end{aligned}$$

ce qui montre que X et Y sont indépendantes. La réciproque est évidente. \square

Exemple 6.3. On lance deux dés, on appelle X et Y les résultats respectifs du premier et deuxième dé. On a

$$\mathbb{P}(X = 2, Y = 3) = \frac{1}{36} = \mathbb{P}(X = 2)\mathbb{P}(Y = 3).$$

Donc $\{X = 2\}$ est indépendant de $\{Y = 3\}$ et de même pour les autres valeurs de X et Y . Ainsi X est indépendant de Y .

Proposition 6.4. Si X et Y sont indépendantes alors

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)],$$

pour toutes fonctions f et g (telles que les espérances précédentes soient bien définies).

Démonstration. C'est toujours la même histoire : on a la propriété pour les fonctions indicatrices, par linéarité on l'obtient pour les fonctions étagées, puis pour les fonctions positives par convergence monotone, etc... \square

Proposition 6.5. Soient X, Y des variables aléatoires indépendantes. On a $M_{X+Y} = M_X M_Y$.

Démonstration. On utilise la proposition précédente :

$$M_{X+Y}(t) = \mathbb{E}(e^{t(X+Y)}) = \mathbb{E}(e^{tX}e^{tY}) = \mathbb{E}(e^{tX})\mathbb{E}(e^{tY}) = M_X(t)M_Y(t). \quad \square$$

Exemple 6.6. Soit X, Y des variables indépendantes suivant des lois de Poisson de paramètres λ et μ , respectivement. On calcule la fonction génératrice des moments de $X + Y$:

$$\begin{aligned} M_{X+Y}(t) &= M_X(t)M_Y(t) \\ &= \exp(\lambda(e^t - 1)) \exp(\mu(e^t - 1)) \\ &= \exp((\lambda + \mu)(e^t - 1)). \end{aligned}$$

On reconnaît la fonction génératrice des moments de la loi de Poisson de paramètre $\lambda + \mu$. Comme la fonction génératrice des moments caractérise la loi, la variable $X + Y$ suit la loi de Poisson de paramètre $\lambda + \mu$.

Définition 6.7. On dit que $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes si pour toute suite finie d'indices $0 \leq i_1 < i_2 < \dots < i_n$ et de Boréliens B_1, \dots, B_n on a

$$\mathbb{P}(X_{i_1} \in B_1, \dots, X_{i_n} \in B_n) = \mathbb{P}(X_{i_1} \in B_1) \times \dots \times \mathbb{P}(X_{i_n} \in B_n).$$

Si on a seulement

$$\mathbb{P}(X_i \in A, X_j \in B) = \mathbb{P}(X_i \in A)\mathbb{P}(X_j \in B)$$

pour tout $i \neq j$ on dit que les variables $(X_i)_{i \geq 1}$ sont deux à deux indépendantes.

6.2 Covariance

Définition 6.8. On appelle covariance de deux variables de carré intégrable la quantité

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Remarque. Comme $2|XY| \leq X^2 + Y^2$, la covariance de X et Y est définie dès que X et Y sont de carré intégrable.

La covariance sert à calculer la variance de la somme :

$$\begin{aligned} \text{var}(X + Y) &= \mathbb{E}[(X + Y)^2] - [\mathbb{E}(X + Y)]^2 \\ &= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - (\mathbb{E}X)^2 - 2(\mathbb{E}X)(\mathbb{E}Y) + (\mathbb{E}Y)^2 \\ &= \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y). \end{aligned}$$

Plus généralement on a

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

pour toute famille X_1, \dots, X_n de variables de carré intégrable.

Lemme 6.9. Si X, Y sont indépendantes et de carré intégrable alors $\text{cov}(X, Y) = 0$.

Démonstration. Il suffit d'appliquer la Proposition 6.4 : on a $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. □

Remarque. La réciproque n'est pas vraie, on peut avoir $\text{cov}(X, Y) = 0$ sans que X et Y soient indépendantes.

Par conséquent si X et Y sont indépendantes on a $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$. Plus généralement, si X_1, X_2, \dots, X_n sont indépendantes

$$\text{var}(X_1 + \dots + X_n) = \text{var}(X_1) + \dots + \text{var}(X_n).$$

Exemple 6.10. On peut retrouver la variance de la binomiale ainsi. Soient X_1, \dots, X_n des variables indépendantes suivant toutes la loi de Bernoulli de paramètre p . Alors $X = \sum_i X_i$ suit une binomiale de paramètres n et p et

$$\text{var}(X) = \text{var}(X_1) + \dots + \text{var}(X_n) = np(1 - p).$$

Des variables qui, comme dans l'exemple précédent, sont indépendantes et de même loi sont dites i.i.d. (pour "indépendantes identiquement distribuées").

6.3 Exercices

Exercice 6.1. On considère une main de poker, le nombre d'as est-il indépendant du nombre de rois ? du nombre de trèfles ?

Exercice 6.2. Soit X et Y des variables aléatoires indépendantes, prenant toutes les deux les valeurs 1 et -1 avec probabilité $1/2$. Montrer que X, Y et XY sont deux à deux indépendantes. Sont-elles mutuellement indépendantes ?

Exercice 6.3 (loi faible des grands nombres). Soit $(X_i)_{i \geq 1}$ une suite de variables i.i.d. de carré intégrable. On pose $m = \mathbf{E}(X_1)$ et $\sigma = \mathbf{var}(X_1)$.

1. Exprimer la variance de $(X_1 + \dots + X_n)/n$ en fonction de n et σ .
2. Montrer que pour tout $\epsilon > 0$ on a

$$\lim_{n \rightarrow +\infty} \mathbf{P} \left(\left| \frac{X_1 + \dots + X_n}{n} - m \right| \geq \epsilon \right) = 0.$$

On dit que $(X_1 + \dots + X_n)/n$ converge vers m en probabilité.

Exercice 6.4. Soient X, Y des variables indépendantes.

1. Exprimer la fonction de répartition de $\max(X, Y)$ en fonction des répartitions de X et Y .
Même question avec $\min(X, Y)$.
2. On suppose maintenant que X et Y sont uniformes sur $[0, 1]$, déterminer la loi de $\max(X, Y)$.

Exercice 6.5. Soient n, m des entiers, et $p \in [0, 1]$. Soient X et Y des variables indépendantes et binomiales de paramètres n, p et m, p , respectivement.

1. Calculer les fonctions génératrices des moments de X et Y .
2. Déterminer la fonction génératrice des moments de $X + Y$.
3. En déduire la loi de $X + Y$.

Exercice 6.6. Soient X, Y des variables aléatoires indépendantes et de même loi géométrique de paramètre p .

1. Déterminer la loi de $\min(X, Y)$.
2. Calculer la fonction de masse de $X - Y$.
3. Montrer que $\min(X, Y)$ est indépendante de $X - Y$.

Exercice 6.7. Soient X_1, X_2 des variables indépendantes de lois respectives $\Gamma(n_1, \lambda)$ et $\Gamma(n_2, \lambda)$.

1. Calculer la fonction génératrice des moments de $X_1 + X_2$ et en déduire sa loi.
2. Quelle est la loi de $Y_1 + \dots + Y_n$ si les Y_i sont i.i.d de loi exponentielle ?

Exercice 6.8. On considère n lancers de dés successifs. On appelle A_{ij} l'événement "les résultats des i -ème et j -ème lancers sont égaux" et on pose

$$X = \sum_{1 \leq i < j \leq n} \mathbf{1}_{A_{ij}}.$$

1. Calculer l'espérance de X .
2. Montrer que les événements $(A_{ij})_{1 \leq i < j \leq n}$ sont deux à deux indépendants. Sont-ils mutuellement indépendants ?
3. Déterminer la variance de X .

Exercice 6.9. On modélise les durées qui s'écoulent entre deux passages successifs de bus à un arrêt par une suite $(\tau_i)_{i \geq 1}$ de variables i.i.d. de loi exponentielle de paramètre $\lambda > 0$: le premier bus met un temps τ_1 à arriver, puis il s'écoule un temps τ_2 avant le passage du deuxième, etc... Pour tout entier $n \geq 1$ on appelle

$$T_n = \sum_{i=1}^n \tau_i$$

le temps de passage du n -ième bus.

1. Soit $n \geq 1$. Déterminer la loi de T_n .
Indication : On pourra calculer sa fonction génératrice des moments.
2. Montrer que pour tout entier $n \geq 1$ et pour tout $t > 0$, on a

$$\mathbf{P}(T_n \leq t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} + \mathbf{P}(T_{n+1} \leq t).$$

Indication : On pourra intégrer par parties.

3. On fixe $t > 0$. Déduire de la question précédente la loi du nombre $N(t)$ de bus déjà passés à l'instant t .

7 Marche aléatoire simple

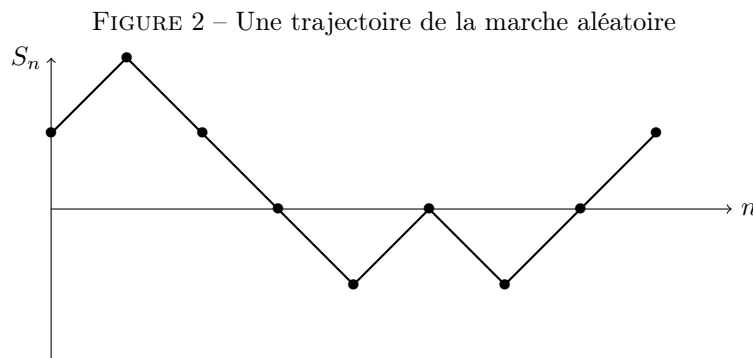
Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées à valeurs dans $\{-1, 1\}$. On se donne un entier relatif S_0 et pour $n \geq 1$ on pose

$$S_n = S_0 + \sum_{k=1}^n X_k.$$

La suite de variables $(S_n)_{n \geq 0}$ est appelée marche aléatoire simple sur \mathbb{Z} . On notera p le paramètre

$$p = P(X_i = 1).$$

Dans la suite il sera utile de représenter graphiquement les trajectoires de la marche aléatoire en mettant en abscisse le temps n et en ordonnées la position S_n de la marche.



7.1 Dénombrement de trajectoires

Soient a, b des entiers relatifs et n un entier naturel. On note

$$N_n(a, b)$$

le nombre de trajectoires allant de la hauteur a à la hauteur b en n pas.

Lemme 7.1. On a

$$N_n(a, b) = \binom{n}{\frac{1}{2}(n+b-a)}.$$

Remarque. Dans la suite, et pour ne pas avoir à distinguer les cas sans arrêt, on adoptera la convention

$$\binom{n}{x} = 0$$

dès que x n'est pas un entier compris entre 0 et n . Par exemple, dans le résultat précédent, le nombre de trajectoire est nul si $n + b - a$ est impair, ou si $|b - a| > n$.

Démonstration. Soit x le nombre de pas vers le haut que fait une trajectoire allant de a à b en n pas. On a alors

$$0 \leq x \leq n.$$

De plus, la trajectoire faisant $n - x$ pas vers le bas on a

$$x - (n - x) = b - a,$$

ce qui implique que $n + b - a$ est un entier naturel pair et que

$$x = \frac{b - a + n}{2}.$$

Pour dénombrer les nombres de telles trajectoires il suffit ensuite de choisir les positions des x pas vers le haut parmi les n pas, d'où le résultat. \square

On en déduit facilement la fonction de masse de S_n .

Corollaire 7.2. *Supposons que $S_0 = 0$. On a pour tout $b \in \mathbb{Z}$*

$$\mathbb{P}(S_n = a) = \binom{n}{\frac{1}{2}(n+b)} (1-p)^{\frac{1}{2}(n-b)} p^{\frac{1}{2}(n+b)}$$

Démonstration. Une trajectoire allant de 0 à b en n pas fait $(n+b)/2$ pas vers le haut et $(n-b)/2$ pas vers le bas. La probabilité individuelle d'une telle trajectoire est donc

$$(1-p)^{\frac{1}{2}(n-b)} p^{\frac{1}{2}(n+b)}.$$

Ensuite on utilise le lemme précédent. □

On utilisera très souvent le lemme suivant.

Lemme 7.3 (Principe de réflexion). *Soient a, b des entiers strictement positifs. On appelle*

$$N_n^0(a, b)$$

Le nombre de trajectoires menant de a à b en n pas qui touchent l'axe des abscisses. Alors

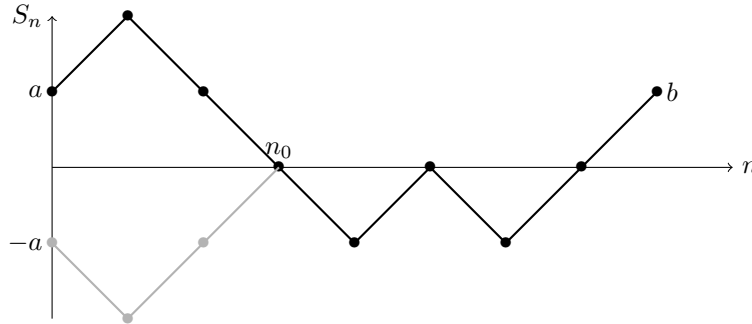
$$N_n^0(a, b) = N_n(-a, b).$$

Démonstration. Étant donnée une trajectoire $(x_k)_{k=0}^n$ allant de a à b en touchant 0, on appelle n_0 le premier instant où elle touche 0. On considère ensuite

$$x'_k = \begin{cases} -x_k & \text{si } k \leq n_0, \\ x_k & \text{sinon.} \end{cases}$$

Alors x'_k est une trajectoire de $-a$ à b et de plus l'application qui à (x_k) associe (x'_k) est clairement bijective. □

FIGURE 3 – Principe de réflexion



7.2 Temps de retour en 0 pour la marche symétrique

On suppose désormais que $p = 1/2$ (marche symétrique) et que la marche part de 0. On s'intéresse au premier temps T de retour en 0 de la marche :

$$T = \inf\{n \geq 1, S_n = 0\}.$$

Théorème 7.4.

$$\mathbb{P}(T > 2n) = \binom{2n}{n} 2^{-2n} = \mathbb{P}(S_{2n} = 0).$$

Remarque. Comme S_{2n+1} est impair on a $S_{2n+1} \neq 0$ et donc

$$\mathbb{P}(T > 2n + 1) = \mathbb{P}(T > 2n).$$

Démonstration. Soit k un entier strictement positif. Calculons $P(T > 2n, S_{2n} = 2k)$, la probabilité que la marche soit égale à $2k$ au temps $2n$ sans avoir retouché 0. Toute trajectoire de longueur $2n$ a probabilité 2^{-2n} . Le nombre de trajectoires allant de 0 à $2k$ en $2n$ pas sans retoucher 0 est égal au nombre de trajectoires allant de 1 à $2k$ en $2n - 1$ pas sans toucher 0. D'après le principe de réflexion, ce nombre est égal à

$$\begin{aligned} N_{2n-1}(1, 2k) - N_{2n-1}^0(1, 2k) &= N_{2n-1}(1, 2k) - N_{2n-1}(-1, 2k) \\ &= \binom{2n-1}{n+k-1} - \binom{2n-1}{n+k} \end{aligned}$$

Par conséquent pour tout $k \geq 1$

$$P(T > 2n, S_{2n} = 2k) = 2^{-2n} \left(\binom{2n-1}{n+k-1} - \binom{2n-1}{n+k} \right).$$

D'après la symétrie de la marche, on obtient le même résultat pour $P(T > 2n, S_{2n} = -2k)$. Par conséquent

$$\begin{aligned} P(T > 2n) &= \sum_{k \neq 0} P(T > 2n, S_{2n} = 2k) \\ &= 2 \sum_{k=1}^n P(T_0^* > 2n, S_{2n} = 2k) \\ &= 2 \cdot 2^{-2n} \sum_{k=1}^n \left(\binom{2n-1}{n+k-1} - \binom{2n-1}{n+k} \right) \\ &= 2 \cdot 2^{-2n} \binom{2n-1}{n} = 2^{-2n} \binom{2n}{n}, \end{aligned}$$

ce qu'il fallait démontrer. \square

Donnons maintenant quelques conséquences de ce résultat. Rappelons la formule de Stirling

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n}.$$

On en déduit facilement que

$$\binom{2n}{n} \sim \frac{2^{2n}}{\sqrt{\pi n}}.$$

Corollaire 7.5.

$$P(T > 2n) \sim \frac{1}{\sqrt{\pi n}}.$$

On a donc en particulier

$$P(T > 2n) \rightarrow 0,$$

ce qui montre que T est fini presque sûrement. Autrement dit, presque sûrement la marche revient en 0. On peut même en déduire assez facilement que S_n revient en 0 une infinité de fois. En effet, appelons T_n le temps du n -ième retour en 0 de la marche. Il est clair que pour tout $n \geq 1$

$$T_{n+1} - T_n = T, \quad \text{en loi.}$$

Par conséquent on a $P(T_{n+1} - T_n < +\infty) = 1$ pour tout n . Et comme une intersection dénombrable d'événements quasi-certains est un événement quasi-certain, on en déduit que presque sûrement

$$T_n < +\infty, \quad \forall n \geq 1.$$

On dit que la marche est *récurrente*.

En revanche T n'est pas intégrable. En effet comme la somme des $1/\sqrt{n}$ diverge

$$\begin{aligned} E[T] &= \sum_{n=0}^{\infty} P(T > n) \\ &= 2 \sum_{n=0}^{\infty} P(T > 2n) = +\infty. \end{aligned}$$

La marche retourne en 0 certes, mais le temps moyen de retour en 0 est infini.

7.3 Loi de l'arc sinus

On suppose toujours que la marche est symétrique et part de 0. Étant donné un entier n fixé on cherche maintenant la loi du dernier passage en 0 avant le temps n :

$$D_n = \sup\{k \leq n, S_k = 0\}$$

Théorème 7.6. Pour $k \in \{0, \dots, n\}$

$$\begin{aligned} \mathbb{P}(D_{2n} = 2k) &= \mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2n-2k} = 0) \\ &= \binom{2k}{k} \binom{2(n-k)}{n-k} 2^{-2n}. \end{aligned}$$

Démonstration. Il s'agit de dénombrer les trajectoires de longueur $2n$ partant de 0 et touchant 0 pour la dernière fois au temps $2k$. Une telle trajectoire est formée d'une trajectoire allant de 0 à 0 en $2k$ pas, suivie d'une trajectoire de longueur $2(n-k)$ partant de 0 et ne retouchant pas 0. On a vu précédemment qu'il $\binom{2k}{k}$ possibilités pour le premier morceau, et $\binom{2(n-k)}{n-k}$ pour le deuxième. Le résultat s'en suit facilement. \square

On peut remarquer que la loi de R_{2n} est symétrique par rapport à n , en particulier R_{2n} a autant de chances d'être proche de 0 que de $2n$. Par ailleurs si que n , k et $n-k$ tendent vers l'infini, on obtient en utilisant la formule de Stirling

$$\mathbb{P}(R_{2n} = 2k) \sim \frac{1}{\pi \sqrt{k(n-k)}},$$

qui est d'autant plus grand que k des bords 0 et n ; et d'autant plus petit que k est proche du milieu $n/2$. En travaillant un peu plus on peut montrer que la variable $R_{2n}/2n$ converge en un certain sens vers une variable continue de densité

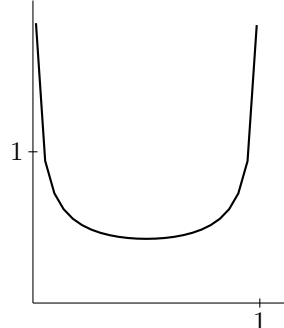
$$\frac{1}{\pi x(1-x)} \mathbf{1}_{[0,1]}(x).$$

Cette loi est appelée loi de l'arc sinus, parce que sa fonction de répartition F vaut

$$F(x) = \frac{1}{\pi} \int_0^x \frac{1}{\sqrt{y(1-y)}} dy = \frac{2}{\pi} \arcsin(\sqrt{x})$$

pour tout $x \in [0,1]$. La loi de l'arc sinus peut paraître surprenante. On pourrait penser que dans

FIGURE 4 – Densité de la loi de l'arc sinus



une suite de jeu de pile ou face les instants où on a obtenu autant de piles que de faces reviennent assez fréquemment, et qu'il y a donc de fortes chances que l'instant en question le plus récent soit proche de la fin. Le résultat précédent montre qu'il n'en est rien, cet instant a autant de chance d'être proche du début que de la fin, et il a beaucoup plus de chances d'être proche du début ou de la fin que du milieu.

7.4 Loi des grands nombres

Théorème 7.7. *Presque sûrement*

$$\frac{S_n}{n} \longrightarrow 0$$

Autrement dit dans une suite infinie jeu de pile ou face, la proportion de piles obtenus tend vers $1/2$, presque sûrement.

Démonstration. Il suffit de montrer que pour tout $\epsilon > 0$ l'événement

$$\{|S_n| > n\epsilon\}$$

ne se produit que pour un nombre fini d'entier n , presque sûrement. En effet, en prenant le complémentaire on a alors presque sûrement

$$\left| \frac{S_n}{n} \right| \leq \frac{1}{k}, \text{ pour } n \text{ suffisamment grand.}$$

Et ce pour tout entier $k \geq 1$. Enfin comme une intersection dénombrable d'événements quasi-certains est un événement quasi-certain, on en déduit que presque sûrement

$$\forall k \geq 1, \left| \frac{S_n}{n} \right| \leq \frac{1}{k}, \text{ pour } n \text{ suffisamment grand,}$$

ce qui est le résultat. D'après le lemme de Borel–Cantelli ci-dessous il suffit de montrer que pour tout $\epsilon > 0$

$$\sum_{n \geq 1} \mathbb{P}(|S_n| > n\epsilon) < +\infty. \quad (6)$$

Or d'après l'inégalité de Markov

$$\mathbb{P}(|S_n| > n\epsilon) = \mathbb{P}(|S_n|^4 > n^4 \epsilon^4) \leq \frac{1}{n^4 \epsilon^4} \mathbb{E}[S_n^4].$$

Rappelons que $S_n = X_1 + \dots + X_n$, on a en développant et en utilisant la linéarité de l'espérance

$$\mathbb{E}[S_n^4] = \sum \mathbb{E}[X_i^4] + 3 \sum \mathbb{E}[X_i^3 X_j] + 3 \sum \mathbb{E}[X_i^2 X_j^2] + \sum \mathbb{E}[X_i X_j X_k^2] + \sum \mathbb{E}[X_i X_j X_k X_l],$$

avec la convention que les sommations sont faites sur des indices i, j, k deux à deux distincts. Comme les X_i sont d'espérance nulle et indépendantes on a

$$\mathbb{E}[X_i^3 X_j] = \mathbb{E}[X_i^3] \mathbb{E}[X_j] = 0.$$

Et de même pour les derniers termes. De plus $\mathbb{E}[X_i^2] = 1$ et $\mathbb{E}[X_i^4] = 1$ donc

$$\mathbb{E}[S_n^4] = \sum_i \mathbb{E}[X_i^4] + \sum_{i \neq j} \mathbb{E}[X_i^2]^2 \leq Cn^2,$$

pour une certaine constante C qui n'a pas d'importance. On en déduit que

$$\mathbb{P}(|S_n| > n\epsilon) \leq \frac{C}{n^2 \epsilon^4},$$

ce qui implique (6) et termine la preuve. \square

On a seulement utilisé le fait que les X_i sont indépendantes de moyennes nulles et admettent un moment d'ordre 4. La même démonstration donnerait donc le théorème suivant.

Théorème 7.8. *Soit (X_i) une suite de variables i.i.d. telles que $\mathbb{E}[X_i^4] < +\infty$. Alors presque sûrement*

$$\frac{1}{n} \sum_{i=1}^n X_i \longrightarrow \mathbb{E}[X_1].$$

En fait ce théorème est vrai en supposant seulement que les X_i sont intégrables, mais la démonstration est plus difficile.

Lemme 7.9 (lemme de Borel–Cantelli). *Soit (A_n) une suite d'événements vérifiant*

$$\sum_{n \geq 1} \mathbb{P}(A_n) < +\infty,$$

alors presque sûrement l'événement A_n ne se produit que pour un nombre fini d'entiers n .

Démonstration. Il s'agit de montrer que l'événement

$$\begin{aligned} \limsup A_n &= \bigcap_{k \geq 1} \bigcup_{n \geq k} A_n \\ &= \llcorner A_n \text{ se produit pour une infinité de } n \end{aligned}$$

est négligeable. Comme

$$\bigcup_{n \geq k} A_n \searrow \limsup A_n$$

on a par convergence monotone

$$\mathbb{P}\left(\bigcup_{n \geq k} A_n\right) \searrow \mathbb{P}(\limsup A_n).$$

Mais comme $\sum \mathbb{P}(A_n)$ converge

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

tend vers 0. D'où le résultat. □

7.5 Théorème de la limite centrale

Revenons à la marche aléatoire. On a vu précédemment que quand n tend vers l'infini S_n/n tend vers la variable 0 presque sûrement. On va maintenant montrer que S_n/\sqrt{n} converge vers la Gaussienne $\mathcal{N}(0, 1)$, en un certain sens.

Théorème 7.10. *Soit $x < y$ des réels. Alors quand n tend vers l'infini*

$$\mathbb{P}\left(\frac{S_n}{\sqrt{n}} \in [x, y]\right) \longrightarrow \frac{1}{2\pi} \int_x^y e^{-t^2/2} dt.$$

On dit que S_n/\sqrt{n} converge en loi vers la Gaussienne $\mathcal{N}(0, 1)$.

Lemme 7.11. *Si $n \rightarrow +\infty$ et $|k| = o(n^{2/3})$ alors*

$$\mathbb{P}(S_{2n} = 2k) = \frac{1}{\pi n} e^{-k^2/n} (1 + o(1)).$$

Démonstration. Comme la loi de S_{2n} est symétrique on peut supposer que k est positif. Soit $k \in \{0, \dots, n\}$

$$\mathbb{P}(S_{2n} = 2k) = \binom{2n}{n+k} 2^{-2n} = \binom{2n}{n} 2^{-2n} \frac{\prod_{i=0}^{k-1} (n-i)}{\prod_{i=1}^k (n+i)} = \binom{2n}{n} 2^{-2n} \frac{\prod_{i=1}^{k-1} (1 - \frac{i}{n})}{\prod_{i=1}^k (1 + \frac{i}{n})}$$

En utilisant

$$\left(1 + \frac{i}{n}\right) = \exp\left(\frac{i}{n} + O\left(\frac{i^2}{n^2}\right)\right),$$

on obtient

$$\begin{aligned} \frac{\prod_{i=1}^{k-1} \left(1 - \frac{i}{n}\right)}{\prod_{i=1}^k \left(1 + \frac{i}{n}\right)} &= \exp\left(-2 \sum_{i=1}^{k-1} \frac{i}{n} - \frac{k}{n} + k O\left(\frac{k^2}{n^2}\right)\right) \\ &= e^{-k^2/n} \left(1 + O\left(\frac{k^3}{n^2}\right)\right). \end{aligned}$$

L'hypothèse assure que $k^3/n^2 \rightarrow 0$. De plus on a déjà vu que

$$\binom{2n}{n} 2^{-2n} = \frac{1}{\sqrt{\pi n}} (1 + o(1)).$$

D'où le résultat. \square

Démonstration du Théorème ??. Si $2k/\sqrt{2n} \in [x, y]$ alors $k = O(\sqrt{n})$ et le lemme précédent s'applique. Par conséquent

$$\begin{aligned} \mathbb{P}\left(\frac{S_{2n}}{\sqrt{2n}} \in [x, y]\right) &= \sum_{\sqrt{\frac{n}{2}}x \leq k \leq \sqrt{\frac{n}{2}}y} \mathbb{P}(S_{2n} = 2k) \\ &= \sum_{\sqrt{\frac{n}{2}}x \leq k \leq \sqrt{\frac{n}{2}}y} \frac{1}{\pi n} e^{k^2/n} (1 + o(1)). \end{aligned}$$

Appelons

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

la densité de la Gaussienne $\mathcal{N}(0, 1)$ et posons $a_k = \sqrt{\frac{2}{n}} k$. Quand k parcourt les entiers entre $\sqrt{\frac{n}{2}}x$ et $\sqrt{\frac{n}{2}}y$ la suite (a_k) parcourt une subdivision de l'intervalle $[x, y]$ de pas $\sqrt{\frac{2}{n}}$. Comme ϕ est une fonction continue, la somme de Riemann

$$\sum_{\sqrt{\frac{n}{2}}x \leq k \leq \sqrt{\frac{n}{2}}y} \frac{1}{\pi n} e^{k^2/n} = \sqrt{\frac{2}{n}} \sum_{\sqrt{\frac{n}{2}}x \leq k \leq \sqrt{\frac{n}{2}}y} \phi(a_k)$$

converge vers l'intégrale de ϕ sur $[x, y]$. On en déduit

$$\mathbb{P}\left(\frac{S_{2n}}{\sqrt{2n}} \in [x, y]\right) \rightarrow \int_x^y \phi(t) dt,$$

ce qui est le résultat cherché pour les entiers pairs. Les calculs sont similaires pour les entiers impairs. \square

8 Intégrales doubles

8.1 Théorème de Fubini

Dans ce chapitre on va intégrer des fonctions sur $\mathbb{R} \times \mathbb{R}$. Pour ce faire on doit d'abord définir une tribu.

Définition 8.1. On note $\mathcal{B}(\mathbb{R}^2)$ la plus petite tribu contenant

$$\{A \times B, A \in \mathcal{B}(\mathbb{R}), B \in \mathcal{B}(\mathbb{R})\}.$$

C'est la tribu des Boréliens de \mathbb{R}^2 .

Théorème 8.2. Soient μ et ν deux mesures sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ vérifiant $\mu(A \times B) = \nu(A \times B)$ pour tous $A, B \in \mathcal{B}(\mathbb{R})$. Alors $\mu = \nu$.

Théorème 8.3. Il existe une unique mesure \mathcal{L}^2 sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ vérifiant

$$\mathcal{L}^2(A \times B) = \mathcal{L}(A)\mathcal{L}(B),$$

pour tous Boréliens A, B de \mathbb{R} . Cette mesure est appelée mesure de Lebesgue sur \mathbb{R}^2 .

Encore une fois nous admettrons ces deux théorèmes, mais nous pouvons remarquer que l'unicité dans le deuxième théorème est une conséquence du premier. La mesure \mathcal{L}^2 est donc l'unique mesure sur \mathbb{R}^2 vérifiant

$$\mathcal{L}^2([a, b] \times [c, d]) = (b - a)(d - c),$$

pour tous réels a, b, c, d (vérifiant $a \leq b$ et $c \leq d$).

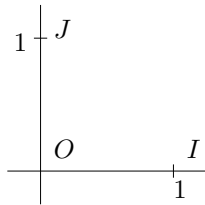
Théorème 8.4 (Théorème de Fubini). Soit $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction mesurable. L'égalité

$$\begin{aligned} \int_{\mathbb{R}^2} f d\mathcal{L}^2 &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dx \right) dy \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dy \right) dx, \end{aligned}$$

est vérifiée si f est positive, ou si f est intégrable pour la mesure \mathcal{L}^2 .

Démonstration. Par définition de la mesure \mathcal{L}^2 l'égalité est vraie si $f = \mathbf{1}_{A \times B}$ pour $A, B \in \mathcal{B}(\mathbb{R})$. En utilisant le Théorème 8.2 on en déduit que l'égalité est vraie pour $f = \mathbf{1}_C$ pour tout $C \in \mathcal{B}(\mathbb{R}^2)$. Ensuite il suffit de lancer la machine habituelle : le résultat est alors vrai pour les fonctions étagées par linéarité, puis pour les fonctions positives par convergence monotone, puis pour les fonctions intégrables. \square

Exemple 8.5. Considérons par exemple le triangle OIJ ci-dessous.



Bien sûr son aire est $1/2$. Mais on peut la calculer de la manière suivante : on remarque que (x, y) appartient au triangle si et seulement si $x \geq 0$, $y \geq 0$ et $x + y \leq 1$. L'aire du triangle vaut donc

$$\int_{\mathbb{R}^2} \mathbf{1}_{\{x \geq 0, y \geq 0, x+y \leq 1\}} dx dy = \int_0^1 \left(\int_0^{1-y} dx \right) dy = \int_0^1 (1 - y) dy = \frac{1}{2}.$$

8.2 Lois jointes

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, soient X et Y des variables aléatoires $\Omega \rightarrow \mathbb{R}$. Alors pour tout Borélien B de \mathbb{R}^2 l'ensemble $\{(X, Y) \in B\}$ appartient à \mathcal{A} .

Proposition 8.6. *On définit une probabilité sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ appelée loi jointe du couple (X, Y) en posant*

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

Si on connaît la loi jointe $\mathbb{P}_{X,Y}$ on peut retrouver les lois de X et Y . En effet si B est un Borélien de \mathbb{R} alors $B \times \mathbb{R}$ est un Borélien de \mathbb{R}^2 et on a

$$\mathbb{P}_{X,Y}(B \times \mathbb{R}) = \mathbb{P}(X \in B, Y \in \mathbb{R}) = \mathbb{P}(X \in B) = \mathbb{P}_X(B).$$

De même $\mathbb{P}_{X,Y}(\mathbb{R} \times B) = \mathbb{P}_Y(B)$. On dit que \mathbb{P}_X et \mathbb{P}_Y sont les lois marginales de $\mathbb{P}_{X,Y}$. En revanche la connaissance de \mathbb{P}_X et \mathbb{P}_Y ne détermine pas $\mathbb{P}_{X,Y}$ comme le montre l'exemple suivant.

Exemple 8.7. On lance deux dés. On appelle X et Y les résultats respectifs du premier et du deuxième dé. Clairement $\mathbb{P}_X = \mathbb{P}_Y$ ce qui montre que les couples $\mathbb{P}_{X,Y}$ et $\mathbb{P}_{X,X}$ ont les mêmes marginales. En revanche $\mathbb{P}_{X,Y}(1, 1) = \mathbb{P}(X = 1, Y = 1) = 1/36$ alors que $\mathbb{P}_{X,X}(1, 1) = \mathbb{P}(X = 1) = 1/6$.

Il y a donc strictement plus d'information dans $\mathbb{P}_{X,Y}$ que dans le couple $(\mathbb{P}_X, \mathbb{P}_Y)$. La connaissance de la loi jointe de (X, Y) permet de calculer l'espérance d'une fonction de (X, Y) .

Proposition 8.8. *Soient X et Y des variables aléatoires, et $\phi: \mathbb{R}^2 \rightarrow \mathbb{R}$ mesurable, on a*

$$\mathbb{E}\phi(X, Y) = \int_{\mathbb{R}^2} \phi(x, y) \mathbb{P}_{X,Y}(dx, dy),$$

dès que la variable $\phi(X, Y)$ est positive ou intégrable.

Démonstration. Il suffit de lancer la machine habituelle : indicatrices, étagées, positives, intégrables. \square

Si les variables X et Y sont discrètes, par exemple à valeurs dans \mathbb{N} le résultat précédent devient

$$\mathbb{E}\phi(X, Y) = \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \phi(i, j) \mathbb{P}(X = i, Y = j).$$

8.3 Couples continus

Définition 8.9. Une densité de probabilité sur \mathbb{R}^2 est une fonction $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ mesurable, positive et vérifiant

$$\int_{\mathbb{R}^2} f(x, y) dx dy = 1.$$

Définition 8.10. On dit qu'une mesure de probabilité π sur $\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)$ est absolument continue s'il existe une densité de probabilité $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ telle que

$$\pi(B) = \int_B f(x, y) dx dy$$

pour tout $B \in \mathcal{B}(\mathbb{R}^2)$. On dit alors que f est la densité de π .

Proposition 8.11. *Soit π une mesure sur \mathbb{R}^2 ayant une densité f . Alors*

$$\int_{\mathbb{R}^2} g d\pi = \int_{\mathbb{R}^2} g(x, y) f(x, y) dx dy$$

pour toute fonction $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ positive ou π -intégrable.

Démonstration. Machine habituelle. \square

Définition 8.12. On dit qu'un couple (X, Y) de variables aléatoires est continu si leur loi jointe $P_{X,Y}$ possède une densité f . On dit aussi que f est la densité de (X, Y) .

De manière informelle $f(x, y)dxdy$ représente la probabilité que le couple (X, Y) appartienne au rectangle $[x, x + dx] \times [y, y + dy]$.

Proposition 8.13. Soit (X, Y) un couple continu de densité $f_{X,Y}$. Alors X et Y sont des variables continues de densités respectives

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) dy$$

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y) dx.$$

On dit que f_X et f_Y sont les marginales de $f_{X,Y}$.

Démonstration. Soit $B \in \mathcal{B}(\mathbb{R})$. On applique Fubini

$$P(X \in B) = P((X, Y) \in B \times \mathbb{R}) = \int_{B \times \mathbb{R}} f_{X,Y}(x, y) dxdy = \int_B \left(\int_{\mathbb{R}} f_{X,Y}(x, y) dy \right) dx,$$

ce qui montre que $x \mapsto \int_{\mathbb{R}} f_{X,Y}(x, y) dy$ est une densité de X . □

Remarque. La réciproque n'est pas vraie : ce n'est pas parce que les variables X et Y sont continues que le couple (X, Y) est continu. Par exemple, le couple (X, X) ne peut pas être continu. En effet, s'il possédait une densité f , on aurait la contradiction suivante

$$1 = P(X = X) = \int_{\mathbb{R}^2} \mathbf{1}_{\{x=y\}} f(x, y) dxdy = 0.$$

Exemple 8.14. On considère les densités suivantes

$$f(x, y) = \mathbf{1}_{\{0 < x, 0 < y\}} e^{-x} y e^{-y}$$

$$, g(x, y) = \mathbf{1}_{\{0 < x < y\}} e^{-y}.$$

Les marginales de f sont $f_1(x) = \mathbf{1}_{\{0 < x\}} e^{-x}$ et $f_2(y) = \mathbf{1}_{\{0 < y\}} y e^{-y}$, c'est-à-dire respectivement la densité exponentielle de paramètre 1 et la densité $\Gamma(2, 1)$. Pour g on a

$$g_1(x) = \int_{\mathbb{R}} \mathbf{1}_{\{0 < x < y\}} e^{-y} dy = \mathbf{1}_{\{0 < x\}} e^{-x}$$

$$g_2(y) = \int_{\mathbb{R}} \mathbf{1}_{\{0 < x < y\}} e^{-y} dx = \mathbf{1}_{\{0 < y\}} y e^{-y}.$$

Donc f et g ont les mêmes marginales. Cet exemple montre que la connaissance des densités marginales ne détermine pas la densité jointe.

8.4 Retour sur l'indépendance

On a déjà vu que deux variables discrètes X, Y étaient indépendantes si et seulement si

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

pour tous réels x, y . On peut maintenant donner l'analogie continue de ce résultat.

Proposition 8.15. Soient X, Y deux variables continues de densités respectives f et g . Les variables X, Y sont indépendantes si et seulement si le couple (X, Y) est continu et admet comme densité

$$(x, y) \in \mathbb{R}^2 \mapsto f(x)g(y).$$

Démonstration. Supposons X et Y indépendantes. Soient A, B des Boréliens de \mathbb{R} , on a par Fubini

$$\begin{aligned} \mathbb{P}((X, Y) \in A \times B) &= \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \\ &= \left(\int_A f(x) dx \right) \left(\int_B g(y) dy \right) \\ &= \int_{A \times B} f(x)g(y) dx dy \end{aligned}$$

D'après le Théorème 8.2 ceci implique

$$\mathbb{P}((X, Y) \in C) = \int_C f(x)g(y) dx dy$$

pour tout Borélien C de \mathbb{R}^2 , ce qui signifie que

$$(x, y) \mapsto f(x)g(y)$$

est une densité de (X, Y) . La réciproque se démontre de manière analogue. \square

8.5 Changement de variable II

En dimension 2 la formule de changement de variable fait intervenir le Jacobien de ϕ .

Définition 8.16. Soit U un ouvert de \mathbb{R}^2 et $\phi: U \rightarrow \mathbb{R}^2$ une fonction de classe \mathcal{C}^1 . Posons $\phi = (\phi_1, \phi_2)$, rappelons que la différentielle de ϕ est la fonction matricielle

$$D\phi = \begin{pmatrix} \partial_x \phi_1 & \partial_y \phi_1 \\ \partial_x \phi_2 & \partial_y \phi_2 \end{pmatrix}.$$

On appelle Jacobien de ϕ le déterminant de cette matrice

$$J(\phi) = \partial_x \phi_1 \partial_y \phi_2 - \partial_x \phi_2 \partial_y \phi_1.$$

Théorème 8.17. Soit U, V des ouverts et soit $\phi: U \rightarrow V$ une fonction bijective de classe \mathcal{C}^1 . Alors

$$\int_{\phi(U)} h(s, t) ds dt = \int_U h \circ \phi(x, y) |J(\phi)(x, y)| dx dy,$$

pour toute fonction h (pourvu que la première intégrale soit bien définie).

Exemple 8.18 (Intégration en coordonnées polaires). La fonction $\phi(r, \theta) = (r \cos(\theta), r \sin(\theta))$ réalise une bijection entre $]0, +\infty[\times]0, 2\pi[$ et $\mathbb{R}^2 \setminus D$ en appelant D la demi-droite $\{(x, 0), x > 0\}$. On a

$$D(\phi)(r, \theta) = \begin{pmatrix} \cos(\theta) & -r \sin(\theta) \\ \sin(\theta) & r \cos(\theta) \end{pmatrix},$$

et donc $J(\phi)(r, \theta) = r \cos^2(\theta) + r \sin^2(\theta) = r$. La formule de changement de variable donne donc

$$\int_0^{+\infty} \int_0^{2\pi} f(r \cos(\theta), r \sin(\theta)) r dr d\theta = \int_{\mathbb{R}^2 \setminus D} f(x, y) dx dy = \int_{\mathbb{R}^2} f(x, y) dx dy.$$

pour toute fonction f (positive ou intégrable).

À titre d'exemple, appliquons cette formule à $f(x, y) = e^{-x^2/2 - y^2/2}$. Le membre de gauche devient

$$\int_0^{+\infty} \int_0^{2\pi} e^{-r^2/2} r dr d\theta = 2\pi [-e^{-r^2/2}]_0^{+\infty} = 2\pi.$$

Par Fubini le membre de droite vaut

$$\int_{\mathbb{R}^2} e^{-x^2/2 - y^2/2} dx dy = \left(\int_{\mathbb{R}} e^{-x^2/2} dx \right)^2.$$

Nous avons donc montré que $\int_{\mathbb{R}} e^{-x^2/2} dx = \sqrt{2\pi}$.

Rappelons qu'un couple (X, Y) de variables aléatoires possède une densité f si pour tout Borélien C de \mathbb{R}^2

$$\mathbf{P}((X, Y) \in C) = \int_C f(x, y) dx dy.$$

ou, de manière équivalente, si pour toute fonction $h: \mathbb{R}^2 \rightarrow \mathbb{R}_+$

$$\mathbf{E}h(X, Y) = \int_{\mathbb{R}^2} h(x, y) f(x, y) dx dy.$$

Rappelons qu'un \mathcal{C}^1 -difféomorphisme est une fonction bijective de classe \mathcal{C}^1 et telle que sa réciproque soit aussi \mathcal{C}^1 .

Proposition 8.19. Soit (X, Y) un couple continu à valeurs dans un ouvert U . Soit $\phi: U \rightarrow V$ un \mathcal{C}^1 -difféomorphisme. Alors le couple $\phi(X, Y)$ est continu.

Démonstration. Soit h une fonction positive (fonction test). En appliquant la formule de changement de variables à ϕ^{-1} on obtient

$$\begin{aligned} \mathbf{E}h \circ \phi(X, Y) &= \int_U h \circ \phi(x, y) f(x, y) dx dy \\ &= \int_V h(s, t) f \circ \phi^{-1}(s, t) |J(\phi^{-1})(s, t)| ds dt. \end{aligned}$$

Ce qui montre que $\phi(X, Y)$ admet comme densité la fonction

$$f \circ \phi^{-1} J(\phi^{-1}) \mathbf{1}_V. \quad \square$$

Exemple 8.20. Soit X et Y deux variables indépendantes de lois respectives $\Gamma(a, \lambda)$ et $\Gamma(b, \lambda)$. On pose $U = X + Y$ et $V = X/X + Y$. On s'intéresse à la loi de (U, V) . On remarque que

$$\begin{cases} u = x + y \\ v = x/(x + y) \\ x > 0, y > 0 \end{cases} \Leftrightarrow \begin{cases} x = uv \\ y = u(1 - v) \\ u > 0, 0 < v < 1 \end{cases}$$

On a montré que $\phi: (x, y) \mapsto (x + y, x/(x + y))$ réalise une bijection de $]0, +\infty[^2$ sur $]0, +\infty[\times]0, 1[$ et que $\phi^{-1}(u, v) = (uv, u(1 - v))$. On calcule le Jacobien $J(\phi^{-1})(u, v) = -u$. D'après la formule de changement de variables on obtient

$$\begin{aligned} \mathbf{E}h(U, V) &= \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} \int_{\mathbb{R}_+^2} h(x + y, x/(x + y)) x^{a-1} y^{b-1} e^{-\lambda x} e^{-\lambda y} dx dy \\ &= \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} \int_0^{+\infty} \int_0^1 h(u, v) (uv)^{a-1} (u(1 - v))^{b-1} e^{-\lambda u} u du dv. \end{aligned}$$

Ceci montre que (U, V) a pour densité

$$(u, v) \mapsto \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} \mathbf{1}_{\{u>0\}} u^{a+b-1} e^{-\lambda u} \mathbf{1}_{\{0<v<1\}} v^{a-1} (1 - v)^{b-1}.$$

On voit donc que U et V sont indépendants, que U suit la loi $\Gamma(a + b, \lambda)$ et que V admet la densité suivante

$$v \mapsto \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} v^{a-1} (1 - v)^{b-1} \mathbf{1}_{\{0<v<1\}}.$$

Au passage nous avons montré que la fonction précédente est une densité, et donc que

$$\int_0^1 v^{a-1} (1 - v)^{b-1} dv = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}.$$

Ce nombre est appelé $\beta(a, b)$. On dit aussi que V suit la loi $\beta(a, b)$. On peut calculer l'espérance de V :

$$\mathbf{E}(V) = \frac{1}{\beta(a, b)} \int_0^1 v v^{a-1} (1 - v)^{b-1} dv = \frac{\beta(a + 1, b)}{\beta(a, b)}.$$

En utilisant la propriété $\Gamma(x + 1) = x\Gamma(x)$ on obtient $\mathbf{E}(V) = a/(a + b)$.

Définition 8.21. Soit f et g des densités (sur \mathbb{R}). On appelle convolution de f et g la fonction

$$f * g(x) = \int_{\mathbb{R}} f(x-y)g(y) dy = \int_{\mathbb{R}} f(y)g(x-y) dy.$$

Une application simple de Fubini montre que $f * g$ est encore une densité. En fait on a le résultat suivant.

Proposition 8.22. Soit X et Y des variables indépendantes de densités respectives f et g , alors $X + Y$ admet $f * g$ comme densité.

Démonstration. La densité jointe du couple (X, Y) est la fonction $(x, y) \mapsto f(x)g(y)$. Soit $h: \mathbb{R} \rightarrow \mathbb{R}_+$ une fonction test. On a

$$\mathbb{E}h(X + Y) = \int_{\mathbb{R}^2} h(x + y)f(x)g(y) dx dy.$$

La fonction $\phi(x, y) = (x, x + y)$ réalise une bijection de \mathbb{R}^2 dans \mathbb{R}^2 . Sa réciproque est $\phi^{-1}(u, v) = (u, v - u)$ est de classe \mathcal{C}^1 et son Jacobien est constant (égal à 1). Par changement de variable et Fubini on a

$$\begin{aligned} \mathbb{E}h(X + Y) &= \int_{\mathbb{R}^2} h(x + y)f(x)g(y) dx dy \\ &= \int_{\mathbb{R}^2} h(v)f(u)g(v - u) dudv = \int_{\mathbb{R}} h(v) \left(\int_{\mathbb{R}} f(u)g(v - u) du \right) dv. \end{aligned}$$

On en déduit que $X + Y$ admet $f * g$ comme densité. □

Exemple 8.23. Soient X et Y des lois $\mathcal{N}(0, 1)$ indépendantes. On calcule la densité de $X + Y$

$$\frac{1}{2\pi} \int_{\mathbb{R}} e^{-y^2/2} e^{-(x-y)^2/2} dy = \frac{1}{2\pi} \left(\int_{\mathbb{R}} e^{-(y-x/2)^2} dy \right) e^{-x^2/4} = \frac{e^{-x^2/4}}{2\sqrt{\pi}}.$$

Ceci montre que la somme de deux $\mathcal{N}(0, 1)$ indépendantes est une $\mathcal{N}(0, 2)$.

8.6 Exercices

Exercice 8.1. On lance une infinité de fois une pièce truquée tombant sur « pile » avec probabilité p (paramètre compris entre 0 et 1 strictement). On appelle *série* une succession de piles (ou de faces) interrompue par le résultat contraire. Par exemple, pour l'événement $PPFFFFPFFF\dots$, la première série est une série de piles et a pour longueur 2, la deuxième a pour longueur 3, etc... Soit X la longueur de la première série et Y la longueur de la deuxième série.

1. Déterminer la loi de X .
2. Déterminer la loi du couple (X, Y) . En déduire la loi de Y .

Exercice 8.2. Soit n un entier non nul. On dispose d'une urne contenant 2 boules blanches et n noires. On retire les boules une par une (jusqu'à ce que l'urne soit vide) en notant X le rang de tirage de la première boule blanche et Y le rang de la deuxième.

1. Déterminer la loi du couple (X, Y) .
2. En déduire la loi de Y .
3. Montrer que

$$\mathbb{E}(Y) = \frac{2(n+3)}{3}.$$

Exercice 8.3. Soit (X, Y) un couple aléatoire à valeurs dans $\{0, 1\} \times \mathbb{N}^*$, dont la loi est définie par :

$$\mathbb{P}(X = 0, Y = k) = \frac{2^k - 1}{4^k}, \quad \mathbb{P}(X = 1, Y = k) = \frac{1}{4^k}, \quad \forall k \in \mathbb{N}^*.$$

1. Déterminer les lois de X et Y , donner leurs espérances et leurs variances.

2. On pose $S = X + Y$ et $T = XY + 1$. Déterminer les lois de S et de T .
3. Trouver la loi du couple (S, T) .
4. Calculer $\mathbb{P}(S = T)$.

Exercice 8.4. Soit (X, Y) un couple de loi continue uniforme sur $[0, 1] \times [0, 1]$, c'est-à-dire de densité $f(x, y) = 1$ si $(x, y) \in [0, 1] \times [0, 1]$ et $f(x, y) = 0$ sinon. On considère les variables aléatoires suivantes :

$$U = \min(X, Y) \quad \text{et} \quad V = \max(X, Y).$$

1. Déterminer les fonctions de répartitions des variables U et V .
2. En déduire que U et V sont des variables continues et déterminer leurs densités.
3. Calculer l'espérance et la variance de U et V .
4. Calculer la covariance de U, V .

Exercice 8.5. Soit (X, Y) un couple continu de densité $f(x, y) = \frac{1}{2x} \mathbf{1}_D(x, y)$ avec

$$D = \{(x, y) \in \mathbb{R}^2 : 0 < y < x, 0 < y < 1/x\}.$$

On considère les variables aléatoires $U = \sqrt{XY}$ et $V = \sqrt{Y/X}$.

1. Montrer que U et V sont i.i.d. de loi uniforme sur $[0, 1]$.
2. En déduire $E(Y)$.

Exercice 8.6. Soit (X, Y) un couple de variables aléatoires admettant pour densité

$$f(x, y) = \begin{cases} 4xe^{-(x+y)} & \text{si } 0 < x < y, \\ 0 & \text{sinon} \end{cases}$$

1. Déterminer les densités des marginales de X et Y .
2. Calculer $\text{cov}(X, Y)$.

Exercice 8.7. Soit (X, Y) un couple de densité

$$f(x, y) = (x + y) \mathbf{1}_{[0,1] \times [0,1]}(x, y).$$

1. Vérifier que f est bien une densité et calculer les densités marginales de X et Y .
2. Calculer $\mathbb{P}(X + Y \leq 1)$.
3. Déterminer une densité de $U = \max(X, Y)$.

Exercice 8.8. Soient X, Y i.i.d. de loi $\mathcal{N}(0, 1)$. Soit M le point de coordonnées (X, Y) dans un repère orthonormé (O, i, j) . Soit R la distance de M à O et soit θ l'angle (i, \overrightarrow{OM}) .

1. Montrer que R et θ sont indépendants.
2. Déterminer la loi de R^2 .
3. Déterminer la loi de $\tan(\theta)$.

Exercice 8.9. Soit M un point aléatoire uniforme sur le disque D de centre $O = (0, 0)$ et de rayon 1. Quelle est la distance moyenne de M à O ?

9 Conditionnement

On a vu que que la connaissance des lois de X et Y ne suffisait pas à déterminer la loi du couple (X, Y) . Pour connaître la loi de (X, Y) il faut connaître la loi de Y et la loi de X *conditionnellement* à Y .

9.1 Cas discret

Définition 9.1. Soit (X, Y) un couple de variable aléatoires et on suppose que Y est discrète. Pour $y \in \mathbb{R}$ vérifiant $P(Y = y) > 0$ on définit la loi conditionnelle de X sachant $Y = y$ ainsi

$$P_{X|Y=y}(A) = P(X \in A | Y = y).$$

Si on se donne un événement A et une variable aléatoire discrète Y , la quantité $u(y) = P(A | Y = y)$ est définie pour sur l'ensemble y_1, y_2, \dots des valeurs prises par Y . Par conséquent la variable aléatoire $u(Y)$ est bien définie. On notera $P(A | Y)$ cette variable. Il faut bien retenir que $P(A | Y = y)$ est un nombre réel, tandis que $P(A | Y)$ est une variable aléatoire.

Exemple 9.2. Une poule pond un nombre aléatoire N d'œufs qui suit une loi de Poisson de paramètre λ . Les œufs éclosent avec probabilité p indépendamment les uns des autres. Quelle est la loi du nombre X de poussins ?

Conditionnellement à $N = n$ le nombre X suit une loi de binomiale de paramètres n et p . C'est-à-dire

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

Par conséquent

$$P(X = k, N = n) = P(X = k | N = n)P(N = n) = \binom{n}{k} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!}.$$

Puis

$$P(X = k) = \sum_{n=k}^{+\infty} P(X = k, N = n) = \sum_{n=k}^{+\infty} \frac{e^{-\lambda} \lambda^n p^k (1-p)^{n-k}}{k!(n-k)!}.$$

En faisant le changement d'indice $m = n - k$ on trouve

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

Donc X suit une loi de Poisson de paramètre λp .

Rappelons que pour X positive ou intégrable on a l'égalité suivante

$$E(X) = \int_{\mathbb{R}} x P_X(dx).$$

Ceci motive la définition suivante.

Définition 9.3. On suppose toujours Y discrète et on se donne y tel que $P(Y = y) > 0$. L'espérance conditionnelle de X sachant $Y = y$ est définie par

$$E(X | Y = y) = \int_{\mathbb{R}} x P_{X|Y=y}(dx).$$

Cette quantité est définie si X est positive, ou si $x \mapsto x$ est intégrable pour $P_{X|Y=y}$.

Proposition 9.4. L'espérance conditionnelle vérifie les mêmes propriétés que l'espérance.

- *Linéarité* : $E(a_1 X_1 + a_2 X_2 | Y = y) = a_1 E(X_1 | Y = y) + a_2 E(X_2 | Y = y)$.
- *Croissance* : si $X_1 \leq X_2$ alors $E(X_1 | Y = y) \leq E(X_2 | Y = y)$.

On a aussi la formule suivante.

Proposition 9.5. Pour toute fonction f

$$\mathbb{E}(f(X) \mid Y = y) = \int_{\mathbb{R}} f(x) \mathbb{P}_{X|Y=y}(dx).$$

Si la fonction $u: y \mapsto \mathbb{E}(X \mid Y = y)$ est bien définie pour tout y tel que $\mathbb{P}(Y = y) > 0$, alors la variable $u(Y)$ est bien définie. Cette variable est notée $\mathbb{E}(X \mid Y)$. Encore une fois il faut retenir que $\mathbb{E}(X \mid Y = y)$ est un nombre tandis que $\mathbb{E}(X \mid Y)$ est une variable aléatoire. On a alors résultat suivant.

Proposition 9.6.

$$\mathbb{E}(\mathbb{E}(X \mid Y)) = \mathbb{E}(X).$$

Démonstration. La variable Y est discrète. Supposons par exemple qu'elle est à valeurs dans \mathbb{N} . Il s'agit de montrer que

$$\sum_{n=0}^{+\infty} \left(\int_{\mathbb{R}} x \mathbb{P}_{X|Y=n}(dx) \right) \mathbb{P}(Y = n) = \mathbb{E}(X).$$

On va en fait montrer que

$$\sum_{n=0}^{+\infty} \left(\int_{\mathbb{R}} f(x) \mathbb{P}_{X|Y=n}(dx) \right) \mathbb{P}(Y = n) = \mathbb{E}(f(X)).$$

pour toute fonction f (telle que $f(X)$ soit bien définie). Si f est une indicatrice le résultat est vrai puisque

$$\begin{aligned} \sum_{n=0}^{+\infty} \left(\int_{\mathbb{R}} \mathbf{1}_B(x) \mathbb{P}_{X|Y=n}(dx) \right) \mathbb{P}(Y = n) &= \sum_{n=0}^{+\infty} \mathbb{P}_{X|Y=n}(B) \mathbb{P}(Y = n) \\ &= \sum_{n=0}^{+\infty} \mathbb{P}(X \in B \mid Y = n) \mathbb{P}(Y = n) \\ &= \mathbb{P}(X \in B) = \mathbb{E} \mathbf{1}_B(X). \end{aligned}$$

On obtient ensuite le résultat pour les fonctions étagées par linéarité, puis pour les fonctions positives par convergence monotone, puis pour les fonctions intégrables en écrivant $f = f_+ - f_-$. \square

Exemple 9.7. Si N suit une loi de Poisson de paramètre λ et si X sachant N est binomiale (N, p) , alors on vu que X suit une loi de Poisson de paramètre λp . En particulier $\mathbb{E}(X) = \lambda p$. On peut retrouver ce résultat en utilisant la proposition précédente : sachant $N = n$ la variable X est binomiale (n, p) donc

$$\mathbb{E}(X \mid N = n) = np.$$

Par conséquent $\mathbb{E}(X \mid N) = Np$. Et donc

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X \mid N)) = \mathbb{E}(Np) = \mathbb{E}(N)p = \lambda p.$$

On peut aussi retrouver la loi de X de la manière suivante : comme X sachant $N = n$ est binomiale n, p , on a

$$\mathbb{E}(e^{tX} \mid N = n) = (pe^t + 1 - p)^n.$$

Donc $\mathbb{E}(e^{tX} \mid N) = (pe^t + 1 - p)^N$. Donc

$$\mathbb{E}(e^{tX}) = \mathbb{E}[\mathbb{E}(e^{tX} \mid N)] = \mathbb{E}[(pe^t + 1 - p)^N] = \exp[\lambda(pe^t + 1 - p - 1)] = \exp[\lambda p(e^t - 1)]$$

On reconnaît la fonction génératrice des moments d'une variable de Poisson de paramètre λp .

9.2 Cas continu

Définition 9.8. Soit X, Y un couple continu, de densité $f_{X,Y}$. On appelle densité conditionnelle de X sachant $Y = y$ la fonction

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)},$$

définie pour tout y telle que $f_Y(y) > 0$.

Définition 9.9. On appelle espérance conditionnelle de X sachant $Y = y$ la quantité

$$\mathbf{E}(X | Y = y) = \int_{\mathbb{R}} x f_{X|Y=y}(x) dx$$

également définie pour les y tels que $f_Y(y) > 0$.

Exemple 9.10. Soit (X, Y) ayant pour densité la fonction $f_{X,Y}(x, y) = (1/x)\mathbf{1}_{\{0 < y < x < 1\}}$. On en déduit que X a pour densité

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) dy = \frac{1}{x} \left(\int_0^x dy \right) \mathbf{1}_{\{0 < x < 1\}} = \mathbf{1}_{\{0 < x < 1\}}.$$

Autrement dit X suit une loi uniforme sur $[0, 1]$. La densité conditionnelle de Y sachant X est donnée par

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{\mathbf{1}_{\{0 < y < x\}}}{x}$$

pour tout $y \in \mathbb{R}$ et tout $x \in]0, 1[$. Autrement dit, conditionnellement à X , la variable Y est uniforme sur $[0, X]$. En particulier $\mathbf{E}(Y | X) = X/2$.

Les propriétés de linéarité et de croissance sont toujours vraies. On a aussi la formule

$$\mathbf{E}(g(X) | Y = y) = \int_{\mathbb{R}} g(x) f_{X|Y=y}(x) dx.$$

pour toute fonction g . La fonction $u: y \mapsto \mathbf{E}(X | Y = y)$ est définie pour les y tels que $f_Y(y) > 0$. Comme $f_Y(Y) > 0$ presque sûrement on peut poser

$$\mathbf{E}(X | Y) = u(Y).$$

Encore une fois $\mathbf{E}(X | Y = y)$ est un nombre tandis que $\mathbf{E}(X | Y)$ est une variable aléatoire. On a toujours l'égalité fondamentale

$$\mathbf{E}(\mathbf{E}(X | Y)) = \mathbf{E}(X).$$

En effet par Fubini

$$\begin{aligned} \mathbf{E}(\mathbf{E}(X | Y)) &= \int_{\mathbb{R}} \mathbf{E}(X | Y = y) f_Y(y) dy \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} x f_{X|Y=y}(x) dx \right) f_Y(y) dy \\ &= \int_{\mathbb{R}^2} x f_{X,Y}(x, y) dx dy = \mathbf{E}(X). \end{aligned}$$

Jusqu'à maintenant on a défini la loi conditionnelle de X sachant Y dans deux cas :

- La variable Y est discrète
- Le couple (X, Y) est continu.

En fait on peut aller au-delà de ces deux cas. On ne donnera pas de définition générale mais seulement un exemple.

Exemple 9.11. Soit Y de loi uniforme sur $[0, 1]$ et soit X suivant, conditionnellement à Y , une loi binomiale de paramètres (n, Y) . On cherche la loi de X .
On sait que pour $k \in \{0, \dots, n\}$

$$P(Y = k | X) = \binom{n}{k} X^k (1 - X)^{n-k}.$$

Par conséquent

$$P(Y = k) = EP(Y = k | X) = \binom{n}{k} E[X^k (1 - X)^{n-k}] = \binom{n}{k} \int_0^1 x^k (1 - x)^{n-k} dx.$$

De plus

$$\int_0^1 x^k (1 - x)^{n-k} dx = \beta(k + 1, n - k + 1) = \frac{k!(n - k)!}{(n + 1)!}.$$

On en déduit que $P(X = k) = 1/(n + 1)$ pour tout $k \in \{0, \dots, n\}$. Autrement dit X est uniforme sur $\{0, \dots, n\}$.

9.3 Exercices

Exercice 9.1. On tire un nombre N suivant une loi de Poisson et on lance la même pièce N fois successivement. Montrer que le nombre de piles obtenus est indépendant du nombre de faces.

Exercice 9.2. On effectue une suite infinie de lancers de dés. Pour $n \in \mathbb{N}^*$ on note X_n le temps d'attente de la première série de n six consécutifs : X_1 est le temps auquel le six sort pour la première fois, X_2 le temps auquel deux six d'affilée sortent pour la première fois et ainsi de suite. Par exemple, pour la suite 23644155366613... les variables X_1, X_2, X_3 valent respectivement 3, 11, 12.

1. Déterminer EX_1 .
2. Soit $n \geq 2$, montrer que

$$E(X_n | X_{n-1}) = \frac{1}{6}(X_{n-1} + 1) + \frac{5}{6}(X_{n-1} + 1 + E(X_n)).$$

3. En déduire une relation de récurrence pour $E(X_n)$.
4. Déterminer $E(X_n)$.

Exercice 9.3. Soit (X, Y) un couple de densité

$$f(x, y) = e^{-y} \mathbf{1}_{\{0 < x < y\}}.$$

1. Déterminer les lois de X et de Y .
2. Déterminer les lois conditionnelles de X sachant Y et de Y sachant X .
3. Calculer les espérances conditionnelles $E(X | Y)$ et $E(Y | X)$.
4. Vérifier que $E(E(X | Y)) = E(X)$ et $E(E(Y | X)) = E(Y)$.

Exercice 9.4. Soit X, Y des variables i.i.d. de loi exponentielle. Déterminer la loi conditionnelle de X sachant $X + Y$.