PROBABILITÉS

Raphaël Alexandre

2 avril 2015

Sommaire

Chapitre 1

Notions fondamentales

1 Espace de probabilité

Intuition. Exemple d'affirmation :

« La probabilité d'obtenir 7 lorsqu'on lance deux dés est 1/6 . »

Le lancer des deux dés est l'expérience aléatoire dont le résultat n'est pas prévisible. L'obtention du 7 est un événement qui se produit ou non selon le résultat de l'expérience. Enfin, 1/6 quantifie la vraisemblance de l'événement.

1.1 Ensemble fondamental

> Définition 1.1
> L'ensemble fondamental est l'ensemble de tous les résultats possibles d'une expérience aléatoire.

Notation. On le note généralement Ω. Un élément de Ω est souvent noté ω.

Exemples.

1. Le lancer d'une pièce :
$$\Omega = \{P, F\}\,.$$
2. Le lancer d'un dé :
$$\Omega = \{1, 2, 3, 4, 5, 6\}\,.$$
3. Battage de cartes (d'un jeu de n cartes numérotées de 1 à n), si on note $\omega(i)$ la place de la carte i après battage alors l'application $\omega : \{1, \ldots, n\} \to \{1, \ldots, n\}$ est une bijection (i.e. permutation) qui caractérise le résultat du battage. Finalement :
$$\Omega = S_n.$$
4. Le lancer d'une pièce jusqu'à obtention de la face P,
$$\omega_1 = P,\ \omega_2 = FP,\ \ldots,\ \omega_k = \underbrace{F\ldots}_{k-1}FP,\ \ldots,\ \omega_\infty = F\ldots\underbrace{F\ldots}_{\infty}$$
Ici,
$$\Omega = \{\omega_1, \omega_2, \ldots\} \cup \{\omega_\infty\}\,.$$
5. Durée de vie d'une ampoule,
$$\Omega = \mathbf{R}_+ \cup \{+\infty\}\,.$$

1.2 Évenement

INTUITION. Événement qui arrive ou pas selon le résultat de l'expérience.

1. A l'événement « le résultat est pair ».
2. B_i l'événement « la i-ième carte est à sa place », C : « aucune carte n'est à sa place » .
3. D : « l'ampoule dure au moins un an ».

À chaque événement, A, on peut associer un sous-ensemble de Ω :

$$\{\omega \in \Omega \,|\, A \text{ se produit si } \omega \text{ est le résultat de l'expériece}\}\,.$$

EXEMPLES. Dans l'ordre des exemples précédents :

1. $A = \{2, 4, 6\}$;
2. $B_i = \{\omega \in \Omega \,|\, \omega(i) = i\}$;
   $C = \{\omega \in \Omega \,|\, \forall i \in \{1, \ldots, n\}\,,\ \omega(i) \neq i\} = \bigcap_{i=1}^{n} \{\omega \in \Omega \,|\, \omega(i) \neq i\} = \bigcap_{i=1}^{n} B_i^c$.

Dorénavant, on identifie un événement avec le sous-ensemble de Ω qui lui correspond.

Vocabulaire des événements	Vocabulaire de la théorie des ensembles
A l'événement	A le sous-ensemble de Ω
$\emptyset$ l'événement impossible	$\emptyset$ l'ensemble vide
Ω l'événement certain	Ω l'ensemble fondamental
$A \cup B$ l'événement « A ou B »	$A \cup B$ l'union de deux ensembles
$A \cap B$ l'événement « A et B »	$A \cap B$ l'intersection de deux sous-ensembles A et B
A^c l'événement contraire de A	A^c le complémentaire d'un sous-ensemble A

On dira de manière équivalente :

— $A \cap B = \emptyset$;
— A et B sont disjoints ;
— A et B sont incompatibles.

DÉFINITION 1.2
Une partition de Ω est une famille d'événements $(A_i)_{i \in I}$ telle que

$$\bigcup_{i \in I} A_i = \Omega$$

et les A_i sont deux à deux disjoints.

1.3 Tribu

On pourrait le considérer comme l'ensemble de tous les événements, $\mathcal{P}(\Omega)$. On le fera lorsque Ω est « petit » (au plus dénombrable). Ça devient pénible lorsque Ω est infini non-dénombrable.

DÉFINITION 1.3
Soit Ω un ensemble fondamental supposé non vide. Soit $\mathcal{A} \subset \mathcal{P}(\Omega)$. On dit que $\mathcal{A}$ est une tribu si :

1. $\Omega \in \mathcal{A}$;
2. pour tout $A \in \mathcal{A}$, $A^c \in \mathcal{A}$;
3. si I est au plus dénombrable et $(A_i)_{i \in I}$ une famille à valeurs dans $\mathcal{A}$ alors

$$\bigcup_{i \in I} A_i \in \mathcal{A}.$$

Exemples. Soit Ω un ensemble non vide.

1. $\{\emptyset, \Omega\}$ est une tribu ;
2. si $A \in \mathcal{P}(\Omega)$ alors $\{\emptyset, \Omega, A, A^c\}$ est une tribu ;
3. $\mathcal{P}(\Omega)$ est une tribu (pour Ω au plus dénombrable, c'est la « tribu naturelle »).
4. pour $\Omega = \mathbf{R}$ on prend $\mathcal{B}(\mathbf{R})$ qui est la tribu des boréliens, c'est-à-dire la plus petit tribu qui contient les intervalles de $\mathbf{R}$ (existence admise).

1.4 Probabilités

Définition 1.4
Soit Ω un ensemble non vide et soit $\mathcal{A}$ une tribu sur Ω. Une probabilité sur $(\Omega, \mathcal{A})$ est une fonction :

$$\mathbf{P} : \mathcal{A} \to [0,1]$$

vérifiant :

1. $\mathbf{P}(\Omega) = 1$;
2. si I est au plus dénombrable, $(A_i)_{i \in I}$ une famille à valeurs dans $\mathcal{A}$ avec les A_i disjoints deux à deux alors

$$\mathbf{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mathbf{P}(A_i).$$

Le second axiome s'appelle « σ-additivité ».

Définition 1.5
On appelle $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité.

Exemple. Pour le lancer de pièce,

$$\Omega = \{P, F\}\,,\ \mathcal{A} = \{\emptyset, \{P\}, \{F\}, \{P, F\}\}\,,\ \mathbf{P}(\{F\}) = \mathbf{P}(\{P\}) = 1/2, \mathbf{P}(\{P, F\}) = 1, \mathbf{P}(\emptyset) = 0.$$

Proposition 1.6
Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité.

1. Si $A \in \mathcal{A}$ alors $\mathbf{P}(A) + \mathbf{P}(A^c) = 1$. En particulier, $\mathbf{P}(\emptyset) = 0$.
2. Si $A, B \in \mathcal{A}$ avec $A \subset B$ alors $\mathbf{P}(A) \leq \mathbf{P}(B)$.
3. Si $A, B \in \mathcal{A}$ alors $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.
4. Si $(A_i)_{i \in \{1, \ldots, n\}}$ est une famille à valeurs dans $\mathcal{A}$ alors

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbf{P}(A_i).$$

5. Si $(A_i)_{i \in \mathbf{N}}$ est une famille à valeurs dans $\mathcal{A}$ alors

$$\mathbf{P}\left(\bigcup_{i \in \mathbf{N}} A_i\right) \leq \sum_{i \in \mathbf{N}} \mathbf{P}(A_i).$$

Démonstration

Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité.

1. Soit $A \in \mathcal{A}$, on a
$$\mathbf{P}(\Omega) = \mathbf{P}(A \cup A^c) = \mathbf{P}(A) + \mathbf{P}(A^c) = 1.$$
2. Soient $A, B \in \mathcal{A}$, on a
$$\mathbf{P}(B) = \mathbf{P}(A \cup B \setminus A) = \mathbf{P}(A) + \mathbf{P}(B \setminus A) \geq \mathbf{P}(A).$$
3. Soient $A, B \in \mathcal{A}$,
$$\mathbf{P}(A \cup B) = \mathbf{P}(A \setminus B) + \mathbf{P}(B \setminus A) + \mathbf{P}(B \cap A) = \mathbf{P}((A \setminus B) \cup (B \cap A)) + \mathbf{P}(B \setminus A) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B).$$
4. Par récurrence sur n :
$$\mathbf{P}\left(\bigcup_{i=1}^{n+1} A_i\right) \leq \mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) + \mathbf{P}(A_{n+1}) \leq \sum_{i=1}^{n+1} \mathbf{P}(A_i).$$
5. Posons
$$A = \bigcup_{i \in \mathbf{N}} A_i \subset \mathcal{A}.$$
Soit $\omega \in A$, on note $\nu(\omega)$ le premier indice i tel que $w \in A_i$. On pose
$$B_n = \{w \in \Omega \,|\, \nu(\omega) = n\} = A_n \bigcap \left(\bigcup_{i=0}^{n-1} A_i\right)^c.$$
On a :
   — $$\bigcup_{n \in \mathbf{N}} B_n = A\ ;$$
   — les B_n sont deux à deux disjoints ;
   — $B_n \subset A_n$.

   On a donc
$$\mathbf{P}\left(\bigcup_{n \in \mathbf{N}} B_n\right) = \sum_{n \in \mathbf{N}} \mathbf{P}(B_n).$$
Enfin
$$\mathbf{P}(A) = \sum_{n \in \mathbf{N}} \mathbf{P}(B_n) \leq \sum_{n \in \mathbf{N}} \mathbf{P}(A_n).$$

Définition 1.7

Soit $(A_n)_{n \in \mathbf{N}}$ une suite croissante d'événements. On définit :
$$\lim_{n \to \infty} A_n = \bigcup_{n \in \mathbf{N}} A_n.$$

Proposition 1.8

Soit $(A_n)_{n \in \mathbf{N}}$ une suite croissante d'événements. On a :

1. $\lim_{n \to \infty} A_n \in \mathcal{A}$;
2. $\mathbf{P}(\lim A_n) = \lim \mathbf{P}(A_n)$.

Démonstration

Soit $(B_n)_{n \in \mathbf{N}}$ définie par :
$$B_0 = A_0,\ B_n = A_n \cap A_{n-1}^c.$$

On a
$$\bigcup_{m=0}^{n} B_m = A_n$$
et les B_n sont deux-à-deux disjoints. Or
$$\bigcup_{n\in\mathbf{N}} A_n = \bigcup_{n\in\mathbf{N}} B_n.$$
En effet, $B_n \subset A_n$ et donc $B_n \subset \bigcup_{n\in\mathbf{N}} A_n$ et donc $\bigcup_{n\in\mathbf{N}} B_n \subset \bigcup_{n\in\mathbf{N}} A_n$; de plus $A_n = \bigcup_{m=0}^{n} B_m \subset \bigcup_{m\in\mathbf{N}} B_m$ et donc $\bigcup_{n\in\mathbf{N}} A_n \subset \bigcup_{n\in\mathbf{N}} B_n$.
Ainsi, $\lim A_n \in \mathcal{A}$ et
$$\begin{aligned}\mathbf{P}(\lim A_n) &= \mathbf{P}\left(\bigcup_{n\in\mathbf{N}} B_n\right)\\ &= \sum_{n=0}^{\infty} \mathbf{P}(B_n)\\ &= \lim_{n\to\infty} \sum_{m=0}^{n} \mathbf{P}(B_m)\\ &= \lim \mathbf{P}\left(\bigcup_{m=0}^{n} B_m\right)\\ &= \lim \mathbf{P}(A_n).\end{aligned}$$

Définition 1.9
Soit $(A_n)_{n\in\mathbf{N}}$ une suite décroissante d'événements. On définit :
$$\lim_{n\to\infty} A_n = \bigcap_{n\in\mathbf{N}} A_n.$$

Proposition 1.10
Soit $(A_n)_{n\in\mathbf{N}}$ une suite décroissante d'événements. On a :
1. $\lim_{n\to\infty} A_n \in \mathcal{A}$;
2. $\mathbf{P}(\lim A_n) = \lim \mathbf{P}(A_n)$.

Démonstration
Par passage au complémentaire : en posant $B_n = A_n^c$ on a une famille croissante d'événements. Ainsi
$$1 - \mathbf{P}\left(\bigcap_{n\in\mathbf{N}} A_n\right) = \mathbf{P}\left(\left(\bigcap_{n\in\mathbf{N}} A_n\right)^c\right) = \mathbf{P}\left(\bigcup_{n\to\infty} B_n\right) = \lim_{n\to infty} \mathbf{P}(B_n) = \lim_{n\to\infty} 1 - \mathbf{P}(A_n).$$

1.5 Probabilités uniformes sur un ensemble fini

Soit Ω un ensemble fini non vide. On prend $\mathcal{A} = \mathcal{P}(\Omega)$ et
$$\mathbf{P} \colon \begin{cases} \mathcal{A} \to [0,1] \\ A \mapsto \dfrac{\operatorname{card} A}{\operatorname{card} \Omega} \end{cases}.$$

$\mathbf{P}$ ainsi défini est bien une probabilité. C'est la probabilité uniforme sur $(\Omega, \mathcal{P}(\Omega))$. On la dit « uniforme » car tous les événements élémentaires sont équiprobables. En effet, pour tout $\omega \in \Omega$, $\mathbf{P}(\{\omega\}) = 1/\operatorname{card}\Omega$ est égal à une constante indépendante de ω.

Dans ce cadre, le calcul des probabilités se réduit à du dénombrement. Quelques résultats :

Proposition 1.11
On a :

1. soient E et F deux ensembles finis s'il existe une bijection entre E et F, alors $\operatorname{card} E = \operatorname{card} F$;
2. si $\operatorname{card} E = n$ alors $\operatorname{card} \mathcal{P}(E) = 2^n$;
3. si $\operatorname{card} E = \operatorname{card} F = n$ alors il y a $n!$ bijections entre E et F (si $E = F$ on dit que de telles bijections sont des permutations) ;
4. il y a
$$A_n^k = \frac{n!}{(n-k)!}$$
injections d'un ensemble à k éléments dans un ensemble à n éléments ($0 \leq k \leq n$) ; A_n^k désigne également le nombre d'arrangements à k éléments dans un ensemble à n éléments ;
5. le nombre de sous-ensembles à k éléments dans un ensemble à n éléments est $\binom{n}{k}$;
6. formule du binôme :
$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$
et en particulier :
$$2^n = \sum_{k=0}^{n} \binom{n}{k}.$$

Exemples.

1. On effectue un tirage de n pièces. On a $\Omega = \{P, F\}^n$, $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathbf{P}$ est la probabilité uniforme, $\operatorname{card}\Omega = 2^n$ pour A un événement quelconque et $\operatorname{card}\mathcal{A} = 2^{2^n}$. Si A est l'événement « obtenir k fois pile exactement » alors
$$\mathbf{P}(A) = \frac{\operatorname{card} A}{\operatorname{card}\Omega} = \frac{\binom{n}{k}}{2^n}.$$
2. Il y a n personnes dans une salle et on regarde l'événement « il y a au moins deux personnes nées le même jour ». On a
$$\Omega = \{1, \ldots, 365\}^n,$$
$\mathcal{A} = \mathcal{P}(\Omega)$ et $\mathbf{P}$ est toujours la probabilité uniforme. On regarde l'événement complémentaire qui consiste à l'ensemble des cas où toutes les personnes sont nées à des jours différents. Or le cardinal de A^c est $A_{365}^n = 365!/(365-n)!$. Ainsi
$$\mathbf{P}(A) = 1 - \frac{A_{365}^n}{365^n}.$$

1.6 Rappels sur les séries

Soit $(a_n)_{n\in\mathbf{N}}$ une séries à valeurs réelles ou complexes.

DÉFINITION 1.12
$\sum a_n$ est convergetnte si $b_n = \sum_{k=0}^{n} a_k$ est une suite convergente. La valeur de la somme de la série est alors $\lim_{n\to\infty} b_n$.

DÉFINITION 1.13
$\sum a_n$ est absolument convergente si $\sum |a_n|$ est convergente.

PROPOSITION 1.14
On a :

1. Une série absolument convergente est convergente.
2. Si $\sum a_n$ est absolument convergente alors pour toute bijection de $\mathbf{N}$ la série $\sum a_{\sigma(n)}$ est absolument convergente et la valeur de la série ne dépend pas de σ.
3. Si (a_n) est à termes positifs alors la valeur de la série $\sum a_{\sigma(n)}$ (éventuellement infinie) est indépendante de σ.

1.7 Probabilités sur un ensemble au plus dénombrable

Soit Ω un ensemble non vide au plus dénombrable. Une probabilité $\mathbf{P}$ sur $(\Omega, \mathcal{P}(\Omega))$ est entièrement caractérisée par les valeurs de $\mathbf{P}(\{\omega\})$ pour tout $\omega \in \Omega$.

PROPOSITION 1.15
Soit Ω non vide au plus dénombrable :

1. soit $(p_i)_{i\in\mathbf{N}^*}$ (resp. $(p_i)_{i/leqn}$) une suite de réels positifs tels que $\sum_{i=1}^{\infty} p_i = 1$ (resp. $\sum_{i=1}^{n} p_i = 1$) avec Ω dénombrable (resp. $\operatorname{card}\Omega = n$) alors
$$\mathbf{P} : A \mapsto \mathbf{P}(a) = \sum_{i \text{ tel que } \omega_i \in A} p_i \ ;$$
2. si $\mathbf{P}$ est une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ alors $\mathbf{P}$ vérifie l'égalité précédente pour une certaine famille $(p_i)_{i\in I}$ telle que $p_i = \mathbf{P}(\{\omega_i\})$.

DÉMONSTRATION
Dans l'ordre :

1. $\mathbf{P}(\Omega) = 1$. Soit $(A_j)_{j\in\mathbf{N}}$ une famille d'événements deux) deux disjoints, on a
$$\mathbf{P}\left(\bigcup_{j\in\mathbf{N}} A_j\right) = \sum_{i \text{ tel que } \omega_i \in \bigcup_{j\in\mathbf{N}} A_j} p_i = \sum_j \sum_{i \text{ tel que } \omega_i \in A_j} p_i = \sum_j \mathbf{P}(A_j).$$
2. Soit $A \in \mathcal{P}(\Omega)$. On a
$$\mathbf{P}(A) = \mathbf{P}\left(\bigcup_{i \text{ tel que } \omega_i \in A} \{\omega_i\}\right) = \sum_{i \text{ tel que } \omega_i \in A} \mathbf{P}(\{\omega_i\}).$$

EXEMPLE. On lance une pièce jusqu'à obtenir la face pile.

$$\Omega = \{\omega_1, \ldots\} \cup \{\omega_\infty\}.$$

On définit une probabilité sur $(\Omega, \mathcal{P}(\Omega))$:

$$\mathbf{P}(\{\omega_j\}) = \frac{1}{2^j},\ \mathbf{P}(\{\omega_\infty\}) = 0.$$

On vérifie la valeur de la somme :

$$\sum_{j=1}^{\infty} \mathbf{P}(\{\omega_j\}) + \mathbf{P}(\{\omega_\infty\}) = \sum_{j=1}^{\infty} 2^{-j} + 0 = \frac{1/2}{1-1/2} = 1.$$

1.8 Probabilités uniformes sur un intervalle de R

THÉORÈME 1.16 (Existence et unicité de la mesure de LEBESGUE)
On a :

1. il existe une tribu $\mathcal{B}(\mathbf{R})$ qui contient les intervalles de $\mathbf{R}$ et il existe une fonction $\lambda : \mathcal{B}(\mathbf{R}) \to \mathbf{R}_+ \cup \{\infty\}$ qui est σ-additive et qui vérifie $\lambda([a,b]) = b - a$;
2. $\mathcal{B}(\mathbf{R}) \subsetneq \mathcal{P}(\mathbf{R})$;
3. le résultat est faux avec $\mathcal{P}(\mathbf{R})$.

PROBABILITÉ SUR $([0,1], \mathcal{A})$. Avec $\mathcal{A} = \{A \in \mathcal{B}(\mathbf{R}) \mid A \subset [0,1]\}$, la fonction $\lambda : \mathcal{A} \to [0,1]$ ainsi définie est appelée probabilité uniforme sur $[0,1]$ car si I est un intervalle inclus dans $[0,1]$ alors $\mathcal{P}(I)$ ne dépend que de la longueur de l'intervalle.

REMARQUES. On a $\mathbf{P}(\{\omega\}) = 0$ pour tout $\omega \in [0,1]$, $\mathbf{P}(\mathbf{Q} \cap [0,1]) = 0$.

2 PROBABILITÉS CONDITIONNELLES

2.1 Intuition et définition

OBJECTIF. Recalculer la probabilité d'un événement en fonction d'une information partielle.

INTUITION. Considérons une expérience aléatoire et un événement A (par exemple battre des cartes et « la première carte est un as ») et considérons un second événement (« la dernière carte est un as »). On effectue N fois l'expérience, on ne retient que les $N(B)$ expériences où B s'est produit. Sur ces $N(B)$ expériences, on compte le nombre de fois où A s'est produit. Le rapport $N(A \cap B)/N(B)$ est la fréquence empirique de réalisation de A sachant que B s'est produit. Or

$$\frac{N(A \cap B)}{N(B)} = \frac{N(A \cap B)}{N} \frac{N}{N(B)}$$

et c'est donc le produit de la probabilité de $A \cap B$ et l'inverse de la probabilité de B.

DÉFINITION 2.1
Soient $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité, soient A et B deux événements avec $\mathbf{P}(B) > 0$. On définit la probabilité conditionnelle de A sachant B par :

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

Exemple. Soit $\Omega = \{1, \ldots, 6\}^2$ un double lancer de dés. Soit A l'événement « la somme est 5 », B « le premier dé donne 3 » et C « le premier dé donne au moins 3 ».

$$\mathbf{P}(A) = \frac{4}{36} = \frac{1}{9},\ \mathbf{P}(A|B) = \frac{1}{6},\ \mathbf{P}(A|C) = \frac{\frac{2}{36}}{\frac{24}{36}} = \frac{1}{12}.$$

Proposition 2.2
Soit $B \in \mathcal{A}$ et tel que $\mathbf{P}(B) > 0$. Alors la fonction $\mathbf{P}(\cdot|B) : \mathcal{A} \to [0,1]$ et qui $A \mapsto \mathbf{P}(A|B)$ est une probabilité sur $(\Omega, \mathcal{A})$.

Démonstration — On vérifie que pour tout $A \in \mathcal{A}$, $\mathbf{P}(A|B) \in [0,1]$. En effet :

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

et comme $A \cap B \subset B$ on a $\mathbf{P}(A \cap B) \leq \mathbf{P}(B)$.

— On vérifie que $\mathbf{P}(\Omega|B) = 1$, en effet

$$\mathbf{P}(\Omega|B) = \frac{\mathbf{P}(\Omega \cap B)}{\mathbf{P}(B)} = 1.$$

— On vérifie que $\mathbf{P}(\cdot|B)$ est σ-additive : soit $(A_i)_{i\in I}$ une famille au plus dénombrable d'événements deux à deux disjoints. On a

$$\mathbf{P}\left(\bigcup_{i\in I} A_i|B\right) = \frac{\mathbf{P}\left(\left(\bigcup_{i\in I} A_i\right) \cap B\right)}{\mathbf{P}(B)}$$

or

$$\left(\bigcup_{i\in I} A_i\right) \cap B = \bigcup_{i\in I}(A_i \cap B).$$

De plus $A_i \cap B \subset A_i$ et A_i deux à deux disjoints et donc $A_i \cap B$ sont deux à deux disjoints. D'où :

$$\mathbf{P}\left(\bigcup_{i\in I}(A_i \cap B)\right) = \sum_{i\in I} \mathbf{P}(A_i \cap B)$$

et donc

$$\mathbf{P}\left(\bigcup_{i\in I} A_i|B\right) = \sum_{i\in I} \frac{\mathbf{P}(A_i \cap B)}{\mathbf{P}(B)} = \sum_{i\in I} \mathbf{P}(A_i|B).$$

Théorème 2.3 (Formule des probabilités totales)
Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité.

1. Soient $A, B \in \mathcal{A}$ tel que $\mathbf{P}(B) \in]0,1[$ alors

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

2. Soit $(B_i)_{i\in I}$ avec I au plus dénombrable, une partition de Ω avec $\mathbf{P}(B_i) \in]0,1[$ pour tout i. Alors

$$\mathbf{P}(A) = \sum_{i\in I} \mathbf{P}(A|B_i)\mathbf{P}(B_i).$$

Démonstration
Le second point entraine le premier. En effet, (B, B^c) est une partition de Ω.

Pour le second point, on considère la famille $(B_i \cap A)_{i\in I}$ les événements de cette famille sont deux à deux disjoints et

$$\bigcup_{i\in I}(B_i \cap A) = \left(\bigcup_{i\in I} B_i\right) \cap A = \Omega \cap A = A.$$

Ainsi

$$\mathbf{P}(A) = \sum_{i\in I} \mathbf{P}(B_i \cap A)$$

or

$$\mathbf{P}(A|B_i) = \frac{\mathbf{P}(A\cap B_i)}{\mathbf{P}(B_i)}$$

et donc

$$\mathbf{P}(A\cap B_i) = \mathbf{P}(A|B_i)\mathbf{P}(B_i).$$

Finalement :

$$\mathbf{P}(A) = \sum_{i\in I} \mathbf{P}(A|B_i)\mathbf{P}(B_i).$$

Exemples. On a deux pièces, une pièce honnête et une pièce truquée avec deux piles. L'expérience consiste à choisir l'une des pièces au hasard et la lancer trois fois. Quelle est la probabilité d'avoir PPP ? On considère l'événement A : « obtenir PPP » et B : « on choisit la pièce honnête ».

$$\mathbf{P}(A) = \mathbf{P}(A|B)\mathbf{P}(B) + \mathbf{P}(A|B^c)\mathbf{P}(B^c) = \frac{1}{2}\mathbf{P}(A|B) + \frac{1}{2}\mathbf{P}(A\cap B^c).$$

Or $\mathbf{P}(A|B) = 1/8$ et $\mathbf{P}(A|B^c) = 1$. Finalement :

$$\mathbf{P}(A) = \frac{9}{16}.$$

Proposition 2.4 (Formule de Bayes)
Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité.

1. Soient $A, B \in \mathcal{A}$ tel que $\mathbf{P}(B) \in]0,1[$ et $\mathbf{P}(A) > 0$.

$$\mathbf{P}(B|A) = \frac{\mathbf{P}(A|B)\mathbf{P}(B)}{\mathbf{P}(A|B)\mathbf{P}(B) + \mathbf{P}(A|B^c)\mathbf{P}(B^c)}.$$

2. Soient $(B_i)_{i\in I}$ avec I au plus dénombrable, une partition de Ω, $\mathbf{P}(B_i) \in]0,1[$ pour tout i. Soit $A \in \mathcal{A}$ tel que $\mathbf{P}(A) > 0$.

$$\begin{aligned}\forall i \in I,\ \mathbf{P}(B_i|A) &= \frac{\mathbf{P}(A|B_i)\mathbf{P}(B_i)}{\sum_{k\in I}\mathbf{P}(A|B_k)\mathbf{P}(B_k)} \\ &= \frac{\mathbf{P}(A|B_i)\mathbf{P}(B_i)}{\mathbf{P}(A)}.\end{aligned}$$

Démonstration
La deuxième point entraine le premier en prenant (B, B^c) comme partition.

Pour le second point, soit $i \in I$ fixé.

$$\begin{aligned}\mathbf{P}(B_i|A) &= \frac{\mathbf{P}(B_i \cap A)}{\mathbf{P}(A)} \\ &= \frac{\mathbf{P}(A|B_i)\mathbf{P}(B_i)}{\mathbf{P}(A)} \\ &= \frac{\mathbf{P}(A|B_i)\mathbf{P}(B_i)}{\sum_{k\in I}\mathbf{P}(A|B_k)\mathbf{P}(B_k)}.\end{aligned}$$

Exemple. Le test d'une maladie rare (la vache folle). Un laboratoire produit un test de détection de la maladie sur la notice il est indiqué :

— si le test est appliqué sur une vache malade, alors le test est positif avec une probabilité de 99.8% ;

— si le test est appliqué sur une vache saine, alors le test est négatif avec probabilité de 99.6%.

Le test est-il bon ? Il y a une vache malade sur 10^5. On fait le test sur une vache, le test est positif. Quelle est la probabilité que la vache soit malade ? On considère les événements : M : « la vache est malade », T^+ : « le test est positif » et T^- : « le test est négatif ».

$$\mathbf{P}(T^+|M) = 0.998,\ \mathbf{P}(T^-|M^c) = 0.996\ ;$$

$$\mathbf{P}(T^-|M) = 0.002,\ \mathbf{P}(T^+|M^c) = 0.004\ ;$$

et

$$\mathbf{P}(M) = 10^{-5}.$$

On cherche $\mathbf{P}(M|T^+)$:

$$\begin{aligned}\mathbf{P}(M|T^+) &= \frac{\mathbf{P}(T^+|M)\mathbf{P}(M)}{\mathbf{P}(T^+|M)\mathbf{P}(M) + \mathbf{P}(T^+|M^c)\mathbf{P}(M^c)} \\ &= \frac{0.998 \cdot 10^{-5}}{0.998 \cdot 10^{-5} + 0.004 \cdot (1 - 10^{-5})} \\ &\simeq \frac{10^{-5}}{10^{-5} + 0.004} = 0.0025 = 0.25\%.\end{aligned}$$

On a vu

$$\mathbf{P}(A \cap B) = \mathbf{P}(A|B)\mathbf{P}(B)$$

si $\mathbf{P}(B) > 0$. Généralisation :

Proposition 2.5 (Formule de conditionnement multiple)
Soit $(A_i)_{i\le n}$ une suite d'événements avec

$$\mathbf{P}\left(\bigcap_{i=1}^{n-1} A_i\right) > 0.$$

— Si $n = 2$:

$$\mathbf{P}(A_1 \cap A_2) = \mathbf{P}(A_2|A_1)\mathbf{P}(A_1)\ ;$$

— pour $n \ge 3$:

$$\mathbf{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbf{P}\left(A_n|\bigcap_{i=1}^{n-1} A_i\right)\mathbf{P}\left(A_{n-1}|\bigcap_{i=1}^{n-2} A_i\right) \times \ldots \times \mathbf{P}(A_2|A_1)\mathbf{P}(A_1).$$

Démonstration
Par récurrence sur n.

- Si $n = 2$ alors c'est vérifié.
- Au rang $n = (n-1) + 1$ on a :

$$\begin{aligned}\mathbf{P}\left(\bigcap_{i=1}^{n} A_n\right) &= \mathbf{P}\left(\left(\bigcap_{i=1}^{n-1} A_i\right) \cap A_n\right) \\ &= \mathbf{P}\left(A_n | \bigcap_{i=1}^{n-1} A_i\right) \mathbf{P}\left(\bigcap_{i=1}^{n-1} A_i\right).\end{aligned}$$

Exemple. On a une urne avec 6 boules blanches et 4 boules noires. On tire sans remise 3 boules. Quelle est la probabilité d'avoir la série BBN. On note A_1 l'événement « la première boule est blanche », A_2 : « la deuxième boule est blanche », A_3 : « la troisième est noire ».

$$\mathbf{P}(A_1 \cap A_2 \cap A_3) = \mathbf{P}(A_3 | A_1 \cap A_2)\mathbf{P}(A_2|A_1)\mathbf{P}(A_1) = \frac{4}{8} \cdot \frac{5}{9} \cdot \frac{6}{10} = \frac{2}{9}.$$

3 Indépendance

Intuition. Deux événements 1 et B sont indépendants si la réalisation de B n'influe pas la probabilité de réalisation de A. C'est-à-dire

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

Et si la réalisation de A n'influe pas la probabilité de réalisation de B, on a $\mathbf{P}(B|A) = \mathbf{P}(B)$. Or

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

et donc

$$\mathbf{P}(A|B) = \mathbf{P}(A) \iff \mathbf{P}(A \cap B)\mathbf{P}(A)\mathbf{P}(B).$$

Définition 3.1
Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité, soient $A, B \in \mathcal{A}$. On dit que A et B sont indépendantes si $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$.

Exemple. Lancer de deux dés $\Omega = \{1, \dots, 6\}^2$. A est l'événement « le premier dé donne 4 », B « le secondé donne 3 ».

$$\mathbf{P}(A) = \frac{1}{6},\ \mathbf{P}(B) = \frac{1}{6},\ \mathbf{P}(A \cap B) = \frac{1}{36}$$

et donc A et B sont indépendants. Soit C l'événement : « la somme veut 7 ».

$$\mathbf{P}(C) = \frac{6}{36} = \frac{1}{6};\ \mathbf{P}(A) = \frac{1}{6},\ \mathbf{P}(A \cap C) = \frac{1}{36}$$

et donc A et C sont indépendants. Si D est l'événement « la somme vaut 6 » alors

$$\mathbf{P}(D) = \frac{5}{36},\ \mathbf{P}(A) = \frac{1}{6},\ \mathbf{P}(A \cap D) = \frac{1}{36}$$

et donc A et D ne sont pas indépendants.

REMARQUES.

- Indépendance n'est pas la disjonction des événements. Ainsi, A et B sont indépendants et compatibles implique que $\mathbf{P}(A) = 0$ ou $\mathbf{P}(B) = 0$.
- L'indépendance est commutative mais pas transitive.

  EXEMPLE. Soit E l'événement « la différence est paire ».
  $$\mathbf{P}(E) = \frac{1}{2},\ \mathbf{P}(A) = \frac{1}{6},\ \mathbf{P}(E \cap A) = \frac{3}{36} = \frac{1}{12}.$$
  A et E sont donc indépendants.
  $$\mathbf{P}(C) = \frac{1}{6},\ \mathbf{P}(C \cap A) = \frac{1}{36}$$
  et donc A et C sont indépendants.
  $$\mathbf{P}(E) = \frac{1}{2},\ \mathbf{P}(C) = \frac{1}{6},\ \mathbf{P}(E \cap C) = 0$$
  et donc C et E ne sont pas indépendants.
- La notion d'indépendance dépend de la probabilité. Si A et B sont indépendants, pour $\mathbf{P}$, alors on n'a pas forcément
  $$\mathbf{P}(A \cap B|C) = \mathbf{P}(A|C)\mathbf{P}(B|C).$$

> DÉFINITION 3.2
> Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité. Soient $(A_i)_{i \le n}$ n événements ; on dit que les événements sont indépendants deux à deux si pour tout $i \neq j$ avec $i, j \le n$ on a
> $$\mathbf{P}(A_i \cap A_j) = \mathbf{P}(A_i)\mathbf{P}(A_j).$$

EXEMPLE. On considère les dates d'anniversaire de 3 personnes prises au hasard, appelées X, Y, Z. On appelle A l'événement « X et Y sont nés le même jour », B : « X et Y nés le même jour » et C : « Y et Z nés le même jour ».
$$\mathbf{P}(A) = \frac{1}{365},\ \mathbf{P}(B) = \mathbf{P}(C) = \frac{1}{365}.$$
Comme $\mathbf{P}(A \cap B) = 1/365^2$, les événements A, B, C sont indépendants deux à deux. Mais si A et B se produisent alors C se produit.
$$\mathbf{P}((A \cap B) \cap C) = \mathbf{P}(A \cap B) = \frac{1}{365^2}$$
et donc $A \cap B$ et C ne sont pas indépendants.
$$\frac{1}{365^2} = \mathbf{P}(A \cap B \cap C) \neq \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C) = \frac{1}{365^3}.$$
La notion d'indépendance deux à deux n'est pas très ferme.

> DÉFINITION 3.3
> Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité. Soit $(A_i)_{i \le n}$ une famille de n événements, on dit que les événements sont mutuellement indépendants si pour tous $i_1 < \ldots < i_k$ entre 1 et n et pour tout $2 \le k \le n$, on a
> $$\mathbf{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbf{P}(A_{i_1}) \ldots \mathbf{P}(A_{i_k}).$$

Par convention, « indépendants » veut dire « mutuellement indépendants ».

Remarques.

— L'indépendance entraine l'indépendance deux à deux.
— Il y a $2^n - n - 1$ égalités à vérifier.

Définition 3.4

Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité, soit $(A_i)_{i\in\mathbf{N}}$ une famille d'événements. On dit que les $(A_i)_{i\in\mathbf{N}}$ sont indépendants si pour tout $n \geq 2$, $(A_i)_{i\leq n}$ est une famille d'événements indépendant si, et seulement si, pour tout $k \geq 2$, et pour tous $i_1 < \ldots < i_k \in \mathbf{N}$ on a

$$\mathbf{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbf{P}(A_{i_1}) \ldots \mathbf{P}(A_{i_k}).$$

Chapitre 2

Variables aléatoires discrètes

1 Définitions

On s'intéresse à une quantité dont la valeur dépend du résultat de l'expérience aléatoire : c'est une variable aléatoire, $X : \Omega \to E,\ \omega \mapsto X(\omega)$. On considère E au plus dénombrable dans ce chapitre. Il faut également que Ω, qui est l'ensemble des événements où X prend une valeur $x \in E$, appartienne à $\mathcal{A}$.

Définition 1.1
Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité. Une variable aléatoire discrète, X, est une application de $\Omega \to E$ avec E un ensemble au plus dénombrable et tel que

$$\forall A \in \mathcal{P}(E),\ \{\omega \in \Omega \,|\, X(\omega) \in A\} \in \mathcal{A}.$$

On note

$$\{X \in A\} = \{\omega \in \Omega \,|\, X(\omega) \in A\}.$$

Exemple. On considère le lancer de deux dés. Soit $\Omega = \{1, \ldots, 6\}^2$ et $A = \mathcal{P}(\Omega)$. On considère l'application :

$$S\colon \begin{cases} \Omega \to \{2, \ldots, 12\} \\ (\omega_1, \omega_2) \mapsto \omega_1 + \omega_2 \end{cases}.$$

Remarque. Si $\mathcal{A} = \mathcal{P}(\Omega)$ alors pour tout $A \in \mathcal{P}(E)$, $\{X \in A\} \in \mathcal{A}$. En effet il suffit de vérifier que pour tout $x \in E$, $\{X = x\} \in \mathcal{A}$. Mais E est au plus dénombrable et donc

$$\{x \in A\} = \bigcup_{x \in A} \{X = x\} \in \mathcal{A}.$$

Définition 1.2
Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité et soit $X : \Omega \to E$ une variable aléatoire discrète. La loi de X est la fonction :

$$p_X\colon \begin{cases} \mathcal{P}(E) \to [0, 1] \\ A \mapsto \mathbf{P}(X \in A) \end{cases}$$

avec

$$\mathbf{P}(X \in A) = \mathbf{P}(\{X \in A\}) = \mathbf{P}(\{\omega \in \Omega \,|\, X(\omega) \in A\}.$$

PROPOSITION 1.3
Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité et $X : \Omega \to E$ une variable aléatoire discrète.

1. La loi p_X de X est une probabilité sur $(E, \mathcal{P}(E))$.
2. La loi p_X de X est caractérisée par les nombres :

$$\{p_X(x) \,|\, x \in E\}$$

où

$$l_X(x) = \mathbf{P}(X = x) = \mathbf{P}(\{X = x\}) = \mathbf{P}_X(x).$$

DÉMONSTRATION
Dans l'ordre :

1. On a

$$p_X(E) = \mathbf{P}(X \in E) = \mathbf{P}(\Omega) = 1.$$

Soit (A_i) une famille d'événements de $\mathcal{P}(E)$ deux) deux disjoints (au plus dénombrable).

$$p_X\left(\bigcup A_i\right) = \mathbf{P}\left(\left\{X \in \bigcup A_i\right\}\right) = \mathbf{P}\left(\bigcup \{X \in A_i\}\right) = \sum \mathbf{P}(X \in A_i).$$

2. Soit $A \in \mathcal{P}(E)$. On a

$$p_X(A) = \mathbf{P}(\{X \in A\}) = \mathbf{P}\left(\bigcup_{x \in A} \{X = x\}\right) = \sum_{x \in A} \mathbf{P}(X = x) = \sum_{x \in A} p_x(x).$$

HISTOGRAMME. C'est la représentation graphique de la loi d'une variable discrète à valeurs dans $E \subset \mathbf{Z}$. p_X est caractérisée par $\{p_X(x) \,|\, x \in E\}$ avec $p_X(x) = \mathbf{P}(X = x)$. Chaque bâton a pour base un intervalle de longueur 1, centré sur x et de hauteur $p_X(x)$.

LOI DE BERNOULLI. Soit $p \in [0, 1]$. La loi de BERNOULLI est la probabilité sur $(\{0, 1\}, \mathcal{P}(\{0, 1\}))$ telle que $p_X(1) = p$. On a bien défini une probabilité car $p_X(0) = 1 - p_X(1) = 1 - p$. On note cette loi $\mathcal{B}(p)$.

DÉFINITION 1.4
On dit que X est une variable de BERNOULLI de paramètre $p \in [0, 1]$ si la loi p_X de X est une loi $\mathcal{B}(p)$.

PROPOSITION 1.5
Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité et $X : \Omega \to \{0, 1\}$ une variable aléatoire discrète. Alors la loi l_X de X est $\mathcal{B}(p)$ avec $p = \mathbf{P}(X = 1)$.

NOTATION. On note $X \sim \mathcal{B}(p)$ lorsque X suit la loi de BERNOULLI de paramètre p.

DÉFINITION 1.6 (Loi binomiale)
Soit $p \in [0, 1]$ et soit $n \in \mathbf{N}^*$. La loi binomiale de paramètre (n, p) est la probabilité sur $\{0, \ldots, n\}$ définie par

$$l_X(\{k\}) = \binom{n}{k} p^k (1 - p)^{n-k}$$

pour tout $k \leq n$. On note $\mathcal{B}(n, p)$ cette loi et $\mathcal{B}(1, p) = \mathcal{B}(p)$.

Si X est une variable aléatoire discrète de loi $\mathcal{B}(n,p)$ alors on dit que X est une variable binomiale de paramètre (n,p).

DÉFINITION 1.7 (Loi géométrique)
Soit $0 < p \leq 1$. La loi géométrique de paramètre p est la probabilité sur $(\mathbf{N}^*, \mathcal{P}(\mathbf{N}^*))$ définie par

$$p_X(n) = (1-p)^{n-1}p$$

pour tout $n \in \mathbf{N}^*$. On la note $\mathcal{G}(p)$.

1.1 Espérance

DÉFINITION 1.8
Soit X une variable aléatoire discrète à valeurs dans $E \subset \mathbf{R}$ au plus dénombrable. On note $E = \{x_n \,|\, n \in I\}$ où $I \subset \mathbf{N}$. Sa loi est caractérisée par $\{p_X(x) \,|\, x \in E\}$.
On dit que la variable aléatoire X est *intégrable* si

$$\sum_{i \in I} |x_i|\, p_X(x_i) < \infty.$$

REMARQUE. C'est toujours intégrable quand I est fini, ce n'est pas toujours le cas lorsque I est infini.

CONTRE-EXEMPLE. Avec $E = \mathbf{N}^*$ et X une variable aléatoire à valeurs dans E de loi :

$$\forall x \in \mathbf{N}^*,\ p_X(x) = \frac{1}{x(x+1)}.$$

On vérifie que $p_X(x) \geq 0$ pour tout $x \in E$. De plus :

$$\sum_{x \in E} p_X(x) = \sum_{x=1}^{\infty} \frac{1}{x} - \frac{1}{x+1} = 1.$$

Mais X n'est pas intégrable car la série

$$\sum_{x=1}^{\infty} x \left(\frac{1}{x} - \frac{1}{x+1} \right)$$

n'est pas convergente.

DÉFINITION 1.9
Soit X une variable aléatoire discrète intégrable. *L'espérance* de X est définie par :

$$\mathbf{E}[X] = \sum_{i \in I} x_i \cdot p_X(x_i).$$

Comme X est intégrable, la série définie est absolument convergente. De plus, l'espérance ne dépend pas de la numérotation, I, choisie puisque la série est absolument convergente.

DÉFINITION 1.10 (Extension)
Si X est une variable aléatoire discrète à valeurs dans $\mathbf{R}_+$, non nécessairement intégrable, alors son espérance

$$\mathbf{E}[X] = \sum_{i \in I} x_i \cdot p_X(x_i).$$

La somme de la série est bien définie, éventuellement infinie mais ne dépend pas de l'ordre des termes choisi.

Exemple, variable aléatoire binomiale. Soit $n \in \mathbf{N}^*$ et $p \in [0,1]$. On prend $X \sim \mathcal{B}(n,p)$.

$$\begin{aligned}
\mathbf{E}[X] &= \sum_{k=0}^{n} k \cdot p_X(k) \\
\mathbf{E}[X] &= \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} \\
\mathbf{E}[X] &= (1-p)^n \sum_{k=0}^{n} k \binom{n}{k} \left(\frac{p}{1-p}\right)^k .
\end{aligned}$$

Mais

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k$$

et en dérivant :

$$n(1+x)^{n-1} = \frac{\mathrm{d}}{\mathrm{d}x}(1+x)^n = \sum_{k=0}^{n} k \binom{n}{k} x^{k-1}$$

et finalement :

$$\begin{aligned}
\mathbf{E}[X] &= (1-p)^n \left(\frac{p}{1-p}\right) \sum_{k=0}^{n} k \binom{n}{k} \left(\frac{p}{1-p}\right)^{k-1} \\
\mathbf{E}[X] &= (1-p)^n n \frac{p}{1-p} \left(\frac{1}{1-p}\right)^{n-1} \\
\mathbf{E}[X] &= np.
\end{aligned}$$

Exemple, variable aléatoire géométrique. Soit $p \in]0,1]$ et $X \sim \mathcal{G}(p)$ définie par

$$\forall x \in \mathbf{N}^*,\ p_X(j) = p(1-p)^{j-1}.$$

Par définition,

$$\mathbf{E}[X] = \sum_{j=1}^{\infty} jp(1-p)^{j-1}$$

or on sait que pour $x \in [0,1[$

$$\begin{aligned}
\sum_{n=0}^{\infty} x^n &= \frac{1}{1-x} \\
\frac{\mathrm{d}}{\mathrm{d}x} \sum_{n=0}^{\infty} x^n &= \frac{1}{(1-x)^2}
\end{aligned}$$

et on a

$$\frac{\mathrm{d}}{\mathrm{d}x} \sum_{n=0}^{\infty} x^n = \sum_{n=0}^{\infty} \frac{\mathrm{d}}{\mathrm{d}x}(x^n) = \sum_{n=1}^{\infty} nx^{n-1}$$

et donc

$$\begin{aligned}
\mathbf{E}[X] &= p \sum_{j=1}^{\infty} j(1-p)^{j-1} \\
\mathbf{E}[X] &= p \frac{1}{(1-(1-p))^2} \\
\mathbf{E}[X] &= \frac{1}{p}.
\end{aligned}$$

DÉFINITION 1.11
Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité. On définit $L^1_d(\Omega, \mathcal{A}, \mathbf{P})$ comme l'ensemble des variables aléatoires discrètes intégrables définies sur $(\Omega, \mathcal{A}, \mathbf{P})$.

PROPOSITION 1.12
On a :

1. L^1_d a une structure d'espace vectoriel sur $\mathbf{R}$. De plus, $\mathbf{E}$ est une application linéaire de L^1_d dans $\mathbf{R}$.
2. Si $X \geq 0$ [§1] alors $\mathbf{E}[X] \geq 0$.
3. Si $X \geq Y$ alors $\mathbf{E}[X] \geq \mathbf{E}[Y]$.

DÉMONSTRATION
Dans l'ordre :

1. Si $E = X(\Omega)$ et $F = Y(\Omega)$ alors $\lambda X + \mu Y$ est à valeurs dans $G = \{\lambda x + \mu y \,|\, x \in E, y \in E\}$ et donc G est au plus dénombrable car il existe une surjection de $E \times F$ dans G.
   $\lambda X + \mu Y$ reste intégrable.
2. La somme convergente de termes positifs est positive.
3. En posant $Z = X - Y$, d'après le premier point $\mathbf{E}[Z] = \mathbf{E}[X] - \mathbf{E}[Y]$ on a par le second point $\mathbf{E}[Z] \geq 0$ et donc $\mathbf{E}[X] \geq \mathbf{E}[Y]$.

PROPOSITION 1.13
Soit $X : \Omega \to E$ une variable aléatoire discrète (E au plus dénombrable) et soit $f : E \to \mathbf{R}$. On pose $Y = f(x)$.

1. Y est une variable aléatoire discrète, à valeurs dans $F = f(E)$. Sa loi est caractérisée par :
$$\{p_Y(y) \,|\, y \in F\}, \; p_Y(y) = \sum_{x \in f^{-1}(y)} p_X(x).$$
2. Y est intégrable si, et seulement si,
$$\sum_{x \in E} |f(x)| \, p_X(x) < \infty.$$
   Dans ce cas,
$$\mathbf{E}[Y] = \sum_{x \in E} f(x) p_X(x).$$

DÉMONSTRATION
On a :

1. Y est à valeurs dans F et
$$\{Y = y\} = \bigcup_{x \in f^{-1}(y)} \{X = x\} \in \mathcal{A}$$
   et donc Y est une variable discrète. L'union est disjointe et donc :
$$p_Y(y) = \mathbf{P}(Y = y) = \sum_{x \in f^{-1}(y)} \mathbf{P}(X = x).$$

§1. *I.e.* pour tout $\omega \in \Omega$, $X(\Omega) \geq 0$.

2. Y est intégrable à l'unique condition :

$$\sum_{y\in F} |y|\, p_Y(y) < \infty.$$

Mais :

$$\begin{aligned}\sum_{y\in F} |y|\, p_Y(y) &= \sum_{y\in F} |y| \left(\sum_{x\in f^{-1}(y)} p_X(x)\right)\\ &= \sum_{y\in F}\sum_{x\in f^{-1}(y)} |f(x)|\, p_X(x)\\ &= \sum_{x\in E} |f(x)|\, p_X(x).\end{aligned}$$

Dans ce cas, par arrangement des termes,

$$\begin{aligned}\mathbf{E}[Y] &= \sum_{y\in F} y p_Y(y)\\ &= \sum_{y\in F} y \sum_{x\in f^{-1}(y)} p_X(x)\\ &= \sum_{x\in E} f(x) p_X(x).\end{aligned}$$

Proposition 1.14
Soient $X:\Omega\to E$ une variable aléatoire discrète à valeurs dans $E\subset\mathbf{R}$ et $Y:\Omega\to E$ de même dans $F\subset\mathbf{R}^+$ et intégrable. Si $|X|\le Y$ alors X est intégrable et

$$|\mathbf{E}[X]| \le \mathbf{E}[|X|] \le \mathbf{E}[Y].$$

Application de la linéarité de l'espérance : la formule d'inclusion-exclusion.
Soient $A_1,\dots,A_n$ n événements définis sur un même espace de probabilité. On a

$$\mathbf{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k+1} \sum_{1\le i_1<\ldots<i_k\le n} \mathbf{P}(A_{i_1}\cap\ldots\cap A_{i_k}).$$

On peut faire une preuve à partir des indicatrices. Pour tous $1\le i_1<i_2<\ldots<i_k\le n$ on a

$$1_{\bigcap_{i=1}^k A_{i_k}} = \prod_{i=1}^k 1_{A_{i_k}}.$$

Ainsi,

$$\begin{aligned}1_{\bigcap_{i=1}^k A_{i_k}}(\omega) = 1 &\iff \omega\in\bigcap_{i=1}^k A_{i_k}\\ &\iff \prod_{j=1}^k 1_{A_{i_j}}(\omega) = 1.\end{aligned}$$

On a

$$1 - 1_{\bigcup_{i=1}^n A_i} = \prod_{i=1}^n (1-1_{A_i}).$$

En effet,

$$\begin{aligned}1 - 1_{\bigcup_{i=1}^n A_i} &= 1_{(\bigcup_{i=1}^n A_i)^c}\\ &= 1_{\bigcap_{i=1}^n A_i^c}\\ &= \prod_{i=1}^n 1_{A_i^c}\\ &= \prod_{i=1}^n (1 - 1_{A_i}).\end{aligned}$$

Ainsi,

$$\begin{aligned}1_{\bigcup_{i=1}^n A_i} &= 1 - \prod_{i=1}^n (1 - 1_{A_i})\\ &= \sum_{i=1}^n 1_{A_i} - \sum_{1\le i<j\le n} 1_{A_i} 1_{A_j} + \dots\end{aligned}$$

Or on a

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

Donc

$$\begin{aligned}\mathbf{E}\left[1_{\bigcup_{i=1}^n A_i}\right] &= \sum_{i=1}^n \mathbf{E}[1_{A_i}] - \sum_{i<j} \mathbf{E}[1_{A_i\cap A_j}] + \dots\\ \mathbf{P}\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n \mathbf{E}[1_{A_i}] - \sum_{i<j} \mathbf{E}[1_{A_i\cap A_j}] + \dots\end{aligned}$$

Application à la formule d'inclusion-exclusion : problème des arrangements. On s'intéresse au battage de N cartes. $\Omega = S_N$, $\mathcal{A} = \mathcal{P}(\Omega)$ et $\mathbf{P}$ la probabilité uniforme. Soit A l'événement « aucune carte n'est à sa place après le battage ». C'est-à-dire :

$$A = \bigcap_{i=1}^N \{\omega \in \Omega \,|\, \omega(i) \neq i\}.$$

Soit

$$A_i = \{\omega \in \Omega \,|\, \omega(i) = i\}.$$

Alors

$$A^c = \bigcup_{i=1}^N A_i.$$

On a

$$\mathbf{P}(A_i) = \frac{\operatorname{card} A_i}{\operatorname{card} \Omega}.$$

Or $\operatorname{card}\Omega = N!$ et $\operatorname{card} A_i = \operatorname{card} A_n$. En effet la conjugaison par la transposition (i, n) envoie A_i sur A_n univoquement. Mais $\operatorname{card} A_N = (N-1)!$. Finalement $\mathbf{P}(A_i) = 1/N$.

Pour $i < j$ on a $\operatorname{card}(A_i \cap A_j) = \operatorname{card}(A_n \cap A_{n-1})$ et donc $\mathbf{P}(A_i \cap A_j) = 1/n(n-1)$.

Plus généralement, $1 \le i_1 < \ldots < i_k \le N$ on a

$$\mathbf{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \frac{\operatorname{card} A_{i_1} \cap \ldots \cap A_{i_k}}{N!}.$$

On a

$$\operatorname{card} A_{i_1} \cap \ldots \cap A_{i_k} = \operatorname{card} A_{N-k+1} \cap \ldots \cap A_N = (N-k)!.$$

Finalement,

$$\mathbf{P}(A_{i_1} \cap \dots A_{i_k}) = \frac{(N-k)!}{N!} = \frac{1}{N(N-1)\dots(N-k+1)}.$$

En conclusion

$$\begin{aligned}
\mathbf{P}(A) &= 1 - \mathbf{P}\left(\bigcup_{i=1}^{N} A_i\right) \\
&= 1 - \sum_{k=1}^{N} (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq N} \mathbf{P}\left(\bigcap_{j=1}^{k} A_{i_k}\right) \\
&= 1 + \sum_{k=1}^{N} (-1)^{k} \sum_{1 \leq i_1 < \dots < i_k \leq N} \frac{(N-k)!}{N!} \\
&= 1 + \sum_{k=1}^{N} (-1)^{k} \binom{N}{k} \frac{(N-k)!}{N!} \\
&= 1 + \sum_{k=1}^{N} (-1)^{k} \frac{1}{k!}
\end{aligned}$$

On a

$$\left|\mathbf{P}(A) - e^{-1}\right| \leq \frac{1}{(N+1)!}$$

et donc $\mathbf{P}(A) \simeq 0.37$.

1.2 Variance

Définition 1.15
La variance est l'écart quadratique moyen. Si $X \in L^1_d$ alors

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

Remarque. $\mathrm{Var}(X) \in [0, +\infty]$.

Définition 1.16
L'écart-type est la racine carrée de la variance. Si $X \in L^1_d$ alors

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

Remarque. $\sigma(X)$ mesure la dispersion de X et a la même dimension que X (si X s'exprime mètres, $\sigma(X)$ et $\mathbf{E}[X]$ s'expriment en mètres).

Proposition 1.17
On a :

1. $\mathrm{Var}(X) \in [0, \infty[$ si, et seulement si X est de carré intégrable $X^2 \in L^1_d$.
2. $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$ pour X de carré intégrable et $a, b \in \mathbf{R}$. En particulier, Var n'est pas linéaire.

3. Formule de HUYGENS :

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

pour tout X de carré intégrable.

4. Si $\text{Var}(X) = 0$ alors $\mathbf{P}(X = \mathbf{E}[X]) = 1$.

DÉMONSTRATION
Dans l'ordre :

1. On a
$$(X - \mathbf{E}[X])^2 \leq 2X^2 + 2\mathbf{E}[X]^2$$
mais X est de carré intégrable donc $X^2 \in L^1_d, \mathbf{E}[X]^2 \in L^d_1$ donc la somme est encore dans L^1_d et donc $(X - \mathbf{E}[X])^2 \in L^d_1$.
2. On a
$$\begin{aligned}\text{Var}(aX + b) &= \mathbf{E}[(aX + b - \mathbf{E}[aX + b])^2] \\ &= \mathbf{E}[(aX + b - a\mathbf{E}[X] - b)^2] \\ &= a^2\mathbf{E}[(X - \mathbf{E}[X])^2] \\ &= a^2\text{Var}(X).\end{aligned}$$
3. On a
$$\begin{aligned}\text{Var}(X) &= \mathbf{E}[(X - \mathbf{E}[X])^2] \\ &= \mathbf{E}[X^2 - 2X\mathbf{E}[X] + \mathbf{E}[X]^2] \\ &= \mathbf{E}[X^2] - 2\mathbf{E}[X]\mathbf{E}[X] + \mathbf{E}[X]^2 \\ &= \mathbf{E}[X^2] - \mathbf{E}[X]^2\end{aligned}$$
4. Supposons que $\text{Var}(X) = 0$. Soit $Y = (X - \mathbf{E}[X])^2$. Y est une variable aléatoire discrète à valeurs dans $F \subset \mathbf{R}^+$. On a $\mathbf{E}[Y] = 0$ car $\text{Var}(X) = 0$. Mais
$$\mathbf{E}[Y] = \sum_{y \in Y} y\mathbf{P}(Y = y)$$
et donc pour tout $y \in F$, $y\mathbf{P}(Y = y) = 0$ et donc pour tout y non nul $\mathbf{P}(Y = y) = 0$. Finalement,
$$\mathbf{P}(Y = 0) = 1 - \sum_{0 \neq y \in F} \mathbf{P}(Y = y) = 1.$$
C'est-à-dire
$$\mathbf{P}(X = \mathbf{E}[X]) = 1.$$

EXEMPLE 1. Pour $X = a \in \mathbf{R}$ on a $\mathbf{E}[X] = a$ et $\text{Var}(X) = 0$. En effet, $\mathbf{E}[X^2] = a^2$ et $\mathbf{E}[X]^2 = a^2$.

EXEMPLE 2. Soit $X \sim \mathcal{B}(p)$. On a $\mathbf{E}[X] = p$, $\mathbf{E}[X^2] = \mathbf{P}(X = 0)0^2 + \mathbf{P}(X = 1)1^2 = p$. Finalement, $\text{Var}(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = p - p^2 = p(1 - p)$.

REMARQUE. $\mathbf{E}[X]$ est la valeur de $a \in \mathbf{R}$ telle que $\mathbf{E}[(X - a)^2]$ soit minimal.

Variance d'une variable binomiale. Soit $X \sim \mathcal{B}(n,p)$ pour $n \in \mathbf{N}^*, p \in [0,1]$. $\mathbf{E}[X] = np$.

$$\begin{aligned}\mathbf{E}[X^2] &= \sum_{k=0}^{n} k^2 \mathbf{P}(X=k)\\ &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k}.\end{aligned}$$

Si $p=1$ alors $\mathbf{E}[X]=n, \mathbf{E}[X^2]=n^2$ et $\mathrm{Var}(X)=0$. Si $p \in [0,1[$ alors

$$\mathbf{E}[X^2] = (1-p)^n \sum_{k=0}^{n} \binom{n}{k} k^2 q^k$$

avec $q = p/(1-p)$. Mais

$$\begin{aligned}q\frac{\mathrm{d}}{\mathrm{d}q}\left[q\frac{\mathrm{d}}{\mathrm{d}q}(1+q)^n\right] &= \sum_{k=0}^{n}\binom{n}{k} k^2 q^k\\ &= q\frac{\mathrm{d}}{\mathrm{d}q}\left[nq(1+q)^{n-1}\right]\\ &= nq(1+q)^{n-2}(1+q+q(n-1))\\ &= nq(1+q)^{n-2}(1+nq).\end{aligned}$$

Ainsi,

$$\begin{aligned}\mathbf{E}[X^2] &= (1-p)^n nq(1+q)^{n-2}(1+nq)\\ &= (1-p)^2 n\frac{p}{1-p}\left(1+n\frac{p}{1-p}\right)\\ &= np(1-p)\left(1+n\frac{p}{1-p}\right)\\ \mathrm{Var}(X) &= \mathbf{E}[X^2]-\mathbf{E}[X]^2\\ &= np(1-p)+n^2p^2-n^2p^2\\ &= np(1-p).\end{aligned}$$

Variance d'une variable géométrique. Soit $X \sim \mathcal{G}(p)$ pour $p \in]0,1]$. $\mathbf{E}[X] = 1/p$. On a

$$\begin{aligned}\mathbf{E}[X^2] &= \sum_{j=1}^{\infty} j^2 p(1-p)^{j-1}\\ &= p\sum_{j=1}^{\infty} j^2 q^{j-1}.\end{aligned}$$

Mais

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}q}\left[q\frac{\mathrm{d}}{\mathrm{d}q}\frac{1}{1-q}\right] &= \sum_{j=1}^{\infty} j^2 q^{j-1}\\ q\frac{\mathrm{d}}{\mathrm{d}q}\left[nq(1+q)^{n-1}\right] &= nq(1+q)^{n-2}[1+q+(n-1)q]\\ &= nq(1+q)^{n-2}(1+nq).\end{aligned}$$

Donc

$$\begin{aligned}
\mathbf{E}[X^2] = p\frac{\mathrm{d}}{\mathrm{d}q}\left[q\frac{\mathrm{d}}{\mathrm{d}q}\frac{1}{1-q}\right] \\
&= p\frac{\mathrm{d}}{\mathrm{d}q}\left[\frac{q}{(1-q)^2}\right] \\
&= p\frac{(1-q)^2+2(1-q)q}{(1-q)^4} \\
&= p\frac{(1-q)+2q}{(1-q)^3} \\
&= p\frac{1+q}{(1-q)^3} = \frac{2-p}{p^2}
\end{aligned}$$

Donc

$$\mathrm{Var}(X) = \frac{2-p}{p^2 - \frac{1}{p^2}} = \frac{1-p}{p^2}.$$

1.3 Loi de Poisson de paramètre $\lambda \in]0, +\infty[$

> Définition 1.18
> La loi de Poisson de paramètre $\lambda \in]0, +\infty[$ est la probabilité $\mathbf{P}$ sur $(\mathbf{N}, \mathcal{P}(\mathbf{N}))$ définie par :
> $$\forall k \in \mathbf{N},\ \mathbf{P}(\{k\}) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

On a bien

$$\sum_{k=0}^{\infty}\mathbf{P}(\{k\}) = 1.$$

> Définition 1.19
> Soit $X : (\Omega, \mathcal{A}, \mathbf{P}) \to \mathbf{N}$. On dit que X est une variable aléatoire de Poisson de paramètre λ si la loi de X est une loi de Poisson de paramètre λ.

Espérance-variance. Soit X une variable aléatoire de Poisson de paramètre $\lambda \in]0, +\infty[$.

$$\begin{aligned}
\mathbf{E}[X] &= \sum_{k=0}^{\infty} k\mathbf{P}(\{k\}) \\
&= \sum_{k=1}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} \\
&= \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}e^{-\lambda} \\
&= \lambda e^{-\lambda}e^{\lambda} = \lambda.
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}[X^2] &= \sum_{k=0}^{\infty} k^2 \mathbf{P}(\{k\}) \\
&= \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} e^{-\lambda} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda} \sum_{k=0}^{\infty} (k+1) \frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda} \left(\sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} + \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \right) \\
&= \lambda e^{-\lambda} \left(\lambda e^{\lambda} + e\lambda \right) \\
&= \lambda^2 + \lambda. \\
\mathrm{Var}(X) &= \mathbf{E}[X^2] - \mathbf{E}[X]^2 \\
&= \lambda^2 + \lambda - \lambda^2 = \lambda.
\end{aligned}$$

Lien avec la loi binomiale. Soit $X \sim \mathcal{B}(n, p_n)$ avec $p_n = \lambda/n$ pour $\lambda \in]0, \infty[$. X est le nombre de succès dans une suite d'un grand nombre d'expériences dont chaque expérience a une faible probabilité de succès.

$$\mathbf{E}[X] = np_n = \lambda$$

et

$$\mathrm{Var}(X) = np_n(1 - p_n) = \lambda(1 - \lambda/n) \underset{n\to\infty}{\longrightarrow} \lambda.$$

Soit $n \geq k$,

$$\begin{aligned}
\mathbf{P}(X = k) &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\
&= \frac{n!}{(n-k)} \frac{1}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&= \underbrace{\left(\prod_{k=0}^{k-1} \frac{n-k}{n} \right)}_{\to 1} \underbrace{\frac{\lambda^k}{k!}}_{\to \lambda^k/k!} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\to 1}.
\end{aligned}$$

2 Famille de variables aléatoires discrètes

2.1 Vecteur aléatoire

Définition 2.1
Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité. Soient $X_i : \Omega \to E_i$ des variables aléatoires discrètes, E_i au plus dénombrable, pour $i = 1, \ldots, n$.
On appelle $(X_1, \ldots, X_n)$ *vecteur aléatoire*. C'est une fonction de Ω dans $E_1 \times \ldots \times E_n$.

Remarque. Si $n = 2$ on parle plutôt de couple aléatoire.

PROPOSITION 2.2
Un vecteur aléatoire $(X_1, \ldots, X_n)$ est une variable aléatoire discrète.

DÉMONSTRATION
On a deux axiomes à vérifier.

1. $(X_1, \ldots, X_n)$ est à valeurs dans un espace, E, au plus dénombrable. En effet, $E = E_1 \times \ldots \times E_n$ est au plus dénombrable en tant que produit cartésien d'espaces au plus dénombrables.
2. Pour tout $x \in E$, $\{(X_1, \ldots, X_n) = x\} \in \mathcal{A}$. En effet, ici $x = (x_1, \ldots, x_n)$ avec $x_i \in E_i$. Mais

$$\{(X_1, \ldots, X_n) = x\} = \{(X_1, \ldots, X_n)\} = \{x_i, \ldots, x_n)\} = \bigcap_{i=1}^{n} \{X_i = x_i\} \in \mathcal{A}.$$

NOTATION. Pour simplifier, on note $X = (X_1, \ldots, X_n) : \Omega \to E$.

LOI DE X. Comme X est une variable aléatoire discrète, on peut parler de sa loi. La loi P_X de X est la probabilité sur $(E, \mathcal{P}(E))$ définie par :

$$\forall x \in E,\ P_X(\{x\}) = \mathbf{P}(X = x).$$

DÉFINITION 2.3
On appelle P_X la *loi jointe* de X. P_{X_i} est la i-ième loi marginale de X.

LEMME 2.4 (Formule des marginales)
Soit $X = (X_1, \ldots, X_n)$ un vecteur aléatoire de loi jointe P_X. On a :

$$P_{X_1}(\{x_1\}) = \sum_{x_2 \in E_2, \ldots, x_n \in E_n} P_X(\{(x_1, \ldots, x_n)\}$$

et

$$P_{(X_1, X_2)}(\{x_1, x_2\}) = \sum_{x_3 \in E_3, \ldots, x_n \in E_n} P_X(\{(x_1, \ldots, x_n)\}).$$

DÉMONSTRATION
Soit $x_1 \in E_1$,

$$\begin{aligned}
\{X_1 = x_1\} &= \bigcup_{x_2 \in E_2, \ldots, x_n \in E_n} \{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\}, \\
\mathbf{P}(X_1 = x_1) &= \mathbf{P}\left(\bigcup_{x_2 \in E_2, \ldots, x_n \in E_n} \{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\} \right) \\
&= \sum_{x_2 \in E_2, \ldots, x_n \in E_n} \mathbf{P}(X_1 = x_1, \ldots, X_n = x_n).
\end{aligned}$$

PROPOSITION 2.5
Soit X un vecteur aléatoire à valeurs dans E, soit $f : E \to \mathbf{R}$. $f(X)$ est une variable aléatoire discrète, elle est intégrable si, et seulement si,

$$\sum_{x \in E} |f(x)|\, P_X(\{x\}) < \infty$$

et dans ce cas,

$$\mathbf{E}[f(X)] = \sum_{x\in E} f(x)P_X(\{x\}).$$

Exemple. Soient $\Omega = \{P, F\}^n$ avec n fini, $\mathcal{A} = \mathcal{P}(\Omega)$ et $\mathbf{P}$ la probabilité uniforme. Si $S(\omega)$ désigne le nombre de piles dans ω alors $S \sim \mathcal{B}(n, 1/2)$. Si $T(\omega)$ est le nombre de faces alors $T \sim \mathcal{B}(n, 1/2)$. (S, T) est un couple aléatoire de loi disjointe :

$$P_{(S,T)}(\{k, k'\}) = \mathbf{P}(S = k, T = k') = \begin{cases} 0 \text{ si } k + k' \neq n \\ \binom{n}{k} 2^{-n} \text{ sinon} \end{cases}$$

Soient $X_i = 1_{A_i}$ pour $i = 1, \ldots, n$ avec $A_i = \{\omega_i = P\}$. $(X_1, \ldots, X_n)$ est un vecteur aléatoire de Ω dans $\{0, 1\}^n$. Pour tout $i = 1, \ldots, n$ on a $X_i \sim \mathcal{B}(1/2)$.

$$\mathbf{P}((X_1, \ldots, X_n) = (x_1, \ldots, x_n)) = \frac{1}{2^n}.$$

2.2 Covariance

Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité. Pour rappel, L^1_d désigne l'ensemble des variables aléatoires discrètes réelles intégrables définies sur $(\Omega, \mathcal{A}, \mathbf{P})$.

Définition 2.6
L^2_d est l'ensemble des variables aléatoires discrètes réelles de carré intégrable définies sur $(\Omega, \mathcal{A}, \mathbf{P})$. C'est-à-dire l'ensemble des $X : \Omega \to E$ avec E au plus dénombrable inclus dans $\mathbf{R}$ tel que

$$\sum_{x\in E} x^2 p_X(x) < \infty.$$

Proposition 2.7
On a :

1. si $x \in L^2_d$ alors $X \in L^1_d$;
2. L^2_d a une structure de $\mathbf{R}$-espace vectoriel ;
3. si $X, Y \in L^2_d$ alors $XY \in L^2_d$ et

$$|\mathbf{E}[XY]| \leq \mathbf{E}[X^2]^{1/2}\mathbf{E}[Y^2]^{1/2}.$$

Démonstration
Dans l'ordre :

1. On a pour tout $x \in \mathbf{R}$, $x \leq 1/2 + x^2/2$ d'où

$$|x| \leq \frac{1}{2} + \frac{x^2}{2}$$

ce qui montre que $|x|$ est intégrable si x^2 l'est.

2. Pour tous $(X, Y) \in L^2_d, (\lambda, \mu) \in \mathbf{R}^2$ on a

$$(\lambda X + \mu Y)^2 \leq 2\lambda^2 X^2 + 2\mu^2 Y^2$$

donc $\lambda X + \mu Y$ appartient à L^2_d.

3. On a $|XY| \leq X^2/2 + Y^2/2$. Donc $XY \in L^2_d$ si $X, Y \in L^2_d$.
Si $\mathbf{E}[X^2] = 0$ alors $\mathbf{P}(X^2 = 0) = 1$ et donc $\mathbf{E}[XY] = 0$.
Supposons $\mathbf{E}[X^2] > 0$ et $\mathbf{E}[Y^2] > 0$. On a alors pour tout $\lambda \geq 0$,

$$0 \leq \mathbf{E}[(X - \lambda Y)^2] = \mathbf{E}[X^2] - 2\lambda \mathbf{E}[XY] + \lambda^2 \mathbf{E}[Y^2]$$

on choisit $\lambda = \mathbf{E}[XY]/\mathbf{E}[Y^2]$ et donc on a :

$$0 \leq \mathbf{E}[X^2] - \frac{\mathbf{E}[XY]^2}{\mathbf{E}[Y^2]}.$$

Définition 2.8
Soient $X, Y \in L^2_d$. Pour rappel, $\mathrm{Var}(X) = \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 \geq 0$. On définit la covariance de X, Y :

$$\mathrm{Cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y]) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y] \in \mathbf{R}.$$

On définit également la corrélation de X, Y : si $\mathrm{Var}(X)$ et $\mathrm{Var}(Y)$ sont strictement positives alors

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)^{1/2}\mathrm{Var}(Y)^{1/2}} \in [-1, 1].$$

Remarque. Si $\mathrm{Var}(X) = 0$ ou $\mathrm{Var}(Y) = 0$ alors $\mathrm{Cov}(X, Y) = 0$. $\mathrm{Cov}(X, Y)$ et sa forme normalisée, $\mathrm{Corr}(X, Y)$, décrivent les variations simultanées de X et Y. Si $\mathrm{Cov}(X, Y) > 0$ ou $\mathrm{Corr}(X, Y) > 0$ alors cela signifie que X et Y « varient dans le même sens ».

Exemple. $\Omega = \{P, F\}^n$, $\mathcal{A} = \mathcal{P}(\omega)$ et $\mathbf{P}$ la probabilité uniforme. On pose S le nombre de piles et T le nombre de faces.

$$\mathrm{Cov}(S, S) = \mathrm{Var}(S), \ \mathrm{Corr}(S, S) = 1.$$

Comme $T = n - S$ on a

$$\begin{aligned}
\mathrm{Cov}(S, T) &= \mathbf{E}[ST] - \mathbf{E}[S]\mathbf{E}[T] \\
&= n\mathbf{E}[S] - \mathbf{E}[S^2] - \mathbf{E}[S]\mathbf{E}[T] \\
&= n\mathbf{E}[S] - \left(\mathrm{Var}(S) + \mathbf{E}[S]^2\right) - \mathbf{E}[S^2] - \mathbf{E}[S]\mathbf{E}[T] \\
&= \frac{n^2}{2} - \frac{n^2 + n}{4} - \frac{n^2}{4} = -\frac{n}{4}.
\end{aligned}$$

Or $\mathrm{Var}(S) = \mathrm{Var}(T) = n/4$ et donc

$$\mathrm{Corr}(S, T) = \frac{-n/4}{n/4} = -1.$$

Démonstration
Montrons que $\mathrm{Corr}(X, Y) \in [-1, 1]$.
On a

$$|\mathrm{Cov}(X, Y)| = |\mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y]))]| \leq \mathbf{E}[(X - \mathbf{E}[X])^2]^{1/2}\mathbf{E}[(Y - \mathbf{E}[Y])^2]^{1/2}$$

et donc

$$|\mathrm{Cov}(X, Y)| \leq \mathrm{Var}(X)^{1/2}\mathrm{Var}(Y)^{1/2}$$

Remarque. L'espérance est additive mais la variance n'est pas additive. Cependant, pour $X, Y \in L^2_d$
$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y).$$

Proposition 2.9
Soient $X_1, \ldots, X_n \in L^2_d$. On a
$$\mathrm{Var}(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j).$$

Démonstration
$$\begin{aligned}
\mathrm{Var}(X_1 + \ldots + X_n) &= \mathbf{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] - \mathbf{E}\left[\sum_{i=1}^{n} X_i\right]^2 \\
&= \mathbf{E}\left[\sum_{i,j=1}^{n} X_i X_j\right] - \left(\sum_{i=1}^{n} \mathbf{E}[X_i]\right)^2 \\
&= \sum_{i,j=1}^{n} \mathbf{E}[X_i X_j] - \sum_{i,j=1}^{n} \mathbf{E}[X_i]\mathbf{E}[X_j] \\
&= \sum_{i,j=1}^{n} \mathbf{E}[X_i X_j] - \mathbf{E}[X_i]\mathbf{E}[X_j] \\
&= \sum_{i=j=1}^{n} \mathbf{E}[X_i^2] - \mathbf{E}[X_i]^2 + 2 \sum_{1 \leq i < j \leq n} \mathbf{E}[X_i X_j] - \mathbf{E}[X_i]\mathbf{E}[X_j] \\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \leq i < j} \mathrm{Cov}(X_i, X_j).
\end{aligned}$$

2.3 Variables aléatoires discrètes indépendantes

Définition 2.10
Soient $X_1, \ldots, X_N$ des variables aléatoires indépendantes définie sur l'espace $(\Omega, \mathcal{A}, \mathbf{P})$ dans E_i avec $i \in [n]$.
On dit que $(X_i)_{i \in [n]}$ est une *famille de variables aléatoires indépendantes* si
$$\forall (x_1, \ldots, x_n) \in E_1 \times \ldots \times E_n,\ \mathbf{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbf{P}(X_i = x_i).$$

Exemple. Soit $\Omega = \{P, F\}^n$ avec $\mathbf{P}$ la probabilité uniforme. Soient S le nombre de piles, T de faces et $X_i = \mathbb{1}_{A_i}$ avec A_i l'événement « obtenir pile au n-ième lancer ».
$$\mathbf{P}(S = k, T = k') = \begin{cases} 0 \text{ si } k + k' \neq n \\ \binom{n}{k} 2^{-n} \text{ sinon} \end{cases}$$

On a $S \sim \mathcal{B}(n, 1/2)$ et $T \sim \mathcal{B}(n, 1/2)$. De plus $\mathbf{P}(S = 0, T = 0) = 0$ pour $n \neq 0$ mais $\mathbf{P}(S = 0) = 2^{-n}$ et $\mathbf{P}(T = 0) = 2^{-n}$ ce qui montre que T et S ne sont pas indépendantes.

Pour tout $(x_1, \ldots, x_n) \in \{0,1\}^n$ on a $\mathbf{P}(X_1 = x_1, \ldots, X_n = x_n) = 2^{-n}$ et pour tout $i \leq n$, $\mathbf{P}(X_i = x_i) = 1/2$. Donc les $(X_i)_{i \in \mathbf{N}}$ sont indépendants.

Proposition 2.11
Soient $X_1, \ldots, X_n$ des variables aléatoires discrètes de $(\Omega, \mathcal{A}, \mathbf{P}) \to E_i$.
La famille $(X_i)_{i \in [n]}$ est indépendante si, et seulement si, pour toute famille $(f_i)_{i \in [n]}$, $f_i : E_i \to \mathbf{R}$ avec $f_i(X_i)$ intégrable on a

$$\mathbf{E}\left(\prod_{i=1}^n f_i(X_i)\right) = \prod_{i=1}^n \left(\mathbf{E}(f_i(X_i))\right).$$

Démonstration
Supposons les X_i indépendantes.
Soient f_i telles que $f_i(X_i)$ intégrable pour tout $i \leq n$. Montrons que $\prod_{i=1}^n f_i(X_i)$ est intégrable, c'est-à-dire que

$$I = \sum_{(x_1, \ldots, x_n)} \left| \prod_{i=1}^n f_i(x_i) \right| p_{X_1, \ldots, X_n}(x_1, \ldots, X_n) < \infty.$$

On a

$$p_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \prod_{i=1}^n p_{X_i}(x_i)$$

d'où

$$\begin{aligned} I &= \sum_{(x_1, \ldots, x_n)} \prod_{i=1}^n |f_i(x_i)| \, p_{X_i}(x_i) \\ &= \prod_{i=1}^n \sum_{x_i \in E_i} |f_i(x_i)| \, p_{X_i}(x_i) \end{aligned}$$

d'après le théorème de Fubini pour les séries à termes positifs. Donc

$$I = \prod_{i=1}^n \mathbf{E}(|f_i(X_i)|) < \infty.$$

La variable aléatoire $\prod f_i(X_i)$ est donc intégrable, son espérance est

$$\begin{aligned} \mathbf{E}\left(\prod_{i=1}^n f_i(X_i)\right) &= \sum_{(x_1, \ldots, x_n)} \left(\prod_{i=1}^n f_i(x_i) p_{X_1, \ldots, X_n}(x_1, \ldots, x_n)\right) \\ &= \prod_{i=1}^n \mathbf{E}(f_i(X_i)). \end{aligned}$$

Réciproquement, supposons que pour toutes $f_i : E_i \to \mathbf{R}$ avec $f_i(X_i)$ intégrable on ait $\mathbf{E}(\prod f_i(X_i)) = \prod \mathbf{E}(f_i(X_i))$.
Soit $(x_1, \ldots, x_n) \in E_1 \times \ldots \times E_n$. On pose $f_i = \mathbb{1}_{\{x_i\}}$. $f_i(X_i)$ est intégrable car bornée. Donc

$$\begin{aligned} \mathbf{E}\left(\prod_{i=1}^n \mathbb{1}_{X_i = x_i}\right) &= \prod_{i=1}^n \mathbf{E}(\mathbb{1}_{X_i = X_i}) \\ \iff \quad \mathbf{E}\left(\mathbb{1}_{\cap \{X_i = x_i\}}\right) &= \prod_{i=1}^n \mathbf{P}(X_i = x_i) \\ \iff \quad \mathbf{P}\left(\bigcap_{i=1}^n \{X_i = x_i\}\right) &= \prod_{i=1}^n \mathbf{P}(X_i = x_i) \end{aligned}$$

donc les X_i sont indépendants.

Exemple. Indépendant deux à deux est différent d'indépendant. On considère $\Omega = \{-1,1\}^2$ et $\mathbf{P}$ la probabilité uniforme sur Ω. On pose $X : (\omega_1, \omega_2) \mapsto \omega_1$, $Y : (\omega_1, \omega_2) \mapsto \omega_2$ et $Z = XY$.

X et Y sont indépendantes, en effet, $\mathbf{P}(X = x, Y = y) = \frac{1}{4} = \mathbf{P}(X = x)\mathbf{P}(Y = y)$.

X et Z (et de même, Y et Z) sont indépendantes.

Cependant, X, Y et Z ne sont pas simultanément indépendantes.

Proposition 2.12

Soient $(X_i)_{i\in[n]}$ des variables aléatoires discrètes indépendantes de carré intégrable. Alors

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

Démonstration

Dans tous les cas (indépendance non supposée),

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

Or si les X_i sont indépendantes,

$$\mathrm{Cov}(X_i, X_j) = \mathbf{E}[X_iX_j] - \mathbf{E}[X_i]\mathbf{E}[X_j] = 0.$$

Application. La variance d'une variable aléatoire binomiale $\mathcal{B}(n,p)$ est $np(1-p)$. Posons $\Omega = \{P, F\}^n$, $\mathbf{P}(P) = p$ et $\mathbf{P}(F) = q = 1 - p$. Soit S le nombre de piles, qui suit $\mathcal{B}(n, p)$. On considère pour $i \in \mathbf{N}$, $X_i = \mathbb{1}_{A_i}$ avec A_i l'événement « le i-ième lancer donne pile ». Les X_i sont indépendantes et $S = \sum X_i$. Or $\mathbf{E}[S] = np$ et donc

$$\mathrm{Var} S = \sum_{i=1}^{n} \mathrm{Var}(X_i) = np(1-p).$$

2.4 Fonctions génératrices de variables aléatoires

Soient X et Y deux variables aléatoires discrètes définies sur un même espace de probabilité, à valeurs dans $\mathbf{N}$, indépendantes et soit $Z = X + Y$. Si on connait p_x la loi de X et p_y celle de Y. Peut-on écrire simplement celle de Z ? Soit $z \in \mathbf{N}$,

$$p_Z(z) = \mathbf{P}(Z = z) = \mathbf{P}(X + Y = z) = \mathbf{E}(\mathbb{1}_z(X + Y)).$$

Ainsi,

$$\begin{aligned} p_Z(z) &= \sum_{(x,y)\in\mathbf{N}^2} \mathbb{1}_z(x+y)p_{(X,Y)}(x,y) \\ &= \sum \mathbb{1}_{y=z-x} p_X(x)p_Y(z-x) \\ &= \sum p_X(x)p_Y(z-x). \end{aligned}$$

Définition 2.13

Soit X une variable aléatoire discrète à valeurs dans $\mathbf{N}$. Soit p_X la loi de X.

La *fonction génératrice*, $G_X(z)$, est la somme de la série

$$\sum_{x=0}^{\infty} p_X(x)z^x.$$

On note R_x le rayon de convergence de la série.

Proposition 2.14
On a :

1. $R_X \geq 1$ et la série $\sum p_X(x) z^x$ converge absolument pour tout $z \in [-1,1]$ et la somme de la série est
$$G_X(z) = \mathbf{E}[z^X].$$
2. G_X est continue sur $[-1,1]$ et C^∞ sur $]-1,1[$.
3. G_x (*i.e.* $G_x(z), z \in [-1,1]$) caractérise la loi de X, pour tout $n \in \mathbf{N}$:
$$p_X(n) = \frac{1}{n!} \frac{\mathrm{d}^n}{\mathrm{d}z^n} G_X(0).$$

Démonstration
Dans l'ordre :

1. Pour tout $r \leq 1$,
$$|p_X(x) r_x| \leq p_X(x) \leq 1$$
donc $R_X \geq 1$. Soit $z \in [-1,1]$, on a $|z^n p_X(x)| \leq p_X(x)$ or
$$\sum_{x=0}^{\infty} p_X(x) = 1$$
donc $\sum z^x p_X(x)$ est absolument convergente.
Donc
$$G_X(z) = \sum_{x=0}^{\infty} z^x p_X(x)$$
existe et est finie, de plus
$$\sum_{x=0}^{\infty} |z^x| \, p_X(x) < \infty$$
et donc z^x est intégrable.
$$\mathbf{E}(z^X) = \sum_{x=0}^{\infty} z^x p_X(x) = G_X(z).$$
2. On utilise le fait que G_X est C^∞ sur $]-R_X, R_X[$ et de plus $G'_X(z) = \sum x z^{x-1} p_X(x)$. Donc G_X est C^∞ sur $]-1,1[$. Montrons que G_X est continue à gauche en 1 et à droite en -1.
On a :
$$\forall x \in \mathbf{N}, \forall z \in [-1,1],\ |z^x p_X(x)| \leq p_X(x)\ ;\ z^x p_X(x) \underset{z \to 1^-}{\longrightarrow} p_X(x)$$
donc
$$\lim_{z \to 1^-} G_X(z) = G_X(1).$$
Donc G_X est continue à gauche en 1 et de même à droite en -1.
3. On a pour tout $z \in]-1,1[$,
$$\frac{\mathrm{d}^n}{\mathrm{d}z^n} G_X(z) = \sum_{x=0}^{\infty} \frac{\mathrm{d}^n}{\mathrm{d}z^n}(z^x) p_X(x)$$
mais $\mathrm{d}^n / \mathrm{d}z^n (z^x)$ en $z = 0$ est non nul à l'unique condition $x = n$. D'où
$$G_X(0) = \frac{\mathrm{d}^n}{\mathrm{d}z^n}(z^n) p_X(n) + 0.$$
Finalement,
$$\frac{\mathrm{d}^n}{\mathrm{d}z^n}(z^n) G_X(0) = n! p_X(n).$$

PROPOSITION 2.15
G_X est dérivable à gauche en 1 si, et seulement si, X est intégrable et dans ce cas $G'_X(1) = \mathbf{E}[X]$.

DÉMONSTRATION
Voir polycopié pour $R_X = 1$. Si $R_X > 1$ alors G_X est C^∞ sur $[-1, 1] \subset]-R_X, R_X[$. En particulier, G_X est dérivable et

$$\frac{\mathrm{d}}{\mathrm{d}z} G_X(z) = \sum_{x=1}^{\infty} x z^{x-1} p_X(x)$$

et donc

$$G'_X(1) = \sum_{x=1}^{\infty} x p_X(x) = \mathbf{E}[X].$$

PROPOSITION 2.16
G_X est $k \in \mathbf{N}^*$ fois dérivable à gauche en 1 si, et seulement si, X^k est intégrable et alors

$$G_X^{(k)}(1) = \mathbf{E}[X(X-1) \times \ldots \times (X - 1 + k)].$$

EXEMPLES.

1. Pour $X \sim \mathcal{B}(p)$ avec $p \in [0, 1]$,
$$G_X(z) = z^0 \mathbf{P}(X = 0) + z^1 \mathbf{P}(X = 1) = 1 - p + pz.$$
2. Pour $X =\sim \mathcal{B}(n, p)$ pour $n \in \mathbf{N}^*$ et $p \in [0, 1]$,
$$G_X(z) = \sum_{k=0}^{n} z^k \mathbf{P}(X = k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} z^k = (1 - p + pz)^n.$$
Ici
$$G'_X(z) = np(1 - p + pz)^{n-1},\ G'_X(1) = np = \mathbf{E}[X].$$
3. Pour $X \sim \mathcal{G}(p)$ avec $p \in]0, 1]$. On a
$$G_X(z) = \sum_{k=0}^{+\infty} z^k \mathbf{P}(X = k) = \sum_{k=1}^{+\infty} z^k p(1-p)^{k-1} = pz \sum_{k=1}^{+\infty} (z(1-p))^{k-1} = \frac{pz}{1 - (1-p)z}.$$
4. Pour $X \sim \mathcal{P}(\lambda)$ avec $\lambda \in]0, +\infty[$. On a
$$G_X(z) = \sum_{k=0}^{+\infty} z^k \mathbf{P}(X = k) = \sum_{k=0}^{+\infty} z^k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda z} = e^{\lambda(z-1)}.$$

2.5 Fonctions génératrices de vecteurs aléatoires

DÉFINITION 2.17
Soit $X = (X_1, \ldots, X_n)$ tel que les X_i sont définies sur le même espace de probabilité et X_i à valeurs dans $\mathbf{N}$ pour tout $i \in \{1, \ldots, n\}$.
La *fonction génératrice*, G_X, de X est la somme de la série entière :

$$G_X(z) = \sum_{x_1, \ldots, x_n \in \mathbf{N}} p_X(x_1, \ldots, x_n) z_1^{x_1} \ldots z_n^{x_n}$$

avec $z = (z_1, \ldots, z_n) \in \mathbf{R}^n$.

PROPOSITION 2.18
1. G_X est bien définie et continue sur $D = [-1,1]^n$;
2. G_X est C^∞ sur $\mathring{D} =]-1,1[^n$;
3. G_X caractérise la loi p_X de X :

$$\forall x_1, \dots, x_n \in \mathbf{N},\ p_X(x_1, \dots, x_n) = \frac{1}{x_1! \dots x_n!} \frac{\mathrm{d}^{x_1}}{\mathrm{d}z_1^{x_1}} \dots \frac{\mathrm{d}^{x_n}}{\mathrm{d}z_n^{x_n}} G_X(z_1, \dots, z_n)\Big|_{z_1 = \dots = z_n = 0} ;$$

4. on peut retrouver la fonction génératrice G_{X_1} de X_1 à partir de la fonction génératrice, G_X, de X :

$$\forall z \in [-1,1],\ G_{X_1}(z) = G_X(z, 1, \dots, 1).$$ [§2]

PROPOSITION 2.19
Soit $X = (X_1, \dots, X_n)$ est un vecteur aléatoire avec X_i des variables aléatoires entières. Les X_i sont indépendantes si, et seulement si, la fonction génératrice de X est factorisable, *i.e.* pour tout $(z_1, \dots, z_n) \in D$,

$$G_X(z_1, \dots, z_n) = \prod_{i=1}^n G_{X_i}(z_i).$$

DÉMONSTRATION
Si les X_i sont indépendantes,

$$\begin{aligned} G_X(z_1, \dots, z_n) &= \mathbf{E}[z_1^{X_1} \dots z_n^{X_n}] \\ &= \mathbf{E}\left[\prod_{i=1}^n f_i(X_i)\right] \\ &= \prod_{i=1}^n \mathbf{E}[f_i(X_i)] \\ &= \prod_{i=1}^n G_{X_i}(z_i). \end{aligned}$$

Avec $f_i(x) = z_i^x$.
Réciproquement, il faut montrer que $p_X(x_1, \dots, x_n) = \prod p_{X_i}(x_i)$. On a :

$$\begin{aligned} p_X(x_1, \dots, x_n) &= \frac{1}{x_1! \dots x_n!} \frac{\mathrm{d}^{x_1}}{\mathrm{d}z_1^{x_1}} \dots \frac{\mathrm{d}^{x_n}}{\mathrm{d}z_n^{x_n}} G_X(z_1, \dots, z_n)\Big|_{z_1 = \dots = z_n = 0} \\ p_{X_i}(x_i) &= \frac{1}{x_i!} \frac{\mathrm{d}^{x_i}}{\mathrm{d}z_i^{x_i}} G_{X_i}(z_i)\Big|_{z_i = 0}. \end{aligned}$$

Ici :

$$\begin{aligned} p_X(x_1, \dots, x_n) &= \frac{1}{x_1! \dots x_n!} \frac{\mathrm{d}^{x_1}}{\mathrm{d}z_1^{x_1}} \dots \frac{\mathrm{d}^{x_n}}{\mathrm{d}z_n^{x_n}} \prod_{i=1}^n G_{X_i}(z_i)\Big|_{z_1 = \dots = z_n = 0} \\ &= \prod_{i=1}^n p_{X_i}(x_i). \end{aligned}$$

§2. En effet, on remarque que

$$G_X(z_1, \dots, z_n) = \mathbf{E}[z_1^{X_1} \dots z_n^{X_n}].$$

Corollaire 2.20
Soit $X = (X_1, \dots, X_n)$ est un vecteur aléatoire avec X_i des variables aléatoires entières. Si les X_i sont indépendantes alors $Y = \sum X_i$ est une variable aléatoire entière donc la fonction génératrice, G_Y, est

$$\forall z \in [-1,1],\ G_Y(z) = \prod_{i=1}^{n} G_{X_i}(z).$$

Démonstration

$$\begin{aligned} G_Y(z) &= \mathbf{E}[z^Y] \\ &= \mathbf{E}[z^{X_1} \dots z^{X_n}] \\ &= G_X(z, \dots, z) \\ &= G_{X_1}(z) \dots G_{X_n}(z). \end{aligned}$$

Exemple. Soit $X \sim \mathcal{B}(n,p)$ pour $n \in \mathbf{N}^*$ et $p \in [0,1]$. On sait que $X = \sum X_i$ pour $X_i \sim \mathcal{B}(p)$ et les X_i sont indépendantes. Donc

$$G_X(z) = \prod_{i=1}^{n} G_{X_i}(z) = G_{X_1}(z)^n = (1 - p + pz)^n.$$

Remarque pratique. Pour identifier la loi d'une variable aléatoire entière on peut essayer la stratégie suivante : 1. on calcule la fonction génératrice de X ; 2. on essaie d'identifier la loi qui correspond à la fonction génératrice trouvée.

Exemple. Soient $X \sim \mathcal{P}(\lambda)$, $Y \sim \mathcal{P}(\mu)$ avec $\lambda, \mu \in]0, +\infty[$. X et Y sont indépendantes. Soit $Z = X + Y$, on veut la loi de Z. On peut procéder « brutalement » :

$$\begin{aligned} p_Z(z) &= \mathbf{P}(Z = z) \\ &= \sum_{k=0}^{z} p_X(k) p_Y(z-k) \\ &= \sum_{k=0}^{z} e^{-\lambda} \frac{\lambda^k}{k!} e^{-\mu} \frac{\mu^{z-k}}{(z-k)!}. \end{aligned}$$

On peut également remarquer :

$$G_X(z) = e^{\lambda(z-1)},\ G_y(z) = e^{\mu(z-1)}$$

d'où

$$G_Z(z) = G_X(z) G_Y(z) = e^{(\lambda+\mu)(z-1)}$$

et on reconnait la fonction génératrice de la loi $\mathcal{P}(\lambda + \mu)$ donc $Z \sim \mathcal{P}(\lambda + \mu)$.

Chapitre 3

Variables aléatoires réelles continues

1 Variables aléatoires réelles

Définition 1.1
Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité. Soit $X : \Omega \to \mathbf{R}$. X est une *variable aléatoire réelle* si pour tout intervalle I de $\mathbf{R}$,

$$\{\omega \in \Omega \,|\, X(\omega) \in I\} \in \mathcal{A}$$

et on note $\{X \in I\}$ cet ensemble.

Cette définition permet de pouvoir considérer $\mathbf{P}(X \geq 0)$, $\mathbf{P}(X \in I)$ pour I un intervalle.

Remarque. Soit X une variable aléatoire discrète réelle, alors X est une variable aléatoire réelle au sens de la définition précédente.

Proposition 1.2
Soit $X : \Omega \to \mathbf{R}$.

1. X est une variable aléatoire réelle dès que $\{X \leq x\} \in \mathcal{A}$ pour tout $x \in \mathbf{R}$.
2. Si X est une variable aléatoire réelle, alors $\{X \in B\} \in \mathcal{A}$ pour tout $B \in \mathcal{B}(\mathbf{R})$.

Remarque. En pratique, on montre que pour tout x, $\{X \leq x\} \in \mathcal{A}$ et on a alors $\{X \in B\} \in \mathcal{A}$ pour tout $B \in \mathcal{B}(\mathbf{R})$.

Démonstration
On montre le premier point.

1. On remarque que
$$\left\{X \in \bigcup_{i \in \mathbf{N}} A_i\right\} = \bigcup_{i \in \mathbf{N}} \{X \in A_i\}$$
et
$$\{X \in A^c\} = \{X \in A\}^c .$$
On sait que $\{X \leq x\} \in \mathcal{A}$ pour tout $x \in \mathbf{R}$. Mais tous les intervalles de $\mathbf{R}$ peuvent s'écrire en fonctions d'intervalles de la forme $]-\infty, x]$. Par exemple,
$$]a, b[= \bigcup_{n \in \mathbf{N}}]a, b - 1/n].$$

Exemple. Soit X une variable aléatoire réelle et soit $f : \mathbf{R} \to \mathbf{R}$ continue par morceaux. Alors $f(X) = Y$ est une variable aléatoire réelle.

En effet, soit $y \in \mathbf{R}$, on veut montrer que $\{Y \leq y\} \in \mathcal{A}$.

$$\begin{aligned}\{Y \leq y\} &= \{f(X) \leq y\} \\ &= \left\{X \in f^{-1}(]-\infty, y])\right\}\end{aligned}$$

mais il existe des f_i continues et $(I_i)_{i\in J}$ une partitition de $\mathbf{R}$ telles que

$$f(x) = \sum_{i\in J} 1_{I_i}(x) f_i(x)$$

et alors

$$\begin{aligned}\{Y \leq y\} &= \bigcup_{i\in J} \{X \in I_i\}\, X \in f_i^{-1}(]-\infty, y]) \\ &= \bigcup_{i\in J} \{X \in I_i\} \bigcap \left\{X \in f_i^{-1}(]-\infty, y])\right\} \in \mathcal{A}\end{aligned}$$

Définition 1.3

Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité. Soit $X : \Omega \to \mathbf{R}$ une variable réelle. La loi, Q, de X est la probabilité sur $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ définie par

$$Q(B) = \mathbf{P}(X \in B)$$

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

Démonstration

Q est bien une probabilité.

1. Q est définie sur $\mathcal{B}(\mathbf{R})$;
2. $Q(\mathbf{R}) = 1$;
3. si $(B_i)_{i\in J}$ est une famille au plus dénombrable de boréliens deux à deux disjoints, alors

$$\begin{aligned}Q\left(\bigcup_{i\in J} B_i\right) &= \mathbf{P}\left(X \in \bigcup_{i\in J} B_i\right) \\ &= \mathbf{P}\left(\bigcup_{i\in J} \{X \in B_i\}\right) \\ &= \sum_{i\in J} \mathbf{P}(X \in B_i) \\ &= \sum_{i\in J} Q(B_i).\end{aligned}$$

On veut montrer que la loi d'une variable réelle est caractérisée par la donnée de $Q(]-\infty, x])$ pour $x \in \mathbf{R}$.

Définition 1.4

Soit $F : \mathbf{R} \to [0, 1]$. On dit que F est une *fonction de répartition* si elle satisfait les propriétés suivantes :

1. F est croissante ;
2. $\lim_{x\to-\infty} F(x) = 0, \lim_{x\to+\infty} F(x) = 1$;
3. F est continue à droite.

Proposition 1.5
Si X est une variable aléatoire réelle alors $F_X : \mathbf{R} \to [0,1]$ définie par $F_X(x) = \mathbf{P}(X \leq x)$ est une fonction de répartition.

Exemple. Si $X \sim \mathcal{B}(p)$ alors

$$F_X(x) = \mathbf{P}(X \leq x) = \begin{cases} 0 \text{ si } x \in]-\infty, 0[\\ 1-p \text{ si } x \in [0,1[\\ 1 \text{ sinon} \end{cases}$$

Proposition 1.6
On a :

1. Soit Q une probabilité sur $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. Si on pose $F(x) = Q(]-\infty, x])$ pour tout x réel alors F est une fonction de répartition.
2. Soit $F : \mathbf{R} \to [0,1]$ une fonction de répartition. Alors il existe une unique probabilité Q sur $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ telle que $Q(]-\infty, x]) = F(x)$ pour tout x réel.

Démonstration
Dans l'ordre :

1. (a) $F : \mathbf{R} \to [0,1]$ est croissante. En effet, pour $x \leq y$ on a $]-\infty, x] \subset]-\infty, y]$ et donc $F(x) = Q(]-\infty, x]) \leq Q([-\infty, y]) = F(y)$.
   (b) Soit (x_n) une suite tendant vers $-\infty$. On montre que $F(x_n)$ tend vers 0. On pose $A_n =]-\infty, x_n]$. Si x_n est décroissante, alors A_n est une suite décroissante d'événement et d'intersection vide. On a alors $\lim Q(A_n) = Q(\emptyset) = 0$. Si X_n n'est pas décroissante, soit $\varepsilon > 0$. Il existe N tel que $Q(]-\infty, -N]) \leq \varepsilon$ et il existe M tel que $n \geq M$ implique $x_n \leq -N$ et alors $Q(]-\infty, x_n]) \leq Q(]-\infty, -N]) \leq \varepsilon$.
   (c) Soit $x \in \mathbf{R}$, on pose $A_n =]-\infty, x + 1/n]$ une suite décroissante d'événements. L'intersection de A_n est $]-\infty, x]$ et alors $Q(A_n)$ tend vers $Q(]-\infty, x])$ donc $F(x+1/n)$ tend vers $F(x)$.
   Soit (h_n) une suite de termes positifs tendant vers 0. On veut montrer que $F(x+h_n)$ tend vers $F(x)$. Soit $\varepsilon > 0$, il existe N tel que $|F(x+1/N) - F(x)| \leq \varepsilon$. Mais F est croissante donc il existe M tel que $h_n \leq 1/N$ pour tout $n \geq M$ et

$$F(x) \leq F(x+h_n) \leq F(x+1/N) \leq F(x) + \varepsilon$$

   donc c'est vérifié.
2. Admis.

Corollaire 1.7
Soit X une variable réelle, alors sa loi est caractérisée par sa fonction de répartition $F(x) = \mathbf{P}(X \leq x)$.

Définition 1.8
Une fonction $p : \mathbf{R} \to \mathbf{R}_+$ est appelée densité (de probabilité) si elle est continue par morceaux (avec un nombre fini de points de discontinuité) et telle que $\int_{-\infty}^{+\infty} p(x)\,\mathrm{d}x = 1$.

Définition 1.9
Une loi à densité est une probabilité Q sur $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ dont la fonction de répartition est de la forme

$$F(x) = \int_{-\infty}^{x} p(y)\,\mathrm{d}y,$$

où p est une densité et pour tout intervalle I de $\mathbf{R}$ (pour tout borélien de $\mathbf{R}$), on a

$$Q(I) = \int_I p(y)\,\mathrm{d}y.$$

Définition 1.10
Une variable réelle est dite à densité si sa loi est à densité.

Proposition 1.11
Soit X une variable réelle à densité p.

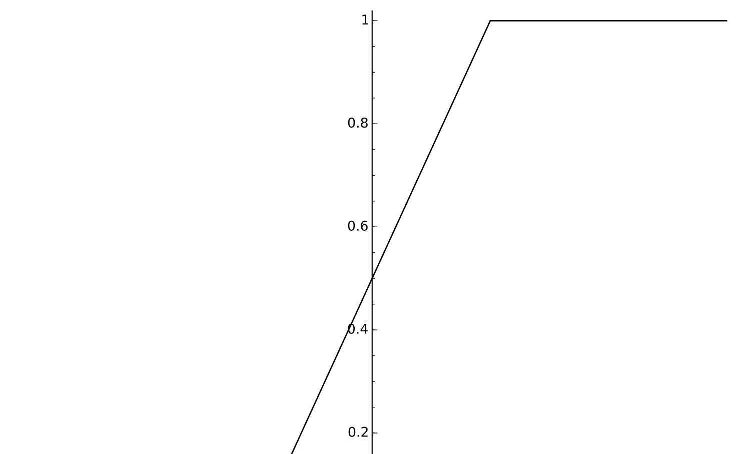
1. La fonction de répartition de X est continue et dérivable en tout point de continuité de la densité.
2. La loi de X est diffuse :
$$\forall x \in \mathbf{R},\ \mathbf{P}(X = x) = 0.$$
3. Pour tous $a, b \in \mathbf{R}$ avec $a < b$, on a :
$$\mathbf{P}(X \in [a,b]) = \mathbf{P}(x \in]a,b[) = \int_a^b p(x)\,\mathrm{d}x.$$

Démonstration
Montrons le second point. Soit $x \in \mathbf{R}$. Pour $A_n =]x - 1/n, x + 1/n[$ une suite décroissante on a $\{x\} = \bigcap A_n$. On a alors

$$\mathbf{P}(X = x) = \lim_{n\to+\infty} \mathbf{P}(X \in A_n) = \lim_{n\to+\infty} F(x + 1/n) - F(x - 1/n) = 0.$$

Remarque. La densité au point x, $p(x)$, n'est pas $\mathbf{P}(X = x)$.

Exemple, loi uniforme. La loi uniforme, $\mathcal{U}(a,b)$, sur $[a,b]$ avec $a < b$ deux réels. $\mathcal{U}([-1,1])$ a pour fonction de répartition :



Pour $\mathcal{U}(a,b)$ la fonction de répartition, F, est :

$$F(x) = \begin{cases} 0 \text{ si } x < a \\ \dfrac{x-a}{b-a} \text{ si } a \leq x < b \\ 1 \text{ si } x \geq b \end{cases}$$

Cette loi est à densité p :

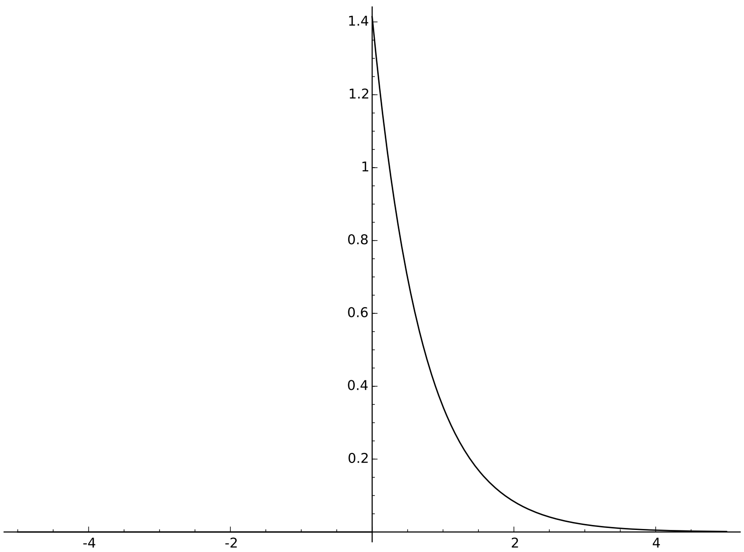
$$p(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x).$$

Une variable aléatoire uniforme, X, sur $[a,b]$ est une variable réelle dont la loi est $\mathcal{U}(a,b)$. Si $X \sim \mathcal{U}(0,1)$ alors $\mathbf{P}(X \in [c,d]) = d - c$.

Exemple, loi exponentielle. La loi exponentielle de paramètre $\lambda \in \mathbf{R}_+^*$, notée $\mathcal{E}(\lambda)$, c'est la loi à densité p :

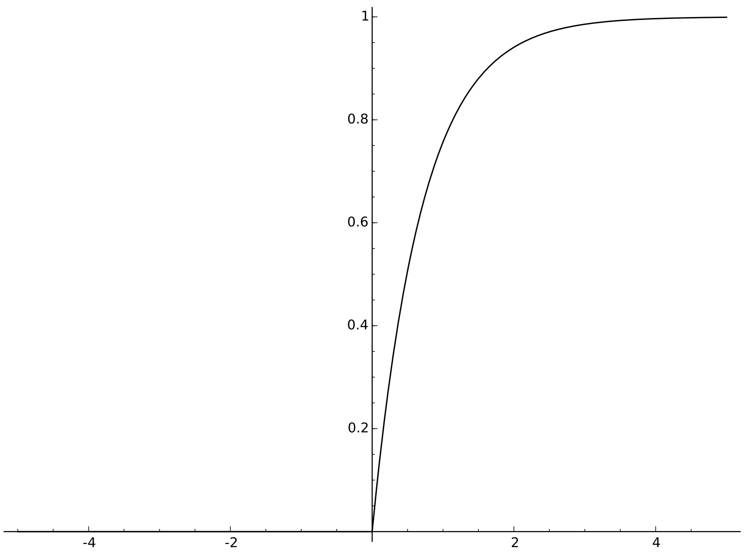
$$p(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0,+\infty[}(x)$$

qui vérifie $\int_{\mathbf{R}} p(x)\,\mathrm{d}x = 1$.



La fonction de répartition est alors

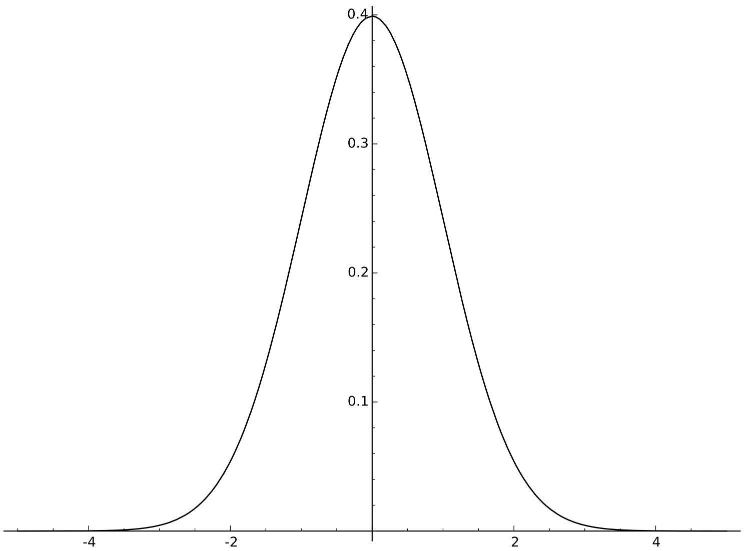
$$F(x) = \begin{cases} 0 \text{ si } x < 0 \\ 1 - e^{-\lambda x} \text{ si } x \geq 0 \end{cases}$$



Un variable, X, exponentielle de paramètre λ est une variable aléatoire réelle de loi $\mathcal{E}(\lambda)$, on note $X \sim \mathcal{E}(\lambda)$.

Exemple, loi gaussienne. On l'appelle aussi loi normale, c'est la loi à densité :

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



La fonction de répartition associée n'admet pas d'expression en termes simples.

Proposition 1.12
Une variable aléatoire réelle, X, est à densité si, et seulement si, sa fonction de répartition est de la forme

$$F(x) = \int_{-\infty}^{x} p(y)\, \mathrm{d}y$$

où p est une densité.

En pratique. Pour montrer qu'une variable aléatoire réelle est à densité :

1. On calcule sa fonction de répartition.
2. On dérive F et on prend $p = F'$ et on vérifie que $F(x) = \int_{-\infty}^{x} p(y)\, \mathrm{d}y$.

EXEMPLE. Soit X une variable aléatoire réelle à densité p. Soit $f : \mathbf{R} \to \mathbf{R}$ continue par morceaux, on pose $Y = f(X)$. On a déjà que Y est une variable aléatoire réelle, est-ce que Y est à densité ? Généralement non ($f = 1$). Mais avec $X \sim \mathcal{U}(0,1)$ et

$$f(x) = \begin{cases} -\ln x \text{ si } x \in]0,1[\\ 0 \text{ sinon} \end{cases}$$

alors pour $Y = f(X)$ on a :

$$F_Y(y) = \mathbf{P}(Y \leq y) = \mathbf{P}(-\ln X \leq y) = \begin{cases} 0 \text{ si } y < 0 \\ \mathbf{P}(X \geq e^{-y}) = \displaystyle\int_{e^{-y}}^{1} 1 \,\mathrm{d}z = 1 - e^{-y} \text{ si } y \geq 0 \end{cases}$$

F_Y est bien continue par morceaux et dérivable et pour $y < 0$, $F_Y'(0) = 0$ et pour $y > 0$, $F_Y'(y) = e^{-y}$. Le candidat comme densité est

$$p(y) = e^{-y} \mathbb{1}_{[0,+\infty[}(y).$$

On vérifie que $F_Y(y) = \int_{-\infty}^{y} p(z) \,\mathrm{d}z$: Y est une variable aléatoire à densité p et $Y \sim \mathcal{E}(1)$.

1.1 Espérance

REMARQUE. On peut définir l'espérance pour une variable aléatoire réelle arbitraire (mais cela nécessite la théorie de la mesure). Ici on va juste définir la notion d'espérance pour des variables aléatoires réelles à densité.

DÉFINITION 1.13
Soit X une variable aléatoire réelle à densité p. Soit $\varphi : \mathbf{R} \to \mathbf{R}$ continue par morceaux telle que

$$\int_{-\infty}^{+\infty} |\varphi(x)| \, p(x) \,\mathrm{d}x < \infty.$$

On dit alors que la variable aléatoire réelle $\varphi(X)$ est intégrable et

$$\mathbf{E}[\varphi(X)] = \int_{-\infty}^{+\infty} \varphi(x) p(x) \,\mathrm{d}x.$$

Dans le cas $\varphi = \mathrm{id}$, et si $\int_{\mathbf{R}} |x| \, p(x) \,\mathrm{d}x < \infty$ alors on dit que X est intégrable et

$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} x p(x) \,\mathrm{d}x.$$

On vérifie que cette définition est compatible avec celle donnée pour les variables aléatoires discrètes. Soit X une variable aléatoire réelle à densité p, soit φ une fonction constante par morceaux :

$$\varphi : x \mapsto \sum_{i=1}^{n} \alpha_i \mathbb{1}_{I_i}(x)$$

avec $(I_i)_{1 \leq i \leq n}$ une partition de $\mathbf{R}$ et les $\alpha_i \in \mathbf{R}$.

$$\varphi(X) = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{I_i}(X)$$

est une variable aléatoire réelle, c'est même une variable aléatoire de BERNOULLI. Donc $\varphi(X)$ est une variable aléatoire discrète. On peut calculer son espérance des deux manières :

1. En tant que variable discrète c'est :
$$\mathbf{E}[\varphi(X)] = \sum_{i=1}^{n} \alpha_i \mathbf{E}[\mathbb{1}_{I_i}(X)] = \sum_{i=1}^{n} \alpha_i p_i$$
où $p_i = \mathbf{P}(X \in I_i)$.
2. En tant que variable réelle :
$$\mathbf{E}[\varphi(X)] = \int_{-\infty}^{+\infty} \varphi(x)p(x)\,\mathrm{d}x = \sum_{i=1}^{n} \alpha_i \int_{-\infty}^{+\infty} \mathbb{1}_{I_i}(x)p(x)\,\mathrm{d}x = \sum_{i=1}^{n} \alpha_i \mathbf{P}(X \in I_i).$$

Remarque. Si $\varphi(x) = a$ pour tout $x \in \mathbf{R}$ alors $\mathbf{E}[\varphi(X)] = a$ car
$$\int_{-\infty}^{+\infty} ap(x)\,\mathrm{d}x = a.$$

Proposition 1.14
On a :
1. Soit X une variable aléatoire réelle à densité p. Soient $\varphi_1, \varphi_2 : \mathbf{R} \to \mathbf{R}$ continues par morceaux telles que
$$\int_{\mathbf{R}} |\varphi_i(x)|\, p(x)\,\mathrm{d}x < \infty$$
pour $i = 1, 2$.
Soient $\alpha, \beta \in \mathbf{R}$ alors $Y = \alpha\varphi_1(X) + \beta\varphi_2(X)$ est une variable réelle intégrable et
$$\mathbf{E}[Y] = a\mathbf{E}[\varphi_1(X)] + \beta\mathbf{E}[\varphi_2(X)].$$
2. Soit X une variable aléatoire réelle à densité p, soit $\varphi : \mathbf{R} \to \mathbf{R}_+$ telle que $\varphi(X)$ est intégrable. Alors $\mathbf{E}[\varphi(X)] \geq 0$ et si $\mathbf{E}[\varphi(X)] = 0$ alors $\mathbf{P}(\varphi(X) = 0) = 1$.

Démonstration
On montre le second point. Comme $\varphi(x)p(x)$ est positif,
$$\int_{-\infty}^{+\infty} \varphi(x)p(x)\,\mathrm{d}x \geq 0.$$
Si de plus, $\mathbf{E}[\varphi(X)] = 0$ alors on pose $A_n = \{\varphi(X) \geq 1/n\}$, cela constitue une suite d'événements croissante. On a $\bigcup A_n = \{\varphi(X) > 0\}$. $A_n = \{\varphi(X) \geq 1/n\} = \{\Psi(\varphi(x)) \geq 0\}$.
$$\mathbf{P}(A_n) = \mathbf{E}[\mathbb{1}_{\varphi(X) \geq 1/n}] = \mathbf{E}[\Psi(\varphi(X))] \leq \mathbf{E}[\Theta(\varphi(X))] = n\mathbf{E}[\varphi(X)] = 0.$$
Donc $\mathbf{P}(\varphi(X) > 0) = \lim_{n\to+\infty} \mathbf{P}(A_n) = 0$ et $\mathbf{P}(\varphi(X) = 0) + \mathbf{P}(\varphi(X) > 0) = 1$ donc $\mathbf{P}(\varphi(X) = 0) = 1$.

Définition 1.15
Soit $k \in \mathbf{N}^*$. Soit X une variable aléatoire réelle à densité p. Le k-ième moment de X existe si
$$\int_{\mathbf{R}} |x|^k\, p(x)\,\mathrm{d}x < \infty$$
et alors
$$\mathbf{E}[X^k] = \int_{\mathbf{R}} x^k p(x)\,\mathrm{d}x.$$