Introduction aux Probabilités discrètes

林琴　Sorbonne Université

整理：钟启帆　校对：王道越

Préface

Ouvrages de référence

1) C.Deschamps, F.Maulin etc **Maths MPSI tout-en-un**, 2018
2) C.Deschamps, F.Maulin, A.Warrasfel etc **Mathématiques tout en un**, 2016

Errata

Si vous trouvez des erreurs, n'hésitez pas à nous contacter.
e-mail : wdy_math@mail.ustc.edu.cn

Table des matières

Chapitre 1

Ensembles, dénombrement et dénombrabilité

1.1 Opératious sur les ensembles

Soit E un ensemble, on ecrit $x \in E$: « x approtient à E ». Un ensemble A tel que $A \subset E$ est appelé un **sous-ensemble** ou une **partie de** E.

L'ensemble qui ne contient aucun élément est appelé l'ensemble vide, noté $\varnothing$.

Soit A, $B \subset E$, on définit

— Union $\quad A \cup B := \{\omega \in E : \omega \in A \text{ ou } \omega \in B\}$,

— Intersection $\quad A \cap B := \{\omega \in E : \omega \in A \text{ et } \omega \in B\}$,

— Complimentaire $\quad A^C := \{\omega \in E : \omega \notin A\}$,

— Différence de A avec B $\quad A \backslash B := A \cap B^C$,

— Différence symétrique $\quad A \triangle B := (A \backslash B) \cup (B \backslash A)$.

Attention! $\quad A \backslash B \neq B \backslash A$.

Soit E un ensemble, I un ensemble d'indice arbitraire, et $(A_i)_{i \in I}$ une famille de sous-ensembles de E, on définit

— Union $\quad \bigcup\limits_{i \in I} A_i := \{\omega \in E : \exists i \in I, \text{ tel que } \omega \in A_i\}$,

— Intersection $\quad \bigcap\limits_{i \in I} A_i := \{\omega \in E : \forall i \in I, \omega \in A_i\}$.

$\bigcup\limits_{i \in I} A_i$ et « disjointe » si les $(A_i)_{i \in I}$ sont 2 à 2 disjoint. (c-à-d : $A_i \cap A_j = \varnothing$ si $i \neq j$, i, $j \in I$) Dans ce cas, on écrit aussi « $\bigsqcup\limits_{i \in I} A_i$ ».

Disbutivité :

— $\left(\bigcup\limits_{i \in I} A_i\right) \bigcap B = \bigcup\limits_{i \in I} (A_i \cap B)$,

— $\left(\bigcap\limits_{i \in I} A_i\right) \bigcup B = \bigcap\limits_{i \in I} (A_i \cap B)$,

— $\left(\bigcap\limits_{i \in I} A_i\right)^C = \bigcup\limits_{i \in I} A_i^C$,

— $\left(\bigcup\limits_{i \in I} A_i\right)^C = \bigcap\limits_{i \in I} A_i^C$.

Produit cartésien :

$$A \times B := \{ \underbrace{(a, b)}_{\text{couple ordonné}} : a \in A \text{ et } b \in B\}.$$

$n \geqslant 2$ entier

$$A^n := \text{ le produit } \underbrace{A \times \cdots \times A}_{n \text{ fois}}.$$

$\mathcal{P}(E) := \{A : A \subset E\}$ est appelé l'ensemble de tous les sous-ensembles de E. C'est-à-dire

$$A \in \mathcal{P}(E) \Leftrightarrow A \subset E.$$

$\forall A \in \mathcal{P}(E)$, on définit **la fonction indicatrice de A**

$$\mathbb{1}_A : E \to \mathbb{R}$$
$$x \mapsto \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \in A^C \end{cases}.$$

Définition 1.1.1 (Injective, surjective, bijective) $f : A \to B$ une application est

— **injective** si tout élément de B est l'image d'au plus un élément de A,

— **surjective** si tout élément de B est l'image d'au moins un élément de A,

— **bijective** si elle est injective et surjective.

Si f est bijective, on peut définir l'application inverse $f^{-1} : B \to A$.

Attention! Pour une partie S de B, **l'image réciproque de** S **par** f est l'ensemble

$$f^{-1}(S) := \{\omega \in A : f(\omega) \in S\}.$$

1.2 Ensembles finis

Définition 1.2.1 Eensemble fini, cardinal] Un ensemble E est **finie** s'il existe un entier $n \geqslant 0$ et une bijection de $[\![1, n]\!]$ sur E. (convention « $[\![1, 0]\!] = \varnothing$ »)

Cet entier n est appelé **cardinal** de E, on encore nombre d'élément de E. On note

$$n = \text{Card}\,(E) = |E| = \#E.$$

Notation

— $\mathbb{N} = \{0, 1, 2, \dots\}$ l'ensemble des entiers naturels,

— $\mathbb{N}^* = \mathbb{N}\backslash\{0\}$,

— $\mathbb{Z}$ l'ensemble des entiers relatifs,

— $\mathbb{Q}$ l'ensemble des nombres rationnels,

— $\mathbb{R}$ l'ensemble des nombres réels,

— $\mathbb{C}$ l'ensemble des nombres complexes.

Lemme 1.2.1 Le cardinal d'un ensemble fini E est bien défini.

Démonstration. Supposons que E soit un ensemble fini. Par définition, il existe au moins un entier naturel n et une bijection $f : [\![1, n]\!] \to E$. Supposons qu'il existe un autre entier naturel m et une autre bijection $g : [\![1, m]\!] \to E$.

Considérons l'application composée

$$h = g^{-1} \circ f : [\![1, n]\!] \to [\![1, m]\!].$$

Puisque f et g (et donc g^{-1}) sont des bijections, h est également une bijection.

($\leqslant$) Comme h est injective, les images $h(1), h(2), \ldots, h(n)$ sont n éléments distincts de $[\![1, m]\!]$. Le cardinal de $[\![1, m]\!]$ étant m, on en déduit nécessairement que $n \leqslant m$.

($\geqslant$) Réciproquement, considérons l'application réciproque $h^{-1} = f^{-1} \circ g : [\![1, m]\!] \to [\![1, n]\!]$. Celle-ci est aussi une bijection. Par le même argument, l'injectivité de h^{-1} implique que $m \leqslant n$.

Des deux inégalités $n \leqslant m$ et $m \leqslant n$, on conclut que $n = m$. Ainsi, l'entier naturel n associé à toute bijection d'un intervalle entier sur E est unique. Ceci prouve que le cardinal de E est bien défini. □

Remarque Soit E un élément fini de cardinal $n \geqslant 1$.

Une bijection $i \in [\![1, n]\!] \to a_i \in E$ permet de numérotir les éléments de E et d'écrire $E = \{a_1, \ldots, a_n\}$.

Définition 1.2.2 (Ensemble infini) Un ensemble E est **infini** s'il n'est pas fini.

Proposition 1.2.1 Si A et B sont des ensembles finis, alors $A \cup B$ l'est aussi et l'on a la formule

$$\mathrm{Card}\,(A \cup B) = \mathrm{Card}\,(A) + \mathrm{Card}\,(B) - \mathrm{Card}\,(A \cap B)\,.$$

Démonstration. On a $A \cup B = A \cup \underbrace{(B \backslash A \cap B)}_{=B \backslash A}$. Or A et $B' := B \backslash A \cap B$ sont disjoints. Ils sont des ensembles finis.

Si $[\![1, n_1]\!] \xrightarrow{f} A$, $[\![1, m_1]\!] \xrightarrow{g} B'$ deux bijections, alors

$$[\![1, n_1 + m_1]\!] \xrightarrow{h} A \cup B' = A \sqcup B'$$

$$a \mapsto \begin{cases} f(a) \in A & \text{si } a \leqslant n_1 \\ g(a - n_1) \in B' & \text{si } a > n_1 \end{cases}$$

est une bijection.

Donc $\mathrm{Card}\,(A \cup B) = n_1 + m_1 = \mathrm{Card}\,(A) + \mathrm{Card}\,(B')$.

Mais $B = (A \cap B) \sqcup B'$. Par le même argument ci-dessus, on sait que $\mathrm{Card}\,(B) = \mathrm{Card}\,(A \cap B) + \mathrm{Card}\,(B')$.

Par conséquent, $\mathrm{Card}\,(A \cup B) = \mathrm{Card}\,(A) + \mathrm{Card}\,(B) - \mathrm{Card}\,(A \cap B)$. □

Proposition 1.2.2 (Formule d'inclusion-exclusion) Soient $A_1, \dots, A_n$ des parties d'un ensemble E fini, alors

$$\mathrm{Card}\,(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \sum_{1 \leqslant r_1 < \cdots < r_i \leqslant n} (-1)^{i-1} \mathrm{Card}\,(A_{r_1} \cap \cdots \cap A_{r_i})\,.$$

Remarque Rappel :

— $\mathrm{Card}\,(A_1 \cup A_2) = \mathrm{Card}\,(A_1) + \mathrm{Card}\,(A_2) - \mathrm{Card}\,(A_1 \cap A_2)$,

— $\mathrm{Card}\,(A_1 \cup A_2 \cup A_3) = \mathrm{Card}\,(A_1) + \mathrm{Card}\,(A_2) + \mathrm{Card}\,(A_3) - \mathrm{Card}\,(A_1 \cap A_2) - \mathrm{Card}\,(A_2 \cap A_3) - \mathrm{Card}\,(A_1 \cap A_3) + \mathrm{Card}\,(A_1 \cap A_2 \cap A_3)$.

Démonstration. On peut montrer que

$$\mathbb{1}_{A_1 \cup \cdots \cup A_n} = \sum_{j=1}^{n} \sum_{J \subset [\![1,n]\!], |J|=j} (-1)^{j-1} \mathbb{1}_{\bigcap_J A_i}.$$

En prenant l'espérance des deux côtés, on obtient la formule d'inclusion-exclusion.

On peut aussi démontrer cette formule par récurrence sur n. □

Notation Soit $f : E \to F$, $B \subset F$ **l'image réciproque de B par f** :

$$f^{-1}(B) := \{x \in E : f(x) \in B\}.$$

Si $A \subset E$, **l'image directe de A par f** :

$$f(A) := \{f(a) : a \in A\}.$$

Proposition 1.2.3 Soient E et $\mathcal{F}$ deux ensembles avec E fini. Si $f : E \to F$ est une application, alors $f(E)$ est fini et $\mathrm{Card}\,(f(E)) \leqslant \mathrm{Card}\,(E)$.

De plus, $\mathrm{Card}\,(f(E)) = \mathrm{Card}\,(E)$ si et seulement si f est injective.

Démonstration. Pour chaque $y \in f(E)$, on choisit un $x_y \in f^{-1}(y)$.

Alors l'application

$$\begin{aligned} g : f(E) \to E' := \{x_y \in E : y \in f(E)\} \subset E \\ y \mapsto x_y \end{aligned}$$

est surjective par définition.

Elle est injective : Si $x_{y_1} = x_{y_2}$ avec y_1, $y_2 \in f(E)$, alors $y_1 = f(x_{y_1}) = f(x_{y_2}) = y_2$.

Donc g est bien bijective et alors

$$\mathrm{Card}\,(f(E)) = \mathrm{Card}\,(E') \leqslant \mathrm{Card}\,(E) < +\infty.$$

* Si f est injective, alors $f : E \to f(E)$ est bijective et donc $\mathrm{Card}\,(f(E)) = \mathrm{Card}\,(E)$.

* Si f n'est pas injective, $\exists y \in f(E)$ et $x_1 \neq x_y \in E$ t.q. $f(x_1) = f(x_y)$. Donc $x_1 \notin E'$ or $x_1 \in E$. Donc $E' \subsetneq E$ ce qui implique que $\mathrm{Card}\,(E') = \mathrm{Card}\,(f(E)) < \mathrm{Card}\,(E)$.

Par conséquent, $\mathrm{Card}\,(f(E)) = \mathrm{Card}\,(E)$ si et seulement si f est injective. □

Théorème 1.2.1 (Principe des tiroirs) Si l'on place n objets dans m tiroirs avec $n > m$, alors il y a un tiroir qui contient au moins deux objets.

Démonstration. Soit

$$\begin{aligned}\varphi : E = \{\text{objets}\} &\to F = \{\text{tiroirs}\}\\ x &\mapsto \varphi(x).\end{aligned}$$

Comme

$$\mathrm{Card}\,(\varphi(E)) \leqslant \mathrm{Card}\,(F) = m < n = \mathrm{Card}\,(E)\,,$$

on sait que φ n'est pas injective. Donc $\exists y \in F$ tel que $f^{-1}(y)$ a au moins deux éléments. □

Exemple 1.2.1 Dans une pièce où il y a $n \geqslant 2$ personnes, ces gens se serrent la main (握手) pour se saluer. Alors à chaque instant, il y a deux personnes qui ont serré la main à un nombre identique de personnes.

Démonstration. Si à un instant donné, $k \geqslant 2$ personnes ont déjà serré la main, alors le nombre de mains serrées par l'une de ces k personnes est un entier entre 1 et $k-1$. Donc par le principe des tiroirs, au moins deux de ces personnes ont serré le même nombre de mains. □

Exemple 1.2.2 Soit $n \in \mathbb{N}^*$. On considère des entiers $a_1, \ldots, a_n$ non neccessairement distincts. Montrer qu'on peut trouver un sous-ensemble $\neq \varnothing$ des ces n entiers dont la somme divisible par n.

Démonstration. On pose $S_1 = a_1, S_2 = a_1 + a_2, \ldots, S_n = a_1 + \cdots + a_n$.

* Si l'un des $S_j \equiv 0 \mod n$, c'est bon.
* Sinon, ils sont dans $1, \ldots, n-1$ modulo n, $n-1$ tiroirs. Par le principe des tiroirs, il existe S_{j_1}, S_{j_2} avec $j_1 < j_2$ dans le même tiroir. Donc $S_{j_2} - S_{j_1} \equiv 0 \mod n$.

□

1.3 Dénombrement

Proposition 1.3.1 Si E est un ensemble de cardinal n, alors $\mathcal{P}(E)$ a pour cardinal 2^n.

Démonstration. On définit

$$\begin{aligned}\varphi : \mathcal{P}(E) &\to \{0,1\}^E = \{f : E \to \{0,1\} \text{ application}\}\\ A &\mapsto \mathbb{1}_A\end{aligned}$$

Alors, φ est une bijection :

Surjective : Pour $f : E \to \{0,1\}$, on pose $A_f := \{x \in E : f(x) = 1\}$, alors $\varphi(A_f) = \mathbb{1}_{A_f} = f$.

Injective : Si A, $A' \subset E$ tels que $\mathbb{1}_A = \mathbb{1}_{A'}$, alors $x \in A \Leftrightarrow \mathbb{1}_A(x) = 1 \Leftrightarrow \mathbb{1}_{A'}(x) = 1 \Leftrightarrow x \in A'$. Donc $A = A'$.

Ainsi on a

$$\mathrm{Card}\left(\mathcal{P}(E)\right) = \mathrm{Card}\left(\{0,1\}^E\right) = 2^{\mathrm{Card}(E)} = 2^n.$$

□

Soient $E_1, \ldots, E_k$ des ensembles. On note **le produit cartésien**

$$E_1 \times \cdots \times E_k = \{ \underbrace{(x_1, \ldots, x_k)}_{\text{une } k\text{-liste ou un } k\text{-uplet (avec ordre)}} : x_j \in E_j, 1 \leqslant j \leqslant k\}$$

Définition 1.3.1 Soit $p \in \mathbb{N}^*$ et E un ensemble. On appelle **p-arrangement** d'éléments de E toute p-liste d'éléments de E 2 à 2 distincts.

Théorème 1.3.1 Si $\mathrm{Card}\,(E) = n$, alors

$$A_n^p := \mathrm{Card}\left(\{(x_1, \ldots, x_p) \in E^p : x_i \neq x_j, i \neq j\}\right) = \begin{cases} \dfrac{n!}{(n-p)!} & \text{si } p \leqslant n \\ 0 & \text{si } p > n \end{cases}.$$

Démonstration. Soit $p \leq n$. Pour construire un p-uplet d'éléments distincts dans E :

- Pour x_1 : n choix possibles,
- Pour x_2 : $n-1$ choix, (car $x_2 \neq x_1$)
- Pour x_3 : $n-2$ choix, (car $x_3 \neq x_1, x_2$)
- ...
- Pour x_p : $n-p+1$ choix.

Par le principe multiplicatif, le nombre total est

$$A_n^p = n(n-1)(n-2)\ldots(n-p+1) = \frac{n!}{(n-p)!}.$$

Si $p > n$, il est impossible d'avoir p éléments distincts dans un ensemble de cardinal n, donc $A_n^p = 0$. □

Corollaire Le nombre d'injections d'un ensemble E de cardinal p dans un ensemble $\mathcal{F}$ de cardinal n est A_n^p.

Démonstration. On pose $E = \{a_1, \ldots, a_p\}$, alors $f : E \to F$ est injective $\Leftrightarrow x_j = f(a_j) \in F$ sont 2 à 2 distincts $\Leftrightarrow (x_j)_{1 \leqslant j \leqslant p}$ est un p-arrangements d'éléments de $\mathcal{F}$.

Donc on a une bijection entre les injections de E à $\mathcal{F}$ et les p-arrangements d'éléments de $\mathcal{F}$. □

Définition 1.3.2 Une **permutation** d'un ensemble E est une bijection de E sur lui-même.

Corollaire Le nombre de permutations d'un ensemble de cardinal n est $n!$.

Rappel **Coefficient binomial** : $\binom{n}{p} = C_n^p := \frac{A_n^p}{p!}$.

Théorème 1.3.2 Le nombre de sous-ensembles de cardinal p dans un ensemble E de cardinal n est $\binom{n}{p}$.

Démonstration. Tout p-arrangement $(x_1, \ldots, x_p)$ de E induit un sous-ensemble $\{x_1, \ldots, x_p\} \subset E$ (non ordonné).

Soit $F = \{x_1, \ldots, x_p\}$ un tel sous-ensemble. D'après le théorème précédent, le nombre de p-arrangements de $\mathcal{F}$ est $p!$. Par conséquent, le nombre de combinaisons de p éléments de E vaut $A_n^p/p!$. □

Exercice 1.3.1 (Anagrammes) Étant donné un mot m, on appelle anagramme de m tout mot formé par les mêmes lettres que m.

1) Quel est le nombre d'anagrammes du mot orange ?
2) Quel est le nombre d'anagrammes du mot ananas ?
3) Quel est le nombre d'anagrammes d'un mot contenant n_1 fois la lettre l_1, n_2 fois la lettre l_2, …, n_p fois la lettre l_p ?

Solution.

1) Il y a 6! permutations des 6 lettres distinctes. On a donc 6! anagrammes du mot orange.
2) On peut permuter les 3 apparitions de la lettre a sans changer l'anagramme associé. Donc on a $\frac{6!}{3!2!} = 60$ anagrammes.
3) $\frac{(n_1 + \cdots + n_p)!}{n_1! n_2! \ldots n_p!}$.

□

Lemme 1.3.1 (Formule du binôme) Pour tous $a, b \in \mathbb{C}, n \in \mathbb{N}^*$, on a

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Démonstration. On a $(a+b)^n = (a+b) \times \cdots \times (a+b) = \sum_{k=0}^{\infty} c_k a^k b^{n-k}$.

Un terme du produit s'obtient en prenant un facteur dans chaque parenthèse.

Soit a soit b s'il y a k facteurs a, alors il y a $n-k$ facteurs b.

Chaque terme est de la forme $a^k b^{n-k}$ où $0 \leqslant k \leqslant n$.

Pour chaque k fixé, il y a autant de termes de la forme $a^k b^{n-k}$ qu'il y a de façons de choisir les k parenthèses où l'on prend les facteurs a. □

Théorème 1.3.3 (Formule de Vandermonde) Soient $m, n, p \in \mathbb{N}$, on a

$$\sum_{k=0}^{p} \binom{m}{k}\binom{n}{p-k} = \binom{m+n}{p}.$$

Démonstration. Soient $a, b \in \mathbb{C}$, par formule du binôme, on a

$$(a+b)^{m+n} = \sum_{p=0}^{m+n} \binom{m+n}{p} a^p b^{m+n-p}.$$

et

$$\begin{aligned}(a+b)^{m+n} &= (a+b)^m (a+b)^n \\ &= \left(\sum_{k=0}^{m} \binom{m}{k} a^k b^{m-k}\right)\left(\sum_{j=0}^{n} \binom{n}{j} a^j b^{n-j}\right) \\ &= \sum_{p=0}^{m+n} \sum_{\substack{0\leqslant k\leqslant \min(m,p) \\ p-k\leqslant n}} \binom{m}{k}\binom{n}{p-k} a^p b^{m+n-p}. \quad (\text{soit } p = k+j)\end{aligned}$$

Mais $\binom{m}{k} = 0$ si $k > m$ et $\binom{n}{p-k} = 0$ si $p - k > n$.
Donc

$$(a+b)^{m+n} = \sum_{p=0}^{m+n} \sum_{k=0}^{p} \binom{m}{k}\binom{n}{p-k} a^p b^{m+n-p}.$$

En comparant les coefficients de $a^p b^{m+n-p}$, on obtient la formule de Vandermonde. □

1.4 Dénombrabilité

Définition 1.4.1 Un ensemble est **dénombrable** s'il est en bijection avec $\mathbb{N}$. Il est **au plus dénombrable** s'il est fini ou dénombrable.

Si E est fini, on peut énumérer ses éléments comme une suite finie $x_1, x_2, ..., x_n$.
Si E est dénombrable, on peut énumérer ses éléments comme une suite infinie $(x_n)_{n\geqslant 0}$.

Proposition 1.4.1 Un ensemble A est au plus dénombrable s'il existe une injection de A dans $\mathbb{N}$. Donc un ensemble est au plus dénombrable si et seulement si il est en bijection avec une partie de $\mathbb{N}$.

Démonstration.

($\Rightarrow$) Par définition.

($\Leftarrow$) Soit $f : A \to \mathbb{N}$ une injection. Si $f(A)$ est un ensemble fini, alors A est un ensemble fini car $\tilde{f} : A \to f(A)$ est bijective.

Si $f(A)$ n'est pas fini, alors on construit par récurrence une suite $(a_n)_{n\in\mathbb{N}}$ par $a_0 = \min f(A)$ et $\forall n \in \mathbb{N}, a_{n+1} = \min\{x \in f(A) : x > a_n\}$.

Alors

$$\begin{array}{ccccc} \mathbb{N} & \xrightarrow[\text{bijection}]{\tilde{f}} & f(A) & \xrightarrow{\tilde{f}^{-1}} & A \\ n & \mapsto & a_n & \mapsto \tilde{f}^{-1}(a_n) & \end{array}$$

est une bijection. Donc A est dénombrable.

□

Corollaire Un sous-ensemble infini d'un ensemble dénombrable est dénombrable.

Un sous-ensemble d'un ensemble au plus dénombrable est au plus dénombrable.

Lemme 1.4.1

1) $\exists f : A \to B$ injective avec B au plus dénombrable $\Rightarrow A$ est au plus dénombrable.
2) $\exists f : A \to B$ surjective avec A au plus dénombrable $\Rightarrow B$ est au plus dénombrable.

Démonstration.

1) Comme la composition de deux injections reste injective, cela résulte de la proposition précédente.
2) Si $f : A \to B$ surjective, $\forall b \in B$, on choisit un antécédent $a_b \in A$ de b par f, c-à-d $f(a_b) = b$.
   On note $A' = \{a_b : b \in B\} \subset A$. Alors $f : A' \to B$ est une bijection.
   Si A est au plus dénombrable, A' l'est aussi (par le corollaire ci-dessus) et donc B l'est aussi.

□

Théorème 1.4.1

1) Si $A_1, \ldots, A_n$ sont (au plus) dénombrables, alors le produit $A_1 \times A_2 \times \cdots \times A_n$ est (au plus) dénombrable.
2) Si $(A_i)_{i \in I}$ est une famille au plus dénombrable ($I \neq \varnothing$ est au plus dénombrable) d'ensembles au plus dénombrables (chaque A_i), alors l'union $\bigcup_{i \in I} A_i$ est également au plus dénombrable. Si l'un des A_i est dénombrable, $\bigcup_{i \in I} A_i$ est alors dénombrable.

Démonstration.

1) On peut supposer que $A_1 = A_2 = \cdots = A_n = \mathbb{N}^* = \{1, 2, \ldots\}$.
   Notons $p_1 = 2, p_2 = 3, \ldots, p_n$ les n plus petits nombres premiers.
   L'application

$$\begin{aligned} f : (\mathbb{N}^*)^n &\to \mathbb{N}^* \\ (i_1, \ldots, i_n) &\mapsto p_1^{i_1} p_2^{i_2} \ldots p_n^{i_n} \end{aligned}$$

   est injective grâce à l'unicité de la décomposition en facteurs premiers d'un entier donné.
   Par le lemme, on voit que $(\mathbb{N}^*)^n$ est dénombrable.

2) Si I est fini, on peut ajouter des A_i' vides sans changer l'union $\bigcup_{i\in I} A_i$.

Donc sans perte de généralité, on peut supposer que $I = \mathbb{N}^*$.

L'application

$$\begin{aligned}\mathbb{N}^* \times \mathbb{N}^* &\to \bigcup_{i\in I} A_i \\ (n,m) &\mapsto \text{le } m\text{-ième élément de l'ensemble } A_n\end{aligned}$$

est surjective.

Le lemme précédent permet de conclure.

Si l'un des A_i n'est pas fini, alors $\bigcup_{i\in I} A_i$ ne l'est pas non plus.

□

Remarque

Si l'un des A_i, $1 \leqslant i \leqslant n$ est vide, $A_1 \times \cdots \times A_n$ est vide.

Sinon, si l'un des A_i n'est pas fini, alors $A_1 \times \cdots \times A_n$ n'est pas fini.

Exemple 1.4.1

1) $\mathbb{Z}$ est dénombrable, car

$$\begin{aligned}\{1,-1\} \times \mathbb{Z} &\to \mathbb{Z} \\ (\varepsilon, n) &\mapsto \varepsilon n\end{aligned}$$

est une surjection et $\mathbb{Z}$ est infini.

2) $\mathbb{N}^2$ est dénombrable, car

$$\mathbb{N}^2 = \bigcup_{j\in\mathbb{N}} A_j,$$

où $A_j = \{(a,b) \in \mathbb{N}^2 : a + b \leqslant j\}$ sont ensembles finis.

3) $\mathbb{Q}$ est dénombrable.

4) $\mathbb{Z}[X]$ est dénombrable. (Indice : $\mathbb{Z}[X] = \bigcup_{n\in\mathbb{N}} \mathbb{Z}_n[X]$.)

Théorème 1.4.2 $\mathbb{R}$ n'est pas dénombrable.

Méthode 1. Raisonnement par l'absurde en utilisant le procédé diagonal de Cantor. □

Méthode 2. Comme $\mathbb{R}$ est infini, pour montrer qu'il n'est pas dénombrable, il suffit de montrer qu'il n'existe pas de bijection de $\mathbb{N}$ dans $\mathbb{R}$.

Soit $f : \mathbb{N} \to \mathbb{R}$ une bijection. On a donc une suite $(u_n)_{n\in\mathbb{N}}$ d'éléments 2 à 2 distincts dans $\mathbb{R}$.

On note $E = \{u_n : n \in \mathbb{N}\} = f(\mathbb{N}) \subset \mathbb{R}$.

On va montrer que f n'est pas surjective en construisant un $y \in \mathbb{R}\backslash E$.

En fait, on va construire deux suites de Cauchy $(a_n), (b_n)$ avec $a_n < b_n, \forall n \in \mathbb{N}$ et $u_n \notin [a_n, b_n]$ pour tout $n \in \mathbb{N}$.

Pour $n = 0$, on prend $a_0 < b_0, a_0, b_0 \in \mathbb{Q}$ tels que $u_0 \neq 0$, $\exists j \in \{1, 2, 3\}$ tel que $u_1 \notin I_0^j := [a_1, b_1]$.

Alors $a_0 \leqslant a_1 < b_1 \leqslant b_0$ et $b_1 - a_1 = \dfrac{1}{3}(b_0 - a_0)$.

On continue comme ça, alors $u_n \notin [a_n, b_n]$ et $a_n, b_n \in \mathbb{Q}$.

$(a_n)_{n\geqslant 0}$ est croissante, $(b_n)_{n\geqslant 0}$ est décroissante et

$$0 < a_n - a_{n-1} < b_{n-1} - a_{n-1} = \left(\frac{1}{3}\right)^{n-1}(b_0 - a_0),$$

$$0 < b_{n-1} - b_n < b_{n-1} - a_{n-1} = \left(\frac{1}{3}\right)^{n-1}(b_0 - a_0).$$

Donc $(a_n), (b_n)$ sont des suites de Cauchy dans $\mathbb{R}$.

$y := \lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} b_n$ existe dans $\mathbb{R}$ et $\forall n \in \mathbb{N}, y \in [a_n, b_n]$.

Mais $u_n \notin [a_n, b_n]$ pour tout $n \in \mathbb{N}$. On en déduit que $\forall n \in \mathbb{N}, y \neq u_n$.

Autrement dit, f n'est pas surjective. Donc $\mathbb{R}$ n'est pas dénombrable. □

Corollaire $\mathbb{R} = \bigcup\limits_{n\in\mathbb{Z}} [n, n+1[\Rightarrow [0, 1[$ n'est pas dénombrable.

$\{0, 1\}^{\mathbb{N}}$ n'est pas dénombrable. (Indice : utiliser la représentation binaire)

Proposition 1.4.2 Soit E un ensemble. Il n'existe pas de surjection de E dans $\mathcal{P}(E)$.

Démonstration. Soit $f : E \to \mathcal{P}(E)$ application. Posons $X = \{x \in E : x \notin f(x)\} \subset E$.

S'il existe $a \in E$ tel que $f(a) = X$:

— Soit $a \in X$ alors $a \notin f(a) = X$ ce qui est contradictoire.

— Soit $a \notin X$ alors $a \in f(a) = X$ ce qui est contradictoire.

Donc $X \notin f(E)$ c'est-à-dire f n'est pas surjective. □

Corollaire $\mathbb{R}$ et $\mathcal{P}(\mathbb{R})$ ne sont pas en bijection.

$$\underset{\text{distinct par leurs cardinals}}{\text{Ensembles finis}} \to \underset{\text{dénombrable}}{\mathbb{N}} \overset{?}{\to} \underset{\text{non-dénombrable en bijection } \mathcal{P}(\mathbb{N})}{\mathbb{R}} \to \dots$$

Définition 1.4.2 (Hypothèse du continu) Il n'y a pas de cardinaux strictement compris entre celui du dénombrable $\aleph_0$ et celui du continu $\aleph_1$.

Théorème 1.4.3 (Paul Cohen, 1963) Si le système des axiomes de la théorie des ensembles de Zermelo-Fraenkel (ZF) avec l'axiome du choix est non contradictoire, alors l'hypothèse du continu n'est pas prouvable à partir de ZFC.

Chapitre 2

Espace de probabilité

2.1 Espaces de probabilité (ou Espaces probabilisés)

— L'espace d'état « univers » Ω est un ensemble contenant tous les résultats possibles de l'expérience aléatoires.

— L'ensemble des événements $\mathcal{F}$ contient certaines parties du Ω,
un événement $= \{\omega \in \Omega$ tel que « ω vérifie une certaine proprété »$\}$.

— La probabilité $\mathbb{P}$ est une application de $\mathcal{F}$ à $[0,1]$.

— Triplet $(\Omega, \mathcal{F}, \mathbb{P})$ est appelé **espace probabilisé**.

Définition 2.1.1 (Tribu(σ-algèbre), espace probabilisable) Soit Ω un ensemble non vide (quelconque). On dit que $\mathcal{F} \subset \mathcal{P}(\Omega)$ une **tribu** (ou une **σ-algèbre**) sur Ω si

1) $\Omega \in \mathcal{F}$
2) $\mathcal{F}$ est stable pour passage au complémentaire : $A \in \mathcal{F} \Rightarrow A^C = \Omega \backslash A \in \mathcal{F}$
3) $\mathcal{F}$ est stable pour union dénombrable : si $(A_i)_{i\in\mathbb{N}}$ est une suite d'éléments de $\mathcal{F}$, alors $\bigcup\limits_{i\in\mathbb{N}} A_i \in \mathcal{F}$

Dans ce cas, on dit que $(\Omega, \mathcal{F})$ est un **espace probabilisable**.

Proposition 2.1.1 Soit $(\Omega, \mathcal{F})$ un espace probabilisable. Alors

1) $\varnothing \in \mathcal{F}$,
2) Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'éléments de $\mathcal{F}$, alors $\bigcap\limits_{n\in\mathbb{N}} A_n \in \mathcal{F}$,
3) Si $A_1, \ldots, A_n \in \mathcal{F}$, alors $\bigcup\limits_{i=1}^{n} A_i \in \mathcal{F}$ et $\bigcap\limits_{i=1}^{n} A_i \in \mathcal{F}$,
4) $\forall A, B \in F, A\backslash B \in \mathcal{F}$.

Démonstration.

1) $\Omega \in \mathcal{F} \Rightarrow \varnothing \in \mathcal{F}$,
2) $\left(\bigcap\limits_{n\in\mathbb{N}} A_n\right)^C = \bigcup\limits_{n\in\mathbb{N}} \left(A_n^C\right) \in \mathcal{F}$,
3) Prenons $A_m = A_n$ pour tout $m \geqslant n$, $\mathcal{F} \ni \bigcup\limits_{n\in\mathbb{N}} A_n = A_1 \cup \ldots A_n \in \mathcal{F}$,
4) $A\backslash B = A \cap (\Omega\backslash B) = A \cap B^C \in \mathcal{F}$.

□

Définition 2.1.2 (Axiomes des probabilités) Soit $\Omega(\neq \varnothing)$ un univers muni d'une tribu. Une fonction $\mathbb{P} : \mathcal{F} \to [0,1]$ est appelée **(mesure de) probabilité** si

— **A1** : $\mathbb{P}(\Omega) = 1$,

— **A2** (σ-additivité) : Pour toute suite $(A_n)_{n\geqslant 1}$ d'élément de $\mathcal{F}$ 2 à 2 disjoints, on a

$$\mathbb{P}\left(\bigsqcup_{n\geqslant 1} A_n\right) = \mathbb{P}\left(\bigcup_{n\geqslant 1} A_n\right) = \sum_{n\geqslant 1} \mathbb{P}(A_n).$$

$\varnothing$ est appelé **l'événement impossible**, Ω est appelé **l'événement certain**.

Pour $A \in \mathcal{F}$, $\mathbb{P}(A)$ est appelée **probabilité de l'événement** A.

Remarque Si $\mathbb{P} : \mathcal{F} \to \overline{\mathbb{R}}^+ = [0, +\infty]$ vérifie seulement **A2**, alors $\mathbb{P}$ est une **mesure** sur l'espace mesurable $(\Omega, \mathcal{F})$.

Définition 2.1.3 (espace probabilisé) Un **espace probabilisé** est un triplet $(\Omega, \mathcal{F}, \mathbb{P})$ où $(\Omega, \mathcal{F})$ espace probabilisable et $\mathbb{P}$ une probabilité sur $(\Omega, \mathcal{F})$

Remarque

Si Ω est au plus dénombrable, on prend $\mathcal{F} = \mathcal{P}(\Omega)$ la plus grande tribu possible sur Ω.

Si Ω est non dénombrable, il est possible qu'il existe aucune probabilité sur $(\Omega, \mathcal{P}(\Omega))$.

Théorème 2.1.1 Soit Ω au plus dénombrable.

$$\text{Une probabilité } \mathbb{P} \text{ sur } (\Omega, \mathcal{P}(\Omega)) \overset{\text{1-1}}{\longleftrightarrow} \text{une fonction } \begin{matrix} p : \Omega \to [0,1] \\ \omega \mapsto p_\omega \end{matrix} \text{ vérifiant } \begin{cases} 1)\ \forall \omega \in \Omega,\ p_\omega \geqslant 0 \\ 2)\ \sum\limits_{\omega\in\Omega} p_\omega = 1 \end{cases}$$

$$\begin{aligned} \mathbb{P} &\longrightarrow p_\omega = \mathbb{P}(\{\omega\}) \\ \mathbb{P}(A) = \sum_{\omega\in A} p_\omega &\longleftarrow (p_\omega)_{\omega\in\Omega} \end{aligned}$$

est une bijection.

Démonstration.

($\Rightarrow$) Pour tout $\omega \in \Omega$, le singleton (単点集) $\{\omega\}$ appartient à $\mathcal{P}(\Omega)$.

Si $\mathbb{P}$ est une proba,

$$\forall \omega \in \Omega,\ p_\omega = \mathbb{P}(\{\omega\}) \geqslant 0 \text{ et } \sum_{\omega\in\Omega} p_\omega = \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) \overset{\sigma-\text{add}}{=} \mathbb{P}\left(\bigsqcup_{\omega\in\Omega} \{\omega\}\right) = \mathbb{P}(\Omega) = 1.$$

($\Leftarrow$) Si $p_\omega \geqslant 0$, $\omega \in \Omega$ et $\sum\limits_{\omega\in\Omega} p_\omega = 1$, on définit

$$\begin{aligned} \mathbb{P} : \mathcal{P}(\Omega) &\to [0,1] \\ A &\mapsto \mathbb{P}(A) = \sum_{\omega\in A} p_\omega. \end{aligned}$$

Alors,

$$\forall A \in \mathcal{P}(\Omega),\ 0 \leqslant \mathbb{P}(A) = \sum_{\omega\in A} p_\omega \leqslant \sum_{\omega\in\Omega} p_\omega = 1.$$

Particulièrement, on a

$$\mathbb{P}(\Omega) = \sum_{\omega\in\Omega} p_\omega = 1.$$

Si $A_i \in \mathcal{P}(\Omega)$, $i \in \mathbb{N}$, $A_i \cap A_j = \varnothing, i \neq j$, alors

$$\mathbb{P}\left(\bigsqcup_{i\in\mathbb{N}} A_i\right) = \sum_{\omega\in\bigsqcup_{i\in\mathbb{N}} A_i} p_\omega = \sum_{i\in\mathbb{N}}\left(\sum_{\omega\in A_i} p_\omega\right) = \sum_{i\in\mathbb{N}} \mathbb{P}(A_i).$$

Donc $\mathbb{P}$ est une probabilité sur $(\Omega, \mathcal{P}(\Omega))$.

□

Théorème 2.1.2 Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Alors,

1) $\mathbb{P}(\varnothing) = 0$,
2) Pour toute famille $(A_1, \ldots, A_n)$ d'événements 2 à 2 incompatibles,

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \mathbb{P}(A_i),$$

3) $\forall A \in \mathcal{F}$, $\mathbb{P}\left(A^C\right) = 1 - \mathbb{P}(A)$,
4) $\forall A, B \in \mathcal{F}$, $A \subset B$, on a $\mathbb{P}(B\backslash A) = \mathbb{P}(B) - \mathbb{P}(A)$. En particulier, $\mathbb{P}(B) \geqslant \mathbb{P}(A)$,
5) $\forall A, B \in \mathcal{F}$, on a $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Démonstration.

1) Soient $A_n = \varnothing$, $\forall n \geqslant 1$. Par σ-additivité,

$$\sum_{n=1}^{\infty} \mathbb{P}(\varnothing) = \sum_{n=1}^{\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}(\varnothing) \Rightarrow \mathbb{P}(\varnothing) = 0.$$

2) On applique σ-additivité à $(B_m)_{m\geqslant 1}$ où $B_m = A_m$ si $m \leqslant n$, $B_m = \varnothing$ si $m > n$.

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \mathbb{P}\left(\bigcup_{m\geqslant 1} B_m\right) = \sum_{m\geqslant 1} \mathbb{P}(B_m) = \sum_{m=1}^{n} \mathbb{P}(B_m) + 0 = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

3) $\Omega = A \sqcup A^C$, on a alors

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(A \sqcup A^C\right) = \mathbb{P}(A) + \mathbb{P}\left(A^C\right).$$

4) $B = A \sqcup (B\backslash A) \Rightarrow \mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B\backslash A)$.
5) $A \cup B = A \sqcup (B\backslash A) = A \sqcup (B\backslash(A \cap B))$, donc

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B\backslash(A \cap B)) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

□

Notation

— $\mathcal{F}$: σ-field,

— $\mathcal{A}$: σ-algèbre.

Si Ω est au plus dénombrable, on prend $\mathcal{F} = \mathcal{P}(\Omega)$,

$$\forall \omega \in \Omega,\ \{\omega\} \in \mathcal{F} \Rightarrow \forall A \in \mathcal{P}(\Omega),\ A = \bigcup_{\omega \in A} \{\omega\} \in \mathcal{F}.$$

Si Ω n'est pas au plus dénombrable, on prend $\mathcal{F}$ une tribu construite à partir des événements de base en utilisant passage au complémentaire / union / intersection dénombrable.

Potentiellement $\mathcal{F} \neq \mathcal{P}(\Omega)$ lorsqu'il n'est pas possible d'étendre une probabilité définie sur ces événements de base à une probabilité définie sur $\mathcal{P}(\Omega)$.

Exemple 2.1.1 (Ex9 TD3)

$$\nexists\, \mathbb{P} \text{ probabilité sur } ([0,1], \mathcal{P}([0,1])) \text{ t.q. } \begin{cases} 1)\ \forall 0 \leqslant a < b \leqslant 1\, \mathbb{P}([a,b]) = b - a, \\ 2) \text{ invariance par translation.} \end{cases}$$

La tribu borélienne $\mathcal{B}([0,1])$, définie comme la plus petite tribu contenant les intervalles $[a,b]$ pour tous $0 \leqslant a < b \leqslant 1$, est appelée **tribu de Borel**. On va construire une telle probabilité $\mathbb{P}$ sur celle-ci.

Exemple 2.1.2 Soit $\Omega = \mathbb{N}$ et $\mathcal{F} = \mathcal{P}(\Omega)$.

1) **Loi géométrique, $\mathcal{G}$éom(p)**

Si $p \in]0,1[$, on pose

$$\mathbb{P}(\{n\}) = p_n = (1-p)^n p$$

pour $n \geqslant 0$.

On a $p_n > 0$ et $\sum_{n \geqslant 0} p_n = 1$. La famille $(p_n)_{n \geqslant 0}$ définit une probabilité sur $\mathbb{N}$.

2) **Loi de Poisson, $\mathcal{P}$oi(λ) ou $\mathcal{P}(\lambda)$**

Si $\lambda > 0$, on pose

$$\mathbb{P}(\{n\}) = p_n = e^{-\lambda} \cdot \frac{\lambda^n}{n!}$$

pour $n \geqslant 0$.

On a $p_n > 0$ et $\sum_{n \geqslant 0} p_n = 1$. La famille $(p_n)_{n \geqslant 0}$ définit une probabilité sur $\mathbb{N}$.

Définition 2.1.4 (Probabilité uniforme) Soit Ω un univers **fini**. On prend toujours $\mathcal{F} = \mathcal{P}(\Omega)$. On définit

$$\mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}, \quad \text{pour tout } A \subset \Omega.$$

$\mathbb{P}$ est une probabilité appelée **probabilité uniforme sur** Ω. On a $\mathbb{P}(\{\omega\}) = \dfrac{1}{\text{Card}(\Omega)}$ pour tout $\omega \in \Omega$, ce qui ne dépend pas de ω.

Exercice 2.1.1 Considérons un jeu de cartes : 13 valeurs $(A, 2, 3, \ldots, J, Q, K)$ et 4 enseignes (carreau $\diamondsuit$, cœur $\heartsuit$, trèfle $\clubsuit$, pique $\spadesuit$). Dans ce jeu, 5 cartes sont distribuées au hasard à un joueur.

Modélisation : $\Omega =$ l'ensemble des parties à 5 éléments de l'ensemble des 52 cartes. On a

$$\mathrm{Card}\,(\Omega) = \binom{52}{5} < +\infty, \quad \mathcal{F} = \mathcal{P}(\Omega).$$

On munit Ω de la probabilité uniforme $\mathbb{P}$.

Q1 : Soit $A =$ « le joueur a exactement 3 cartes de carreau ».

$$\mathrm{Card}\,(A) = \binom{13}{3} \times \binom{39}{2}, \quad \mathbb{P}\,(A) = \frac{2717}{33320} \approx 0,08.$$

Q2 : Soit $B =$ « le joueur a au moins 2 cartes de même valeur ». Alors $B^C =$ « le joueur a 5 cartes de valeurs différentes ».

$$\mathrm{Card}\left(B^C\right) = \binom{13}{5} \times 4^5, \quad \mathbb{P}\,(B) = 1 - \mathbb{P}\left(B^C\right) = \frac{2053}{4165} \approx 0,49.$$

Exercice 2.1.2 Soient n personnes nées au hasard une année ayant 365 jours. Soit p_n la probabilité qu'au moins deux personnes aient leur anniversaire le même jour.

On pose

$$\Omega = [\![1, 365]\!]^n = \{\omega = (\omega_1, \ldots, \omega_n) : \forall 1 \leqslant i \leqslant n,\, \omega_i \in [\![1, 365]\!]\}.$$

Alors $\mathrm{Card}\,(\Omega) = 365^n$.

Soit $A = \{\omega \in \Omega : \exists\, i \neq j,\, \omega_i = \omega_j\}$. On a $\mathrm{Card}\left(A^C\right) = A_{365}^n$. Ainsi,

$$\mathbb{P}\,(A) = 1 - \mathbb{P}\left(A^C\right) = 1 - \frac{\mathrm{Card}\left(A^C\right)}{\mathrm{Card}\,(\Omega)}.$$

Théorème 2.1.3 (Continuité par le bas et par le haut des probabilités)

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

1) Soit $(A_n)_{n\geqslant 1}$ une suite croissante d'événements (c-à-d $A_n \subset A_{n+1}$ pour tout $n \geqslant 1$). Alors

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}\,(A_n)\,.$$

2) Soit $(A_n)_{n\geqslant 1}$ une suite décroissante d'événements (c-à-d $A_{n+1} \subset A_n$ pour tout $n \geqslant 1$). Alors

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}\,(A_n)\,.$$

Démonstration.

1) On écrit

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup \left(\bigcup_{n=2}^{\infty} (A_n \setminus A_{n-1})\right),$$

qui est une union d'événements deux à deux disjoints. Par σ-additivité,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}(A_1) + \sum_{n=2}^{\infty} \mathbb{P}(A_n \setminus A_{n-1}).$$

D'autre part, par additivité finie,

$$\mathbb{P}(A_k) = \mathbb{P}(A_1) + \sum_{n=2}^{k} \mathbb{P}(A_n \setminus A_{n-1}) \quad \text{pour tout } k \geqslant 2.$$

Ainsi, la limite $\lim_{k\to\infty} \mathbb{P}(A_k)$ existe et est égale à $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right)$.

2) La suite $(A_n^C)_{n\geqslant 1}$ est croissante. D'après 1),

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n^C\right) = \lim_{n\to\infty} \mathbb{P}\left(A_n^C\right).$$

Par conséquent,

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = 1 - \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n^C\right) = 1 - \lim_{n\to\infty} \mathbb{P}\left(A_n^C\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

□

Corollaire Soit $(A_n)_{n\geqslant 1}$ une suite d'événements (non nécessairement monotone). Alors

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right), \quad \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k=1}^{n} A_k\right).$$

Démonstration. Posons $B_n = \bigcup_{k=1}^{n} A_k$. La suite (B_n) est croissante. D'après le théorème,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n).$$

Or, on vérifie que $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$. La preuve de la deuxième formule est similaire. □

Sous-additivité de $\mathbb{P}$ Soit $(A_i)_{i\in I}$ une famille au plus dénombrable d'événements. Alors

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

Démonstration. Comme I est au plus dénombrable, on peut supposer $I \subset \mathbb{N}$. Alors

$$\bigcup_{i\in I} A_i = A_{\min I} \cup \left(\bigcup_{\substack{i\in I \\ i>\min I}} \left(A_i \setminus \bigcup_{\substack{j\in I \\ j<i}} A_j\right)\right)$$

est une union disjointe. Par σ-additivité,

$$\mathbb{P}\left(\bigcup_{i\in I} A_i\right) = \mathbb{P}\left(A_{\min I}\right) + \sum_{\substack{i\in I\\ i>\min I}} \mathbb{P}\left(A_i \setminus \bigcup_{\substack{j\in I\\ j<i}} A_j\right).$$

Comme $A_i \setminus \bigcup\limits_{j<i} A_j \subset A_i$, on a

$$\mathbb{P}\left(A_i \setminus \bigcup_{\substack{j\in I\\ j<i}} A_j\right) \leqslant \mathbb{P}\left(A_i\right).$$

On en déduit

$$\mathbb{P}\left(\bigcup_{i\in I} A_i\right) \leqslant \mathbb{P}\left(A_{\min I}\right) + \sum_{\substack{i\in I\\ i\neq\min I}} \mathbb{P}\left(A_i\right) = \sum_{i\in I} \mathbb{P}\left(A_i\right). \qquad \square$$

Définition 2.1.5 (Négligeable, presque sûr)

Un événement de probabilité nulle est dit **négligeable**.

Un événement de probabilité 1 est dit **presque sûr** ou **presque certain**. Une propriété vérifiée sur un événement de probabilité 1 est dite **presque sûre (p.s.)**.

Corollaire Une réunion au plus dénombrable d'événements négligeables est négligeable.

Démonstration. Soit $(A_i)_{i\in I}$ une famille au plus dénombrable d'événements tels que $\mathbb{P}\left(A_i\right) = 0$ pour tout $i \in I$. Par sous-additivité,

$$0 \leqslant \mathbb{P}\left(\bigcup_{i\in I} A_i\right) \leqslant \sum_{i\in I} \mathbb{P}\left(A_i\right) = 0. \qquad \square$$

Exemple 2.1.3 (Jeu de Pile ou Face) Lancer une pièce équilibrée une infinité de fois. Soit $A_k =$ « le k-ème lancer est Pile » pour $k \geqslant 1$. On pose

$$\Omega = \{0,1\}^{\mathbb{N}^*} = \{(a_n)_{n\geqslant 1} : a_n = 0 \text{ ou } 1 \text{ pour tout } n \geqslant 1\}.$$

On a $\mathbb{P}\left(A_k\right) = \dfrac{1}{2}$ pour tout $k \geqslant 1$ et $\mathbb{P}\left(\bigcap\limits_{k=1}^{n} A_k\right) = \left(\dfrac{1}{2}\right)^n$ pour tout $n \geqslant 1$.

Soit $A = \bigcap\limits_{k=1}^{\infty} A_k =$ « on n'obtient que des Piles » $= \{(0,0,\dots)\} \neq \varnothing$. Alors

$$\mathbb{P}\left(A\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k=1}^{n} A_k\right) = \lim_{n\to\infty} \left(\frac{1}{2}\right)^n = 0.$$

Ainsi, A est négligeable.

De même, $\mathbb{P}\left(\bigcap\limits_{k=n}^{\infty} A_k\right) = 0$ pour tout $n \geqslant 1$. Par sous-additivité,

$$\mathbb{P}\left(\bigcup_{n\geqslant 1}\left(\bigcap_{k=n}^{\infty} A_k\right)\right) \leqslant \sum_{n\geqslant 1} \mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k\right) = 0.$$

En passant au complémentaire,

$$\mathbb{P}\left(\bigcap_{n\geqslant 1}\left(\bigcup_{k=n}^{\infty} A_k^C\right)\right) = 1.$$

Ce dernier événement signifie « il y a une infinité de Faces dans la suite ».

2.2 Probabilité conditionnelle et Indépendance

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

Théorème 2.2.1 Pour $A \in \mathcal{F}$ avec $\mathbb{P}(A) > 0$, l'application

$$\begin{aligned} \mathbb{P}_A : \mathcal{F} &\to \mathbb{R} \\ B &\mapsto \mathbb{P}_A(B) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} \end{aligned}$$

est une probabilité sur $(\Omega, \mathcal{F})$.

Démonstration. D'abord, $B \cap A \subset B$, donc $0 \leqslant \mathbb{P}(B \cap A) \leqslant \mathbb{P}(A)$, ce qui implique $\mathbb{P}_A(B) \in [0,1]$ pour tout $B \in \mathcal{F}$. Ensuite, $\mathbb{P}_A(\Omega) = \dfrac{\mathbb{P}(\Omega \cap A)}{\mathbb{P}(A)} = 1$. Enfin, pour la σ-additivité, soit $(B_j)_{j\geqslant 1} \subset \mathcal{F}$ une famille d'événements deux à deux disjoints. Alors

$$\mathbb{P}_A\left(\bigcup_{j\geqslant 1} B_j\right) = \frac{\mathbb{P}\left(\left(\bigcup_{j\geqslant 1} B_j\right) \cap A\right)}{\mathbb{P}(A)} = \frac{\mathbb{P}\left(\bigcup_{j\geqslant 1}(B_j \cap A)\right)}{\mathbb{P}(A)} = \sum_{j\geqslant 1} \frac{\mathbb{P}(B_j \cap A)}{\mathbb{P}(A)} = \sum_{j\geqslant 1} \mathbb{P}_A(B_j).$$

□

Définition 2.2.1 (Probabilité conditionnelle) Dans ce cas, $\mathbb{P}_A$ est appelée **probabilité conditionnelle sachant** A. Pour tout $B \in \mathcal{F}$, $\mathbb{P}(B \mid A) := \mathbb{P}_A(B)$ est appelée **probabilité conditionnelle de** B **sachant** A.

Remarque $\mathbb{P}\left(A^C \mid B\right) = 1 - \mathbb{P}(A \mid B) = 1 - \mathbb{P}_B(A)$, mais $\mathbb{P}\left(A \mid B^C\right) \neq 1 - \mathbb{P}(A \mid B)$ en général.

Proposition 2.2.1 (Probabilités conditionnelles en chaîne / Formule des probabilités composées) Soit $n \geqslant 2$ et $A_1, \ldots, A_n$ des événements tels que $\mathbb{P}(A_1 \cap \cdots \cap A_n) > 0$. Alors

$$\begin{aligned} \mathbb{P}\left(\bigcap_{k=1}^{n} A_k\right) &= \mathbb{P}(A_1) \times \mathbb{P}(A_2 \mid A_1) \times \mathbb{P}(A_3 \mid A_1 \cap A_2) \times \cdots \times \mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}) \\ &= \mathbb{P}(A_1) \times \prod_{k=2}^{n} \mathbb{P}(A_k \mid A_1 \cap \cdots \cap A_{k-1}). \end{aligned}$$

Démonstration. On procède par récurrence. Pour $n = 2$, c'est la définition de la probabilité conditionnelle. Supposons la formule vraie pour $n-1$. Alors

$$\mathbb{P}\left(\bigcap_{k=1}^{n} A_k\right) = \mathbb{P}\left(\left(\bigcap_{k=1}^{n-1} A_k\right) \cap A_n\right) = \mathbb{P}\left(\bigcap_{k=1}^{n-1} A_k\right) \times \mathbb{P}\left(A_n \,\middle|\, \bigcap_{k=1}^{n-1} A_k\right).$$

Par hypothèse de récurrence,

$$\mathbb{P}\left(\bigcap_{k=1}^{n-1} A_k\right) = \mathbb{P}(A_1) \times \prod_{k=2}^{n-1} \mathbb{P}(A_k \mid A_1 \cap \cdots \cap A_{k-1}).$$

Donc

$$\mathbb{P}\left(\bigcap_{k=1}^{n} A_k\right) = \mathbb{P}(A_1) \times \prod_{k=2}^{n} \mathbb{P}(A_k \mid A_1 \cap \cdots \cap A_{k-1}).$$

□

Définition 2.2.2 (Système quasi-complet) Un système $(A_i)_{i\in I}$ est dit **quasi-complet** pour $(\Omega, \mathcal{F}, \mathbb{P})$ si $A_i \in \mathcal{F}$, $I \subset \mathbb{N}$, $A_i \cap A_j = \varnothing$ pour tout $i \neq j$, et $\sum_{i\in I} \mathbb{P}(A_i) = 1$. C'est une partition (划分) au plus dénombrable de Ω (à un ensemble négligeable près).

Théorème 2.2.2 (Formule des probabilités totales) Soit $(A_i)_{i\in I}$ un système quasi-complet d'événements de $(\Omega, \mathcal{F}, \mathbb{P})$ tel que $\mathbb{P}(A_i) > 0$ pour tout $i \in I$. Alors, pour tout $B \in \mathcal{F}$,

$$\mathbb{P}(B) = \sum_{i\in I} \mathbb{P}(B \mid A_i)\,\mathbb{P}(A_i).$$

Démonstration. On a $B = \bigcup_{i\in I}(A_i \cap B)$, et cette union est disjointe. Donc

$$\mathbb{P}(B) = \mathbb{P}\left(\bigsqcup_{i\in I}(A_i \cap B)\right) = \sum_{i\in I} \mathbb{P}(A_i \cap B) = \sum_{i\in I} \mathbb{P}(B \mid A_i)\,\mathbb{P}(A_i).$$

□

Théorème 2.2.3 (Formule de Bayes) Si A, $B \in \mathcal{F}$ avec $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$, alors

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

Démonstration. On a $\mathbb{P}(A)\,\mathbb{P}(B \mid A) = \mathbb{P}(A \cap B) = \mathbb{P}(B)\,\mathbb{P}(A \mid B)$. □

Exercice 2.2.1 Soient $p \in]0,1[$ et $\alpha > 0$ tels que la probabilité p_n qu'une famille ait n enfants vérifie

$$\forall n \geqslant 1, \quad p_n = \alpha \cdot p^n.$$

De plus, un enfant a la même probabilité d'être un garçon ou une fille.

On pose

— $\Omega = \{\text{toutes les familles}\}$,

— $E_j = \{\text{familles ayant } j \text{ enfants}\}$ pour $j \in \mathbb{N}$,

— $\Omega = \bigsqcup_{j\in\mathbb{N}} E_j$ avec $p_n = \mathbb{P}(E_n)$.

Soient $F_k = \{\text{familles ayant } k \text{ filles}\}$ et $G_k = \{\text{familles ayant } k \text{ garçons}\}$ pour $k \in \mathbb{N}$.

Questions :

1) Calculer $p_0 = \mathbb{P}(E_0)$.

2) Calculer $\mathbb{P}(E_2 \mid F_2)$.

3) Calculer $\mathbb{P}(G_2 \mid F_2)$.

Solution.

1)
$$1 = \sum_{j\in\mathbb{N}} \mathbb{P}(E_j) = \sum_{j\in\mathbb{N}} p_j = p_0 + \sum_{n\geqslant 1} \alpha p^n = p_0 + \frac{\alpha p}{1-p}.$$
Il faut que $\dfrac{\alpha p}{1-p} \in [0,1]$, c'est-à-dire $0 \leqslant \alpha \leqslant \dfrac{1-p}{p}$. Alors $p_0 = 1 - \dfrac{\alpha p}{1-p}$.

2)
$$\mathbb{P}(E_2 \mid F_2) = \frac{\mathbb{P}(E_2 \cap F_2)}{\mathbb{P}(F_2)} = \frac{\mathbb{P}(F_2 \mid E_2)\,\mathbb{P}(E_2)}{\mathbb{P}(F_2)}.$$
On a $\mathbb{P}(F_2 \mid E_2) = \left(\dfrac{1}{2}\right)^2$, $\mathbb{P}(E_2) = p_2 = \alpha p^2$. Pour $\mathbb{P}(F_2)$, on utilise la formule des probabilités totales :
$$\mathbb{P}(F_2) = \sum_{n\geqslant 0} \mathbb{P}(F_2 \mid E_n)\,\mathbb{P}(E_n).$$
Si $n < 2$, $\mathbb{P}(F_2 \mid E_n) = 0$. Si $n \geqslant 2$, $\mathbb{P}(F_2 \mid E_n) = \binom{n}{2}\left(\dfrac{1}{2}\right)^n$. Donc
$$\mathbb{P}(F_2) = \sum_{n\geqslant 2} \binom{n}{2}\left(\frac{1}{2}\right)^n \alpha p^n = \frac{\alpha}{2}\sum_{n\geqslant 2} n(n-1)\left(\frac{p}{2}\right)^n.$$
En utilisant $\sum_{n=2}^{\infty} n(n-1)x^{n-2} = \dfrac{2}{(1-x)^3}$ pour $0 \leqslant x < 1$, on trouve
$$\mathbb{P}(F_2) = \frac{2\alpha p^2}{(2-p)^3}.$$
Ainsi,
$$\mathbb{P}(E_2 \mid F_2) = \frac{\left(\dfrac{1}{4}\right)\alpha p^2}{\dfrac{2\alpha p^2}{(2-p)^3}} = \frac{(2-p)^3}{8}.$$

3)
$$\mathbb{P}(G_2 \mid F_2) = \frac{\mathbb{P}(G_2 \cap F_2)}{\mathbb{P}(F_2)} = \frac{\mathbb{P}(G_2 \cap F_2 \mid E_4)\,\mathbb{P}(E_4)}{\mathbb{P}(F_2)}.$$
On a $\mathbb{P}(G_2 \cap F_2 \mid E_4) = \binom{4}{2}\left(\dfrac{1}{2}\right)^4 = \dfrac{3}{8}$, $\mathbb{P}(E_4) = \alpha p^4$. Donc
$$\mathbb{P}(G_2 \mid F_2) = \frac{\dfrac{3}{8}\alpha p^4}{\dfrac{2\alpha p^2}{(2-p)^3}} = \frac{3p^2(2-p)^3}{16}.$$

□

Définition 2.2.3 (Indépendance d'évenements) Deux événements $A, B \in \mathcal{F}$ sont dits **indépendants** si

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

Remarque Si $\mathbb{P}(A) > 0$, alors A et B sont indépendants si et seulement si $\mathbb{P}(B \mid A) = \mathbb{P}(B)$.

Remarque L'indépendance n'implique pas l'incompatibilité. Si $\mathbb{P}(A)\,\mathbb{P}(B) > 0$, alors l'indépendance implique $\mathbb{P}(A \cap B) > 0$, donc $A \cap B \neq \varnothing$. Si A et B sont indépendants, alors A^C et B le sont aussi, de même que A et B^C, ainsi que A^C et B^C.

Définition 2.2.4 (2-à-2 indépendants, mutuellement indépendants) Soit $(A_i)_{i\in I}$ une famille d'événements d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, où I est un ensemble d'indices arbitraire.

— On dit que $(A_i)_{i\in I}$ sont **deux à deux indépendants** si pour tout $i, j \in I$ avec $i \neq j$, on a $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\,\mathbb{P}(A_j)$.

— On dit que $(A_i)_{i\in I}$ sont **(mutuellement) indépendants** si pour toute sous-famille finie $J \subset I$ avec $\mathrm{Card}(J) \geqslant 2$, on a

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

Attention! Pour l'indépendance de n événements $A_1, \ldots, A_n$, il ne suffit pas de vérifier que $\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \ldots \mathbb{P}(A_n)$. Il faut mentrer que cette propriété de factorisation soit vraie pour toute sous-famille.

Exemple 2.2.1 On lance deux dés équilibrés (均匀的骰子) à 6 faces. Soient

— $A =$ « le premier dé amène un nombre pair »,

— $B =$ « le second dé amène un nombre pair »,

— $C =$ « les deux dés amènent des nombres de même parité ».

Alors A, B, C sont deux à deux indépendants, mais A et $B \cap C$ ne sont pas indépendants, et A, B, C ne sont pas mutuellement indépendants.

Démonstration. On a $\mathbb{P}(A) = \mathbb{P}(B) = \dfrac{1}{2}$, $\mathbb{P}(C) = \dfrac{\mathrm{Card}(C)}{\mathrm{Card}(\Omega)} = \dfrac{3\times 3 + 3\times 3}{6\times 6} = \dfrac{1}{2}$.

Puisque $A \cap B = A \cap C = B \cap C$, on a

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{3\times 3}{6\times 6} = \frac{1}{4}.$$

Du coup,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B), \quad \mathbb{P}(A \cap C) = \mathbb{P}(A)\,\mathbb{P}(C), \quad \mathbb{P}(B \cap C) = \mathbb{P}(B)\,\mathbb{P}(C).$$

Ainsi, A, B, C sont deux à deux indépendants.

Cependant, on a $A \cap B \cap C = A \cap B$ (car C est automatiquement vrai lorsque A et B le sont), donc

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap B) = \frac{1}{4}.$$

Mais $\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} \neq \frac{1}{4}$. Par conséquent, $\mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$, et A, B, C ne sont pas mutuellement indépendants.

Maintenant, considérons $B \cup C$. On a :

$$\mathbb{P}(B \cup C) = \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(B \cap C) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}.$$

De plus, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$. Mais puisque $A \cap B = A \cap C$ (car $A \cap B = A \cap C$ comme noté précédemment), on a :

$$\mathbb{P}(A \cap (B \cup C)) = \mathbb{P}(A \cap B) = \frac{1}{4}.$$

Or,

$$\mathbb{P}(A)\mathbb{P}(B \cup C) = \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8} \neq \frac{1}{4}.$$

Ainsi, $\mathbb{P}(A \cap (B \cup C)) \neq \mathbb{P}(A)\mathbb{P}(B \cup C)$, donc A et $B \cup C$ ne sont pas indépendants. □

Remarque L'indépendance mutuellement implique l'indépendance deux à deux. Mais cet exemple montre que l'indépendance deux à deux n'implique pas l'indépendance mutuellememt.

Proposition 2.2.2 Si $(A_i)_{i \in I}$ sont indépendants, et si pour chaque $i \in I$ on pose $B_i = A_i$ ou A_i^C, alors $(B_i)_{i \in I}$ sont également indépendants.

Démonstration. Il suffit de montrer que si $(A_1, \ldots, A_n)$ est une famille finie d'événements indépendants, il en est de même pour les 2^n familles $(A'_1, \ldots, A'_n)$ avec $A'_i = A_i$ ou A_i^C, $1 \leqslant i \leqslant n$. On procède par récurrence sur le nombre d'événements complémentaires :

Soit J une sous-ensemble fini de I, Sans perte de généralité, on suppose $J \subset [\![1, n]\!]$.

* Si $A'_i = A_i$ pour tout $i \in J$, alors

$$\mathbb{P}\left(\bigcap_{i \in J} A'_i\right) = \prod_{i \in J} \mathbb{P}(A'_i).$$

* Si $\text{Card}\left(\{i \in J : A'_i = A_i^C\}\right) = 1$, on peut supposer qu'il s'agit de $A'_1 = A_1^C$.
  On note $J\{1\} = \{i_1, \ldots, i_p\} \subset [\![2, n]\!]$. Alors

$$\mathbb{P}\left(A_1^C \cap (A_{i_1} \cap \cdots \cap A_{i_p})\right) = \mathbb{P}\left(A_1^C\right)\mathbb{P}\left(A_{i_1} \cap \cdots \cap A_{i_p}\right) = \mathbb{P}\left(A_1^C\right)\mathbb{P}(A_{i_1}) \ldots \mathbb{P}\left(A_{i_p}\right).$$

* Si le cas $\text{Card}\left(i \in J : A'_i = A_i^C\right) = \ell$ est vrai, on applique l'étape ci-dessus pour déduire que le cas $\text{Card}\left(i \in J : A'_i = A_i^C\right) = \ell + 1$ est également vrai.

□

Exemple 2.2.2 (Loi binomiale) **Schéma de Bernoulli :** $A_1, \dots, A_n$ événements indépendants avec $\mathbb{P}(A_i) = p \in]0,1[$ pour tout $1 \leqslant i \leqslant n$. Soit $B_k =$ « exactement k succès parmi ces n épreuves ». Alors

$$B_k = \bigsqcup_{\substack{I \subset [\![1,n]\!] \\ \mathrm{Card}(I)=k}} \left(\bigcap_{i \in I} A_i \right) \cap \left(\bigcap_{i \notin I} A_i^C \right),$$

$$\mathbb{P}(B_k) = \sum_{\substack{I \subset [\![1,n]\!] \\ \mathrm{Card}(I)=k}} \mathbb{P}\left(\left(\bigcap_{i \in I} A_i \right) \cap \left(\bigcap_{i \notin I} A_i^C \right) \right) = \sum_{\substack{I \subset [\![1,n]\!] \\ \mathrm{Card}(I)=k}} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}.$$

Modèle de Pile ou Face infini $\Omega = \{0,1\}^{\mathbb{N}^*} = \{(\omega_1, \omega_2, \dots) : \omega_i \in \{0,1\}\}$. Soit

$$\begin{aligned} A_n &= \text{« le } n\text{-ième lancer donne pile »} \\ &= \{(\omega_k)_{k \geqslant 1} \in \{0,1\}^{\mathbb{N}^*} : \omega_n = 0\}. \end{aligned}$$

On pose $\mathcal{A} = \{A_n : n \geqslant 1\}$ et $\mathcal{F} = \sigma(A) =$ la plus petite tribu contient $\mathcal{A}$.
Il existe une probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{F})$ vérifie que

$$\begin{cases} \forall n \geqslant 1,\ \mathbb{P}(A_n) = p \in]0,1[, \\ (A_n)_{n \geqslant 1} \text{ est une famille d'événements indépendant.} \end{cases}$$

L'existence de $\mathbb{P}$ est non triviale. (la constuction de mesure générale)

Cependant, on ne peut pas prendre $\mathcal{F} = \mathcal{P}(\Omega)$.

Démonstration.

Idée : Par l'absurde, nous construirons l'espace de probabilité quotient $(\widetilde{\Omega}, \mathcal{P}(\widetilde{\Omega}), \widetilde{\mathbb{P}})$ et une action de $\mathbb{Z}$ sur $\widetilde{\Omega}$ par translation, $\widetilde{g} : \mathbb{Z} \times \widetilde{\Omega} \to \widetilde{\Omega}$, $(k, \widetilde{\omega}) \mapsto \widetilde{g}_k(\widetilde{\omega})$ qui préservera $\widetilde{\mathbb{P}}$. De plus, nous trouverons un sous-ensemble $\widetilde{\Omega}' \subset \widetilde{\Omega}$ presque sûr t.q. l'action $\widetilde{g}$ agira librement sur celui-ci, c-à-d si $\widetilde{\omega} \in \widetilde{\Omega}'$, alors $\widetilde{g}_k(\widetilde{\omega}) = \widetilde{\omega}$ ssi $k = 0$. Nous trouverons encore un ensemble $E \subset \widetilde{\Omega}'$ t.q. $\widetilde{\Omega}' = \bigsqcup_{k \in \mathbb{Z}} \widetilde{g}_k(E)$. Cela conduira à la contradiction $1 = \widetilde{\mathbb{P}}(\widetilde{\Omega}') = \sum_{k \in \mathbb{Z}} \widetilde{\mathbb{P}}(E)$.

Détails : Soit $\mathbb{P} : (\Omega, \mathcal{P}(\Omega)) \to [0,1]$ une telle probabilité, on definit *une relation d'équivalence* $\sim$ sur Ω par

$$(\omega_k)_{k \geqslant 1} \sim (\omega'_k)_{k \geqslant 1} :\Leftrightarrow \exists\, N \in \mathbb{N}^* \text{ t.q. } \forall j \geqslant N,\ \omega_j = \omega'_j. \quad \text{(égalité à partir d'un certain rang)}$$

On pose $\widetilde{\Omega} = \Omega / \sim$ et $\pi : \Omega \to \widetilde{\Omega}$, $\omega \mapsto \widetilde{\omega}$ la projection. De plus, l'application

$$\begin{aligned} \widetilde{\mathbb{P}} : \mathcal{P}(\widetilde{\Omega}) &\to [0,1] \\ E &\mapsto \mathbb{P}\left(\pi^{-1}(E)\right) \end{aligned}$$

est une probabilité sur $(\widetilde{\Omega}, \mathcal{P}(\widetilde{\Omega}))$. L'espace de probabilité induit $(\widetilde{\Omega}, \mathcal{P}(\widetilde{\Omega}), \widetilde{\mathbb{P}})$ est appelé **l'espace de probabilité quotient**.

Ensuite, on note $\omega = (\omega_i)_{i\geqslant 1} \in \Omega$ et définit la translation sur Ω :

$$\begin{aligned} g : \mathbb{Z} \times \Omega &\to \Omega \\ (k, \omega) &\mapsto g_k(\omega) = \begin{cases} (\underbrace{0, \dots, 0}_{k \text{ fois } 0}, \omega_1, \omega_2, \dots) & \text{si } k > 0 \\ (\omega_{-k+1}, \omega_{-k+2}, \dots) & \text{si } k \leqslant 0 \end{cases}. \end{aligned}$$

Mais cette translation n'est pas une action de groupe (群作用) sur Ω, parce que

$$g_k \circ g_{-k}(\omega) = (\underbrace{0, \dots, 0}_{k \text{ fois } 0}, \omega_{k+1}, \omega_{k+2}, \dots) \neq \omega = g_0(\omega).$$

Malgré cela, cette translation g peut induire naturellement une action $\widetilde{g}$ de $\mathbb{Z}$ sur $\widetilde{\Omega}$.

$$\begin{aligned} \widetilde{g} : \mathbb{Z} \times \widetilde{\Omega} &\to \widetilde{\Omega} \\ (k, \widetilde{\omega}) &\mapsto \widetilde{g}_k(\widetilde{\omega}) := g_k(\widetilde{\omega}) \quad \text{(l'image directe de } \widetilde{\omega} \text{ par } g_k) \end{aligned}$$

Il faut montrer que $\widetilde{g}$ est bien défini et qu'il vérifie

$$\forall k_1, k_2 \in \mathbb{Z},\ \forall \widetilde{\omega} \in \widetilde{\Omega},\ \widetilde{g}_{k_1} \circ \widetilde{g}_{k_2}(\widetilde{\omega}) = \widetilde{g}_{k_1+k_2}(\widetilde{\omega}).$$

Si $\widetilde{A} \in \mathcal{P}(\widetilde{\Omega})$, on prend $A = \pi^{-1}(\widetilde{A})$. Alors, pour tout $k > 0$, on a

$$\pi^{-1}\left(\widetilde{g}_k(\widetilde{A})\right) = \{0,1\}^k \times A.$$

Par la propriété du schéma de Bernoulli, on a

$$\mathbb{P}\left(\{0,1\}^k \times A\right) = \mathbb{P}(A). \quad (\textit{pourquoi ?})$$

Donc

$$\widetilde{\mathbb{P}}\left(\widetilde{g}_k(\widetilde{A})\right) = \mathbb{P}\left(\pi^{-1}\left(\widetilde{g}_k(\widetilde{A})\right)\right) = \mathbb{P}\left(\{0,1\}^k \times A\right) = \mathbb{P}(A) = \mathbb{P}\left(\pi^{-1}(\widetilde{A})\right) = \widetilde{\mathbb{P}}(\widetilde{A}).$$

En utilisant $\widetilde{g}_{-k} = (\widetilde{g}_k)^{-1}$, on a alors

$$\widetilde{\mathbb{P}}\left(\widetilde{g}_k(\widetilde{A})\right) = \widetilde{\mathbb{P}}\left(\widetilde{g}_k\left(\widetilde{g}_{-k}(\widetilde{A})\right)\right) = \widetilde{\mathbb{P}}(\widetilde{A}).$$

Et pour $k = 0$, $\widetilde{g}_0 = \mathrm{id}_{\widetilde{\Omega}}$. De manière évidente, on a $\widetilde{\mathbb{P}}\left(\widetilde{g}_0(\widetilde{A})\right) = \widetilde{\mathbb{P}}(\widetilde{A})$. Cela prouve que $\widetilde{\mathbb{P}}$ est préservée par l'action $\widetilde{g}$.

Soit $\widetilde{\omega} \in \widetilde{\Omega}$ et $\omega = (\omega_i)_{i\geqslant 1} \in \pi^{-1}(\{\widetilde{\omega}\})$. Alors,

$$\widetilde{g}_k(\widetilde{\omega}) = \widetilde{\omega} \Leftrightarrow \exists\, N \in \mathbb{N} \text{ t.q. } \forall j > N,\ \omega_{j+|k|} = \omega_j. \quad (|k|\text{-périodique à partir d'un certain rang})$$

On pose pour tout $k \geqslant 1$,

$$\begin{aligned} D_k &= \{\omega \in \Omega : \exists\, N \in \mathbb{N} \text{ t.q. } \forall j > N,\ \omega_{j+k} = \omega_j\} \\ &= \bigcup_{N\in\mathbb{N}} \{\omega \in \Omega : \forall j > N,\ \omega_{j+k} = \omega_j\}. \quad (\text{dénombrable}) \end{aligned}$$

Ainsi, $D := \bigcup_{k\geqslant 1} D_k$ est aussi dénombrable.

Lemme 2.2.1 $\forall \omega \in \Omega, \mathbb{P}(\{\omega\}) = 0$. C'est-à-dire que $\mathbb{P}$ n'a pas d'atome.

Démonstration. Soit $\omega = (\omega_i)_{i \geqslant 1} \in \Omega$. On pose pour $i \geqslant 1$,

$$B_i = \begin{cases} A_i & \text{si } \omega_i = 0 \\ A_i^C & \text{si } \omega_i = 1 \end{cases}.$$

On a alors $\{\omega\} = \bigcap_{i \geqslant 1} B_i$. On pose $r = \max(p, 1-p) \in]0,1[$, donc

$$0 \leqslant \mathbb{P}(\{\omega\}) = \mathbb{P}\left(\bigcap_{i \geqslant 1} B_i\right) \leqslant \mathbb{P}\left(\bigcap_{i=1}^{k} B_i\right) = \prod_{i=1}^{k} \mathbb{P}(B_i) \leqslant r^k.$$

En faisant tendre k vers ∞, on obtient $\mathbb{P}(\omega) = 0$. □

D étant dénombrable, on en déduit que $\mathbb{P}(D) = 0$. On pose $\widetilde{D} = \pi(D)$. On a aussi $\pi^{-1}(\widetilde{D}) = D$, parce que $\omega \in D \Rightarrow \widetilde{\omega} \subset D$. Donc

$$\widetilde{\mathbb{P}}(\widetilde{D}) = \mathbb{P}\left(\pi^{-1}(\widetilde{D})\right) = \mathbb{P}(D) = 0.$$

On prend $\widetilde{\Omega}' = \widetilde{\Omega} \backslash \widetilde{D}$, alors $\widetilde{\mathbb{P}}(\widetilde{\Omega}') = \widetilde{\mathbb{P}}(\widetilde{\Omega}) - \widetilde{\mathbb{P}}(\widetilde{D}) = 1 - 0 = 1$. Si $\widetilde{\omega} \in \widetilde{\Omega}'$, alors $\widetilde{g}_k(\widetilde{\omega}) = \widetilde{\omega}$ implique $k = 0$, ce qui signifie que $\mathbb{Z}$ agit librement sur $\widetilde{\Omega}'$ par $\widetilde{g}$.

On définit *une relation d'équivalence* $\sim'$ sur $\widetilde{\Omega}'$:

$$\begin{aligned} \widetilde{\omega_1} \sim' \widetilde{\omega_2} : &\Leftrightarrow \exists\, k \in \mathbb{Z} \text{ t.q. } \widetilde{g}_k(\widetilde{\omega_1}) = \widetilde{\omega_2} \\ &\Leftrightarrow \widetilde{\omega_1} \text{ et } \widetilde{\omega_2} \text{ appartiennent à la même } \textbf{orbite} \text{ sous l'action de } \widetilde{g}. \end{aligned}$$

Dans chaque orbite, on choisit un représentant pour former un ensemble $E \subset \widetilde{\Omega}'$, c'est-à-dire que E intersecte chaque classe d'équivalence par $\sim'$ (chaque orbite sous l'action de $\widetilde{g}$) en exactement un point.

Si $k_1, k_2 \in \mathbb{Z}$, avec $k_1 \neq k_2$, et si $\widetilde{g}k_1(E) \cap \widetilde{g}k_2(E) \neq \varnothing$, alors il existe $\widetilde{\omega_1}, \widetilde{\omega_2} \in E$ tels que $\widetilde{g}_{k_1}(\widetilde{\omega_1}) = \widetilde{g}_{k_2}(\widetilde{\omega_2})$. En posant $k_0 = k_1 - k_2 \neq 0$, on obtient

$$\widetilde{g}_{k_0}(\widetilde{\omega_1}) = \widetilde{\omega_2} \text{ et donc } \widetilde{\omega_1} \neq \widetilde{\omega_2},\ \widetilde{\omega_1} \sim' \widetilde{\omega_2}.$$

Ceci contredit la définition de E comme ensemble de représentants des orbites. Ainsi, les images $\widetilde{g}_k(E)$, pour $k \in \mathbb{Z}$, sont 2-à-2 disjoints.

On a aussi

$$\widetilde{\Omega}' = \bigcup_{k \in \mathbb{Z}} \widetilde{g}_k(E).$$

Donc

$$1 = \widetilde{\mathbb{P}}(\widetilde{\Omega}') = \widetilde{\mathbb{P}}\left(\bigsqcup_{k \in \mathbb{Z}} \widetilde{g}_k(E)\right) = \sum_{k \in \mathbb{Z}} \widetilde{\mathbb{P}}\left(\widetilde{g}_k(E)\right) = \sum_{k \in \mathbb{Z}} \widetilde{\mathbb{P}}(E).$$

Ce qui est impossible. Par conséquent, on ne peut pas prendre $\mathcal{F} = \mathcal{P}(\Omega)$. □

Chapitre 3

Variables aléatoires discrétes

3.1 Variables aléatoires discrètes

Soit $(\Omega, \mathcal{F})$ un espace probabilisable.

Définition 3.1.1 (Variable aléatoire discrète) Une **variable aléatoire discrète (v.a.d.)** est une application $X : \Omega \to E$ telle que :

1) $X(\Omega) = \{X(\omega) : \omega \in \Omega\} \subset E$ est au plus dénombrable.

2) $\forall x \in X(\Omega)$, on a $X^{-1}(\{x\}) = \{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}$.

Si $E = \mathbb{R}$, on dit que X est une **v.a.d. réelle**. Si $X(\Omega)$ est fini, on dit que X est une **v.a.d. finie**.

Proposition 3.1.1 Soit $X : \Omega \to E$ une v.a.d. sur $(\Omega, \mathcal{F})$. Alors $\forall A \subset E$, on a $X^{-1}(A) \in \mathcal{F}$.

Démonstration. On a

$$\begin{aligned} X^{-1}(A) &= \{\omega \in \Omega : X(\omega) \in A\} = \{\omega \in \Omega : X(\omega) \in A \cap X(\Omega)\} \\ &= \bigcup_{x \in A \cap X(\Omega)} \{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}. \end{aligned}$$

□

Exemple 3.1.1

1) **Fonction constante** : Soit $c \in E$ fixé, on définit $X(\omega) = c$ pour tout $\omega \in \Omega$.

2) Soit $A \in \mathcal{F}$, alors $X = \mathbb{1}_A : \Omega \to \mathbb{R}$ est une v.a.d.. Toute v.a. réelle Y qui ne prend que les valeurs 0 et 1 est nécessairement une fonction indicatrice. En effet, $Y = \mathbb{1}_B$ où $B = \{\omega \in \Omega : Y(\omega) = 1\} \in \mathcal{F}$. On l'appelle une v.a. de **Bernoulli**.

Notation On note :

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

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

Ce type d'événements est appelé **événements générés par** X. En particulier,

$$\{X = x\} = \{X \in \{x\}\} = X^{-1}(x),$$

$$\mathbb{P}(X = x) = \mathbb{P}(\{X = x\}),$$

$$\{X > a\} = \{X \in]a, \infty[\,\} = X^{-1}(\,]a, \infty[\,).$$

Désormais, $(\Omega, \mathcal{F}, \mathbb{P})$ désigne un espace de probabilité.

Si X, Y sont des v.a. réelles définies sur le même $(\Omega, \mathcal{F}, \mathbb{P})$, alors

$$\{X = Y\} = \{X - Y = 0\} = \bigcup_{x \in X(\Omega) \cup Y(\Omega)} (\{X = x\} \cap \{Y = x\}) \in \mathcal{F}.$$

Théorème 3.1.1 (Loi d'une variable aléatoire discrète) Soit $X : \Omega \to E$ une v.a.d. Alors l'application

$$\begin{aligned} \mathcal{P}(E) &\to [0,1] \\ A &\mapsto \mathbb{P}(X \in A) \end{aligned}$$

est une probabilité sur $(E, \mathcal{P}(E))$, appelée **loi de** X et notée $\mathbb{P}_X$. On a

$$\mathbb{P}(X \in A) = \mathbb{P}_X(A) = \sum_{x \in A \cap X(\Omega)} \mathbb{P}(X = x).$$

Démonstration. On a $\mathbb{P}_X(E) = \mathbb{P}(X \in E) = \mathbb{P}(\Omega) = 1$.

Soit $(A_i)_{i \geqslant 1}$ une famille de parties de E deux à deux disjointes, c'est-à-dire $A_i \cap A_j = \varnothing$ pour tout $i \neq j$. Alors les événements $\{X \in A_i\}$ sont deux à deux disjoints. Donc

$$\mathbb{P}_X\left(\bigsqcup_{i \geqslant 1} A_i\right) = \mathbb{P}\left(X \in \bigsqcup_{i \geqslant 1} A_i\right) = \mathbb{P}\left(\bigsqcup_{i \geqslant 1} \{X \in A_i\}\right) = \sum_{i \geqslant 1} \mathbb{P}(X \in A_i) = \sum_{i \geqslant 1} \mathbb{P}_X(A_i).$$

Ainsi, $\mathbb{P}_X$ est une probabilité. □

Définition 3.1.2 (Variable presque sûrement constante) $X : \Omega \to E$ v.a.d. est dite **presque sûrement constante** s'il existe $a \in E$ tel que $\mathbb{P}(X = a) = 1$. On note $X = a$ **p.s.** .

Remarque

1) Si $X : \Omega \to E$ est une v.a.d., alors pour $x \neq y$, on a $\{X = x\} \cap \{X = y\} = \varnothing$ et

$$\bigcup_{x \in X(\Omega)} \{X = x\} = \Omega.$$

On en déduit que $\{X = x\}, x \in X(\Omega)$ est un **système complet d'événements associés à** X. On a

$$\sum_{x \in X(\Omega)} \mathbb{P}(X = x) = 1.$$

2) Toute probabilité sur un ensemble E au plus dénombrable peut être considérée comme la loi d'une v.a.d..

Démonstration. 2) Soit $\mathbb{P}$ une probabilité sur E. Alors il existe $(p_x)_{x \in E}$ tel que $p_x \in [0,1]$, $\sum_{x \in E} p_x = 1$ et $\mathbb{P}(A) = \sum_{x \in A} p_x$. On prend $(\Omega, \mathcal{F}) = (E, \mathcal{P}(E))$ et $X = \mathrm{id} : \Omega \to E$. Alors $X(\Omega) = E$ et $\mathbb{P}_X(\{x\}) = \mathbb{P}(X = x) = p_x$. □

Définition 3.1.3 (Variables de même loi) Soient $X, Y : \Omega \to E$ deux v.a.d.. On dit qu'elles **ont même loi** (ou sont **équidistribuées**, ou **identiquement distribuées**) si $\mathbb{P}_X = \mathbb{P}_Y$. On note $X \sim Y$ ou $X \overset{(\text{loi})}{=} Y$ ou $X \overset{(\text{d})}{=} Y$. Si X est une v.a.d. et L une loi discrète, on note $X \sim L$ si $\mathbb{P}_X = L$.

Exemple 3.1.2 Soit $X = \mathbb{1}_A$. Alors $X(\Omega) = \{0, 1\}$, $\mathbb{P}_X(\{1\}) = \mathbb{P}(X = 1) = \mathbb{P}(A)$, $\mathbb{P}_X(\{0\}) = \mathbb{P}(X = 0) = \mathbb{P}\left(A^C\right) = 1 - \mathbb{P}(A)$, et $\mathbb{P}_X(\{x\}) = 0$ pour tout $x \notin \{0, 1\}$. Si $\mathbb{P}_X(\{1\}) = \mathbb{P}_X(\{0\}) = \frac{1}{2} = \mathbb{P}(A)$, alors $X = \mathbb{1}_A$ et $1 - X = \mathbb{1}_{A^C}$ ont la même loi, c'est-à-dire $\mathbb{P}_X = \mathbb{P}_{1-X}$. Mais X et $1 - X$ ne sont jamais égales.

Exercice 3.1.1 Soit $X : \Omega \to \mathbb{R}$ une v.a.d.. Calculez :

1) $\lim\limits_{x \to +\infty} \mathbb{P}(X \leqslant x)$,
2) $\lim\limits_{x \to -\infty} \mathbb{P}(X \leqslant x)$.

Solution.

1) On note $A_n = \{X \leqslant n\}$ pour $n \in \mathbb{Z}$. Alors $(A_n)_{n \in \mathbb{N}}$ est une suite croissante et $\bigcup\limits_{n \in \mathbb{N}} A_n = \Omega$. Donc, par continuité par le bas de $\mathbb{P}$,

$$\lim_{n \to +\infty} \mathbb{P}(A_n) = \mathbb{P}(\Omega) = 1.$$

Or, pour tout $x > 0$, on note $A_x = \{X \leqslant x\}$. On a $\mathbb{P}\left(A_{\lfloor x \rfloor}\right) \leqslant \mathbb{P}(A_x) \leqslant \mathbb{P}\left(A_{\lfloor x \rfloor + 1}\right)$. Donc $\lim\limits_{x \to \infty} \mathbb{P}(A_x) = 1$.

2) Pour $n \in \mathbb{N}$, on note $A_{-n-1} \subset A_{-n}$. La suite $(A_{-n})_{n \in \mathbb{N}}$ est décroissante et $\bigcap\limits_{n \in \mathbb{N}} A_{-n} = \varnothing$. La continuité par le haut de $\mathbb{P}$ implique que

$$\lim_{k \to -\infty} \mathbb{P}(A_k) = \mathbb{P}\left(\bigcap_{n \in \mathbb{N}} A_{-n}\right) = \mathbb{P}(\varnothing) = 0.$$

Par le même argument ci-dessus, on a $\lim\limits_{x \to -\infty} \mathbb{P}(A_x) = 0$.

□

Proposition 3.1.2 Soit $X : \Omega \to E$ une v.a.d. et $f : E \to F$ une application.
Alors $f \circ X : \Omega \to F$ est une v.a.d. notée $f(X)$ et

$$\forall y \in f(X)(\Omega),\ \mathbb{P}(f(X) = y) = \sum_{x \in f^{-1}(y)} \mathbb{P}(X = x).$$

Démonstration. D'abord, $f \circ X(\Omega)$ est au plus dénombrable car $X(\Omega)$ est au plus dénombrable. Ensuite, pour tout $y \in f(X(\Omega))$, on a

$$(f \circ X)^{-1}(y) = X^{-1}(f^{-1}(y)) = \bigsqcup_{x \in X(\Omega) \cap f^{-1}(y)} X^{-1}(x) \in \mathcal{F}.$$

Donc

$$\mathbb{P}\left(f(X)=y\right)=\mathbb{P}\left((f\circ X)^{-1}(y)\right)=\sum_{x\in X(\Omega)\cap f^{-1}(y)}\mathbb{P}\left(X^{-1}(x)\right)=\sum_{x\in f^{-1}(y)}\mathbb{P}\left(X=x\right).$$

□

Définition 3.1.4 (Variables presque sûrement égales) Deux v.a.d. X, Y définies sur le même $(\Omega,\mathcal{F},\mathbb{P})$ et à valeurs dans le même ensemble E sont dites **presque sûrement égales** si

$$\mathbb{P}\left(X=Y\right)=\mathbb{P}\left(\{\omega\in\Omega : X(\omega)=Y(\omega)\}\right)=1.$$

Autrement dit, X et Y diffèrent sur un ensemble négligeable.

Exercice 3.1.2 Montrer que si X et Y sont p.s. égales, alors X et Y ont la même loi.

Remarque X et Y ont même loi $\not\Rightarrow$ X et Y sont p.s. égales. Deux v.a.d. réelles définies sur des espaces de probabilité différents peuvent avoir la même loi.

Exemple 3.1.3 Considérons le lancer d'une pièce : Pile (0) et Face (1).

Soit

$$\begin{aligned} X:\Omega=\{0,1\}&\to\{0,1\}\\ \omega&\mapsto\omega. \end{aligned}$$

Alors $\mathbb{P}\left(X=0\right)=\mathbb{P}\left(X=1\right)=\dfrac{1}{2}$.

Considérons maintenant un dé à 6 faces équilibré. Soit

$$\begin{aligned} Y:\Omega'=[\![1,6]\!]&\to\{0,1\}\\ \omega&\mapsto \mathbb{1}_{\{\omega \text{ pair}\}}. \end{aligned}$$

Soit $A=$ « le résultat est pair ». Alors $\mathbb{P}'(Y=0)=\mathbb{P}'(A^C)=\dfrac{1}{2}$, $\mathbb{P}'(Y=1)=\dfrac{1}{2}$. Ainsi, X et Y ont la même loi.

3.2 Lois discrètes usuelles

3.2.1 Loi uniforme discrète

Soit $X:\Omega\to E$ avec $\mathrm{Card}\,(E)<+\infty$. On dit que X est une **v.a. uniforme** (discrète) sur E (notée $X\sim\mathcal{U}(E)$ ou $X\sim\mathcal{U}\mathrm{nif}(E)$) si et seulement si $\mathbb{P}_X$ est la probabilité uniforme sur E.

$$\mathbb{P}\left(X=x\right)=\frac{1}{\mathrm{Card}\,(E)},\quad\forall x\in E.$$

Cas particuliers :

— $E=[\![1,n]\!]$,

— $E=[\![a,b]\!]$ avec $a,b\in\mathbb{Z}$, $a<b$: $X(\Omega)=[\![a,b]\!]$ et

$$\mathbb{P}\left(X=k\right)=\frac{1}{b-a+1},\quad\forall a\leqslant k\leqslant b,\quad X\sim\mathcal{U}([\![a,b]\!]).$$

3.2.2 Loi de Bernoulli

X est une **v.a. de Bernoulli** si $X(\Omega) = \{0,1\}$, ce qui implique $X = \mathbb{1}_{\{X=1\}}$. La loi $\mathbb{P}_X$ est caractérisée par $\mathbb{P}(X=1) = p \in [0,1]$. On note $X \sim \mathcal{B}(p)$ ou $X \sim \mathcal{B}\text{ern}(p)$.

3.2.3 Loi binomiale

Soient $n \geqslant 1$ et $p \in [0,1]$. On dit que X suit une loi binomiale de paramètres n et p ($X \sim \mathcal{B}\text{in}(n,p)$ ou $X \sim \mathcal{B}(n,p)$) si $X(\Omega) = [\![0,n]\!]$ et

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

On a $\mathcal{B}(1,p) = \mathcal{B}(p)$.

Proposition 3.2.1 Si $X_1, \dots, X_n$ sont des v.a. de Bernoulli $\mathcal{B}(p)$ indépendantes, alors $X_1 + \cdots + X_n \sim \mathcal{B}(n,p)$.

3.2.4 Loi géométrique

Soit $p \in]0,1[$. On dit que $X \sim \mathcal{G}(p)$ ou $\mathcal{G}\text{eom}(p)$ si $X(\Omega) = \mathbb{N}^*$ et

$$\mathbb{P}(X=n) = (1-p)^{n-1}p = q^{n-1}p, \quad \forall n \geqslant 1.$$

Une variante : $\mathcal{G}_0(p)$ ou $\mathcal{G}\text{eom}_0(p)$ si $X(\Omega) = \mathbb{N}$ et $\mathbb{P}(X=n) = q^n p$ pour tout $n \geqslant 0$.

Proposition 3.2.2 (Absence de mémoire) Si $X \sim \mathcal{G}(p)$, alors $\forall k, \ell \in \mathbb{N}$, on a $\mathbb{P}(X > k) = q^k$ et $\mathbb{P}(X > k+\ell \mid X > k) = \mathbb{P}(X > \ell)$. Cette propriété caractérise la loi géométrique.

Démonstration. On a

$$\mathbb{P}(X > k) = \mathbb{P}\left(\bigsqcup_{n=k+1}^{\infty} \{X=n\} \right) = \sum_{n \geqslant k+1} \mathbb{P}(X=n) = \sum_{n \geqslant k+1} q^{n-1}p = q^k.$$

De plus,

$$\mathbb{P}(X > k+\ell \mid X > k) = \frac{\mathbb{P}(X > k+\ell)}{\mathbb{P}(X > k)} = q^\ell = \mathbb{P}(X > \ell).$$

□

Temps d'attente du r-ième succès, $r \geqslant 1$ On considère une suite infinie d'épreuves de Bernoulli indépendantes et de même probabilité de succès $p \in]0,1[$. Soit $Y_r =$ « le nombre de tirages pour obtenir r succès pour la première fois ». On pose $Y_r = +\infty$ si on n'obtient jamais r succès. Ainsi, $Y_r(\Omega) = \mathbb{N}^* \cup \{+\infty\}$.

Vérifions d'abord que Y_r est bien une v.a.d. Ici $\Omega = \{0,1\}^{\mathbb{N}^*}$, $\mathcal{F} = \sigma(\{A_n\}_{n \geqslant 1})$, où $A_n =$ « la n-ième épreuve est un succès ».

— Si $k < r$, $\{Y_r = k\} = \varnothing \in \mathcal{F}$.

— Si $k \geqslant r$,

$$\{Y_r = k\} = \bigcup_{\substack{I \subset [\![1,k-1]\!] \\ \mathrm{Card}(I)=r-1}} \left(\left(\bigcap_{i \in I} A_i \right) \cap \left(\bigcap_{i \in [\![1,k-1]\!] \setminus I} A_i^C \right) \right) \cap A_k.$$

— Si $k = +\infty$, $\{Y_r = +\infty\} = \left(\bigcup_{k \in \mathbb{N}^*} \{Y_r = k\} \right)^C \in \mathcal{F}$.

Ainsi, Y_r est bien une v.a.d. à valeurs dans $\mathbb{N}^* \cup \{+\infty\}$.

On remarque que

$$\{Y_r = +\infty\} \subsetneq \bigcup_{N \in \mathbb{N}^*} \bigcap_{k \geqslant N} A_k^C$$

$$= \text{« à partir d'un certain rang, toutes les épreuves donnent échecs »}.$$

Si $k < r$, $\mathbb{P}(Y_r = k) = 0$. Si $k \geqslant r$,

$$\mathbb{P}(Y_r = k) = \sum_{\substack{I \subset [\![1,k-1]\!] \\ \mathrm{Card}(I)=r-1}} p^{\mathrm{Card}(I)} q^{k-1-\mathrm{Card}(I)} p = \binom{k-1}{r-1} p^r q^{k-r}.$$

Cette loi s'appelle **la loi de Pascal**.

On peut montrer que $\mathbb{P}(Y_r = +\infty) = 0$ en calculant :

$$\begin{aligned}
\sum_{k \in \mathbb{N}^*} \mathbb{P}(Y_r = k) &= \sum_{k \geqslant r} \binom{k-1}{r-1} p^r q^{k-r} \\
&= \frac{p^r}{(r-1)!} \cdot \frac{\partial^{r-1}}{\partial q^{r-1}} \sum_{k \geqslant r} q^{k-1} \\
&= \frac{p^r}{(r-1)!} \cdot \frac{\partial^{r-1}}{\partial q^{r-1}} \left(\sum_{k \geqslant 1} q^{k-1} - C_r \right) \quad (\text{où } C_r = \sum_{k=1}^{r-1} q^{k-1} \text{ est une constante}) \\
&= \frac{p^r}{(r-1)!} \cdot \frac{\partial^{r-1}}{\partial q^{r-1}} \frac{1}{1-q} \\
&= \frac{p^r}{(1-q)^r} = 1.
\end{aligned}$$

Donc $\mathbb{P}(Y_r = +\infty) = 1 - \sum_{k \in \mathbb{N}^*} \mathbb{P}(Y_r = k) = 0$.

3.2.5 Loi de Poisson

Définition 3.2.1 Soit $\lambda \in]0, +\infty[$. On dit que $X \sim \mathcal{P}(\lambda)$ ou $\mathcal{P}\mathrm{oi}(\lambda)$ si $X(\Omega) = \mathbb{N}$ et

$$\mathbb{P}(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \quad k \geqslant 0.$$

Une binomiale avec beaucoup de tentatives mais une probabilité de succès « faible » est approximativement une loi de Poisson.

Proposition 3.2.3 Soit $(p_n)_{n\geqslant 1}$ une suite de réels dans $]0,1[$ telle que $p_n \sim \dfrac{\lambda}{n}$ quand $n \to \infty$, où $\lambda > 0$. Soit $X_n \sim \mathcal{B}\text{in}(n, p_n)$. Alors pour tout $k \geqslant 0$, on a

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}.$$

Démonstration. Cas où $p_n = \dfrac{\lambda}{n}$: Soit $k \geqslant 0$ fixé et $n \geqslant k$.

$$\begin{aligned}
\mathbb{P}(X_n = k) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\
&= \frac{n!}{k!(n-k)!} \cdot \frac{\lambda^k}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} \cdot \frac{n(n-1)\dots(n-k+1)}{n^k} \cdot \frac{1}{\left(1-\frac{\lambda}{n}\right)^k} \cdot \left(1-\frac{\lambda}{n}\right)^n \\
&\xrightarrow[n\to\infty]{} e^{-\lambda} \cdot \frac{\lambda^k}{k!}.
\end{aligned}$$

□

3.3 Couple de variables aléatoires discrétes

Définition 3.3.1 (Couple de v.a.d.) Soient $X : (\Omega, \mathcal{F}) \to E$ et $Y : (\Omega, \mathcal{F}) \to E'$ deux v.a.d. L'application

$$\begin{aligned}
\Omega &\to E \times E' \\
\omega &\mapsto (X(\omega), Y(\omega))
\end{aligned}$$

est appelée **couple de v.a.d. sur** $(\Omega, \mathcal{F})$. On le note (X, Y). Si $E = E' = \mathbb{R}$, (X, Y) est un **couple de v.a.d. réelles**.

Proposition 3.3.1

- Si $X : (\Omega, \mathcal{F}) \to E$ et $Y : (\Omega, \mathcal{F}) \to E'$ sont deux v.a.d., alors (X, Y) est une v.a.d..
- Si $Z : (\Omega, \mathcal{F}) \to E \times E'$ est une v.a.d., elle peut s'écrire comme (X, Y), où X et Y sont des v.a.d..

Démonstration.

- On a $(X, Y)(\Omega) \subset X(\Omega) \times Y(\Omega)$, qui est au plus dénombrable. De plus, pour tout $(x, y) \in (X, Y)(\Omega)$,
$$(X, Y)^{-1}((x, y)) = X^{-1}(x) \cap Y^{-1}(y) \in \mathcal{F}.$$
- Réciproquement, si $Z : \Omega \to E \times E'$ est une v.a.d., on note $\pi_1 : E \times E' \to E$ et $\pi_2 : E \times E' \to E'$ les projections canoniques.
  Alors $Z(\Omega)$ est au plus dénombrable. On prend $X = \pi_1 \circ Z$, $Y = \pi_2 \circ Z$, donc $Z = (X, Y)$.

On a $X(\Omega) = \pi_1(Z(\Omega))$, $Y(\Omega) = \pi_2(Z(\Omega))$ le sont également.

Pour tout $x \in E$, on a $X^{-1}(x) = Z^{-1}(\{x\} \times E') \in \mathcal{F}$. Idem pour Y.

□

Remarque Soit $(X, Y) : (\Omega, \mathcal{F}) \to E \times E'$ un couple de v.a.d..

Alors $(\{X = x\} \cap \{Y = y\})_{(x,y)\in X(\Omega)\times Y(\Omega)}$ est un système complet d'événement de Ω.

Proposition 3.3.2 Si $X : (\Omega, \mathcal{F}) \to \mathbb{R}$, $Y : (\Omega, \mathcal{F}) \to \mathbb{R}$ sont sont deux v.a.d. réelles et définies sur le même $(\Omega, \mathcal{F}, \mathbb{P})$. Alors $XY : \Omega \to \mathbb{R}$, $\omega \mapsto X(\omega) \cdot Y(\omega)$ l'est également.

Démonstration.

$(XY)(\Omega) \subset \bigcup_{y\in Y(\Omega)} yX(\Omega)$ est au plus dénombrable, car $Y(\Omega)$, $yX(\Omega) = \{yX(\omega) : \omega \in \Omega\}$ sont au plus dénombrables. Pour tout $r \in \mathbb{R}$,

* si $r = 0$, $(XY)^{-1}(0) = \{X = 0\} \cup \{Y = 0\} \in \mathcal{F}$,
* si $r \neq 0$, $(XY)^{-1}(r) = \bigcup_{y\in Y(\Omega)\setminus\{0\}} \left\{X = \frac{r}{y}\right\} \cap \{Y = y\} \in \mathcal{F}$.

Donc XY est une v.a.d. réelle. □

Désormais, $(\Omega, \mathcal{F}, \mathbb{P})$ désigne un espace probabilisé.

Définition 3.3.2 (Loi conjointe) Soient $X : (\Omega, \mathcal{F}) \to E$ et $Y : (\Omega, \mathcal{F}) \to E'$ deux v.a.d.. La **loi conjointe de X et Y** est la loi du couple (X, Y). Elle est déterminée par

$$p_{x,y} = \mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(X = x, Y = y), \quad (x, y) \in X(\Omega) \times Y(\Omega).$$

Définition 3.3.3 (Lois marginales) Si (X, Y) est un couple de v.a.d., alors la loi de X est appelée **première loi marginale du couple**, et la loi de Y est appelée **deuxième loi marginale du couple**.

Théorème 3.3.1 Si (X, Y) est un couple de v.a.d.. Alors, pour tout $x \in X(\Omega)$,

$$\mathbb{P}(X = x) = \sum_{y\in Y(\Omega)} \mathbb{P}(X = x, Y = y),$$

et pour tout $y \in Y(\Omega)$,

$$\mathbb{P}(Y = y) = \sum_{x\in X(\Omega)} \mathbb{P}(X = x, Y = y).$$

Démonstration. On a $\Omega = \bigcup_{y\in Y(\Omega)} \{Y = y\}$, et $(\{Y = y\}, y \in Y(\Omega))$ est un système complet d'événements sur Ω. Donc

$$\{X = x\} = \bigsqcup_{y\in Y(\Omega)} \{X = x\} \cap \{Y = y\}.$$

La formule pour $\mathbb{P}(X = x)$ découle de la σ-additivité de $\mathbb{P}$. La preuve est similaire pour la seconde formule. □

Exercice 3.3.1 On considère une suite d'épreuves de Bernoulli indépendantes de probabilité de succès $p \in]0,1[$. On note $X =$ « le rang du 1er succès », $Y =$ « le rang du 2e succès ».

1) Déterminer la loi du couple (X, Y).

2) En déduire les lois de X et Y.

Solution.

1) Pour $k, m \in \mathbb{N}^*$:

— Si $k \geqslant m$, alors $\{X = k\} \cap \{Y = m\} = \varnothing$ car $X < Y$.

— Si $m = k + l$ avec $l > 0$, alors

$$\mathbb{P}(X = k, Y = k + l) = (1-p)^{k-1} p (1-p)^{l-1} p = (1-p)^{m-2} p^2.$$

2)

$$\begin{aligned}
\mathbb{P}(X = k) &= \sum_{l \geqslant 1} \mathbb{P}(X = k, Y = k + l) \\
&= \sum_{l \geqslant 1} (1-p)^{k+l-2} p^2 \\
&= p(1-p)^{k-1}, \quad k \geqslant 1.
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(Y = m) &= \sum_{k=1}^{+\infty} \mathbb{P}(X = k, Y = m) \\
&= \sum_{k=1}^{m-1} \mathbb{P}(X = k, Y = m) \\
&= \sum_{k=1}^{m-1} (1-p)^{m-2} p^2 \\
&= (m-1) p^2 (1-p)^{m-2}, \quad m \geqslant 2.
\end{aligned}$$

□

Définition 3.3.4 (Loi conditionnelle) Soit (X, Y) un couple de v.a.d.. Pour $y \in Y(\Omega)$ tel que $\mathbb{P}(Y = y) \neq 0$, la **loi conditionnelle de X sachant** $\{Y = y\}$ est la loi de X dans $(\Omega, \mathcal{F}, \mathbb{P}_{Y=y})$.

$$\mathbb{P}_{Y=y}(X = x) = \mathbb{P}(X = x \mid Y = y).$$

De même, on définit la loi conditionnelle de Y sachant $\{X = x\}$.

Exemple 3.3.1 Reprenons l'exercice précédent.

$$\mathbb{P}_{\{Y=j\}}(X = k) = \begin{cases} 0 & \text{si } k \geqslant j, \\ \dfrac{p^2(1-p)^{j-2}}{(j-1)p^2(1-p)^{j-2}} = \dfrac{1}{j-1} & \text{si } k < j. \end{cases}$$

Ainsi, la loi conditionnelle de X sachant $\{Y = j\}$ est la loi uniforme sur $[\![1, j-1]\!]$.

$$\mathbb{P}_{\{X=i\}}(Y - i = k) = \frac{p^2(1-p)^{k+i-2}}{p(1-p)^{i-1}} = p(1-p)^{k-1}.$$

Donc la loi conditionnelle de $Y - i$ sachant $\{X = i\}$ est la loi $\mathcal{G}\text{eom}(p)$.

Plus généralement :

Définition 3.3.5 (n-uplet de v.a. / vecteur aléatoire discret)

Soit $n \in \mathbb{N}^*$. Si $X_i : \Omega \to E_i$ $(i = 1, \ldots, n)$ sont des v.a.d. sur l'espace $(\Omega, \mathcal{F})$, alors l'application

$$\begin{aligned} \Omega &\to E_1 \times \cdots \times E_n \\ \omega &\mapsto (X_1(\omega), \ldots, X_n(\omega)) \end{aligned}$$

est appelée **n-uplet de v.a. sur** Ω. On le note $(X_1, \ldots, X_n)$. Si $E_1 = \cdots = E_n = \mathbb{R}$, $(X_1, \ldots, X_n)$ est appelé un **vecteur aléatoire discret**.

Proposition 3.3.3 Le n-uplet de v.a.d. $(X_1, \ldots, X_n)$ est une v.a.d. à valeurs dans $E_1 \times \cdots \times E_n$.

Démonstration. $(X_1, \ldots, X_n)(\Omega) \subset X_1(\Omega) \times \cdots \times X_n(\Omega)$ est au plus dénombrable. De plus,

$$\{(X_1, \ldots, X_n) = (x_1, \ldots, x_n)\} = \bigcap_{i=1}^{n} \{X_i = x_i\} \in \mathcal{F}.$$

□

Définition 3.3.6 (Loi conjointe, lois marginales) **La loi conjointe de** $X_1, \ldots, X_n$ est la loi du n-uplet $(X_1, \ldots, X_n)$. **Les lois marginales** du n-uplet $(X_1, \ldots, X_n)$ sont les lois des v.a.d. $X_1, \ldots, X_n$.

$$\mathbb{P}(X_i = x_i) = \sum_{j \neq i} \sum_{x_j \in X_j(\Omega)} \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n).$$

3.4 Indépendance des variables aléatoires

Définition 3.4.1 Deux v.a.d X et Y définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ sont dites **indépendantes** (noté $X \perp\!\!\!\perp Y$) si

$$\forall (x, y) \in X(\Omega) \times Y(\Omega), \quad \{X = x\} \text{ et } \{Y = y\} \text{ sont indépendants,}$$

c'est-à-dire $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y)$.

Exercice 3.4.1 On suppose que la loi conjointe de (X, Y) s'écrit comme

$$\mathbb{P}(X = x, Y = y) = \varphi(x)\psi(y),$$

où $\varphi : X(\Omega) \to \mathbb{R}$ et $\psi : Y(\Omega) \to \mathbb{R}$. Montrer que X et Y sont indépendantes et qu'il existe des constantes c, c' telles que $cc' = 1$, $\mathbb{P}(X = x) = c\varphi(x)$ et $\mathbb{P}(Y = y) = c'\psi(y)$.

Démonstration. Comme

$$\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} \mathbb{P}(X=x,Y=y)=1$$

et $0\leqslant\varphi(x)\psi(y)\leqslant 1$, il existe $(x_0,y_0)\in X(\Omega)\times Y(\Omega)$ tel que $\varphi(x_0)\psi(y_0)>0$. Alors pour tout $x\in X(\Omega)$, $\varphi(x)\psi(y_0)=\mathbb{P}(X=x,Y=y_0)\geqslant 0$ implique que $\varphi(x)$ a le même signe que $\varphi(x_0)$. De même, pour tout $y\in Y(\Omega)$, $\psi(y)$ a le même signe que $\psi(y_0)$.

Pour tout $x\in X(\Omega)$,

$$1\geqslant\mathbb{P}(X=x)=\sum_{y\in Y(\Omega)}\mathbb{P}(X=x,Y=y)=\varphi(x)\sum_{y\in Y(\Omega)}\psi(y).$$

Donc $c:=\sum_{y\in Y(\Omega)}\psi(y)<+\infty$ et $\mathbb{P}(X=x)=c\varphi(x)$ pour tout $x\in X(\Omega)$.

De même, $c':=\sum_{x\in X(\Omega)}\varphi(x)<+\infty$ et $\mathbb{P}(Y=y)=c'\psi(y)$ pour tout $y\in Y(\Omega)$.

Finalement,

$$1=\sum_{(x,y)\in X(\Omega)\times Y(\Omega)}\mathbb{P}(X=x,Y=y)=(\sum_{y\in Y(\Omega)}\psi(y))(\sum_{x\in X(\Omega)}\varphi(x))=cc'.$$

Donc $\mathbb{P}(X=x,Y=y)=\mathbb{P}(X=x)\,\mathbb{P}(Y=y)$, c'est-à-dire $X\perp\!\!\!\perp Y$. □

Proposition 3.4.1 Si $X:\Omega\to E$ et $Y:\Omega\to F$ sont deux v.a.d. indépendantes, alors

$$\mathbb{P}(X\in A,Y\in B)=\mathbb{P}(X\in A)\,\mathbb{P}(Y\in B),\quad\forall A\subset E,B\subset F.$$

C'est-à-dire, $\{X\in A\}$ et $\{Y\in B\}$ sont indépendants.

Démonstration.

$$\begin{aligned}\mathbb{P}(X\in A,Y\in B)&=\mathbb{P}(X\in A\cap X(\Omega),Y\in B\cap Y(\Omega))\\&=\sum_{x\in A\cap X(\Omega),\ y\in B\cap Y(\Omega)}\mathbb{P}(X=x,Y=y)\\&=\sum_{x\in A\cap X(\Omega),\ y\in B\cap Y(\Omega)}\mathbb{P}(X=x)\,\mathbb{P}(Y=y)\\&=\left(\sum_{x\in A\cap X(\Omega)}\mathbb{P}(X=x)\right)\left(\sum_{y\in B\cap Y(\Omega)}\mathbb{P}(Y=y)\right)\\&=\mathbb{P}(X\in A)\,\mathbb{P}(Y\in B).\end{aligned}$$

□

Proposition 3.4.2 Soit (X,Y) un couple de v.a.d.. Alors $X\perp\!\!\!\perp Y$ si et seulement si pour tout $y\in Y(\Omega)$ tel que $\mathbb{P}(Y=y)>0$, on a

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

Démonstration. Si $\mathbb{P}(Y=y)>0$, alors

$$\mathbb{P}_{\{Y=y\}}(X=x)=\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)}=\mathbb{P}(X=x)\Leftrightarrow\mathbb{P}(X=x,Y=y)=\mathbb{P}(X=x)\,\mathbb{P}(Y=y).$$

□

Théorème 3.4.1 (Indépendance par transformation) Si X, Y des v.a.d. indépandences sur $(\Omega,\mathcal{F},\mathbb{P})$, alors $\forall f: X(\Omega)\to E$, $g: X(\Omega)\to F$, les v.a.d. $f(X)$ et $g(Y)$ sont indépandentes.

Démonstration. $\{f(X)=x\}=\{X\in f^{-1}(x)\}$, $\{g(Y)=y\}=\{Y\in g^{-1}(y)\}$. Donc

$$\begin{aligned}\mathbb{P}(f(X)=x,g(Y)=y)&=\mathbb{P}\left(X\in f^{-1}(x),Y\in g^{-1}(y)\right)\\&=\mathbb{P}\left(X\in f^{-1}(x)\right)\mathbb{P}\left(Y\in g^{-1}(y)\right)\quad(\text{car } X\perp\!\!\!\perp Y)\\&=\mathbb{P}(f(X)=x)\,\mathbb{P}(g(Y)=y).\end{aligned}$$

□

Plus généralement, soient $X_j:\Omega\to E_j$ des v.a.d., pour $1\leqslant j\leqslant n$.

Définition 3.4.2 (Indépendantes 2 à 2, mutuellement indépendantes)

Les $(X_j)_{1\leqslant j\leqslant n}$ sont **indépendantes 2 à 2** si

$$\forall i\neq j,\quad X_i\perp\!\!\!\perp X_j.$$

Elles sont **(mutuellement) indépendantes** si

$$\forall(x_1,\dots,x_n)\in E_1\times\cdots\times E_n,\quad\mathbb{P}(X_1=x_1,\dots,X_n=x_n)=\prod_{j=1}^{n}\mathbb{P}(X_j=x_j).$$

Proposition 3.4.3 Si $(X_j)_{1\leqslant j\leqslant n}$ sont indépendantes, alors

$$\forall A_j\subset E_j,\quad\mathbb{P}(X_1\in A_1,\dots,X_n\in A_n)=\prod_{j=1}^{n}\mathbb{P}(X_j\in A_j).$$

En particulier, $(X_j)_{1\leqslant j\leqslant n}$ sont indépendantes 2 à 2.

Démonstration.

$$\begin{aligned}
\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) &= \sum_{x_j \in A_j \cap X_j(\Omega),\ 1 \leqslant j \leqslant n} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\
&= \sum_{x_j \in A_j \cap X_j(\Omega),\ 1 \leqslant j \leqslant n} \mathbb{P}(X_1 = x_1) \dots \mathbb{P}(X_n = x_n) \\
&= \prod_{j=1}^{n} \left(\sum_{x_j \in A_j \cap X_j(\Omega)} \mathbb{P}(X_j = x_j) \right) = \prod_{j=1}^{n} \mathbb{P}(X_j \in A_j).
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(X_i = x_i, X_j = x_j) &= \sum_{x_k \in X_k(\Omega),\ k \neq i,j} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\
&= \sum_{x_k \in X_k(\Omega),\ k \neq i,j} \mathbb{P}(X_1 = x_1) \dots \mathbb{P}(X_n = x_n) \\
&= \mathbb{P}(X_i = x_i)\, \mathbb{P}(X_j = x_j) \prod_{k \neq i,j} \left(\sum_{x_k \in X_k(\Omega)} \mathbb{P}(X_k = x_k) \right) \\
&= \mathbb{P}(X_i = x_i)\, \mathbb{P}(X_j = x_j).
\end{aligned}$$

□

Proposition 3.4.4 Si $(X_j)_{1 \leqslant j \leqslant n}$ sont indépendantes.

1) Soient $f_j : E_j \to F_j$ des fonctions. Alors $(f_j(X_j))_{1 \leqslant j \leqslant n}$ sont aussi indépendantes.

2) Soient $f : E_1 \times \cdots \times E_p \to H_1$ et $g : E_{p+1} \times \dots E_n \to H_2$ deux fonctions. Alors $f(X_1, \dots, X_p) \perp\!\!\!\perp g(X_{p+1}, \dots, X_n)$.

Démonstration.

1)

$$\begin{aligned}
&\mathbb{P}(f_1(X_1) = y_1, \dots, f_n(X_n) = y_n) \\
&\qquad = \mathbb{P}\left(\bigcap_{j=1}^{n} \{X_j \in f_j^{-1}(y_j)\} \right) = \prod_{j=1}^{n} \mathbb{P}\left(X_j \in f_j^{-1}(y_j)\right) = \prod_{j=1}^{n} \mathbb{P}(f_j(X_j) = y_j).
\end{aligned}$$

2)

$$\begin{aligned}
&\mathbb{P}(f(X_1, \dots, X_p) = h_1,\ g(X_{p+1}, \dots, X_n) = h_2) \\
&= \sum_{\substack{(x_1, \dots, x_p) \in f^{-1}(h_1) \\ (x_{p+1}, \dots, x_n) \in g^{-1}(h_2)}} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\
&= \sum_{(x_1, \dots, x_p) \in f^{-1}(h_1)} \mathbb{P}(X_1 = x_1, \dots, X_p = x_p) \sum_{(x_{p+1}, \dots, x_n) \in g^{-1}(h_2)} \mathbb{P}(X_{p+1} = x_{p+1}, \dots, X_n = x_n) \\
&= \mathbb{P}(f(X_1, \dots, X_p) = h_1)\, \mathbb{P}(g(X_{p+1}, \dots, X_n) = h_2).
\end{aligned}$$

□

Remarque Il est facile de vérifier que $(X_1, \dots, X_p) \perp\!\!\!\perp (X_{p+1}, \dots, X_n)$. Ensuite, on peut utiliser indépendance par transformation.

Lemme 3.4.1 Si X et Y sont deux v.a.d. de même loi, alors pour toute fonction $f : E \to F$, $f(X)$ et $f(Y)$ ont la même loi.

Exercice 3.4.2 Soient $X \sim \mathcal{B}(n,p)$, $Y \sim \mathcal{B}(m,p)$ avec $X \perp\!\!\!\perp Y$. Montrer que $X+Y \sim \mathcal{B}(n+m,p)$.

Solution 1. Calculer $\mathbb{P}(X+Y=k)$. □

Solution 2. Soient $(X_i)_{1\leqslant i\leqslant n+m}$ des v.a. de loi de Bernoulli $\mathcal{B}(p)$ et indépendantes. Posons

$$X' = \sum_{i=1}^{n} X_i \sim \mathcal{B}(n,p) \quad \text{et} \quad Y' = \sum_{i=n+1}^{n+m} X_i \sim \mathcal{B}(m,p).$$

Alors (X', Y') a même loi que (X, Y), donc $X' + Y' \overset{(loi)}{=} X + Y$. Or

$$X' + Y' = \sum_{i=1}^{n+m} X_i \sim \mathcal{B}(n+m,p).$$

□

Exercice 3.4.3 Soient $X_1, \dots, X_n$ des v.a.d. indépendantes avec $X_k \sim \mathcal{P}(\lambda_k)$ pour tout $1 \leqslant k \leqslant n$. Montrer que $X_1 + \cdots + X_n \sim \mathcal{P}(\lambda_1 + \cdots + \lambda_n)$.

Démonstration. Cas $n = 2$: Pour tout $k \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{P}(X_1 + X_2 = k) &= \mathbb{P}\left(\bigsqcup_{j=0}^{k} \{X_1 = j\} \cap \{X_2 = k-j\}\right) \\
&= \sum_{j=0}^{k} \mathbb{P}(X_1 = j)\, \mathbb{P}(X_2 = k-j) \\
&= \sum_{j=0}^{k} e^{-\lambda_1} \frac{\lambda_1^j}{j!} \cdot e^{-\lambda_2} \frac{\lambda_2^{k-j}}{(k-j)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{j=0}^{k} \binom{k}{j} \lambda_1^j \lambda_2^{k-j} \\
&= e^{-(\lambda_1+\lambda_2)} \cdot \frac{(\lambda_1+\lambda_2)^k}{k!}.
\end{aligned}$$

Par récurrence, on obtient $X_1 + \cdots + X_n \sim \mathcal{P}(\lambda_1 + \cdots + \lambda_n)$, pour tout $n \geqslant 2$. □

Exercice 3.4.4 Soient $X_1, \dots, X_n$ des v.a. indépendantes de loi géométrique de paramètres respectifs $p_1, \dots, p_n$.

1) Montrer que $Y = \min\{X_1, \dots, X_n\}$ suit une loi géométrique.
2) Déterminer la loi de $Z = \max\{X_1, \dots, X_n\}$.

Démonstration. $(X_1, \dots, X_n)$ est un n-uplet de v.a.d. Y et Z sont des transformations de ce n-uplet :

$$Y : \Omega \to (\mathbb{N}^*)^n \to \mathbb{N}^*, \quad \omega \mapsto (X_1(\omega), \dots, X_n(\omega)) \mapsto \min(X_1(\omega), \dots, X_n(\omega)),$$
$$Z : \Omega \to (\mathbb{N}^*)^n \to \mathbb{N}^*, \quad \omega \mapsto (X_1(\omega), \dots, X_n(\omega)) \mapsto \max(X_1(\omega), \dots, X_n(\omega)).$$

Donc Y et Z sont également des v.a.d..

1) Pour tout $k \geqslant 1$,
$$\{Y \geqslant k\} = \{X_1 \geqslant k\} \cap \dots \cap \{X_n \geqslant k\},$$
$$\mathbb{P}(Y \geqslant k) = \prod_{i=1}^{n} \mathbb{P}(X_i \geqslant k) = \prod_{i=1}^{n} (1 - p_i)^{k-1}.$$
Donc
$$\forall k \geqslant 1,\ \mathbb{P}(Y = k) = \mathbb{P}(Y \geqslant k) - \mathbb{P}(Y \geqslant k+1) = \left(1 - \prod_{i=1}^{n}(1-p_i)\right)\left(\prod_{i=1}^{n}(1-p_i)\right)^{k-1}.$$
C'est-à-dire $Y \sim \mathcal{G}\left(1 - \prod_{i=1}^{n}(1-p_i)\right)$.
2) Pour tout $k \geqslant 1$,
$$\{Z \leqslant k\} = \{X_1 \leqslant k\} \cap \dots \cap \{X_n \leqslant k\},$$
$$\mathbb{P}(Z \leqslant k) = \prod_{i=1}^{n} \mathbb{P}(X_i \leqslant k) = \prod_{i=1}^{n} (1 - (1-p_i)^k).$$
Donc
$$\forall k \geqslant 1,\ \mathbb{P}(Z = k) = \mathbb{P}(Z \leqslant k) - \mathbb{P}(Z \leqslant k-1) = \prod_{i=1}^{n}(1-(1-p_i)^k) - \prod_{i=1}^{n}(1-(1-p_i)^{k-1}).$$

□

Théorème 3.4.2 (Admis) $\forall n \in \mathbb{N}$, soit $\mathbb{P}_n$ une probabilité (discrète) sur $(E_n, \mathcal{P}(E_n))$. Il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et une suite de v.a. (discrète) $(X_n)_{n\in\mathbb{N}}$ indépandentes sur $(\Omega, \mathcal{F}, \mathbb{P})$ telle que pour tout $n \in \mathbb{N}$, la loi de X_n est $\mathbb{P}_n$.

Remarque Une probabilité discrète sur $(E_n, \mathcal{P}(E_n))$ est une fonction $\mathbb{P}_n : \mathcal{P}(E_n) \to [0,1]$ qui vérifie $\exists\, F_n \subset E_n$ au plus dénombrable t.q. $\mathbb{P}_n(F_n) = 1$.

Remarque Pour chaque n, on peut construire $(E_n, \mathcal{P}(E_n), \mathbb{P}_n)$ avec E_n un espace « discret » (au plus dénombré) et $X_n = id_{E_n}$. Si on prend $\Omega = \prod_{n\in\mathbb{N}} E_n$, on note $\pi_n : \Omega \to E_n$ projection conorique.

Comment choisis F ? On peut prendre $F = \sigma\left(\bigcup_{n\in\mathbb{N}} \pi_n^{-1}(\mathcal{P}(E_n))\right)$. (la plus petite tribu contenant $\{\pi_n^{-1}(\mathcal{P}(E_n)) : n \in \mathbb{N}\}$)

Vrac déja est de construire une probabilité $\mathbb{P}$ (mesure produit) à partir de $\{\mathbb{P}_n : n \in \mathbb{N}\}$, $\pi_n \circ \mathbb{P} = \mathbb{P}_n$.

$$\forall A \subset E_n,\ \mathbb{P}(\pi_n^{-1}(A)) = \mathbb{P}(E_n \times \cdots \times E_{n-1} \times A \times E_{n+1} \times \dots) = \mathbb{P}_n(A).$$

Exercice 3.4.5 Soit $(X_n)_{n\geqslant 1}$ une suite de v.a.d. indépendantes sur $(\Omega, \mathcal{F}, \mathbb{P})$ suivant toutes la loi de Bernoulli de paramètre $p \in]0,1[$. Pour $k \in \mathbb{N}^*$, on note $Y_k =$ « le temps d'attente du k-ième 1 ». On pose $Z_1 = Y_1$ et $Z_k = Y_k - Y_{k-1}$ pour $k \geqslant 2$. Montrer que $(Z_n)_{n\geqslant 1}$ est une suite de v.a.d. indépendantes de même loi.

Démonstration. Soient $k_j \in \mathbb{N}^*$ pour $1 \leqslant j \leqslant n$. On a

$$\begin{aligned}\{Z_1 = k_1, \dots, Z_n = k_n\} &= \{Y_1 = k_1, Y_2 = k_1 + k_2, \dots, Y_n = k_1 + \cdots + k_n\} \\ &= \left(\bigcap_{i=1}^{n} \{X_{l_i} = 1\}\right) \cap \left(\bigcap_{j\in[\![1,l_n]\!]\setminus\{l_1,\dots,l_n\}} \{X_j = 0\}\right). \quad (l_i := \sum_{r=1}^{i} k_r)\end{aligned}$$

Alors

$$\begin{aligned}&\mathbb{P}(Z_1 = k_1, \dots, Z_n = k_n) \\ &= \prod_{i=1}^{n} \mathbb{P}(X_{l_i} = 1) \times \prod_{j\in[\![1,l_n]\!]\setminus\{l_1,\dots,l_n\}} \mathbb{P}(X_j = 0) \quad \text{(par indépendance)} \\ &= p^n(1-p)^{l_n - n} \\ &= \prod_{i=1}^{n} p(1-p)^{k_i - 1}.\end{aligned}$$

De plus,

$$\begin{aligned}&\mathbb{P}(Z_2 = k_2, \dots, Z_n = k_n) \\ &= \sum_{k_1=1}^{\infty} \mathbb{P}(Z_1 = k_1, \dots, Z_n = k_n) \\ &= \sum_{k_1=1}^{\infty} \left(\prod_{i=1}^{n} p(1-p)^{k_i-1}\right) \\ &= \left(\prod_{i=2}^{n} p(1-p)^{k_i-1}\right) \sum_{k_1=1}^{\infty} p(1-p)^{k_1-1} \\ &= \prod_{i=2}^{n} p(1-p)^{k_i-1}.\end{aligned}$$

On peut continuer ainsi pour obtenir la loi de Z_1 :

$$\forall k_1 \in \mathbb{N}^*, \quad \mathbb{P}(Z_1 = k_1) = p(1-p)^{k_1-1},$$

c'est-à-dire que $Z_1 \sim \mathcal{G}(p)$. Idem pour $2 \leqslant i \leqslant n$.

Donc

$$\mathbb{P}(Z_1 = k_1, \ldots, Z_n = k_n) = \mathbb{P}(Z_1 = k_1) \ldots \mathbb{P}(Z_n = k_n) \quad \text{pour tout } n.$$

Par conséquent, $(Z_n)_{n\in\mathbb{N}}$ est une suite de v.a.d. indépendantes de même loi $\mathcal{G}(p)$. □

3.5 Espérance, variance, covariance

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.

Définition 3.5.1 (Espérance) Soit $X : \Omega \to \mathbb{R}_+ = \{x \in \mathbb{R} : x \geqslant 0\}$ une v.a.d. positive. **L'espérance de** X est définie par

$$\mathbb{E}[X] = \sum_{x\in X(\Omega)\subset\mathbb{R}_+} x\mathbb{P}(X = x) \in [0, +\infty].$$

Définition 3.5.2 (Espérance finie) Soit $X : \Omega \to \mathbb{R}$ une v.a.d. On dit que X est **d'espérance finie** si la famille $(x\mathbb{P}(X = x))_{x\in X(\Omega)}$ est sommable. C'est-à-dire,

$$\sum_{x\in X(\Omega)\subset\mathbb{R}} |x|\,\mathbb{P}(X = x) = \sum_{x\in X(\Omega)} |x|\,\mathbb{P}(X = x) < +\infty.$$

Si c'est le cas, on définit

$$\mathbb{E}[X] = \sum_{x\in X(\Omega)} x\mathbb{P}(X = x) \in \mathbb{R}.$$

Notation Pour $x \in \mathbb{R}$, on définit :

— $x^+ := \max\{x, 0\}$: partie positive de x,

— $x^- := \max\{-x, 0\} \geqslant 0$: partie négative de x.

On a $x^+ - x^- = x$ et $x^+ + x^- = |x|$.

Définition 3.5.3 (v.a.d. admettant une espérance) Soit $X : \Omega \to \mathbb{R}$ une v.a.d. On dit que X **admet une espérance** si

$$\sum_{x\in X(\Omega)} x\mathbb{P}(X = x)$$

est bien définie en tant que

$$\sum_{x\in X(\Omega)} x^+\mathbb{P}(X = x) - \sum_{x\in X(\Omega)} x^-\mathbb{P}(X = x),$$

c'est-à-dire si $\sum_{x\in X(\Omega)} x^+\mathbb{P}(X = x) < +\infty$ ou $\sum_{x\in X(\Omega)} x^-\mathbb{P}(X = x) < +\infty$.

Si X admet une espérance, on définit l'espérance de X par

$$\mathbb{E}(X) = \mathbb{E}[X] = \sum_{x\in X(\Omega)} x^+\mathbb{P}(X = x) - \sum_{x\in X(\Omega)} x^-\mathbb{P}(X = x) \in [-\infty, +\infty].$$

Attention! $\mathbb{E}[X]$ n'est pas toujours définie à cause du problème $(+\infty, -\infty)$. Si X ne prend qu'un nombre fini de valeurs réelles, alors X admet une espérance finie.

Remarque Pour toute v.a.d. réelle X, on considère les v.a.d. positives :

$$\begin{aligned} X^+ &= \max\{0, X\} \geqslant 0, \\ X^- &= \max\{0, -X\} \geqslant 0, \\ |X| &= X^+ + X^-. \end{aligned}$$

$\mathbb{E}[X^+]$, $\mathbb{E}[X^-]$, $\mathbb{E}[|X|]$ sont toujours bien définies.

On a

$$\begin{aligned} \mathbb{E}\left[X^+\right] &= \sum_{x \in X^+(\Omega)} x\mathbb{P}\left(X^+ = x\right) \quad \left(= \sum_{x \in \mathbb{R}} x\mathbb{P}\left(X^+ = x\right)\right) \\ &= \sum_{x \in X(\Omega)} x\mathbb{P}\left(X^+ = x\right) \\ &= \sum_{x \in X(\Omega)} x^+\mathbb{P}\left(X = x\right), \end{aligned}$$

et de même,

$$\mathbb{E}\left[X^-\right] = \sum_{x \in X(\Omega)} x^-\mathbb{P}\left(X = x\right).$$

— X admet une espérance $\Leftrightarrow$ au moins l'une des deux espérances $\mathbb{E}[X^+]$, $\mathbb{E}[X^-]$ est finie. Alors $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$.

— X admet une espérance finie

$$\begin{aligned} &\Leftrightarrow \mathbb{E}\left[X^+\right] \text{ et } \mathbb{E}\left[X^-\right] \text{ sont finies} \\ &\Leftrightarrow \mathbb{E}[|X|] = \sum_{x \in |X|(\Omega)} x\mathbb{P}\left(|X| = x\right) \\ &\qquad = \sum_{x \in X(\Omega)} |x|\,\mathbb{P}\left(X = x\right) \\ &\qquad = \sum_{x \in X(\Omega)} x^+\mathbb{P}\left(X = x\right) + \sum_{x \in X(\Omega)} x^-\mathbb{P}\left(X = x\right) \\ &\qquad = \mathbb{E}\left[X^+\right] + \mathbb{E}\left[X^-\right] < +\infty \\ &\Leftrightarrow \sum_{x \in X(\Omega)} x\mathbb{P}\left(X = x\right) \text{ est absolument convergente,} \\ &\quad \text{c-à-d} \sum_{x \in X(\Omega)} |x|\,\mathbb{P}\left(X = x\right) < +\infty. \end{aligned}$$

Proposition 3.5.1 Soit X une v.a.d. admettant une espérance. Alors on a

$$|\mathbb{E}[X]| \leqslant \mathbb{E}[|X|].$$

Démonstration. Si X admet une espérance finie, alors

$$|\mathbb{E}[X]| = \left| \sum_{x \in X(\Omega)} x\mathbb{P}(X = x) \right| \leqslant \sum_{x \in X(\Omega)} |x| \, \mathbb{P}(X = x) = \mathbb{E}[|X|].$$

Sinon,

$$|\mathbb{E}[X]| = \left|\mathbb{E}\left[X^+\right] - \mathbb{E}\left[X^-\right]\right| \leqslant \left|\mathbb{E}\left[X^+\right]\right| + \left|\mathbb{E}\left[X^-\right]\right| = \mathbb{E}\left[X^+\right] + \mathbb{E}\left[X^-\right] = \mathbb{E}[|X|].$$

□

Exemple 3.5.1

1) $X : \Omega \to \mathbb{N}^*$, $\mathbb{P}(X = n) = \dfrac{1}{n(n+1)}$.

$$\mathbb{E}[X] = \sum_{n \geqslant 1} n\mathbb{P}(X = n) = +\infty.$$

2) $Y : \Omega \to \mathbb{Z}^*$, $\mathbb{P}(Y = n) = \dfrac{1}{2|n|(|n|+1)}$ pour $n \neq 0$. Alors

$$\mathbb{E}[|Y|] = \sum_{n \in \mathbb{Z}^*} |n|\mathbb{P}(Y = n) = \sum_{n \geqslant 1} \frac{1}{n+1} = +\infty.$$

3) Si $X \sim \mathcal{U}([1, n])$,

$$\mathbb{E}[X] = \sum_{k=1}^{n} k \cdot \mathbb{P}(X = k) = \frac{n+1}{2}.$$

4) Si $X \sim \mathcal{B}(p)$,

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

5) Si $X \sim \mathcal{G}(p)$,

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=1}^{\infty} k \cdot \mathbb{P}(X = k) \\
&= \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} p \\
&= p \frac{\partial}{\partial q} \sum_{k=1}^{\infty} q^k \quad (q = 1-p) \\
&= p \cdot \frac{\partial}{\partial q} \frac{q}{1-q} \\
&= p \cdot \frac{1}{(1-q)^2} = \frac{1}{p}.
\end{aligned}$$

6) Si $X \sim \mathcal{P}(\lambda)$,

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

Exercice 3.5.1 Si $X \sim B(n,p)$, calculer $\mathbb{E}[X]$.

Théorème 3.5.1 (Formule de transfert) Soit X une v.a.d. sur $(\Omega, \mathcal{F}, \mathbb{P})$ et $f : X(\Omega) \to \mathbb{R}$. Alors la v.a.d. $f(X)$ admet une espérance si et seulement si $\sum_{x \in X(\Omega)} f(x)\mathbb{P}(X=x)$ est bien définie. Dans ce cas, on a

$$\mathbb{E}[f(X)] = \sum_{x \in X(\Omega)} f(x)\mathbb{P}(X=x).$$

Remarque Cette formule s'applique toujours à $f(X)^+$, $f(X)^-$, $|f(X)|$. $f(X)$ admet une espérance finie $\Leftrightarrow \sum_{x \in X(\Omega)} |f(x)|\,\mathbb{P}(X=x) < +\infty$, c'est-à-dire que la famille $\{f(x)\mathbb{P}(X=x)\}_{x \in X(\Omega)}$ est sommable.

Démonstration. On note $Y = f(X) : \Omega \to f(X(\Omega))$.

Supposons d'abord que f est à valeurs dans $\mathbb{R}_+$. Dans ce cas, Y admet une espérance car elle est positive. Pour tout $y \in Y(\Omega)$,

$$\mathbb{P}(Y=y) = \mathbb{P}\left(X \in f^{-1}(y)\right) = \sum_{x \in X(\Omega), f(x)=y} \mathbb{P}(X=x).$$

Par définition,

$$\mathbb{E}[Y] = \sum_{y \in Y(\Omega)} y\mathbb{P}(Y=y) = \sum_{y \in Y(\Omega)} \sum_{x \in X(\Omega), f(x)=y} f(x)\mathbb{P}(X=x).$$

En remarquant que

$$X(\Omega) = \bigsqcup_{y \in Y(\Omega)} \{x \in X(\Omega) : f(x) = y\},$$

on applique la sommation par paquets et on obtient donc

$$\mathbb{E}[Y] = \sum_{x \in X(\Omega)} f(x)\mathbb{P}(X=x).$$

Enlevons maintenant l'hypothèse que f est positive. On applique le résultat qu'on vient de démontrer aux fonctions positives f^+ et f^- :

$$\mathbb{E}\left[f(X)^+\right] = \mathbb{E}\left[f^+(X)\right] = \sum_{x\in X(\Omega)} f^+(x)\mathbb{P}\left(X=x\right),$$
$$\mathbb{E}\left[f(X)^-\right] = \mathbb{E}\left[f^-(X)\right] = \sum_{x\in X(\Omega)} f^-(x)\mathbb{P}\left(X=x\right).$$

La famille de réels $\{f(x)\mathbb{P}\left(X=x\right)\}_{x\in X(\Omega)}$ admet une somme si et seulement si une des deux sommes $\sum f^+(x)\mathbb{P}\left(X=x\right)$ et $\sum f^-(x)\mathbb{P}\left(X=x\right)$ est finie, auquel cas

$$\sum f(x)\mathbb{P}\left(X=x\right) = \sum f^+(x)\mathbb{P}\left(X=x\right) - \sum f^-(x)\mathbb{P}\left(X=x\right).$$

Donc $\sum_{x\in X(\Omega)} f(x)\mathbb{P}\left(X=x\right)$ est bien définie $\Leftrightarrow$ au moins une des deux espérances $\mathbb{E}\left[f(X)^+\right]$ et $\mathbb{E}\left[f(X)^-\right]$ est finie, auquel cas

$$\sum_{x\in X(\Omega)} f(x)\mathbb{P}\left(X=x\right) = \mathbb{E}\left[f(X)^+\right] - \mathbb{E}\left[f(X)^-\right] = \mathbb{E}\left[f(X)\right].$$

□

Proposition 3.5.2 Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ une v.a.d. avec Ω au plus dénombrable. Alors X admet une espérance si et seulement si la somme

$$\sum_{\omega\in\Omega} X(\omega)\mathbb{P}\left(\{\omega\}\right)$$

est définie. Si c'est le cas,

$$\mathbb{E}\left[X\right] = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}\left(\{\omega\}\right).$$

Démonstration. Considérons la v.a.d. $Y : \Omega \to \Omega, \omega \mapsto \omega$ identité, et $X : \Omega \to \mathbb{R}, \omega \mapsto X(\omega)$ vue comme une fonction. On a $X(Y)(\omega) = X(Y(\omega)) = X(\omega)$ pour tout $\omega \in \Omega$, donc $\mathbb{E}\left[X(Y)\right] = \mathbb{E}\left[X\right]$. Or,

$$\mathbb{E}\left[X(Y)\right] = \sum_{y\in Y(\Omega)} X(y)\mathbb{P}\left(Y=y\right) = \sum_{y\in Y(\Omega)} X(y)\mathbb{P}\left(\{y\}\right).$$

Donc,

$$\mathbb{E}\left[X\right] = \sum_{y\in Y(\Omega)} X(y)\mathbb{P}\left(\{y\}\right) = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}\left(\{\omega\}\right).$$

□

Théorème 3.5.2 (Linéarité de l'espérance) Si X, Y sont des v.a.d. réelles d'espérance finie sur $(\Omega, \mathcal{F}, \mathbb{P})$ et $\lambda \in \mathbb{R}$, alors $\lambda X + Y$ est d'espérance finie et

$$\mathbb{E}\left[\lambda X + Y\right] = \lambda\mathbb{E}\left[X\right] + \mathbb{E}\left[Y\right].$$

C'est-à-dire, l'ensemble des v.a.d. réelles sur $(\Omega, \mathcal{F}, \mathbb{P})$ d'espérance finie est un $\mathbb{R}$-espace vectoriel, noté $L^1(\Omega, \mathcal{F}, \mathbb{P})$, et l'espérance est une forme linéaire sur $L^1(\Omega, \mathcal{F}, \mathbb{P})$.

Démonstration. $\lambda X + Y$ est d'espérance finie si et seulement si $\mathbb{E}\left[|\lambda X + Y|\right] < +\infty$.

$$\begin{aligned}
\mathbb{E}\left[|\lambda X + Y|\right] &= \sum_{z \in |\lambda X+Y|(\Omega)} z \sum_{\substack{x \in X(\Omega) \\ y \in Y(\Omega) \\ |\lambda x + y| = z}} \mathbb{P}\left(X = x, Y = y\right) \\
&= \sum_{x \in X(\Omega), y \in Y(\Omega)} |\lambda x + y| \, \mathbb{P}\left(X = x, Y = y\right) \\
&\leqslant \sum_{x \in X(\Omega), y \in Y(\Omega)} (|\lambda| \, |x| + |y|) \mathbb{P}\left(X = x, Y = y\right) \\
&= \sum_{x \in X(\Omega)} |\lambda| \, |x| \, \mathbb{P}\left(X = x\right) + \sum_{y \in Y(\Omega)} |y| \, \mathbb{P}\left(Y = y\right) \\
&= |\lambda| \mathbb{E}\left[|X|\right] + \mathbb{E}\left[|Y|\right] < +\infty.
\end{aligned}$$

par hypothèse.

De même manière, on obtient

$$\begin{aligned}
\mathbb{E}\left[\lambda X + Y\right] &= \sum_{x \in X(\Omega), y \in Y(\Omega)} (\lambda x + y) \mathbb{P}\left(X = x, Y = y\right) \\
&= \lambda \sum_{x \in X(\Omega)} x \mathbb{P}\left(X = x\right) + \sum_{y \in Y(\Omega)} y \mathbb{P}\left(Y = y\right) \\
&= \lambda \mathbb{E}\left[X\right] + \mathbb{E}\left[Y\right].
\end{aligned}$$

□

Remarque Si $X, Y \geqslant 0$ et $a, b \in \mathbb{R}^+$, alors $aX + bY$ admet une espérance et

$$\mathbb{E}\left[aX + bY\right] = a\mathbb{E}\left[X\right] + b\mathbb{E}\left[Y\right].$$

Exemple 3.5.2

1) $X \sim \mathcal{B}\text{in}(n, p)$. On peut écrire $X = X_1 + \cdots + X_n$ où $X_i \sim \mathcal{B}\text{ern}(p)$ pour $1 \leqslant i \leqslant n$ sont indépendantes. Alors
$$\mathbb{E}\left[X\right] = \mathbb{E}\left[X_1\right] + \cdots + \mathbb{E}\left[X_n\right] = np.$$

2) Soient $A_1, \ldots, A_n$ des événements (pas nécessairement indépendants). Soit $X = \mathbb{1}_{A_1} + \cdots + \mathbb{1}_{A_n}$ = nombre de A_i vérifiés. La loi de X peut être très compliquée, mais
$$\mathbb{E}\left[X\right] = \mathbb{E}\left[\mathbb{1}_{A_1}\right] + \cdots + \mathbb{E}\left[\mathbb{1}_{A_n}\right] = \mathbb{P}\left(A_1\right) + \cdots + \mathbb{P}\left(A_n\right).$$

Proposition 3.5.3 (Monotonie de l'espérance) Soient $X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ deux v.a.d. réelles. Si $\forall \omega \in \Omega, X(\omega) \leqslant Y(\omega)$, alors
$$\mathbb{E}\left[X\right] \leqslant \mathbb{E}\left[Y\right].$$

Démonstration. $Y - X \geqslant 0$ est une v.a. positive. Par linéarité,

$$\mathbb{E}\left[Y - X\right] + \mathbb{E}\left[X\right] = \mathbb{E}\left[Y\right].$$

Comme $\mathbb{E}\left[Y - X\right] \geqslant 0$, on a $\mathbb{E}\left[X\right] \leqslant \mathbb{E}\left[Y\right]$. □

Exemple 3.5.3 Si X est bornée, c'est-à-dire il existe $M < +\infty$ tel que $|X(\omega)| \leqslant M$ pour tout $\omega \in \Omega$, alors

$$|\mathbb{E}[X]| \leqslant \mathbb{E}[|X|] \leqslant M.$$

En particulier, X admet une espérance finie.

On définit une relation d'équivalence sur $L^1(\Omega, \mathcal{F}, \mathbb{P})$ par $X \sim Y$ si $X - Y = 0$ presque sûrement.

Théorème 3.5.3 Sur $L^1(\Omega, \mathcal{F}, \mathbb{P})/\sim$, $\mathbb{E}[|\cdot|]$ est une norme.

Démonstration.

— **Séparation** :

$$\mathbb{E}[|X|] = 0 \Leftrightarrow \sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) = 0 \Leftrightarrow \text{si } x \neq 0 \text{ alors } \mathbb{P}(X = x) = 0 \Leftrightarrow X = 0 \text{ p.s.}$$

— **Homogénéité** : $\forall \lambda \in \mathbb{R}$,

$$\mathbb{E}[|\lambda X|] = |\lambda| \mathbb{E}[|X|].$$

— **Inégalité triangulaire** :

$$|X + Y| \leqslant |X| + |Y| \Rightarrow \mathbb{E}[|X + Y|] \leqslant \mathbb{E}[|X|] + \mathbb{E}[|Y|].$$

De plus, si $X \sim Y$, alors $\mathbb{P}(X = x) = \mathbb{P}(Y = x)$ pour tout $x \in X(\Omega) \cup Y(\Omega)$, donc $\mathbb{E}[X] = \mathbb{E}[Y]$ et $\mathbb{E}[|X|] = \mathbb{E}[|Y|]$. Ainsi, $\mathbb{E}[|\cdot|]$ est bien définie sur $L^1(\Omega, \mathcal{F}, \mathbb{P})/\sim$ et c'est une norme. □

Remarque On a vu que si $X : \Omega \to \mathbb{R}^+$ est une v.a. positive, alors

$$\mathbb{E}[X] = 0 \Rightarrow X \text{ est presque sûrement égale à } 0.$$

Théorème 3.5.4 Si $X, Y : \Omega \to \mathbb{R}$ sont des v.a.d. **indépendantes** et $X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, alors $XY \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ et

$$\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y].$$

Démonstration.

$$\begin{aligned}
\mathbb{E}[|XY|] &= \sum_{x \in X(\Omega), y \in Y(\Omega)} |xy| \mathbb{P}(X = x, Y = y) \\
&= \sum_{x \in X(\Omega), y \in Y(\Omega)} |x| \mathbb{P}(X = x) \cdot |y| \mathbb{P}(Y = y) \\
&= \left(\sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) \right) \left(\sum_{y \in Y(\Omega)} |y| \mathbb{P}(Y = y) \right) \\
&= \mathbb{E}[|X|] \mathbb{E}[|Y|] < +\infty.
\end{aligned}$$

Donc $XY \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.

On utilise le même argument (sans valeur absolue) pour avoir

$$\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y].$$

□

Remarque Si $X, Y \geqslant 0$ et indépendantes, alors $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$ même sans l'hypothèse que $\mathbb{E}[X]$ et $\mathbb{E}[Y]$ soient finies.

Conventions : $0 \times (+\infty) = 0$, $(+\infty) \times (+\infty) = +\infty$.

Définition 3.5.4 (v.a.d. centrée) Toute v.a.d. réelle d'espérance nulle est dite **centrée**.

Proposition 3.5.4 Si X est une v.a.d. réelle d'espérance finie, alors $X - \mathbb{E}[X]$ est centrée.

Démonstration.

$$\mathbb{E}[X - \mathbb{E}[X]] = \mathbb{E}[X] - \mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X] - \mathbb{E}[X] = 0.$$

□

Exercice 3.5.2 Soit $n \in \mathbb{N}^*$, $(X_k)_{k\in\mathbb{N}^*}$ une suite de v.a. indépendantes sur $(\Omega, \mathcal{F}, \mathbb{P})$, toutes de loi uniforme sur $[\![1, n]\!]$. On cherche à calculer le nombre moyen de tirages nécessaires pour obtenir tous les numéros de 1 à n.

1) Pour $1 \leqslant k \leqslant n$, on note $T_k =$ « le nombre de tirages nécessaires pour obtenir k numéros distincts » et $Z_1 = T_1 = 1$, $Z_k = T_k - T_{k-1}$ pour $2 \leqslant k \leqslant n$. Montrer que Z_k pour $2 \leqslant k \leqslant n$ suivent toutes une loi géométrique.
2) Calculer $\mathbb{E}[T_n]$ et trouver un équivalent de $\mathbb{E}[T_n]$ lorsque $n \to \infty$.

Solution.

1) Soit $k \in [\![1, n]\!]$. Si $\mathbb{P}(T_{k-1} = i) > 0$, alors

$$\mathbb{P}(Z_k = j \mid T_{k-1} = i) = \mathbb{P}(T_k = i + j \mid T_{k-1} = i) = \left(\frac{k-1}{n}\right)^{j-1} \times \left(1 - \frac{k-1}{n}\right).$$

Donc

$$\begin{aligned}
\mathbb{P}(Z_k = j) &= \sum_{i=1}^{\infty} \mathbb{P}(Z_k = j \mid T_{k-1} = i)\,\mathbb{P}(T_{k-1} = i) \\
&= \sum_{i=1}^{\infty} \left(\frac{k-1}{n}\right)^{j-1} \left(1 - \frac{k-1}{n}\right) \mathbb{P}(T_{k-1} = i) \\
&= \left(\frac{k-1}{n}\right)^{j-1} \left(1 - \frac{k-1}{n}\right) \sum_{i=1}^{\infty} \mathbb{P}(T_{k-1} = i) \\
&= \left(\frac{k-1}{n}\right)^{j-1} \left(1 - \frac{k-1}{n}\right).
\end{aligned}$$

C'est-à-dire, $Z_k \sim \mathcal{G}\left(\dfrac{n-k+1}{n}\right)$ pour tout $2 \leqslant k \leqslant n$.

2)

$$\mathbb{E}[Z_k] = \frac{n}{n-k+1}, \quad \forall 2 \leqslant k \leqslant n.$$

$$\mathbb{E}[T_n] = \sum_{k=1}^{n} \mathbb{E}[Z_k] = n \sum_{j=1}^{n} \frac{1}{j}.$$

Donc $\mathbb{E}[T_n] \sim n \log n$ quand $n \to \infty$.

□

Définition 3.5.5 (Moment d'ordre r) Soit X une v.a.d. réelle et $r \in \mathbb{Z}$. Si X^r est d'espérance finie (c'est-à-dire $\mathbb{E}[|X^r|] < +\infty$), on dit que X admet **un moment d'ordre** r et le moment d'ordre r est le réel $\mathbb{E}[X^r]$.

Proposition 3.5.5 Si X est une v.a.d. admettant un moment d'ordre $r \geqslant 2$, alors elle admet un moment d'ordre k pour tout $0 \leqslant k \leqslant r$. En particulier, X est d'espérance finie.

Démonstration. On a

$$|x|^k \leqslant |x|^r + 1, \quad \forall x \in \mathbb{R},\ 0 \leqslant k \leqslant r.$$

Donc $|X|^k \leqslant |X|^r + 1$. On en déduit que

$$\mathbb{E}\left[|X|^k\right] \leqslant \mathbb{E}[|X|^r] + 1 < +\infty.$$

□

Proposition 3.5.6 Les v.a.d. réelles sur $(\Omega, \mathcal{F}, \mathbb{P})$ admettant un moment d'ordre 2 forment un $\mathbb{R}$-espace vectoriel, noté $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Démonstration. Si X, Y admettent un moment d'ordre 2, c'est-à-dire $\mathbb{E}\left[|X|^2\right] < +\infty$ et $\mathbb{E}\left[|Y|^2\right] < +\infty$, alors pour tout $\lambda \in \mathbb{R}$, on a

$$\mathbb{E}\left[|\lambda X + Y|^2\right] \leqslant 2|\lambda|^2 \mathbb{E}\left[|X|^2\right] + 2\mathbb{E}\left[|Y|^2\right] < +\infty,$$

c'est-à-dire $\lambda X + Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. □

Théorème 3.5.5 (Inégalité de Cauchy-Schwarz) Si $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, alors $XY \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ et

$$\mathbb{E}[XY]^2 \leqslant \mathbb{E}\left[X^2\right] \mathbb{E}\left[Y^2\right].$$

Remarque On a aussi

$$\mathbb{E}[|XY|] \leqslant \mathbb{E}\left[X^2\right]^{\frac{1}{2}} \mathbb{E}\left[Y^2\right]^{\frac{1}{2}}.$$

Lorsque $\mathbb{E}[Y^2] > 0$ (c'est-à-dire Y n'est pas p.s. égale à 0),

$$|\mathbb{E}[XY]| = \mathbb{E}\left[X^2\right]^{\frac{1}{2}} \mathbb{E}\left[Y^2\right]^{\frac{1}{2}} \Leftrightarrow \exists \lambda \in \mathbb{R},\ X = \lambda Y \text{ p.s.}$$

Démonstration. On a $|XY| \leqslant \frac{1}{2}\left(|X|^2 + |Y|^2\right)$, donc

$$\mathbb{E}[|XY|] \leqslant \frac{1}{2}\mathbb{E}\left[|X|^2\right] + \frac{1}{2}\mathbb{E}\left[|Y|^2\right] < +\infty,$$

c'est-à-dire $XY \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.

Pour tout $\lambda \in \mathbb{R}$,

$$0 \leqslant \mathbb{E}\left[(\lambda X + Y)^2\right] = \lambda^2 \mathbb{E}\left[X^2\right] + 2\lambda \mathbb{E}[XY] + \mathbb{E}\left[Y^2\right].$$

Ce polynôme en λ de degré 2 est toujours positif, donc son discriminant est négatif :

$$\mathbb{E}[XY]^2 - \mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right] \leqslant 0.$$

Si $\mathbb{E}[XY]^2 = \mathbb{E}[X^2]\,\mathbb{E}[Y^2]$, alors soit $\mathbb{E}[Y^2] = 0$ (auquel cas $Y^2 = 0$ p.s. et donc $Y = 0$ p.s.), soit $\mathbb{E}[Y^2] \neq 0$. Dans ce dernier cas, on prend $\lambda_0 = -\dfrac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}$ et on observe que

$$\mathbb{E}\left[(\lambda_0 X + Y)^2\right] = 0,$$

donc $\lambda_0 X + Y = 0$ p.s., c'est-à-dire $Y = -\lambda_0 X$ p.s.. □

Pour $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, on définit une relation d'équivalence $X \sim Y \Leftrightarrow X = Y$ p.s. On définit également

$$\|X\|_2 = \left(\mathbb{E}\left[X^2\right]\right)^{\frac{1}{2}}.$$

Théorème 3.5.6 $(L^2(\Omega, \mathcal{F}, \mathbb{P})/\sim, \|\cdot\|_2)$ est un espace vectoriel normé.

Démonstration. Si $X \sim Y$, alors $\mathbb{P}(X = x) = \mathbb{P}(Y = x)$ pour tout $x \in X(\Omega) \cup Y(\Omega)$, donc

$$\mathbb{E}\left[X^2\right] = \sum_{x \in X(\Omega)\cup Y(\Omega)} x^2 \mathbb{P}(X = x) = \sum_{x \in X(\Omega)\cup Y(\Omega)} x^2 \mathbb{P}(Y = x) = \mathbb{E}\left[Y^2\right].$$

Ainsi, $\|\cdot\|_2$ est bien définie sur $L^2(\Omega, \mathcal{F}, \mathbb{P})/\sim$.

— **Séparation** :

$$\|X\|_2 = 0 \Leftrightarrow \mathbb{E}\left[X^2\right] = 0 \Leftrightarrow X^2 = 0 \text{ p.s.} \Leftrightarrow X = 0 \text{ p.s.}$$

— **Homogénéité** : Pour tout $\lambda \in \mathbb{R}$,

$$\|\lambda X\|_2 = \left(\mathbb{E}\left[(\lambda X)^2\right]\right)^{\frac{1}{2}} = \left(\lambda^2 \mathbb{E}\left[X^2\right]\right)^{\frac{1}{2}} = |\lambda| \cdot \|X\|_2.$$

— **Inégalité triangulaire** : Soit $Z = X + Y$. Alors

$$\|Z\|_2 \leqslant \|X\|_2 + \|Y\|_2 \Leftrightarrow \mathbb{E}\left[Z^2\right] \leqslant \left(\mathbb{E}\left[X^2\right]^{\frac{1}{2}} + \mathbb{E}\left[Y^2\right]^{\frac{1}{2}}\right)^2 = \mathbb{E}\left[X^2\right] + \mathbb{E}\left[Y^2\right] + 2\sqrt{\mathbb{E}[X^2]\,\mathbb{E}[Y^2]}.$$

Mais par l'inégalité de Cauchy-Schwarz,

$$\mathbb{E}[XY] \leqslant \sqrt{\mathbb{E}[X^2]\,\mathbb{E}[Y^2]},$$

donc

$$\mathbb{E}\left[Z^2\right] = \mathbb{E}\left[X^2\right] + \mathbb{E}\left[Y^2\right] + 2\mathbb{E}[XY] \leqslant \mathbb{E}\left[X^2\right] + \mathbb{E}\left[Y^2\right] + 2\sqrt{\mathbb{E}[X^2]\,\mathbb{E}[Y^2]}.$$

Ainsi, $\|\cdot\|_2$ est une norme.

□

Remarque Si $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, alors $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, donc $\mathbb{E}[X]$ est bien définie dans $\mathbb{R}$. En tant que constante, $\mathbb{E}[X] \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Comme $L^2(\Omega, \mathcal{F}, \mathbb{P})$ est un espace vectoriel, on a $X - \mathbb{E}[X] \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Définition 3.5.6 (Variance, écart-type) Si $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, on appelle **la variance de** X

$$\mathbb{V}(X) = \mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right],$$

et **l'écart-type de** X

$$\sigma(X) = \sqrt{\mathbb{V}(X)}.$$

Ceci mesure la distance de X à $\mathbb{E}[X]$ dans $(L^2(\Omega, \mathcal{F}, \mathbb{P}), \|\cdot\|_2)$.

Théorème 3.5.7 (Formule de Kœnig-Huygens) Si $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, on a

$$\mathbb{V}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2.$$

Démonstration.

$$\begin{aligned}\mathbb{E}\left[(X - \mathbb{E}[X])^2\right] &= \mathbb{E}\left[X^2 - 2\mathbb{E}[X]X + \mathbb{E}[X]^2\right] \\ &= \mathbb{E}\left[X^2\right] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 \\ &= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2.\end{aligned}$$

□

Exemple 3.5.4

1) Si $X \sim \mathcal{U}(\llbracket 1, n \rrbracket)$,

$$\mathbb{E}\left[X^2\right] = \sum_{k=1}^{n} k^2 \cdot \frac{1}{n} = \frac{(n+1)(2n+1)}{6}, \quad \mathbb{V}(X) = \frac{n^2-1}{12}.$$

2) Si $X \sim \mathcal{B}(p)$,

$$\mathbb{E}\left[X^2\right] = p, \quad \mathbb{V}(X) = p(1-p).$$

3) Si $X \sim \mathcal{G}(p)$,

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

$$\mathbb{E}[X(X-1)] = \sum_{k=1}^{\infty} k(k-1)(1-p)^{k-1}p = \frac{2(1-p)}{p^2},$$

$$\mathbb{E}\left[X^2\right] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \frac{2-p}{p^2},$$

$$\mathbb{V}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \frac{1-p}{p^2}.$$

4) Si $X \sim \mathcal{P}(\lambda)$,

$$\mathbb{E}[X] = \lambda,$$

$$\mathbb{E}\left[X^2\right] = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!}e^{-\lambda} = \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!}e^{-\lambda} + \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \lambda^2 + \lambda,$$

$$\mathbb{V}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \lambda.$$

Proposition 3.5.7 Si $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, alors pour tout $a, b \in \mathbb{R}$, on a

$$\mathbb{V}(aX + b) = a^2 \mathbb{V}(X).$$

Démonstration.

$$\begin{aligned}
\mathbb{V}(aX + b) &= \mathbb{E}\left[(aX + b - \mathbb{E}[aX + b])^2\right] \\
&= \mathbb{E}\left[(aX + b - a\mathbb{E}[X] - b)^2\right] \\
&= \mathbb{E}\left[a^2(X - \mathbb{E}[X])^2\right] \\
&= a^2 \mathbb{V}(X).
\end{aligned}$$

□

Proposition 3.5.8 Si $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ et $\mathbb{V}(X) \neq 0$, alors

$$X^* = \frac{X - \mathbb{E}[X]}{\sigma(X)}$$

est une v.a.d. réelle centrée réduite (de variance 1) appelée **la v.a.d. centrée réduite associée à** X.

Démonstration.

$$\mathbb{E}[X^*] = \mathbb{E}\left[\frac{X - \mathbb{E}[X]}{\sigma(X)}\right] = 0.$$

$$\mathbb{V}(X^*) = \mathbb{E}\left[(X^*)^2\right] = \mathbb{E}\left[\frac{(X - \mathbb{E}[X])^2}{\sigma(X)^2}\right] = \frac{1}{\mathbb{V}(X)} \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = 1. \qquad \square$$

Définition 3.5.7 (Covariance) Pour X, $Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, **la covariance de X et Y** (ou du couple (X, Y)) est définie par

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

Si $\mathrm{Cov}(X, Y) = 0$, on dit que X et Y sont **non-corrélées**.

Proposition 3.5.9 $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$. En particulier, si $X = Y$, $\mathrm{Cov}(X, Y) = \mathbb{V}(X) \geqslant 0$.

Démonstration.

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\
&= \mathbb{E}[XY - \mathbb{E}[X]\,Y - X\mathbb{E}[Y] + \mathbb{E}[X]\,\mathbb{E}[Y]] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] - \mathbb{E}[X]\,\mathbb{E}[Y] + \mathbb{E}[X]\,\mathbb{E}[Y] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y].
\end{aligned}$$

□

Corollaire Si X, $Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ indépendantes, alors $\mathrm{Cov}(X, Y) = 0$. C'est-à-dire, $X \perp\!\!\!\perp Y \Rightarrow \mathrm{Cov}(X, Y) = 0$.

Remarque $\mathrm{Cov}(X, Y) = 0 \not\Rightarrow X \perp\!\!\!\perp Y$: voir l'exemple suivant.

Exemple 3.5.5 Soit $X \sim \mathcal{U}(\{-1,0,1\})$, $Y = \mathbb{1}_{\{X=0\}}$.

— X, Y ne sont pas indépendantes :

$$\mathbb{P}(X=1,Y=1) = \mathbb{P}(X=1,X=0) = \mathbb{P}(\varnothing) = 0, \quad \mathbb{P}(X=1)\,\mathbb{P}(Y=1) = \frac{1}{9} > 0.$$

— X, Y sont non-corrélées car $XY = 0$ par définition. $\mathbb{E}[XY] = 0$. Or $\mathbb{E}[X] = \frac{1}{3}(-1+0+1) = 0$, donc $\mathrm{Cov}(X,Y) = 0$.

Proposition 3.5.10 La covariance $\mathrm{Cov}(\cdot,\cdot)$ est une forme symétrique et bilinéaire non-négative sur $L^2(\Omega,\mathcal{F},\mathbb{P})/\sim$. $\forall$ v.a. $X,Y,Z \in L^2(\Omega,\mathcal{F},\mathbb{P})$, $a,b \in \mathbb{R}$, on a

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X),$$

$$\mathrm{Cov}(X,X) \geqslant 0,$$

$$\mathrm{Cov}(aX+bY,Z) = a\ \mathrm{Cov}(X,Z) + b\ \mathrm{Cov}(Y,Z).$$

De plus, si X ou Y est presque sûrement constante, on a $\mathrm{Cov}(X,Y) = 0$.

Démonstration.

— **Symétrie** : par définition.

— **Non-négativité** : $\mathrm{Cov}(X,X) = \mathbb{V}(X) \geqslant 0$.

— **Bilinéarité** :

$$\begin{aligned}\mathrm{Cov}(aX+bY,Z) &= \mathbb{E}[(aX+bY)Z] - (a\mathbb{E}[X] + b\mathbb{E}[Y])\mathbb{E}[Z] \\ &= a\mathbb{E}[XZ] + b\mathbb{E}[YZ] - a\mathbb{E}[X]\,\mathbb{E}[Z] - b\mathbb{E}[Y]\,\mathbb{E}[Z] \\ &= a\ \mathrm{Cov}(X,Z) + b\ \mathrm{Cov}(Y,Z).\end{aligned}$$

— Si X est p.s. égale à $c \in \mathbb{R}$, alors $\mathbb{E}[X] = c$ et $X - \mathbb{E}[X]$ est p.s. égale à 0, donc $(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])$ est p.s. égal à 0, d'où $\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = \mathrm{Cov}(X,Y) = 0$.

□

Proposition 3.5.11 Si $X_1,\dots,X_n \in L^2(\Omega,\mathcal{F},\mathbb{P})$, alors

$$\mathbb{V}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{V}(X_i) + \sum_{\substack{1\leqslant i,j\leqslant n \\ i\neq j}} \mathrm{Cov}(X_i,X_j) = \sum_{i=1}^n \mathbb{V}(X_i) + 2\sum_{1\leqslant i<j\leqslant n} \mathrm{Cov}(X_i,X_j).$$

Démonstration.

$$\begin{aligned}\mathbb{V}\left(\sum_{i=1}^n X_i\right) &= \mathrm{Cov}\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \mathrm{Cov}(X_i,X_j) \\ &= \sum_{i=1}^n \mathbb{V}(X_i) + \sum_{\substack{1\leqslant i,j\leqslant n \\ i\neq j}} \mathrm{Cov}(X_i,X_j).\end{aligned}$$

□

Corollaire Si $X_1, \ldots, X_n \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ deux à deux indépendantes, alors

$$\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}(X_i),$$

car $\mathrm{Cov}(X_i, X_j) = 0$ pour $i \neq j$.

Exemple 3.5.6 Si $X = \sum_{i=1}^{n} X_i$ avec X_i indépendantes, $X_i \sim \mathcal{B}(p)$, alors

$$X \sim \mathcal{B}\mathrm{in}(n,p), \quad \mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = np, \quad \mathbb{V}(X) = \sum_{i=1}^{n} \mathbb{V}(X_i) = np(1-p).$$

Exercice 3.5.3 Soient $X_1, \ldots, X_n$ v.a.d. $\in L^2(\Omega, \mathcal{F}, \mathbb{P})$. La matrice $M = (\mathrm{Cov}(X_i, X_j))_{1 \leqslant i,j \leqslant n} \in \mathcal{M}_n(\mathbb{R})$ est appelée **la matrice de covariance**.

1) Montrer que M est diagonalisable (可対角化的) et qu'elle a des valeurs propres (特征值) positives ou nulles.
2) Si $\mathbb{E}[X_i] = 0$ pour tout $1 \leqslant i \leqslant n$, à quelle condition 0 est-il une valeur propre de M ?

Solution.

1) Comme $\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$, M est une matrice réelle symétrique, donc M est diagonalisable.

   Soit $\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{R}^n$, alors

   $$0 \leqslant \mathbb{V}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i,j=1}^{n} \mathrm{Cov}(a_i X_i, a_j X_j) = \vec{a}^T M \vec{a}.$$

   Donc M est une matrice semi-définie positive. On en déduit que ses valeurs propres sont non-négatives.

   Si λ est une valeur propre et $\vec{a}$ un vecteur propre associé ($M\vec{a} = \lambda\vec{a}$), alors $0 \leqslant \lambda \vec{a}^T \vec{a}$.

2) 0 est une valeur propre de $M \Leftrightarrow \exists\, a \in \mathbb{R}^n, a \neq 0$, tel que $Ma = 0$.
   $\mathbb{V}\left(\sum_{i=1}^{n} a_i X_i\right) = \vec{a}^T M \vec{a} = 0$ implique $\sum_{i=1}^{n} a_i X_i = 0$ p.s.. (car $\mathbb{E}[X_i] = 0$, $1 \leqslant i \leqslant n$)
   Réciproquement, si $\sum_{i=1}^{n} a_i X_i = 0$ p.s. avec $\vec{a} = (a_1, \ldots, a_n)^T \neq 0$, alors

   $$\begin{pmatrix} \mathrm{Cov}\left(X_1, \sum_{i=1}^{n} a_i X_i\right) \\ \vdots \\ \mathrm{Cov}\left(X_n, \sum_{i=1}^{n} a_i X_i\right) \end{pmatrix} = M \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = 0,$$

donc 0 est une valeur propre de M et $\vec{a}$ est un vecteur propre associé.

Conclusion : 0 est une valeur propre de M avec $\vec{a} = (a_1, \ldots, a_n)^T \neq 0$ un vecteur propre associé si et seulement si $\sum_{i=1}^{n} a_i X_i = 0$ p.s..

□

Exercice 3.5.4 Soit $(X_n)_{n\geqslant 1}$ une suite de variables de Bernoulli indépendantes de paramètre $p \in]0,1[$. On pose $q = 1-p$. On dit que la première série est de longueur n si les premières n épreuves ont donné le même résultat et la $(n+1)$-ième épreuve donne un résultat différent. De même, la deuxième série commence à l'épreuve qui suit la dernière épreuve de la première série et s'arrête avant le changement suivant. On note L_1 et L_2 respectivement la longueur de la première et de la seconde série. Montrer que $\mathrm{Cov}\,(L_1, L_2) = -\dfrac{(p-q)^2}{pq} \leqslant 0$.

Solution. Soient $m, n \in \mathbb{N}^*$.

$$\begin{aligned}\mathbb{P}\,(L_1 = m, L_2 = n) &= \mathbb{P}\,(X_1 = \cdots = X_m = 1, X_{m+1} = \cdots = X_{m+n} = 0, X_{m+n+1} = 1)\\ &\quad + \mathbb{P}\,(X_1 = \cdots = X_m = 0, X_{m+1} = \cdots = X_{m+n} = 1, X_{m+n+1} = 0)\\ &= p^{m+1}q^n + q^{m+1}p^n.\end{aligned}$$

$$\mathbb{P}\,(L_1 = m) = p^m q + q^m p.$$

Alors

$$\begin{aligned}\mathbb{P}\,(L_2 = n) &= \sum_{m=1}^{\infty} \mathbb{P}\,(L_1 = m, L_2 = n)\\ &= q^n \cdot \frac{p^2}{1-p} + p^n \cdot \frac{q^2}{1-q}\\ &= p^2 q^{n-1} + q^2 p^{n-1},\end{aligned}$$

et

$$\sum_{m=1}^{\infty} \mathbb{P}\,(L_1 = m) = \sum_{n=1}^{\infty} \mathbb{P}\,(L_2 = n) = 1.$$

Donc $\mathbb{P}\,(L_1 = \infty) = \mathbb{P}\,(L_2 = \infty) = 0$. L_1 et L_2 sont des v.a. à valeurs dans $\mathbb{N}^*$.

$$\mathbb{E}\,[L_1] = \sum_{m=1}^{\infty} m(p^m q + q^m p) = pq\left(\frac{1}{(1-p)^2} + \frac{1}{(1-q)^2}\right) = \frac{p}{q} + \frac{q}{p}.$$

$$\mathbb{E}\,[L_2] = \sum_{n=1}^{\infty} n(p^2 q^{n-1} + q^2 p^{n-1}) = \frac{p^2}{(1-q)^2} + \frac{q^2}{(1-p)^2} = 2.$$

$$\begin{aligned}\mathbb{E}[L_1L_2] &= \sum_{m,n\in\mathbb{N}^*} mn\mathbb{P}(L_1=m,L_2=n)\\ &= \sum_{m,n\in\mathbb{N}^*} mn(p^{m+1}q^n+q^{m+1}p^n)\\ &= \left(\sum_m mp^{m+1}\right)\left(\sum_n nq^n\right)+\left(\sum_m mq^{m+1}\right)\left(\sum_n np^n\right)\\ &= \frac{p^2}{(1-p)^2}\cdot\frac{q}{(1-q)^2}+\frac{q^2}{(1-q)^2}\cdot\frac{p}{(1-p)^2}\\ &= \frac{1}{q}+\frac{1}{p}.\end{aligned}$$

$$\mathrm{Cov}(L_1,L_2)=\mathbb{E}[L_1L_2]-\mathbb{E}[L_1]\mathbb{E}[L_2]=\left(\frac{1}{q}+\frac{1}{p}\right)-\left(\frac{p}{q}+\frac{q}{p}\right)\cdot 2=-\frac{(p-q)^2}{pq}\leqslant 0.$$

On observe que $\mathbb{E}[L_1L_2]<\mathbb{E}[L_1]\mathbb{E}[L_2]$. □

Exemple 3.5.7 Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. indépendantes suivant toutes la loi $\mathcal{B}(p)$, $p\in\,]0,1[$. $\forall k\in\mathbb{N}^*, Y_k :=$ « le temps d'attente du k-ième 1 ». On pose $Z_1=Y_1$, $Z_k=Y_k-Y_{k-1}$ pour $k\geqslant 2$. On a vu que $(Z_n)_{n\in\mathbb{N}}$ est une suite de v.a. indépendantes, suivant toutes la loi $\mathcal{G}(p)$. Donc

$$\mathbb{E}[Z_n]=\frac{1}{p},\quad \mathbb{V}(Z_n)=\frac{1-p}{p^2}.$$

$$\mathbb{E}[Y_k]=\mathbb{E}\left[\sum_{i=1}^k Z_i\right]=\sum_{i=1}^k\mathbb{E}[Z_i]=\frac{k}{p}.$$

$$\mathbb{V}(Y_k)=\sum\mathbb{V}(Z_i)=\frac{k(1-p)}{p^2}.$$

On peut généraliser la définition de l'espérance à une v.a.d. $X:(\Omega,\mathcal{F},\mathbb{P})\to\overline{\mathbb{R}}=\mathbb{R}\cup\{\pm\infty\}$.

$$\mathbb{E}[X]=\sum_{x\in X(\Omega)}x^+\mathbb{P}(X=x)-\sum_{x\in X(\Omega)}x^-\mathbb{P}(X=x),\quad X(\Omega)\subset\overline{\mathbb{R}}.$$

Conventions :

$$+\infty\times p=\begin{cases}+\infty & \text{si } p>0,\\ 0 & \text{si } p=0,\end{cases}\qquad -\infty\times p=\begin{cases}-\infty & \text{si } p>0,\\ 0 & \text{si } p=0.\end{cases}$$

$$(+\infty)^+=+\infty,\quad (+\infty)^-=0,\quad (-\infty)^+=0,\quad (-\infty)^-=+\infty.$$

Si $X:\Omega\to\overline{\mathbb{R}}_+$, $\mathbb{E}[X]$ est toujours bien définie dans $\overline{\mathbb{R}}_+$.

3.6 Inégalités probabilistes

Théorème 3.6.1 (Inégalité de Markov) Soit $X:(\Omega,\mathcal{F},\mathbb{P})\to\mathbb{R}_+$ une v.a.d. positive. Alors pour tout $\varepsilon>0$,

$$\mathbb{P}(X\geqslant\varepsilon)\leqslant\frac{\mathbb{E}[X]}{\varepsilon}.$$

Application : Si $\mathbb{E}[X] < +\infty$, alors pour tout $c > 0$,

$$\mathbb{P}(X \geqslant c\mathbb{E}[X]) \leqslant \frac{1}{c}.$$

Démonstration. On a

$$X\mathbb{1}_{\{X \geqslant \varepsilon\}} \geqslant \varepsilon\mathbb{1}_{\{X \geqslant \varepsilon\}},$$

car sur $\{X \geqslant \varepsilon\}$, $X(\omega) \geqslant \varepsilon$. De plus,

$$X\mathbb{1}_{\{X < \varepsilon\}} = X\mathbb{1}_{\{0 \leqslant X < \varepsilon\}} \geqslant 0.$$

Comme $\mathbb{1} = \mathbb{1}_{\{X \geqslant \varepsilon\}} + \mathbb{1}_{\{X < \varepsilon\}}$, on a

$$X = X\mathbb{1}_{\{X \geqslant \varepsilon\}} + X\mathbb{1}_{\{X < \varepsilon\}} \geqslant \varepsilon\mathbb{1}_{\{X \geqslant \varepsilon\}} + 0.$$

En prenant l'espérance,

$$\mathbb{E}[X] \geqslant \mathbb{E}\left[\varepsilon\mathbb{1}_{\{X \geqslant \varepsilon\}}\right] = \varepsilon\mathbb{P}(X \geqslant \varepsilon).$$

D'où

$$\mathbb{P}(X \geqslant \varepsilon) \leqslant \frac{\mathbb{E}[X]}{\varepsilon}.$$

□

Remarque Si X est une v.a. positive, alors pour tout $\varepsilon > 0$,

$$\mathbb{P}(X > \varepsilon) \leqslant \frac{\mathbb{E}[X]}{\varepsilon}.$$

Théorème 3.6.2 (Inégalité de Bienaymé-Tchebychev) Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ une v.a. admettant un moment d'ordre 2. Alors pour tout $\varepsilon > 0$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geqslant \varepsilon) \leqslant \frac{\mathbb{V}(X)}{\varepsilon^2}.$$

Application : Pour tout $c > 0$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geqslant c \cdot \sigma(X)) \leqslant \frac{\mathbb{V}(X)}{c^2\sigma^2(X)} = \frac{1}{c^2}.$$

Typiquement, X s'écarte de son espérance $\mathbb{E}[X]$ d'au plus une constante fois $\sigma(X)$.

Démonstration. On a

$$\{|X - \mathbb{E}[X]| \geqslant \varepsilon\} = \{(X - \mathbb{E}[X])^2 \geqslant \varepsilon^2\}.$$

Donc

$$\mathbb{P}(|X - \mathbb{E}[X]| \geqslant \varepsilon) = \mathbb{P}\left((X - \mathbb{E}[X])^2 \geqslant \varepsilon^2\right) \leqslant \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{\varepsilon^2} = \frac{\mathbb{V}(X)}{\varepsilon^2}.$$

□

Méthode des moments

Exemple 3.6.1 Considérons une suite d'urnes numérotées de 1 à n. Dans l'urne i, il y a 1 boule noire et i boules blanches. On tire successivement et indépendamment une boule uniformément au hasard de chaque urne.

Soit $X_n =$ « nombre de boules noires tirées lors des n premiers tours ». Quelle est la vitesse de croissance de X_n ?

Solution. On écrit $X_n = \sum_{k=1}^{n} \mathbb{1}_{A_k}$, où $A_k =$ « on tire la boule noire de la k-ième urne ». Par linéarité de l'espérance,

$$\mathbb{E}[X_n] = \sum_{k=1}^{n} \mathbb{E}[\mathbb{1}_{A_k}] = \sum_{k=1}^{n} \mathbb{P}(A_k) = \sum_{k=1}^{n} \frac{1}{k+1} \sim \log n. \quad (\text{lorsque } n \to +\infty)$$

Les A_k $(k \geqslant 1)$ sont indépendantes, ainsi que les $\mathbb{1}_{A_k}$ $(k \geqslant 1)$.

$$\mathbb{V}(X_n) = \sum_{k=1}^{n} \mathbb{V}(\mathbb{1}_{A_k}) = \sum_{k=1}^{n} \frac{1}{k+1}\left(1 - \frac{1}{k+1}\right) \leqslant \sum_{k=1}^{n} \frac{1}{k+1} = \mathbb{E}[X_n] \leqslant \log(n+1).$$

Par l'inégalité de Bienaymé-Tchebychev,

$$\mathbb{P}\left(|X_n - \mathbb{E}[X_n]| \geqslant 10\sqrt{\log(n+1)}\right) \leqslant \frac{\mathbb{V}(X_n)}{100\log(n+1)} \leqslant \frac{1}{100}.$$

Pour tout n fixé, il y a au moins 99% de chance que

$$X_n \in \left[\mathbb{E}[X_n] - 10\sqrt{\log(n+1)},\ \mathbb{E}[X_n] + 10\sqrt{\log(n+1)}\right].$$

Pour n grand, $\mathbb{E}[X_n] \approx \log n$. □

Remarque On peut raffiner l'estimation de X_n en utilisant la méthode des moments d'ordre supérieur. Par exemple, on peut utiliser l'inégalité $\mathbb{P}(|X - \mathbb{E}[X]| \geqslant \varepsilon) \leqslant \dfrac{\mathbb{E}[(X - \mathbb{E}[X])^4]}{\varepsilon^4}$.

Remarque Si X admet un moment exponentiel fini, c'est-à-dire si $\mathbb{E}\left[e^X\right] < +\infty$, on peut utiliser l'inégalité suivante :

$$\mathbb{P}(X \geqslant \varepsilon) = \mathbb{P}\left(e^X \geqslant e^\varepsilon\right) = \frac{\mathbb{E}\left[e^X\right]}{e^\varepsilon}.$$

3.7 Fonctions génératrices

Définition 3.7.1 (Fonction génératrice) Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{N}$ une v.a.. **La fonction génératrice de (la loi de)** X est

$$G_X(t) := \mathbb{E}\left[t^X\right] = \sum_{n \in \mathbb{N}} t^n \mathbb{P}(X = n),$$

et $G_X(t)$ est définie si et seulement si $\sum_{n=0}^{\infty} t^n \mathbb{P}(X = n)$ converge absolument.

Convention : $0^0 = 1$, donc $G_X(0) = \mathbb{P}(X = 0)$.

Théorème 3.7.1 Deux v.a. à valeurs dans $\mathbb{N}$ ayant la même fonction génératrice ont la même loi.

Démonstration. Les probabilités $\mathbb{P}(X = n)$ pour $n \geqslant 0$ sont les coefficients du développement de Taylor de G_X en 0. La loi de X est déterminée par G_X grâce à l'unicité de la décomposition en série entière. □

Proposition 3.7.1 Soit $X : \Omega \to \mathbb{N}$ une v.a. Le rayon de convergence R de la série entière $\sum_{n=0}^{\infty} \mathbb{P}(X = n)\, t^n$ est supérieur ou égal à 1. La fonction G_X est (au moins) définie et continue sur $[-1, 1]$ et pour tout $k \in \mathbb{N}$,

$$\mathbb{P}(X = k) = k! \cdot G_X^{(k)}(0),$$

où $G_X^{(k)}$ est la k-ième dérivée de G_X.

Démonstration. Pour tout $t \in \mathbb{C}$ avec $|t| \leqslant 1$,

$$\sum_{n=0}^{\infty} \mathbb{P}(X = n)\, |t^n| \leqslant \sum_{n=0}^{\infty} \mathbb{P}(X = n) = 1 < +\infty.$$

Donc le rayon de convergence $R \geqslant 1$ et $\sum_{n=0}^{\infty} \mathbb{P}(X = n)\, t^n$ converge normalement sur $[-1, 1]$. Par les propriétés des séries entières, la fonction G_X est définie et continue sur $[-1, 1]$. Le théorème de dérivation des séries entières implique que G_X est C^∞ sur $]-R, R[$.

Pour tout $k \geqslant 0$ et $-1 < t < 1$,

$$G_X^{(k)}(t) = \sum_{n=k}^{+\infty} n(n-1)\dots(n-k+1) t^{n-k} \mathbb{P}(X = n) = \sum_{n=k}^{\infty} \frac{n!}{(n-k)!} t^{n-k} \mathbb{P}(X = n).$$

En prenant $t = 0$,

$$G_X^{(k)}(0) = k! \mathbb{P}(X = k).$$

□

Lemme 3.7.1 Soit $(a_n)_{n\in\mathbb{N}}$ une suite à termes positifs (c'est-à-dire $a_n \geqslant 0$) telle que $\sum a_n$ converge. On note $f(t) = \sum_{n\in\mathbb{N}} a_n t^n$. Alors f est dérivable à gauche en 1 si et seulement si $\sum n a_n$ converge. On a alors $f'(1-) = \sum_{n\geqslant 1} n a_n$.

Démonstration. On note R le rayon de convergence de la série $\sum a_n t^n$. Comme $\sum a_n < +\infty$ et $a_n \geqslant 0$, on sait que $R \geqslant 1$. La fonction f est C^∞ sur $]-R, R[$ et

$$f'(t) = \sum_{n=1}^{\infty} n a_n t^{n-1}, \quad \forall t \in]-R, R[.$$

— Si $R > 1$, alors f est dérivable en 1 et $f'(1) = \sum n a_n$ qui converge.

— Si $R = 1$, pour tout $t \in [0, 1[$,

$$\frac{f(t) - f(1)}{t - 1} = \sum_{n=1}^{\infty} a_n \frac{t^n - 1}{t - 1} = \sum_{n=1}^{\infty} a_n (t^{n-1} + \cdots + t + 1).$$

La fonction $t \mapsto \dfrac{f(t) - f(1)}{t - 1}$ est donc croissante sur $[0, 1[$.
Si f est dérivable à gauche en 1, alors

$$\forall 0 < t < 1,\ \forall N \in \mathbb{N}^*, \quad \sum_{n=1}^{N} a_n (t^{n-1} + \cdots + 1) \leqslant \frac{f(t) - f(1)}{t - 1} \leqslant f'(1-).$$

En prenant $t \uparrow 1$, on a

$$\sum_{n=1}^{N} n a_n \leqslant f'(1-).$$

et donc $\sum\limits_{n=1}^{\infty} n a_n \leqslant f'(1-) < +\infty$.
Si $\sum\limits_{n=1}^{\infty} n a_n < +\infty$, alors

$$\forall 0 < t < 1, \quad \frac{f(t) - f(1)}{t - 1} \leqslant \sum_{n=1}^{\infty} n a_n.$$

Lorsque $t \uparrow 1$, $\dfrac{f(t) - f(1)}{t - 1}$ est croissante et majorée, donc la limite $\lim\limits_{t \to 1^-} \dfrac{f(t) - f(1)}{t - 1} = f'(1-)$ existe dans $\mathbb{R}$ et $f'(1-) \leqslant \sum\limits_{n=1}^{\infty} n a_n$.
Donc $f'(1-)$ existe si et seulement si $\sum\limits_{n=1}^{\infty} n a_n < +\infty$. Dans ce cas, $f'(1-) = \sum\limits_{n=1}^{\infty} n a_n$.

□

Théorème 3.7.2 Soit $X : \Omega \to \mathbb{N}$ une v.a..

1) $\mathbb{E}[X] < +\infty$ si et seulement si G_X est dérivable à gauche en 1. On a alors $\mathbb{E}[X] = G'_X(1-)$.
2) $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ si et seulement si G_X est deux fois dérivable à gauche en 1. On a alors $\mathbb{E}[X(X-1)] = G''_X(1-)$.

Démonstration. On pose $0 \leqslant p_n = \mathbb{P}(X = n) \leqslant 1$. Alors

$$\sum_{n=0}^{\infty} p_n = 1, \quad \mathbb{E}[X] = \sum_{n=0}^{\infty} n p_n.$$

1) Le résultat découle directement du lemme ci-dessus.
2) Si $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, alors $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Par 1), G_X est dérivable à gauche en 1 et

$$G'_X(t) = \sum_{n \geqslant 1} n p_n t^{n-1} \quad \text{pour } t \in [0, 1[, \quad G'_X(1-) = \sum_{n \geqslant 1} n p_n < +\infty.$$

En appliquant le lemme ci-dessus à G'_X, on voit que G'_X est dérivable à gauche en 1 si et seulement si

$$\sum_{n\geqslant 2} n(n-1)p_n < +\infty \Leftrightarrow \sum_{n\geqslant 2} n^2 p_n < +\infty.$$

En particulier, $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ implique que G_X est deux fois dérivable à gauche en 1.

Si $G''_X(1-)$ existe, alors $\sum n^2 p_n < +\infty$, on en déduit que $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Donc $G''_X(1-)$ existe $\Leftrightarrow X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Dans ce cas,

$$G''_X(1-) = \sum_{n\geqslant 2} n(n-1)p_n = \mathbb{E}\left[X(X-1)\right].$$

□

Exemple 3.7.1

1) $X \sim \mathcal{B}(p)$:

$$G_X(t) = 1 - p + pt, \quad \text{rayon de convergence } R = +\infty.$$

$$\mathbb{E}[X] = G'_X(1) = p, \quad \mathbb{E}\left[X^2\right] = G''_X(1) + G'_X(1) = p, \quad \mathbb{V}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = p(1-p).$$

2) $X \sim \mathcal{G}(p)$ avec $p \in]0,1[$: Soit $q = 1-p$,

$$G_X(t) = \sum_{n=1}^{\infty} t^n p q^{n-1} = \frac{p}{q} \cdot \frac{qt}{1-qt} = \frac{pt}{1-qt},$$

avec rayon de convergence $R = \dfrac{1}{q} > 1$.

$$G'_X(t) = \frac{p}{(1-qt)^2}, \quad G''_X(t) = \frac{2pq}{(1-qt)^3}.$$

$$\mathbb{E}[X] = G'_X(1) = \frac{1}{p}, \quad \mathbb{E}\left[X^2\right] = G''_X(1) + G'_X(1) = \frac{1+q}{p^2}, \quad \mathbb{V}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \frac{q}{p^2}.$$

3) $X \sim \mathcal{P}(\lambda)$ avec $\lambda > 0$:

$$G_X(t) = \sum_{n=0}^{\infty} t^n \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda+\lambda t}, \quad R = +\infty.$$

$$G'_X(t) = \lambda e^{-\lambda+\lambda t}, \quad G''_X(t) = \lambda^2 e^{-\lambda+\lambda t}.$$

$$\mathbb{E}[X] = G'_X(1) = \lambda, \quad \mathbb{E}\left[X^2\right] = \lambda^2 + \lambda, \quad \mathbb{V}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \lambda.$$

Plus généralement, pour tout $t \in]-R, R[$ avec $R \geqslant 1$,

$$G_X^{(k)}(t) = \sum_{n=0}^{\infty} n(n-1)\dots(n-k+1)t^{n-k}\mathbb{P}(X=n),$$

et

$$\mathbb{E}\left[X(X-1)\dots(X-k+1)\right] = \sum_{n=0}^{\infty} n(n-1)\dots(n-k+1)\mathbb{P}(X=n).$$

Proposition 3.7.2 Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{N}$ une v.a..

— Pour tout $k \geqslant 1$,

$$\lim_{t \to 1-} G_X^{(k)}(t) = \mathbb{E}\left[X(X-1)\dots(X-k+1)\right] \in [0, +\infty].$$

— X admet un moment d'ordre k si et seulement si G_X est k fois dérivable à gauche en 1, c-à-d

$$G_X^{(k)}(1-) = \lim_{t \to 1-} G_X^{(k)}(t) < +\infty.$$

Proposition 3.7.3 Si X, Y sont des v.a. indépendantes à valeurs dans $\mathbb{N}$, alors pour tout $t \in \mathbb{R}$ tel que $G_X(t)$ et $G_Y(t)$ sont définies, $G_{X+Y}(t)$ l'est aussi et

$$G_{X+Y}(t) = G_X(t) G_Y(t).$$

Plus généralement, soient $X_1, \dots, X_n : \Omega \to \mathbb{N}$ des v.a. indépendantes, alors pour tout $t \in \mathbb{R}$ tel que $G_{X_i}(t)$ pour $1 \leqslant i \leqslant n$ sont définies, $G_{\sum_1^n X_i}(t)$ est aussi définie et

$$G_{\sum_1^n X_i}(t) = \prod_{i=1}^{n} G_{X_i}(t).$$

Démonstration. Supposons que $G_X(t)$ et $G_Y(t)$ sont définies, c'est-à-dire t^X et t^Y ont une espérance finie. $X \perp\!\!\!\perp Y \Rightarrow t^X \perp\!\!\!\perp t^Y$. Donc $t^X \cdot t^Y = t^{X+Y} \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ et

$$G_{X+Y}(t) = \mathbb{E}\left[t^{X+Y}\right] = \mathbb{E}\left[t^X\right] \mathbb{E}\left[t^Y\right] = G_X(t) G_Y(t).$$

On peut démontrer le résultat général par récurrence. □

Exemple 3.7.2

1) Si $X = \sum_{i=1}^{n} X_i$ avec $X_i \sim \mathcal{B}(p)$ indépendantes, alors $X \sim \mathcal{B}(n, p)$, donc

$$G_X(t) = \prod_{i=1}^{n} G_{X_i}(t) = (1 - p + pt)^n.$$

2) Si $X_j \sim \mathcal{B}(n_j, p)$ pour $1 \leqslant j \leqslant k$ sont indépendantes et $X = \sum_{j=1}^{k} X_j$, alors

$$G_X(t) = \prod_{j=1}^{k} G_{X_j}(t) = \prod_{j=1}^{k} (1 - p + pt)^{n_j} = (1 - p + pt)^{\sum_{j=1}^{k} n_j},$$

donc $X \sim \mathcal{B}\left(\sum_{j=1}^{k} n_j, p\right)$.

3) Si $X_j \sim \mathcal{P}(\lambda_j)$ pour $1 \leqslant j \leqslant k$ sont indépendantes et $X = \sum_{j=1}^{k} X_j$, alors

$$G_X(t) = \prod_{j=1}^{k} G_{X_j}(t) = \prod_{j=1}^{k} e^{\lambda_j (t-1)} = e^{\sum_{j=1}^{k} \lambda_j (t-1)},$$

donc $X \sim \mathcal{P}\left(\sum_{j=1}^{k} \lambda_j\right)$.

Chapitre 4

Loi des grands nombres

4.1 Loi des grands nombres

Théorème 4.1.1 (Loi faible des grands nombres) Soit $X_n : \Omega \to \mathbb{R}, n \geqslant 1$ une suite de v.a.d. réelles 2 à 2 indépendantes et de même loi. On suppose que $X_n \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. On pose $m = \mathbb{E}[X_1] \in \mathbb{R}$ et $S_n = \sum_{k=1}^{n} X_k, n \geqslant 1$. Alors

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - m\right| > \varepsilon\right) = 0.$$

Démonstration. On a

$$\mathbb{E}\left[\frac{S_n}{n}\right] = \frac{1}{n}\mathbb{E}[S_n] = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[X_k] = \frac{1}{n}\sum_{k=1}^{n} m = m.$$

Comme $X_k \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ et sont 2 à 2 indépendantes, on a

$$\mathbb{V}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\mathbb{V}(S_n) = \frac{1}{n^2}\sum_{k=1}^{n}\mathbb{V}(X_k) = \frac{\sigma^2}{n}, \quad \text{où } \sigma^2 = \mathbb{V}(X_1).$$

Donc pour tout $\varepsilon > 0$, par l'inégalité de Bienaymé-Tchebychev, on a

$$0 \leqslant \mathbb{P}\left(\left|\frac{S_n}{n} - m\right| > \varepsilon\right) = \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| > \varepsilon\right) \leqslant \frac{\mathbb{V}\left(\frac{S_n}{n}\right)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2 n}.$$

En prenant $n \to \infty$, on a

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - m\right| > \varepsilon\right) = 0.$$

□

Théorème 4.1.2 (Loi forte des grands nombres) Avec les mêmes hypothèses, il existe un événement négligeable A (c'est-à-dire $\mathbb{P}(A) = 0$) tel que

$$\forall \omega \in \Omega \backslash A, \quad \lim_{n\to\infty} \frac{S_n(\omega)}{n} = m = \mathbb{E}[X_1].$$

Autrement dit, $\frac{S_n}{n} \xrightarrow[n\to\infty]{} m$ p.s..

Démonstration. Fixons $\varepsilon > 0$. Pour $n \geqslant 1$, on pose

$$A_n = \left\{\omega \in \Omega : \exists\, p \in [\![n^2, (n+1)^2 - 1]\!], \text{ t.q. } \left|\frac{S_p(\omega)}{n^2} - m\right| > \varepsilon\right\} \in \mathcal{F}.$$

Si $\omega \in A_n$, alors

$$\text{soit } \left|\frac{S_{n^2}(\omega)}{n^2} - m\right| > \frac{\varepsilon}{2}, \quad \text{soit } \exists\, p \in [\![n^2+1, (n+1)^2-1]\!] \text{ t.q. } \left|\frac{S_p - S_{n^2}}{n^2}(\omega)\right| > \frac{\varepsilon}{2}.$$

Par l'inégalité de Bienaymé-Tchebychev,

$$\mathbb{P}\left(\left|\frac{S_{n^2}}{n^2} - m\right| > \frac{\varepsilon}{2}\right) \leqslant \frac{\mathbb{V}\left(\frac{S_{n^2}}{n^2}\right)}{\left(\frac{\varepsilon}{2}\right)^2} = \frac{\frac{1}{n^4}n^2\sigma^2}{\frac{\varepsilon^2}{4}} = \frac{4\sigma^2}{n^2\varepsilon^2},$$

et pour $p \in [\![n^2+1, (n+1)^2-1]\!]$,

$$\mathbb{P}\left(\left|\frac{S_p - S_{n^2}}{n^2} - \frac{(p-n^2)m}{n^2}\right| > \frac{\varepsilon}{4}\right) \leqslant \frac{\mathbb{V}\left(\frac{S_p - S_{n^2}}{n^2}\right)}{\left(\frac{\varepsilon}{4}\right)^2} = \frac{16(p-n^2)\sigma^2}{n^4\varepsilon^2} \leqslant \frac{32\sigma^2}{n^3\varepsilon^2},$$

car $p - n^2 \leqslant 2n$.

Si $\dfrac{2|m|}{n} < \dfrac{\varepsilon}{4}$, alors

$$\forall p \in [\![n^2+1, n^2+2n]\!], \quad \frac{(p-n^2)|m|}{n^2} \leqslant \frac{2|m|}{n} < \frac{\varepsilon}{4},$$

et donc

$$\left\{\omega \in \Omega : \left|\frac{S_p - S_{n^2}}{n^2}(\omega)\right| > \frac{\varepsilon}{2}\right\} \subset \left\{\omega \in \Omega : \left|\frac{S_p - S_{n^2}}{n^2}(\omega) - \frac{(p-n^2)m}{n^2}\right| > \frac{\varepsilon}{4}\right\}.$$

On obtient donc

$$\begin{aligned}
\mathbb{P}(A_n) &\leqslant \mathbb{P}\left(\left|\frac{S_{n^2}}{n^2} - m\right| > \frac{\varepsilon}{2}\right) + \mathbb{P}\left(\exists\, p \in [\![n^2+1, n^2+2n]\!], \left|\frac{S_p - S_{n^2}}{n^2}\right| > \frac{\varepsilon}{2}\right) \\
&\leqslant \mathbb{P}\left(\left|\frac{S_{n^2}}{n^2} - m\right| > \frac{\varepsilon}{2}\right) + \sum_{p=n^2+1}^{n^2+2n} \mathbb{P}\left(\left|\frac{S_p - S_{n^2}}{n^2}\right| > \frac{\varepsilon}{2}\right) \\
&\leqslant \frac{4\sigma^2}{n^2\varepsilon^2} + 2n \cdot \frac{32\sigma^2}{n^3\varepsilon^2} = \frac{68\sigma^2}{n^2\varepsilon^2}, \quad \text{si } \frac{2|m|}{n} < \frac{\varepsilon}{4}.
\end{aligned}$$

Si on choisit $\varepsilon = n^{-\frac{1}{4}}$, il existe $N \in \mathbb{N}$ tel que

$$\forall n > N, \quad \frac{2|m|}{n} < \frac{1}{4n^{\frac{1}{4}}} = \frac{\varepsilon}{4}.$$

Par le premier lemme de Borel-Cantelli,

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{j\geqslant n}^{\infty} A_j\right) = 0,$$

c'est-à-dire p.s. seul un nombre fini de A_n est réalisé.

Si on pose

$$B = \left\{\omega \in \Omega : \exists\, n_0 > 0,\ \forall n \geqslant n_0,\ \forall p \in [\![n^2, (n+1)^2-1]\!],\ \left|\frac{S_p}{n^2} - m\right| \leqslant \frac{1}{n^{\frac{1}{4}}}\right\},$$

alors $\mathbb{P}(B) = 1$ car

$$B = \left(\bigcap_{n=1}^{\infty} \bigcup_{j \geqslant n} A_j \right)^C.$$

En particulier,

$$\forall \omega \in B, \quad \lim_{p \to \infty} \frac{S_p(\omega)}{(\lfloor \sqrt{p} \rfloor)^2} = m.$$

Mais $\dfrac{(\lfloor \sqrt{p} \rfloor)^2}{p} \xrightarrow[p \to \infty]{} 1$.

Par conséquent,

$$\forall \omega \in B, \quad \lim_{p \to \infty} \frac{S_p(\omega)}{p} = m.$$

□

D'autres versions de la loi des grands nombres à voir en TD :

Soient $X_n \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, $n \geqslant 1$ indépendantes et de même loi. $S_n = \sum_{k=1}^{n} X_k$, $n \geqslant 1$, $m = \mathbb{E}[X_1] \in \mathbb{R}$.

— **Loi forte** : $\mathbb{P}\left(\left\{ \lim_{n \to \infty} \dfrac{S_n}{n} = m \right\} \right) = 1.$

— **Loi faible** : $\forall \varepsilon > 0, \ \lim_{n \to \infty} \mathbb{P}\left(\left| \dfrac{S_n}{n} - m \right| > \varepsilon \right) = 0.$

Remarque Si $X_n \geqslant 0$, $n \geqslant 1$ sont indépendantes, de même loi, mais $\mathbb{E}[X_1] = +\infty$. On pose $S_n = \sum_{k=1}^{n} X_k$. Alors

$$\forall M > 0, \quad \lim_{n \to \infty} \mathbb{P}\left(\frac{S_n}{n} < M \right) = 0.$$

De plus, $\mathbb{P}\left(\left\{ \lim_{n \to \infty} \dfrac{S_n}{n} = +\infty \right\} \right) = 1.$