Chapitre 7 : Probabilité sur un univers fini

Table des matières

Dans tout le chapitre, Ω est un ensemble fini.

1 Rappels sur les univers probabilisés

1.1 Vocabulaire

On retiendra le tableau suivant qui fait le parallèle entre le vocabulaire ensembliste de la théorie des ensembles et le vocabulaire probabiliste.

Notation	**Vocabulaire ensembliste**	**Vocabulaire probabiliste**
Ω	ensemble plein	événement certain
$\varnothing$	ensemble vide	événement impossible
ω	élément de Ω	événement élémentaire
A	sous-ensemble de Ω	événement
$\overline{A}$	complémentaire de A dans Ω	événement contraire
$A \cup B$	réunion de A et B	événement A ou B
$A \cap B$	intersection de A et B	événement A et B
$A \cap B = \varnothing$	A et B sont disjoints	A et B sont incompatibles

1.2 Probabilité sur un univers fini

Définition 1. *Ensemble des parties de Ω.*
*On définit l'**ensemble des parties** de Ω, noté $\mathscr{P}(\Omega)$, l'ensemble*

$$\mathscr{P}(\Omega) = \{A,\ A \subset \Omega\}.$$

Définition 2. *Probabilité.*
*On appelle **probabilité** sur Ω toute application $\mathbf{P} : \mathscr{P}(\Omega) \longrightarrow [0,1]$ vérifiant :*

(i) $\mathbf{P}(\Omega) = 1$ *;*

(ii) *pour tous événements $A_1, \dots, A_n$ deux à deux incompatibles, on a*

$$\mathbf{P}(A_1 \cup \cdots \cup A_n) = \sum_{k=1}^{n} \mathbf{P}(A_k).$$

Définition 3. *Univers probabilisé.*
*Un **espace probabilisé** (fini) est un couple $(\Omega, \mathbf{P})$ où Ω est un univers fini et $\mathbf{P}$ est une probabilité sur Ω.*

Proposition 1. *Propriétés fondamentales.*
Soit $(\Omega, \mathbf{P})$ *un espace probabilisé. Alors :*

(i) *pour tout* $A \in \mathscr{P}(\Omega)$, $\mathbf{P}\left(\overline{A}\right) = 1 - \mathbf{P}(A)$, *en particulier,* $\mathbf{P}(\varnothing) = 0$ *;*

(ii) *pour tout* $(A, B) \in \mathscr{P}(\Omega)^2$, *si* $A \subset B$, *alors* $\mathbf{P}(B \backslash A) = \mathbf{P}(B) - \mathbf{P}(A)$, *en particulier* $\mathbf{P}(A) \leq \mathbf{P}(B)$ *;*

(iii) *pour tout* $(A, B) \in \mathscr{P}(\Omega)^2$, $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

Démonstration. (i) Comme les événements A et $\overline{A}$ et $A \cup \overline{A} = \Omega$, d'après la définition 2, on a

$$1 = \mathbf{P}(\Omega) = \mathbf{P}(A) + \mathbf{P}\left(\overline{A}\right),$$

soit $\mathbf{P}\left(\overline{A}\right) = 1 - \mathbf{P}(A)$.

Si l'on remarque que $\varnothing = \overline{\Omega}$, alors on a $\mathbf{P}(\varnothing) = 1 - 1 = 0$.

(ii) Il est clair que les événements A et $B \backslash A$ sont incompatibles et vérifient $B = A \cup B \backslash$. Ainsi, d'après (ii) de la définition 2, on a $\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B \backslash A)$.

Comme $\mathbf{P}(B \backslash A) \geq 0$, on a en particulier, $\mathbf{P}(A) \leq \mathbf{P}(B)$.

(iii) Il suffit de remarquer que $A \cup B = (A \backslash (A \cap B)) \cup (A \cap B) \cup (B \backslash (A \cap B))$. Les événements étant deux à deux incompatibles, on a

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

Comme $A \cap B \backslash A$ et $A \cap B \backslash B$, en utilisant (ii), on a $\mathbf{P}(A \backslash (A \cap B)) = \mathbf{P}(A) - \mathbf{P}(A \cap B)$ et $\mathbf{P}(B \backslash (A \cap B)) = \mathbf{P}(B) - \mathbf{P}(A \cap B)$, ce qui donne finalement

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

□

Proposition 2. *Si* $\Omega = \{\omega_1, \ldots, \omega_n\}$ *et si* $p_1, \ldots, p_n$ *sont des nombres positifs tels que* $p_1 + \cdots + p_n = 1$, *alors il existe une unique probabilité* $\mathbf{P}$ *sur* Ω *telle que*

$$\forall i \in [\![1, n]\!], \quad \mathbf{P}(\{\omega_i\}) = p_i.$$

Démonstration. L'unicité est claire.

Réciproquement, si $\mathbf{P}$ est une application définie sur $\mathscr{P}(\Omega)$ telle que pour tout $i \in [\![1, n]\!]$, $\mathbf{P}(\{\omega_i\}) = p_i$, il est aisé de vérifier que $\mathbf{P}$ vérifie (i) et (ii) de la définition 2.

□

Remarque 1. Autrement dit, pour définir une probabilité $\mathbf{P}$ sur un univers fini Ω, il suffit de définir $\mathbf{P}$ sur chaque singleton de Ω (aussi appelé **événement élémentaire**) de sorte que la somme donne 1. La formule générale est alors donnée par :

$$\forall A \in \mathscr{P}(\Omega), \quad \mathbf{P}(A) = \sum_{w \in A} \mathbf{P}(\{\omega\}).$$

Exemple 1. Si $\Omega = [\![1, n]\!]$, il existe une unique probabilité $\mathbf{P}$ sur Ω telle que

$$\forall k \in \Omega, \quad \mathbf{P}(\{k\}) = \frac{1}{n}.$$

Ainsi, pour tout $A \subset \Omega$, on a

$$\mathbf{P}(A) = \frac{\operatorname{card}(A)}{n}.$$

Cette probabilité s'appelle la **probabilité uniforme** sur Ω.

1.3 Conditionnement par un événement

Définition 4. *Probabilité conditionnelle.*

Soient A et B deux événements avec $\mathbf{P}(B) \neq 0$. On définit la ***probabilité de A sachant B****, notée $\mathbf{P}_B(A)$, par*

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

Remarque 2. On utilise aussi parfois la notation $\mathbf{P}(A|B)$.

Proposition 3. *Si B est un événement de probabilité non nulle, l'application $\mathbf{P}_B : A \in \mathscr{P}(\Omega) \longmapsto \mathbf{P}_B(A)$ est une probabilité sur Ω.*

Remarque 3. Comme l'application $\mathbf{P}_B : A \in \mathscr{P}(\Omega) \longmapsto \mathbf{P}_B(A)$ est une probabilité sur Ω, toutes les propriétés établies à la proposition 1 sont vérifiées par $\mathbf{P}_B$.

Démonstration. On vérifie les points (i) et (ii) de la définition 2.

(i) On a $\mathbf{P}_B(\Omega) = \dfrac{\mathbf{P}(B \cap \Omega)}{\mathbf{P}(B)} = \dfrac{\mathbf{P}(B)}{\mathbf{P}(B)} = 1$.

(ii) Soient $A_1, \dots, A_n$ des événements deux à deux incompatibles. En utilisant la distributivité de l'intersection sur l'union, on a

$$\mathbf{P}_B(A_1 \cup \dots \cup A_n) = \frac{\mathbf{P}((A_1 \cup \dots \cup A_n) \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}((A_1 \cap B) \cup \dots \cup (A_n \cap B))}{\mathbf{P}(B)}.$$

Comme les événements $A_1, \dots, A_n$ sont deux à deux incompatibles, les événements $A_1 \cap B, \dots, A_n \cap B$ le sont aussi. Comme $\mathbf{P}$ est une probabilité, on en déduit que

$$\mathbf{P}_B(A_1 \cup \dots \cup A_n) = \frac{\mathbf{P}(A_1 \cap B) + \dots + \mathbf{P}(A_n \cap B)}{\mathbf{P}(B)} = \sum_{k=1}^{n} \mathbf{P}_B(A_k).$$

□

Proposition 4. *Formule des probabilités composées.*

Soient $A_1, \dots, A_n$ des événements tels que $\mathbf{P}(A_1 \cap \dots \cap A_{n-1}) \neq 0$, alors

$$\mathbf{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbf{P}(A_1) \times \mathbf{P}_{A_1}(A_2) \times \mathbf{P}_{A_1 \cap A_2}(A_3) \times \dots \times \mathbf{P}_{A_1 \cap \dots \cap A_{n-1}}(A_n).$$

Démonstration. On prouve la proposition par récurrence sur le nombre d'événements.

Si $n = 1$, il n'y a rien à faire.

On suppose que pour tout n-uplet d'événements $(A_1, \dots, A_n) \in \mathscr{P}(\Omega)^n$ avec $\mathbf{P}(A_1 \cap \dots \cap A_{n-1}) \neq 0$,

$$\mathbf{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbf{P}(A_1) \times \mathbf{P}_{A_1}(A_2) \times \mathbf{P}_{A_1 \cap A_2}(A_3) \times \dots \times \mathbf{P}_{A_1 \cap \dots \cap A_{n-1}}(A_n).$$

Soient $A_1, \dots, A_n, A_{n+1}$ $(n+1)$ événements tels que $\mathbf{P}(A_1 \cap \dots \cap A_n) > 0$.

En utilisant la définition 4, on a

$$\mathbf{P}(A_1 \cap \dots \cap A_n \cap A_{n+1}) = \mathbf{P}((A_1 \cap \dots \cap A_n) \cap A_{n+1}) = \mathbf{P}_{A_1 \cap \dots \cap A_n}(A_{n+1}) \times \mathbf{P}(A_1 \cap \dots \cap A_n).$$

En utilisant l'hypothèse de récurrence, on a :

$$\mathbf{P}(A_1 \cap \dots \cap A_n \cap A_{n+1}) = \mathbf{P}(A_1) \times \mathbf{P}_{A_1}(A_2) \times \mathbf{P}_{A_1 \cap A_2}(A_3) \times \dots \times \mathbf{P}_{A_1 \cap \dots \cap A_{n-1}}(A_n) \times \mathbf{P}_{A_1 \cap \dots \cap A_n}(A_{n+1}).$$

□

Exemple 2. Une urne contient n boules vertes et n boules rouges. On pioche successivement et sans remise les boules, que l'on suppose indiscernable au toucher, de l'urne jusqu'à vider l'urne. Calculer la probabilité de vider l'urne en alternant boule verte et boule rouge.

Solution. Soit A l'événement dont on souhaite calculer la probabilité. Pour tout $i \in [\![1, 2n]\!]$, on introduit les événements R_i : « on pioche une boule rouge au i-ème tirage » et V_i : « on pioche une boule verte au i-ème tirage ». Ainsi,

$$A = (R_1 \cap V_2 \cap R_3 \cap \cdots \cap V_{2n}) \cup (V_1 \cap R_2 \cap V_3 \cap \cdots \cap R_{2n}).$$

Par incompatibilité, on a

$$\mathbf{P}(A) = \mathbf{P}(R_1 \cap V_2 \cap R_3 \cap \cdots \cap V_{2n}) + \mathbf{P}(V_1 \cap R_2 \cap V_3 \cap \cdots \cap R_{2n}).$$

On peut se convaincre que $\mathbf{P}(R_1 \cap V_2 \cap R_3 \cap \cdots \cap V_{2n}) = \mathbf{P}(V_1 \cap R_2 \cap V_3 \cap \cdots \cap R_{2n})$.

Comme $\mathbf{P}(R_1 \cap V_2 \cap \cdots \cap R_{2n-1}) > 0$, la formule des probabilités composées donne

$$\begin{aligned} \mathbf{P}(R_1 \cap V_2 \cap R_3 \cap \cdots \cap V_{2n}) &= \mathbf{P}(R_1) \times \mathbf{P}_{R_1}(V_2) \times \cdots \times \mathbf{P}_{R_1 \cap V_2 \cap R_3 \cap \cdots \cap R_{2n-1}}(R_1 \cap V_2 \cap R_3 \cap \cdots \cap V_{2n}) \\ &= \frac{n}{2n} \times \frac{n}{2n-1} \times \cdots \times \frac{1}{1} \\ &= \frac{(n!)^2}{(2n!)} \end{aligned}$$

Finalement $\mathbf{P}(A) = \dfrac{2(n!)^2}{(2n)!}$.

Définition 5. *Système complet d'événements.*

*Soient $A_1, \ldots, A_n$ des évéments de $\mathbf{\Omega}$. On dit que la famille $(A_1, \ldots A_n)$ est un **système complet d'événements** si :*

(i) *les événements $A_1, \ldots, A_n$ sont deux à deux incompatibles ;*

(ii) $\displaystyle\bigcup_{i=1}^{n} A_i = \Omega$.

Proposition 5. *Formule des probabilités totales.*

Soit $(A_1, \ldots, A_n)$ un système complet d'événements. Alors, pour tout événement B, on a

$$\mathbf{P}(B) = \sum_{k=1}^{n} \mathbf{P}(B \cap A_k).$$

Si en plus on suppose que pour tout $k \in [\![1, n]\!]$, on a $\mathbf{P}(A_k) \neq 0$, la relation précédente se réécrit en

$$\mathbf{P}(B) = \sum_{k=1}^{n} \mathbf{P}_{A_k}(B) \times \mathbf{P}(A_k).$$

Démonstration. Comme $\mathbf{\Omega} = \displaystyle\bigcup_{i=1}^{n} A_i$, et par distributivité de l'intersection sur l'union, on a $B = \displaystyle\bigcup_{i=1}^{n} (B \cap A_i)$. Comme les événements $A_1, \ldots, A_n$ sont deux à deux incompatibles, les événements $A_1 \cap B, \ldots, A_n \cap B$ le sont aussi. Ainsi, par définition, on a

$$\mathbf{P}(B) = \sum_{k=1}^{n} \mathbf{P}(A_k \cap B).$$

Lorsque pour tout $k \in [\![1, n]\!]$, $\mathbf{P}(A_k) \neq 0$, il suffit d'utiliser la formule des probabilités totales pour conclure sur la seconde forme.

□

Exemple 3. Soit $n \in \mathbf{N}$, $n \geq 2$. On dispose de n urnes numérotées de 1 à n. L'urne $k \in [\![1, n]\!]$ contient k boules blanches et $n - k$ boules noires. On choisit une urne au hasard et on pioche simultanément deux boules dans cette urne.

Calculer la probabilité d'obtenir deux boules blanches.

Solution. Soit A l'événement dont on souhaite calculer la probabilité. Pour $k \in [\![1, n]\!]$, on note U_k : « on choisit l'urne k ». Comme $(U_1, \dots, U_n)$ est un système complet d'événements dont les probabilités sont toutes non nulles, la formule des probabilités totales donne

$$\mathbf{P}(A) = \sum_{k=1}^{n} \mathbf{P}_{U_k}(A) \times \mathbf{P}(U_k) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{P}_{U_k}(A).$$

Lorsque l'on pioche deux boules dans l'urne k, le nombre total de tirages est $\binom{n}{2}$ et le nombre de cas favorable est $\binom{k}{2}$. Ainsi,

$$\mathbf{P}(A) = \frac{1}{n} \sum_{k=1}^{n} \frac{\binom{k}{2}}{\binom{n}{2}} = \frac{1}{n^2(n-1)} \sum_{k=1}^{n} k(k-1).$$

Or, $\displaystyle\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$ et $\displaystyle\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$, après calculs, on a

$$\mathbf{P}(A) = \frac{n+1}{3n}.$$

Proposition 6. *Formule de Bayès.*
Soient A et B deux événements de probabilité non nulle. Alors

$$\mathbf{P}_B(A) = \frac{\mathbf{P}_A(B)\,\mathbf{P}(A)}{\mathbf{P}(B)}.$$

Démonstration. Il suffit d'utiliser deux fois la formule des probabilités conditionnelles. □

1.4 Indépendance d'événements

Définition 6. *Indépendance de deux événements.*
*Soient A et B deux événements. On dit que A et B sont **indépendants** si $\mathbf{P}(A \cap B) = \mathbf{P}(A) \times \mathbf{P}(B)$.*

Remarque 4. Lorsque B est de probabilité non nulle, l'indépendance de A et B est équivalente à $\mathbf{P}_B(A) = \mathbf{P}(A)$.

Définition 7. *Indépendance mutuelle d'événements.*
*Soient $A_1, \dots, A_n$ des événements. On dit que les événements sont **mutuellement indépendants** si*

$$\forall I \in \mathscr{P}([\![1, n]\!]), \quad \mathbf{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbf{P}(A_i).$$

Remarque 5. L'indépendance mutuelle entraîne l'indépendance deux à deux, mais la réciproque est fausse comme le montre l'exemple ci-dessous.

Exemple 4. Soit $\Omega = [\![1, 4]\!]$ muni de la probabilité uniforme. Soient $A = \{1, 2\}$, $B = \{1, 3\}$ et $C = \{1, 4\}$.
On vérifie facilement que les événements A, B et C sont deux à deux indépendants mais ils ne sont pas mutuellement indépendants car

$$\mathbf{P}(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8} = \mathbf{P}(A)\,\mathbf{P}(B)\,\mathbf{P}(C).$$

2 Rappels sur les variables aléatoires réelles

2.1 Généralités

Définition 8. *Variable aléatoire.*
*Une **variable aléatoire réelle** est une application $X : \Omega \longrightarrow \mathbf{R}$.*

Notation :
Si $U \subset \mathbf{R}$, on note $(X \in U)$ l'événement

$$(X \in U) = X^{-1}(U) = \{\omega \in \Omega,\ X(\omega) \in U\}.$$

En particulier, $(X \leq x)$ $(x \in \mathbf{R})$ est l'événement $\{\omega \in \Omega,\ X(\omega) \leq x\}$.

Définition 9. *Image d'une variable aléatoire par une application.*
*Soit $f : X(\Omega) \longrightarrow \mathbf{R}$ une application. On définit la **variable aléatoire image** de X par f comme la composée $f \circ X$, i.e.*

$$\forall \omega \in \Omega, \quad (f \circ X)(\omega) = f(X(\omega)).$$

2.2 Loi d'une variable aléatoire réelle

Proposition 7. *L'application $\mathbf{P}_X$ définie sur $\mathscr{P}(X(\Omega))$ (qui est fini car Ω l'est) par :*

$$\forall U \in \mathscr{P}(X(\Omega)), \quad \mathbf{P}_X(U) = \mathbf{P}(X \in U)$$

est une probabilité sur $\mathscr{P}(X(\Omega))$.

Démonstration. On vérifie les points (i) et (ii) de la définition 2. On a

(i) On a
$$\mathbf{P}_X(X(\Omega)) = \mathbf{P}\left(X^{-1}(X(\Omega))\right) = \mathbf{P}(\Omega) = 1.$$

(ii) Soient $A_1, \ldots, A_n$ des événements deux à deux incompatibles de $X(\Omega)$. On a
$$\mathbf{P}_X\left(\bigcup_{i=1}^{n} A_i\right) = \mathbf{P}\left(X^{-1}\left(\bigcup_{i=1}^{n} A_i\right)\right) = \mathbf{P}\left(\bigcup_{i=1}^{n} X^{-1}(A_i)\right).$$

Comme les $A_1, \ldots, A_n$ sont deux à deux incompatibles, les événements $X^{-1}(A_1), \ldots, X^{-1}(A_n)$ le sont aussi. Comme $\mathbf{P}$ est une probabilité de Ω, on a
$$\mathbf{P}_X\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbf{P}\left(X^{-1}(A_i)\right) = \sum_{i=1}^{n} \mathbf{P}(X \in A_i) = \sum_{i=1}^{n} \mathbf{P}_X(A_i).$$

□

Définition 10. *Loi d'une variable aléatoire.*
*La variable aléatoire $\mathbf{P}_X$ s'appelle la **loi** de X.*

Proposition 8. *La loi $\mathbf{P}_X$ est entièrement déterminée par la donnée de $\mathbf{P}(X = x)$ pour $x \in X(\Omega)$.*

Démonstration. Supposons connues toutes les probabilités $\mathbf{P}(X = x)$ pour $x \in X(\Omega)$.
Soit $U \in \mathscr{P}(X(\Omega))$. On écrit $U = \{y_1, \ldots, y_m\}$. On a alors :

$$\mathbf{P}_X(U) = \mathbf{P}\left(X^{-1}(U)\right) = \mathbf{P}\left(\bigcup_{i=1}^{m} X^{-1}(\{y_i\})\right).$$

Comme les événements $X^{-1}(\{y_1\}), \ldots, X^{-1}(\{y_m\})$ sont deux à deux incompatibles, on a

$$\mathbf{P}_X(U) = \sum_{i=1}^{m} \mathbf{P}\left(X^{-1}(\{y_i\})\right) = \sum_{i=1}^{m} \mathbf{P}(X = y_i).$$

Or, $\mathbf{P}(X = y_i)$ est supposé connue pour tout $i \in [\![1, m]\!]$, donc $\mathbf{P}_X(U)$ l'est aussi.

□

2.3 Fonction de répartition

Définition 11. *Fonction de répartition.*
On définit la fonction de répartition F_X sur $\mathbf{R}$ par

$$\forall x \in \mathbf{R}, \quad F_X(x) = \mathbf{P}(X \leq x).$$

Proposition 9. *Fonction de répartition d'une variable aléatoire finie.*
Soit X une variable aléatoire sur Ω. On note $X(\Omega) = \{x_1, \ldots, x_n\}$ où $x_1 < \cdots < x_n$. On note aussi, pour tout $i \in [\![1, n]\!]$, $p_i = \mathbf{P}(X = x_i)$, alors on a

$$\forall x \in \mathbf{R}, \quad F_X(x) = \begin{cases} 0 & \text{si } x < x_1 \\ p_1 + \cdots + p_i & \text{si } x \in [x_i, x_{i+1}[\ \text{ et } i \in [\![1, n-1]\!] \\ 1 & \text{si } x \geq x_n \end{cases}.$$

Démonstration. On distingue trois cas :

(i) Comme $X(\Omega) = \{x_1, \ldots, x_n\}$, il est clair que si $x < x_1$, $F_X(x) = \mathbf{P}(X \leq x) = \mathbf{P}(\varnothing) = 0$.

(ii) Soit $x \in [x_i, x_{i+1}[$. On a $(X \leq x_i) = (X = x_1) \cup \cdots \cup (X = x_i)$. Ces événements étant deux à deux incompatibles, on a

$$F_X(x) = \mathbf{P}((X = x_1) \cup \cdots \cup (X = x_i)) = \sum_{k=1}^{i} \mathbf{P}(X = x_k) = \sum_{k=1}^{i} p_k.$$

(iii) Si $x \geq x_n$, on a $(X \leq x_n) = \Omega$, donc $F_X(x) = \mathbf{P}(\Omega) = 1$.

□

Corollaire 1. *Propriétés de la fonction de répartition.*
La fonction de répartition a les propriétés suivantes :

(i) *F_X est croissante sur $\mathbf{R}$;*

(ii) *$\lim\limits_{x \to -\infty} F_X(x) = 0$ et $\lim\limits_{x \to +\infty} F_X(x) = 1$;*

(iii) *F_X est continue à droite et admet une limite à gauche en tout point (fonction càdlàg) ;*

(iv) *deux variables aléatoires finies suivent la même loi si, et seulement si, elles ont la même fonction de répartition.*

Démonstration. La preuve est claire à l'aide de la proposition 9.

□

2.4 Espérance d'une variable aléatoire réelle

Lorsque Ω est fini, $X(\Omega)$ l'est aussi. La notion d'espérance ne pose pas de problème à définir.

Définition 12. *Espérance d'une variable aléatoire.*
*On définit l'**espérance** de X par :*

$$\mathbf{E}(X) = \sum_{x \in X(\Omega)} x \, \mathbf{P}(X = x).$$

Remarque 6. L'espérance correspond à la moyenne théorique.

Proposition 10. *Formule du transfert.*
Si $f : X(\Omega) \longrightarrow \mathbf{R}$, alors

$$\mathbf{E}(f(X)) = \sum_{x \in X(\Omega)} f(x) \, \mathbf{P}(X = x).$$

Démonstration. Par définition, on a

$$\mathbf{E}\left(f\left(X\right)\right) = \sum_{y \in f(X(\Omega))} y \, \mathbf{P}\left(f\left(X\right) = y\right).$$

On écrit $f\left(X\left(\Omega\right)\right) = \{y_1, \ldots, y_m\}$ et pour tout $i \in [\![1, m]\!]$, on pose $\mathscr{N}_i = f^{-1}\left(\{y_i\}\right)$ (ensemble des antécédents de y_i par f). En particulier,

$$\forall i \in [\![1, m]\!], \quad \mathbf{P}\left(f\left(X\right) = y_i\right) = \mathbf{P}\left(X \in \mathscr{N}_i\right) = \sum_{x \in \mathscr{N}_i} \mathbf{P}\left(X = x\right).$$

En utilisant le fait que $y_i = f\left(x\right)$ pour tout $x \in \mathscr{N}_i$, on a

$$\mathbf{E}\left(f\left(X\right)\right) = \sum_{i=1}^{m} y_i \sum_{x \in \mathscr{N}_i} \mathbf{P}\left(X = x\right) = \sum_{i=1}^{m} \sum_{x \in \mathscr{N}_i} f\left(x\right) \mathbf{P}\left(X = x\right).$$

Comme $\displaystyle\bigcup_{i=1}^{m} \mathscr{N}_i = X\left(\Omega\right)$, on en déduit que

$$\mathbf{E}\left(f\left(X\right)\right) = \sum_{x \in X(\Omega)} f\left(x\right) \mathbf{P}\left(X = x\right).$$

□

Proposition 11. *Linéarité de l'espérance.*
Si X est une variable aléatoire et $(a, b) \in \mathbf{R}^2$, alors

$$\mathbf{E}\left(aX + b\right) = a\,\mathbf{E}\left(X\right) + b.$$

Démonstration. D'après la formule du transfert, on a

$$\begin{aligned} \mathbf{E}\left(aX + b\right) &= \sum_{x \in X(\Omega)} \left(ax + b\right) \mathbf{P}\left(X = x\right) \\ &= a \sum_{x \in X(\Omega)} x\,\mathbf{P}\left(X = x\right) + b \sum_{x \in X(\Omega)} \mathbf{P}\left(X = x\right) \\ &= a\,\mathbf{E}\left(X\right) + b. \end{aligned}$$

□

Le résultat suivant généralise le résultat précédent.

Proposition 12. *Soient X et Y deux variables aléatoires et $a \in \mathbf{R}$. Alors*

$$\mathbf{E}\left(aX + Y\right) = a\,\mathbf{E}\left(X\right) + \mathbf{E}\left(Y\right).$$

Démonstration. Admis.

□

On en déduit les propriétés suivantes de l'espérance.

Proposition 13. *Propriétés de l'espérance.*
Soient X et Y deux variables aléatoires. Alors :

(i) *si X prend des valeurs positives, alors $\mathbf{E}\left(X\right) \geq 0$ (positivité de l'espérance) ;*
(ii) *si $X \leq Y$, alors $\mathbf{E}\left(X\right) \leq \mathbf{E}\left(Y\right)$ (croissance de l'espérance) ;*
(iii) $|\mathbf{E}\left(X\right)| \leq \mathbf{E}\left(|X|\right)$.

Démonstration. (i) C'est clair par définition de l'espérance.

(ii) La variable aléatoire $Y - X$ est positive, donc d'après (i), $\mathbf{E}(Y - X) \geq 0$. La linéarité permet de conclure.

(iii) Il suffit d'utiliser l'inégalité triangulaire pour les sommes et le théorème de transfert.

□

La proposition suivante, bien que hors programme, est fondamentale.

Proposition 14. *Inégalité de Markov.*
*Soit X une variable aléatoire sur Ω à **valeurs positives**. Alors*

$$\forall a > 0, \quad \mathbf{P}(X \geq a) \leq \frac{\mathbf{E}(X)}{a}.$$

Démonstration. Soit $a > 0$. On écrit :

$$\mathbf{E}(X) = \sum_{x \in X(\Omega)} x\,\mathbf{P}(X = x) = \sum_{\substack{x \in X(\Omega) \\ x < a}} x\,\mathbf{P}(X = x) + \sum_{\substack{x \in X(\Omega) \\ x \geq a}} x\,\mathbf{P}(X = x).$$

Comme $X(\Omega) \subset \mathbf{R}_+$, on a $\displaystyle\sum_{\substack{x \in X(\Omega) \\ x < a}} x\,\mathbf{P}(X = x) \geq 0$ et $\displaystyle\sum_{\substack{x \in X(\Omega) \\ x \geq a}} x\,\mathbf{P}(X = x) \geq a \sum_{\substack{x \in X(\Omega) \\ x \geq a}} \mathbf{P}(X = x) = a\,\mathbf{P}(X \geq a)$.

On en déduit ainsi

$$\mathbf{E}(X) \geq a\,\mathbf{P}(X \geq a).$$

□

Exemple 5. Soit $n \in \mathbf{N}^\star$. On lance n fois une pièce équilibrée et on suppose que les lancers sont mutuellement indépendants. On note X le rang du premier pile s'il existe et on pose $X = 0$ si l'on n'obtient pas de pile.

Calculer l'espérance de X.

Solution. Déjà $X(\Omega) = [\![0, n]\!]$. Pour $j \in [\![1, n]\!]$, on pose P_j : « le j-ème lancer est un pile » et on définit de même F_j.

Soit $k \in [\![1, n]\!]$. On a $(X = k) = F_1 \cap \cdots \cap F_{k-1} \cap P_k$. Par indépendance mutuelle des lancers, on a

$$\mathbf{P}(X = k) = \left(\frac{1}{2}\right)^k.$$

On en déduit

$$\mathbf{P}(X = 0) = 1 - \sum_{k=1}^{n} \mathbf{P}(X = k) = 1 - \sum_{k=1}^{n} \left(\frac{1}{2}\right)^k = 1 - \frac{1}{2} \times \frac{1 - \left(\frac{1}{2}\right)^n}{1 - \frac{1}{2}} = \left(\frac{1}{2}\right)^n.$$

On a alors

$$\mathbf{E}(X) = \sum_{k=0}^{n} k\,\mathbf{P}(X = k) = \sum_{k=1}^{n} k \left(\frac{1}{2}\right)^k = \frac{1}{2} \sum_{k=1}^{n} k \left(\frac{1}{2}\right)^{k-1}.$$

Pour calculer cette somme, on se souvient que

$$\forall x \in \mathbf{R} \setminus \{1\}, \quad \sum_{k=0}^{n} x^k = \frac{1 - x^{n+1}}{1 - x},$$

ainsi en dérivant cette relation, il vient que

$$\sum_{k=1}^{n} k x^{k-1} = \frac{(n+1)\,x^n\,(x - 1) - x^{n+1} + 1}{(1 - x)^2}.$$

Finalement

$$\mathbf{E}(X) = 2\left(1 - \frac{1}{2}(n+1)\left(\frac{1}{2}\right)^n - \left(\frac{1}{2}\right)^{n+1}\right).$$

2.5 Variance et écart type d'une variable aléatoire réelle

Définition 13. *Variance.*
On définit la ***variance*** *de X, notée $\mathbf{V}(X)$ par :*

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

Remarque 7. (i) La variance est un réel positif.
(ii) La variance est une mesure de la dispersion de X autour de sa moyenne $\mathbf{E}(X)$.
(iii) Une variable aléatoire est constante si, et seulement si, $\mathbf{V}(X) = 0$.

Définition 14. *Écart type.*
*On définit l'**écart type**, noté $\sigma(X)$, par $\sigma(X) = \sqrt{\mathbf{V}(X)}$.*

Proposition 15. *Formule de König-Huygens.*
On a $\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$.

Démonstration. En utilisant la linéarité de l'espérance, on a

$$\begin{aligned}
\mathbf{V}(X) &= \mathbf{E}\left((X - \mathbf{E}(X))^2\right) \\
&= \mathbf{E}\left(X^2 - 2X\,\mathbf{E}(X) + \mathbf{E}(X)^2\right) \\
&= \mathbf{E}(X^2) - 2\,\mathbf{E}(X)^2 + \mathbf{E}(X)^2 \\
&= \mathbf{E}(X^2) - \mathbf{E}(X)^2.
\end{aligned}$$

□

Proposition 16. *Soit $(a,b) \in \mathbf{R}^2$. On a $\mathbf{V}(aX+b) = a^2\mathbf{V}(X)$.*

Démonstration. En utilisant la formule de König-Huygens et la linéarité de l'espérance, on a

$$\begin{aligned}
\mathbf{V}(aX+b) &= \mathbf{E}\left((aX+b)^2\right) - \mathbf{E}(aX+b)^2 \\
&= \mathbf{E}(a^2X^2 + 2abX + b^2) - \left(a^2\,\mathbf{E}(X)^2 + 2ab\,\mathbf{E}(X) + b^2\right) \\
&= a^2\left(\mathbf{E}(X^2) - \mathbf{E}(X)^2\right) \\
&= a^2\mathbf{V}(X).
\end{aligned}$$

□

Proposition 17. *Inégalité de Bienaymé-Tchebyshev.*
On a

$$\forall \varepsilon > 0, \quad \mathbf{P}(|X - \mathbf{E}(X)| \geq \varepsilon) \leq \frac{\mathbf{V}(X)}{\varepsilon^2}.$$

Démonstration. La variable aléatoire $(X - \mathbf{E}(X))^2$ est à valeurs positives admet une espérance

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

L'inégalité de Markov (que l'on rappelle, est hors programme) donne

$$\mathbf{P}\left((X - \mathbf{E}(X))^2 \geq \varepsilon^2\right) \leq \frac{\mathbf{E}\left((X - \mathbf{E}(X))^2\right)}{\varepsilon^2}$$

soit

$$\mathbf{P}(|X - \mathbf{E}(X)| \geq \varepsilon) \leq \frac{\mathbf{V}(X)}{\varepsilon^2}.$$

□

Exemple 6. On dispose d'un dé équilibré à 6 faces. Deux jeux s'offrent à vous : le premier consiste à lancer le dé : si vous obtenez un nombre pair, vous gagnez un euro et vous perdez un euro sinon. Le second consiste à lancer le dé : si vous obtenur un nombre pair, vous gagner 10 000 euros et vous perdez 10 000 euros sinon.

Quel jeu choisissez-vous ?

Solution. Soit X la variable aléatoire donnant le gain algébrique dans le premier jeu et Y la variable aléatoire donnant le gain algébrique du second jeu.

On a $X(\Omega) = \{-1, 1\}$ et $Y(\Omega) = \{-10\,000, 10\,000\}$ avec $\mathbf{P}(X = 1) = \mathbf{P}(X = -1) = \dfrac{1}{2}$ et $\mathbf{P}(Y = 10\,000) = \mathbf{P}(Y = -10\,000) = \dfrac{1}{2}$.

On a alors $\mathbf{E}(X) = \mathbf{E}(Y) = 0$. Mais $X^2 = 1$ et $Y^2 = 10^8$ ainsi, d'après la formule de König-Huygens, on a $\mathbf{V}(X) = \mathbf{E}\left(X^2\right) - (\mathbf{E}(X))^2 = 1$ et $\mathbf{V}(Y) = \mathbf{E}\left(Y^2\right) - (\mathbf{E}(Y))^2 = 10^8$.

Pour répondre à la question, cela dépend de vous ! Ce que l'on remarque c'est que $\mathbf{V}(Y) > \mathbf{V}(X)$ ainsi, même si l'espérance de gain est la même, le second jeu est plus risqué !

Fallait-il faire tout cela pour s'en convaincre ? !

2.6 Lois usuelles

2.6.1 Loi uniforme

Définition 15. *Loi uniforme.*

*Soit $n \in \mathbf{N}^\star$. On dit que X suit le **loi uniforme** sur l'ensemble $\{x_1, \dots, x_n\}$ si*

$$X(\Omega) = \{x_1, \dots, x_n\} \quad \text{et} \quad \forall k \in [\![1, n]\!],\ \mathbf{P}(X = x_k) = \frac{1}{n}.$$

On note $X \hookrightarrow \mathscr{U}(\{x_1, \dots, x_n\})$.

Proposition 18. *Si $X \hookrightarrow \mathscr{U}([\![1, n]\!])$, alors*

$$\mathbf{E}(X) = \frac{n+1}{2} \quad \text{et} \quad \mathbf{V}(X) = \frac{n^2 - 1}{12}.$$

Démonstration. On sépare les deux calculs.

Espérance. On a

$$\begin{aligned} \mathbf{E}(X) &= \sum_{k=1}^{n} k\, \mathbf{P}(X = k) \\ &= \frac{1}{n} \sum_{k=1}^{n} k \\ &= \frac{1}{n} \times \frac{n(n+1)}{2} \\ &= \frac{n+1}{2}. \end{aligned}$$

Variance. On utilise la formule de König-Huygens :

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

D'après la formule du transfert, on a

$$\begin{aligned} \mathbf{E}\left(X^2\right) &= \sum_{k=1}^{n} k^2\, \mathbf{P}(X = k) \\ &= \frac{1}{n} \sum_{k=1}^{n} k^2 \\ &= \frac{1}{n} \times \frac{n(n+1)(2n+1)}{6} \\ &= \frac{(n+1)(2n+1)}{6}. \end{aligned}$$

Ainsi,

$$\begin{aligned}\mathbf{V}(X) &= \mathbf{E}\left(X^2\right) - \mathbf{E}(X)^2 \\ &= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \\ &= \frac{2(n+1)(2n+1)}{12} - \frac{3(n+1)}{12} \\ &= \frac{n^2-1}{12}.\end{aligned}$$

□

2.6.2 Loi de Bernoulli

Définition 16. *Loi de Bernoulli.*

Soit $p \in [0,1]$. Une variable aléatoire X suit la ***loi de Bernoulli*** *de paramètre p, et on note $X \hookrightarrow \mathscr{B}(p)$ si :*

$$X(\Omega) = \{0,1\} \quad \text{et} \quad \mathbf{P}(X=1) = p \text{ et } \mathbf{P}(X=0) = 1-p.$$

Remarque 8. Une variable aléatoire suivant une loi de Bernoulli modélise une expérience ayant deux issues possibles : succès (avec probabilité de p) et échec (probabilité de $1-p$).

Proposition 19. *Soit X une variable aléatoire qui suit une loi de Bernoulli de paramètre $p \in [0,1]$. Alors*

$$\mathbf{E}(X) = p \quad \text{et} \quad \mathbf{V}(X) = p(1-p).$$

Démonstration. On sépare les deux calculs.

Espérance. On a

$$\mathbf{E}(X) = 1 \times \mathbf{P}(X=1) + 0 \times \mathbf{P}(X=0) = p.$$

Variance. Comme $X^2 = X$, on a $\mathbf{E}\left(X^2\right) = \mathbf{E}(X) = p$. La formule de König-Huygens donne

$$\mathbf{V}(X) = \mathbf{E}\left(X^2\right) - \mathbf{E}(X)^2 = p - p^2 = p(1-p).$$

□

2.6.3 Loi binomiale

Définition 17. *Loi binomiale.*

Soient $n \in \mathbf{N}^\star$ et $p \in [0,1]$. Une variable aléatoire X suit la ***loi binomiale*** *de paramètres n et p, et on note $X \hookrightarrow \mathscr{B}(n,p)$ si :*

$$X(\Omega) = [\![0,n]\!] \quad \text{et} \quad \forall k \in [\![0,n]\!],\ \mathbf{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Remarque 9. Une loi binomiale de paramètres n et p compte le nombre de succès au cours d'une répétition de n épreuves de Bernoulli indépendantes de même paramètre p.

Proposition 20. *Soit $X \hookrightarrow \mathscr{B}(n,p)$ ($n \in \mathbf{N}^\star$ et $p \in [0,1]$), on a*

$$\mathbf{E}(X) = np \quad \text{et} \quad \mathbf{V}(X) = np(1-p).$$

On utilisera le lemme suivant, parfois appelé « lemme du pion » :

Lemme 1. *Soit $n \in \mathbf{N}^\star$ et $k \in [\![1,n]\!]$, on a*

$$k\binom{n}{k} = n\binom{n-1}{k-1}.$$

Démonstration. Il suffit d'écrire

$$k\binom{n}{k} = k \times \frac{n!}{(n-k)!k!} = n\frac{(n-1)!}{(n-1-(k-1))!\,(k-1)!} = n\binom{n-1}{k-1}.$$

□

On prouve la proposition 20.

Démonstration. On sépare les deux calculs.

Espérance. On a

$$\begin{array}{rcll} \mathbf{E}(X) & = & \displaystyle\sum_{k=0}^{n} k\,\mathbf{P}(X=k) & \\ & = & \displaystyle\sum_{k=1} k\binom{n}{k} p^k (1-p)^{n-k} & \\ & = & n\displaystyle\sum_{k=1}^{n} \binom{n-1}{k-1} p^k (1-p)^{n-k} & \text{lemme 1} \\ & = & n\displaystyle\sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^{\ell+1} (1-p)^{n-1-\ell} & \text{on pose } \ell = k-1 \\ & = & np\displaystyle\sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^{\ell} (1-p)^{n-1-\ell} & \\ & = & np\,(p+1-p)^{n-1} & \\ & = & np. & \end{array}$$

Variance. On commence par remarquer que

$$\mathbf{E}\left(X^2\right) = \mathbf{E}(X(X-1)) + \mathbf{E}(X).$$

D'après la formule de transfert, on a

$$\begin{array}{rcll} \mathbf{E}(X(X-1)) & = & \displaystyle\sum_{k=0}^{n} k(k-1)\,\mathbf{P}(X=k) & \\ & = & \displaystyle\sum_{k=2}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k} & \\ & = & n(n-1)\displaystyle\sum_{k=0}^{n-2} \binom{n-2}{k-2} p^k (1-p)^{n-k} & \text{lemme 1} \\ & = & n(n-1)\displaystyle\sum_{\ell=0}^{n-2} \binom{n-2}{\ell} p^{\ell+2} (1-p)^{n-2-\ell} & \text{on pose } \ell = k-2 \\ & = & n(n-1)\,p^2\,(p+1-p)^{n-2} & \\ & = & n(n-1)\,p^2. & \end{array}$$

Grâce à la formule de König-Huygens, on en déduit

$$\begin{array}{rcl} \mathbf{V}(X) & = & \mathbf{E}\left(X^2\right) - \mathbf{E}(X)^2 \\ & = & \mathbf{E}(X(X-1)) + \mathbf{E}(X) - \mathbf{E}(X)^2 \\ & = & n(n-1)\,p^2 + np - n^2p^2 \\ & = & np(1-p). \end{array}$$

□

3 Couples de variables aléatoires

On fixe X et Y deux variables aléatoires sur Ω. Le couple (X, Y) est un couple de variables aléatoires.

3.1 Généralités

Définition 18. *Loi conjointe.*
*La **loi conjointe** du couple (X, Y) est la donnée de $\mathbf{P}\left((X = x) \cap (Y = y)\right)$ pour tout $(x, y) \in X(\Omega) \times Y(\Omega)$.*

Définition 19. *Lois marginales.*
*La loi X et la loi Y s'appellent les **lois marginales** du couple (X, Y).*

Remarque 10. La loi du couple donne la loi des marginales. En effet, la formule des probabilités totales donne

$$\forall x \in X(\Omega), \quad \mathbf{P}(X = x) = \sum_{y \in Y(\Omega)} \mathbf{P}\left((X = x) \cap (Y = y)\right)$$

et

$$\forall y \in Y(\Omega), \quad \mathbf{P}(Y = x) = \sum_{x \in X(\Omega)} \mathbf{P}\left((X = x) \cap (Y = y)\right).$$

⚠La réciproque est fausse : la loi des marginales **ne donne pas** la loi du couple. Par exemple, si X et Y sont telles que $X(\Omega) = Y(\Omega) = \{0, 1\}$, on peut définir deux lois de couple différentes par :

$$\forall (x, y) \in \{0, 1\}^2, \quad \mathbf{P}\left((X = x) \cap (Y = y)\right) = \frac{1}{4}$$

et

$$\forall (x, y) \in \{0, 1\}^2, \quad \mathbf{P}\left((X = x) \cap (Y = y)\right) = \begin{cases} \frac{1}{2} & \text{si } (x, y) = (0, 0) \text{ ou } (1, 1) \\ 0 & \text{sinon} \end{cases},$$

alors que, dans les deux cas, les lois marginales X et Y suivent toutes les deux une loi uniforme sur l'ensemble $\{0, 1\}$.

Remarque 11. Les définitions précédentes s'étendent naturellement au n-uplet de variables aléatoires.

Définition 20. *Loi conditionnelle.*
*Soit $y \in Y(\Omega)$ tel que $\mathbf{P}(Y = y) > 0$. La **loi conditionnelle** de X sachant $(Y = y)$ est la donnée de $\mathbf{P}_{(Y=y)}(X = x)$ pour $x \in X(\Omega)$.*

3.2 Indépendance de deux variables aléatoires

Définition 21. *Indépendance de deux variables aléatoires.*
*On dit que X et Y sont des variables aléatoires **indépendantes** si :*

$$\forall (x, y) \in X(\Omega) \times Y(\Omega), \quad \mathbf{P}\left((X = x) \cap (Y = y)\right) = \mathbf{P}(X = x) \times \mathbf{P}(Y = y).$$

Proposition 21. *On suppose que les variables aléatoires X et Y sont indépendantes. Alors :*

(i) *pour tout $(A, B) \in \mathscr{P}(X(\Omega)) \times \mathscr{P}(X(\Omega))$, on a*

$$\mathbf{P}\left((X, Y) \in A \times B\right) = \mathbf{P}(X \in A) \times \mathbf{P}(Y \in B).$$

(ii) *pour toutes applications $f, g : \mathbf{R} \longrightarrow \mathbf{R}$, les variables aléatoires $f(X)$ et $g(Y)$ sont indépendantes (lemme des coalitions).*

Démonstration. Admis.

□

3.3 Indépendance mutuelle de variables aléatoires

On fixe $n \in \mathbf{N}^\star$ et on se donne un n-uplet de variables aléatoires $(X_1, \ldots, X_n)$.

Définition 22. *Indépendance mutuelle.*
On dit que les variables aléatoires $X_1, \ldots, X_n$ sont ***mutuellement indépendantes*** *si*

$$\forall (x_1, \ldots, x_n) \in X_1(\Omega) \times \cdots \times X_n(\Omega), \quad \mathbf{P}\left(\bigcap_{i=1}^{n} (X_i = x_i)\right) = \prod_{i=1}^{n} \mathbf{P}(X_i = x_i).$$

Proposition 22. *On suppose que $X_1, \ldots, X_n$ sont mutuellement indépendantes. Alors :*
(i) *pour tout $(A_1, \ldots, A_n) \in \mathscr{P}(X_1(\Omega)) \times \cdots \times \mathscr{P}(X_n(\Omega))$, on a*

$$\mathbf{P}((X_1, \ldots, X_n) \in A_1 \times \cdots \times A_n) = \mathbf{P}(X_1 \in A_1) \times \cdots \times \mathbf{P}(X_n \in A_n).$$

(ii) *pour toutes applications $f : \mathbf{R}^p \longrightarrow \mathbf{R}$ et $g : \mathbf{R}^{n-p} \longrightarrow \mathbf{R}$, les variables aléatoires $f(X_1, \ldots, X_p)$ et $g(X_{p+1}, \ldots, X_n)$ sont indépendantes (lemme des coalitions).*

Démonstration. Admis.

□

Proposition 23. *Lien entre la loi binomiale et la loi de Bernoulli.*
Soient $X_1, \ldots, X_n$ des variables aléatoires mutuellement indépendantes qui suivent une loi de Bernoulli de paramètre $p \in [0, 1]$. Alors $X_1 + \cdots + X_n \hookrightarrow \mathscr{B}(n, p)$.

Démonstration. On fait une preuve par récurrence sur n.

Lorsque $n = 1$, le résultat est clair.

On suppose que pour tout n-uplet $(X_1, \ldots, X_n)$ de variables aléatoires mutuellement indépendantes qui suivent une loi de Bernoulli de paramètre p, $X_1 + \cdots + X_n \hookrightarrow \mathscr{B}(n, p)$.

Soient $X_1, \ldots, X_n, X_{n+1}$ $(n+1)$ variables aléatoires mutuellement indépendantes qui suivent toutes une loi de Bernoulli de paramètre p.

L'hypothèse de récurrence assure que $X_1 + \cdots + X_n$ suit une loi binomiale $\mathscr{B}(n, p)$. La proposition 22 assure que les variables aléatoires $X_1 + \cdots + X_n$ et X_{n+1} sont indépendantes.

On pose $U = X_1 + \cdots + X_n$ et $S = X_1 + \cdots + X_n + X_{n+1}$. Comme $U(\Omega) = [\![0, n]\!]$ et $X_{n+1}(\Omega) = \{0, 1\}$, on a $S(\Omega) = [\![0, n+1]\!]$. Soit $k \in [\![0, n+1]\!]$.

(a) Si $k = 0$. Alors $(S = k) = (U = 0) \cap (X_{n+1} = 0)$, ainsi par indépendance,

$$\mathbf{P}(S = k) = \mathbf{P}(U = 0) \times \mathbf{P}(X_{n+1} = 0) = \binom{n}{0} p^0 (1-p)^n \times (1-p) = \binom{n+1}{0} p^0 (1-p)^{n+1}.$$

(b) Si $k \in [\![1, n+1]\!]$. La formule des probabilités totales utilisée avec le système complet d'événements $(X_{n+1} = 0, X_{n+1} = 1)$ donne

$$\begin{aligned}
\mathbf{P}(S = k) &= \mathbf{P}((U = k) \cap (X_{n+1} = 0)) + \mathbf{P}((U = k) \cap (X_{n+1} = 1)) \\
&= \mathbf{P}((U = k) \cap (X_{n+1} = 0)) + \mathbf{P}((U = k-1) \cap (X_{n+1} = 1)) \\
&= \mathbf{P}(U = k) \times \mathbf{P}(X_{n+1} = 0) + \mathbf{P}(U = k-1) \times \mathbf{P}(X_{n+1} = 1) \quad \text{par indépendance} \\
&= \binom{n}{k} p^k (1-p)^{n-k} \times (1-p) + \binom{n}{k-1} p^{k-1} (1-p)^{n-(k-1)} \times p \quad \text{hypothèse de récurrence} \\
&= \left(\binom{n}{k} + \binom{n}{k-1}\right) p^k (1-p)^{n+1-k} \\
&= \binom{n+1}{k} p^k (1-p)^{n+1-k} \quad \text{formule de Pascal.}
\end{aligned}$$

□

3.4 Covariance de deux variables aléatoires

Proposition 24. *Espérance d'un produit de variables aléatoires.*
On a

$$\mathbf{E}(XY) = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} xy\, \mathbf{P}((X = x) \cap (Y = y)).$$

Démonstration. Soit $Z = XY$.
On note $Z(\Omega) = \{z_1, \ldots, z_n\}$ et pour tout $i \in [\![1, n]\!]$, $\mathscr{N}_i = \{(x, y) \in X(\Omega) \times Y(\Omega),\ xy = z_i\}$. Par définition, on a

$$\mathbf{E}(Z) = \sum_{i=1}^{n} z_i\, \mathbf{P}(Z = z_i) = \sum_{i=1}^{n} z_i \sum_{(x,y) \in \mathscr{N}_i} \mathbf{P}((X = x) \cap (Y = y)) = \sum_{i=1}^{n} \sum_{(x,y) \in \mathscr{N}_i} xy\, \mathbf{P}((X = x) \cap (Y = y)).$$

Comme la famille $(\mathscr{N}_1, \ldots, \mathscr{N}_n)$ est une partition de $\boldsymbol{\Omega}$, on en déduit que

$$\mathbf{E}(XY) = \mathbf{E}(Z) = \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} xy\, \mathbf{P}((X = x) \cap (Y = y)).$$

□

Définition 23. *Covariance du couple (X, Y).*
On définit la ***covariance*** *du couple (X, Y), notée $\operatorname{Cov}(X, Y)$ par :*

$$\operatorname{Cov}(X, Y) = \mathbf{E}((X - \mathbf{E}(X))(Y - \mathbf{E}(Y))).$$

Remarque 12. On peut interpréter la covariance de la façon suivante :

(i) Si $\operatorname{Cov}(X, Y) > 0$, alors Y a tendance à augmenter lorsque X augmente ;
(ii) Si $\operatorname{Cov}(X, Y) < 0$, alors Y a tendance à diminuer lorsque X augmente ;
(iii) Si $\operatorname{Cov}(X, Y) = 0$, il n'y pas de lien entre X et Y.

Proposition 25. *Formule de König-Huygens.*
On a

$$\operatorname{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y).$$

Démonstration. Par définition, on a

$$\begin{aligned}
\operatorname{Cov}(X, Y) &= \mathbf{E}((X - \mathbf{E}(X))(Y - \mathbf{E}(Y))) \\
&= \mathbf{E}(XY - \mathbf{E}(X)Y - \mathbf{E}(Y)X + \mathbf{E}(X)\mathbf{E}(Y)) \\
&= \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y) \quad \text{linéarité de l'espérance}
\end{aligned}$$

□

Proposition 26. *Lien variance/covariance.*
Soit $(a, b) \in \mathbf{R}^2$. On a

$$\mathbf{V}(aX \pm bY) = a^2\mathbf{V}(X) \pm 2ab\operatorname{Cov}(X, Y) + b^2\mathbf{V}(Y).$$

Remarque 13. Cette formule est plus facile à retenir qu'elle n'y paraît. Il faut se convaincre que c'est une identité remarquable. Le terme quadratique étant la variance et le double produit, la covariance.

Démonstration. Il suffit d'utiliser la définition de la variance et la linéarité de l'espérance.

□

Proposition 27. *Formules pour des variables aléatoires indépendantes.*
On suppose que les variables aléatoires sont indépendantes. On a :

(i) $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$ *;*
(ii) $\operatorname{Cov}(X, Y) = 0$ *;*

(iii) $\mathbf{V}(X+Y) = \mathbf{V}(X) + \mathrm{V}(Y)$.

Démonstration. L'argument important est à chaque fois la proposition 26.

(i) D'après la proposition 24, on a

$$\mathbf{E}(XY) = \sum_{x\in X(\Omega)} \sum_{y\in Y(\Omega)} xy\,\mathbf{P}\left((X=x)\cap(Y=y)\right).$$

En utilisant l'indépendance de X et Y (définition 21), on a

$$\begin{aligned}\mathbf{E}(XY) &= \sum_{x\in X(\Omega)} \sum_{y\in Y(\Omega)} xy\,\mathbf{P}(X=x)\times\mathbf{P}(Y=y)\\ &= \sum_{x\in X(\Omega)} x\,\mathbf{P}(X=x) \sum_{y\in Y(\Omega)} y\times\mathbf{P}(Y=y)\\ &= \mathbf{E}(X)\,\mathbf{E}(Y).\end{aligned}$$

(ii) C'est immédiat à l'aide de la proposition 25.

(iii) On utilise la proposition 26 avec $a=b=1$. Comme $\mathrm{Cov}(X,Y)=0$ (point (ii)), on a donc

$$\mathbf{V}(X+Y) = \mathbf{V}(X) + \mathbf{V}(Y).$$

□

⚠Les réciproques sont fausses. Par exemple, soient X et Y sont deux variables aléatoires indépendantes telles que $X(\Omega)=\{0,1\}$ et $Y(\Omega)=\{-1,1\}$ avec

$$\mathbf{P}(X=0) = \mathbf{P}(X=1) = \mathbf{P}(Y=-1) = \mathbf{P}(Y=1) = \frac{1}{2}.$$

Soit $Z=XY$. Les variables aléatoires X et Z ne sont pas indépendantes car

$$\mathbf{P}\left((X=1)\cap(Z=1)\right) \neq \mathbf{P}(X=1)\,\mathbf{P}(Z=1)$$

et $\mathbf{E}(XZ) = \mathbf{E}(X)\,\mathbf{E}(Z)$.

3.5 Coefficient de corrélation linéaire

3.5.1 Généralités

Définition 24. *Coefficient de corrélation linéaire.*

*On suppose que les variables aléatoires ne sont pas constantes. On définit le **coefficient de corrélation linéaire**, noté $\rho(X,Y)$, par :*

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\,\sigma(Y)}.$$

Proposition 28. *Propriétés du coefficient de corrélation linéaire.*

On suppose que les variables aléatoires X et Y ne sont pas constantes. On a :

(i) *si X et Y sont indépendantes, alors $\rho(X,Y)=0$;*

(ii) $|\rho(X,Y)| \leq 1$ *;*

(iii) $|\rho(X,Y)| = 1$ *si, et seulement si, il existe $(a,b)\in\mathbf{R}^\star\times\mathbf{R}$ tel que $Y=aX+b$.*

Démonstration. (i) C'est clair d'après la proposition 27.

(ii) On reverra cette preuve lorsque l'on prouvera l'inégalité de Cauchy-Schwarz un peu plus tard dans l'année. Soit $P : t \longmapsto \mathbf{V}(tX+Y)$. D'après la proposition 26, on a

$$\forall t\in\mathbf{R},\quad P(t) = \mathbf{V}(X)\,t^2 + 2t\mathrm{Cov}(X,Y) + \mathbf{V}(Y).$$

Comme $\mathbf{V}(X) \neq 0$ (car X n'est pas constante), P est une fonction polynomiale de degré 2, de signe positivf (par définition et car la variance est positive), ainsi son discriminant est négatif, soit

$$4\text{Cov}(X,Y)^2 - 4\mathbf{V}(X)\mathbf{V}(Y) \leq 0 \iff \text{Cov}(X,Y)^2 \leq \mathbf{V}(X)\mathbf{V}(Y).$$

En composant par la fonction $\sqrt{\cdot}$ croissante sur $\mathbf{R}_+$ et comme $\sigma(X)\sigma(Y) \neq 0$ (car X et Y ne sont pas constantes), on en déduit que

$$|\rho(X,Y)| \leq 1.$$

(iii) La preuve est en deux temps.

$\boxed{\Longleftarrow}$ De simples calculs montrent que (utiliser les définitions 13, 14 et 23)

$$\text{Cov}(X, aX+b) = a\text{Cov}(X,X) = |a|\,\mathbf{V}(X) \quad \text{et} \quad \sigma(aX+b) = |a|\,\sigma(X).$$

On en déduit que

$$|\rho(X, aX+b)| = \left|\frac{\text{Cov}(X, aX+b)}{\sigma(X)\,\sigma(aX+b)}\right| = \left|\frac{|a|\,\mathbf{V}(X)}{|a|\,\sigma(X)^2}\right| = 1$$

car $\mathbf{V}(X) = \sigma(X)^2$.

$\boxed{\Longrightarrow}$ Si $|\rho(X,Y)| = 1$, alors le discriminant de la fonction polynomiale de degré 2 introduite ci-dessus vaut 0. En particulier, il existe (unique mais c'est inutile) $a \in \mathbf{R}$ tel que $P(a) = 0$. Ainsi $\mathbf{V}(aX+Y) = 0$. Il s'ensuit que la variable aléatoire $aX+Y$ est constante. Notons b cette constante. On a donc $Y = -aX + b$.

□

3.5.2 Corrélation n'est pas causalité !

On termine ce cours avec une note plus légère ! Il ne faut pas confondre corrélation et causalité, ce ne sont pas les mêmes notions. Quelques exemples où les deux notions sont bien distinctes.

1. Une étude anglaise a prouvé que les gens habitant près de pylônes à haute tension étaient significativement plus souvent malades que le reste de la population. Est-ce la faute du courant électrique ? Ce n'est pas évident parce qu'une autre étude a révélé que les habitants sous les pylônes étaient en moyenne plus pauvres ; et on sait les liens santé-pauvreté... À elle seule, cette étude ne permet pas de conclure.
2. Les assurances ont établi que 50% des accidents arrivaient sur un trajet de moins de 30 km. On en a conclu, un peu vite, que l'habitude des courts trajets pour aller travailler favorisait le manque d'attention des conducteurs. Il est possible que ce soit vrai, mais la « démonstration » est fausse : la plupart des trajets font moins de 30 km !
3. Le conseil de l'Ordre des médecins a publié une étude prouvant que ceux qui pratiquaient régulièrement le jogging à l'âge de 60 ans avaient une probabilité de se trouver en bonne santé à l'âge de 70 ans plus grande que la population normale. Conclusion de l'Ordre, le jogging est une bonne pratique. Il est encore possible que ce soit vrai, mais ce n'est pas une démonstration : la population qui pratique le jogging à 60 ans concentre ceux qui sont déjà en bonne santé. On a donc seulement prouvé que ceux qui sont en bonne santé à 60 ans ont plus de chance de l'être encore 10 ans plus tard.

4 Compléments

Ces compléments sont des notions, qui sans être au programme, peuvent servir (ou pas !) aux concours.

4.1 Loi hypergéométrique

Définition 25. *Définition de la loi hypergéométrique.*

On tire simultanément n boules dans une urne contenant pA boules gagnantes et qA boules perdantes (avec $q = 1-p$, soit un nombre total de boules valant $pA + qA = A$). On compte alors le nombre de boules gagnantes extraites et on appelle X la variable aléatoire donnant ce nombre.

On dit que X suit la ***loi hypergéométrique*** *de paramètres n, p et A et on note $X \hookrightarrow \mathscr{H}(n,p,A)$.*

On peut préciser le support et calculer les probabilités associées.

Proposition 29. *Loi hypergéométrique.*
Soit $X \hookrightarrow \mathscr{H}(n,p,A)$.
On a $X(\Omega) = [\![\max\{0, n-qA\}, \min\{n, pA\}]\!]$ *et*

$$\forall k \in [\![\max\{0, n-qA\}, \min\{n, pA\}]\!], \quad \mathbf{P}(X=k) = \frac{\binom{pA}{k}\binom{qA}{n-k}}{\binom{A}{n}}.$$

Démonstration. On sépare l'étude du support et le calcul des probabilités.

Support. La valeur minimale de X est atteinte lorsque que l'on pioche un maximum de boules perdantes.
Ainsi, si $qA \geq n$, le minimum est 0, alors que si $n > qA$, le minimum est qA. Dans tout les cas, le minimum est $\max\{0, n-qA\}$.
La valeur maximale de X est atteinte lorsque que l'on pioche un maximum de boules gagnantes.
Ainsi, si $pA \geq n$, le maximum est pA, alors que si $n > pA$, le maximum est pA. Dans tout les cas, le minimum est $\in \{n, pA\}$.

Probabilités. Soit $k \in [\![\max\{0, n-qA\}, \min\{n, pA\}]\!]$.
Comme on pioche simultanément n boules parmi A boules, le nombre total de tirages est $\binom{A}{n}$.
L'événement $(X=k)$ est l'événement « obtenir k boules gagnantes et $n-k$ boules perdantes », ce que l'on peut faire de $\binom{pA}{k}\binom{qA}{n-k}$ façons, ainsi $\mathbf{P}(X=k) = \dfrac{\binom{pA}{k}\binom{qA}{n-k}}{\binom{A}{n}}$.

□

Corollaire 2. *Formule de Vandermonde.*
Soient $p \in]0,1[$, $A \in \mathbf{N}$ *tel que* $pA \in \mathbf{N}$. *Soit aussi* $n \in \mathbf{N}$ *avec* $n \leq A$. *On a*

$$\sum_{k=\max\{0,n-qA\}}^{\min\{n,pA\}} \binom{pA}{k}\binom{(1-p)A}{n-k} = \binom{A}{n}.$$

Démonstration. Soit $X \hookrightarrow \mathscr{H}(n,p,A)$. On a $\displaystyle\sum_{x \in X(\Omega)} \mathbf{P}(X=x) = 1$, ce qui donne la relation souhaitée en utilisant la proposition 29.

□

Proposition 30. *Espérance d'une loi hypergéométrique.*
Soit $X \hookrightarrow \mathscr{H}(n,p,A)$. *On a* $\mathbf{E}(X) = np$.

Démonstration. Numérotons de 1 à pA les boules de type « réussite » et définissons pour tout $k \in [\![1, pA]\!]$ l'événement E_k : « on a tiré parmi les n boules la boule « réussite » numéro k ».

Comme le nombre de boules « réussites » est $X = \displaystyle\sum_{k=1}^{pA} \mathbf{1}_{E_k}$ ($\mathbf{1}_{E_k}$ est la fonction indicatrice de E_k), par linéarité de l'espérance, on a

$$\mathbf{E}(X) = \mathbf{E}\left(\sum_{k=1}^{pA} \mathbf{1}_{E_k}\right) = \sum_{k=1}^{pA} \mathbf{E}\left(\mathbf{1}_{E_k}\right) = \sum_{k=1}^{pA} \mathbf{P}(E_k) = pA\,\mathbf{P}(E_1)$$

car tous les événements E_k ont la même probabilité.

Pour conclure, il suffit de calculer $\mathbf{P}(E_1)$. On calcule $\mathbf{P}\left(\overline{E_1}\right)$. Ne pas tirer la boule « réussite » numéro 1 revient à piocher les n boules parmi les $A-1$ disponibles, soit

$$\mathbf{P}\left(\overline{E_1}\right) = \frac{\binom{A-1}{n}}{\binom{A}{n}} = \frac{A-n}{A}.$$

On en déduit que $\mathbf{P}(E_1) = \dfrac{n}{A}$, puis que $\mathbf{E}(X) = np$.

□

4.2 Polynômes de Bernstein et théorème de Weierstrass

Le but de cette sous-partie est d'établir le théorème de Weierstrass qui assure que toute fonction continue sur $[0,1]$ peut être « approchée » par une fonction polynomiale.

Proposition 31. *Théorème de Weierstrass polynomiale.*

Soit $f : [0,1] \longrightarrow \mathbf{R}$ une fonction continue. Soit $\varepsilon > 0$. Il existe une fonction polynomiale $P : [0,1] \longrightarrow \mathbf{R}$ telle que

$$\forall x \in [0,1], \quad |f(x) - P(x)| \leq \varepsilon.$$

Démonstration. Pour tout $n \in \mathbf{N}^\star$ et pour tout $x \in [0,1]$, on définit la variable aléatoire $X_{n,x}$. On suppose que $X_{n,x}$ suit une loi binomiale de paramètres n et x. Soit $f : [0,1] \longrightarrow \mathbf{R}$ une fonction continue.

Pour tout $n \in \mathbf{N}^\star$, on définit $Y_{n,x} = f\left(\dfrac{X_{n,x}}{n}\right)$. La formule du transfert donne

$$\mathbf{E}(Y_{n,x}) = \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

On définit ainsi sur $[0,1]$ la fonction $B_n(f)$ par

$$\forall x \in [0,1], \quad B_n(f)(x) = \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

Soit $\varepsilon > 0$. On admet qu'il existe $\alpha > 0$ tel que

$$\forall (x,y) \in [0,1]^2, \quad (|x-y| < \alpha) \Longrightarrow \left(|f(x) - f(y)| < \frac{\varepsilon}{2}\right).$$

Cette propriété porte le nom d'**uniforme continuité**. Soit $x \in [0,1]$, on définit les ensembles suivants :

$$J_{1,x} = \left\{k \in [\![0,n]\!],\ \left|f\left(\frac{k}{n}\right) - f(x)\right| < \frac{\varepsilon}{2}\right\} \quad \text{et} \quad J_{2,x} = \left\{k \in [\![0,n]\!],\ \left|f\left(\frac{k}{n}\right) - f(x)\right| \geq \frac{\varepsilon}{2}\right\}.$$

Comme f est continue sur le segment $[0,1]$, elle est bornée : on note $\|f\|_\infty = \sup\limits_{x \in [0,1]} |f(x)|$. On a

$$\begin{aligned} |B_n(f)(x) - f(x)| &= \left|\sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k} - \sum_{k=0}^{n} f(x) \binom{n}{k} x^k (1-x)^{n-k}\right| \\ &\leq \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \left|f(x) - f\left(\frac{k}{n}\right)\right|. \end{aligned}$$

Comme $J_{1,x}$ et $J_{2,x}$ forment une partition de $[\![0,n]\!]$, on en déduit que

$$|B_n(f)(x) - f(x)| \leq \sum_{k \in J_{1,x}} \binom{n}{k} x^k (1-x)^{n-k} \left|f(x) - f\left(\frac{k}{n}\right)\right| + \sum_{k \in J_{2,x}} \binom{n}{k} x^k (1-x)^{n-k} \left|f(x) - f\left(\frac{k}{n}\right)\right|.$$

Par définition de $J_{1,x}$, on a

$$\begin{aligned}\sum_{k\in J_{1,x}}\binom{n}{k}x^k(1-x)^{n-k}\left|f(x)-f\left(\frac{k}{n}\right)\right| &\leq \frac{\varepsilon}{2}\sum_{k\in J_{1,x}}\binom{n}{k}x^k(1-x)^{n-k}\\ &\leq \frac{\varepsilon}{2}\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}\\ &\leq \frac{\varepsilon}{2}.\end{aligned}$$

De plus, par utilisation de l'inégalité triangulaire et par définition de $\|f\|_\infty^{[0,1]}$, on a

$$\begin{aligned}\sum_{k\in J_{2,x}}\binom{n}{k}x^k(1-x)^{n-k}\left|f(x)-f\left(\frac{k}{n}\right)\right| &\leq 2\|f\|_\infty^{[0,1]}\sum_{k\in J_{2,x}}\binom{n}{k}x^k(1-x)^{n-k}\\ &\leq 2\|f\|_\infty^{[0,1]}\sum_{k\in J_{2,x}}\binom{n}{k}x^k(1-x)^{n-k}\\ &\leq 2\|f\|_\infty^{[0,1]}\sum_{k\in J_{2,x}}\mathbf{P}(X_{n,x}=k).\end{aligned}$$

Aussi

$$X_{n,x}\in J_{2,x} \iff \left|f\left(\frac{X_{n,}}{n}\right)-f(x)\right|\geq\frac{\varepsilon}{2} \iff |Y_{n,x}-f(x)|\geq\frac{\varepsilon}{2}.$$

On en déduit que

$$\mathbf{P}(X_{n,x}\in J_{2,x})=\sum_{k\in J_{2,x}}\mathbf{P}(X_{n,x}=k)=\mathbf{P}\left(|Y_{n,x}-f(x)|\geq\frac{\varepsilon}{2}\right).$$

Il en résulte que

$$|B_n(f)(x)-f(x)|\leq\frac{\varepsilon}{2}+2\|f\|_\infty^{[0,1]}\mathbf{P}\left(|Y_{n,x}-f(x)|\geq\frac{\varepsilon}{2}\right). \tag{1}$$

D'après la définition de α, pour tout $(x,y)\in[0,1]^2$,

$$\left(|f(x)-f(y)|\geq\frac{\varepsilon}{2}\right)\Longrightarrow(|x-y|\geq\alpha).$$

Comme

$$\left(|Y_{n,x}-f(x)|\geq\frac{\varepsilon}{2}\right)=\left(\left|f\left(\frac{X_{n,x}}{n}\right)-f(x)\right|\geq\frac{\varepsilon}{2}\right),$$

on récupère

$$\left(|Y_{n,x}-f(x)|\geq\frac{\varepsilon}{2}\right)\subset\left(\left|\frac{X_{n,x}}{n}-x\right|\geq\alpha\right).$$

En utilisant le fait que

$$\left(\left|\frac{X_{n,x}}{n}-x\right|\geq\alpha\right)=(|X_{n,x}-nx|\geq n\alpha),$$

on a montré que

$$\left(|Y_{n,x}-f(x)|\geq\frac{\varepsilon}{2}\right)\subset(|X_{n,x}-nx|\geq n\alpha).$$

On en déduit

$$\mathbf{P}\left(|Y_{n,x}-f(x)|\geq\frac{\varepsilon}{2}\right)\leq\mathbf{P}(|X_{n,x}-nx|\geq n\alpha),$$

puis en utilisant l'inégalité 1, on a

$$|B_n(f)(x)-f(x)|\leq\frac{\varepsilon}{2}+2\|f\|_\infty^{[0,1]}\mathbf{P}(|X_{n,x}-nx|\geq n\alpha). \tag{2}$$

On utilise l'inégalité de Bienaymé-Tchebyshev avec la variable aléatoire $X_{n,x}$ qui admet une variance valant $nx(1-x)$. On a donc :

$$\begin{aligned}\mathbf{P}\left(|X_{n,x}-nx| \geq n\alpha\right) &= \mathbf{P}\left(|X_{n,x}-\mathbf{E}\left(X_{n,x}\right)| \geq n\alpha\right)\\ &\leq \frac{nx(1-x)}{n^2\alpha^2}\\ &\leq \frac{x(1-x)}{n\alpha^2}.\end{aligned}$$

Une simple étude de fonction montre que pour tout $x \in [0,1]$, $x(1-x) \leq \frac{1}{4}$. On en déduit, en utilisant l'inégalité (2)

$$|B_n(f)(x) - f(x)| \leq \frac{\varepsilon}{2} + \frac{\|f\|_\infty^{[0,1]}}{2n\alpha^2}.$$

Comme $\lim\limits_{n\to+\infty} \frac{\|f\|_\infty^{[0,1]}}{2n\alpha^2} = 0$, il existe un entier naturel N tel que pour tout $n \geq N$,

$$0 \leq \frac{\|f\|_\infty^{[0,1]}}{2n\alpha^2} \leq \frac{\varepsilon}{2}.$$

On en déduit que pour tout $n \geq N$,

$$|B_n(f)(x) - f(x)| \leq \varepsilon.$$

□

4.3 Formule du crible

Le but de cette sous-partie est de montrer la formule du crible.

Proposition 32. *Formule du crible.*
Soit $n \in \mathbf{N}^\star$. Soient $A_1, \ldots, A_n$ des ensembles finis. Alors,

$$\operatorname{card}\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n \left((-1)^{k+1} \sum_{1\leq i_1<i_2<\cdots<i_k\leq n} \operatorname{card}\left(\bigcap_{j=1}^k A_{i_j}\right)\right).$$

Démonstration. Pour $A \subset E$, on introduit la fonction caractéristique χ_A de A définie sur $\mathbf{R}$ par

$$\chi_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}.$$

On commence par remarquer que

$$\sum_{x\in E} \chi_A(x) = \sum_{x\in A} \chi_A(x) + \sum_{x\notin A} \chi_A(x) = \sum_{x\in A} 1 + \sum_{x\notin A} 0 = \operatorname{card}(A). \tag{3}$$

De plus, pour tout $x \in E$, on a

$$1 - \chi_{\bigcup_{i=1}^n A_i}(x) = \chi_{\overline{\bigcup_{i=1}^n A_i}}(x) = \chi_{\bigcap_{i=1}^n \overline{A_i}}(x) = \prod_{i=1}^n \chi_{\overline{A_i}}(x) = \prod_{i=1}^n \left(1 - \chi_{A_i}(x)\right). \tag{4}$$

On prouve par récurrence que pour tout $n \in \mathbf{N}^\star$, la proposition $\mathscr{P}_n$: « pour tout $(A_1, \ldots, A_n) \in \mathscr{P}(E)$, $\prod\limits_{i=1}^n (1-\chi_{A_i}) = 1 + \sum\limits_{k=1}^n \left((-1)^k \prod\limits_{1\leq i_1\leq i_2\leq\cdots\leq i_k\leq n} \chi_{A_{i_1}} \cdots \chi_{A_{i_k}}\right)$ » est vraie.

$\mathscr{P}_1$ est clairement vraie.

On suppose $\mathscr{P}_n$ vraie pour un certain entier naturel non nul n, montrons que $\mathscr{P}_{n+1}$ est vraie.

Soit $(A_1, \dots, A_n, A_{n+1}) \in \mathscr{P}(E)^{n+1}$. En utilisant l'hypothèse de récurrence, on a

$$\begin{aligned}
\prod_{i=1}^{n+1} (1 - \chi_{A_i}) &= (1 - \chi_{A_{n+1}}) \prod_{i=1}^{n} (1 - \chi_{A_i}) \\
&= (1 - \chi_{A_{n+1}}) \times \left(1 + \sum_{k=1}^{n} (-1)^k \prod_{1 \leq i_1 < \cdots < i_k \leq n} \chi_{A_{i_1}} \cdots \chi_{A_{i_k}} \right) \\
&= 1 + \sum_{k=1}^{n+1} \left((-1)^{k+1} \prod_{1 \leq i_1 < \cdots < i_k \leq n+1} \chi_{A_{i_1}} \cdots \chi_{A_{i_{k+1}}} \right).
\end{aligned}$$

On a montré que $\mathscr{P}_{n+1}$ est vraie, ce qui termine la récurrence.

La précédente récurrence et la ligne (4) montrent que

$$\chi_{\bigcup\limits_{i=1}^{n} A_i} = \sum_{k=1}^{n} (-1)^{k+1} \left(\prod_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} \chi_{A_{i_1}} \cdots \chi_{A_{i_k}} \right).$$

On en déduit : pour tout $x \in E$,

$$\begin{aligned}
\sum_{x \in E} \chi_{\bigcup\limits_{i=1}^{n} A_i}(x) &= \sum_{k=1}^{n} (-1)^{k+1} \sum_{x \in E} \prod_{1 \leq i_1 < \cdots < i_k \leq n} \chi_{A_{i_1}}(x) \dots \chi_{A_{i_k}}(x) \\
&= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} \sum_{x \in E} \chi_{A_{i_1} \cap \dots \cap A_{i_k}}(x).
\end{aligned}$$

En utilisant le résultat de la ligne (3), on obtient finalement

$$\operatorname{card}\left(\bigcup_{i=1}^{n} A_i \right) = \sum_{k=1}^{n} \left((-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} \operatorname{card}\left(\bigcap_{j=1}^{k} A_{i_j} \right) \right).$$

□