CHAPITRE I

Introduction à la théorie des probabilités

1. Rappels sur les espaces probabilisés

Les éléments ci-dessous ne sont que brièvement rappelés car ils sont normalement traités dans un cours d'intégration.

Définition 1. *Soit E un ensemble. On dit que $\mathcal{A} \subset \mathcal{P}(E)$ est une* tribu *(ou une σ-algèbre) si :*

- $E \in \mathcal{A}$ *;*
- $\forall A \in \mathcal{A}, A^c \in \mathcal{A}$ *;*
- $\forall n \geq 0, (\ A_n \in \mathcal{A} \Rightarrow \bigcup_{n\geq 0} A_n \in \mathcal{A})$*.*

Les éléments de $\mathcal{A}$ sont appelés *parties mesurables* (ou $\mathcal{A}$-mesurables). On dit alors que $(E, \mathcal{A})$ est un *espace mesurable.*

Il est facile de vérifier que l'ensemble vide appartient toujours à une tribu, et qu'une tribu est stable par intersection dénombrable.

Exemples de tribus

- $\mathcal{P}(E)$, $\{\emptyset, E\}$;
- $\mathcal{C} \subset \mathcal{P}(E)$ alors la plus petite tribu contenant $\mathcal{C}$, appelée *tribu engendrée par* $\mathcal{C}$, est $\sigma(\mathcal{C}) = \bigcap_{\mathcal{A} \text{ tribu },\ \mathcal{C}\subset\mathcal{A}} \mathcal{A}$;
- La *tribu des boréliens* $\mathcal{B}(E)$ est la tribu engendrée par les ouverts d'un espace topologique ; si $E = \mathbb{R}$, $\mathcal{B}(\mathbb{R})$ est la tribu engendrée par les $]a, b[$ ou $]-\infty, a[$, pour $a \in \mathbb{R}$ (ou $\mathbb{Q}$).

En probabilité on utilise un vocabulaire particulier :

- l'ensemble mesurable E, noté en général Ω, est appelé espace des réalisations ou univers ;
- un élément ω de Ω est appelé une réalisation ;
- un élément A de $\mathcal{A}$ est appelé événement ;
- $\emptyset$ est l'événement impossible ;
- $A \cap B$ est l'événement « A et B » ; $A \cup B$ est l'événement « A ou B ».

Définition 2. *On appelle* mesure positive *sur un espace mesurable* $(E, \mathcal{A})$ *une application* $\mu : \mathcal{A} \to [0, +\infty]$ *telle que :*

- $\mu(\emptyset) = 0$
- *pour toute famille* $(A_n)_{n\geq 0}$ *de parties mesurables disjointes :*

$$\mu\left(\bigcup_{n\geq 0} A_n\right) = \sum_{n\geq 0} \mu(A_n)$$

On dit que μ est finie si $\mu(E) < +\infty$, et que μ est σ-finie s'il existe une suite croissante de parties mesurées $(A_n)_{n\geq 0}$ telle que $\bigcup_{n\geq 0} A_n = E$ et $\mu(A_n) < +\infty$ pour tout $n \geq 0$.

Un élément x de E est un *atome* si $\mu(\{x\}) > 0$. Une mesure est *diffuse* si elle n'a pas d'atome.

On appelle *mesure de Dirac* en x_0 la mesure δ_{x_0} telle que $\delta(A) = 1$ si A contient x_0, 0 sinon.
On appelle *mesure de Lebesgue*, l'unique mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que

$$\lambda(]a, b[) = b - a.$$

Remarque : il est possible de montrer, en utilisant l'axiome du choix, qu'on ne peut pas étendre la mesure de Lebesgue à toutes les parties de $\mathbb{R}$.)

Définition 3. *Une mesure positive est une* probabilité *si* $\mu(E) = 1$.

On dit que $(\Omega, \mathcal{A}, \mathbb{P})$ est un *espace probabilisé* si $(\Omega, \mathcal{A})$ est un espace mesurable sur lequel on a défini une probabilité $\mathbb{P}$. Une propriété qui est vraie sur une partie de Ω de probabilité 1 est dite *vraie presque sûrement (p.s).*

Exemples

- On lance un dé (non pipé) :

$$\Omega = \{1, 2, \dots, 6\},\ \mathcal{A} = \mathcal{P}(\Omega), \mathbb{P}_1(A) = \frac{\text{Card}(A)}{6}$$

La probabilité choisie rend tous les tirages possibles équiprobables. On peut également noter $\mathbb{P}_1 = \sum_{i=1}^{6} \frac{1}{6}\delta_i$.

- On lance un dé (non pipé) deux fois :

$$\Omega = \{1, 2, \dots, 6\}^2,\ \mathcal{A} = \mathcal{P}(\Omega), \mathbb{P}(A) = \frac{\text{Card}(A)}{36}$$

La probabilité choisie rend là encore tous les tirages possibles équiprobables. C'est en fait le produit $\mathbb{P}_1 \otimes \mathbb{P}_1$ car $\mathbb{P}(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_1(A_2)$.

- On lance un dé jusqu'à obtenir 6. L'expérience définie ici peut amener à une infinité de lancers, le plus simple est de ne considérer que des suites infinies et de poser :

$$\Omega = \{1, 2, \dots, 6\}^{\mathbb{N}^*}.$$

Les éléments de Ω sont donc les suites $\omega = (\omega_1, \omega_2, \dots)$ qui représentent les tirages successifs. La tribu $\mathcal{A}$ sur Ω est la plus petite tribu qui rend mesurable tous les ensembles :

$$\{\omega\ :\ \omega_1 = i_1, \omega_2 = i_2, \dots, \omega_n = i_n\}$$

où $n \geq 1$ et $i_1, \ldots, i_n \in \{1, \ldots, 6\}$, et $\mathbb{P}$ est l'unique probabilité sur Ω telle que

$$\mathbb{P}(\{\omega : \omega_1 = i_1, \omega_2 = i_2, \ldots, \omega_n = i_n\}) = \left(\frac{1}{6}\right)^n .$$

Propriétés

Soit $\mathbb{P}$ une probabilité sur $(\Omega, \mathcal{A})$, soient A, B, $(A_n)_{n\in\mathbb{N}}$ et $(B_n)_{n\in\mathbb{N}}$ des éléments de $\mathcal{A}$. On peut vérifier les propriétés suivantes :

- $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- Formule de Poincaré
$$\mathbb{P}\Big(\bigcup_{1\leq j\leq n} A_j\Big) = \sum_{1\leq k\leq n} (-1)^{k-1}\Big(\sum_{1\leq i_1<\cdots<i_k\leq n} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k})\Big),$$
- Si la suite (A_n) est croissante , $\mathbb{P}(\bigcup_{n\geq 0} A_n) = \lim_{n\to+\infty} \mathbb{P}(A_n)$
- Si la suite (B_n) est décroissante , $\mathbb{P}(\bigcap_{n\geq 0} B_n) = \lim_{n\to+\infty} \mathbb{P}(B_n)$
- $\mathbb{P}(\bigcup_{n\geq 0} A_n) \leq \sum_{n\geq 0} \mathbb{P}(A_n)$

Nous rappelons ci-dessous le lemme de classe monotone, normalement vu en cours d'intégration, et son corollaire, outil utile en probabilité.

Lemme 4.*[Lemme de classe monotone] On appelle classe monotone une classe contenant Ω, stable par différence et union croissante.*
Soit $\mathcal{C}$ une classe stable par intersection finie. Alors la plus petite classe monotone contenant $\mathcal{C}$ est $\sigma(C)$.

Lemme 5.*[Lemme d'unicité des mesures de probabilité] Si deux mesures de probabilités coïncident sur un ensemble d'événements $\mathcal{C}$ stable par intersection finie, elles coïncident sur la tribu $\sigma(C)$.*
En particulier, si deux mesures de probabilités coïncident sur $\{]-\infty, x], x \in \mathbb{R}\}$, elles coïncident sur $\mathcal{B}(\mathbb{R})$.
Si deux mesures de probabilités coïncident sur $\{A\times B, A \in \mathcal{A}, B \in \mathcal{B}\}$, elles coïncident sur la tribu produit $\mathcal{A} \otimes \mathcal{B}$.

En effet, l'ensemble des événements sur lesquelles coïncident deux lois de probabilités est une classe monotone.

2. Probabilités discrètes

C'est un cas particulier important qui consiste à considérer un univers Ω dénombrable. On prend alors généralement $\mathcal{A} = \mathcal{P}(\Omega)$. Pour caractériser une probabilité sur $(\Omega, \mathcal{A})$ il suffit alors de se donner les probabilités $\mathbb{P}(\{\omega\})$ pour toute réalisation $\omega \in \Omega$, qui doivent vérifier $\sum \mathbb{P}(\{\omega\}) = 1$. On a alors de façon immédiate que pour tout événement A, $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$.

Ceci s'applique en particulier lorsque Ω est fini et que les ω sont équiprobables, c'est-à-dire que

$$\forall \omega \in \Omega, \mathbb{P}(\{\omega\}) = \frac{1}{\mathrm{Card}(\Omega)}$$

On obtient alors la règle d'équiprobabilité

$$\mathbb{P}(A) = \frac{\text{nombre d'issues favorables à } A}{\text{nombre d'issues possibles}} = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)}$$

Pour calculer explicitement des probabilités dans ce cadre, on utilise souvent des résultats de combinatoire, dont on rappelle les plus classiques.

- Nombre de permutations d'un ensemble à n éléments : $n!$
- Nombre de $p-$uplets dans un ensemble à n éléments : n^p
- Nombre de $p-$uplets d'éléments distincts dans un ensemble à n éléments : $A_n^p = n(n-1)\dots(n-p+1)$
- Nombre de parties d'un ensemble à n éléments : 2^n
- Nombre de parties à p éléments dans un ensemble à n éléments : $\binom{n}{p} = \frac{n}{p!(n-p)!} = \frac{A_n^p}{p!}$.

3. Probabilités conditionnelles, événements indépendants

Dans une expérience réelle, le fait de savoir qu'un événement s'est produit, peut modifier la probabilité qu'on souhaite affecter à un autre événement. Il est donc naturel pour décrire cela de définir une nouvelle probabilité.

Définition 6. *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé et* A *un événement de probabilité non nulle. On définit une nouvelle probabilité sur* $(\Omega, \mathcal{A})$*, appelée* probabilité conditionnelle *sachant* A*, en posant pour tout* $B \in \mathcal{A}$*,*

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

On peut facilement voir que l'on a

$$\forall B \in \mathcal{A}, \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A)$$

Cela se généralise en

$$\mathbb{P}(\cap_{i=1}^n A_i) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2)\dots\mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1})$$

De plus, si $\mathbb{P}(A^c) > 0$ alors

$$\forall B \in \mathcal{A},\ \mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(A^c)\mathbb{P}(B|A^c)$$

et on en déduit la formule de Bayes permettant d'inverser les conditionnements :

$$\text{Si } \mathbb{P}(B) > 0,\ \mathbb{P}(A|B) = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(A^c)\mathbb{P}(B|A^c)}.$$

Il est possible de généraliser ces dernières relations lorsqu'on considère une partition dénombrable $(A_n)_{n\in\mathbb{N}}$ de Ω ($A_k \cap A_l = \emptyset$, $\cup A_n = \Omega$ et $\mathbb{P}(A_n) > 0$) :

$$\forall B \in \mathcal{A},\ \mathbb{P}(B) = \sum_{n\geq 0} \mathbb{P}(A_n)\mathbb{P}(B|A_n) \quad \text{(formule des probabilités totales)}$$

$$\mathbb{P}(A_n|B) = \frac{\mathbb{P}(A_n)\mathbb{P}(B|A_n)}{\sum_{n\geq 0}\mathbb{P}(A_n)\mathbb{P}(B|A_n)} \quad \text{(formule de Bayes)}$$

On peut remarquer que lorsque $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ alors $\mathbb{P}(B|A) = \mathbb{P}(B)$, autrement dit le fait de savoir que A est réalisé ne donne pas d'information sur la réalisation ou non de l'événement B ; ceci motive la définition suivante.

Définition 7. *On considère un espace probabilisé* $(\Omega, \mathcal{A}, \mathbb{P})$*. Si* $(A, B) \in \mathcal{A}^2$ *sont deux événements, on dit que* A *et* B *sont* indépendants *si*

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

Plus généralement on a :

Définition 8. *On dit que* n *événements* $A_1, \dots, A_n$ *sont indépendants si, pour tout sous-ensemble non vide* $\{j_1, \dots, j_p\}$ *de* $\{1, \dots, n\}$*, on a*

$$\mathbb{P}(A_{j_1} \cap \dots \cap A_{j_p}) = \mathbb{P}(A_{j_1}) \dots \mathbb{P}(A_{j_p}).$$

Attention ceci n'est pas équivalent à la seule égalité $\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \dots \mathbb{P}(A_n)$ ni à, « pour chaque paire $\{i, j\} \subset \{1, \dots, n\}$ les événements A_i et A_j sont indépendants » (on dit alors que les événements sont *deux-à-deux indépendants*).

Exemple
Considérons l'espace correspondant à deux lancers de pile ou face et prenons

$A = \{\text{pile au premier lancer}\}$, $B = \{\text{pile au second lancer}\}$ et

$C = \{\text{même résultats aux deux lancers}\}$.

Les trois événements sont 2 à 2 indépendants, mais ne sont pas indépendants.

Proposition 9. *Les* n *événements* $A_1, \dots, A_n$ *sont indépendants si et seulement si*

$$\mathbb{P}(B_1 \cap \dots \cap B_n) = \mathbb{P}(B_1) \dots \mathbb{P}(B_n)$$

pour tout choix des $B_i \in \sigma(A_i) = \{\emptyset, \Omega, A_i, A_i^c\}$.

Proposition 10. *Si* A *et* B *sont deux événements indépendants,* A *et* B^c *sont indépendants.*
Les événements de probabilité nulle et de probabilité 1 sont indépendants de tous les autres événements.

4. Lemme de Borel-Cantelli

Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements on note

$$\limsup_{n\to\infty} A_n = \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

De façon équivalente on a

$$\omega \in \limsup_{n\to\infty} A_n \Leftrightarrow \omega \text{ appartient à une infinité d'événements } A_n.$$

On peut rappeler qu'on a aussi la définition $\liminf_{n\to\infty} A_n = \bigcup_{n=0}^{\infty} \bigcap_{k=n}^{\infty} A_k$.

De façon équivalente,

$$\omega \in \liminf_{n\to\infty} A_n \Leftrightarrow \omega \text{ appartient à tous les } A_n \text{ sauf éventuellement un nombre fini.}$$

Enfin, puisqu'une suite à valeurs dans $\{0,1\}$ a pour limite supérieure 1 si et seulement si elle prend une infinité de fois la valeur 1, on a une interprétation simple en terme de fonctions indicatrices :

$$1_{\limsup A_n} = \limsup(1_{A_n}) \quad \text{et } 1_{\liminf A_n} = \liminf(1_{A_n})$$

Le résultat suivant est le *lemme de Borel-Cantelli* (vue son importance cela pourrait être un théorème.)

Lemme 11.*[Borel Cantelli] Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements*

(i) Si $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$ alors

$$\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$$

ou de manière équivalente

$$\{n \in \mathbb{N} : \omega \in A_n\} \text{ est fini p.s.}$$

(ii) Si $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$ et si les événements A_n sont indépendants, alors

$$\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$$

ou de manière équivalente

$$\{n \in \mathbb{N} : \omega \in A_n\} \text{ est infini p.s.}$$

L'hypothèse d'indépendance est nécessaire dans le (ii), comme le montre l'exemple trivial où $A_n = A$ pour tout $n \in \mathbb{N}$, avec $0 < \mathbb{P}(A) < 1$.

Démonstration. (i) Supposons $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$. On a pour tout entier m

$$\limsup_{n\to\infty} A_n = \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} A_k \subset \bigcup_{k=m}^{\infty} A_k$$

et donc $\mathbb{P}(\limsup_{n\to\infty} A_n) \leq \sum_{k\geq m} \mathbb{P}(A_m)$ qui tend vers 0 avec m si la série converge.

(ii) Pour le deuxième point, remarquons d'abord que pour tout n et tout $N \geq n$

$$\begin{aligned}\mathbb{P}(\bigcup_{n\leq k\leq N} A_k) &= 1 - \mathbb{P}(\bigcap_{n\leq k\leq N} A_k^c) \\ &= 1 - \prod_{n\leq k\leq N} (1 - \mathbb{P}(A_k))\end{aligned}$$

Comme $1 - x \leq e^{-x}$ pour tout $x \geq 0$, on en déduit

$$\mathbb{P}\left(\bigcup_{n\leq k\leq N} A_k\right) \geq 1 - \exp\left(-\sum_{n\leq k\leq N} \mathbb{P}(A_k)\right)$$

Lorsque N tend vers l'infini, la somme dans l'exponentielle tend, pour tout n, vers l'infini par hypothèse, et donc $\mathbb{P}(\bigcup_{k\geq n} A_k) = 1$. Il ne reste alors plus qu'à remarquer que $\mathbb{P}(\limsup_{n\to\infty} A_n) = \lim_{n\to\infty} \mathbb{P}(\bigcup_{k\geq n} A_k)$. □

5. Exercices

Exercice 1.

Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements d'un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$. Ecrire les événements suivants à l'aide d'opérations ensemblistes :

E_1=« au moins un des A_n est réalisé » ; E_2=« tous les A_n sont réalisés » ;

E_3 = « à partir de $n = 500$ aucun A_n n'est réalisé » ;

E_4 = « un nombre fini de A_n est réalisé » ;

E_5=« une infinité de A_n sont réalisés » ; E_6=« tous les A_n sauf un nombre fini sont réalisés ».

Exercice 2.

Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements d'un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$.

1. Démontrer la formules de Poincaré :

$$\mathbb{P}\Big(\bigcup_{1\leq j\leq n} A_j\Big) = \sum_{1\leq k\leq n} (-1)^{k-1}\Big(\sum_{1\leq i_1<\cdots<i_k\leq n} \mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k})\Big)$$

2. Démontrer les égalités suivantes :

$$\begin{aligned}\liminf\,(A_n^c) &= (\limsup A_n)^c, \\ \limsup\,(A_n\cup B_n) &= \limsup A_n \cup \limsup B_n.\end{aligned}$$

3. Démontrer les inégalités suivantes :

$$\mathbb{P}(\liminf A_n) \leq \liminf \mathbb{P}(A_n) \leq \limsup \mathbb{P}(A_n) \leq \mathbb{P}(\limsup A_n).$$

4. On suppose que les événements $(A_n)_{n\geq 0}$ sont presque certains, i.e. $\mathbb{P}(A_n) = 1$. Démontrer que $\cap_{n\geq 1} A_n$ est presque certain.

Exercice 3.

Dans un espace probabilisé on suppose qu'on dispose d'une suite $(A_n)_{n\geq 1}$ d'événements indépendants. Démontrer :

$1 - P(\cup_{j=1}^n A_j) \leq \exp\big(- \sum_{j=1}^n P(A_j)\big)$.

Exercice 4.

N personnes déposent leur parapluie à l'entrée d'un restaurant. En partant, chacune récupère un parapluie au hasard : la première choisit le sien de manière uniforme parmi les N parapluies, la deuxième en prend un de façon uniforme parmi les $N-1$ restants, etc.
1. Montrer qu'on peut prendre comme espace de probabilités Ω correspondant à cette situation l'ensemble S_n des permutations de n éléments, muni de la loi uniforme.
2. Quelle est la probabilité que la première personne récupère son parapluie ? Quelle est la probabilité que la ième personne récupère son parapluie ? Ces événements sont-ils indépendants ?
3. Quelle est l'espérance du nombre de personnes qui récupèrent leur propre parapluie ?
4. Déterminer la probabilité p_N qu'au moins une personne récupère son parapluie. Que devient p_N lorsque N tend vers l'infini ?

Exercice 5.

Les gènes se présentent le plus souvent en paire et sous deux formes d'allèles que nous noterons A et B. Cela donne donc trois génotypes possibles : AA, AB et BB. Chaque individu reçoit au hasard un gène de chacun de ses parents. Chaque allèle composant le gène d'un des parents ayant la probabilité $1/2$ d'être transmis à l'enfant. On suppose que les génotypes des deux parents sont indépendants et de même loi : soit x la probabilité d'avoir le génotype AA, $2y$ celle d'avoir le génotype AB, z celle d'avoir le génotype BB (remarque évidente : $x + 2y + z = 1$).

1. Calculer la probabilité de chacun des génotypes pour un enfant.
2. Vérifier que, si $x = y = z = 1/4$, alors les génotypes de l'enfant ont les mêmes probabilités que ceux des parents.
3. Calculer la probabilité de chacun des génotypes pour un enfant de la seconde génération. Que constatez-vous ? La loi de répartition des génotypes s'appelle la loi de Hardy-Weinberg.

Exercice 6.

On teste deux médicaments I et II sur des femmes et des hommes. On trouve les résultats suivants :