L’espace de probabilités $(\Omega,\mathcal{A},P)$

1 Introduction

Le calcul des probabilités est la science qui modélise les phénomènes aléatoires. Une modélisation implique donc certainement une simplification des phénomènes, mais cette simplification conduit à une quantification, donc à la possibilité de faire des calculs et à prédire. Le jet d’un dé, le tirage du Loto pourraient être analysés par les lois de la mécanique, mais ce serait trop compliqué pour être utile. La modélisation du calcul des probabilités a été inventée par A. N. Kolmogorov dans un livre paru en 1933. Cette modélisation est faite à partir de 3 objets $(\Omega,\mathcal{A},P)$ que nous allons décrire.

2 L’espace des observables Ω.

Nous conviendrons que effectuer une expérience, c’est sélectionner par un procédé quelconque un élément ω dans un ensemble Ω: jeter un dé revient à sélectionner un élément de $\Omega = \{1,2,3,4,5,6\}$; jeter ensemble deux dés rouge et vert revient à sélectionner un élément de l’ensemble $\Omega = \{1,2,3,4,5,6\}$ des couples ordonnés (i,j) avec $1 \leq i \leq 6$ et $1 \leq j \leq 6$ (ici Ω a 36 points). Plus délicat: jeter ensemble deux dés indiscernables revient à sélectionner un élément de l’ensemble Ω des couples (i,j) avec $1 \leq i \leq j \leq 6$ (ici Ω a $6+\frac{1}{2}6\times 5 = 21$ points). Observer la durée de vie d’une ampoule de 100 watts revient à sélectionner un élément de $\Omega = [0,+\infty[$. Mesurer la durée de vie de 12 ampoules de 100 watts est sélectionner un élément de $\Omega = [0,+\infty[^{12}$.

Cet ensemble Ω est appelé l’**espace des observables**. On dit aussi dans la littérature l’espace échantillon, l’espace des évènements - élémentaires, l’expérimental ou encore l’évènementiel. Ses points ω sont appelés observables ou évènements-élémentaires. Il est très important qu’il soit clairement défini. On peut s’exercer à définir Ω dans les 2 cas suivants : jeter 12 fois de suite la même pièce de monnaie, jeter en même temps 12 pièces de monnaie identiques (on admet que la pièce tombe sur pile ou sur face, et jamais sur la tranche).

3 La tribu des évènements $\mathcal{A}$.

Les questions qu’on se pose sur le résultat d’une expérience sont systématiquement du type suivant: on choisit un sous ensemble A de l’espace d’observables Ω et on se demande: le résultat ω de l’expérience va-t-il tomber dans A ou non? Les parties de Ω pour lesquelles on se pose ce genre de question sont appelées des **évènements**. Un des premiers points délicats de la théorie est que on ne va pas toujours considérer tous les sous ensembles de Ω comme des évènements. Dans l’exemple de la lampe de 100 watts, il parait inintéressant de se demander si sa durée de vie, mesurée en heures, est un nombre irrationnel, et intéressant de se demander si elle tombe dans l’intervalle

$[300{,}400]$. L'idée de Kolmogorov est que l'ensemble $\mathcal{A}$ des évènements a une structure de tribu:

Définition Soit Ω un ensemble et soit $\mathcal{A}$ une partie de $\mathcal{P}(\Omega)$. $\mathcal{A}$ a une structure de **tribu** si il satisfait aux trois axiomes:

1. Si $A \in \mathcal{A}$, alors son complémentaire $A^c = \Omega \setminus A$ est aussi dans $\mathcal{A}$.
2. Si on a une suite finie ou dénombrable $A_1, \ldots, A_n, \ldots$ d'éléments de $\mathcal{A}$, alors leur réunion $\bigcup_{n \geq 1} A_n$ est aussi dans $\mathcal{A}$.
3. L'ensemble vide $\emptyset$ est dans $\mathcal{A}$.

Un élément de $\mathcal{A}$ est appelé un **événement**.

Tirons quelques conséquences de ces axiomes.

Proposition 1.1 Soit $\mathcal{A}$ une tribu de parties de l'ensemble Ω. Alors $\Omega \in \mathcal{A}$. De plus, si on a une suite finie ou dénombrable $A_1, \ldots, A_n, \ldots$ d'éléments de $\mathcal{A}$, alors leur intersection $\bigcap_{n \geq 1} A_n$ est aussi dans $\mathcal{A}$.

Démonstration En appliquant les axiomes 1 et 3, on a le premier résultat. Pour le second, il suffit de se rappeler que le complémentaire d'une réunion finie ou infinie d'ensembles est l'intersection des complémentaires ("Loi de Morgan"). Donc

$$\bigcap_{n \geq 1} A_n = \Big(\bigcup_{n \geq 1} A_n^c\Big)^c,$$

et le deuxième membre de cette égalité est donc dans $\mathcal{A}$: on applique successivement l'axiome 1, puis 2, puis 1 à nouveau.

Le langage de la théorie des ensembles permet des calculs systématiques sur les évènements. Toutefois, il faut savoir que le langage courant, que nous utilisons dans une première étape pour décrire des évènements a sa traduction ensembliste. Voici un petit dictionnaire :

Ensemble Ω:	évènement certain
Ensemble vide:	évènement impossible
$A \cup B$:	A *ou* B sont réalisés ("ou" non exclusif)
$A \cap B$:	A *et* B sont réalisés
A et B sont disjoints:	les évènements A et B sont incompatibles
$A^c = \Omega \setminus A$:	évènement contraire de A.

Le fait que on ne sorte pas de la famille des évènements intéressants à considérer en prenant une intersection ou une réunion d'évènements est raisonnable si ceux ci sont en nombre fini. Le fait de se permettre ceci également quand on en a une infinité est plus subtil: les mathématiques ne maniant que des ensembles finis sont élémentaires mais les résultats exacts auquels elles conduisent sont trop compliqués pour être utilisables. Le passage à l'infini est le passage de l'algèbre à l'analyse, donc à des approximations maniables et à de puissantes techniques issues du calcul différentiel et intégral. Quant au fait que dans ce passage à l'infini, on se limite à une infinité *dénombrable* d'évènements, c'est un point technique qu'on ne justifiera que dans un cours de 3 ème année d'université. Rappelons qu'un ensemble E avec une infinité d'éléments est dit dénombrable si il existe une bijection entre E et l'ensemble $\mathbf{N}$ des entiers positifs: l'ensemble $\mathbf{Q}$ des nombres rationnels est dénombrable, le segment $[0,1]$ ne l'est pas, comme nous l'avons vu en première année.

Finalement, ce point délicat: "on ne considère pas nécessairement tout sous ensemble A de Ω comme un élément de la tribu $\mathcal{A}$ des évènements" ne jouera pas un grand rôle dans la suite. Typiquement, nous envisagerons deux cas particuliers importants:

- Le cas où Ω lui même est dénombrable, et nous prendrons comme tribu $\mathcal{A}$ la famille $\mathcal{P}(\Omega)$ de tous les sous ensembles de Ω.
- Le cas où Ω est la droite réelle $\mathbb{R}$. Nous prendrons alors pour tribu $\mathcal{A}$ la tribu $\mathcal{B}$ (dite tribu de Borel, dont les éléments sont appelés des *boréliens*) qui est la plus petite tribu qui contient tous les intervalles de $\mathbb{R}$.
  On peut laborieusement démontrer que $\mathcal{B} \neq \mathcal{P}(\mathbb{R})$; toutefois, une description complète des éléments de $\mathcal{B}$ n'est pas possible, et en fait pas très utile en pratique: les seuls boréliens que nous aurons à manipuler seront les intervalles (attention, $\mathbb{R}$ ou une demi droite sont aussi des intervalles) ou des réunions finies, ou plus rarement, dénombrables, d'intervalles.

Ce ne sont pas les seuls espaces de probabilité utilisés: on verra le schéma Succès Echec à la section 2 et le cas $\Omega = \mathbb{R}$ ou $\mathbb{R}^n$ plus tard.

Définition La plus petite tribu qui contient les ouverts de $\mathbb{R}$ muni de sa topologie canonique est appelée **tribu de Borel**. Les éléments de cette tribu sont appelés les **boréliens** de $\mathbb{R}$.

4 La probabilité P

Définition Etant donnés un espace d'observables Ω et une tribu d'évènements $\mathcal{A}$ formée de certains sous ensembles de Ω, **une probabilité** P est une application de $\mathcal{A}$ dans $[0,1]$, donc un procédé qui associe à tout évènement A un nombre $P(A)$ compris entre 0 et 1 **appelé probabilité de** A, et qui satisfait aux axiomes suivants

- L' évènement certain est de probabilité 1: $P(\Omega) = 1$.
- Axiome d'additivité dénombrable: pour toute suite $A_1,A_2,\dots,A_n\dots$ d'évènements de $\mathcal{A}$ qui sont de plus deux à deux disjoints, c'est à dire tels que $A_k \cap A_j =$

$\emptyset$ si $k \neq j$, alors la série
$$\sum_{k=1}^{\infty} P(A_k)$$
converge et a pour somme $P(\bigcup_{k\geq 1} A_k)$.
Le triplet $(\Omega,\mathcal{A},P)$ est alors appelé un **espace de probabilité**.

Voici quelques conséquences immédiates des axiomes.

Théorème 1.2 Soit $(\Omega,\mathcal{A},P)$ un espace de probabilité. Alors

1. $P(\emptyset) = 0$.
2. Si $A_1,A_2,\ldots,A_n$ dans $\mathcal{A}$ sont deux à deux disjoints, alors
$$P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n);$$
en particulier $P(A^c) = 1 - P(A)$.
3. Si A et B sont dans $\mathcal{A}$ et si $A \subset B$ alors $P(A) \leq P(B)$.

Démonstration

1) L'axiome d'additivité dénombrable est appliquable à la suite constante définie par $A_n = \emptyset$, qui est effectivement formée d'évènements deux à deux disjoints. La série dont le terme général $P(\emptyset)$ est constant ne peut converger que si ce terme général est 0.

2) Sa première partie se démontre en appliquant l'axiome d'additivité dénombrable à $A_1,A_2,\ldots,A_n$ continuée par $\emptyset = A_{n+1} = A_{n+2} = \cdots$, et en utilisant le 1). Appliquer ça à $n = 2$, $A_1 = A$ et $A_2 = A'$ fournit $1 = P(\Omega) = P(A) + P(A^c)$ en utilisant le premier axiome d'une probabilité.

3) On écrit $B = A \cup (B \setminus A)$ comme réunion de deux ensembles disjoints (notez que $B \setminus A = B \cap A'$ est bien dans $\mathcal{A}$), et on applique le 2): $P(B) = P(A) + P(B \setminus A) \geq P(A)$.

Théorème 1.3 Soit $(\Omega,\mathcal{A},P)$ un espace de probabilité. Alors

1. Si A et B sont dans $\mathcal{A}$, mais ne sont pas nécessairement disjoints, alors
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$
Si les $A_1,A_2,\ldots,A_n$ dans $\mathcal{A}$ ne sont pas nécessairement deux à deux disjoints, alors
$$P(A_1 \cup \cdots \cup A_n) \leq P(A_1) + \cdots + P(A_n).$$
2. Continuités croissante et décroissante: Soit une suite $B_1,B_2,\ldots,B_n\ldots$ d'évènements de $\mathcal{A}$ qui soit ou bien croissante (c'est à dire que pour tout $n \geq 1$ on a $B_n \subset B_{n+1}$) ou bien décroissante (c'est à dire que pour tout $n \geq 1$ on a $B_n \supset B_{n+1}$). Alors, dans le cas croissant:
$$\lim_{n\to+\infty} P(B_n) = P(\bigcup_{n\geq 1} B_n);$$

et dans le cas décroissant:

$$\lim_{n\to+\infty} P(B_n) = P(\bigcap_{n\geq 1} B_n).$$

3. Sous additivité dénombrable: Soit une suite $B_1,B_2,\ldots,B_n\ldots$ d'évènements de $\mathcal{A}$. Alors ou bien la série $\sum_{k=1}^{\infty} P(B_k)$ diverge; ou bien elle converge et dans ce cas sa somme est $\geq P(\bigcup_{n\geq 1} B_n)$.

Démonstration

1. On écrit comme dans la démonstration précédente:

$$P(B) = P(A\cap B) + P(B\setminus A),\, P(A) = P(A\cap B) + P(A\setminus B),$$

puis on écrit $A\cup B = (A\cap B)\cup(B\setminus A)\cup(A\setminus B)$ comme réunion de trois ensembles deux à deux disjoints et on applique le 2):

$$P(A\cup B) = P(A\cap B) + P(B\setminus A) + P(A\setminus B) =$$
$$P(A\cap B)+(P(B)-P(A\cap B))+(P(A)-P(A\cap B)) = P(A)+P(B)-P(A\cap B);$$

Pour terminer le 1) on démontre le résultat par récurrence sur n. C'est trivial pour $n = 1$. Si c'est démontré pour n, appliquons la première partie de ce 1) à $A = A_1\cup\cdots\cup A_n$ et à $B = A_{n+1}$. On obtient, à l'aide de l'hypothèse de récurrence

$$P(A\cup B) = P(A)+P(B)-P(A\cap B) \leq P(A)+P(B) \leq (\sum_{k=1}^{n} P(A_k))+P(A_{n+1}).$$

2. Dans le cas croissant, posons $A_1 = B_1$ et, pour $n\geq 2$, $A_n = B_n\setminus B_{n-1}$. Les $A_1,A_2,\ldots,A_n\ldots$ sont alors deux à deux disjoints. La série $\sum_{k=1}^{\infty} P(A_k)$ est donc convergente. D'après la partie 2) de la proposition précédente, on a

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

Passons à la limite dans l'égalité ci dessus; on obtient

$$\lim_{n\to+\infty} P(B_n) = \sum_{k=1}^{\infty} P(A_k).$$

Or d'après l'axiome d'additivité dénombrable, le second membre est $P(\bigcup_{k\geq 1} A_k)$, qui est aussi par définition des A_n égal à $P(\bigcup_{n\geq 1} B_n)$.
Dans le cas décroissant, on se ramène au cas précédent par passage au complémentaires, à l'aide de la loi de Morgan: le complémentaire d'une union est l'intersection des complémentaires:

$$\lim P(B_n) = 1-\lim P(B_n^c) = 1-P(\cup_{n\geq 1}B_n^c) = 1-(1-P(\cap_{n\geq 1}B_n)) =$$
$$P(\cap_{n\geq 1}B_n).$$

3. La suite d'évènements définie par $C_n = B_1 \cup \cdots \cup B_n$ est croissante et on peut lui appliquer le 2). En utilisant aussi la sous additivité finie on a donc

$$P(\bigcup_{n\geq 1} B_n) = \lim_{n\to+\infty} P(C_n) \leq \lim_{n\to+\infty}(P(B_1) + \cdots + P(B_n)) = \sum_{k=1}^{\infty} P(B_k).$$

Exercices sur la section 1.

1. Soit A,B,C trois évènements d'un espace de probabilité. Montrer à l'aide du Th. 1.2) que
$$P(A\cup B\cup C) =$$
$$P(A)+P(B)+P(C)-P(A\cap B)-P(B\cap C)-P(C\cap A)+P(A\cap B\cap C).$$
Etablir une formule de ce genre pour une réunion de 4 évènements.
2. Soit $\mathcal{A}$ une tribu d'évènements sur Ω, et soit f une fonction positive sur $\mathcal{A}$ ayant les propriétés suivantes: $f(\Omega) = 1$, $f(A\cup B) = f(A) + f(B)$ si A et B sont des évènements disjoints et , si (B_n) est une suite décroissante de $\mathcal{A}$ telle que $\cap_{n\geq 1} B_n = \emptyset$ alors
$$\lim_{n\to+\infty} f(B_n) = 0.$$
Montrer qu'alors f est une probabilité. Méthode: si (A_n) est une suite d'évènements deux à deux disjoints, considérer
$$B_n = \cup_{k\geq n+1} A_k.$$