Chapitre 1

Le calcul d'une probabilité

Dans ce premier chapitre, nous allons présenter les concepts et les formules qui permettent de mener à bien le calcul d'une probabilité. On s'accorde à penser que les débuts du calcul des probabilités remontent au XVII^e siècle, plus précisément à l'occasion d'un échange de correspondance initié en 1654 entre Pascal, vivant à Paris, et Fermat, résidant à Toulouse. Cet échange se rapporte à la résolution du problème « des partis » c'est-à-dire de la répartition équitable entre les joueurs des enjeux d'un jeu interrompu. Huygens, dans son ouvrage *De ratiociniis in ludo aleae*, reconnaît en Pascal et Fermat les inventeurs du calcul des probabilités. Mais dans cet ouvrage, Huygens en a lui-même énoncé et correctement utilisé les concepts fondamentaux. Vinrent ensuite, au XVIII^e siècle, de nombreux contributeurs parmi lesquels on peut citer Jacques Bernoulli, avec son ouvrage *Ars conjectandi* paru en 1713 et dans lequel il déclare : « La probabilité est en effet un degré de la certitude et en diffère comme la partie diffère d'un tout » , puis de Moivre, Laplace, Gauss, Poisson. C'est ensuite avec Borel et Lebesgue que la théorie de la mesure et de l'intégration s'est développée et, enfin, avec Kolmogorov que l'axiomatique des probabilités actuelles a été formulée. Nous renvoyons à [10], [26] ou encore à [4] chapitre XI, pour un historique plus détaillé.

1.1 L'espace fondamental

Tout d'abord, nous parlons d'expérience aléatoire pour indiquer qu'on ne peut prévoir de façon précise son résultat et que, répétée dans des conditions identiques, elle peut conduire à des résultats différents. Cette dernière propriété constitue, en effet, la négation d'une expérience déterministe où les mêmes causes produisent les mêmes effets. On note alors Ω l'espace décrivant les résultats possibles de cette expérience.

Définition 1. On appelle Ω *l'espace fondamental* ou encore l'univers des états possibles. Un événement A est un sous-ensemble de Ω. Les événements $\{\omega\}$ réduits à un point sont appelés *événements élémentaires*. Un événement $A \subset \Omega$ est réalisé si l'expérience aléatoire donne lieu à un résultat ω qui appartient à A.

L'ensemble Ω peut être :

— fini : par exemple, pour représenter l'expérience aléatoire d'un jet de 2 dés différenciés, l'un rouge et l'autre jaune, on peut considérer Ω comme étant l'ensemble des couples (i, j), où i est le résultat du dé rouge et j celui du dé jaune. On obtient alors :

$$\Omega = \{(i, j) \mid 1 \leqslant i, j \leqslant 6\}.$$

On peut également représenter Ω par une matrice à 6 lignes et 6 colonnes :

$$\Omega = \begin{pmatrix} (1,1) & \cdots & (1,6) \\ \vdots & & \vdots \\ (6,1) & \cdots & (6,6) \end{pmatrix};$$

qui permet de calculer facilement, par exemple, la loi de la somme des points obtenus.

— dénombrable : par exemple, si on s'intéresse au nombre d'appels à un standard téléphonique durant un temps T, aucun argument ne nous permet de borner supérieurement le nombre d'appels possibles. On retient alors pour Ω l'ensemble des entiers naturels $\mathbb{N}$;

— infini non dénombrable : par exemple, l'ensemble des couples de réels représentant la position d'une particule sur une surface plane ou encore le temps de bon fonctionnement d'un dispositif qui conduit alors à $\Omega = \mathbb{R}_+$.

1.1.1 L'ensemble des événements

On a vu que lorsqu'on s'intéresse à un sous-ensemble de Ω non réduit à un point on parle alors d'événement. Par exemple, si on s'intéresse au jet de dés, on peut considérer l'événement « la somme des points obtenus est paire » ou encore « les dés donnent le même chiffre » et ainsi de suite. Dans le cas où Ω est fini ou dénombrable l'ensemble des événements coïncident avec l'ensemble des parties de Ω. Dans le cas infini non dénombrable, par exemple si $\Omega = \mathbb{R}$, il se peut qu'un sous-ensemble quelconque de Ω ne soit pas un événement : on ne peut pas lui affecter une probabilité. On se limite alors à une classe de sous-ensembles de Ω que l'on va noter $\mathcal{T}$. On rappelle que $\overline{A}$ représente le complémentaire de A dans Ω.

Définition 2. Soit Ω un espace fondamental. Tout sous-ensemble non vide $\mathcal{T}$ de $\mathcal{P}(\Omega)$, ensemble des parties de Ω, qui vérifie :

- **(i)** Si $A \in \mathcal{T}$, alors $\overline{A} \in \mathcal{T}$;
- **(ii)** Si $(\forall n \in \mathbb{N})$, $A_n \in \mathcal{T}$, alors $\bigcup_n A_n \in \mathcal{T}$.

est une tribu. Le couple $(\Omega, \mathcal{T})$ est appelé espace probabilisable.

On en conclut immédiatement que $\Omega \in \mathcal{T}$, $\emptyset \in \mathcal{T}$, et si $(\forall n \in \mathbb{N})$, $A_n \in \mathcal{T}$, alors $\bigcap_n A_n \in \mathcal{T}$. L'ensemble des événements, ou tribu, est donc stable par union finie ou dénombrable, intersection finie ou dénombrable et par passage au complémentaire. Notons également que l'intersection de deux tribus est aussi une tribu.

Définition 3. Soit $\mathcal{A}$ un sous-ensemble de l'ensemble $\mathcal{P}(\Omega)$. On appelle *tribu engendrée* par $\mathcal{A}$ l'intersection de toutes les tribus contenant $\mathcal{A}$.

Définition 4. (Tribu borélienne) Si Ω est muni d'une topologie, on appelle *tribu borélienne* la plus petite tribu engendrée par les ouverts. L'utilisation la plus fréquente est $\mathcal{B}(\mathbb{R})$ la tribu borélienne de $\mathbb{R}$ qui est aussi engendrée par les intervalles ouverts.

Les événements associés à une expérience aléatoire sont énoncés en langage courant. Il en faut une traduction mathématique ensembliste. Nous rappelons ici que la conjonction « et » se traduit par l'intersection, que la conjonction « ou » se traduit par l'union, que la négation d'un énoncé se traduit par le passage au complémentaire et que l'implication se traduit par l'inclusion. On comprend alors qu'il est important d'énoncer un événement de façon très claire pour en faciliter sa transcription ensembliste.

1.1.2 Quelques rappels sur les ensembles

Soient Ω un ensemble, $A \subset \Omega$ et, pour $n \in \mathbb{N}$, $B_n \subset \Omega$ des parties de Ω. On a :

$$\begin{aligned}
\overline{\left(\bigcap_n B_n\right)} &= \bigcup_n \overline{B_n}\ , \\
\overline{\left(\bigcup_n B_n\right)} &= \bigcap_n \overline{B_n}\ , \\
A \cup \left(\bigcap_n B_n\right) &= \bigcap_n (A \cup B_n)\ , \\
A \cap \left(\bigcup_n B_n\right) &= \bigcup_n (A \cap B_n)\ .
\end{aligned}$$

1.2 Probabilité

Définition 5. Soit $(\Omega, \mathcal{T})$ un espace probabilisable. Une *probabilité* est une application $\mathbb{P}\colon \mathcal{T} \longrightarrow [0,1]$ qui vérifie :

(i) $\mathbb{P}(\Omega) = 1$

(ii) Pour toute suite d'événements $(A_n)_{n\in\mathbb{N}}$, deux à deux disjoints, on a

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n).$$

Le triplet $(\Omega,\ \mathcal{T},\ \mathbb{P})$ est un *espace probabilisé.*

Outre le fait que la masse totale vaut 1, il est remarquable que la notion même de probabilité ne repose que sur l'axiome fondamental **(ii)** : celui de la σ-additivité. Cette propriété généralise au cas infini dénombrable le fait que la longueur totale d'un segment constitué de sous-segments deux à deux disjoints est la somme des longueurs des sous-segments. De cette définition résulte un certain nombre de règles simples qui sont d'un usage constant en probabilités :

(i) $\mathbb{P}(\emptyset) = 0$,

(ii) Axiome des probabilités totales : si $A \cap B = \emptyset$, alors :
$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$,

(iii) $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$,

(iv) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$,

(v) Continuité croissante : si pour tout $n \in \mathbb{N}$, $A_n \subset A_{n+1}$, alors

$$\mathbb{P}\left(\bigcup_{n=0}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n) \quad .$$

(vi) Continuité décroissante : si pour tout $n \in \mathbb{N}$, $A_{n+1} \subset A_n$, alors

$$\mathbb{P}\left(\bigcap_{n=0}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n) \quad .$$

(vii) Formule de Poincaré : Soient $A_1, \ldots, A_n$ n événements, on a

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k-1} S_{k,n}$$

où

$$S_{k,n} = \sum_{i_1 < i_2 < \ldots < i_k} \mathbb{P}\left(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}\right);$$

la somme portant alors sur toutes les sous-suites croissantes de taille k de $[\![1;n]\!]$.

Démonstration. (i) Prenons tout d'abord la suite $(A_n)_{n\geqslant 0}$ telle que $A_0=\Omega$ et $(\forall n \in \mathbb{N}^\star)\ A_n=\emptyset$. Tous ces ensembles sont deux à deux disjoints puisque leur intersection est vide. La σ-additivité nous donne :

$$\mathbb{P}(\Omega)=\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right)=\mathbb{P}(\Omega)+\sum_{n\geqslant 1}\mathbb{P}(\emptyset)$$

ce qui conduit forcément à $\mathbb{P}(\emptyset)=0$.

(ii) Posons alors $A_0=A,\ A_1=B$ où A et B sont disjoints et, pour $n\geqslant 2$, $A_n=\emptyset$. On obtient, toujours par σ-additivité,

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

(iii) Prenant ensuite dans l'égalité ci-dessus $B=\overline{A}$ on trouve :
$1=\mathbb{P}(A)+\mathbb{P}(\overline{A})$. On en déduit que $\mathbb{P}(\overline{A})=1-\mathbb{P}(A)$.

(iv) En appliquant l'axiome des probabilités totales à $B\cap A$ et $B\cap\overline{A}$, dont l'union vaut B, on obtient : $\mathbb{P}(B)=\mathbb{P}(B\cap A)+\mathbb{P}(B\cap\overline{A})$ ce qui conduit à $\mathbb{P}(B\cap\overline{A})=\mathbb{P}(B)-\mathbb{P}(A\cap B)$. On écrit alors $A\cup B=A\cup(B\cap\overline{A})$ qui se traduit alors par :

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

(v) La continuité croissante s'obtient en transformant une suite croissante d'événements en suite disjointe en posant $(\forall n\geqslant 1)\ B_n=A_n\setminus A_{n-1}$ et $B_0=A_0$. On peut alors appliquer la σ-additivité à la suite (B_n).

(vi) Pour la continuité décroissante on commence par observer que la généralisation des formules de Morgan nous donne $\bigcap\limits_{n=0}^{\infty} A_n=\overline{\left(\bigcup\limits_{n=0}^{\infty}\overline{A_n}\right)}$. Passant aux probabilités on obtient :

$$\mathbb{P}\left(\bigcap_{n=0}^{\infty} A_n\right)=1-\mathbb{P}\left(\bigcup_{n=0}^{\infty}\overline{A_n}\right).$$

Comme la suite (A_n) est décroissante la suite $(\overline{A_n})$ est croissante et on peut lui appliquer la propriété de continuité croissante. Ce qui donne :

$$\mathbb{P}\left(\bigcap_{n=0}^{\infty} A_n\right)=1-\lim_{n\to+\infty}\mathbb{P}(\overline{A_n})=\lim_{n\to+\infty}\mathbb{P}(A_n)\quad .$$

(vii) Notons enfin que la formule de Poincaré peut encore s'écrire :

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} \mathbb{P}(A_k) - \sum_{1\leqslant i_1<i_2\leqslant n} \mathbb{P}(A_{i_1}\cap A_{i_2}) + \cdots$$
$$+(-1)^k \sum_{1\leqslant i_1<i_2\ldots<ik\leqslant n} \mathbb{P}(A_{i_1}\cap A_{i_2}\ldots\cap A_{i_k}) + \cdots$$
$$+(-1)^n \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right)$$

$\square$

Comme $\mathbb{P}(\Omega)=1$, on dit souvent que Ω est l'événement certain alors que $\emptyset$ est l'événement impossible. Par analogie avec la théorie de la mesure, on dit aussi qu'un événement de probabilité nulle est *négligeable*. En revanche, lorsqu'une propriété est vraie sur un sous-ensemble de probabilité 1, on dit qu'elle est presque sûre et on note : $p.s.$

Exemple :

— Si Ω est fini, on définit *l'équiprobabilité* comme étant l'unique probabilité sur $(\Omega, \mathcal{P}(\Omega))$ telle que pour tout $\omega\in\Omega$, $\mathbb{P}[\{\omega\}] = \frac{1}{\text{card }\Omega}$. Pour un sous-ensemble quelconque A de Ω, on a alors $\mathbb{P}(A) = \dfrac{\text{card } A}{\text{card }\Omega}$, ce que l'on traduit souvent en disant que la probabilité de A est le quotient :

$$\frac{\text{nombre de cas favorables à la réalisation de A}}{\text{nombre de cas possibles}}.$$

— Si Ω est dénombrable, $\Omega = \{\omega_k \mid k\in\mathbb{N}\}$, on peut utiliser comme tribu $\mathcal{P}(\Omega)$. Une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ est alors caractérisée par une suite de réels p_k tels que $0\leqslant p_k\leqslant 1$ et $\sum_{k=0}^{\infty} p_k = 1$ où $p_k = \mathbb{P}[\{\omega_k\}]$.

— Si $\Omega=[0,1]$ l'espace est alors infini non-dénombrable et on peut prendre pour $\mathbb{P}$ la mesure de Lebesgue.

1.3 Probabilité conditionnelle

Il faut se situer dans la problématique suivante. Lors d'une expérience aléatoire, on dispose d'une information partielle sur la nature du résultat : on sait qu'un résultat $\omega\in B$ a été obtenu et donc que B est réalisé. Ceci nous invite à recalculer les probabilités qui nous intéressent en tenant compte de cette information. Pour un événement A on ne va donc retenir que ce qui est commun à A et à B en considérant que le nouvel univers des possibles se résume à B. On obtient alors :

Définition 6. Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $B \in \mathcal{T}$ un événement de probabilité non nulle. On appelle *probabilité conditionnelle* de A sachant B, que l'on note $\mathbb{P}(A|B)$, le réel défini par :

$$\boxed{\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}\cdot}$$

Proposition 1. L'application $\mathbb{P}(.|B) : \mathcal{T} \longrightarrow [0,1]$ est une probabilité sur $(\Omega, \mathcal{T})$.

Cette proposition, relativement simple à démontrer, se montre extrêmement fertile à l'usage. Elle signifie que tous les résultats obtenus pour une probabilité s'appliquent à une probabilité conditionnelle. Par exemple on a :

$$\mathbb{P}(\overline{A}|B) = 1 - \mathbb{P}(A|B)$$

et ainsi de suite.

Remarque : Si A et B sont de probabilités non nulles, on a :

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A|B) \ = \mathbb{P}(A)\mathbb{P}(B|A). \tag{1.1}$$

Conséquences : Trois formules fondamentales

1. **Formule des probabilités composées** : Il s'agit d'une généralisation de (1.1) : soient $A_1, \ldots, A_n$ n événements tels que $\mathbb{P}\left[\bigcap_{k=1}^{n-1} A_k\right] \neq 0$. On a le résultat suivant :

$$\boxed{\mathbb{P}\left[\bigcap_{k=1}^{n} A_k\right] = \mathbb{P}(A_1)\ \mathbb{P}(A_2|A_1)\ \mathbb{P}(A_3|A_1 \cap A_2)\ \ldots\ \mathbb{P}(A_n|\bigcap_{k=1}^{n-1} A_k).}$$

2. **Formule des probabilités totales** : Soit $(A_k)_{k=1,\ldots,n}$ un système complet d'événements, c'est-à-dire si $i \neq j$, alors $A_i \cap A_j = \emptyset$ et $\bigcup_{k=1}^{n} A_k = \Omega$. On suppose de plus que pour tout $k \in [\![1;n]\!]$, $\mathbb{P}(A_k) > 0$. Soit $B \in \mathcal{T}$. On a alors :

$$\boxed{\mathbb{P}(B) = \sum_{k=1}^{n} \mathbb{P}(B|A_k)\ \mathbb{P}(A_k).}$$

3. **Formule de Bayes** : Dans la formule qui suit, les événements A_k ci-dessus vont être considérés comme des causes dont on connait les probabilités. Cette connaissance résulte soit d'une étude statistique soit d'un jugement d'expert. On les appelle probabilités *a priori*. On suppose aussi connaître les probabilités conditionnelles $\mathbb{P}(B|A_k)$ de l'apparition de l'événement B lorsque la cause A_k est réalisée. On suppose enfin que ce qu'on observe est l'événement B avec $\mathbb{P}(B) \neq 0$. On va alors corriger les probabilités *a priori* en tenant compte de ce qui s'est passé. On est amené à calculer les probabilités $\mathbb{P}(A_k|B)$ que l'on appelle alors probabilités *a posteriori* :

$$\boxed{\forall k \in [\![1;n]\!], \quad \mathbb{P}(A_k|B) = \frac{\mathbb{P}(B|A_k)\ \mathbb{P}(A_k)}{\sum_{j=1}^{n} \mathbb{P}(B|A_j)\ \mathbb{P}(A_j)}.}$$

Le numérateur est l'un des termes de la somme qui se trouve au dénominateur. Cette formule, aussi appelée *formule de probabilités des causes*, peut s'appliquer dans de nombreux cas, dans des domaines aussi divers que le contrôle de qualité, la médecine, l'actuariat, etc.

1.4 Indépendance

Définition 7. Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé. Deux événements A et B sont dits *indépendants* si et seulement si

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

Il ne faut pas confondre événements incompatibles, qui veut dire que $A \cap B = \emptyset$, avec événements indépendants. En particulier, deux événements incompatibles, de probabilités non nulles, sont dépendants : si l'un est réalisé, l'autre ne l'est pas.
Si $\mathbb{P}(A)$ ou $\mathbb{P}(B)$ est nul, alors A et B sont indépendants.
Si $\mathbb{P}(B) \neq 0$, A et B indépendants entraîne que $\mathbb{P}(A|B) = \mathbb{P}(A)$ et donc la réalisation de B ne modifie pas la probabilité de A.

Remarque : En ne considérant que des événements A et B de probabilités non nulles, on peut définir l'indépendance de A par rapport à B par la condition $\mathbb{P}(A|B) = \mathbb{P}(A)$, qui est peut-être plus intuitive. On montre alors que $\mathbb{P}(B|A) = \mathbb{P}(B)$ et on aboutit à la définition ci-dessus comme condition nécessaire et suffisante d'indépendance.

Proposition 2. Si A et B sont indépendants, alors A et $\overline{B}$ sont indépendants (donc aussi B et $\overline{A}$) et $\overline{A}$ et $\overline{B}$ sont indépendants.