Probabilités discrètes et calcul intégral

Cours de probabilités

Tome 1

Sommaire

Chapitre 1

Espaces probabilisés

1. Un peu d'histoire

Siméon Poisson (1781-1840) a écrit : "Un problème relatif aux jeux de hasard proposé à un austère janséniste par un homme du monde a été à l'origine du calcul des probabilités". Il s'agissait du chevalier de Méré (1607-1684), homme du monde, qui proposa à Blaise Pascal (1623-1662), austère janséniste, des problèmes sur des jeux de hasard, avec, entre autres, le célèbre "problème des parties" : le prix d'un tournoi est gagné par le premier participant qui remporte un nombre fixé de parties. Si on interrompt le jeu avant la fin, comment répartir équitablement le prix entre les participants ?
De nombreuses fausses solutions ont été proposées durant deux siècles pour ce problème. Pascal, dans son *Traité du triangle arithmétique* publié en 1665, en donna une solution correcte qu'il proposa à Pierre de Fermat (1601-1665).
Dans son ouvrage *De ratiocinus in ludo alae*, Christian Huygens (1629-1695) exposa les concepts fondamentaux du calcul des probabilités, comme par exemple le calcul de l'espérance d'une variable aléatoire ne prenant qu'un nombre fini de valeurs.
Jacques Bernoulli (1654-1705), dans son livre posthume *Ars conjectandi* (1713), a approfondi les résultats de Huygens. Mais il est surtout le premier à démontrer la loi des grands nombres (qui prouve la convergence de la moyenne empirique vers la moyenne), via une approche combinatoire, résultat qui est à l'origine du réel essor des probabilités.
Abraham de Moivre (1667-1754), dans *Doctrines of chances* (1733), précisa la vitesse de convergence dans la loi des grands nombres, établissant ainsi la toute première version du théorème central limite (TCL). Le très célèbre Pierre-Simon Laplace (1749-1827) étendra ce résultat en utilisant de nouveaux outils comme le calcul infinitésimal, en développant les fonctions génératrices et caractéristiques, dans son traité *Théorie analytique des probabilités* (1812). Son ouvrage dépasse le cadre strict de la combinatoire, et donne un élan nouveau au calcul des probabilités.
Les résultats très généraux sur la loi des grands nombres et le théorème central limite seront établis au XIX^{e} siècle par Siméon Poisson, Irénée-Jules Bienaymé (1796-1878), et l'école russe de Saint-Pétersbourg avec Tchebychev (1821-1894), Andrei Markov (1856-1922) et Aleksandr Mikhailovich Lyapunov (1857-1918).
La théorie de la mesure et de l'intégration, due essentiellement à Henri Lebesgue (1875-1941), permet d'asseoir complètement la théorie du calcul des probabilités modernes. La très célèbre monographie de Andrey Nikolaevich Kolmogorov (1903-1987) *Grundbegriffe der Wahrscheinlichkeitsrechnung* (1933) fournit enfin le cadre théorique dans lequel s'exprime l'actuel calcul des probabilités.
Cette première moitié du XX^{e} siècle voit aussi l'essor des processus stochastiques et de leurs applications. Le phénomène s'accélère encore dans la seconde moitié du XX^{e} siècle. Les applications du calcul des probabilités sont très nombreuses, et en faire une liste exhaustive serait impossible. Mais les probabilités interviennent par exemple dans la plupart des modélisations de phénomènes physiques complexes, en démographie, en épidémiologie, en médecine, en biologie, dans les techniques d'analyse d'ADN, en analyse d'image, en reconnaissance des formes, en fiabilité, en assurance, dans les banques, sur les marchés financiers et

boursiers, dans des simulations comme le pilotage des centrales nucléaires, etc.
Depuis l'avènement des outils informatiques, le calcul des probabilités a pris un essor vertigineux, lié à la puissance de calcul des machines. Les simulations, les méthodes de type Monte-Carlo, sont devenues un domaine incontournable du calcul des probabilités.

2. Introduction

L'objet de la théorie des probabilités est de décrire et d'étudier divers modèles mathématiques de phénomènes aléatoires d'un point de vue théorique.

L'étude des théories des probabilités est nécessaire à l'étude de la Statistique, laquelle est plus concernée par la création de certains principes et certains critères pour permettre de traiter des données issues de phénomènes aléatoires. La Statistique inférentielle utilise pleinement la théorie des probabilités.
Les modèles probabilistes ont pour but de décrire les expériences aléatoires – expériences que, théoriquement, on pourrait répéter indéfiniment, et dont les résultats futurs ne peuvent être prédits exactement, même si les conditions expérimentales sont complètement contrôlées.
Comme on le verra, la base de la théorie des probabilités est l'espace probabilisé. L'idée forte derrière cette notion d'espace probabilisé est la stabilisation des fréquences relatives. Supposons que nous répétions une expérience aléatoire un très grand nombre de fois, de manière indépendante, et que, pour chaque expérience, nous soyons intéressés à la réalisation (ou non) d'un certain événement A (même si nous n'avons pas encore défini mathématiquement les notions d'indépendance et d'événement).

Soit $N_n(A)$ le nombre de fois où A est réalisé au cours des n essais indépendants, et notons $r_n(A)$ la fréquence relative correspondante :

$$r_n(A) = \frac{N_n(A)}{n}.$$

Depuis la nuit des temps, on a observé que, dans ces conditions, on observe que la fréquence relative $r_n(A)$ se stabilise, au sens où il existe un réel λ $(0 \leqslant \lambda \leqslant 1)$ tel que :

$$r_n(A) \xrightarrow[n\to+\infty]{} \lambda.$$

L'interprétation intuitive du concept de probabilité est que la probabilité de l'événement A est λ, et on peut raisonnablement espérer que la fréquence relative observée au cours de n expériences indépendantes soit approximativement égale à λ.
Cette approche est appelée « approche fréquentiste des probabilités », et est universellement adoptée.

L'étape suivante est l'axiomatisation complète de cette théorie. Après bien des tergiversations, indiquées dans le paragraphe précédent, c'est finalement A. N. Kolmogorov en 1933 qui a assis complètement l'axiomatique de la théorie moderne des probabilités.
Nous allons développer cette théorie dans les chapitres à venir.

3. Notion d'expériences aléatoires

Cette première notion de la théorie des probabilités n'a vu le jour que vers le XVII^{e} siècle dans l'étude des jeux de hasard (jeux de dés, de cartes, de loterie, etc.). Ces différents jeux, aisément modélisables, obéissent à des lois mathématiques que l'on précisera plus loin.

Il existe bien des situations où l'aléatoire intervient. Citons quelques exemples :

a) l'observation des durées de vie des puces électroniques, ou des humains ;

b) l'observation du volume des transactions en bourse ;

c) l'observation d'un électroencéphalogramme, d'un signal radar, d'un signal sismique ;

d) l'observation de la propagation d'une maladie, etc.

La première étape de la formalisation consiste à préciser le cadre dans lequel on va observer ces différentes actions dues au hasard, ce que l'on nommera *expériences aléatoires*.

Une expérience aléatoire se décrit mathématiquement par la donnée de l'ensemble de tous les résultats possibles de l'expérience en question. On notera ω un tel résultat, qu'on nommera aussi *épreuve* ou *issue*. On notera Ω l'ensemble des résultats possibles de l'expérience.

Il reste cependant une part d'arbitraire dans le choix de Ω. En effet, si on considère un jet d'une pièce, on peut proposer comme espaces :

$\Omega_1 = \{\text{Pile}, \text{Face}\}$;

$\Omega_2 = \{\text{Pile}, \text{Face}, \text{Tranche}\}$;

$\Omega_3 = \mathbb{R}^3 = \{(x, y, z) : \text{coordonnées du centre de gravité de la pièce}\}$;

$\Omega_4 = (\mathbb{R}^3)^{[0,T]} = \{\text{trajectoires de la pièce pendant } [0, T]\}$.

Cela peut paraître surprenant *a priori*, mais cela apparaît à chaque fois que l'on veut donner une formalisation mathématique d'un phénomène réel.
Les expériences aléatoires correspondant à des espaces Ω finis sont particulièrement simples à expliciter. Par exemple, si on lance deux dés distinguables, l'espace Ω se compose alors des coupes $(x, y) = \omega$ tels que $1 \leqslant x, y \leqslant 6$, et alors :

$$\Omega = \{(x, y) : (x, y) \in \mathbb{N}^2; 1 \leqslant x, y \leqslant 6\}.$$

Mais, dans la plupart des cas, les espaces Ω seront de cardinaux infinis. Dans l'exemple b) précédent, on pourra prendre $\mathbb{R}_+^d$ comme espace Ω. L'espace Ω_4 précédent est un espace fonctionnel, espace des fonctions continues $\mathscr{C}([0, T])$ définies sur $[0, T]$ à valeurs dans $\mathbb{R}^3$.

Finalement, la complexité de l'espace Ω est directement liée à celle de l'expérience aléatoire étudiée.
Souvent, on considérera que Ω représente l'ensemble des états de la nature à défaut de savoir préciser plus avant l'ensemble des résultats possibles d'une expérience.

4. Notion d'événement

La seconde étape de la formalisation sera celle d'*événement aléatoire*, c'est-à-dire événement lié à une certaine expérience aléatoire. Dans le langage courant, dire qu'un événement est réalisé revient alors à énoncer une propriété : le dé marque un chiffre pair, l'enfant est de sexe masculin, il pleut...

On considère que la réalisation ou non d'un événement dépend exclusivement du résultat de l'expérience à laquelle il est attaché. Ainsi, un événement A sera toujours représenté par l'ensemble des résultats ω de l'expérience qui le réalisent. A est réalisé si et seulement si le résultat de l'expérience $\omega \in A$.
Par exemple, si l'expérience aléatoire consiste à jeter un dé, alors Ω peut être identifié à l'ensemble $\{1, 2, 3, 4, 5, 6\}$. L'événement A "on a obtenu un nombre pair" s'écrit : $A = \{2, 4, 6\}$.
À toute propriété définie sur Ω, on associe un sous-ensemble de Ω : l'ensemble de tous les $\omega \in \Omega$ qui vérifient la propriété en question.
Réciproquement, tout sous-ensemble de Ω définit une propriété par l'intermédiaire de la notion d'expérience. Cela nous conduit à appeler *provisoirement* événement tout sous-ensemble de Ω, et à dire que l'événement $A \in \mathscr{P}(\Omega)$ (où $\mathscr{P}(\Omega)$ est l'ensemble de toutes les parties de Ω) est réalisé si et seulement si le résultat ω de l'expérience aléatoire appartient à A. Il serait naturel de prendre $\mathscr{P}(\Omega)$ comme ensemble des événements

aléatoires. Mais c'est un ensemble souvent trop vaste pour que l'on puisse le probabiliser. On préférera, en général, une classe de parties de Ω, strictement incluses dans $\mathscr{P}(\Omega)$.

Donc, tout événement A est identifié à une partie de Ω, partie dont les éléments réalisent A.

4.1. Algèbre de Boole des événements

On imposera cependant à cette classe de parties, des conditions de stabilité de façon à ce que les opérations logiques usuelles, ou mieux encore, les opérations ensemblistes correspondantes, ne fassent pas sortir de la classe. De manière plus précise, on a la définition suivante :

> **Définition 1.**
> La classe $\mathscr{E}$ des événements est appelée algèbre de Boole de parties de Ω (c'est donc une classe de parties de Ω), si elle contient Ω et est stable par intersection, union et complémentation.

On dit souvent *algèbre* au lieu de *algèbre de Boole*.

4.2. Théorème de représentation

Décrivons les opérations logiques que l'on peut effectuer sur les événements. D'ailleurs, de manière axiomatique, ce sont plus les opérations et leurs règles de maniement qui définissent la notion d'événement aléatoire.

1. Soient $A \subset \Omega$ et $B \subset \Omega$ deux événements, alors $A \cup B$ est un événement réalisé si et seulement si l'un des deux au moins est réalisé. $A \cap B$ est un événement réalisé si et seulement si les deux sont réalisés simultanément.

2. $\varnothing$ est un événement qui ne peut être réalisé ; on l'appelle l'événement impossible. Par contre, à chaque expérience, Ω est toujours réalisé ; on appelle l'événement Ω l'événement certain.

3. Si $A \subset \Omega$ est un événement, son complémentaire est noté A^c ou $\overline{A}$, et est appelé événement contraire de A. Il est réalisé si et seulement si le résultat ω de l'expérience n'appartient pas à A.

4. La différence de deux événements A et B, notée $A \setminus B$, est un événement qui est défini par :
$$A \setminus B = A \cap B^c,$$
et qui est réalisé si A est réalisé et pas B.

5. La différence symétrique de A et de B, notée $A \Delta B$, est un événement défini par :
$$A \Delta B = (A \setminus B) \cup (B \setminus A),$$
et qui est réalisé si l'un des deux événements est réalisé et pas l'autre.

6. Si, pour tout n de $\mathbb{N}$, l'événement A_n est l'événement « avoir n enfants », alors $\bigcup_{n=1}^{\infty} A_n$ est l'événement signifiant « avoir un ou plusieurs enfants » (ou encore « avoir au moins un enfant ») et $\bigcap_{n=1}^{\infty} A_n^c$ est l'événement signifiant « ne pas avoir d'enfants », car $\left(\bigcup_{n=1}^{\infty} A_n\right)^c = \bigcap_{n=1}^{\infty} A_n^c$.
Cet événement est appelé limité inférieure de la suite (A_n^c) et noté : $\liminf A_n^c$. Donc $\liminf A_n^c = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n$ est l'événement « tous les A_n sont réalisés sauf un nombre fini ».

Ces deux derniers points conduisent à définir les limites de suites d'événements comme suit :

> **Définition 2.**
> Soit $(A_n)_{n \in \mathbb{N}^*}$ une suite d'événements de Ω. On définit alors les limites inférieure et supérieure d'événe-

ments par :

$$A_* = \liminf A_n = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n$$

$$A^* = \limsup A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$$

Si, de plus, les ensembles A_* et A^* coïncident, alors on écrit :

$$A = A_* = A^* = \lim_{n \to +\infty} A_n$$

La proposition qui suit se démontre aisément.

Proposition 1.
Soit $(A_n)_{n \in \mathbb{N}^*}$ une suite d'événements de Ω.

(i) Si $A_1 \subset A_2 \subset \ldots$, alors :

$$\lim_{n \to +\infty} A_n = \bigcup_{n=1}^{\infty} A_n$$

(ii) Si $A_1 \supset A_2 \supset \ldots$, alors :

$$\lim_{n \to +\infty} A_n = \bigcap_{n=1}^{\infty} A_n$$

Tout ce qui précède permet de constater que l'on peut identifier une algèbre de Boole d'événements à une algèbre de parties d'un ensemble. Cette propriété est d'ailleurs très générale.

Théorème 1 (Stone - 1936).
Toute algèbre de Boole d'événements est isomorphe à une algèbre de parties d'un ensemble.

Pour une démonstration, voir A. Renyi, *Calcul des probabilités* (Dunod), p. 13-16.
Ce théorème justifie donc les notations ensemblistes utilisées précédemment.

5. Espace probabilisable

Certains événements font apparaître des opérations plus compliquées. Par exemple, si on joue à Pile ou Face jusqu'à ce que Pile apparaisse pour la première fois, et si on considère l'événement A "le nombre de coups nécessaires pour obtenir Pile est pair", A est alors réunion dénombrables des événements "Pile apparaît pour la première fois au $(2p)$-ième lancer", $p \in \mathbb{N}^*$.
Pour certaines raisons, qui pourraient être de bon sens, mais qui ont en fait une justification mathématique, lorsque Ω n'est pas fini ou dénombrable, on restreint l'ensemble des événements à un certain sous-ensemble $\mathscr{A}$ de $\mathscr{P}(\Omega)$. Ce sous-ensemble $\mathscr{A}$ doit être stable par réunion, intersection et complémentation.
Pour cela, on supposera que $\mathscr{A}$ est une **tribu** d'événements comme définie ci-dessous.

Définition 3.
Soit Ω un ensemble. On appelle tribu (ou σ-algèbre) sur Ω, un ensemble $\mathscr{A}$ de parties de Ω tel que :

(i) $\Omega \in \mathscr{A}$.

(ii) Si $A \in \mathscr{A}$, alors $A^c \in \mathscr{A}$.

(iii) $\forall (A_n)_{n \in \mathbb{N}^*}$ une suite d'éléments de A, alors l'événement $\bigcup_{n=1}^{\infty} A_n \in \mathscr{A}$.

Cette définition, due à A.N. Kolmogorov (1903-1987) dans une célèbre monographie de 1933, fut à la base de tout le calcul des probabilités modernes.
Il y a de nombreuses manières de choisir une tribu. La tribu la plus grossière est : $\mathscr{A} = \{\emptyset, \Omega\}$, et la plus "grosse" est : $\mathscr{A} = \mathscr{P}(\Omega)$. Si Ω est fini ou dénombrable, on prendra souvent $\mathscr{A} = \mathscr{P}(\Omega)$, ensemble des parties de Ω. Si Ω a une puissance supérieure au dénombrable, $\mathscr{P}(\Omega)$ est une tribu beaucoup plus compliquée, et le choix de la tribu dépendra alors des événements auxquels on s'intéresse.

Définition 4.
On appelle espace probabilisable (ou espace mesurable) le couple $(\Omega, \mathscr{A})$ constitué par un ensemble Ω et une tribu $\mathscr{A}$ sur Ω. Les éléments de Ω sont appelés éventualités, les éléments de $\mathscr{A}$ sont appelés événements.

Donnons quelques propriétés sur les tribus.

Proposition 2.
Soit $\mathscr{A}$ une tribu sur Ω. Alors :

(a) $\emptyset \in \mathscr{A}$.

(b) $\forall A_1, \ldots, A_k \in \mathscr{A}$, *alors :* $\displaystyle\bigcup_{i=1}^{k} A_i \in \mathscr{A}$.

(c) $\forall A_1, \ldots, A_k \in \mathscr{A}$, *alors :* $\displaystyle\bigcap_{i=1}^{k} A_i \in \mathscr{A}$.

(d) $\forall (A_n)_{n\in\mathbb{N}^*}$ *suite d'événements de* $\mathscr{A}$, *alors :* $\displaystyle\bigcap_{n\in\mathbb{N}^*} A_n \in \mathscr{A}$.

(e) $\forall (A_n)_{n\in\mathbb{N}^*}$ *suite d'événements de* $\mathscr{A}$, *alors :* $\liminf A_n \in \mathscr{A}$.

(f) $\forall (A_n)_{n\in\mathbb{N}^*}$ *suite d'événements de* $\mathscr{A}$, *alors :* $\limsup A_n \in \mathscr{A}$.

Démonstration. (a) $\emptyset = \Omega^c \in \mathscr{A}$ grâce à la définition (ii) d'une tribu.

(b) Soit $A_i = \emptyset$, $\forall i > k$; alors : $\displaystyle\bigcup_{i=1}^{k} A_i = \bigcup_{n\in\mathbb{N}^*} A_n \in \mathscr{A}$ grâce au (a) de la proposition ci-dessus et au (iii) de la définition d'une tribu.

(c) Soit $A_i = \Omega$, $\forall i > k$; alors : $\displaystyle\overline{\bigcap_{i=1}^{k} A_i} = \overline{\bigcap_{n\in\mathbb{N}^*} A_n} = \bigcup_{n\in\mathbb{N}^*} \overline{A_n} \in \mathscr{A}$ grâce aux (i), (ii) et (iii) de la définition d'une tribu, et une nouvelle fois grâce au (ii), car :

$$\overline{\bigcap_{i=1}^{k} A_i} \in \mathscr{A} \implies \bigcap_{i=1}^{k} A_i \in \mathscr{A}.$$

(d) $\displaystyle\overline{\bigcap_{n\in\mathbb{N}^*} A_n} = \bigcup_{n\in\mathbb{N}^*} \overline{A_n} \in \mathscr{A}$ grâce aux points (ii) et (iii) de la définition d'une tribu et :

$$\overline{\bigcap_{n\in\mathbb{N}^*} A_n} \in \mathscr{A} \implies \bigcap_{n\in\mathbb{N}^*} A_n \in \mathscr{A} \text{ grâce au point (ii) de la définition.}$$

(e) $\forall n \in \mathbb{N}^*$, $A_n \in \mathscr{A} \implies A_n^c \in \mathscr{A}$ grâce au point (ii) de la définition d'une tribu. D'où :

$\displaystyle\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n^c \in \mathscr{A}$ grâce au point (d) de la proposition ci-dessus et au (iii) de la définition d'une tribu.

(f) $\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \in \mathscr{A}$ grâce au (iii) de la définition d'une tribu et au point (d) de la proposition ci-dessus.

□

6. Variable aléatoire

La troisième étape de la modélisation consiste à remarquer que, très souvent, un événement s'énonce de manière numérique (par ex. : « le chiffre marqué sur le dé est 5 »; « le niveau sonore est supérieur à 80 décibels »; etc.).
C'est aussi le cas pour des événements du style « il a plu hier ». En effet, si N désigne le niveau des précipitations de la veille, l'événement s'écrit $\{N > 0\}$.

De manière précise, à toute expérience ω, on associe un nombre $X(\omega)$ ou un n-uple de nombres $(X_1(\omega), \ldots, X_n(\omega))$ mesurant un caractère, ou un ensemble de n caractères du résultat de l'expérience.
Supposons que X désigne une application : $\Omega \to \mathbb{R}$ et que $(X_1, \ldots, X_n)$ désigne une application : $\Omega \to \mathbb{R}^n$.
Les événements les plus simples seront de la forme $\{X \in I\}$ où I est un intervalle réel. Il s'agit d'une **notation** abrégée signifiant :

$$\{X \in I\} = \{\omega \in \Omega : X(\omega) \in I\} = X^{-1}(I).$$

Comme nous avons convenu de ne nous intéresser qu'aux événements faisant partie d'une tribu $\mathscr{A} \subset \mathscr{P}(\Omega)$, on pose la définition suivante :

> **Définition 5.**
> On appelle variable aléatoire réelle toute application à valeurs réelles $X : \Omega \to \mathbb{R}$ telle que, pour tout intervalle I de $\mathbb{R}$, $\{X \in I\}$ soit un événement de la tribu $\mathscr{A}$.

Nous avons aussi besoin de définir la notion de tribu borélienne.

> **Définition 6.**
> On appelle tribu borélienne de $\mathbb{R}$, notée $\mathscr{B}_{\mathbb{R}}$, une tribu sur $\mathbb{R}$ qui soit telle que $\mathscr{B}_{\mathbb{R}}$ soit la plus petite tribu de $\mathbb{R}$ contenant tous les intervalles de $\mathbb{R}$. Les éléments de $\mathscr{B}_{\mathbb{R}}$ seront appelés les boréliens de $\mathbb{R}$.

Nous en admettrons spontanément l'existence, ainsi que la propriété suivante :

> **Propriété.**
> *Soit X une variable aléatoire réelle, alors :*
> $$\forall B \in \mathscr{B}_{\mathbb{R}} \quad , \quad \{X \in B\} \in \mathscr{A}.$$

À ce point de l'exposé, le modèle se présente comme suit :

$$(\Omega, \mathbb{A}) \xrightarrow{X} (\mathbb{R}, \mathscr{B}_{\mathbb{R}}),$$

où Ω représente l'espace des résultats possibles de l'expérience aléatoire, ou des états possibles du phénomène aléatoire, $\mathbb{R}$ représente l'espace des valeurs de ces résultats ou de ces états, et X représente cette manière de mesurer elle-même.
En général, l'observateur ne connaît pas ω lui-même, mais $X(\omega)$; les événements de $\mathscr{A}$ qu'il peut concevoir ne sont que ceux s'exprimant à l'aide de X, c'est-à-dire les événements de la forme $\{X \in B\}$, pour $B \in \mathscr{B}_{\mathbb{R}}$.

> **Propriété.**
> *$X^{-1}(\mathscr{B}_{\mathbb{R}})$ est une tribu sur Ω; elle est appelée la tribu des événements engendrés par X.*

Démonstration. On doit démontrer les trois points de la définition d'une tribu pour $X^{-1}(\mathscr{B}_{\mathbb{R}})$.

(i) $\Omega = X^{-1}(\mathbb{R}) \in X^{-1}(\mathscr{B}_{\mathbb{R}})$, car $\mathbb{R} \in \mathscr{B}_{\mathbb{R}}$.

(ii) Soit $A \in X^{-1}(\mathscr{B}_{\mathbb{R}})$. Alors :

$$\exists B \in \mathscr{B}_{\mathbb{R}} \quad \text{tel que :} \quad A = X^{-1}(B).$$

D'où :

$$\overline{A} = \overline{X^{-1}(B)} = X^{-1}(\overline{B}).$$

Comme $B \in \mathscr{B}_{\mathbb{R}}$, alors $\overline{B} \in \mathscr{B}_{\mathbb{R}}$ et, par suite, $X^{-1}(\overline{B}) \in X^{-1}(\mathscr{B}_{\mathbb{R}})$, c'est-à-dire :

$$\overline{A} \in X^{-1}(\mathscr{B}_{\mathbb{R}}).$$

(iii) Soit $(A_n)_{n\in\mathbb{N}^*}$ une suite d'événements de $X^{-1}(\mathscr{B}_{\mathbb{R}})$. Alors :

$$\forall n \in \mathbb{N}^*, \quad \exists B_n \in \mathscr{B}_{\mathbb{R}} \ \text{ tel que :} \quad A_n = X^{-1}(B_n).$$

D'où :

$$\bigcup_{n\in\mathbb{N}^*} A_n = \bigcup_{n\in\mathbb{N}^*} X^{-1}(B_n) = X^{-1}\Big(\bigcup_{n\in\mathbb{N}^*} B_n\Big).$$

Comme $B_n \in \mathscr{B}_{\mathbb{R}}$, $\forall n \in \mathbb{N}^*$, alors : $\bigcup\limits_{n\in\mathbb{N}^*} B_n \in \mathscr{B}_{\mathbb{R}}$.

D'où :

$$X^{-1}\Big(\bigcup_{n\in\mathbb{N}^*} B_n\Big) \in X^{-1}(\mathscr{B}_{\mathbb{R}}), \quad \text{c'est-à-dire :} \ \bigcup_{n\in\mathbb{N}^*} A_n \in X^{-1}(\mathscr{B}_{\mathbb{R}}).$$

□

Nous sommes maintenant en mesure de définir la notion de probabilité introduite par A. N. Kolmogorov en 1933.

7. Probabilité

7.1. Notion de probabilité

Considérons une expérience aléatoire telle qu'il soit possible de la répéter un très grand nombre de fois dans des conditions identiques et indépendantes les unes des autres. Soit A un événement associé à cette expérience. Si, lors de n répétition de l'expérience, A s'est produit k fois exactement, on dira que k est sa *fréquence absolue*, et que k/n est sa *fréquence relative*.
Quand n devient très grand, on peut constater expérimentalement que k/n se stabilise autour d'une valeur bien déterminée. Le nombre $p(A)$ ainsi mis en évidence s'appellera probabilité de l'événement A.
À partir des fréquences relatives, on voit alors que $p(A)$ possède les propriétés suivantes :

(*i*) $0 \leqslant p(A) \leqslant 1$,

(*ii*) $A \subset B \implies p(A) \leqslant p(B)$,

(*iii*) $p(\Omega) = 1$,

(*iv*) $A \cap B = \varnothing \implies p(A \cup B) = p(A) + p(B)$.

Si l'on se limite à ce point de vue, on est donc amené à définir la théorie des probabilités comme l'étude de la répétition d'expériences ou d'observations effectuées sous des conditions invariantes. La probabilité d'un événement apparaît alors comme une constante physique. Cette interprétation dite fréquentiste (ou objectiviste) de la notion de probabilité a été développée par Von Mises (1931).
Tout cela nous amène à la définition suivante (Kolmogorov, 1933) :

> **Définition 7.**
> Soit $(\Omega, \mathscr{A})$ un espace probabilisable. On appelle probabilité sur $(\Omega, \mathscr{A})$ toute application $P : \mathscr{A} \to [0\,;1]$ telle que :

(i) $P(\Omega) = 1$

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

$$P\left(\bigcup_{n\in\mathbb{N}^*} A_n\right) = \sum_{n\in\mathbb{N}^*} P(A_n) \quad \text{(propriété dite de } \sigma\text{-additivité).}$$

Il est utile de remarquer que, pour une expérience aléatoire décrite par un espace probabilisable $(\Omega, \mathscr{A})$, il existe un grand nombre de probabilités P possibles. Le choix de cette probabilité résulte d'hypothèses faites sur l'expérience aléatoire, ou est elle-même une hypothèse dont les conséquences théoriques seront à confronter avec les résultats expérimentaux.

Définition 8.
Le triplet $(\Omega, \mathscr{A}, P)$ s'appelle un espace probabilisé.

Exemple 1 (Le cas fini)**.**
Soit Ω un ensemble fini muni de la tribu $\mathscr{P}(\Omega)$ de ses parties. Soit P l'application définie sur $\mathscr{P}(\Omega)$ à valeurs dans $[0\,;1]$ telle que :

$$P(A) = \frac{\operatorname{Card} A}{\operatorname{Card} \Omega}, \quad \forall A \in \mathscr{P}(\Omega).$$

Il est clair que $(\Omega, \mathscr{P}(\Omega), P)$ est un espace probabilisé. Dans cet exemple, les événements élémentaires $\{\omega\}$ sont mesurables et équiprobables au sens où l'on a :

$$P(\{\omega\}) = \frac{1}{\operatorname{Card} \Omega}, \quad \forall \omega \in \Omega.$$

Le calcul des probabilités sur l'espace $(\Omega, \mathscr{P}(\Omega), P)$ se ramène alors à des problèmes de dénombrement. Ces problèmes font l'objet de l'*analyse combinatoire* dont nous supposerons connus les résultats essentiels.

Le couple $(\mathbb{R}, \mathscr{B}_\mathbb{R})$ est lui-aussi un espace probabilisable. Cela peut ici se faire de manière tout à fait naturelle, en considérant une variable aléatoire $X : (\Omega, \mathscr{A}, P) \to (\mathbb{R}, \mathscr{B}_\mathbb{R})$ et en posant :

$$\forall B \in \mathscr{B}_\mathbb{R} \quad , \quad P_X(B) = P(X \in B) = P(X^{-1}(B)).$$

P_X est une application définie sur $(\mathbb{R}, \mathscr{B}_\mathbb{R})$, à valeurs dans $[0\,;1]$ telle que :

(i) $P_X(\mathbb{R}) = P(X \in \mathbb{R}) = P(\Omega) = 1$.

(ii) Pour toute suite $(B_n)_{n\in\mathbb{N}^*}$ d'éléments de $\mathscr{B}_\mathbb{R}$ deux à deux disjoints, $(X \in B_n)_{n\in\mathbb{N}^*}$ est une suite d'éléments de $\mathscr{A}$ deux à deux disjoints, donc :

$$\begin{aligned} P_X\left(\bigcup_{n\in\mathbb{N}^*} B_n\right) &= P\left(X \in \bigcup_{n\in\mathbb{N}^*} B_n\right) &&= P\left(\bigcup_{n\in\mathbb{N}^*} \{X \in B_n\}\right) \\ &= P\left(\bigcup_{n\in\mathbb{N}^*} X^{-1}(B_n)\right) &&= P\left(X^{-1}\left(\bigcup_{n\in\mathbb{N}^*} B_n\right)\right) \\ &= \sum_{n\in\mathbb{N}^*} P(X \in B_n) &&= \sum_{n\in\mathbb{N}^*} P_X(B_n) \end{aligned}$$

et alors P_X est une probabilité sur $(\mathbb{R}, \mathscr{B}_\mathbb{R})$.

Définition 9.
P_X est appelée loi de probabilité de X.

Le schéma complet de modélisation se présente alors finalement sous la forme : $(\Omega, \mathbb{A}, P) \xrightarrow{X} (\mathbb{R}, \mathscr{B}_\mathbb{R}, P_X)$.

7.2. Propriétés des probabilités

À partir de la définition d'une probabilité, on peut en déduire un certain nombre de relations, très utiles en pratique, entre les probabilités d'union, de sous-ensembles, de complémentaires, etc., comme le montre le théorème qui suit.

Théorème 2.
Soit $(\Omega, \mathscr{A}, P)$ un espace probabilisé. Alors :

a) $P(\varnothing) = 0$.

b) Si A et B sont deux événements disjoints, alors :
$$P(A \cup B) = P(A) + P(B).$$

c) Si A et B sont deux événements quelconques, alors :
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad \textit{(formule de Poincaré)}.$$

d) Si A et B sont deux événements tels que : $A \subset B$, alors :
$$P(B \setminus A) = P(A) - P(B).$$

e) $\forall A \in \mathscr{A}, \quad P(A^c) = 1 - P(A)$.

f) Si A et B sont deux événements tels que : $A \subset B$, alors :
$$P(A) \leqslant P(B).$$

g) Si A et B sont deux événements quelconques, alors :
$$P\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) \leqslant \sum_{n \in \mathbb{N}^*} P(A_n) \quad \textit{(propriété de } \sigma\textit{-additivité)}.$$

h) Si $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'événements tels que : $A_n \downarrow \varnothing$, alors :
$$P(A_n) \downarrow 0.$$

i) Si $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'événements tels que : $A_n \downarrow A$, alors :
$$P(A_n) \downarrow P(A).$$

j) Si $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'événements tels que : $A_n \uparrow A$, alors :
$$P(A_n) \uparrow P(A).$$

Démonstration. a) Soit $A_1 = \Omega$, et $A_i = \varnothing$, $\forall i \geqslant 2$. Ces événements sont deux à deux disjoints :
$$\forall i \neq j \quad , \quad A_i \cap A_j = \varnothing.$$

Par conséquent,
$$P(\Omega) = P\left(\bigcup_{n \in \mathbb{N}^*}\right) = P(A_1) + \sum_{n=2}^{+\infty} P(A_n).$$

Donc : $\sum_{n=2}^{+\infty} P(A_n) = 0$. Comme, pour tout n, $0 \leqslant P(A_n) \leqslant 1$, on conclut alors que : $P(\varnothing) = 0$.

b) Soit $A_1 = A$, $A_2 = B$, $A_i = \varnothing$, $\forall i \geqslant 3$. Ces événements sont deux à deux disjoints, et $P(A_i) = 0$, $\forall i \geqslant 3$. Donc :
$$P(A \cup B) = P\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) = \sum_{n \in \mathbb{N}^*} P(A_n) = P(A) + P(B).$$

De la même manière, si $A_1,\dots,A_n$ sont n événements deux à deux disjoints, alors :

$$P(A_1\cup A_2\cup\dots\cup A_n)=P(A_1)+\dots+P(A_n).$$

c) On a :

$$\begin{aligned}&A=(A\cap B)\cup(A\cap B^c)\\&B=(A\cap B)\cup(A^c\cap B)\\&(A\cup B)=(A\cap B)\cup(A^c\cap B)\cup(A\cap B^c).\end{aligned}$$

A, B et $A\cup B$ sont écrits sous la forme de réunions d'événements deux à deux disjoints. Donc, d'après (b) :

$$\begin{aligned}&P(A)=P(A\cap B)+P(A\cap B^c)\\&P(B)=P(A\cap B)+P(A^c\cap A)\\&P(A\cup B)=P(A\cap B)+P(A^c\cap B)+P(A\cap B^c).\end{aligned}$$

Alors, on a :

$$\begin{aligned}P(A\cup B)&=P(A\cap B)+P(A\cap B^c)+P(A\cap B)+P(A^c\cap B)-P(A\cap B)\\&=P(A)+P(B)-P(A\cap B).\end{aligned}$$

On peut généraliser à n événements quelconques $A_1,\dots,A_n$:

$$P\left(\bigcup_{i=1}^n A_i\right)=\sum_{i=1}^n P(A_i)-\sum_{1\leqslant i<j\leqslant n}P(A_i\cap A_j)+\sum_{1\leqslant i<j<k\leqslant n}P(A_i\cap A_j\cap A_k)+\dots+(-1)^{n+1}P\left(\bigcap_{i=1}^n A_i\right)$$

appelée formule de Poincaré pour n événements.

d) $B=A\cup(B\setminus A)$ (réunion disjointe). Donc :

$$P(B)=P(A)+P(B\setminus A).$$

D'où :

$$P(B\setminus A)=P(B)-P(A).$$

e) En particulier, en posant $B=\Omega$, on a :

$$P(A^c)=P(\Omega)-P(A)=1-P(A).$$

f) Si $A\subset B$, de (d), on tire que $P(B\setminus A)=P(B)-P(A)\geqslant 0$, soit :

$$P(B)\geqslant P(A).$$

g) La suite : $B_1=A_1,B_2=A_2\cap A_1^c,\dots,B_n=A_n\cap A_{n-1}^c\cap\dots\cap A_1^c,\dots$ est formée d'événements deux à deux disjoints, et telle que :

$$\forall n\in\mathbb{N}^*,\quad B_n\subset A_n\quad\text{et}\quad\bigcup_{n\in\mathbb{N}^*}A_n=\bigcup_{n\in\mathbb{N}^*}B_n.$$

En effet, clairement on a :

$$\bigcup_{n\in\mathbb{N}^*}B_n\subset\bigcup_{n\in\mathbb{N}^*}A_n.$$

De plus, pour tout ω dans $\bigcup\limits_{n\in\mathbb{N}^*}A_n$, il existe n_0 le plus petit entier tel que $\omega\in A_{n_0}$. Alors, *a fortiori*, $\omega\in B_{n_0}\implies\omega\in\bigcup_{n\in\mathbb{N}^*}B_n$. Ainsi :

$$\bigcup_{n\in\mathbb{N}^*}A_n\subset\bigcup_{n\in\mathbb{N}^*}B_n.$$

Donc :

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

h) La notation $A_n\downarrow\varnothing$ signifie que les événements sont emboîtés :

$$A_1\supset A_2\supset A_3\supset\dots\supset A_n\supset A_{n+1}\supset\dots\quad\text{et que}\quad\bigcap_{n\in\mathbb{N}^*}A_n=\varnothing.$$

Alors la suite $(A_n \setminus A_{n+1})_{n\in\mathbb{N}^*}$ est formée d'événements deux à deux disjoints, dont la réunion est A_1.
La série $\displaystyle\sum_{n=1}^{+\infty} P(A_n \setminus A_{n+1}) = P\left(\bigcup_{n\in\mathbb{N}^*}(A_n \setminus A_{n+1})\right) = P(A_1)$ est donc convergente.
Donc son reste d'ordre k tend vers 0 quand $k \to +\infty$.
Or ce reste est : $\displaystyle\sum_{n=k}^{+\infty} P(A_n \setminus A_{n+1}) = P(A_k)$.

La décroissance des $P(A_n)$ provient de la propriété (f) précédente.

i) La notation $A_n \downarrow A$ signifie que les événements sont emboîtés :

$$A_1 \supset A_2 \supset A_3 \supset \cdots \supset A_n \supset A_{n+1} \supset \ldots \text{ et que } \bigcap_{n\in\mathbb{N}^*} A_n = A.$$

Il suffit donc de considérer la suite $(A_n \setminus A)_{n\in\mathbb{N}^*}$ pour se ramener au cas précédent :

$$P(A_n \setminus A) \downarrow 0.$$

Et comme, pour tout n de $\mathbb{N}^*$, $A \subset A_n$, alors : $P(A_n) \downarrow P(A)$.

j) La notation $A_n \uparrow A$ signifie que les événements A_n sont emboîtés :

$$A_1 \subset A_2 \subset A_3 \subset \cdots \subset A_n \subset A_{n+1} \subset \ldots \text{ et que } \bigcup_{n\in\mathbb{N}^*} A_n = A.$$

Il suffit de considérer la suite $(A \setminus A_n)_{n\in\mathbb{N}^*}$ pour se ramener au cas du point h) :

$$P(A \setminus A_n) \downarrow 0,$$

et, comme pour tout n de $\mathbb{N}^*$, $A_n \subset A$, alors : $P(A_n) \uparrow P(A)$.

□

Les propriétés h), i) et j) du théorème précédent sont dites propriétés de la continuité monotone d'une probabilité.
Le premier théorème de Borel-Cantelli qui suit, et la seconde partie qui se trouve à la fin de ce chapitre, seront très utiles dans l'étude de la convergence presque sûre (concept que nous introduirons plus loin) des variables aléatoires.

> **Lemme 1** (de Borel-Cantelli (première partie)).
> *Si $(A_n)_{n\in\mathbb{N}^*}$ est une suite d'événements telle que :* $\displaystyle\sum_{n=1}^{+\infty} P(A_n) < +\infty$*, alors* $P\left(\limsup_{n\to+\infty} A_n\right) = 0$.

Cela signifie que la probabilité qu'une infinité d'événements A_n se réalise est nulle. Autrement dit, la probabilité pour que tous les événements A_n^c soient réalisés, sauf un nombre fini, est 1 :

$$P\left(\liminf_{n\to+\infty} A_n\right) = 1.$$

Cela ne veut pas dire que $\limsup_{n\to+\infty} A_n$ soit l'événement impossible $\varnothing$, ni que $\liminf_{n\to+\infty} A_n^c$ soit l'événement certain Ω. Ce sont des événements appelés respectivement *événement presque impossible* et *événement presque certain.*

Démonstration. $\forall i \in \mathbb{N}^*$, posons $E_i = \displaystyle\bigcup_{n=i}^{+\infty} A_n$. D'après le point g) du théorème précédent, on a :

$$P(E_i) \leqslant \sum_{n=i}^{+\infty} P(A_n).$$

Or : $\limsup_{n\to+\infty} A_n = \bigcap_{k=1}^{+\infty}\bigcup_{n=k}^{+\infty} A_n \subset \bigcup_{n=i}^{+\infty} A_n$ pour tout $i \in \mathbb{N}^*$.

Donc : $\forall i \in \mathbb{N}^*, \quad 0 \leqslant P\left(\limsup_{n\to+\infty} A_n\right) \leqslant \sum_{n=i}^{+\infty} P(A_n)$.

Et comme la série converge, alors : $P\left(\limsup_{n\to+\infty} A_n\right) = 0$. □

8. Probabilité conditionnée par un événement

8.1. Définition

Soit $(\Omega, \mathscr{A}, P)$ un espace probabilisé, et A_0 un événement de probabilité non nulle, donc : $P(A_0) > 0$. Alors on définit la probabilité conditionnelle de A par rapport à A_0 comme :

$$\begin{aligned} P(A \mid A_0) &= \frac{P(A \cap A_0)}{P(A_0)}, \quad \forall A \in \mathscr{A} \\ &= P^{A_0}(A). \end{aligned}$$

$P^{A_0}(\cdot)$ est une probabilité sur $(\Omega, \mathscr{A})$. En effet :

* $P^{A_0}(\Omega) = \dfrac{P(\Omega \cap A_0)}{P(A_0)} = \dfrac{P(A_0)}{P(A_0)} = 1 \quad A_0 \subset \Omega$;
* Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements deux à deux disjoints, alors :

$$\begin{aligned} P^{A_0}\left(\bigcup_n A_n\right) &= \frac{P\left(\left(\bigcup_n A_n\right) \cap A_0\right)}{P(A_0)} = \frac{P\left(\bigcup_n (A_n \cap A_0)\right)}{P(A_0)} \\ &= \frac{\sum_n P(A_0 \cap A_n)}{P(A_0)} = \sum_n P^{A_0}(A_n). \end{aligned}$$

On peut donc ainsi définir un nouvel espace probabilisé : $(\Omega, \mathscr{A}, P^{A_0})$ ou bien $(A_0, \mathscr{A} \cap A_0, P^{A_0})$.

8.2. Formules de Bayes (ou théorème de la probabilité des causes)

Cas de deux événements :

On doit à Thomas Bayes (1702-1761) les résultats suivants sur les probabilités conditionnées par des événements :

Théorème 3 (Formule de Bayes pour deux événements)**.**
Soit $(\Omega, \mathscr{A}, P)$ un espace probabilisé, A et B deux événements de $\mathscr{A}$, de probabilités non nulles. Alors :

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A^c)P(A^c)}$$

(on suppose ici aussi que $P(A^c) \neq 0$).

Démonstration. On a immédiatement :

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

Or :

$$P(B) = P((B \cap A) \cup (B \cap A^c)) = P(B \cap A) + P(B \cap A^c).$$

La conclusion est alors aisée. □

Cas de n événements :

Commençons par un premier résultat, connu sous le vocable de « théorème de la probabilité des causes ».

> **Théorème 4** (Théorèmes des probabilités totales)**.**
> *Soit $(\Omega, \mathscr{A}, P)$ un espace probabilisé, et $(A_i)_{i\in\mathbb{N}}$ une partition de Ω. On suppose de plus que, pour tout i, $P(A_i) \neq 0$.*
> *Alors :*
> $$\forall B \in \mathscr{A}, \qquad P(B) = \sum_{i\in\mathbb{N}} P(B \mid A_i)P(A_i).$$

Démonstration. On a :
$$P(B) = P\left(B \cap \left(\bigcup_{i\in\mathbb{N}} A_i\right)\right) = P\left(\bigcup_{i\in\mathbb{N}} (B \cap A_i)\right) = \sum_{i\in\mathbb{N}} P(B \cap A_i),$$
car les A_i sont deux à deux disjoints.
Il vient alors immédiatement que :
$$P(B) = \sum_{i\in\mathbb{N}} P(B/A_i)P(A_i).$$
Clairement, le résultat reste valide si on choisit une partition finie. □

Exemple 2 (Les daltoniens)**.**
On considère une population composée de 48 % d'hommes et 52 % de femmes. Il y a 5 % d'hommes daltoniens et 0,25 % de femmes daltoniennes.
Quelle est la probabilité qu'un individu pris au hasard soit daltonien ?
On notera les événements :

H "être un homme" ; F "être une femme" ; D "être daltonien".

La traduction de l'énoncé donne :
$$P(D/H) = 0{,}05 \quad ; \quad P(D/F) = 0{,}0025$$
$$P(H) = 0{,}48 \quad ; \quad P(F) = 0{,}52.$$
En utilisant le résultat précédent, on a :
$$P(D) = P(D/H)P(H) + P(D/F)P(F) = 0{,}0253.$$
Il y a donc 2,53 % de daltoniens dans la population totale (hommes et femmes confondus).

On peut maintenant énoncer la formule de Bayes pour n événements.

> **Théorème 5.**
> *Soit $(\Omega, \mathscr{A}, P)$ un espace probabilisé, et $(A_i)_{i=1,\dots,n}$ une partition finie de Ω. On suppose de plus que, pour tout i, $P(A_i) \neq 0$, et que B est un événement de probabilité non nulle. Alors :*
> $$\forall i = 1, \dots, n \qquad P(A_i/B) = \frac{P(B/A_i)P(A_i)}{\displaystyle\sum_{k=1}^{n} P(B/A_k)P(A_k}.$$

Démonstration. Il suffit de reprendre le même schéma que dans la démonstration de la formule de Bayes pour deux événements, en utilisant pour le dénominateur le théorème des probabilités totales. □

Exemple 3 (Les filles en amphi)**.**
Dans un amphi, il y a 20 % d'élèves de 19 ans, 50 % de 20 ans, 20 % de 21 ans, 10 % de 22 ans, et, dans ces tranches d'âges, il y a respectivement 50 %, 20 %, 30 %, 40 % de filles.
Si on appelle une fille au hasard, quelle est la probabilité qu'elle ait 20 ans ?

On définit : A_1, A_2, A_3, A_4 les événements "avoir 19, 20, 21, 22 ans" et B par "être une fille". On cherche ici : $P(A_2/B)$.
La traduction de l'énoncé donne :

$$P(B/A_1) = 0{,}5 \qquad P(B/A_2) = 20$$
$$P(B/A_3) = 0{,}3 \qquad P(B/A_4) = 0{,}4 \quad \text{et}$$
$$P(A_1) = 0{,}2 \qquad P(A_2) = 0{,}5 \qquad P(A_3) = 0{,}2 \qquad P(A_4) = 0{,}1.$$

D'où :

$$P(A_2/B) = \frac{0{,}2 \times 0{,}5}{0{,}5 \times 0{,}2 + 0{,}2 \times 0{,}5 + 0{,}3 \times 0{,}2 + 0{,}4 \times 0{,}1} = \frac{1}{3}.$$

Interprétation : Si l'amphi compte 100 étudiants, il y en a :
20 de 19 ans, dont $20 \times 50\,\% = 10$ filles
50 de 20 ans, dont $50 \times 20\,\% = 10$ filles
20 de 21 ans, dont $20 \times 30\,\% = 6$ filles
10 de 22 ans, dont $10 \times 40\,\% = 4$ filles,
et sur les 30 filles de l'amphi, 10, soit $1/3$ d'entre elles, ont 20 ans.

9. Indépendance

9.1. Indépendance de deux événements

Nous avons défini la probabilité conditionnelle de A sachant B (deux événements de probabilité non nulle) par :

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

On pourrait bien entendu définir l'indépendance de A et B, en disant que la probabilité de A sachant B ne dépend pas de B ; le fait de savoir B réalisé ou non n'a pas d'influence sur la probabilité de A :

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

Mais cette définition de l'indépendance de ces deux événements nécessite l'hypothèse de non nullité des probabilités de A et de B respectivement. De l'égalité précédente, on tire :

$$P(A \cap B) = P(A)P(B).$$

C'est cette dernière égalité qui nous servira de définition de l'indépendance pour deux événements, car elle ne nécessite pas l'hypothèse de non nullité d'événements.

Définition 10.
Soit $(\Omega, \mathscr{A}, P)$ un espace probabilisé, A et B deux événements. On dira que A et B sont indépendants pour la probabilité P si et seulement si :

$$P(A \cap B) = P(A) \cdot P(B).$$

Il faut noter que la notion d'indépendance n'est pas intrinsèque aux événements, mais dépend de la probabilité P choisie sur $(\Omega, \mathscr{A})$. Deux événements indépendants pour une probabilité donnée peuvent ne plus l'être pour une autre probabilité, et réciproquement. Donnons enfin une propriété pour deux événements indépendants.

Propriété.
Soit $(\Omega, \mathscr{A}, P)$ un espace probabilisé, A et B deux événements indépendants, alors on a :

(i) A^c et B sont indépendants.

> *(ii)* A *et* B^c *sont indépendants.*
>
> *(iii)* A^c *et* B^c *sont indépendants.*

Démonstration. (i) Il faut montrer que $P(A^c \cap B) = P(A^c)P(B)$. On a clairement : $B = B \cap \Omega = B \cap (A \cup A^c) = (B \cap A) \cup (B \cap A^c)$ (réunion disjointe).
D'où : $P(B) = P(B \cap A) + P(B \cap A^c)$.
Par suite : $P(A^c \cap B) = P(B) - P(A)P(B) = P(B)[1 - P(A)] = P(B)P(A^c)$.
(ii) et (iii) se démontrent de la même façon...

□

9.2. Indépendance de n événements

On généralise aisément au cas de n événements.

> **Définition 11.**
> Soit $(A_1, \ldots, A_n)$ un n-uple d'événements. On dit qu'ils sont indépendants, ou mutuellement indépendants, si et seulement si, pour tout $k = 1, \ldots, n$, si pour tout sous-ensemble $(A_{i_1}, \ldots, A_{i_k})$ de k événements choisis parmi les $(A_1, \ldots, A_n)$, on a :
> $$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1} \times \cdots \times P(A_{i_k}).$$

Exemple 4 (Indépendance deux à deux, mais pas indépendance mutuelle).
Soit $(\Omega, \mathscr{A}, P)$ un espace probabilisé défini par : $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$; $\mathscr{A} = \mathscr{P}(\Omega)$ et $P\omega_1 = P\omega_2 = P\omega_3 = P\omega_4 = 1/4$.
Soient $A_1 = \{\omega_1, \omega_4\}$; $A_2 = \{\omega_2, \omega_4\}$; $A_3 = \{\omega_3, \omega_4\}$.
Il est aisé de vérifier que A_1 et A_2 sont indépendants, A_2 et A_3 sont indépendants, A_1 et A_3 sont indépendants, mais que A_1, A_2, A_3 ne sont pas indépendants, puisque :
$$P(A_1 \cap A_2 \cap A_3) = P(\{\omega_4\}) = \frac{1}{4} \quad \text{et} \quad P(A_1) \times P(A_2) \times P(A_3) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}.$$
On dit que A_1, A_2, A_3 sont indépendants deux à deux, mais pas mutuellement indépendants.

9.3. Indépendance d'une suite d'événements

> **Définition 12.**
> Soit $(\Omega, \mathscr{A}, P)$ un espace probabilisé, soit