Thème 3 : ensembles, espaces de probabilités finis

Serge Cohen, Monique Pontier, Pascal J. Thomas

Septembre 2004

1 Généralités : ensembles et parties d'un ensemble

Définition 1.1 *On appelle* **ensemble** *une collection finie ou infinie d'objets, mathématiques ou non.*
Tout objet de cette collection s'appelle **élément** *de l'ensemble.*
Une partie de l'ensemble s'appelle un **sous-ensemble**.

1.1 Appartenance et inclusion

Prenons par exemple l'ensemble des étudiants du groupe, noté C ; chaque étudiant est un élément de C, on dit encore que "chaque étudiant appartient à C". Par exemple, si Laure est une étudiante du groupe C, on dit que "Laure appartient à C" et on note Laure $\in C$.

La partie du groupe composé des filles, F, est un sous-ensemble ; on note $F \subset C$ et on dit que F est inclus dans C. Remarquons de plus que

$$\text{Laure } \in F.$$

Si G est le sous-ensemble composé des garçons, Laure $\notin G$.

ATTENTION : l'appartenance et l'inclusion sont deux notions très différentes !!

Définition 1.2 *On dira qu'un ensemble B* est inclus dans *A, ou de façon équivalente que B* est une partie de *A, ou que B* est un sous-ensemble de *A, noté $B \subset A$, si et seulement si : n'importe quel élément de B est aussi un élément de A.*

1.2 Parties d'un ensemble

Soit $\mathbb{N}$ l'ensemble des entiers naturels. La collection des parties (ou sous-ensembles) de $\mathbb{N}$ peut être considérée comme un ensemble, noté $\mathcal{P}(\mathbb{N})$, dont les éléments sont eux-mêmes des ensembles (c'est un ensemble d'ensembles).

Soit par exemple l'ensemble des entiers pairs,

$$P = \{2k, k \in \mathbb{N}\} \in \mathcal{P}(\mathbb{N}).$$

P est un sous-ensemble. Mais attention !!

$$P \subset \mathbb{N},$$

P est un **élément** de $\mathcal{P}(\mathbb{N})$, et un **sous-ensemble** de $\mathbb{N}$!

On note $P_i = \{ik, k \in \mathbb{N}\} \in \mathcal{P}(\mathbb{N})$, l'ensemble des multiples de i où i est un entier non-nul donné.

Soit $\mathcal{C} = \{P_1, \cdots, P_{10}\}$: $\mathcal{C} \subset \mathcal{P}(\mathbb{N})$. C'est un sous-ensemble de $\mathcal{P}(\mathbb{N})$.

Soit $\{2, 4, 6\}$: c'est une partie de $\mathbb{N}$, $\{2, 4, 6\} \subset \mathbb{N}$, mais c'est un élément de $\mathcal{P}(\mathbb{N})$: $\{2, 4, 6\} \in \mathcal{P}(\mathbb{N})$.

Définition 1.3 *Si E est un ensemble, il existe un ensemble, appelé* **ensemble des parties de** E *noté* $\mathcal{P}(E)$, *dont les éléments sont tous les ensembles inclus dans E. Il vérifie donc pour tout ensemble F :*

$$F \in \mathcal{P}(E) \text{ équivaut à } F \subset E.$$

1.3 Quelques sous-ensembles remarquables

On considère des parties d'un ensemble donné E. Remarquons d'abord que l'ensemble tout entier, E, est une partie de lui-même (le mot "partie" n'est pas à comprendre ici au sens strict du langage courant !)

On admet le résultat suivant : il existe un ensemble, appelé **ensemble vide**, noté $\emptyset$, qui ne contient aucun élément, et qui est sous-ensemble de E, et en fait de n'importe quel ensemble. Remarquons donc qu'il est inclus dans tout autre sous-ensemble de E.

Les sous-ensembles ne contenant qu'un seul élément s'apellent des **singletons** et sont notés (comme tous les ensembles) entre accolades : si E et un ensemble et si $x \in E$,

$$\{x\} \text{ note le singleton qui ne contient que l'élément } x \text{ de } E.$$

Attention ! Le singleton $\{x\}$ est distinct de l'élément x; on a

$$2 \in \mathbb{N} \; ; \; \{2\} \in \mathcal{P}(\mathbb{N}).$$

1.4 Méthode générale

Un ensemble est défini par une propriété caractéristique de tous ses éléments ; notons la P, c'est à dire

$$E_P = \{x \text{ tel que } P \text{ est vraie pour } x\}.$$

Un même ensemble peut être défini par une autre propriété Q. Pour voir qu'il s'agit vraiment du même ensemble, cela signifie qu'il faut montrer l'équivalence :

$$x \text{ vérifie } P \text{ si et seulement si } x \text{ vérifie } Q.$$

Une manière d'arriver à ceci est de faire semblant qu'on a deux ensembles distincts, E_P et

$$E_Q = \{x \text{ tel que } Q \text{ est vraie pour } x\},$$

puis de montrer que tout élément de E_P vérifie en fait Q, et aussi que tout élément de E_Q vérifie P. La démarche est donc la suivante : pour montrer que $E_P = E_Q$, on montre que $E_P \subset E_Q$ *et* que $E_Q \subset E_P$. C'est une méthode très habituelle pour montrer que deux ensembles coïncident.

Exemples.

$$\begin{aligned}
E &= \{x \in \mathbb{N} : x^2 - 3x + 2 = 0\} = \{1, 2\}, \\
F &= \{x \in \mathbb{N} : \text{ le dernier chiffre de } x \text{ est } 0, 2, 4, 6, \text{ ou } 8\} = \{2k, k \in \mathbb{N}\}, \\
G &= \{x \in \mathbb{N} : \text{ la somme des chiffres de } x \text{ est divisible par } 3\} = \{3k, k \in \mathbb{N}\}.
\end{aligned}$$

1.5 Tous, aucun, certains

Le langage mathématique a besoin pour fonctionner d'être très rigoureux. Quand on dit que $A \subset E_P$ signifie que "tous les éléments de A satisfont la propriété P", il faut comprendre que (a) les éléments de A, sans aucune exception, vérifient P et que (b) il peut très bien y avoir des éléments qui vérifient la propriété P, mais ne sont pas dans A (puisqu'on n'a **pas** dit que "$E_P \subset A$").

Par exemple : soit A l'ensemble des nombres premiers (c'est-à-dire des nombres entiers p strictement positifs tels que p ne soit divisible que par 1 et par p), et I l'ensemble des nombres entiers impairs. Soit A_1 l'ensemble des nombres premiers supérieurs ou égaux à 3. Alors $A_1 \subset I$.

Par contre le nombre $2 \in A$ et 2 n'est pas impair, donc A n'est pas inclus dans I (noté $A \not\subset I$). Vous voyez qu'il suffit d'une seule exception pour qu'on ne puisse pas affirmer "tous les éléments de A vérifient P", autrement dit, la négation de "tous les éléments de A vérifient P" est : "il existe (au moins) un élément de A qui ne vérifie pas P". Il important de se souvenir de cet usage de la négation.

On peut écrire tout ce qui précède avec des symboles conventionnels appelés *quantificateurs*, dont nous essaierons de ne pas abuser.

Par exemple $\forall$ se lit "pour tout" ou "quel que soit" :

$A_1 \subset E_P$ est équivalent à $\forall x \in A_1$, x vérifie P.

$\exists$ se lit "il existe" :

$A \not\subset E_P$ est équivalent à $\exists x \in A$, x ne vérifie pas P.

1.6 Exercices

1. Soit l'ensemble $\mathbb{N}$ et l'ensemble $\mathcal{P}(\mathbb{N})$. Remplacer les pointillés par l'un des symboles $\in, \subset$, ou $\notin$.

$$2 \ldots \mathbb{N} \ ; \ 2 \ldots \mathcal{P}(\mathbb{N}) \ ; \ \{2\} \ldots \mathcal{P}(\mathbb{N}).$$

$$\{2,4,6\} \ldots \mathbb{N} \ ; \ \{2,4,6\} \ldots . \mathcal{P}(\mathbb{N}).$$

2. Soit E un ensemble à deux éléments : $E = \{a, b\}$, et F un ensemble à quatre éléments : $F = \{d, e, f, g\}$. Décrire totalement l'ensemble des parties de E, noté $\mathcal{P}(E)$, et l'ensemble des parties de F, noté $\mathcal{P}(F)$. Dire combien d'éléments ont chacun de ces deux ensembles de parties.

3. Soient E et F deux ensembles.
3.1 Montrer que $E \subset F$ équivaut à $\mathcal{P}(E) \subset \mathcal{P}(F)$.
3.2 En déduire que $E = F$ équivaut à $\mathcal{P}(E) = \mathcal{P}(F)$.

4. Soit C l'ensemble des étudiants du groupe. On note C_1 le sous-ensemble de ceux qui ont un ordinateur à la maison, C_2 le sous-ensemble de ceux qui habitent Toulouse, F le sous-ensemble des filles et G le sous-ensemble des garçons.
Que signifie (en français !!) :

$$F \subset C_2 \ ; \ \text{non } [C_1 \subset G].$$

5. Soit E l'ensemble des fonctions continues sur les réels, à valeurs réelles. On note $A = \{f \in E \text{ telle que } f(0) = 0\}$. Soit g une fonction sur les réels, à valeurs réelles. Traduire $g \notin A$.

6. On reprendre les notations de l'exemple du paragraphe 1.2
6.1 Montrer que $P_1 = \mathbb{N}$.
6.2 Déterminer l'ensemble des éléments de $\mathbb{N}$ qui sont à la fois dans P_6 et P_4.

2 Opérations sur les sous-ensembles

On considère un ensemble E, l'ensemble de ses parties $\mathcal{P}(E)$, et les sous-ensembles de E : A, B, éléments de $\mathcal{P}(E)$.

Définition 2.1 *On appelle* **union** *de A et B notée $A \cup B$ l'ensemble des éléments de E qui sont soit dans A, soit dans B, soit dans les deux.*

$$A \cup B = \{x \in E \text{ tels que } x \in A \text{ ou } x \in B\}.$$

On appelle **intersection** *de A et B notée $A \cap B$ l'ensemble des éléments de E qui sont à la fois dans A et dans B.*

$$A \cap B = \{x \in E \text{ tels que } x \in A \text{ et } x \in B\}.$$

On appelle **complémentaire** *de A, noté selon les auteurs $\complement A, A^c, \overline{A}$, l'ensemble des éléments de E qui ne sont pas dans A.*

$$\overline{A} = \{x \in E \text{ tels que } x \notin A\}.$$

On définit parfois d'autres opérations à partir de celles qui précèdent. Par exemple, la différence de deux ensembles, qui est l'ensemble de tous les éléments de A qui ne sont pas dans B,

$$A \setminus B = A \cap \overline{B};$$

ou la différence symétrique,

$$A \Delta B = (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B),$$

qui est l'ensemble de tous les éléments qui sont soit dans A, soit dans B, mais pas dans les deux à la fois (vérifiez à titre d'exercice que les définitions données sont bien équivalentes à cela).

Il est parfois utile de définir précisément le fait qu'un ensemble est découpé en un certain nombre de parties qui ne se recoupent pas, comme l'ensemble des communes françaises (métropolitaines) peuvent être réparties selon le département où elles sont situées, par exemple.

On dit que deux ensembles A et B sont *disjoints* si et seulement si $A \cap B = \emptyset$, c'est-à-dire qu'il n'ont aucun élément en commun.

Définition 2.2 *Soit E un ensemble et $A_1, A_2, \ldots, A_n$ des sous-ensembles de E. On dit que $\{A_1, A_2, \ldots, A_n\}$ forme* une partition *de E si et seulement si*

1. $E = A_1 \cup A_2 \cup \cdots \cup A_n$

2. Les sous-ensembles $A_1, \ldots, A_n$ *sont deux à deux disjoints, c'est-à-dire : pour tout* $i \neq j$, $A_i \cap A_j = \emptyset$.

exemples.
1. Soient A_1 l'ensemble des nombres pairs, A_2 l'ensemble des nombres impairs, $\{A_1, A_2\}$ forme une partition de $\mathbb{N}$.
2. Etant donnée une partie quelconque $A \subset E$, $\{A, \overline{A}\}$ forme une partition de E.
3. Pour $i = 0, 1, 2, \ldots, 9$, soit Q_i l'ensemble des entiers naturels dont le dernier chiffre (dans l'écriture décimale) est i. Alors $\{Q_0, Q_1, \cdots, Q_9\}$ forme une partition de $\mathbb{N}$.

2.1 Propriétés élémentaires

à montrer en exercice :

$$A \text{ et } B \subset A \cup B \ ; \ A \cap B \subset A \text{ et } B.$$

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \ ; \ \overline{A \cap B} = \overline{A} \cup \overline{B}.$$

$$A \subset B \text{ équivaut à } A \cap B = A.$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \ ; \ A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

2.2 Exercices

1. Soit A et $B \in \mathcal{P}(E)$; comparer

$$\mathcal{P}(A \cup B) \text{ et } \mathcal{P}(A) \cup \mathcal{P}(B),$$

$$\mathcal{P}(A \cap B) \text{ et } \mathcal{P}(A) \cap \mathcal{P}(B).$$

2. On reprend l'exercice 4. de la section 1. Que signifie

$$C_1 \cap F \neq \emptyset \ ; \ C_2 \cup G = C \ ; \ C_1 \cap C_2 \subset G \ ?$$

3. Montrer que

$$A \cup B = A \cap C \text{ équivaut à : } B \subset A \text{ et } A \subset C.$$

3 Applications

Soient E et F deux ensembles, on va définir la notion d'*application* de E dans F (ce n'est pas tout-à-fait une fonction).

Définition 3.1 *Une application de E dans F, notée $f : E \longrightarrow F$ est une correspondance : à* tout *élément x de E, f fait correspondre un élément et un seul y de F, on note cet élément $y = f(x)$. L'application s'appelle f, on note aussi $x \mapsto f(x)$.*
On dit que E est l'ensemble de départ *de f et que F est* l'ensemble d'arrivée *de f.*

Définition 3.2 *Quand f est une application de E dans F et que $x \in E$, on dit que $f(x)$ est l'*image *de x par f.*
Si $y \in F$, et qu'on trouve $x \in E$ tel que $f(x) = y$, x est un antécédent *(ou une pré-image) de y par f.*

Attention! Certains éléments y n'ont pas d'antécédents ; d'autres peuvent en avoir plusieurs.

Exemples avec $E = F = \mathbb{R}$ (l'ensemble des nombres réels):

1. $x \mapsto \sin x$, ici f est l'application "sinus". L'élément $y = 2$ n'admet aucun antécédent. L'élément $y = 0$ admet une infinité d'antécédents (tous les multiples entiers de π).
2. $x \mapsto \exp(x) = e^x$, ici f est l'application "exponentielle".
3. $x \mapsto x^2 + 1$. Ici f est l'application "prendre le carré et ajouter 1". Un malencontreux désir d'abréviation fait qu'on pourra entendre des gens parler d'"application x^2+1", ce qui est incorrect : $x^2 + 1$ n'est que l'image de l'élément x par l'application qui à x fait correspondre $x^2 + 1$. L'élément $y = 0$ n'admet aucun antécédent. L'élément $y = 2$ admet deux antécédents, $+1$ et -1.
4. Ici $E = \mathbb{R}^* = R \setminus \{0\}$: $x \mapsto 1/x$.
5. $x \mapsto x^3$.

Dans le cas des applications d'une partie de $\mathbb{R}$ dans $\mathbb{R}$, comme le sont tous les exemples ci-dessus, on trouve les antécédents géométriquement à partir du graphe de f en traçant une droite horizontale d'ordonnée y et en prenant les abscisses $x_1, x_2 \ldots$ de ses points d'intersection avec le graphe de f.

Définition 3.3 *Une application f de E dans F est dite* injective *si tout $y \in F$ admet* au plus *un antécédent dans E. On dit aussi dans ce cas que f est une* injection

Exemples : parmi les exemples donnés ci-dessus, $x \mapsto \exp(x)$ et $x \mapsto x^3$ sont des injections de $\mathbb{R}$ dans $\mathbb{R}$; $x \mapsto 1/x$ est une injection de $\mathbb{R}^*$ dans $\mathbb{R}$.

Etre injectif dépend du choix de E : l'application de $\mathbb{R}$ dans $\mathbb{R}$ donnée par $x \mapsto x^2+1$ n'est pas injective. Mais si on prend $E = [0, +\infty[$, l'application de E dans $\mathbb{R}$ donnée par $x \mapsto x^2+1$ est injective.

Traduction pratique de la définition : si x_1, x_2 sont des éléments quelconques de E tels que $x_1 \neq x_2$, alors on doit avoir $f(x_1) \neq f(x_2)$. On peut aussi l'exprimer en disant que

$$\forall x_1, x_2 \in E, \text{ si } f(x_1) = f(x_2) \text{ alors nécessairement } x_1 = x_2.$$

En effet, l'élément $y = f(x_1) = f(x_2) \in F$ a au plus un antécédent. Or x_1 et x_2 sont par hypothèse des antécédents de y, donc ce devait être en fait le même : $x_1 = x_2$.

Définition 3.4 *Une application f de E dans F est dite* surjective *si tout $y \in F$ admet* au moins *un antécédent dans E. On dit aussi dans ce cas que f est une* surjection.

Exemples : parmi les exemples donnés ci-dessus, $x \mapsto x^3$ est une surjection de $\mathbb{R}$ dans $\mathbb{R}$. Les autres ne sont pas surjectives.

Définition 3.5 *Une application f de E dans F est dite* bijective *si elle est à la fois injective et surjective, c'est-à-dire si tout $y \in F$ admet* exactement *un antécédent dans E. On dit aussi dans ce cas que f est une* bijection.

Encore des exemples :

1. $\sin :]-\frac{\pi}{2}, +\frac{\pi}{2}[\longrightarrow \mathbb{R}$ est injective, car sin est strictement croissante sur $]-\frac{\pi}{2}, +\frac{\pi}{2}[$ (sa dérivée, cos, y est strictement positive). Donc si $x_1 \neq x_2$, alors soit on a $x_1 < x_2$, donc $\sin x_1 < \sin x_2$ donc $\sin x_1 \neq \sin x_2$, soit $x_2 < x_1$ et on peut faire le même raisonnement en intervertissant leurs rôles.
2. $\sin :]-\frac{\pi}{2}, +\frac{\pi}{2}[\longrightarrow \mathbb{R}$ n'est pas surjective, par exemple $y = 2$ n'admet aucun antécédent.
3. $\sin :]-\frac{\pi}{2}, +\frac{\pi}{2}[\longrightarrow]-1, +1[$ est bijective.

4 Application inverse, composition d'applications, images réciproques

4.1 Application inverse

Définition 4.1 *Soit f une bijection de E dans F, alors pour tout $y \in F$, y possède un unique antécédent $x \in E$ tel que $f(x) = y$. On appelle* application inverse *de f et on note*

f^{-1} l'application qui admet F pour ensemble de départ et E pour ensemble d'arrivée et telle que $f^{-1}(y) = x$ (autrement dit, $f^{-1}(f(x)) = x$).
On a l'équivalence

$$\big(x = f^{-1}(y), y \in F\big) \quad \Longleftrightarrow \quad (y = f(x), x \in E)\,.$$

Exemple : Si $E = \mathbb{R}$, $F =]0, +\infty[$, l'application exponentielle $x \mapsto e^x$ est une bijection, et son application inverse, qui a pour ensemble de départ $]0, +\infty[$ et pour ensemble d'arrivée $\mathbb{R}$, est donnée par $y \mapsto \ln y$.

Attention ! Quand a est un nombre réel non nul, la notation a^{-1} désigne le réel $1/a$. On peut donc trouver la notation $(f(x))^{-1}$ pour désigner l'inverse du nombre réel $f(x)$, mais il ne faut pas la confondre avec l'application inverse. Par exemple, on pourrait écrire (mais on ne le fera pas en général, car cela prête à confusion !)

$$(e^x)^{-1} = \frac{1}{e^x} = e^{-x},$$

mais cela n'a rien à voir avec le nombre $\ln x$ (si $x > 0$).

4.2 Composition d'applications

L'opération de composition entre f et g consiste à appliquer d'abord f, puis g. Il faut bien entendu que les ensembles d'arrivée de f et de départ de g soient le même.

Définition 4.2 *Soient trois ensembles E, F et G, et deux applications $f : E \longrightarrow F$ et $g : F \longrightarrow G$. On appelle* (application) composée *de f et de g et on note $g \circ f$ l'application $g \circ f : E \longrightarrow G$ telle que $x \mapsto (g \circ f)(x) = g(f(x))$.*

Attention ! La notation $g \circ f$ semble se lire de droite à gauche : c'est f que l'on applique en premier, et g en deuxième. La raison de cette notation réside dans la formule ci-dessus. Et l'ordre est important, cf. exemples ci-dessous.

Exemples. Dans tout ce qui suit, on prend $E = F = G = \mathbb{R}$.

1. Soient f définie par $x \mapsto f(x) = x + 1$, et g définie par $x \mapsto g(x) = x^2$. Alors $(g \circ f)(x) = (x+1)^2 = x^2 + 2x + 1$ et $(f \circ g)(x) = x^2 + 1 \neq (g \circ f)(x)$ (quand $x \neq 0$).
2. Soient f définie par $x \mapsto f(x) = \sin x$, et g définie par $x \mapsto g(x) = 1 - x^2$. Alors $(g \circ f)(x) = \cos^2 x$ (que l'on devrait écrire $(\cos x)^2$, en toute rigueur !), et $(f \circ g)(x) = \sin(1 - x^2)$. On voit que $(g \circ f)(0) = 1$, et que $(f \circ g)(0) = \sin 1 < \frac{\sqrt{3}}{2} < 1$. Donc à nouveau les deux applications $f \circ g$ et $g \circ f$ sont différentes (pour que deux applications soient égales, il faut qu'elles aient même ensembles de départ E et d'arrivée F, et qu'elles fassent correspondre la même valeur à *tout* élément $x \in E$. Donc dès qu'on trouve ne serait-ce qu'un seul $x \in E$ tel que $f \circ g(x) \neq g \circ f(x)$, on conclut que $f \circ g \neq g \circ f$, rappelez-vous le paragraphe 1.5).

Définition 4.3 *Etant donné un ensemble E, l'application qui a pour ensemble départ E et pour ensemble d'arrivée E et qui est donnée par $x \mapsto x$ pour tout $x \in E$ est appelée* application identité *et est notée Id ou Id_E.*

Exercice. Soit f une application de E dans F. Vérifiez que $f \circ Id_E = f$, et que $Id_F \circ f = f$.

Proposition 4.4 *Si f est une bijection de E dans F, alors $f^{-1} \circ f = Id_E$ et $f \circ f^{-1} = Id_F$.*

Démonstration. Si $y = f(x)$, $f^{-1} \circ f(x) = f^{-1}(f(x)) = f^{-1}(y) = x$, car il n'y a qu'un antécédent possible pour y.

Pour la composition dans l'autre sens, $f^{-1}(y)$ doit être un antécédent de y (le seul, en fait), donc $f \circ f^{-1}(y) = f(f^{-1}(y)) = y$. $\square$

4.3 Image et image réciproque d'un ensemble

A partir d'un application f, on définit deux applications, une de $\mathcal{P}(E)$ dans $\mathcal{P}(F)$, et l'autre de $\mathcal{P}(F)$ dans $\mathcal{P}(E)$, notées par abus f et f^{-1}, et ceci même quand f n'est pas bijective et donc que f^{-1} n'a pas de sens, n'existe pas, comme application de F dans E.

Il faut retenir que cette f^{-1} est en fait plus facile à manipuler que f (quand on les applique à des sous-ensembles) ; c'est f^{-1} qui intervient implicitement quand on écrit l'ensemble des solutions d'une équation, ou d'un système d'équations, par exemple.

Définition 4.5 *Soit une application f de E dans F, et $B \subset F$. Alors on appelle* image réciproque *de B et on note $f^{-1}(B)$ l'ensemble de tous les antécédents des éléments $y \in B$, autrement dit*

$$f^{-1}(B) = \{x \in E : f(x) \in B\}.$$

Soit $A \subset E$, on appelle image (directe) *de A et on note $f(A)$ l'ensemble de toutes les images des éléments $x \in A$, autrement dit*

$$f(A) = \{y \in F : \textit{ il existe } x \in A \textit{ tel que } f(x) = y\}.$$

Remarque. Si f est une bijection, alors la notation $f^{-1}(B)$ semble avoir deux sens possibles : l'image réciproque de B par f, ou l'image directe de B par l'application bien définie f^{-1}. Heureusement, ces deux sens coïncident.

Exemples. Dans tout ce qui suit, on prend $E = F = \mathbb{R}$.

1. $x \mapsto f(x) = x^2 + 1$. Alors $f([0, 1]) = [1, 2]$, $f^{-1}([1, 2]) = [-1, +1]$, $f^{-1}([0, 2]) = [-1, +1]$.
2. $x \mapsto f(x) = x^2 - 1$. Alors $f([0, 1]) = [-1, 0]$, $f^{-1}(\{0\}) = \{-1, +1\}$, $f^{-1}(\{-2\}) = \emptyset$.
3. $x \mapsto f(x) = \sin x$. Alors $f([0, \frac{\pi}{2}]) = [0, 1]$. Par contre,
$$f^{-1}([0, 1]) = \dots [-4\pi, -3\pi] \cup [-2\pi, -\pi] \cup [-\pi, +\pi] \cup [2\pi, 3\pi] \cup [4\pi, 5\pi] \dots$$

4.4 Exercices

1. Soient f et g deux bijections, respectivement de E dans F et de F dans G. Montrer que $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

On suppose maintenant que f et g sont seulement deux applications, respectivement de E dans F et de F dans G. Soient $A \subset E$ et $B \subset G$. Montrer que $g(f(A)) = (g \circ f)(A)$ et que $f^{-1}(g^{-1}(B)) = (g \circ f)^{-1}(B)$.

2. Dire si les assertions suivantes sont vraies, auquel cas, les montrer. Si elles sont fausses, donnner un contre-exemple, et préciser lorsqu'une égalité a été proposée si l'une des inclusions est quand même vraie.

$$f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B) \tag{1}$$
$$f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B) \tag{2}$$
$$f^{-1}(\overline{A}) = \overline{f^{-1}(A)} \tag{3}$$

où A et $B \subset F$.

De même, est ce que ??

$$f(A \cup B) = f(A) \cup f(B) \tag{4}$$
$$f(A \cap B) = f(A) \cap f(B) \tag{5}$$
$$f(\overline{A}) = \overline{f(A)} \tag{6}$$

où A et $B \subset E$.

3. Soit $f : \mathbb{R} \to \mathbb{R}, x \mapsto \sin x$. Définir $f(\mathbb{R})$, $f([0, \pi])$, $f([-\pi/2, \pi/2])$, $f^{-1}([2, +\infty))$, $f^{-1}([0, 1])$, $f^{-1}([-6, 0])$.

5 Combinatoire

5.1 Généralités sur le dénombrement

Définition 5.1 *Soit E un ensemble fini, comprenant n éléments, on dit que n est son* cardinal*, on le note card$E$ ou $\#E$.*

Dénombrer E, c'est trouver le cardinal de E.

Pour trouver le cardinal de E, il faut expliquer une méthode systématique qui permette de trouver *tous* les éléments de E; sans compter aucun élément plusieurs fois.

Proposition 5.2 *Soit E un ensemble fini, et F un ensemble tel qu'il existe une bijection de E dans F. Alors card$E$ = card$F$.*

Soient E un ensemble fini et A une partie de E. Alors card$A \leq$ card$E$.

Démonstration. Si f est une bijection de E dans F, alors pour tout $y \in F$, comme f est surjective, on peut trouver un antécédent x dans E, et un seul x ne peut avoir au plus qu'une image $f(x)$, donc il y a au moins autant d'éléments dans E que dans F; mais comme f est injective, un élément $y \in F$ a au plus un antécédent dans E, donc il n'y a pas plus d'éléments dans E que dans F.

Si $A \subset E$, les éventuels éléments de E qui ne sont pas dans A sont en plus dans E, donc il y a plus d'éléments dans E que dans A. □

5.2 Dénombrements classiques

Proposition 5.3 *Si E est un ensemble fini de cardinal n, on note*

$$E^p := \{(x_1, \dots, x_p), x_i \in E, 1 \leq i \leq p\},$$

*l'ensemble des p-*uplets *d'éléments de E (les x_i sont choisis arbitrairement dans E, on peut répéter le même élément, et l'ordre est important).*

Alors card$E^p = n^p$.

Démonstration. Pour construire un p-uplet, on a n choix pour x_1. Puisqu'on admet les répétitions, pour chaque choix de x_1, on a encore n choix pour x_2, donc au total $n \times n$ choix pour (x_1, x_2). Pour chacun de ces choix, on a à nouveau n choix pour x_3, et ainsi de suite : on a au total $n \times n \times \cdots \times n = n^p$ choix. □

Exemples.

1. $E = \{0, 1, 2\}$ et $p = 2$. On donne la liste des $3^2 = 9$ possibilités pour (x_1, x_2) en présentant d'abord tous les cas où $x_1 = 0$, puis tous ceux où $x_1 = 1$, etc.

$$(0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (1, 2), (2, 0), (2, 1), (2, 2).$$

2. Combien y a-t-il de nombres qui s'écrivent avec au plus trois chiffres en écriture binaire ? Comme les nombres à moins de trois chiffres correspondent à des nombres avec un ou plusieurs "0" au début, on a ici le cas où $E = \{0, 1\}$, donc card$E = 2$, et $p = 3$. Il y en a donc $2^3 = 8$ (les octets font trois *bits* !). Voici la liste, qui correspond à la liste des entiers de 0 à 7, écrits en binaire:

$$000, 001, 010, 011, 100, 101, 110, 111.$$

Proposition 5.4 *Si E est un ensemble fini de cardinal n, et $p \leq n$, on appelle* arrangement *à p éléments de E une liste ordonnée $(x_1, \ldots, x_p)$ d'éléments de E tous distincts.*

Le nombre d'arrangements à p éléments de E est noté A_n^p et vaut

$$A_n^p = \frac{n!}{(n-p)!} = n(n-1)(n-2)\cdots(n-p+1).$$

Démonstration. On a n choix pour x_1. Mais pour chaque choix de x_1, on n'a plus que $n-1$ choix pour x_2, puis $n-2$ choix pour x_3 une fois que x_1 et x_2 ont été choisis, et ainsi de suite jusqu'à x_p pour lequel on a $n-(p-1) = n-p+1$ choix. Comme avant, on a donc $n \times (n-1)$ choix pour (x_1, x_2), $n \times (n-1) \times (n-2)$ choix pour (x_1, x_2, x_3), et finalement

$$A_n^p = n(n-1)(n-2)\cdots(n-p+1).$$

Et comme $n! = n \times (n-1) \times (n-2) \cdots \times (n-p+1) \times (n-p) \times \cdots \times 2 \times 1$, et $n! = (n-p) \times \cdots \times 2 \times 1$, on voit finalement que

$$A_n^p = \frac{n!}{(n-p)!}.$$

□

Remarques.

1. Le nombre d'arrangements de n éléments d'un ensemble de cardinal n est $A_n^n = n!$. On énumère tous les éléments de E en mettant des numéros, c'est donc le nombre de manières de numéroter E, ou de classer E (au sens des classements des participants à une compétition sportive, par exemple).

2. Comme on le voit dans l'exemple précédent, l'ordre est donc important aussi pour les arrangements, et si par exemple $E = \{a, b, c, d\}$, les deux arrangements (a, b) et (b, a) sont différents et doivent être comptés 1 chacun.

Proposition 5.5 *Si E est un ensemble fini de cardinal n, et $p \leq n$, le nombre de sous-ensembles de E qui ont p éléments (c'est-à-dire $A \subset E$ et $\mathrm{card}A = p$) est noté C_n^p ou $\begin{pmatrix} n \\ p \end{pmatrix}$. On parle aussi du nombre de* combinaisons *à p éléments dans un ensemble à n éléments.*

Alors

$$C_n^p = \frac{A_n^p}{p!} = \frac{n!}{p!(n-p)!} = \frac{n(n-1)(n-2)\cdots(n-p+1)}{p!}.$$

Remarque. On adopte toujours la convention $0! = 1$. Ainsi

$$C_n^n = \frac{n!}{0!n!} = 1,$$

(le seul sous-ensemble de E de cardinal égal à celui de E est E lui-même), et

$$C_n^n = \frac{n!}{n!0!} = 1,$$

(le seul sous-ensemble de E de cardinal égal à 0 est l'ensemble vide).

Démonstration. Pour les sous-ensembles, l'ordre ne compte pas. Si on reprend l'exemple précédent avec $E = \{a, b, c, d\}$, les notations $\{a, b\}$ et $\{b, a\}$ représentent le même sous-ensemble. Donc il y a plusieurs arrangements possibles correspondant à un même sous-ensemble. Plus précisément, étant donné un sous-ensemble $\{x_1, \ldots, x_p\}$, tous les arrangements mettant en jeu ces mêmes p éléments s'obtiennent en changeant l'ordre (en renumérotant) l'ensemble $\{x_1, \ldots, x_p\}$. Nous venons de voir qu'il y avait $p!$ manières de faire cela. Donc à chaque combinaison (sous-ensemble) de p éléments de E correspondent $p!$ arrangements, et le nombre de sous-ensembles est celui d'arrangements, divisé par $p!$. (Cela prouve au passage que la formule de la proposition donne toujours un nombre entier, ce qui n'était pas évident). □

5.3 Lien avec le cardinal de l'ensemble des injections, des surjections, des bijections.

Ce paragraphe se fait uniquement sous la forme d'exercices.

1. Soient E et F des ensembles finis ($\#E = p$, $\#F = n$) et $\mathcal{A}(E, F)$ l'ensemble des applications de E dans F.
1.1 Montrer que $\#\mathcal{A}(E, F) = n^p$.
1.2 Montrer que $\#\mathcal{P}(E) = 2^p$.
1.3 Montrer que le nombre d'injections de E dans F est égal à A_n^p. Quand $n = p$, quel est le nombre des bijections de E dans F ?

2. On suppose que $p \leq n$ et que E et F sont tous les deux munis d'un ordre total.

Calculer le nombre d'applications strictement croissantes de E dans F.

3. On suppose que $n = 2 \leq p$. Calculer le nombre de surjections de E dans F.

6 Introduction aux Espaces de probabilité

Expériences aléatoires.

On parle d'*expérience aléatoire* quand les résultats de l'expérience dépendent du hasard.

Expérience 1. On jette un dé cubique et on lit le numéro de la face supérieure.

Expérience 2. On jette deux fois un dé et on note les numéros obtenus. (Remarque : peut-on distinguer les dés ?)

Evénements.

Les événements sont les résultats d'une expérience qui peuvent se produire ou non (à distinguer de l'expérience elle-même).

Expérience 1. On peut considérer par exemple l'événement A_1 "le numéro obtenu est pair"; ou l'événement A_2 "le numéro obtenu est supérieur ou égal à 4".

Expérience 2. On peut considérer par exemple l'événement B "la somme des deux numéros obtenus est égale à 6".

Deux événements particuliers.

L'événement *impossible* I n'est jamais réalisé (dans l'expérience 1, on pourrait le définir par "le numéro obtenu est égal à 7"... ou de bien d'autres façons).

L'événement *certain* C est toujours réalisé (dans l'expérience 1, on pourrait le définir par "le numéro obtenu est un entier compris entre 1 et 6").

Evénements et ensembles.

Nous allons interpréter les événements décrits ci-dessus comme des ensembles.

Pour l'expérience 1, posons $\Omega = \{1, 2, 3, 4, 5, 6\}$ et notons $\omega \in \Omega$ un numéro d'une face de dé. Alors

$$A_1 = \{2, 4, 6\}, \quad A_2 = \{4, 5, 6\}, \quad I = \emptyset, \quad C = \Omega.$$

Pour l'expérience 2, posons

$$\Omega = \{1, 2, 3, 4, 5, 6\}^2 = \{(\omega_1, \omega_2), \omega_1, \omega_2 \text{ entiers compris entre 1 et 6}\}.$$

Il s'agit ici de couples ordonnés (de 2-uplets) parce qu'on distingue le premier tirage du deuxième tirage du dé (ou, de façon équivalente, qu'on tire deux dés de couleurs différentes, qui permettent de les individualiser). Alors

$$B = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}.$$

Plus généralement, à chaque expérience aléatoire, on associe un ensemble Ω (qui n'est pas unique) tel que chaque événement observable puisse être décrit comme une partie de Ω. L'ensemble de tous les événements possibles est $\mathcal{P}(\Omega)$; en particulier, l'événement impossible est $\emptyset$, l'événement certain est Ω.

Dictionnaire.

L'événement contraire de A (celui qui se produit exactement quand A ne se produit pas) correspond au complémentaire $\overline{A}$.

Exemples : Le contraire de l'événement certain est l'événement impossible, $\overline{\Omega} = \emptyset$.

Dans l'expérience 1 ci-dessus, "le numéro obtenu est impair" correspond à l'ensemble $\{1,3,5\}$ qui est le complémentaire dans Ω de $A_1 = \{2,4,6\}$.

Si A et B sont deux événements, "A et B se produisent tous les deux" correspond à $A \cap B$.

Quand A et B sont deux événements tels que $A \cap B = \emptyset$ (donc ils ne se produisent jamais simultanément, comme dans le cas d'un événement et de son contraire), on dit que ces deux événements sont *incompatibles*.

Si A et B sont deux événements, "A ou B se produit (un ou l'autre, ou tous les deux)" correspond à $A \cup B$.

Si A et B sont deux événements, on dira que A entraîne B si dans tout les cas où A se produit, alors B se produit, c'est-à-dire que $\omega \in A \Rightarrow \omega \in B$, donc $A \subset B$.

Exemple : dans l'expérience 1, si le nombre obtenu est pair, alors il est supérieur ou égal à 2 : $\{2,4,6\} \subset \{2,3,4,5,6\}$.

Pour tout $\omega \in \Omega$, on appelle le singleton $\{\omega\}$ un événement élémentaire.

Approche intuitive des probabilités.

Si on a une expérience aléatoire qu'on peut répéter dans des conditions 'dentiques" et que A est un événement, on peut considérer la fréquence de A : au bout de n répétitions de l'expérience, si l'événement A s'est produit r fois, on pose $f_n(A) := r/n \in [0,1]$.

Supposons que $f_n(A)$ admette une limite quand $n \to \infty$ (les observations faites dans la pratique suggèrent fortement que c'est le cas pour beaucoup d'expériences typiques, comme les jeux de dés). On appelle cette limite "probabilité de A", et on la note $P(A)$.

Exemples.

1. $f_n(\Omega) = 1$, $f_n(\emptyset) = 0$ pour tout n. Donc $P(\Omega) = 1$, $P(\emptyset) = 0$.

2. Si un événement ne se produit pas, son contraire se produit, et réciproquement ; donc on a toujours $f_n(A) + f_n(\overline{A}) = 1$, donc $P(A) + P(\overline{A}) = 1$.

3. Dans le cas de l'expérience 1, si le dé est bien équilibré, on peut se convaincre que $P(A_1) = P(\overline{A_1}) = \frac{1}{2}$.

7 Espace probabilisé fini

7.1 Cas général

Définition 7.1 *Soit Ω un ensemble fini, $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω. On appelle* probabilité finie *sur $(\Omega, \mathcal{P}(\Omega))$ toute application P de $\mathcal{P}(\Omega)$ dans $[0, 1]$ vérifiant*

1. $P(\Omega) = 1$ et

2. Pour tout A et B tels que $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$.

Le triplet $(\Omega, \mathcal{P}(\Omega), P)$ est appelé *espace de probabilité fini*; pour tout événement A (c'est-à-dire pour tout sous-ensemble A de Ω), le réel positif (ou nul) $P(A)$ est appelé *probabilité* de l'événement A.

Remarques.

1. On pourrait rendre le point (2) plausible à partir d'un raisonnement sur les fréquences, mais dorénavant nous le prenons comme une définition.

2. On déduit facilement du point (2) que si $A_1, A_2, \dots, A_p$ sont des événements deux à deux disjoints, c'est-à-dire qui vérifient $A_i \cap A_j = \emptyset$ pour tout $i \neq j$, alors

$$P(A_1 \cup A_2 \cup \cdots \cup A_p) = P(A_1) + P(A_2) + \cdots P(A_p).$$

Proposition 7.2 *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini, et $A, B \in \mathcal{P}(\Omega)$, alors*

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

2. Si $A \subset B$, alors $P(A) \leq P(B)$,

3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Démonstration.

(1) On a $\Omega = A \cup \overline{A}$, et $A \cap \overline{A} = \emptyset$, donc

$$1 = P(\Omega) = P(A) + P(\overline{A}), \text{ d'où } P(\overline{A}) = 1 - P(A).$$

(2) De $\Omega = A \cup \overline{A}$, on déduit que $B = B \cap \Omega$ (tous les éléments de B sont déjà dans Ω par définition de B), c'est-à-dire $B = (B \cap A) \cup (B \cap \overline{A})$ (tous les éléments de B sont soit dans A, soit hors de A).

Mais on voit que $(B \cap A) \cap (B \cap \overline{A}) = \emptyset$ (un élément ne peut pas être à la fois dans A et hors de A), donc on peut appliquer le point (2) de la définition :

$$P(B) = P(B \cap A) + P(B \cap \overline{A}) \geq P(B \cap A), \tag{7}$$

puisque $P(B \cap \overline{A}) \geq 0$ d'après la définition (une probabilité d'événement est toujours comprise entre 0 et 1).

(3) D'après la décomposition de B faite ci-dessus,

$$A \cup B = A \cup (B \cap A) \cup (B \cap \overline{A}) = (A \cup (B \cap A)) \cup (B \cap \overline{A}) = A \cup (B \cap \overline{A}),$$

où la dernière égalité découle du fait que tous les éléments de $B \cap A$ sont déjà dans A, donc n'ajoutent rien dans l'union $A \cup (B \cap A)$. L'avantage de cette écriture est que nous avons désormais une décomposition en sous-ensembles *disjoints* : comme avant, un élément ne peut ne peut pas être à la fois dans A et hors de A, donc

$$A \cap (B \cap \overline{A}) = A \cap \overline{A} \cap B = \emptyset.$$

Il faut encore calculer $P(B \cap \overline{A})$. Or on déduit de la première égalité de (7) que $P(B \cap \overline{A}) = P(B) - P(B \cap A)$, et finalement

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

Remarque : les rôles de A et B dans cette formule sont bien entendus symétriques ; on peut construire une démonstration plus "symétrique" en découpant chacun des deux ensembles, $A = (A \cap B) \cup (A \cap \overline{B})$ et $(B \cap A) \cup (B \cap \overline{A})$, et en recollant les morceaux de façon appropriée. $\square$

7.2 Equiprobabilité

Définition 7.3 *On dit qu'il y a* équiprobabilité *lorsque les probabilités de tous les singletons ou événements élémentaires sont égales. On dit aussi que P est la* probabilité uniforme *sur $(\Omega, \mathcal{P}(\Omega))$, ou que tous les singletons sont* équiprobables.

Remarque. C'est une hypothèse pour simplifier, ce n'est pas la seule possible !

Théorème 7.4 *Soit* $(\Omega, \mathcal{P}(\Omega))$ *un espace de probabilité fini muni de la probabilité uniforme* P. *Pour tout événement* A, *on a :*

$$P(A) = \frac{\mathrm{card}A}{\mathrm{card}\Omega} = \frac{\text{“\textit{nombre de cas favorables}”}}{\text{“\textit{nombre de cas possibles}”}}.$$

Démonstration. Si $\mathrm{card}\Omega = n$,

$$\Omega = \{\omega_1, \ldots, \omega_n\} = \{\omega_1\} \cup \cdots \cup \{\omega_n\},$$

donc

$$1 = P(\Omega) = P(\{\omega_1\}) + \cdots + P(\{\omega_n\}) = nP(\{\omega_i\}),$$

pour n'importe quel indice i variant entre 1 et n, donc $P(\{\omega_i\}) = \frac{1}{n}$ pour tout i. Si un sous-ensemble $A \subset \Omega$ comporte exactement p éléments, on peut, quitte à renuméroter les ω_i, l'écrire

$$A = \{\omega_1, \ldots, \omega_p\} = \{\omega_1\} \cup \cdots \cup \{\omega_p\},$$

donc

$$P(A) = P(\{\omega_1\}) + \cdots + P(\{\omega_p\}) = \frac{p}{n} = \frac{\mathrm{card}A}{\mathrm{card}\Omega}.$$

□

Exemples.

1. Expérience 1 : on suppose le dé parfaitement équilibré, et donc les événements élémentaires sont équiprobables. Pour tout $i \in \Omega = \{1, 2, 3, 4, 5, 6\}$, nous avons $P(\{i\}) = \frac{1}{6}$, et si nous considérons à nouveau $A_1 = \{2, 4, 6\}$, alors $P(A_1) = \frac{3}{6} = \frac{1}{2}$.

2. Expérience 2 : cette fois-ci,

$$\mathrm{card}\Omega = \mathrm{card}\{1, 2, 3, 4, 5, 6\}^2 = 6^2 = 36,$$

donc pour tout couple $(i, j) \in \Omega$, $P\{(i, j)\} = \frac{1}{36}$. Pour l'événement $B = \{\text{la somme des dés } = 6\}$, $\mathrm{card}B = 5$, donc $P(B) = \frac{5}{36}$.

3. Si on interprète l'expérience 2 avec des dés indistingables (on les lance en même temps, ils ont la même couleur...), alors on considère que $(5, 1) = (1, 5)$ par exemple, et

$$\Omega = \big\{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6), (1,2), (1,3), (1,4), (1,5),$$
$$(1,6), (2,3), (2,4), (2,5), (2,6), (3,4), (3,5), (3,6), (4,5), (4,6), (5,6)\big\},$$

et on voit que $\mathrm{card}\Omega = 6 + \frac{1}{2}(36 - 6)$, ou encore $\mathrm{card}\Omega = 6 + C_6^2 = 21$. Dans ce modèle, l'événement B aura pour éléments $(1, 5), (2, 4), (3, 3)$, et sa probabilité serait $\frac{3}{21} = \frac{1}{7} > \frac{5}{36}$.

Dans ce cas, un rapide raisonnement permet de se convaincre que ce deuxième modèle ne correspond pas à la réalité expérimentale (la couleur des dés peut-elle influer leur mouvement ?), et l'expérience physique confirme que c'est le modèle précédent qui est correct (a-t-on autant de chances de sortir une somme égale à 2 qu'une somme égale à 3 ? Essayez !).

On peut obtenir un modèle valide physiquement en ne considérant pas les singletons de notre Ω comme équiprobables, mais que

$$P(i,j) = \frac{1}{36} \text{ si } i = j, \quad P(i,j) = \frac{2}{36} \text{ si } i \neq j.$$

On obtiendra alors les mêmes probabilités quand le modèle précédent pour les événements qui peuvent être définis dans les deux modèles (c'est-à-dire ceux qui ne dépendent pas de l'ordre des dés).

Ce qui est plus intéressant, c'est que dans des situations de physique des particules, mettant en jeu des objets qui ne peuvent pas être perçus directement (comme des photons), on peut faire des expériences qui permettent de calculer des probabilités empiriques. On peut ensuite voir si elles correspondent, soit au modèle "particules distingables", soit au modèle "particules indistingables" (et on a la surprise, dans certains cas, de constater que des particules peuvent êtres indistingables en ce sens).

4. Expérience du tirage d'une boule dans une boîte comprenant 10 boules rouges et 15 boules blanches ; on a deux manières de voir les choses:

   (i) $\Omega = \{$ blanc, rouge$\}$. Dans ce cas il est clair que les singletons ne sont pas équiprobables, et que l'on a : $\mathbb{P}(\{\text{blanc}\}) = 2/5$; $\mathbb{P}(\{\text{rouge}\}) = 3/5$.

   (ii) Mais on peut aussi choisir de définir Ω comme l'ensemble des tirages équiprobables de chacune 25 boules. Alors, $\Omega = \{b_1, ..., b_{10}, r_1, ..., r_{15}\}$ et pour tout i et tout j $\mathbb{P}(\{r_i\}) = \mathbb{P}(\{b_j\}) = 1/25$.

8 Probabilité conditionnelle

8.1 Définition

La notion de probabilité *conditionnelle* intervient chaque fois que l'on dispose d'une information supplémentaire qui restreint le champ du possible, et qui permet (par exemple) de réviser ses pronostics sur la réalisation ou non d'un événement.

Par exemple : supposons qu'on joue au tiercé (pari sur les noms des trois premiers chevaux dans une course, très populaire en France entre les années 1950 et 1980) et qu'on apprenne que le cheval numéro 10, blessé, doit se retirer de la course au dernier moment. La probabilité de toutes les combinaisons de trois chevaux contenant le 10 devient instantanément nulle.

De même, si on joue aux dés et qu'on lance deux dés, un après l'autre, on sait si on obtient un 1 au premier lancer que la probabilité d'avoir une somme supérieure ou égale à 8 devient nulle, et que la probabilité d'avoir une somme supérieure ou égale à 7 est sans doute fortement réduite.

En effet, si on connaît déjà le résultat du premier dé (1), on peut raisonner sur l'espace de probabilité Ω_1 correspondant au lancement d'un seul dé (le deuxième en l'occurrence). Pour avoir une somme supérieure ou égale à 7, il faut que le résultat du deuxième dé soit supérieur ou égal à 6, donc égal à 6, donc la probabilité de l'événement est $\frac{1}{6}$. Au contraire, quand on n'a encore lancé aucun dé, il y a $6+5+4+3+2+1 = 21$ couples $(i, j) \in \Omega = \{1, 2, 3, 4, 5, 6\}^2$ qui donnent une somme supérieure ou égale à 7, et la probabilité de l'événement est donc $\frac{21}{36} = \frac{7}{12}$.

Pour donner un modèle de telles situations, on considère que l'ensemble plus réduit de possibilités, disons A, qui nous est laissé par l'information supplémentaire est un sous-ensemble de l'espace de probabilité original. Ce $A \subset \Omega$ est un événement en soi, mais nous le considérons comme un nouvel espace de probabilité plus restreint, dans lequel A devient l'événement certain. Dans l'exemple ci-dessus, on avait

$$\Omega = \{1, 2, 3, 4, 5, 6\}^2 = \big\{(1, 1), (1, 2), \ldots, (1, 6), (2, 2) \ldots (2, 6), \ldots (6, 1), \ldots (6, 6)\big\},$$

alors qu'une fois qu'on sait qu'un 1 est sorti pour le premier dé, on peut se restreindre à

$$A := \big\{(1, 1), (1, 2), \ldots, (1, 6)\big\} \subset \Omega.$$

Remarquez que A n'est pas égal au Ω_1 donné ci-dessus, mais qu'ils sont en bijection (il suffit de connaître le résultat du deuxième dé pour tout savoir).

Si on sait déjà que le premier dé donne 1, l'événement A, "le premier dé donne 1" devient certain. L'événement "la somme des dés est supérieure ou égale à 7" devient, une fois que A est réalisé, "le premier dé donne 1 et la somme des dés est supérieure ou égale à 7" : c'est l'intersection de deux événements.

Tout cela motive la définition suivante.

Définition 8.1 *Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace de probabilité fini, et A un événement de probabilité non nulle. Pour tout événement $B \subset \Omega$, on appelle* probabilité conditionnelle de B sachant A, *le nombre*

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

On note aussi, souvent, $P_A(B) = P(B|A)$.

Proposition 8.2 *Le triplet $(\Omega, \mathcal{P}(\Omega), P_A)$ définit un espace de probabilité.*

C'est aussi vrai pour le triplet $(A, \mathcal{P}(A), P_A)$.

Démonstration. Le premier point de la définition 7.1 explique pourquoi nous avons divisé par $P(A)$. En effet,

$$P_A(\Omega) = \frac{P(A \cap \Omega)}{P(A)} = \frac{P(A)}{P(A)} = 1.$$

Si maintenant nous avons deux événements $B_1, B_2 \subset \Omega$ tels que $B_1 \cap B_2 = \emptyset$, alors $(A \cap B_1) \cap (A \cap B_2) = A \cap B_1 \cap B_2 = \emptyset$, donc

$$\begin{aligned} P_A(B_1 \cup B_2) = \frac{P(A \cap (B_1 \cup B_2))}{P(A)} &= \frac{P((A \cap B_1) \cup (A \cap B_2))}{P(A)} \\ &= \frac{P(A \cap B_1) + P(A \cap B_2)}{P(A)} = \frac{P(A \cap B_1)}{P(A)} + \frac{P(A \cap B_2)}{P(A)}. \end{aligned}$$

Le deuxième point est laissé au lecteur. □

8.2 Probabilités composées, formule des probabilités totales

8.2.1 Probabilités composées

Remarque.

On tire de façon évidente de la définition 8.1 l'égalité suivante :

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

C'est ce qu'on appelle la *formule des probabilités composées.* C'est plus utile qu'il n'y paraît.

Exemple

Soit une boîte contenant 4 boules blanches et 3 boules noires. Quelle est la probabilité de tirer (sans remise) une boule blanche suivie d'une noire ?

On considère l'événement $A =$ "la première boule est blanche" et l'événement $B =$ "la deuxième boule est noire". Une fois A réalisé, on est ramené à une boite avec trois boules blanches et trois boules noires : $P(B|A) = \frac{1}{2}$; par ailleurs, lorsque l'on ne tire qu'une boule il est facile d'obtenir $P(A) = \frac{4}{7}$. En appliquant la "formule des probabilités composées" ci-dessus :

$$P(A \cap B) = \frac{1}{2} \cdot \frac{4}{7} = \frac{2}{7}.$$

8.2.2 Formule des probabilités totales

Poursuivons l'étude de cet exemple, et supposons que nous voulions connaître la probabilité de l'événement B "la deuxième boule tirée de l'urne est noire". Cela peut se produire de deux manières : soit la première boule tirée était blanche (et on se retrouve dans la situation ci-dessus, celle de l'événement $A \cap B$), soit la première boule tirée était noire. Cet événement est $\overline{A}$. Alors il restera dans l'urne, après ce premier tirage, 4 boules blanches et 2 boules noires. Donc

$$P(B|\overline{A}) = \frac{2}{6} = \frac{1}{3},$$

et comme $P(\overline{A}) = 1 - P(A) = \frac{3}{7}$,

$$P(B \cap \overline{A}) = \frac{1}{3} \cdot \frac{3}{7} = \frac{1}{7}.$$

Finalement,

$$P(B) = P(B \cap A) + P(B \cap \overline{A}) = \frac{2}{7} + \frac{1}{7} = \frac{3}{7}.$$

Pour généraliser ce genre de situation, posons une définition :

Définition 8.3 *On appelle* système complet d'événements *une partition de Ω, c'est à dire une famille d'événements $A_1, \ldots, A_n$ deux à deux disjoints (incompatibles), tels que $\Omega = A_1 \cup \cdots \cup A_n$.*

L'idée est que pour calculer la probabilité d'un événement, on va étudier séparément la probabilité dans un nombre fini de cas mutuellement exclusifs, puis tout additionner.

Proposition 8.4 *Soit $A_1, \ldots, A_n$ un système complet d'événements de probabilités non-nulles. Alors pour tout événement B, on a :*

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots P(A_n)P(B|A_n).$$

Remarque. Un cas particulier très utile est celui utilisé ci-dessus, à savoir $n = 2$, donc $A_1 = A$, $A_2 = \overline{A}$.

Démonstration.

Comme $A_1 \cup \cdots \cup A_n = \Omega$,

$$B = B \cap \Omega = B \cap (A_1 \cup \cdots \cup A_n) = (B \cap A_1) \cup \cdots \cup (B \cap A_n),$$

et comme les événements $B \cap A_i$ sont deux à deux disjoints (incompatibles),

$$P(B) = P(B \cap A_1) + \cdots + P(B \cap A_n) = P(A_1)P(B|A_1) + \cdots P(A_n)P(B|A_n),$$

par la formule des probabilités composées appliquée à chaque terme de la somme. □

8.2.3 Exercice

On choisit au hasard un individu dans une population où figure une proportion p de tricheurs ($p \in [0, 1]$) et on lui fait tirer une carte au hasard dans un jeu de 52 cartes. S'il s'agit d'un tricheur, il tirera forcément un as. Quelle la probabilité que l'individu choisi tire un as ?

8.3 Formule de Bayes

On souhaite désormais inverser le processus décrit dans les sous-sections précédentes : au lieu de donner la probabilité d'un événement en fonction de celles d'événements qui ont pu l'influer on voudra, connaissant un événement, en inférer quelles causes ont pu conduire à cet événement.

Supposons que dans l'exemple ci-dessus, la proportion de tricheurs dans la population soit de 5%. Une personne inconnue tire dans le jeu de carte et sort un as. Quelle est la probabilité, sachant cela, que ce soit un tricheur ? Si l'on appelle A l'événement "la personne est un tricheur", et B l'événement "la personne tire un as", nous voyons tout d'abord que

$$P(A) = 0,05, \quad P(B|A) = 1, \quad \text{donc } P(A \cap B) = 0,05.$$

Si par contre la personne n'est pas un tricheur,

$$P(\overline{A}) = 0,95, \quad P(B|A) = \frac{1}{13}, \quad \text{donc } P(A \cap B) = 0,073\ldots$$

Donc la probabilité totale de B est $0,05 + 0,073\ldots = 0,123\ldots$.

Nous voulons maintenant connaître la probabilité de A, sachant B :

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = 0,05/0,123\cdots = 0,407\ldots$$

Notez bien qu'il y a quand même à peu près trois chances sur cinq que la personne ne soit *pas* un tricheur, et ne préparez pas le goudron et les plumes ! (Le résultat serait différent dans une société beaucoup moins honnête, par exemple si on supposait $P(A) = 0,2$).

En général, on peut formaliser ce raisonnement :

Proposition 8.5 (Formule de Bayes) *Soit $A_1, \ldots, A_n$ un système complet d'événements de probabilités non-nulles. Alors pour tout événement de probabilité non-nulle B, pour tout $i \in \{1, \ldots, n\}$, on a :*

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + \cdots P(A_n)P(B|A_n)}.$$

En particulier, si A est un événement tel que $0 < P(A) < 1$ (et donc n'est ni impossible ni certain), alors

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\overline{A})P(B|\overline{A})}.$$

Démonstration. C'est une conséquence immédiate de la formule des probabilités composées, de la formule des probabilités totales, et de la définition des probabilités conditionnelles. □

9 Indépendance d'événements

9.1 Indépendance de deux événements

Intuitivement, on dira que deux événements sont indépendants quand la réalisation de l'un n'influe pas la probabilité que l'autre se réalise. Supposons qu'aucun des deux événements considérés, appelons-les A et B, ne soit de probabilité nulle. Dans le langage des probabilités conditionnelles, l'indépendance se traduit par

$$P(A|B) = P(A) \text{ et } P(B|A) = P(B).$$

La deuxième égalité équivaut à $\frac{P(A \cap B)}{P(A)} = P(B)$, donc $P(A \cap B) = P(A)P(B)$, formulation symétrique qui implique donc aussi la première égalité. Nous prendrons donc cette dernière formule, qui a un sens même quand les probabilités peuvent s'annuler, pour définition.

Définition 9.1 *Deux événements A et B sont* indépendants *si et seulement si*

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

Remarque.

Cette notion est très différente de celles d'événement incompatibles : dans ce cas, la réalisation d'un événement influe énormément sur celle de l'autre, puisqu'elle l'exclut ! On démontre aisément que deux événements ne peuvent être à la fois indépendants et incompatibles que si l'un des deux est de probabilité nulle (auquel cas il est indépendant de tout autre événement, et la notion est de peu d'intérêt).

Proposition 9.2 *Si A et B sont indépendants, alors A et $\overline{B}$, $\overline{A}$ et B, et $\overline{A}$ et $\overline{B}$ sont indépendants.*

Démonstration. Il suffit de montrer la première conclusion, les autres s'en déduisent en échangeant les rôles de A et B. Or

$$P(A \cap \overline{B}) + P(A \cap B) = P(A),$$

donc

$$P(A \cap \overline{B}) = P(A) - P(A \cap B) = P(A)(1 - P(B)) = P(A)P(\overline{B}).$$

□

Exemples.

1. On tire une carte au hasard dans un jeu de 52 cartes. Soit A l'événement "la carte est un as" et soit B l'événement "la carte est un pique". Alors $P(A) = \frac{1}{13}$, $P(B) = \frac{1}{4}$, $P(A \cap B) = \frac{1}{52}$: les deux événements sont indépendants (ici l'expression 'àu hasard" doit s'interpréter mathématiquement par : "toutes les cartes sont équiprobables").

2. On reprend le modèle de l'expérience 2 dans la section 6 (lancer de deux dés),

$$\Omega = \{1, 2, 3, 4, 5, 6\}^2 = \{(i, j), i, j \text{ entiers compris entre 1 et 6}\},$$

   dans lequel les singletons $\{(i, j)\}$ sont tous équiprobables. Soit A l'événement "le premier dé donne un résultat pair" et B l'événement "le deuxième dé donne un résultat égal à 1". Alors A et B sont indépendants. De façon générale, tout événement qui peut se décrire uniquement en termes du résultat du premier dé est indépendant de tout événement qui peut se décrire uniquement en termes du résultat du deuxième dé. Ce type d'exemples se généralise, et est très important.

   Par contre, l'événement A = "le premier dé donne un 1" et l'événement B = "la somme des résultats des deux dés est supérieure où égale à 7" ne sont pas indépendants (nous l'avons vu au début de la section 8).

3. Dans le même modèle, soit A l'événement "le premier dé donne un résultat pair", B l'événement "le deuxième dé donne un résultat pair" et C l'événement "la somme des résultats des deux dés est un nombre pair". Alors on voit que A et B sont indépendants, avec $P(A) = P(B) = \frac{1}{2}$. D'autre part, la somme de deux nombres est paire si et seulement si soit les deux nombres sont pairs, soit les deux nombres sont impairs, donc

$$C = (A \cap B) \cup (\overline{A} \cap \overline{B}),$$

   qui sont des événement disjoints, donc

$$P(C) = P(A \cap B) + P(\overline{A} \cap \overline{B}) = P(A)P(B) + P(\overline{A})P(\overline{B}),$$

   par la proposition 9.2, finalement $P(C) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. D'autre part, $A \cap C = A \cap B$, donc $P(A \cap C) = \frac{1}{4}$: A et C sont indépendants, et on démontre de la même manière que B et C sont indépendants.

Nous avons donc démontré que les trois événements A, B et C sont indépendants deux à deux. Peut-on pourtant dire qu'ils sont indépendants ? Sûrement pas ! Vous voyez immédiatement que si A et B sont réalisés, par exemple, alors C devient certain (en fait, si deux de ces trois événements se réalisent, le troisième en découle).

9.2 Indépendance de familles d'événements

L'exemple qui précède met en évidence la nécessité d'une notion d'indépendance de famille quelconque d'événements.

Définition 9.3 *Une famille d'événements $A_1, \ldots, A_n$ sont dits (mutuellement) indépendants si, quelle que soit la sous famille $A_{i_1}, \ldots, A_{i_p}$ d'événements, avec $2 \leq p \leq n$, alors*

$$P(A_{i_1} \cap \cdots \cap A_{i_p}) = P(A_{i_1}) \times \cdots \times P(A_{i_p}).$$

Dans le cas particulier où $n = 2$, c'est la définition 9.1 (les seuls choix possibles seront $i_1 = 1$ et $i_2 = 2$, ou vice-versa). Dans le cas particulier où $n = 3$, par exemple, cela donnera les conditions suivantes. Pour $p = 2$:

$$P(A_1 \cap A_2) = P(A_1)P(A_2), P(A_1 \cap A_3) = P(A_1)P(A_3), P(A_2 \cap A_3) = P(A_2)P(A_3);$$

pour $p = 3$:

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3).$$

L'exemple ci-dessus (avec $A_1 = A$, $A_2 = B$, $A_3 = C$) vérifie toutes les conditions pour $p = 2$ (les événements sont indépendants deux à deux), mais $P(A_1 \cap A_2 \cap A_3) = P(A_1 \cap A_2) = \frac{1}{4}$, tandis que $P(A_1)P(A_2)P(A_3) = \frac{1}{8}$.

Proposition 9.4 *Soit $A_1, \ldots, A_n$ une famille d'événements indépendants. Pour chaque i, on prend $A'_i = A_i$ ou $\overline{A}_i$. Alors les événements $A'_1, \ldots, A'_n$ sont indépendants.*

Démonstration. Soit p le nombre d'indices i pour lesquels $A'_i = \overline{A}_i$. On va démontrer la propriété par récurrence sur p. Pour $p = 0$, c'est une évidence (on n'a rien changé). Pour $p = 1$: supposons (au besoin en renumérotant) que $A'_1 = \overline{A}_1$, $A'_i = A_i$ pour $i \geq 2$. Il suffit de considérer des familles d'indices qui contiennent $1 = i_1$. Observons que l'hypothèse d'indépendance de la famille $A_1, \ldots, A_n$ implique en particulier que A_1 est indépendant de l'événement $A_{i_2} \cap \cdots \cap A_{i_p}$. D'après la Proposition 9.2, cela implique que $\overline{A}_1$ est indépendant de l'événement $A_{i_2} \cap \cdots \cap A_{i_p}$. Ce qui donne la formule souhaitée pour $P(\overline{A}_1 \cap A_{i_2} \cap \cdots \cap A_{i_p})$.

Pour passer de p à $p+1$, il suffit de remarquer qu'on n'a changé qu'un seul ensemble, et donc qu'on peut appliquer le cas $p = 1$ en redéfinissant convenablement les A_i. □

9.3 Construction d'un modèle probabiliste fondé sur la notion d'indépendance.

Prenons l'exemple suivant : soit un jeu de pile ou face dans lequel on lance n fois la même pièce de monnaie mal équilibrée. L'espace des résultats possibles du jeu est $\Omega = \{P, F\}^n$, où P signifie "pile" et F "face".

Notons P_k l'événement "le $k$ième lancer donne pile" et F_k l'événement "le $k$ième lancer donne face". On se pose la question de savoir quelle probabilité mettre sur $(\Omega, \mathcal{P}(\Omega)$ en sorte que :

- pour tout k la probabilité de P_k soit égale à p et celle de F_k, égale à $(1-p)$ avec $p \in [0, 1]$.
- les résultats des différents lancers soient indépendants.

Un élément de Ω est un $n-$uplet $\omega = (x_1, ..., x_n)$ où x_i est P ou F. On note pour tout ω :

$$\Pi(\omega) = \{i/x_i = P\} \text{ et } \Phi(\omega) = \{i/x_i = F\}.$$

On a alors :

$$\{\omega\} = (\cap_{i\in\Pi(\omega)} P_i) \cap \ (\cap_{i\in\phi(\omega)} F_i).$$

On note

$$\pi(\omega) = \text{card}(\Pi(\omega)), \quad \phi(\omega) = \text{card}(\Phi(\omega)).$$

Si on a l'hypothèse d'indépendance, on obtient :

$$P(\{\omega\}) = \Pi_{i\in\Pi(\omega)} P(P_i) \times \ \Pi_{i\in\Phi(\omega)} P(F_i).$$

et d'après la première hypothèse :

$$P(\{\omega\}) = p^{\pi(\omega)}(1-p)^{\phi(\omega)}.$$

On peut alors vérifier que cette relation définit bien une probabilité sur Ω.

Exercice : Dans ce modèle, quelle est la probabilité de l'événement : "on observe k piles et $n-k$ faces" ?