Probabilités discrètes et introduction à la statistique

Cours et exercices d'application directe.

François-Xavier Vialard

10 décembre 2018

Table des matières

0.1 Notations

Soit A un ensemble fini, on notera $\#A$ le cardinal de A, c'est-à-dire le nombre d'éléments de A. Soit A, B deux ensembles, on notera $A \cap B$ l'intersection de A et B, c'est-à-dire l'ensemble $\{x \,|\, x \in A \text{ et } x \in B\}$. On notera $A \cup B$ l'union de A et B, c'est-à-dire l'ensemble $\{x \,|\, x \in A \text{ ou } x \in B\}$. Dans la suite, la notation $\mathbb{K}$, lorsqu'elle sera utilisée, recouvrira les (le corps des) nombres réels $\mathbb{R}$ ou $\mathbb{C}$ les (le corps des) nombres complexes. On notera $A \times B$ le produit cartésien entre A et B, c'est-à-dire

$$A \times B = \{(x, y) \,:\, x \in A \text{ et } y \in B\}. \tag{1}$$

Si $f : E \mapsto F$ est une application et $y \in F$, on note

$$f^{-1}(\{y\}) = \{x \in X \,:\, f(x) = y\}, \tag{2}$$

qui s'appelle l'image réciproque de y par F.

Chapitre 1

Rappels de dénombrement

Dans ce cours, on travaillera souvent avec des expériences aléatoires dont les résultats possibles est un ensemble fini. Dans ce chapitre, on introduit/rappelle des notions élémentaires de dénombrement. On commence par des rappels de définitions sur les applications.

1.1 Rappels sur les applications

Dans cette partie, on considérera que les ensembles sont toujours non vide.

Définition 1 (Application injective)**.** *Soit E, F deux ensembles, et $f : E \mapsto F$ une application. On dit que f est injective si $\forall x \neq y \in E$, $f(x) \neq f(y)$.*

Définition 2 (Application surjective)**.** *Soit E, F deux ensembles, et $f : E \mapsto F$ une application. On dit que f est surjective si $\forall y \in F$, il existe $x \in E$ tel que $f(x) = y$.*

Définition 3 (Application bijective)**.** *Soit E, F deux ensembles, et $f : E \mapsto F$ une application. On dit que f est bijective si elle est injective et surjective.*

Exemple 1. *L'application $f : \mathbb{R} \mapsto \mathbb{R}$ définie par $f(x) = x^2$ n'est pas surjective car l'équation $f(x) = -1$ n'a pas de solution réelle, c'est-à-dire de solution dans le domaine de définition de la fonction f. De même, elle n'est pas injective car $f(-1) = f(1) = 1$, ce qui contredit la définition 1.*
On remarque aussi que la réponse pour l'injectivité et la surjectivité dépend du domaine de définition de la fonction. Par exemple, la fonction f ci-dessus est surjective quand on élargit le domaine de définition à $\mathbb{C}$ tout entier et l'espace d'arrivée aussi. En revanche, cette extension de la fonction $f : \mathbb{C} \to \mathbb{C}$ est de toute façon non injective.

Exercice 1 (Laissé au lecteur)**.** *On dit qu'une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ est strictement croissante si quelques soient $x, y \in \mathbb{R}$ tels que $x < y$, alors $f(x) < f(y)$. Démontrer que f est injective.*

On dit qu'une fonction est strictement monotone soit f est strictement croissante, soit $-f$ l'est et dans ce cas, on dit que f est strictement décroissante. Ces propriétés se représentent graphiquement, voir la figure 1.1 pour illustration.

1.1.1 Cas des ensembles finis

Dans le cas où les ensembles E et F sont finis, on a :

Théorème 1. *Les ensembles E et F ont même cardinal si et seulement si il existe une bijection entre E et F.*

C'est un résultat très utile pour calculer le cardinal d'un ensemble.

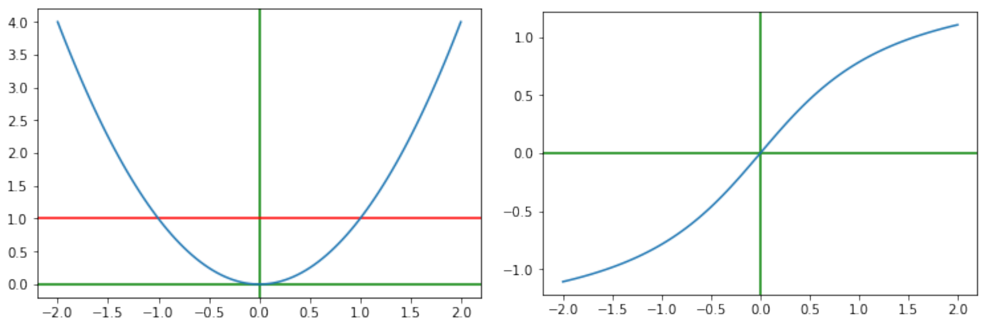


FIGURE 1.1 – À gauche, la fonction $x \to x^2$ n'est pas injective, la droite (rouge) horizontale $y = 1$ coupe le graphe (bleu) de la fonction en exactement deux points, ce qui contredit la définition de l'injectivité. À droite, la fonction $x \to \arctan(x)$ est injective car strictement croissante mais n'est pas surjective car son image est contenue dans l'intervalle $[-\frac{\pi}{2}, \frac{\pi}{2}]$ et ne décrit par $\mathbb{R}$ tout entier
.

Exemple 2. *On considère un ensemble F de cardinal fini n et E un ensemble quelconque. Quel est le cardinal de l'ensemble A défini par*

$$A = \{f : E \mapsto F \ : \ f(x) = f(y) \ \forall x, y \in E\} . \tag{1.1}$$

En d'autres termes, quel est le cardinal des fonctions constantes d'un ensemble quelconque dans un ensemble de cardinal n ?
La réponse est n. Soit un élément $x_0 \in F$, l'application $\varphi : A \mapsto E$ définie par $\varphi(f) = f(x_0)$ est une bijection. Donc par le théorème précédent, on aura $\#A = n$. Le fait que cette application soit une bijection est un exercice élémentaire laissé au lecteur.

En revanche, si on sait que deux ensembles ont même cardinal, alors cette "contrainte" permet d'en savoir plus sur une fonction donnée entre ces deux ensembles.

Théorème 2. *Si $f : E \mapsto F$ est une application entre deux ensembles A et F de même cardinal, alors les propriétés suivantes sont équivalentes :*

1. *f est injective,*
2. *f est surjective,*
3. *f est bijective.*

Là encore, la preuve de ce résultat est laissé au lecteur.

1.2 Dénombrement élémentaire

On commence avec le nombre de fonctions d'une ensemble dans un autre, tous les deux supposé de cardinal fini.

Théorème 3. *Soient E et F des ensembles de cardinal fini alors*

$$\#\{f : E \mapsto F\} = (\#F)^{(\#E)} . \tag{1.2}$$

Le lecteur sera sans doute plus familier avec la notion de tirage avec remise. Ici, il s'agit dont de tirer $\#E$ fois un élément de F. Pour le chaque tirage, il y a $\#F$ choix possibles et donc le nombre de tirages possibles est $\underbrace{\#F \times \ldots \times \#F}_{n \text{ fois.}}$. On voit donc que l'ensemble des fonctions de E dans F est en bijection avec le produit cartésien $\underbrace{F \times \ldots \times F}_{n \text{ fois.}}$

Théorème 4. *Soit $f : E \mapsto F$ une application entre deux ensembles finis. On a,*

$$\#E = \sum_{y \in f(E)} \#f^{-1}(\{y\}) \,. \tag{1.3}$$

Cette propriété est très utile pour dénombrer en construisant des applications intéressantes entre E et un espace F. Si on se donne une liste de n éléments, $(x_1, \ldots, x_n)$, combien y-a-t-il de permutations possibles de cette liste, c'est-à-dire la donnée de la position dans la liste de chaque élément de la liste $x_1, \ldots, x_n$. On a donc une application bijective de $\{1, \ldots, n\}$ dans lui-même. On a en alors :

Théorème 5. *Soit E un ensemble fini alors le cardinal de l'ensemble des bijections de E est égal à $(\#E)!$.*

Remarque 1. *On voit ici que, si E et F ont même cardinal, le cardinal des bijections de E dans F est évidemment aussi égal à $(\#E)!$.*

Preuve: Posons $E = \{e_1, \ldots, e_n\}$. Une bijection σ de E se définit par le choix de l'image des éléments de E à commencer par le premier élément, $\sigma(e_1) \in E$. Il y a donc $\#E$ choix possibles. L'image du second élément doit se faire dans $E \setminus \{\sigma(e_1)\}$ de cardinal $n-1$, et ainsi de suite pour les éléments restants. On a donc $n \times (n-1) \times \ldots \times 1$ façons de construire une bijection. □

Preuve: [Preuve formelle, qu'on pourra omettre en première lecture.] On va noter $\mathrm{Bij}(E, F)$ l'ensemble des bijections de E dans F. La preuve peut se faire par récurrence sur le cardinal de l'ensemble. On considère le résultat du théorème comme hypothèse de récurrence. Pour $\#E = 1$, il n'y a qu'une seule application. Le résultat est donc vrai. Supposons l'hypothèse de récurrence vraie pour tout ensemble dont le cardinal est n et considérons un ensemble E de cardinal $n + 1$. Choisissons un élément dans E, noté x_0 et considérons, pour $\tau_{(x_0,\sigma(x_0))}$ la bijection de E qui échange x_0 et $\sigma(x_0)$,

$$\begin{aligned} \varphi : \mathrm{Bij}(E, E) &\mapsto E \times \mathrm{Bij}(E \setminus \{x_0\}) \\ \sigma &\to (\sigma(x_0), \tau_{(x_0,\sigma(x_0))} \circ \sigma_{E \setminus \{x_0\}}) \,. \end{aligned}$$

L'application φ est une bijection (la preuve détaillée est laissée au lecteur). On a donc une bijection $\mathrm{Bij}(E, E)$ et le produit cartésien entre E et $\mathrm{Bij}(E \setminus \{x_0\})$. Par hypothèse de récurrence, le second ensemble est de cardinal $(n - 1)!$ et on conclut que $\# \mathrm{Bij}(E, E) = \#(E \times \mathrm{Bij}(E \setminus \{x_0\})) = n!$. □

Là encore, le lecteur est sans doute plus familier avec la notion de tirage sans remise. Se donner une bijection de E revient à tirer $\#E$ fois sans remise dans l'ensemble E. On voit donc comment calculer le nombre de fonctions injectives de E dans F.

Proposition 1. *Le nombre de fonctions injectives de E de cardinal p dans F de cardinal n avec $p \leq n$ est*

$$A_n^p \overset{\text{def.}}{=} n \times (n-1) \ldots \times (n+1-p) = \frac{n!}{(n-p)!} \,. \tag{1.4}$$

Ce nombre est appelé $p-$arrangement parmi n.

Évidemment, dans la proposition précédente on remarque que $p \leq n$ et qu'on retombe sur $n!$ quand on a $E = F$.

Preuve: La preuve peut se faire par récurrence et est laissée au lecteur. □

Exemple 3. *On considère le résultat d'une course de 15 concurrents. Combien y-a-t-il de tiercés (ou podiums, ou liste des 3 premiers) possibles ? Il y en a A_{15}^3.*

1.2.1 Les parties d'un ensemble et combinaisons

Définition 4. *Soit E un ensemble fini. On note $\mathcal{P}(E)$ l'ensemble de ses parties, c'est-à-dire l'ensemble formé des sous-ensembles de E.*

Exemple 4. *Soit* $E = \{1, 2, 3\}$ *alors* $\mathcal{P}(E) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$.

Attention, on ne met pas d'accolades autour de l'ensemble vide $\emptyset$ car si c'était le cas on parlerait de l'ensemble composé d'un élément qui est l'ensemble vide.

Proposition 2. *On a* $\#(\mathcal{P}(E)) = 2^{\# E}$.

En effet, ce résultat provient du fait suivant :

Proposition 3. *L'application de* $\mathcal{P}(E)$ *dans les fonctions de* E *dans* $\{0, 1\}$ *définie par*

$$A \in \mathcal{P}(E) \to [x \to \mathbf{1}_A(x)] \tag{1.5}$$

est une bijection (on a noté $\mathbf{1}_A$ *la fonction caractéristique de* A *qui est définie par* $\mathbf{1}_A(x) = 1$ *si* $x \in A$ *et* 0 *sinon).*

Preuve: La preuve est laissée au lecteur. □

Proposition 4. *Si* A, B *sont deux parties d'un ensemble fini* E*, on a*

$$\#(A \cup B) = \#A + \#B - \#(A \cap B)\,. \tag{1.6}$$

On notera que ce résultat peut se comprendre aussi avec l'utilisation de la fonction caractéristique.

Exercice 2. *Donner une formule pour une union de* k *ensembles. Cette formule s'appelle la formule du crible, ou formule de Poincaré,*

$$\#\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{s=1}^{k} (-1)^{s-1} \sum_{1 \le j_1 < \ldots < j_s \le k} \#(A_{j_1} \cap \ldots \cap A_{j_s})\,. \tag{1.7}$$

On peut la démontrer par récurrence.

On s'intéresse maintenant au nombre de parties à k éléments dans un ensemble de cardinal n. C'est un sous-ensemble de $\mathcal{P}(E)$.

Proposition 5. *Le nombre de parties à* k *éléments dans un ensemble* E *de cardinal fini* n *est égal à*

$$\binom{n}{p} \stackrel{\text{def.}}{=} \frac{n!}{p!(n-p)!}\,. \tag{1.8}$$

Dans certain livres, la notation C_n^p est aussi utilisée, mais elle est aujourd'hui moins courante que la notation $\binom{n}{p}$.

Preuve: Notons $\mathcal{P}_k(E)$ le nombre de parties à k éléments. Plutôt que de compter directement les parties à k éléments, on va partir des $k-$arrangements. À un $k-$arrangement, on peut associer une unique partie à k éléments mais deux $k-$arrangements peuvent représenter la même partie. Deux $k-$arrangements notés $(x_1, \ldots, x_k)$ et $(y_1, \ldots, y_k)$ représentent la même partie si et seulement si il existe une bijection σ de $\{1, \ldots, k\}$ telle que $(\sigma(x_1), \ldots, \sigma(x_k)) = (y_1, \ldots, y_k)$. On a que le nombre de $k-$arrangements correspondant à une partie donnée est $k!$. On peut alors utiliser le théorème 4 pour conclure que

$$k! \times \#(\mathcal{P}_k(E)) = A_n^p\,, \tag{1.9}$$

et obtenir le résultat. □

Le nombre de parties à $k-$éléments est donc le coefficient binômial utilisé pour le développement des puissances :

$$. \tag{1.10}$$

On présente quelques propriétés des coefficients binômiaux.

Proposition 6 (Propriétés des coefficients binômiaux)**.** *On a les formules suivantes,* *1.* $p\binom{n}{p} = n\binom{n-1}{p-1}$.

2. $\binom{n}{p} = \binom{n-1}{p-1} + \binom{n-1}{p}$.

3. $(x+y)^n = \sum_{i=1}^{n} \binom{n}{i} x^i y^{n-i}$.

On laisse au lecteur la vérification de ces formules probablement déjà connues.

1.2.2 Les répétitions

On commence par répondre à la question suivante : Combien y-a-t-il de groupes de k lettres (éventuellement les lettres peuvent être identiques) choisies dans un alphabet de n lettres ?

Définition 5. *Soit E un ensemble fini de cardinal n et $k \geq 1$ un entier. On appelle $k-$combinaison avec répétitions un groupe non-ordonné de k éléments de E non nécessairement distincts.*

Une autre façon de décrire une $k-$combinaison avec répétitions est de compter le nombre d'apparition d'un élément donné. C'est ce qui est expliqué dans la proposition ci-dessous, dont on ne donne pas de preuve.

Proposition 7. *Une $k-$combinaison avec répétitions dans un ensemble de cardinal n peut être décrite par un $n-$uplets $(k_1, \dots, k_n)$ d'entiers positifs ou nuls tels que*

$$\sum_{i=1}^{n} k_i = k\,. \tag{1.11}$$

C'est une correspondance qui fournit une bijection entre ces ensembles.

Proposition 8. *Le nombre de $k-$combinaisons avec répétitions parmi n éléments est*

$$\Gamma_n^k = \binom{n+k-1}{k} \tag{1.12}$$

Preuve: Choisir un $n-$uplets d'entiers $(k_1, \dots, k_n)$ tels que $\sum_{i=1}^{n} k_i = k$ est équivalent au problème de comment répartir k boules indistinguables dans n tiroirs différents. Pour répondre à ce problème, il suffit de se donner $n-1$ cloisons (pour partitionner en k tiroirs) et de placer ces cloisons entre les k boules. En prenant en compte la position des cloisons, il y a $\binom{n-1+k}{n-1} = \binom{n-1+k}{k}$ possibilités. $\square$

Exemple 5. *Combien y-a-t-il de possibilités pour répartir 3 ballons entre 5 enfants ? La réponse est Γ_5^3.*

On s'intéresse maintenant au nombre de listes qu'on peut former avec une collection d'éléments de cardinal k d'un ensemble de cardinal n. En d'autres termes, on se donne une $k-$combinaison avec répétitions et on souhaite dénombrer le nombre de listes possibles formé en utilisant tous les éléments de la $k-$combinaison.

Exemple 6. *Si l'ensemble E est choisi comme un alphabet de 26 lettres et qu'on choisit un groupe de lettre A, B, B, C, des listes possibles sont donc $ABBC$, $BABC$, $CABB$, ... La question est donc de savoir combien de mots on peut former à l'aide de ces 4 lettres.*

Exemple 7. *Si les n éléments choisis sont tous distincts, par exemple, A, B, C, D dans l'alphabet, alors le nombre de listes est égal au nombre de permutations $n!$. En général, il y a des répétitions entre plusieurs lettres et donc le nombre de listes sera plus petit que $n!$.*

En d'autres termes, on souhaite dénombrer les permutations d'objets qui contiennent $k_1, \dots, k_n$ répétitions.

Proposition 9. *Le nombre de permutations d'une liste de cardinal n contenant $k_1, \ldots, k_m$ répétitions est $\frac{n!}{k_1! \ldots k_m!}$.*

Preuve: Nommons $x_1, \ldots, x_m$ les m éléments distincts de la liste. Un permutation de la liste est la donnée du choix de la position des objets égaux à x_1, il y en a $\binom{n}{k_1}$, puis de la position des objets égaux à x_2, il y en a $\binom{n-k_1}{k_2}$ et ainsi de suite. Le résultat final est le produit de ces quantités, qui donne le résultat annoncé. □

Exercice 3. *Proposer une preuve en utilisant le théorème 4.*

Exemple 8. *Le nombre d'anagrammes du mot $AFFIRMA$ est égal à $\frac{7!}{(2!)^2} = 1260$, car les lettres A et F sont répétées deux fois chacunes.*

1.3 Quelques exercices de dénombrement

Exercice 4. *Combien y-a-t-il de codes à 3 chiffres contenant le chiffre 1 ?*

Exercice 5. *On considère un jeu de carte de 52 cartes qui contient donc 4 types (pique, coeur, carreau et trèfle) d'une suite de 13 cartes, on les suppose numérotées de* 1 *à* 13*.*

1. *Combien y-a-t-il de mains possibles au poker, c'est-à-dire d'ensemble de 5 cartes ?*
2. *Combien y-a-t-il de mains possibles qui ont des numéros distincts ?*
3. *Combien y-a-t-il de mains qui contiennent exactement une paire (c'est à dire deux cartes qui ont un type différent mais qui sont égales par ailleurs.) ?*

Chapitre 2

Éléments de probabilités

Dans ce cours, on s'intéresse à ce qu'on peut formaliser et quantifier sur les résultats d'une expérience aléatoire. On va attribuer aux résultats possibles de l'expérience aléatoire une probabilité, un nombre entre 0 et 1. Plus ce nombre est proche de 1, plus ce résultat est probable.

2.1 Introduction

On commence par une définition :

Définition 6. *L'ensemble des résultats possible d'une expérience aléatoire est appelé espace des états, ou ensemble fondamental ou encore, univers.*

Notation 1. *Cet ensemble est souvent noté Ω et un élément de cet ensemble ω.*

Exemple 9. *Voici quelques exemples d'univers pour différentes expériences aléatoires.*

1. *Si on lance un dé à 6 faces, les résultats sont $\Omega = \{1, 2, 3, 4, 5, 6\}$.*
2. *Si on lance deux dés cubiques qu'on peut distinguer, par exemple un rouge et un noir, l'univers est $\Omega = \{(i, j) \mid i, j = 1, \dots, 6\}$, où on aura par exemple convenu que le premier chiffre correspond au dé rouge et le second au dé noir.*
3. *On considère le lancer de deux dés et on s'intéresse à la somme obtenue. On peut considérer plusieurs univers. Par exemple $\Omega_1 = \{(i, j) \mid i, j = 1, \dots, 6\}$, mais aussi $\Omega_2 = \{2, 3, \dots, 12\}$, mais encore $\Omega_3 = \{(i, j) \mid i, j = 1, \dots, 6 \text{ et } i \leq j\}$.*
4. *L'ensemble des tiercés parmi k chevaux nommés $c_1, \dots, c_k$ (laissé en exercice au lecteur).*
5. *On tire indéfiniment à pile ou face : $\Omega = \{0, 1\}^{\mathbb{N}}$.*

On va beaucoup parler d'événements dans ce cours. Avant de les définir de manière formelle dans ce chapitre, on peut dire que sur des expériences aléatoires usuelles, les événements est une éventualité qui peut se réaliser. Les événements sont décrits par des conditions sur les résultats et on peut donc savoir si ces conditions sont vérifiées une fois le résultat de l'expérience connu.

Par exemple, on peut considérer un lancer de trois dés cubiques et l'événement *"la somme des trois faces est supérieure à 9"*, ou encore, *"aucune des faces n'est égale à 6"*. On peut donc énoncer :

Définition 7. *Étant donné des conditions définissant un événement, l'événement est une partie de Ω définie par l'ensemble des éléments de l'univers qui vérifient les conditions définissant l'événement.*

Remarque 2. *Soit A un événement, si ω est dans A, on dit que ω est réalisé. L'ensemble Ω est l'événement certain et $\emptyset$ est l'événement impossible.*

On a vu qu'un événement est un sous-ensemble de Ω mais, en général, tout sous-ensemble de Ω ne pourra pas être considéré comme un événement. On y reviendra par la suite.

Définition 8. *Soit Ω un ensemble. On dit qu'un ensemble $\mathcal{A}$ de parties de Ω est une tribu (ou $\sigma-$algèbre) de Ω si*

1. $\Omega \in \mathcal{A}$.
2. *Stabilité par passage au complémentaire : Quelque soit $A \in \mathcal{A}$ alors $A^c = \Omega \setminus A \in \mathcal{A}$.*
3. *Stabilité par union dénombrable : Si $(A_i)_{i\in\mathbb{N}} \in \mathcal{A}$ alors $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{A}$.*

Remarque 3.
- *La plus petite tribu de Ω est $\mathcal{A} = \{\Omega, \emptyset\}$ et la plus grande est $\mathcal{P}(\Omega)$ tout entier.*
- *Soit $\mathcal{C}$ une sous-famille de $\mathcal{P}(\Omega)$, alors on peut considérer la plus petite tribu engendrée par $\mathcal{C}$: il suffit de considérer l'intersection de toutes les tribus contenant $\mathcal{C}$. On la note $\sigma(\mathcal{C})$.*

Exercice 6. *Soit $A \subset \Omega$. Quelle est la tribu engendrée par $\mathcal{A} = \{A\}$?*

Définition 9. *Soit $\mathcal{A}$ une tribu sur Ω. Une probabilité notée $\mathbb{P}$ sur $(\Omega, \mathcal{A})$ est une application $\mathbb{P} : \mathcal{A} \mapsto [0,1]$ vérifiant :*

1. $\mathbb{P}(\Omega) = 1$.
2. $\mathbb{P}(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} \mathbb{P}(A_i)$ *pour $A_i \in \mathcal{A}$ des parties deux à deux disjointes.*

On appelle le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

Définition 10. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.*

1. *On appelle événement quasi-certain, tout événement $A \in \mathcal{A}$ tel que $\mathbb{P}(A) = 1$.*
2. *On appelle événement négligeable, tout événement $B \in \mathcal{A}$ tel que $\mathbb{P}(B) = 0$.*

2.2 Espace probabilisé fini

L'exemple d'espace probabilisé le plus simple est le cas d'un espace fini Ω.

Définition 11. *Pour un univers fini Ω, on considérera (sauf explicitement signalé) que*

1. *La tribu est $\mathcal{P}(\Omega)$, toute partie de Ω est un événement.*
2. *Une probabilité est une application de $\mathbb{P}(\Omega) \mapsto [0,1]$ vérifiant, pour $A, B \in \mathcal{P}(\Omega)$ tels que $A \cap B = \emptyset$ (on dit aussi que les événements A et B sont incompatibles),*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)\,. \tag{2.1}$$

Proposition 10. *Se donner une probabilité sur Ω de cardinal n pour la tribu $\mathcal{P}(\Omega)$ est équivalent à se donner pour chaque $\omega_i \in \Omega$, sa probabilité $\mathbb{P}(\omega_i) \in [0,1]$ vérifiant la condition suivante*

$$\sum_{i=1}^{n} \mathbb{P}(\omega_i) = 1\,. \tag{2.2}$$

On a alors

$$\mathbb{P}(A) = \sum_{\omega\in A} \mathbb{P}(\omega)\,. \tag{2.3}$$

On a aussi la conséquence immédiate suivante :

Proposition 11. *Si $A, B \in \mathcal{P}(\Omega)$ sont tels que $A \subset B$ alors,*

$$\mathbb{P}(A) \leq \mathbb{P}(B)\,. \tag{2.4}$$

Preuve: On a l'égalité $B = A \cup (B \setminus A)$, ce qui implique

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)\,, \tag{2.5}$$

car $\mathbb{P}(A) \geq 0$. □

De manière générale, on a aussi

$$\mathbb{P}(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)\,. \tag{2.6}$$

On introduit maintenant la probabilité uniforme qui est la plus souvent utilisée sur les univers finis.

Définition 12 (Probabilité uniforme)**.** *On dit que la probabilité est uniforme sur* Ω *si* $\mathbb{P}(\omega) = \frac{1}{\#\Omega}$*, ou de manière équivalente,*

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} \,. \tag{2.7}$$

Cette situation de probabilité uniforme correspond à la situation usuelle du tirage à pile ou face. L'univers est $\Omega = \{P, F\}$ et la probabilité est définie par $\mathbb{P}(\{F\}) = \mathbb{P}(\{P\}) = 1/2$. Attention, il est aussi possible de travailler avec des probabilités qui ne sont pas uniformes, comme le montre l'exercice suivant.

Exercice 7. *Prenons l'exemple du lancer de deux dés cubiques avec équiprobabilité pour les deux dés. On a donc* $\Omega = \{(i,j) \,|\, i,j = 1,\ldots,6\}$ *avec* $\mathbb{P}(\{(i,j)\}) = 1/36$*. Si on s'intéresse au couple* (i,j) *à l'ordre près, c'est-à-dire que* (i,j) *est identifié à* (j,i)*, alors on peut considérer* $\Omega = \{(i,j) \,|\, i \leq j\}$*. Quel est la probabilité correspondante ?*

Exercice 8. *On lance* n *boules numérotées dans* n *urnes, quelle est la probabilité qu'il y ait une boule dans chaque urne ?*

Réponse : Dans la plupart des exercices de probabilité de ce type, il faut faire un choix d'espace probabilisé. Ici, on supposera naturellement que chaque lancé de boule est équiprobable. L'univers est $\Omega = \mathcal{F}(\{b_1,\ldots,b_n\},\{u_1,\ldots,u_n\})$, l'ensemble des applications des boules dans les urnes (à chaque boule, on choisit une urne). On a donc $\#\Omega = n^n$. Une boule dans chaque urne signifie que l'application considérée est une bijection. Il y en a donc $n!$. La probabilité p cherchée est donc

$$p = \frac{n!}{n^n} \,. \tag{2.8}$$

Remarque 4. *Quel est le comportement quand* n *est très grand de cette probabilité ? On peut en prendre le logarithme et calculer*

$$\log(p) = \sum_{i=1}^{n} \log(i) - n\log(n) = n(\sum_{i=1}^{n} \frac{1}{n} \log(i/n)) \,. \tag{2.9}$$

Le terme entre parenthèses est une somme de Riemann et donc $\log(p) \simeq n\int_0^1 \log(u)du = -n(1+o(1))$ *et en conclusion,* $p \simeq C(n)e^{-n}$ *avec* $C(n)$ *une fonction de* n *telle que* $C(n)e^{-n} \to 0$ *quand* $n \to +\infty$*. C'est le premier terme de la formule de Stirling qui dit que* $n!/n^n \simeq \sqrt{2\pi n}e^{-n}$*.*

Exercice 9. *Dans une urne, on a* k_1, k_2 *boules respectivement rouges et noires. Notons* $k_1 + k_2 = k$*.*

1. *Quelle est la probabilité de tirer* 3 *boules de la même couleur pour un tirage avec remise ?*
2. *On tire* n *boules avec remise et on considère la probabilité d'avoir obtenu exactement* p *boules rouges.*
3. *On tire sans remise un groupe de* n *boules dans cette urne. Quelle est la probabilité d'obtenir* p *boules rouges ? (on prendra* $n < k$*)*

Exercice 10. *Sur un groupe de* n *personnes tirées au hasard dans la population, quelle est la probabilité que deux personnes aient leur anniversaire le même jour, on supposera qu'il y a* 365 *jours dans l'année.*

On continue avec quelques définitions utiles dans la suite :

Définition 13. *On dit que* $A \in \mathcal{P}(\Omega)$ *est négligeable si* $\mathbb{P}(A) = 0$ *et qu'il est quasi-certain si* $\mathbb{P}(A) = 1$*.*

Évidemment, on a A est quasi-certain si et seulement si A^c est négligeable.

Définition 14. *On dit que* $A_i \in \mathcal{P}(\Omega)$ *pour* $i = 1,\ldots,n$ *est un système complet d'événements si*

1. $A_i \cap A_j = \emptyset$*.*
2. $\bigcup_{i=1}^{n} A_i = \Omega$*.*

On dit qu'il est quasi-complet s'il vérifie

1. $A_i \cap A_j = \emptyset$.
2. $\mathbb{P}(\bigcup_{i=1}^n A_i) = 1$.

Proposition 12. *Si A_i pour $i = 1, \ldots, n$ est quasi-complet alors, quelque soit $B \in \mathcal{P}(\Omega)$, on a*

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i)\,. \tag{2.10}$$

Preuve: On a que $\bigcup_{i=1}^n A_i$ est quasi-certain, notons Z son complémentaire. On a donc $\mathbb{P}(Z) = 0$. On a

$$B = \left(\bigcup_{i=1}^{n} (B \cap A_i)\right) \cup (B \cap Z)\,. \tag{2.11}$$

En utilisant l'additivité de $\mathbb{P}$ car les événements $(B \cap A_i)_{i=1,\ldots,n}$ et $B \cap Z$ sont deux à deux disjoints, on obtient

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i) + \mathbb{P}(B \cap Z)\,. \tag{2.12}$$

Comme $B \cap Z \subset Z$, on a $\mathbb{P}(B \cap Z) = 0$ et le résultat annoncé. $\square$

Chapitre 3

Conditionnement et indépendance

Dans ce chapitre, on introduit une notion de conditionnement sur les probabilités que l'on souhaite estimer : on a une information additionnelle sur le résultat de l'expérience, comment est-ce que cela affecte les probabilités estimées du résultat qui nous intéresse. Si l'information apportée est indépendante (heuristiquement) de l'événement considéré, la probabilité estimée est inchangée. Par exemple, on considère un population de boules et de cubes colorés de façon aléatoire. L'information de la couleur ne nous apportera donc aucune connaissance supplémentaire sur la forme : les événements couleur et forme sont indépendants. En revanche, si les boules sont de couleurs rouge et les cubes de couleur bleue alors ces deux événements sont finalement identiques. Entre ces deux extrêmes, beaucoup de situations sont possibles. On souhaite donc dans ce chapitre quantifié le gain d'information lorsqu'un événement est connu.

3.1 Probabilité conditionnelle

On considère dans la suite un univers Ω de cardinal fini et une probabilité sur Ω.

Définition 15. *Soit $A \in \mathcal{P}(\Omega)$ tel que $\mathbb{P}(A) > 0$. On définit la probabilité conditionnelle à A comme la probabilité sur Ω caractérisée par*

$$\mathbb{P}_A(B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}\,. \tag{3.1}$$

De plus, $\mathbb{P}_A(B)$ est appelé probabilité conditionnelle de B sachant A. On la note aussi $\mathbb{P}(B|A)$.

Exercice 11. *On vérifiera que $\mathbb{P}_A$ est bien une probabilité sur Ω : on doit vérifier que $\mathbb{P}_A$ prend des valeurs positives ou nulles, qu'elle satisfait la propriété d'additivité et enfin que $\mathbb{P}_A(\Omega) = 1$.*

Proposition 13. *On peut itérer la formule précédente pour obtenir, pour des événements $A_1, \ldots, A_k$, $k \geq 2$, tels que $\mathbb{P}(A_1 \cap \ldots \cap A_{k-1}) > 0$,*

$$\begin{aligned}\mathbb{P}(A_1 \cap \ldots \cap A_k) &= \mathbb{P}_{A_1 \cap \ldots \cap A_{k-1}}(A_k)\mathbb{P}(A_1 \cap \ldots \cap A_{k-1}) \\ &= \mathbb{P}(A_1)\mathbb{P}_{A_1}(A_2)\ldots\mathbb{P}_{A_1 \cap \ldots \cap A_{k-2}}(A_{k-1})\mathbb{P}_{A_1 \cap \ldots \cap A_{k-1}}(A_k) \\ &= \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\ldots\mathbb{P}(A_{k-1}|A_1 \cap \ldots \cap A_{k-2})\mathbb{P}(A_k|A_1 \cap \ldots \cap A_{k-1})\,.\end{aligned}$$

Preuve: La preuve se fait par récurrence immédiate. □

Exemple 10. *On lance deux dés distinguables, un bleu et un rouge. Sachant que le dé bleu tombe sur 4, quelle est la probabilité que la somme des deux faces soient supérieure ou égale à 8 ?*
On choisit l'univers $\Omega = \{(i,j)\,|\, i,j = 1, \ldots, 6\}$, muni de la probabilité uniforme. Soit A

l'événement le dé bleu est sur 4, alors, on peut écrire $A = \{(4,j) \,|\, j = 1,\ldots,6\}$. La probabilité de A est donc $1/6$ qui est non nulle. On peut donc conditionner par A. On note B l'événement "somme des dés est supérieure ou égale à 8". On a alors $B = \{(i,j) \,|\, i+j \geq 8\}$ et $A \cap B = \{(4,j) \,|\, j = 4,5,6\}$. On a donc $\mathbb{P}(B|A) = 3/6 = 1/2$.

On introduit maintenant la formule de Bayes et la formule des probabilités totales.

Théorème 6. *Soit A_i pour $i = 1,\ldots,n$ un système quasi-complet d'événements de Ω tel que $\mathbb{P}(A_i) > 0$. Alors on a, pour $B \in \mathcal{P}(\Omega)$,*

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)\,.\ \textit{(formule des probabilités totales).} \tag{3.2}$$

Soient $A, B \in \mathcal{P}(\Omega)$ des événements tels que $0 < \mathbb{P}(A) < 1$ et $\mathbb{P}(B) > 0$, alors

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(A^c)\mathbb{P}(B|A^c)}\ \textit{(formule de Bayes)}\,. \tag{3.3}$$

Preuve: La preuve du premier point découle de la proposition 12 et la formule de Bayes s'obtient en écrivant

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)\,. \tag{3.4}$$

On remarque que l'hypothèse $\mathbb{P}(A) < 1$ n'est utilisée que pour pouvoir conditionner par rapport à A^c qui dans ce cas a une probabilité non nulle. □

3.2 Indépendance

Définition 16. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. On dit que deux événements $A, B \in \mathcal{A}$ sont indépendants si*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\,. \tag{3.5}$$

Immédiatement, on a le lien suivant avec les probabilités conditionnelles : lorsque $\mathbb{P}(A) > 0$, cette propriété d'indépendance est équivalente à $\mathbb{P}_A(B) = \mathbb{P}(B)$.

Remarque 5. *Attention, il ne faut pas faire la confusion suivante : Les événements A, B sont indépendants ne dit pas que $A \cap B = \emptyset$. Au contraire, si on a $A \cap B = \emptyset$ et si A et B ne sont pas des événements négligeables, on a alors $\mathbb{P}(A \cap B) = 0 \neq \mathbb{P}(A)\mathbb{P}(B)$.*

Proposition 14. *Seuls les événements quasi-certains et les événement négligeables sont indépendants de tous les événements.*

Proposition 15. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Si $A, B \in \mathcal{A}$ sont indépendants, alors A^c, B^c le sont aussi. Ainsi que A, B^c et A^c, B.*

Preuve: Montrons par exemple (les autres cas s'en déduisent) que A^c et B sont indépendants. On a

$$\mathbb{P}(B) = \mathbb{P}(A^c \cap B) + \mathbb{P}(A \cap B) = \mathbb{P}(A^c \cap B) + \mathbb{P}(A)\mathbb{P}(B)\,. \tag{3.6}$$

donc

$$\mathbb{P}(A^c \cap B) = (1 - \mathbb{P}(A))\mathbb{P}(B) = \mathbb{P}(A^c)\mathbb{P}(B)\,. \tag{3.7}$$

□

Définition 17. *On dit que les événements $(A_i)_{i=1,\ldots,n}$ sont indépendants si quelque soit la sous famille d'indices $1 \leq i_1 < \ldots < i_k \leq n$, on a*

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \ldots \mathbb{P}(A_{i_k})\,. \tag{3.8}$$

Cette définition est très coûteuse pour la vérifier en pratique. On remarque qu'elle implique l'indépendance deux à deux mais la réciproque est fausse. Il ne suffit pas de vérifier l'indépendance de A_i et A_j pour $i \neq j$ quelconques.

Chapitre 4

Variables aléatoires

Motivation : Dans beaucoup de situations, par exemple les jeux de hasard, on souhaite quantifier le résultat de l'expérience aléatoire, c'est-à-dire associer une valeur, un nombre réel au résultat de l'expérience plutôt que de considérer seulement les probabilités d'occurence des événements. Par exemple, on souhaite calculer le gain moyen d'un joueur au loto. Considérons cet exemple plus en détail : Un joueur doit choisir 6 nombres parmi 49, il gagne 1 million d'euros s'il a choisi les 6 bons numéros et ne gagne rien sinon. Pour jouer à ce jeu, il doit payer 5 euros. Quel est le gain moyen du joueur, on le note m ?

$$m = p(10^6 - 5) - 5(1 - p)\,, \tag{4.1}$$

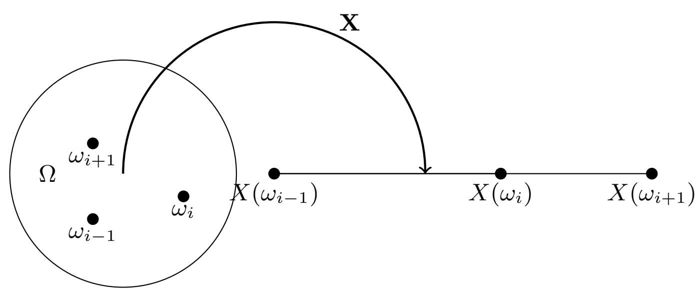
avec p la probabilité de gagner qui est donc de $1/\binom{49}{6}$.

4.1 Variable aléatoire

Dans cette partie, on introduit les concepts mathématiques pour pouvoir manipuler les quantités introduites précédemment.

Définition 18 (Variable aléatoire)**.** *Soit $(\Omega, \mathbb{P})$ un espace probabilisé tel que Ω soit fini, une variable aléatoire X est une application $X : \Omega \mapsto \mathbb{R}$.*

Ci-dessous, on propose un graphique pour illustrer la notion de variable aléatoire. C'est donc bien une application d'un ensemble Ω dans la droite réelle. La flèche en forme de demi-cercle représente l'application X.



Exemple 11. *Soit $A \subset \Omega$ une partie de Ω, on note $\mathbf{1}_A$ la fonction caractéristique de A définie par $\mathbf{1}_A(x) = 1$ si $x \in A$ et 0 sinon.*

Remarque 6. *Pour le lecteur intéressé, qui souhaite aller plus loin, il faut noter que cette définition n'est pas valable dans le cas où on considère un espace Ω qui est infini. Dans ce cas, l'ensemble des parties de Ω que l'on peut mesurer avec la probabilité $\mathbb{P}$ est appelé tribu et est différente de l'ensemble des parties sur Ω. Si on note $\mathcal{F}$ les éléments de cette tribu, aussi appelés ensembles mesurables, $X : \Omega \mapsto \mathbb{R}$ est une variable aléatoire si pour tout intervalle ouvert $I = [a, b]$, $X^{-1}(I) \in \mathcal{F}$. On peut penser a priori que c'est une complication mathématique qui n'est pas nécessaire. En fait, son utilité pratique se voit lorsqu'on souhaite quantifier un très grand nombre d'expériences aléatoires répétées.*

Un avantage de considérer des variables aléatoires, c'est-à-dire des fonctions de Ω dans $\mathbb{R}$ est qu'on peut utiliser les règles de calcul sur $\mathbb{R}$. Par exemple, on a

1. Si X, Y sont deux variables aléatoires et $\lambda \in \mathbb{R}$, alors $X + \lambda Y$ est aussi une variable aléatoire.
2. Plus généralement, si $f : \mathbb{R}^2 \mapsto \mathbb{R}$ est une fonction alors $f(X, Y)$ est une variable aléatoire.

Notation 2. *(Image réciproque ou préimage) Soit $S \subset \mathbb{R}$ un sous-ensemble de $\mathbb{R}$, on note $X^{-1}(S)$ l'ensemble des pré-images de S par X. C'est-à-dire,*

$$X^{-1}(S) = \{\omega \in \Omega \,|\, X(\omega) \in S\} . \tag{4.2}$$

Remarque 7. *(Important) Comme $\Omega = \{\omega_1, \ldots, \omega_n\}$ est un ensemble fini, ici de cardinal n, $X(\Omega)$ est aussi un ensemble fini. On peut donc énumérer l'ensemble des images $X(\Omega) = \{x_1, \ldots, x_k\}$. A priori, $k \leq n$. L'inégalité est stricte si X n'est pas injective car dans ce cas il y a deux éléments de Ω qui ont la même image.*

Proposition 16. *La famille $(X^{-1}(\{x_i\}))_{i=1,\ldots,k}$ forme un système complet d'événements.*

La preuve peut être omise en première lecture.

Preuve: On montre que $X^{-1}(\{x_i\}) \cap X^{-1}(\{x_j\}) = \emptyset$ si $x_i \neq x_j$. En effet, si cette intersection n'est pas vide, alors on considère un élément dedans noté ω_0. On a, par définition de l'image réciproque :

$$X(\omega_0) = x_i \text{ et } X(\omega_0) = x_j \,, \tag{4.3}$$

et donc $x_i = x_j$.

De plus, on a $X^{-1}(\{x_1\}) \cup \ldots \cup X^{-1}(\{x_k\}) = \Omega$. En effet, si $\omega \in \Omega$ alors il existe $i \in 1, \ldots, k$ tel que $X(\omega) = i$ et donc par définition de l'image réciproque, $\omega \in X^{-1}(\{x_i\}$. □

On va maintenant définir la loi d'une variable aléatoire. Comme la définition va le dire, c'est simplement une probabilité sur $X(\Omega)$. Cela reste valable dans le cas des probabilités "continues".

Définition 19 (Loi d'une variable aléatoire)**.** *Soit $X : \Omega \mapsto \mathbb{R}$ une variable aléatoire. La loi de X est la probabilité sur Ω définie par*

$$\mathbb{P}_X : \mathcal{P}(X(\Omega)) \mapsto [0, 1] \tag{4.4}$$

$$S \mapsto \mathbb{P}(X^{-1}(S)) . \tag{4.5}$$

Dans cette définition, on dit implicitement que $\mathbb{P}_X$ est une probabilité sur $X(\Omega)$. Il faut donc s'en convaincre. Avant de commencer la preuve, on introduit une notation agréable.

Notation 3. *On notera $\mathbb{P}(X = x_i) := \mathbb{P}(X^{-1}(x_i))$, c'est-à-dire la probabilité de l'événement que X soit égale à x_i. On notera aussi $\mathbb{P}(X \in S) := \mathbb{P}(X^{-1}(S))$.*

Preuve: *On montre que $\mathbb{P}_X$ est une probabilité.* On a bien que $\mathbb{P}_X(S) = \mathbb{P}(X^{-1}(S)) \in [0, 1]$.

On montre la propriété d'additivité de la probabilité $\mathbb{P}_X$. On considère donc S, S' deux parties d'intersection vide de $X(\Omega)$. On a donc que leur pre-images sont disjointes, c'est-à-dire

$$X^{-1}(S) \cap X^{-1}(S') = \emptyset \,, \tag{4.6}$$

et on peut donc appliquer la propriété d'additivité de $\mathbb{P}$ à ces deux parties de Ω. On a donc

$$\mathbb{P}(X^{-1}(S) \cup X^{-1}(S')) = \mathbb{P}(X^{-1}(S)) + \mathbb{P}(X^{-1}(S')) \,, \tag{4.7}$$

ce qui se réécrit

$$\mathbb{P}_X(S \cup S') = \mathbb{P}_X(S) + \mathbb{P}_X(S') . \tag{4.8}$$

□

On va donner maintenant quelques exemples élémentaires et importants.

Définition 20 (Loi certaine)**.** *On dit qu'une variable aléatoire X suit une loi certaine si il existe un réel x_0 tel que $\mathbb{P}(X = x_0) = 1$.*

Exemple 12 (Variables aléatoires qui ont même loi)**.** *Soit un tirage pile ou face non biaisé. On considère la variable aléatoire* $X : \{P, F\} \mapsto \{-1, 1\}$ *définie par* $X(P) = 1$ *et* $X(F) = -1$*. La loi de* X *est donnée par* $\mathbb{P}_X(1) = \mathbb{P}_X(-1) = 1/2$*. On remarque que la loi de* $-X$ *est la même mais* X *et* $-X$ *ne sont pas les mêmes variables aléatoires. On a donc que* $\mathbb{P}_X = \mathbb{P}_Y$ *n'implique pas que* $X = Y$*, c'est généralement faux.*

L'exemple précédent est semblable à une variable aléatoire dite de Bernoulli

Définition 21 (Loi de Bernoulli)**.** *Une variable aléatoire* X *suit une loi de Bernoulli de paramètre* $p \in [0, 1]$ *si* $\mathbb{P}(X = 1) = p$ *et* $\mathbb{P}(X = 0) = 1 - p$*.*

Exemple 13 (Plusieurs tirages indépendants de Bernoulli)**.** *On tire* n *fois une pièce biaisée dont la probabilité de tomber sur pile est* $p \in]1/2, 1]$*. Les tirages sont indépendants. On considère la variable aléatoire "nombre d'apparitions de pile". Quelle est sa loi ?*

Pour déterminer la loi de cette variable aléatoire, on doit d'abord déterminer $X(\Omega)$ *qui est dans est* $\{0, 1, \dots, n\}$*. Ensuite, pour chaque* $i \in 0, \dots, n$*, on doit calculer* $\mathbb{P}(X = i)$ *qui peut se calculer en définissant d'abord l'espace* Ω*. On peut prendre pour* Ω *l'ensemble des suites de tirages, il y en a donc* 2^n*. Ensuite, pour déterminer la probabilité d'occurence de* i *tirages sur pile, on remarque que la probabilité d'un événement élémentaire dans lequel il y a* i *pile est égale à* $p^i(1-p)^{n-i}$*. Il reste donc à dénombrer l'ensemble de tels tirages. Il y en a* $\binom{n}{i}$*. On a donc*

$$\mathbb{P}(X = i) = \binom{n}{i} p^i (1-p)^{n-i} \,. \tag{4.9}$$

On remarque qu'on a bien

$$\sum_{i=0}^{n} \mathbb{P}(X = i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = (p + 1 - p)^n = 1 \,. \tag{4.10}$$

Cette loi porte le nom de loi binômiale et est un exemple à retenir.

Définition 22 (Loi binômiale)**.** *Une variable aléatoire* X *suit une loi binômiale de paramètres* $n \in \mathbb{N}_{>0}$ *et* $p \in [0, 1]$ *si* $\mathbb{P}(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$ *et* $\mathbb{P}(X = 0) = 1 - p$*.*

Définition 23 (Fonction de répartition)**.** *Soit* X *une variable aléatoire. La fonction de répartition de* X *est notée* $F_X : \mathbb{R} \mapsto [0, 1]$ *et est définie par* $F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X^{-1}(]-\infty, x]))$*.*

Proposition 17. *La fonction de répartition est une fonction croissante.*

Preuve: Si $x \leq y$ sont deux nombre réels, alors on a l'inclusion des événements $(X \leq x) \subset (X \leq y)$ et donc $\mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y)$. $\square$

On peut donner un exemple de graphe d'une fonction de répartition.

Définition 24 (Espérance ou moyenne)**.** *Soit* X *une variable aléatoire. L'espérance de* X *ou sa moyenne est notée* $\mathbb{E}(X)$ *et est définie par*

$$\mathbb{E}(X) = \sum_{i=1}^{n} X(\omega_i)\mathbb{P}(\{\omega_i\}) = \sum_{i=1}^{k} x_i \, \mathbb{P}_X(\{x_i\}) \tag{4.11}$$

avec $X(\Omega) = \{x_1, \dots, x_k\}$*.*

On notera donc que l'espérance peut se calculer sur Ω (première formule) ou sur $X(\Omega)$ (seconde formule).

Proposition 18. *Si* X, Y *sont deux variables aléatoires telles que* $X \leq Y$ *alors on a*

$$E(X) \leq E(Y) \,. \tag{4.12}$$

Preuve: Il suffit d'appliquer la première formule définissant l'espérance et d'utiliser le fait que quelque soit $\omega_i \in \Omega$, on a $X(\omega_i) \leq Y(\omega_i)$ (et le fait que $\mathbb{P}(\{\omega_i\}) \geq 0$). $\square$

Pour décrire une variable aléatoire, la moyenne est une première information qui peut être intéressante. Une autre information intéressante est la variance.

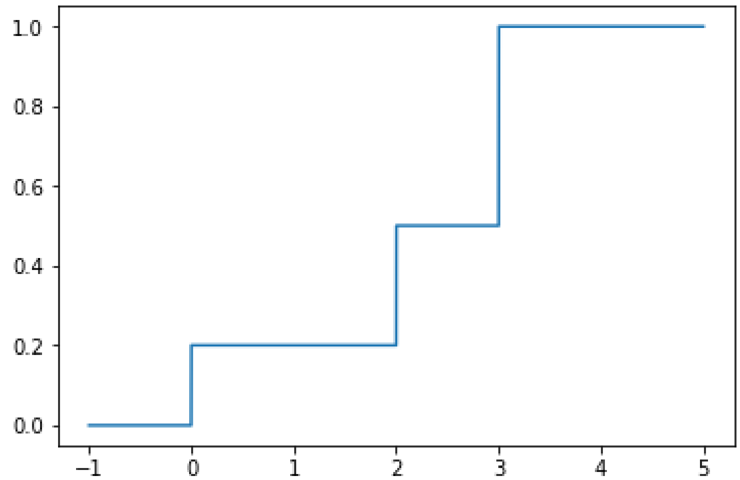


FIGURE 4.1 – C'est un exemple de graphe d'une fonction de répartition d'une variable aléatoire. On peut lire sur le graphe que la probabilité que la variable aléatoire soit strictement plus petite que 0 est nulle. La probabilité qu'elle soit égale à 0 est $0,2$. La probabilité qu'elle soit égale à 2 est $0,5 - 0,2 = 0,3$ et ainsi de suite.

Définition 25 (Variance d'une variable aléatoire)**.** *Soit X une variable aléatoire. La variance de la variable X est notée $\mathsf{Var}(X)$ et est définie par*

$$\mathsf{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \,. \tag{4.13}$$

Remarque 8. *On remarquera que la moyenne et la variance ne dépendent que de la loi de la variable X, i.e. si les variables ont même loi, elles ont même moyenne et variance.*

Exemple 14. *Sur les lois définies précédemment, on peut calculer leurs moyenne et variance respectives.*

1. *La moyenne d'une variable aléatoire suivant une loi de Bernoulli de paramètre p est p et sa variance est $p(1-p)$.*
2. *La variance d'une loi certaine est 0.*
3. *La moyenne d'une loi binômiale de paramètres (n,p) est np et sa variance est $np(1-p)$.*

Théorème 7 (Inégalité de Markov)**.** *Soit X une variable aléatoire réelle et $c > 0$ un réel, alors l'inégalité suivante est vérifiée,*

$$\mathbb{P}(|X| \geq c) \leq \frac{1}{c}\mathbb{E}(|X|) \,. \tag{4.14}$$

Preuve: On considère l'inégalité

$$|X| \geq c\,\mathbf{1}_{|X|>c} \,, \tag{4.15}$$

qui est une inégalité entre variables aléatoires. On peut donc prendre l'espérance et obtenir

$$\mathbb{E}(X) \geq c\,\mathbb{P}(|X| > c) \,, \tag{4.16}$$

qui donne le résultat. □

Théorème 8 (Inégalité de Bienaymé-Tchebychev)**.** *Soit X une variable aléatoire réelle et $c > 0$ un réel, alors l'inégalité suivante est vérifiée,*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq c) \leq \frac{1}{c^2}\mathsf{Var}(X) \,. \tag{4.17}$$

Preuve: On considère une inégalité similaire à celle utilisée dans la preuve précédente :

$$|X|^2 \geq c^2 \, \mathbf{1}_{|X|>c} \,, \tag{4.18}$$

qui donne le résultat en prenant l'espérance pour une variable telle que $\mathbb{E}(X) = 0$. Dans le cas général où la moyenne de X n'est pas nulle, il suffit de considérer la variable aléatoire $Y = X - \mathbb{E}(X)$ et d'appliquer le résultat précédent. □

4.2 Vecteurs aléatoires

Les vecteurs aléatoires sont des variables aléatoires à valeurs dans $\mathbb{R}^d$. La plupart des définitions précédentes s'étendent naturellement au cas des vecteurs aléatoires. On en détaille quelques unes.

Définition 26 (Couple aléatoire)**.** *Soient X et Y deux variables aléatoires réelles, le couple (X, Y) un vecteur aléatoire à valeurs dans $\mathbb{R}^2$. La loi du vecteur aléatoire est une probabilité définie sur $X(\Omega) \times Y(\Omega)$ définie par*

$$\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(\{X = x_i\} \cap \{Y = y_j\}) \tag{4.19}$$

où $X(\Omega) = \{x_1, \dots, x_n\}$ et $Y(\Omega) = \{y_1, \dots, y_m\}$. On notera souvent $p_{ij} = \mathbb{P}(X = x_i, Y = y_j)$.

On souligne le fait que la loi de probabilité du couple (X, Y) est une loi de probabilité sur un ensemble fini inclus dans $\mathbb{R}^2$.

Proposition 19. *Soit $f : \mathbb{R}^2 \mapsto \mathbb{R}$ une fonction et (X, Y) un vecteur aléatoire. Alors, on peut définir*

$$\mathbb{E}[f(X, Y)] = \sum_{\omega \in \Omega} f(X(\omega), Y(\omega)) \mathbb{P}(\{\omega\}) \,, \tag{4.20}$$

qui est aussi égale à

$$\mathbb{E}[f(X, Y)] = \sum_{\omega \in \Omega} f(x_i, y_j)) p_{ij} \,, \tag{4.21}$$

Cette définition s'étend directement aux vecteurs de longueur supérieure à 2. De la loi du couple (X, Y), on peut calculer la loi de X et celle de Y. On appelle ces deux lois, les lois marginales du couple (X, Y).

Définition 27 (Lois marginales)**.** *Étant donnée la loi du vecteur aléatoire (X, Y), on peut calculer la loi marginale de ce vecteur par rapport à la première variable par*

$$\mathbb{P}(X = x_i) = \sum_{j=1}^{m} \mathbb{P}(X = x_i, Y = y_j) \,. \tag{4.22}$$

On la notera p_i s'il n'y a pas de confusion possible.
De même, la loi marginale par rapport à la seconde variable se calcule par

$$\mathbb{P}(Y = y_j) = \sum_{i=1}^{n} \mathbb{P}(X = x_i, Y = y_j) \,. \tag{4.23}$$

On la notera q_j s'il n'y a pas de confusion possible.

On utilise le conditionnement vu au chapitre précédent pour définir la loi conditionnelle de X par rapport à Y.

Définition 28 (Loi conditionnelle de X sachant que $Y = y_j$)**.** *Soit y_j tel que $\mathbb{P}(Y = y_j) > 0$, la loi conditionnelle de X sachant que $Y = y_j$ est la loi de probabilité sur $X(\Omega)$ définie par*

$$\mathbb{P}_{Y=y_j}(X = x_i) = \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(Y = y_j)} \tag{4.24}$$

On la notera parfois $p_{i|j}$.
De la même manière, on peut définir la loi de Y sachant $X = x_i$.

Si on connaît la loi marginale de Y et la loi conditionnelle de X sachant Y, on a accès à la loi du couple et donc celle de X. Par exemple,

Proposition 20. *On a*

$$\mathbb{P}(X = x_i) = \sum_{j \in J} \mathbb{P}_{Y=y_j}(X = x_i)\mathbb{P}(Y = y_j)\,, \tag{4.25}$$

où $J = \{j\,, \mathbb{P}(Y = y_j) > 0\}$.

On peut donc définir l'espérance conditionnelle par rapport à l'événement $Y = y_j$.

Définition 29 (Espérance conditionnelle)**.** *Pour* $y_j \in Y(\Omega)$*, on définit*

$$\mathbb{E}[X|Y = y_j] = \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i|Y = y_j) \tag{4.26}$$

qui est l'espérance conditionnelle de X *par rapport à l'événement* $Y = y_j$.

4.3 Espérance, covariance, etc...

On rappelle la définition de l'espérance appliquée à un vecteur aléatoire : Soient (X, Y) un couple de variables aléatoires réelles (finies) et une fonction $f : \mathbb{R}^2 \mapsto \mathbb{R}$, alors

$$\mathbb{E}[f(X, Y)] = \sum_{i=1}^{n} \sum_{j=1}^{m} f(x_i, y_j)p_{ij}\,. \tag{4.27}$$

On donne maintenant quelques propriétés élémentaires.

Proposition 21. *Par le rappel précédent, on peut déduire les formules suivantes, pour* λ *un réel et* X *et* Y *deux variables aléatoires,*

$$\mathbb{E}[\lambda X + Y] = \lambda \mathbb{E}[X] + \mathbb{E}[Y] \tag{4.28}$$

$$\mathbb{E}[(\lambda X + Y)^2] = \lambda^2 \mathbb{E}[X^2] + \mathbb{E}[Y^2] + 2\lambda \mathbb{E}[XY]\,. \tag{4.29}$$

La seconde égalité donne un polynôme en λ *qui reste positif ou nul ; son discriminant est donc négatif ou nul, ce qui implique l'inégalité suivante (appelée Cauchy-Schwarz),*

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}\,. \tag{4.30}$$

Définition 30. *La covariance entre deux variables aléatoires* X *et* Y *est définie par*

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]\,. \tag{4.31}$$

On a directement la propriété suivante, pour a, b, c, d des réels

$$\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X, Y)\,. \tag{4.32}$$

La conséquence de l'inégalité de Cauchy-Schwarz sur la covariance est l'inégalité suivante :

$$|\mathrm{Cov}(X, Y)| \leq \sqrt{\mathsf{Var}[X]\mathsf{Var}[Y]}\,. \tag{4.33}$$

Parfois, on pourra utiliser le coefficient de corrélation entre X et Y, noté $\rho(X, Y)$ et défini par

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathsf{Var}[X]\mathsf{Var}[Y]}}\,. \tag{4.34}$$

Par l'inégalité précédente, c'est un nombre réel entre -1 et 1 et comme son nom l'indique, il mesure la corrélation entre les variables aléatoires X et Y. On a donné des exemples en cours. Pour comprendre ce que mesure ce coefficient et le mettre en relation avec la notion d'événements indépendants, on définit maintenant la notion d'indépendance pour des variables aléatoires.

Définition 31. *Deux variables aléatoires réelles (finies comme c'est le seul cas traité dans ce cours) X et Y, sont dites indépendantes si quelque soit $(x_i, y_j) \in X(\Omega) \times Y(\Omega)$, on a*

$$\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(\{\omega \in \Omega \,|\, X(\omega) = x_i \text{ et } Y(\omega) = y_j\}) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j)\,. \quad (4.35)$$

Autrement dit, deux variables aléatoires sont indépendantes si tous les couples d'événements $X = x_i$ et $Y = y_j$ sont indépendants. On remarque que le cas d'indépendance de deux variables est un cas pour lequel la connaissance des lois marginales est suffisante pour la connaissance de la loi du couple. Une conséquence de la définition sur la covariance est donnée par la proposition plus générale suivante,

Proposition 22. *Soit $f : \mathbb{R} \mapsto \mathbb{R}$ et $g : \mathbb{R} \mapsto \mathbb{R}$ deux fonctions réelles et X, Y deux variables aléatoires réelles indépendantes, alors*

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]\,. \quad (4.36)$$

La preuve de cette égalité a été proposée en cours. On ne la détaille pas ici mais on insiste sur la conséquence suivante :

Si X et Y sont indépendantes alors $\mathrm{Cov}(X, Y) = 0$.

La réciproque n'est pas vraie, c'est-à-dire qu'il existe des variables aléatoires X, Y qui ne sont pas indépendantes mais pour lesquelles on a quand même $\mathrm{Cov}(X, Y) = 0$. Une autre conséquence importante est que

$$\mathsf{Var}(X + Y) = \mathsf{Var}(X) + \mathsf{Var}(Y)\,. \quad (4.37)$$

Plus généralement, on a

Définition 32. *Les variables aléatoires $X_1, \ldots, X_n$ sont dites indépendantes si*

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2)\ldots\mathbb{P}(X_n = x_n)\,, \quad (4.38)$$

pour tout $(x_1, \ldots, x_n) \in X_1(\Omega) \times \ldots \times X_n(\Omega)$.

Encore une fois, on note que la connaissance des lois marginales suffit à déterminer la loi du vecteur $(X_1, \ldots, X_n)$. Ici aussi, on a (par récurrence par exemple)

$$\mathsf{Var}(X_1 + \ldots + X_n) = \mathsf{Var}(X_1) + \ldots + \mathsf{Var}(X_n)\,. \quad (4.39)$$

Un exemple d'application est le calcul de la variance d'une loi binômiale de paramètres (n, p) qui peut se faire en disant que cette variable aléatoire est la somme de n variables aléatoires indépendantes de loi de Bernoulli de paramètre p. On a

$$\mathbb{E}[X_1 + \ldots + X_n] = n\mathbb{E}[X_1] \quad (4.40)$$
$$\mathsf{Var}[X_1 + \ldots + X_n] = n\mathsf{Var}[X_1]\,. \quad (4.41)$$

On aura noté que l'hypothèse dit que les variables X_i ont même loi et donc elles ont même espérance ce qui donne la première formule. La seconde formule requiert le fait que les lois sont les mêmes, ce qui donne que la variance de X_i est égale à celle de X_1. mais l'indépendance est nécessaire pour obtenir que les termes de covariances qui seraient présent si on développe le carré, disparaissent.

4.4 Loi faible des grands nombres

Dans cette partie, on discute d'un résultat qui est intuitif et on en donne une justification. Pour le motiver, on donne l'exemple du principe des assurances. Supposons qu'un accident arrive à une personne avec une probabilité p qui est faible et mais qui peut coûter cher, posons c son coût en euros. On va supposer qu'il y a un grand nombre N de personnes et que les événements "avoir un accident" sont indépendants entre les individus. On associe donc à la personne i, la variable aléatoire X_i qui vaut 0 si la personne n'a pas d'accident et C sinon. Le coût total, noté C, des accidents sur toute la population est donc

$$C = \sum_{i=1}^{N} X_i \quad (4.42)$$

L'espérance de C est donc Ncp. En terme de moyenne, on a donc que le coût moyen par personne est pc. Le principe de l'assurance est de faire payer à chaque personne une cotisation de l'ordre du coût moyen pour pouvoir payer les éventuels (rares) accidents. Cette idée peut marcher si, la plupart du temps, le coût engendré par les accidents, c'est-à-dire $C(\omega)$ n'est pas trop éloigné de son espérance Ncp. La pire des situations étant $C(\omega) = Nc$ et dans ce cas, il est bien sûr impossible de couvrir ces dépenses avec le coût moyen. Il faut donc essayer de quantifier l'écart à la moyenne. Le théorème de la loi faible des grands nombres est une première étape dans cette direction.

Théorème 9. *Soit X une variable aléatoire finie et $X_1, \dots, X_n$ des variables aléatoires indépendantes de même loi que X. Soit $\varepsilon > 0$ un réel strictement positif,*

$$\mathbb{P}(|S_n - \mathbb{E}[X]| > \varepsilon) \leq \frac{\mathsf{Var}(X)}{n\varepsilon^2}\,, \tag{4.43}$$

où $S_n = \frac{1}{n}\sum_{i=1}^n X_i$. On a donc

$$\lim_{n\to+\infty} \mathbb{P}(|S_n - \mathbb{E}[X]| > \varepsilon) = 0\,. \tag{4.44}$$

En particulier, lorsque n est grand, cette probabilité est très petite. Les événements qui provoqueraient un cas de banqueroute pour cette idée d'assurance ont donc une probabilité faible, quantifiée par l'inégalité (4.43).