Probabilités discrètes

Polycopié de Benoît Laslier. Mis à jour par Noé Cuneo et Cyril Labbé

2022/2023

Table des matières

Ce cours introduit la théorie des probabilités discrètes, c'est-à-dire, sur des espaces dénombrables. Il s'agit d'un cas particulier de la théorie générale des probabilités car on se restreint aux expériences aléatoires pour lesquelles l'ensemble des résultats possibles est au plus dénombrable.

Etant moins lourd techniquement que le cadre général, le cadre discret permet d'introduire de nombreux concepts (espaces de probabilités, variables aléatoires, lois, espérances) sans nécessiter de théorie trop avancée. Par ailleurs, de nombreuses expériences aléatoires sont en fait discrètes par nature, et il est donc pertinent de se limiter à ce cadre dans un premier cours de probabilités.

Chapitre 1

Combinatoire

1.1 Cardinal

Le nombre d'éléments contenus dans un ensemble fini E est appelé **cardinal** de E. Plusieurs notations sont utilisées selon les sources, mais les principales sont $\#E$, $\mathrm{Card}(E)$ et $|E|$. Plus formellement :

Définition 1.1. *Le cardinal d'un ensemble fini E est l'unique entier $n \geq 0$ tel qu'il existe une bijection entre E et $\{1, \ldots, n\}$ (si $n = 0$, ce dernier ensemble est $\emptyset$).*

Autrement dit $\mathrm{Card}(E) = n$ si on peut associer à chaque $x \in E$ un élément de $\{1, \ldots, n\}$, c'est-à-dire si on peut écrire $E = \{x_1, x_2, \ldots, x_n\}$ (cette numérotation n'est pas nécessairement unique).

Lemme 1.2. *Soient A et B deux ensembles finis non vides. Alors* $\mathrm{Card}(A) = \mathrm{Card}(B)$ *si et seulement s'il existe une bijection entre A et B.*

Démonstration. Supposons que $\mathrm{Card}(A) = \mathrm{Card}(B) = n$ et soit f (resp. g) une bijection entre A (resp. B) et $\{1, \ldots n\}$. L'application $g^{-1} \circ f$ est bien définie car l'ensemble de définition de g^{-1} et l'image de f sont tous les deux égaux à $\{1, \ldots, n\}$ et c'est une bijection entre A et B, comme composée de deux bijections.

Réciproquement, si $f : A \to B$ et $g : B \to \{1, \ldots, \mathrm{Card}(B)\}$ sont deux bijections alors $f \circ g$ est aussi une bijection donc $\mathrm{Card}(A) = \mathrm{Card}(B)$. □

Grâce au lemme, on pourra toujours se ramener à des ensembles d'entiers lorsqu'on calculera des cardinaux : au lieu de manipuler des ensembles A, B, on pourra manipuler les ensembles $\{1, \ldots, n\}$, $\{1, \ldots, k\}$ où $n = \mathrm{Card}(A)$ et $k = \mathrm{Card}(B)$.

Maintenant que nous avons introduit la notion de cardinal, nous allons **dénombrer** divers ensembles, c'est-à-dire, déterminer leurs cardinaux. On commence par des résultats généraux sur les produits cartésiens. On rappelle que le produit cartésien $A \times B$ de deux ensembles A et B est l'ensemble des paires ordonnées dont le premier élément est dans A et le deuxième dans B, c.à.d $A \times B = \{(a, b) |\, a \in A,\, b \in B\}$.

Proposition 1.3. *Soient A et B deux ensembles finis. Alors $A \times B$ est un ensemble fini et*

$$\mathrm{Card}(A \times B) = \mathrm{Card}(A)\,\mathrm{Card}(B).$$

Plus généralement si $A_1, \ldots, A_k$ sont des ensembles finis, on a

$$\mathrm{Card}(A_1 \times \ldots \times A_k) = \prod_{i=1}^{k} \mathrm{Card}(A_i).$$

Démonstration. Pour prouver la première égalité, il suffit de construire une bijection entre $\{1, \ldots, n\} \times \{1, \ldots, k\}$ et $\{1, \ldots, nk\}$. La seconde égalité peut se prouver par récurrence. Les détails sont laissés en exercice. □

Exemple 1.3.1. On veut compter le nombre de mots de passe possibles contenant d'abord 3 lettres puis 3 chiffres, sachant que l'ordre des symboles compte.

Soit L l'ensemble des lettres de l'alphabet et C l'ensemble des chiffres. Clairement les mots de passe possibles sont les éléments de $L \times L \times L \times C \times C \times C$ et d'après la proposition il y a $26^3 \times 10^3 = 17\,576\,000$ mots de passe différents.

Dans la suite nous allons dénombrer certains sous-ensembles particuliers de l'ensemble $\{1, \ldots, N\}^n$ où N et n sont deux entiers. Nous verrons que ces ensembles pourront s'interpréter à l'aide d'ensembles de fonctions.

1.2 Fonctions, arrangements et permutations

Proposition 1.4. *Soient A et B deux ensembles finis. L'ensemble $\mathcal{F}(A, B)$ (parfois noté aussi B^A) des fonctions de A dans B est fini et l'on a*

$$\mathrm{Card}\left(\mathcal{F}(A, B)\right) = \mathrm{Card}(B)^{\mathrm{Card}(A)}.$$

En particulier si A est de cardinal n et B est de cardinal N alors $\mathrm{Card}\left(\mathcal{F}(A, B)\right) = N^n$.

Démonstration. Une fonction de A dans B peut être interprétée comme une liste ordonnée de $\mathrm{Card}(A)$ éléments de B : en effet, il suffit d'ordonner les éléments de A et de lister leurs images par f. Ainsi l'ensemble $\mathcal{F}(A, B)$ est en bijection avec $B \times B \times \ldots \times B$ où l'on considère ici $\mathrm{Card}(A)$ occurrences de B. Or le cardinal de ce dernier ensemble est égal à $\mathrm{Card}(B)^{\mathrm{Card}(A)}$ par la proposition précédente. □

Dans la preuve, nous avons utilisé l'existence d'une bijection entre l'ensemble $\{1, \ldots, N\}^n$ et l'ensemble des fonctions d'un ensemble à n éléments dans un ensemble à N éléments. Dans la suite, nous nous reposerons sur cette correspondance, et passerons d'un point de vue à l'autre en fonction des besoins. Examinons à présent quelques sous-ensembles de $\{1, \ldots, N\}^n$ particuliers.

Proposition 1.5. *Soient $1 \leq n \leq N$ des entiers. Le nombre de fonctions injectives d'un ensemble à n éléments dans un ensemble à N éléments est donné par*

$$N \times (N-1) \times \ldots \times (N-n+1) = \prod_{i=0}^{n-1}(N-i) = \frac{N!}{(N-n)!}.$$

Ici on utilise la convention $0! = 1$.

Démonstration. Toute fonction injective de $\{1, \ldots, n\}$ dans $\{1, \ldots, N\}$ peut être vue comme une suite **ordonnée** de n éléments **distincts** de l'ensemble $\{1, \ldots, N\}$. Cette correspondance est bijective. Il suffit donc de dénombrer l'ensemble des telles suites. Or, pour construire une telle suite il suffit de choisir un élément parmi N possibles, puis un second élément parmi $N-1$ possibles, ..., puis un n-ème élément parmi $N-n+1$ possibles. Le résultat s'en suit. □

Dans la preuve, nous avons utilisé le fait suivant : toute fonction injective de $\{1, \ldots, n\}$ dans $\{1, \ldots, N\}$ peut être vue comme une suite **ordonnée** de n éléments **distincts** de l'ensemble $\{1, \ldots, N\}$. Il se trouve que ces suites portent un nom :

Définition 1.6. *Soient $1 \leq n \leq N$ des entiers. Une suite **ordonnée** de n éléments **distincts** d'un ensemble à N éléments est appelée un arrangement à n éléments d'un ensemble à N éléments. Il est usuel de noter A_N^n le nombre total de tels arrangements. La proposition précédente a montré que $A_N^n = \frac{N!}{(N-n)!}$.*

Dans le cas où $N = n$, une fonction injective (d'un ensemble à n éléments dans un ensemble à N éléments) est nécessairement bijective. On retrouve alors le résultat bien connu suivant :

Corollaire 1.7. *Le nombre de bijections d'un ensemble à n éléments dans un ensemble à n éléments est*

$$n! = \prod_{k=1}^{n} k.$$

Dans ce cas, les arrangements sont appelés permutations.

Définition 1.8. *Lorsque l'on ordonne n éléments entre eux, on parle de permutation.*

Le nombre total de permutations est $A_n^n = n!$.

1.3 Lemme des bergers et combinaisons

On va maintenant voir quelques résultats permettant de dénombrer des ensembles plus compliqués à partir des exemples de base de la section précédente.

On commence avec le lemme des bergers, qui tire son nom de l'observation suivante : “pour compter le nombre total de moutons, on peut compter le nombre total de pattes et diviser par 4".

Une version un peu plus générale serait : “si l'on range dans chaque tiroir k paires de chaussettes, alors le nombre total de tiroirs est égal au nombre total de paires de chaussettes divisé par k."

Proposition 1.9 (Lemme des bergers, version formelle)**.** *Soient deux ensembles X, Y finis. S'il existe une fonction surjective f de X dans Y et un entier $k \geq 1$ tels que pour tout $y \in Y$, $\mathrm{Card}(f^{-1}(\{y\})) = k$, alors $\mathrm{Card}(Y) = \mathrm{Card}(X)/k$.*

Démonstration. Posons $n = \mathrm{Card}(Y)$ et notons $Y = \{y_1, \ldots, y_n\}$. On introduit alors $X_i := f^{-1}(\{y_i\})$ pour tout $1 \leq i \leq n$. Par hypothèse, chaque ensemble X_i est de cardinal k. Il n'est pas difficile de prouver que les ensembles X_i sont deux-à-deux disjoints et que $X = \cup_i X_i$. Pour conclure, il suffit d'observer que pour des ensembles disjoints X_i, $\mathrm{Card}(\cup_i X_i) = \sum_i \mathrm{Card}(X_i)$ (une preuve rigoureuse de cette identité sera fournie à la Proposition 1.13). □

Exemple 1.9.1. Le nombre d'anagrammes du mot “ENSEMBLE" est $\frac{8!}{3!}$. En effet, si l'on prend X l'ensemble des permutations de $\{1, \ldots, 8\}$, Y l'ensemble des anagrammes du mot “ENSEMBLE" et f l'application qui associe à une permutation donnée le mot obtenu en permutant les lettres du mot “ENSEMBLE" à l'aide de cette permutation, alors on observe que chaque anagramme admet $k = 3!$ antécédents par f : en effet, on peut permuter les positions des lettres E sans changer l'anagramme obtenu, et il y a $3!$ façons de le faire.

Une conséquence du lemme des bergers est qu'il permet de dénombrer les sous-ensembles à n éléments d'un ensemble à N éléments, que l'on appelle des combinaisons.

Définition 1.10. *Soient $0 \leq n \leq N$. Un sous-ensemble à n éléments d'un ensemble à N éléments est appelé une combinaison à n éléments d'un ensemble à N éléments.*

On notera la différence avec les arrangements : dans le cas des combinaisons les n éléments ne sont pas ordonnés, alors qu'ils le sont dans le cas des arrangements (deux ordres distincts induisent deux arrangements distincts). On peut alors s'appuyer sur la correspondance entre arrangements et combinaisons pour dénombrer ces dernières.

Proposition 1.11. *Le nombre total de combinaisons à n éléments d'un ensemble à N éléments est noté $\binom{N}{n}$ ou parfois C_N^n et vaut*

$$\binom{N}{n} = C_N^n = \frac{N!}{n!(N-n)!}.$$

Démonstration. Soit X l'ensemble des arrangements à n éléments d'un ensemble E à N éléments, et soit Y l'ensemble des sous-ensembles de E à n éléments. L'application f qui associe à chaque arrangement l'ensemble de ses n éléments est une surjection de X dans Y, et chaque élément de Y a exactement $n!$ antécédents. On peut alors appliquer le lemme des bergers et déduire que le cardinal de Y est égal au cardinal de X divisé par $n!$, d'où le résultat. □

Le coefficient $\binom{N}{n}$ est appelé coefficient binomial.

Généralisons. On considère un ensemble E à N éléments, et on se donne $p \geq 1$ entiers $k_1, \ldots, k_p \geq 1$ tels que $k_1 + \ldots + k_p = N$. On s'intéresse au nombre de façons de partitionner E en p parties distinguables contenant respectivement $k_1, \ldots, k_p$ éléments. Le mot distinguable signifie qu'on assigne à chaque partie un numéro entre 1 et p : cela a une importance dans le cas où $k_i = k_j$ pour une certaine paire $i \neq j$.

Proposition 1.12. *Soient E un ensemble à N éléments, et $k_1, \ldots, k_p$ des entiers vérifiant $k_1 + \ldots + k_p = N$. Le nombres de manières de partitionner E en p parties distinguables contenant respectivement $k_1, \ldots, k_p$ éléments est donnée par*

$$\frac{N!}{\prod_{i=1}^{p} k_i!}.$$

Démonstration. Séparer E en p parties est équivalent à associer à chaque élément de E un nombre entre 1 et p, sous la contrainte que l'on doit attribuer k_1 fois le nombre 1, k_2 fois le nombre 2, etc. Ainsi l'ensemble des partitions possibles est en bijection avec l'ensemble des "mots" à N lettres, chaque lettre étant un nombre entre 1 et p, contenant k_1 occurrences de 1, etc. On s'est donc ramené au cas des anagrammes. Comme dans l'exemple ci-dessus, on peut introduire X comme l'ensemble des permutations de $\{1, \ldots, N\}$ et Y comme l'ensemble des anagrammes, et f une application (surjective !) qui associe à chaque permutation un anagramme de sorte que chaque anagramme a exactement $k_1! \cdots k_p!$ antécédents par f. On peut alors conclure par le lemme des bergers. □

1.4 Formule d'inclusion-exclusion

On souhaite à présent déterminer le cardinal d'une union finie d'ensembles finis. Dans le cas particulier où ces ensembles sont disjoints, le résultat est élémentaire.

Proposition 1.13. *Soient A et B deux ensembles disjoints, on a*

$$|A \cup B| = |A| + |B|.$$

Plus généralement si $A_1, \ldots, A_n$ sont des ensembles disjoints, alors

$$|\cup_{i=1}^{n} A_i| = \sum_{i=1}^{n} |A_i|.$$

Démonstration. On se concentre sur le cas à deux ensembles. Il suffit de construire une bijection entre $\{1, \ldots, n+k\}$ et $A \cup B$, où $n = \mathrm{Card}(A)$ et $k = \mathrm{Card}(B)$. On numérote de 1 à n les éléments de A puis de $n+1$ à $n+k$ les éléments de B. La numérotation ainsi obtenue fournit une telle bijection. □

Lorsque les ensembles ne sont pas disjoints, le résultat général est plus compliqué. Commençons par traiter le cas de deux ensembles.

Proposition 1.14. *Soient A et B deux ensembles finis. Alors*

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Démonstration. On remarque que $A \setminus B$, $B \setminus A$ et $A \cap B$ sont trois ensembles disjoints et que $A = (A \setminus B) \cup (A \cap B)$, $B = (B \setminus A) \cup (A \cap B)$ et $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$. Ainsi par la proposition précédente on obtient

$$|A| = |A \setminus B| + |A \cap B| \,, \quad |B| = |B \setminus A| + |A \cap B| \,,$$

ainsi que

$$|A \cup B| = |A \setminus B| + |B \setminus A| + |A \cap B| \,,$$

et le résultat de la proposition s'en suit. □

Passons au cas de trois ensembles.

Proposition 1.15. *Soient A, B, C trois ensembles finis, on a*

$$|A \cup B \cup C| = \Big(|A| + |B| + |C|\Big) - \Big(|A \cap B| + |A \cap C| + |B \cap C|\Big) + \Big(|A \cap B \cap C|\Big).$$

Par ailleurs

$$|A \cup B \cup C| \leq \Big(|A| + |B| + |C|\Big).$$

La preuve formelle est très similaire à celle du résultat précédent et n'est pas très instructive donc nous ne donnerons que l'idée un peu informelle : quand on calcule $|A|+|B|+|C|$, on compte deux fois les éléments qui apparaissent dans deux des ensembles et trois fois ceux qui apparaissent dans les trois en même temps. Pour compenser le premier sur-compte on soustrait les cardinaux des intersections. Cela corrige bien le compte pour les éléments qui sont dans exactement deux ensembles mais pour ceux qui sont dans les trois, on a surcompensé et il ne sont plus comptés du tout. Il faut donc ajouter encore le dernier terme.

Théoreme 1.16. *Soient $A_1, \ldots, A_n$ des ensembles finis. On a*

$$|\cup_{i=1}^n A_i| = \sum_{k=1}^{n} (-1)^{k+1} \Big(\sum_{1 \leq i_1 < \ldots < i_k \leq n} |A_{i_1} \cap \ldots, A_{i_k}| \Big).$$

où la somme sur $1 \leq i_1 < \ldots < i_k \leq n$ désigne une somme sur tous les sous-ensembles de $\{1, \ldots, n\}$ contenant k éléments.

Démonstration. On procède par récurrence. Le cas $n = 2$ a déjà été établi. Supposons la formule vraie au rang n. Alors la formule au rang 2 assure que

$$|\cup_{i=1}^{n+1} A_i| = |\cup_{i=1}^n A_i| + |A_{n+1}| - |(\cup_{i=1}^n A_i) \cap A_{n+1}| \,.$$

L'hypothèse de récurrence assure que

$$|\cup_{i=1}^{n} A_i| = \sum_{k=1}^{n} (-1)^{k+1} \Big(\sum_{1 \leq i_1 < \ldots < i_k \leq n} |A_{i_1} \cap \ldots, A_{i_k}| \Big).$$

Si l'on ajoute au terme $k = 1$ la quantité $|A_{n+1}|$, on obtient

$$|\cup_{i=1}^{n} A_i| + |A_{n+1}| = (-1)^{1+1} \Big(\sum_{1 \leq i_1 \leq n+1} |A_{i_1}| \Big) + \sum_{k=2}^{n} (-1)^{k+1} \Big(\sum_{1 \leq i_1 < \ldots < i_k \leq n} |A_{i_1} \cap \ldots, A_{i_k}| \Big) .$$

Par ailleurs, si l'on pose $B_i := A_i \cap A_{n+1}$ pour tout $1 \leq i \leq n$, on observe que $(\cup_{i=1}^{n} A_i) \cap A_{n+1} = \cup_{i=1}^{n} B_i$ et ainsi l'hypothèse de récurrence montre que

$$\begin{aligned} |(\cup_{i=1}^{n} A_i) \cap A_{n+1}| &= \sum_{k=1}^{n} (-1)^{k+1} \Big(\sum_{1 \leq i_1 < \ldots < i_k \leq n} |B_{i_1} \cap \ldots, B_{i_k}| \Big) \\ &= \sum_{k=1}^{n} (-1)^{k+1} \Big(\sum_{1 \leq i_1 < \ldots < i_k \leq n} |A_{i_1} \cap \ldots, A_{i_k} \cap A_{n+1}| \Big) \\ &= - \sum_{k=2}^{n+1} (-1)^{k+1} \Big(\sum_{1 \leq i_1 < \ldots < i_{k-1} < i_k = n+1} |A_{i_1} \cap \ldots, A_{i_k} \cap A_{n+1}| \Big) \end{aligned}$$

En regroupant les termes obtenus, on obtient la formule au rang $n + 1$. □

Exemple 1.16.1. Un mot de passe est constitué de 6 caractères qui sont soit des lettres soit des chiffres. Le nombre total de mots de passe possibles est ainsi $(26 + 10)^6$. Pour compter le nombre de mots de passe contenant au moins un chiffre, on peut utiliser la Proposition 1.13. En effet, on note A l'ensemble des mots de passe contenant au moins un chiffre, et B l'ensemble complémentaire. Il y a 26^6 mots de passe sans chiffre ainsi $|B| = 26^6$, et donc $|A| = 36^6 - 26^6$.

Exemple 1.16.2. On souhaite calculer le nombre de combinaison à 5 caractères parmi A, B et C n'utilisant que 2 de ces lettres. Par exemple $AABAB$ est autorisé, mais $ACBBC$ ne l'est pas. Soit E_{AB} l'ensemble des combinaisons n'utilisant que A et B, E_{AC} celles n'utilisant que A et C et E_{BC} celles n'utilisant que B et C. On a $|E_{AB}| = |E_{AC}| = |E_{BC}| = 2^5$. Par ailleurs $|E_{AB} \cap E_{AC}| = 1$ car l'intersection ne contient que $AAAAA$. De même $|E_{AB} \cap E_{BC}| = |E_{BC} \cap E_{AC}| = 1$. Par ailleurs $E_{AB} \cap E_{AC} \cap E_{BC} = \emptyset$ et ainsi

$$\begin{aligned} |E_{AB} \cup E_{AC} \cup E_{BC}| &= |E_{AB}| + |E_{AC}| + |E_{BC}| \\ &\quad - |E_{AB} \cap E_{AC}| - |E_{AB} \cap E_{BC}| - |E_{AB} \cap E_{AC}| \\ &\quad + |E_{AB} \cap E_{AC} \cap E_{BC}| \\ &= 3 \cdot 2^5 - 3 + 0 \, . \end{aligned}$$

1.5 Modèles d'urnes

Nous allons maintenant calculer le nombre de façons de répartir des boules dans des urnes. On dira que les boules sont distinguables si chaque boule est différente (par exemple les boules sont numérotées). Dans le cas contraire, on dira qu'elles sont indistinguables. On notera $b \geq 1$ le nombre total de boules. En revanche, on considérera toujours que les urnes sont numérotées, disons de 1 à u.

Par ailleurs, on pourra ajouter une contrainte sur le nombre maximal ou minimal de boules à placer dans une même urne : on dira qu'on est dans le cas libre s'il n'y a aucune contrainte, on écrira $n \geq 1$ si l'on doit placer au moins une boule dans chaque urne, et $n \leq 1$ si l'on ne peut pas placer plus qu'une boule par urne. On notera que des contraintes sur u et b apparaissent implicitement dans les deux derniers cas : lorsque $n \geq 1$, il faut que $b \geq u$ et lorsque $n \leq 1$, il faut que $b \leq u$.

Proposition 1.17. *Le nombre de manières de placer b boules dans u urnes est donné par le tableau suivant :*

	libre	$n \geq 1$	$n \leq 1$
boules distinguables	u^b	*voir Remarque 1.17.1*	$\frac{u!}{(u-b)!}$
boules indistinguables	$\binom{u+b-1}{b}$	$\binom{b-1}{b-u} = \binom{b-1}{u-1}$	$\binom{u}{b}$

Démonstration.

1. Libre/distinguables : cela revient à dénombrer toutes les fonctions de l'ensemble $\{1, \ldots, b\}$ des boules dans l'ensemble $\{1, \ldots, u\}$ des urnes. On applique la Proposition 1.4 et l'on trouve u^b.
2. Libre/indistinguables : on remplit les urnes une par une en commençant par la première. On décompose l'opération de remplissage en deux types d'actions "mettre une boule dans l'urne courante" et "passer à l'urne suivante". Au total on devra faire b actions "mettre une boule" et $u-1$ actions "passer au suivant". Chaque suite d'actions conduit à un remplissage différent donc le nombre de remplissages possibles est $\binom{u+b-1}{b}$ pour le nombre de manières de placer les b actions "mettre une boule" parmi le total de $u+b-1$ actions.
3. $n \geq 1$/distinguables : Il n'existe pas de formule simple et pratique ! Voir la remarque 1.17.1 pour une expression.
4. $n \geq 1$/indistinguables : On commence par mettre une boule dans chaque urne et on est ramené au cas libre avec $b-u$ boules restantes.
5. $n \leq 1$/distinguables : On choisit dans quelle urne mettre la première boule avec u choix, puis la seconde avec $u-1$ choix restants... Il s'agit en fait de l'ensemble des arrangements à b éléments d'un ensemble à u éléments.
6. $n \leq 1$/indistinguables : Il faut choisir quel sous-ensemble d'urne sera occupé : il y a $\binom{u}{b}$ choix au total. Il s'agit en fait de l'ensemble des combinaisons à b éléments d'un ensemble à u éléments.

□

Remarque 1.17.1. Le nombre de manières de placer b boules distinguables dans u urnes avec pour contrainte que chaque urne contient au moins une boule peut être écrit à l'aide d'une formule d'inclusion-exclusion :

$$\sum_{k=0}^{u}(-1)^k\binom{u}{k}(u-k)^b.$$

C'est aussi le nombre de fonctions surjectives d'un ensemble à b éléments dans un ensemble à u éléments.

Chapitre 2

Espace probabilisé

Le but de ce chapitre est d'introduire la définition mathématique d'une probabilité ainsi que les premières propriétés associées.

2.1 Dénombrabilité

Définition 2.1. *Un ensemble E est dit infini dénombrable si et seulement s'il existe une bijection entre E et $\mathbb{N}$. Un ensemble E est dit dénombrable si et seulement s'il est soit fini, soit infini dénombrable.*

L'ensemble des entiers pairs et l'ensemble des rationnels sont infinis dénombrables. En revanche, l'ensemble des nombres réels n'est pas infini dénombrable. On pourra consulter l'Annexe A pour une preuve de ces affirmations. On remarquera qu'un ensemble E est dénombrable si et seulement si il existe une injection de E dans $\mathbb{N}$.

Proposition 2.2. *Soient E et F deux ensembles infinis dénombrables. L'ensemble $E \times F$ est infini dénombrable.*

Démonstration. On considère d'abord le cas $E = F = \mathbb{N}$. On peut énumérer les éléments de $\mathbb{N} \times \mathbb{N}$ de la façon suivante :

$$(0,0), (1,0), (0,1), (2,0), (1,1), (0,2), (3,0), (2,1), \dots \ .$$

Plus formellement on peut construire une bijection f de $\mathbb{N} \times \mathbb{N}$ dans $\mathbb{N}$ en posant

$$f(i,j) = j + \sum_{k=1}^{i+j} k \ , \quad \forall i, j \in \mathbb{N} \ .$$

Pour traiter le cas général, on se donne g et h des bijections de E et F dans $\mathbb{N}$. Alors $F(x,y) := f(g(x), h(y))$ est une bijection entre $E \times F$ et $\mathbb{N}$ ce qui conclut la preuve. □

Proposition 2.3. *Soit I un ensemble dénombrable et soit $(E_i)_{i \in I}$ une suite d'ensembles dénombrables indexés par les éléments de I. L'ensemble $\cup_{i \in I} E_i$ est dénombrable.*

Démonstration. On construit une injection de $\cup_{i\in I}E_i$ dans $\mathbb{N}\times\mathbb{N}$. Soit $f_i : E_i \to \mathbb{N}$ injective pour tout i et soit g injective de I dans $\mathbb{N}$. Pour tout y dans $\cup_{i\in I}E_i$, on pose $n(y) = \inf\{n \in g(I) : y \in E_n\}$ ainsi que $i(y)$, l'unique élément de I tel que $n(y) = g(i(y))$. On pose alors $\phi(y) = \big(n(y), f_{i(y)}(y)\big)$. ϕ est injective, ce qui conclut la preuve. □

2.2 Espace probabilisé

Nous commençons par les définitions formelles. On rappelle que $\mathcal{P}(\Omega)$ désigne l'ensemble des parties d'un ensemble Ω.

Définition 2.4. *Soit Ω un ensemble dénombrable. Une **probabilité** sur Ω (on dit parfois aussi **mesure de probabilité**) est une fonction $\mathbb{P}$ de $\mathcal{P}(\Omega)$ dans $\mathbb{R}$ satisfaisant les propriétés suivantes :*

1. *Pour toute partie E de Ω, $\mathbb{P}(E) \in [0,1]$,*
2. *$\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$,*
3. *Pour toute suite finie ou infinie dénombrable $(E_n)_n$ de parties deux à deux disjointes de Ω ,*
$$\mathbb{P}(\cup_n E_n) = \sum_n \mathbb{P}(E_n).$$

Remarque 2.4.1. La définition est légèrement redondante. En fait, on pourrait seulement imposer 1.′, 2.′ et 3. avec 1′. $\mathbb{P}(E) \geq 0$ pour toute partie E de Ω, et 2.′ $\mathbb{P}(\Omega) = 1$.

Vocabulaire 2.5.
— *Un couple $(\Omega, \mathbb{P})$ est appelé un espace probabilisé.*
— *Ω est appelé univers ou espace fondamental.*
— *un élément $\omega \in \Omega$ est appelé résultat élémentaire ou résultat possible.*
— *Une partie E de Ω est appelée un événement.*

Remarque 2.5.1. Attention à la différence entre un résultat élémentaire $\omega \in \Omega$ et l'événement $\{\omega\}$. D'après nos définitions, ω n'est pas un événement, c'est $\{\omega\}$ qui représente l'événement "le résultat de l'expérience est ω".

Il est parfois assez long et pénible de vérifier qu'une application $\mathbb{P} : \mathcal{P}(\Omega) \to \mathbb{R}$ vérifie bien les axiomes de la définition d'une probabilité. La proposition suivante permet de caractériser une probabilité de manière beaucoup plus directe et compacte. Elle s'appuie sur l'observation suivante. Soit $(\Omega, \mathbb{P})$ un espace probabilisé. Pour tout événement E on a

$$\mathbb{P}(E) = \sum_{\omega\in E} \mathbb{P}(\{\omega\}).$$

En effet, il suffit d'appliquer la propriété 3. à $E_n = \{\omega_n\}$ où $(\omega_n)_n$ est une énumération quelconque des éléments de E.

Proposition 2.6. *Soit $(\Omega, \mathbb{P})$ un espace probabilisé. La fonction $f : \Omega \to \mathbb{R}$ définie par*

$$f(\omega) := \mathbb{P}(\{\omega\}) , \quad \forall \omega \in \Omega ,$$

vérifie

$$\forall \omega, f(\omega) \geq 0 \ , \quad et \quad \sum_{\omega} f(\omega) = 1 \ , \tag{2.2.1}$$

Réciproquement, soit Ω un ensemble dénombrable, et soit $f : \Omega \to \mathbb{R}$ une fonction donnée. Si la fonction f satisfait (2.2.1) alors on peut poser pour tout $E \subset \Omega$

$$\mathbb{P}(E) = \sum_{\omega \in E} f(\omega) \ ,$$

et l'application $\mathbb{P}$ est une probabilité.

Démonstration. Si $\mathbb{P}$ est une probabilité, alors par 1. on sait que $\mathbb{P}(\{\omega\}) \geq 0$ pour tout $\omega \in \Omega$, et par 3. puis 2. on a

$$\sum_{\omega} \mathbb{P}(\{\omega\}) = \mathbb{P}(\Omega) = 1 \ .$$

Réciproquement, supposons que (2.2.1) soit vérifiée. Alors la somme définissant $\mathbb{P}(E)$ a bien du sens car tous les éléments sont positifs. Vérifions l'axiome 2. : on a

$$\mathbb{P}(\Omega) = \sum_{\omega} \mathbb{P}(\{\omega\}) = 1 \ ,$$

et par convention

$$\mathbb{P}(\emptyset) = 0 \ .$$

Concernant l'axiome 3. : pour toute collection E_n d'ensembles deux à deux disjoints la propriété de sommation par paquets des séries positives (voir cours d'analyse),

$$\mathbb{P}(\cup_n E_n) = \sum_{\omega \in \cup_n E_n} \mathbb{P}(\{\omega\}) = \sum_n \sum_{\omega \in E_n} \mathbb{P}(\{\omega\}) = \sum_n \mathbb{P}(E_n) \ .$$

Pour le premier axiome, comme $\mathbb{P}(\{\omega\}) \geq 0$ pour tout $\omega \in \Omega$, on obtient pour tout événement E

$$P(E) = \sum_{\omega \in E} P(\{\omega\}) \geq 0 \ ,$$

ainsi que

$$P(E) = \sum_{\omega \in E} P(\{\omega\}) \leq \sum_{\omega \in \Omega} P(\{\omega\}) = \mathbb{P}(\Omega) = 1 \ .$$

□

A l'avenir, on utilisera implicitement cette proposition pour définir une probabilité et l'on se contentera d'expliciter la fonction $f : \omega \to \mathbb{P}(\{\omega\})$.

Définition 2.7. *Soit Ω un ensemble fini et soit $\mathbb{P}$ la probabilité donnée par*

$$\mathbb{P}(\{\omega\}) := \frac{1}{\#\Omega} \ , \quad \forall \omega \in \Omega \ .$$

La probabilité $\mathbb{P}$ est appelée probabilité uniforme sur Ω.

Dans le cas de la probabilité uniforme, on obtient ainsi l'identité

$$\mathbb{P}(E) = \frac{\text{nombre d'éléments de } E}{\text{nombre d'éléments de } \Omega},$$

pour toute partie E de Ω.

2.3 Exemples

Exemple 2.7.1. On considère l'expérience aléatoire suivante : "on jette successivement deux dés à 6 faces". Chaque résultat de l'expérience peut être représenté par une paire d'entiers entre 1 et 6. On pose donc

$$\Omega = \{1; \dots; 6\}^2 = \{(1,1); (1,2); \dots; (6,5); (6,6)\}.$$

Les résultats élémentaires sont ainsi toutes les paires d'entiers.

Un événement est par définition un ensemble de résultats élémentaires. Ceux-ci peuvent parfois s'exprimer avec des mots, par exemple : "on obtient un double" correspond à $\{(1,1); (2,2); (3,3); (4,4); (5,5); (6,6)\}$, "le résultat du premier dé est 2" s'écrit $\{(2,1); (2,2); (2,3); (2,4); (2,5); (2,6)\}$, et "la somme des résultats vaut 3" est donné par $\{(1,2); (2,1)\}$. Dans chaque cas on a listé tous les résultats élémentaires pour lesquels l'événement s'est produit.

Chaque dé a une chance sur 6 de tomber sur une face donnée. Par ailleurs les lancers ne s'influencent pas. Ainsi chaque résultat élémentaire a une probabilité 1/36. On prend donc pour $\mathbb{P}$ la probabilité uniforme :

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|}.$$

Pour les exemples d'événements donnés ci-dessus, on vérifie que

$$\mathbb{P}(\text{on obtient un double}) = \frac{1}{6}, \quad \mathbb{P}(\text{le résultat du dé rouge est 2}) = \frac{1}{6},$$
$$\mathbb{P}(\text{la somme vaut 3}) = \frac{1}{18}.$$

Exemple 2.7.2. Lancer d'une pièce de monnaie de manière équilibrée : $\Omega = \{P, F\}$ et les deux résultats ont la même probabilité $\mathbb{P}_e(\{P\}) = \mathbb{P}_e(\{F\}) = 1/2$.

Lancer d'une pièce de monnaie de manière biaisée : $\Omega = \{P, F\}, \mathbb{P}_b(\{P\}) = \frac{1}{4}; \mathbb{P}_b(\{F\}) = \frac{3}{4}$. On remarque qu'entre un lancer équilibré et un lancer biaisé, l'ensemble Ω ne change pas. Par contre la probabilité change pour refléter le biais.

Exemple 2.7.3. On lance une pièce de monnaie une infinité de fois. On veut prendre $\Omega = \{P, F\}^{\mathbb{N}}$, c'est-à-dire l'ensemble des suites infinies de P et de F. Cela **sort** du cadre strict de ce cours puisque cet ensemble Ω n'est pas dénombrable. Ceci dit on s'autorisera parfois à considérer cette expérience quand même en faisant comme si toutes nos définitions s'appliquaient telles quelles.

Remarque 2.7.4. En général, le choix d'un ensemble Ω pour modéliser une expérience aléatoire est un peu arbitraire. Par exemple, on peut toujours ajouter des éléments dans Ω et leur donner une probabilité nulle sans que cela ne change les résultats.

2.4 Quelques propriétés élémentaires

On commence par donner quelques propriétés élémentaires sur les paires d'événements. On rappelle que A^c désigne l'ensemble complémentaire de A, c'est-à-dire, $A^c = \Omega \backslash A$.

Proposition 2.8. *Soit $(\Omega, \mathbb{P})$ un espace probabilisé et soient A et B des événements. Alors*

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

En conséquence,

1. *Si A et B sont des événements disjoints (c.à.d $A \cap B = \emptyset$), alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.*
2. *$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.*

Démonstration. On peut écrire les événements $A, B, A \cup B$ sous la forme d'unions disjointes :

$$A = (A \cap B) \cup (A \backslash B) \, , \quad B = (A \cap B) \cup (B \backslash A) \, ,$$

et

$$A \cup B = (A \cap B) \cup (A \backslash B) \cup (B \backslash A) \, .$$

On peut alors appliquer la propriété 3. d'une probabilité et obtenir

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \backslash B) \, , \quad \mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \backslash A) \, ,$$

ainsi que

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

On obtient alors aisément la première identité de l'énoncé. Les deux conséquences sont immédiates. □

Il y existe aussi une version du principe d'inclusion-exclusion pour les probabilités.

Proposition 2.9 (Principe d'inclusion-exclusion)**.** *Soient $E_1, \ldots, E_n$ des événements sur un espace probabilisé $(\Omega, \mathbb{P})$. On a*

$$\mathbb{P}(\cup_{i=1}^n E_i) = \sum_{k=1}^{n} (-1)^{k+1} \Big(\sum_{i_1 < \ldots < i_k} \mathbb{P}(E_{i_1} \cap \ldots \cap E_{i_k}) \Big).$$

Démonstration. On peut procéder par récurrence, les détails sont laissés en exercice. □

2.5 Probabilités et limites

On s'intéresse à présent à des suites d'événements et à la valeur de leurs probabilités.

Proposition 2.10 (Limite croissante)**.** *Soit $(\Omega, \mathbb{P})$ un espace probabilisé, et soit $(E_n)_{n\geq 0}$ une suite croissante d'événements au sens où $E_n \subset E_{n+1}$ pour tout $n \geq 0$. Soit $E_\infty = \cup_{n\geq 0} E_n$, on dit que E_∞ est la* ***limite croissante*** *des E_n et l'on a*

$$\mathbb{P}(E_\infty) = \lim_{n\to\infty} \mathbb{P}(E_n).$$

Démonstration. On pose $E_{-1} = \emptyset$ ainsi que $F_n = E_n \setminus E_{n-1}$ pour tout $n \geq 0$. Comme les E_n forment une suite croissante d'ensembles, les F_n sont deux à deux disjoints et l'on a

$$\forall n \in \mathbb{N},\ E_n = \cup_{k=0}^{n} F_k, \quad E_\infty = \cup_{k=0}^{\infty} F_k.$$

D'après la troisième propriété dans la définition d'une probabilité, on en déduit que

$$\forall n \in \mathbb{N},\ \mathbb{P}(E_n) = \sum_{k=0}^{n} \mathbb{P}(F_k), \quad \mathbb{P}(E_\infty) = \sum_{k=0}^{\infty} \mathbb{P}(F_k),$$

et d'après les propriétés des sommes infinies à termes positifs, $\mathbb{P}(E_n) \to \mathbb{P}(E_\infty)$ quand $n \to \infty$. □

Proposition 2.11 (Limite décroissante)**.** *Soit $(\Omega, \mathbb{P})$ un espace probabilisé, et soit $(E_n)_{n\geq 0}$ une suite décroissante d'événements au sens où $E_n \supset E_{n+1}$ pour tout $n \geq 0$. Soit $E_\infty = \cap_{n\geq 0} E_n$, on dit que E_∞ est la* ***limite décroissante*** *des E_n et l'on a*

$$\mathbb{P}(E_\infty) = \lim_{n\to\infty} \mathbb{P}(E_n).$$

Démonstration. Il suffit d'appliquer la propriété de la limite croissante aux complémentaires des événements E_n. □

2.6 Inégalités

Proposition 2.12 (Croissance)**.** *Soient A et B deux événements tels que $A \subset B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$.*

Démonstration. On a $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$ car les événements à droite sont disjoints et leur union donne B. Or $\mathbb{P}(B \setminus A) \geq 0$ et ainsi $\mathbb{P}(A) \leq \mathbb{P}(B)$. □

Proposition 2.13 (Borne d'union)**.** *Soit $(E_n)_n$ une suite finie ou infinie dénombrable d'événements (pas nécessairement disjoints), on a*

$$\mathbb{P}(\cup_n E_n) \leq \sum_n \mathbb{P}(E_n).$$

Démonstration. On remarque que si $\mathbb{P}(\cup_{n=1}^{N} E_n) \leq \sum_{n=1}^{N} \mathbb{P}(E_n)$ alors

$$\begin{aligned}\mathbb{P}(\cup_{n=1}^{N+1} E_n) &= \mathbb{P}(\cup_{n=1}^{N} E_n) + \mathbb{P}(E_{N+1}) - \mathbb{P}((\cup_{n=1}^{N} E_n) \cap E_{N+1}) \\ &\leq \mathbb{P}(\cup_{n=1}^{N} E_n) + \mathbb{P}(E_{N+1}) \\ &\leq \sum_{n=1}^{N+1} \mathbb{P}(E_n) \ .\end{aligned}$$

On a donc montré par récurrence que pour tout $N \geq 1$

$$\mathbb{P}(\cup_{n=1}^{N} E_n) \leq \sum_{n=1}^{N} \mathbb{P}(E_n) \ .$$

On peut alors passer à la limite $N \to \infty$, en utilisant la propriété de limite croissante vue ci-dessus ainsi que les propriétés des séries à termes positifs. □

Chapitre 3

Conditionnement et indépendance

Dans ce chapitre, on travaillera toujours sur un espace de probabilité $(\Omega, \mathbb{P})$ fixé.

3.1 Définition

Définition 3.1. *Soient A et B deux événements tels que $\mathbb{P}(B) > 0$. La **probabilité conditionnelle de A sachant B**, notée $\mathbb{P}(A \mid B)$ ou parfois $\mathbb{P}_B(A)$, est définie par*

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

Exemple 3.1.1. On tire un dé à 6 faces, et on pose A l'événement "le résultat est pair" et B l'événement "le résultat est supérieur ou égal à 4", c.à.d. $A = \{2, 4, 6\}$ et $B = \{4, 5, 6\}$. On a $\mathbb{P}(A) = 1/2$, $\mathbb{P}(B) = 1/2$ et $\mathbb{P}(A \cap B) = \mathbb{P}(\{4, 6\}) = 1/3$ donc $\mathbb{P}(A \mid B) = 2/3$. Ici on a aussi $\mathbb{P}(B \mid A) = 2/3$, mais les probabilités conditionnelles $\mathbb{P}(A \mid B)$ et $\mathbb{P}(B \mid A)$ ne sont pas égales en général.

La probabilité conditionnelle de A sachant B permet de mesurer si la réalisation de l'événement B influe sur la réalisation de l'événement A. Dans l'exemple précédent, on voit que le dé a plus de chance d'être pair si l'on sait qu'il est supérieur ou égal à 4.

3.2 Propriétés des probabilités conditionnelles

La propriété ci-dessous affirme qu'une probabilité conditionnelle peut s'interpréter comme une probabilité.

Proposition 3.2. *Soit B un événement vérifiant $P(B) > 0$. L'application*

$$\begin{cases} \mathcal{P}(\Omega) & \to \mathbb{R} \\ A & \mapsto \mathbb{P}(A \mid B) \end{cases}$$

est une mesure de probabilité sur Ω.

Démonstration. On doit vérifier les hypothèses de la Définition 2.4.

— Soit E un événement, on a $\mathbb{P}(E \mid B) = \frac{\mathbb{P}(E \cap B)}{\mathbb{P}(B)}$. Comme une probabilité est toujours positive ou nulle on a bien $\mathbb{P}(E \mid B) \geq 0$. Par ailleurs on sait que $\mathbb{P}(E \cap B) \leq \mathbb{P}(B)$ donc $\mathbb{P}(E \mid B) \leq 1$.

— $\mathbb{P}(\emptyset \mid B) = \frac{\mathbb{P}(\emptyset \cap B)}{\mathbb{P}(B)} = \frac{0}{\mathbb{P}(B)} = 0$ et $\mathbb{P}(\Omega \mid B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$.

— Soit $(E_n)_{n\in\mathbb{N}}$ une suite d'événements deux-à-deux disjoints. Clairement les $E_n \cap B$ sont aussi deux-à-deux disjoints et

$$\mathbb{P}\Big((\cup_n E_n) \mid B\Big) = \frac{\mathbb{P}\Big((\cup_n E_n) \cap B\Big)}{\mathbb{P}(B)} = \frac{\mathbb{P}\Big(\cup_n (E_n \cap B)\Big)}{\mathbb{P}(B)} = \frac{\sum_n \mathbb{P}(E_n \cap B)}{\mathbb{P}(B)}$$
$$= \sum_n \mathbb{P}(E_n \mid B).$$

□

En conséquence, les propriétés déjà établies pour les probabilités s'appliquent aux probabilités conditionnelles :

Corollaire 3.3. *Soit B un événement vérifiant $\mathbb{P}(B) > 0$.*

1. *Si E et F sont des événements disjoints (c.à.d $E \cap F = \emptyset$), alors $\mathbb{P}(E \cup F \mid B) = \mathbb{P}(E \mid B) + \mathbb{P}(F \mid B)$.*
2. *$\mathbb{P}(E^c \mid B) = 1 - \mathbb{P}(E \mid B)$.*
3. *$\mathbb{P}(E \cup F \mid B) = \mathbb{P}(E \mid B) + \mathbb{P}(F \mid B) - \mathbb{P}(E \cap F \mid B)$.*
4. *Si $E \subset F$ alors $\mathbb{P}(E \mid B) \leq \mathbb{P}(F \mid B)$.*
5. *Si E_n est une suite croissante ou décroissante, alors $\mathbb{P}(E_\infty \mid B) = \lim_{n\to\infty} \mathbb{P}(E_n \mid B)$.*

Nous abordons à présent une formule importante. Tout d'abord, rappelons qu'une suite dénombrable $(B_n)_n$ est une *partition* de Ω si les B_n sont deux-à-deux disjoints et si $\cup_n B_n = \Omega$. Notons également que tout événement E peut se décomposer sur les B_n, càd $E = \cup_n (E \cap B_n)$, et que les événements $E \cap B_n$ sont deux-à-deux disjoints. Ainsi l'axiome 3. d'une probabilité assure que

$$\mathbb{P}(E) = \sum_n \mathbb{P}(E \cap B_n).$$

Proposition 3.4 (Formule des probabilités totales)**.** *Soit $(B_n)_n$ une suite dénombrable formant une partition de Ω. Supposons que $0 < \mathbb{P}(B_n) < 1$ pour tout n, alors on a pour tout événement E*

$$\mathbb{P}(E) = \sum_n \mathbb{P}(E \mid B_n)\mathbb{P}(B_n).$$

Un cas particulier de cette formule est le suivant : si B est un événement tel que $0 < \mathbb{P}(B) < 1$ alors pour tout événement E

$$\mathbb{P}(E) = \mathbb{P}(E \mid B)\mathbb{P}(B) + \mathbb{P}(E \mid B^c)\mathbb{P}(B^c).$$

Démonstration. On a déjà vu que

$$\mathbb{P}(E) = \sum_n \mathbb{P}(E \cap B_n)$$

et par la définition même d'une probabilité conditionnelle, on a $\mathbb{P}(E \cap B_n) = \mathbb{P}(E \mid B_n)\mathbb{P}(B_n)$ d'où la formule de l'énoncé. □

Pour conclure, on énonce une formule célèbre permettant d'inverser l'ordre dans un conditionnement.

Proposition 3.5 (Formule de Bayes)**.** *Soient A et B deux événements de probabilités strictement positives. On a*

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

Démonstration. Il suffit de remarquer que

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

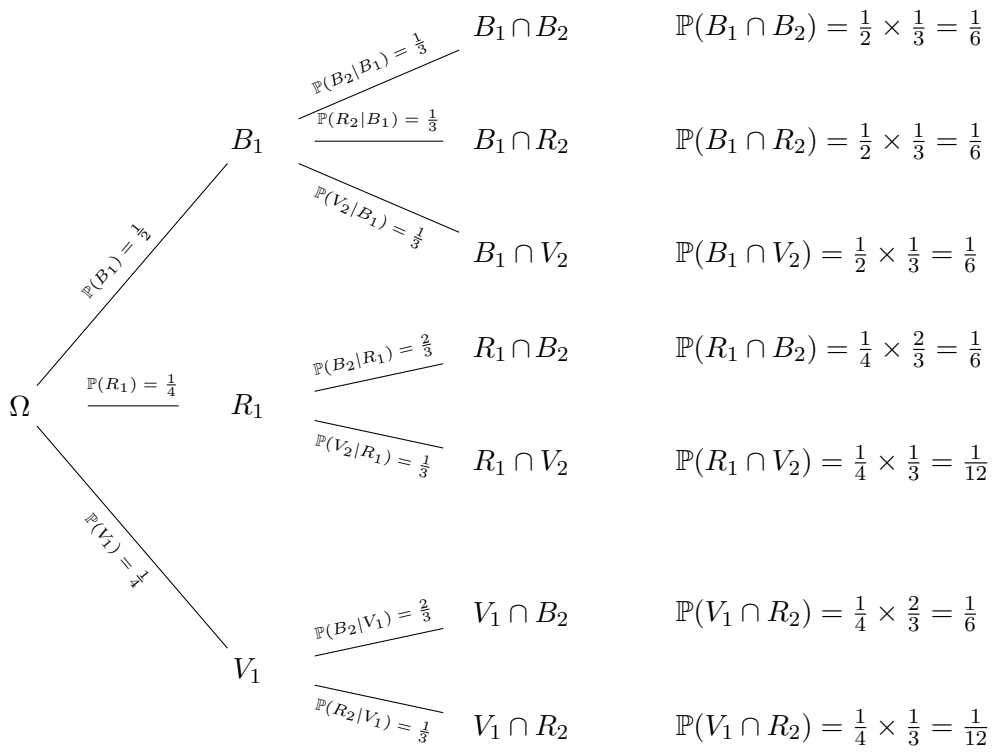
ce qui est la définition de la probabilité conditionnelle. □

3.3 Arbre de probabilité

Certaines expériences aléatoires sont bien décrites au travers de probabilités conditionnelles. Il est alors pratique de représenter l'expérience à l'aide d'un arbre de probabilité. Il existe une théorie formelle des arbres de probabilités, mais nous ne l'introduirons pas dans ce cours. Nous utiliserons en TD les arbres à des fins illustratives, et ils ne se substitueront *pas* aux calculs complets basés sur les formules démontrées (probabilités totales, Bayes, etc.).

Exemple : une urne contient 4 boules : deux bleues, une rouge et une verte. On tire deux boules à la suite. On introduit les événements $B_1, B_2, R_1, R_2, V_1, V_2$ donnés par B_i =“on tire une boule bleue au i'ème tirage”, R_i =“on tire une boule rouge au i'ème tirage”, V_i =“on tire une boule verte au i'ème tirage”.

Il est alors très utile de représenter les données sur un arbre comme ci-dessous :



3.4 Indépendance

Nous abordons à présent la notion d'indépendance, qui est centrale en probabilité.

Définition 3.6. *Deux événements A et B sont **indépendants** si*

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

Exemple 3.6.1. On tire un dé à 6 faces et on considère l'événement A "le résultat est pair" et l'événement B "le résultat est un multiple de 3". On a $\mathbb{P}(A) = 1/2$, $\mathbb{P}(B) = 1/3$ et $\mathbb{P}(A \cap B) = 1/6$ donc A et B sont indépendants.

Remarque 3.6.2. Lorsqu'on manipule plusieurs mesures de probabilités, il est nécessaire de préciser par rapport à quelle mesure de probabilité l'indépendance a lieu, et l'on écrit alors "A et B sont indépendants par rapport à la mesure de probabilité $\mathbb{P}$". On notera que dans l'exemple précédent, si l'on considère la mesure de probabilité $\mathbb{P}'$ définie par

$$\mathbb{P}'(\{1\}) = 0\ , \quad \mathbb{P}'(\{2\}) = \mathbb{P}'(\{3\}) = \mathbb{P}'(\{4\}) = \mathbb{P}'(\{5\}) = \mathbb{P}'(\{6\}) = 1/5\ ,$$

(cette mesure correspond à une situation où le dé n'est pas équilibré) alors on a $\mathbb{P}'(A) = 3/5$, $\mathbb{P}'(B) = 2/5$ et $\mathbb{P}'(A \cap C) = 1/5$ et les événements ne sont pas indépendants par rapport à $\mathbb{P}'$.

Il aurait été naturel de définir l'indépendance en utilisant la notion de probabilité conditionnelle. Plus précisément, on aurait pu opter pour la définition suivante : A et B sont indépendants si $\mathbb{P}(A \mid B) = \mathbb{P}(A)$. En effet, cette identité signifie que la réalisation de l'événement B n'influence pas la probabilité de réalisation de l'événement A ce qui est l'idée intuitive que l'on se fait de l'indépendance.
Malheureusement cette définition n'est pas générale car elle ne couvre pas le cas où $\mathbb{P}(B) = 0$, et par ailleurs elle ne semble pas faire jouer des rôles symétriques aux événements A et B. Cependant, on a le résultat suivant :

Proposition 3.7. *Si A et B sont deux événements tels que $\mathbb{P}(B) > 0$, alors A et B sont indépendants si et seulement si $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.*

Démonstration. Par définition A et B sont indépendants si et seulement si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Comme $\mathbb{P}(B) > 0$, on a équivalence entre $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ et $\mathbb{P}(A \mid B) = \mathbb{P}(A)$, d'où le résultat. □

L'indépendance "ne fait pas de différence" entre un événement et son complémentaire. Ceci est très naturel : savoir si E s'est produit ou pas est équivalent à savoir si E^c s'est produit ou pas.

Proposition 3.8. *Soient A et B deux événements, les quatre propositions suivantes sont équivalentes :*

1. *A et B sont indépendants,*
2. *A^c et B sont indépendants,*
3. *A et B^c sont indépendants,*
4. *A^c et B^c sont indépendants,*

Démonstration. Par symétrie, il suffit de montrer que le premier énoncé implique le deuxième. On note que

$$\begin{aligned}\mathbb{P}(A^c \cap B) &= \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) \\ &= \mathbb{P}(B)(1 - \mathbb{P}(A)) = \mathbb{P}(B)\mathbb{P}(A^c).\end{aligned}$$

□

La notion d'indépendance se généralise à une collection finie d'événements. Cependant, cette généralisation est plus subtile qu'on pourrait le penser.

Définition 3.9. *Soit $(E_1, \ldots, E_n)$ une suite finie d'événements. On dit que les événements $E_1, \ldots, E_n$ sont indépendants (on dit aussi mutuellement indépendants) si pour tout entier $k \leq n$ et pour tous $1 \leq i_1 < \ldots < i_k \leq n$, on a*

$$\mathbb{P}(\cap_{\ell=1}^{k} E_{i_\ell}) = \prod_{\ell=1}^{k} \mathbb{P}(E_{i_\ell}).$$

La subtilité vient du fait que l'on n'impose pas seulement l'égalité $\mathbb{P}(E_1 \cap \ldots \cap E_n) = \prod_{i=1}^n \mathbb{P}(E_i)$, mais également cette égalité pour toute sous-famille $E_{i_1}, \ldots, E_{i_k}$. En cela, le cas $n = 2$ est tout à fait spécial. Une justification pour cette définition plus subtile est qu'elle assure une stabilité de l'indépendance par restriction :

Proposition 3.10. *Soit $(E_1, \ldots, E_n)$ une suite finie d'événements indépendants. Alors toute sous-suite $(E_{i_1}, \ldots, E_{i_k})$ est formée d'événements indépendants.*

Démonstration. Il suffit de constater que toute sous-suite de $(E_{i_1}, \ldots, E_{i_k})$ est elle-même une sous-suite de $(E_1, \ldots, E_n)$. □

Exemple 3.10.1. On considère l'expérience suivante : une pièce est tirée deux fois successivement. On modélise l'expérience par $\Omega = \{P, F\}^2$ et on considère la probabilité uniforme. On pose $A = \{PF, PP\}$, $B = \{FP, PP\}$ et $C = \{FF, PP\}$, c'est-à-dire A dit que le premier résultat est pile, B que le second résultat est pile et C que les résultats des deux tirages sont égaux.

On peut vérifier que chaque paire (A, B), (A, C) et (B, C) est formée d'événements indépendants. Par contre

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{PP\}) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) \ ,$$

et ainsi les événements (A, B, C) ne sont pas indépendants.

On verra en TD qu'il existe des collections d'événements qui vérifient $\mathbb{P}(E_1 \cap \ldots \cap E_n) = \prod_{i=1}^n \mathbb{P}(E_i)$ sans être pour autant mutuellement indépendants.

On énonce à présent la notion d'indépendance pour des suites infinies d'événements.

Définition 3.11. *Soit $(E_n)_{n\in\mathbb{N}}$ une suite infinie d'événements. On dit que les E_n sont indépendants si et seulement si pour tout $k \geq 1$, pour tout $n \geq 1, n_1 < \ldots < n_k$, on a*

$$\mathbb{P}(\cap_{i=1}^k E_{n_i}) = \prod_{i=1}^k \mathbb{P}(E_{n_i}).$$

Remarque 3.11.1. Il y a à nouveau une subtilité par rapport au cas d'un nombre fini de variables : il n'est pas nécessaire de vérifier toutes les intersections possibles mais seulement celles faisant intervenir un nombre fini de variables.

Enfin, même si la définition de l'indépendance pour une suite infinie d'événements ne traite pas des intersections infinies, la proposition suivante montre que l'échange intersection/produit s'étend à ce cadre.

Proposition 3.12. *Soit $(E_n)_{n\in\mathbb{N}}$ une suite infinie d'événements indépendants et soit $(n_i)_{i\in\mathbb{N}}$ une sous-suite infinie strictement croissante. On a*

$$\mathbb{P}(\cap_{i=0}^\infty E_{n_i}) = \prod_{i=0}^\infty \mathbb{P}(E_{n_i}).$$

Démonstration. Pour tout $N > 0$, on remarque que $\mathbb{P}(\cap_{i=0}^{N} E_{n_i}) = \prod_{i=0}^{N} \mathbb{P}(E_{n_i})$ par définition de l'indépendance. Par ailleurs $N \mapsto \cap_{i=0}^{N} E_{n_i}$ est une suite décroissante d'événements dont la limite est $\cap_{i=0}^{\infty} E_{n_i}$ donc par la continuité décroissante des probabilités, $\mathbb{P}(\cap_{i=0}^{\infty} E_{n_i}) = \lim \prod_{i=0}^{N} \mathbb{P}(E_{n_i}) = \prod_{i=0}^{\infty} \mathbb{P}(E_{n_i})$. □

Chapitre 4

Variables aléatoires

4.1 Définitions

Définition 4.1. *Soit $(\Omega, \mathbb{P})$ un espace de probabilité et soit E un ensemble dénombrable. Une application $X : \Omega \to E$ est appelée* ***variable aléatoire*** *à valeurs dans E.*
La mesure de probabilité μ_X sur E définie par

$$\mu_X : \begin{cases} \mathcal{P}(E) & \to [0,1] \\ A & \mapsto \mathbb{P}(\{\omega : X(\omega) \in A\}) \end{cases}$$

est la ***loi*** *de X.*

On écrit parfois $\{\omega : X(\omega) \in A\} = X^{-1}(A)$. Nous avons affirmé que μ_X est une mesure de probabilité sur E, il faut vérifier que c'est bien le cas :

Démonstration. Tout d'abord $\mu_X(A) = \mathbb{P}(\{\omega : X(\omega) \in A\}) \geq 0$ pour tout $A \in \mathcal{P}(E)$. Par ailleurs, les ensembles suivants étant triviaux

$$\{\omega : X(\omega) \in E\} = \Omega\ , \quad \{\omega : X(\omega) \in \emptyset\} = \emptyset\ ,$$

on obtient $\mu_X(E) = 1$ et $\mu_X(\emptyset) = 0$. Enfin soit A_n une suite disjointe d'événements de E, vérifions que $\mu_X(\cup_n A_n) = \sum_n \mu_X(A_n)$. On a $\{\omega : X(\omega) \in \cup_n A_n\} = \cup_n \{\omega : X(\omega) \in A_n\}$ et ces ensembles sont disjoints puisque $X(\omega)$ ne peut pas prendre deux valeurs à la fois. Ainsi en utilisant le fait que $\mathbb{P}$ est une probabilité, on obtient

$$\mu_X(\cup_n A_n) = \mathbb{P}(\{\omega : X(\omega) \in \cup_n A_n\}) = \sum_n \mathbb{P}(\{\omega : X(\omega) \in A_n\}) = \sum_n \mu_X(A_n)\ .$$

□

Par la suite, nous utiliserons les abréviations suivantes :

$$\{X \in A\} = \{\omega : X(\omega) \in A\}\ .$$

De même nous écrirons $\{X = a\} = \{\omega : X(\omega) = a\}$, $\{X \leq a\} = \{\omega : X(\omega) \leq a\}$, etc. Par ailleurs, nous nous permettrons de ne plus écrire les accolades autour des événements dans le calcul de leurs probabilités afin d'alléger les notations :

$$\mathbb{P}(X \in A) = \mathbb{P}(\{X \in A\}) .$$

Remarque 4.1.1. D'après la Proposition 2.6, la loi d'une variable aléatoire X à valeur dans E est entièrement déterminée par la fonction $A \ni a \mapsto \mathbb{P}(X = a)$.

Remarque 4.1.2. La loi d'une variable aléatoire dépend de la mesure de probabilité $\mathbb{P}$ considérée. Si Ω est muni d'une autre mesure de probabilité, alors il faut préciser la mesure par rapport à laquelle on travaille. Nous reviendrons sur ce point quand nous introduirons la loi conditionnelle.

Exemple 4.1.3. Reprenons l'exemple du lancer de deux dés à six faces du Chapitre 2 dans lequel $\Omega = \{1, \dots, 6\}^2$. On note X_1 le résultat du premier lancer, X_2 le résultat du second lancer et S la somme des deux résultats. Plus précisément, pour tout élément $\omega = (\omega_1, \omega_2) \in \Omega$ on pose

$$X_1(\omega) = \omega_1 , \quad X_2(\omega) = \omega_2 , \quad S(\omega) = \omega_1 + \omega_2 .$$

On voit alors que X_1 et X_2 sont des variables aléatoires à valeurs dans $\{1, \dots, 6\}$ alors que S est une variable aléatoire à valeurs dans $\{2, \dots, 12\}$.

Par ailleurs

$$\begin{aligned}
\mu_{X_1}(\{1\}) = \mathbb{P}(X_1 = 1) &= \mathbb{P}\big(\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)\}\big) = 1/6 \\
\mu_{X_1}(\{2\}) = \mathbb{P}(X_1 = 2) &= \mathbb{P}\big(\{(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)\}\big) = 1/6 \\
\mu_{X_1}(\{3\}) = \mu_{X_1}(\{4\}) &= \mu_{X_1}(\{5\}) = \mu_{X_1}(\{6\}) = 1/6 .
\end{aligned}$$

De même

$$\begin{aligned}
\mu_{X_2}(\{1\}) = \mathbb{P}(X_2 = 1) &= \mathbb{P}\big(\{(1,1),(2,1),(3,1),(4,1),(5,1),(6,1)\}\big) = 1/6 \\
\mu_{X_2}(\{2\}) = \mu_{X_2}(\{3\}) &= \mu_{X_2}(\{4\}) = \mu_{X_2}(\{5\}) = \mu_{X_2}(\{6\}) = 1/6 .
\end{aligned}$$

On voit donc que $\mu_{X_1} = \mu_{X_2}$.
Enfin on peut vérifier que

$$\begin{aligned}
&\mathbb{P}(S = 2) = \frac{1}{36} && \mathbb{P}(S = 3) = \frac{2}{36} && \mathbb{P}(S = 4) = \frac{3}{36} \\
&\mathbb{P}(S = 5) = \frac{4}{36} && \mathbb{P}(S = 6) = \frac{5}{36} && \mathbb{P}(S = 7) = \frac{6}{36} \\
&\mathbb{P}(S = 8) = \frac{5}{36} && \mathbb{P}(S = 9) = \frac{4}{36} && \mathbb{P}(S = 10) = \frac{3}{36} \\
&\mathbb{P}(S = 11) = \frac{2}{36} && \mathbb{P}(S = 12) = \frac{1}{36}
\end{aligned}$$

Retenons de l'exemple précédent que deux variables aléatoires distinctes peuvent avoir la même loi. Observons par ailleurs que le choix de l'ensemble d'arrivée E est quelque peu arbitraire : pour X_1 on aurait pu prendre $E = \mathbb{N}$, mais alors la loi de X_1 aurait attribué la valeur 0 à tous les entiers $n \notin \{1, \dots, 6\}$. Il y a donc un choix "naturel" d'ensemble E qui est l'ensemble des valeurs prises avec probabilité positive.

4.2 Conditionnement, indépendance et variables aléatoires

Définition 4.2. *Soit $(\Omega, \mathbb{P})$ un espace probabilisé. Soit X une variable aléatoire à valeurs dans un ensemble E et soit B un événement de probabilité positive. La loi conditionnelle de X sachant B est la mesure de probabilité définie par*

$$\mu_X(\{a\} \mid B) = \mathbb{P}(\{X = a\} \mid B) \ , \quad a \in E \ .$$

A la Proposition 3.2 nous avons vu que $\mathbb{P}(\mid B)$ est une mesure de probabilité. La loi conditionnelle de X sachant B peut s'interpréter comme la loi de X par rapport à la mesure de probabilité $\mathbb{P}(\ \cdot\ \mid B)$ (et non $\mathbb{P}$), voir la Remarque 4.1.2.

Exemple 4.2.1. On reprend le cas du lancer de dé. Si l'on note B l'événement "le premier dé est pair", c'est-à-dire

$$B = \{(\omega_1, \omega_2) \in \Omega : \omega_1 \in \{2, 4, 6\}\} \ ,$$

alors la loi conditionnelle de X_1 sachant B est donnée par

$$\begin{aligned} &\mu_X(\{1\} \mid B) = \mu_X(\{3\} \mid B) = \mu_X(\{5\} \mid B) = 0 \ , \\ &\mu_X(\{2\} \mid B) = \mu_X(\{4\} \mid B) = \mu_X(\{6\} \mid B) = 1/3 \ . \end{aligned}$$

Passons à la notion d'indépendance de variables aléatoires.

Définition 4.3. *Soient $X_1, \ldots, X_n$ des variables aléatoires à valeurs respectivement dans des ensembles dénombrables $E_1, \ldots, E_n$. Ces variables sont indépendantes (on dit parfois mutuellement indépendantes) si pour tous sous-ensembles $F_1 \subset E_1, \ldots, F_n \subset E_n$, on a*

$$\mathbb{P}(\{X_1 \in F_1\} \cap \cdots \cap \{X_n \in F_n\}) = \prod_i \mathbb{P}(X_i \in F_i).$$

La proposition suivante établit un lien avec la notion d'indépendance d'événements, et présente une caractérisation plus simple de l'indépendance.

Proposition 4.4. *Avec les mêmes notations que dans la définition, les variables $X_1, \ldots, X_n$ sont indépendantes si et seulement si l'une des propriétés équivalentes suivantes est vérifiée :*

1. *Pour tous $x_1 \in E_1, \ldots, x_n \in E_n$, $\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod \mathbb{P}(X_i = x_i)$.*
2. *Pour tous ensembles $F_1 \subset E_1, \ldots, F_n \subset E_n$, les événements $\{X_i \in F_i\}$, $1 \leq i \leq n$ sont indépendants.*

Démonstration. La première propriété est clairement un cas particulier de la définition. Montrons qu'elle implique aussi la définition. Soient $F_1 \subset E_1, \ldots, F_n \subset E_n$. On peut écrire

$$\{X_i \in F_i\} = \cup_{x_i \in F_i} \{X_i = x_i\} \ ,$$

et ainsi

$$\{X_1 \in F_1\} \cap \cdots \cap \{X_n \in F_n\} = \cup_{x_1 \in F_1, \ldots, x_n \in F_n} \{X_1 = x_1\} \cap \cdots \cap \{X_n = x_n\} .$$

Dans le terme de droite, l'union porte sur une collection dénombrable d'événements disjoints. Ainsi

$$\mathbb{P}(\{X_1 \in F_1\} \cap \cdots \cap \{X_n \in F_n\}) = \sum_{x_1 \in F_1, \ldots, x_n \in F_n} \mathbb{P}(\{X_1 = x_1\} \cap \cdots \cap \{X_n = x_n\}) .$$

La première propriété implique alors que

$$\begin{aligned}
\mathbb{P}(\{X_1 \in F_1\} \cap \cdots \cap \{X_n \in F_n\}) &= \sum_{x_1 \in F_1, \ldots, x_n \in F_n} \prod_{i=1}^{n} \mathbb{P}(X_i = x_i) \\
&= \prod_{i=1}^{n} \sum_{x_i \in F_i} \mathbb{P}(X_i = x_i) \\
&= \prod_{i=1}^{n} \mathbb{P}(X_i \in F_i)
\end{aligned}$$

et l'indépendance est vérifiée.
La deuxième propriété implique l'indépendance des variables aléatoires. On pourrait penser que cette propriété est même plus forte que l'indépendance puisqu'elle impose des conditions sur toute sous-collection des X_i. Cependant on remarque que pour $i_1 < \ldots < i_k$, on peut écrire $\cap_{\ell=1}^{k} \{X_{i_\ell} \in F_{i_\ell}\} = \cap_{i=1}^{n} \{X_i \in \tilde{F}_i\}$ où $\tilde{F}_i = F_i$ si i appartient à l'ensemble $\{i_1, \ldots i_k\}$ et $\tilde{F}_i = E_i$ sinon. Cela permet de déduire que la deuxième condition est équivalente à l'indépendance des X_i. □

Comme nous l'avions fait pour les suites d'événements, nous énonçons la notion d'indépendance pour des suites de variables aléatoires.

Définition 4.5. *Une suite infinie $(X_i)_{i \in \mathbb{N}}$ de variables aléatoires est indépendante si et seulement si pour tout entier k et pour tous $i_1 < \ldots < i_k$ les variables $X_{i_1}, \ldots, X_{i_k}$ sont indépendantes.*

4.3 Lois usuelles

Dans cette partie, nous donnons une liste de mesures de probabilité qui sont "classiques". Ces mesures sont des lois usuelles au sens où, dans de nombreux problèmes, apparaissent des variables aléatoires dont la loi est donnée par l'une de ces lois usuelles.

Loi de Bernoulli

Définition 4.6. *La loi de Bernoulli de paramètre $p \in [0, 1]$ est la mesure de probabilité sur $E = \{0, 1\}$ définie par*

$$\mu(\{1\}) = p , \quad \mu(\{0\}) = 1 - p.$$

On la note $B(p)$ ou $Ber(p)$.

Loi Uniforme

Définition 4.7. *Soit E un ensemble fini. La loi uniforme sur E est définie par*

$$\mu(\{a\}) = \frac{1}{|E|} , \quad a \in E .$$

On la note $U(E)$ ou $Unif(E)$.

Loi Binomiale

Définition 4.8. *Soit $n \geq 1$ un entier et $p \in [0,1]$. La loi binomiale de paramètres n et p est la mesure de probabilité sur $E = \{0, \dots, n\}$ définie par*

$$\mu(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k} , \quad k \in E .$$

On la note $B(n,p)$ ou $Bin(n,p)$.

Ceci définit bien une probabilité : il suffit d'appliquer la formule du binôme de Newton $(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$ avec $a = p$ et $b = 1-p$.

Proposition 4.9. *Soit $X_1, \dots, X_n$ des variables aléatoires indépendantes de loi $Ber(p)$ pour un certain paramètre $p \in [0,1]$. La variable aléatoire $S = \sum_{i=1}^n X_i$ a pour loi $B(n,p)$.*

Démonstration. Il est clair que S prend ses valeurs dans $\{0, \dots, n\}$. Soit un entier k entre 0 et n, on a

$$\{S = k\} = \bigcup_{\substack{x_1, \dots x_n \in \{0,1\} \\ \sum x_i = k}} \{X_1 = x_1, \cdots, X_n = x_n\} ,$$

et l'union porte sur des événements disjoints. Ainsi, en utilisant l'indépendance à la deuxième ligne

$$\begin{aligned} \mathbb{P}(S = k) &= \sum_{\substack{x_1, \dots x_n \in \{0,1\} \\ \sum x_i = k}} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\ &= \sum_{\substack{x_1, \dots x_n \in \{0,1\} \\ \sum x_i = k}} \mathbb{P}(X_1 = x_1) \dots \mathbb{P}(X_n = x_n) \end{aligned}$$

On observe alors que pour tout n-uplet $x_1, \dots, x_n$ tel que $\sum x_i = k$ on a

$$\mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n) = p^k (1-p)^{n-k} .$$

Or il y a un total de $\binom{n}{k}$ tels n-uplets. On en déduit que

$$\mathbb{P}(S = k) = \binom{n}{k} p^k (1-p)^{n-k} ,$$

ce qui suffit à identifier la loi de S. □

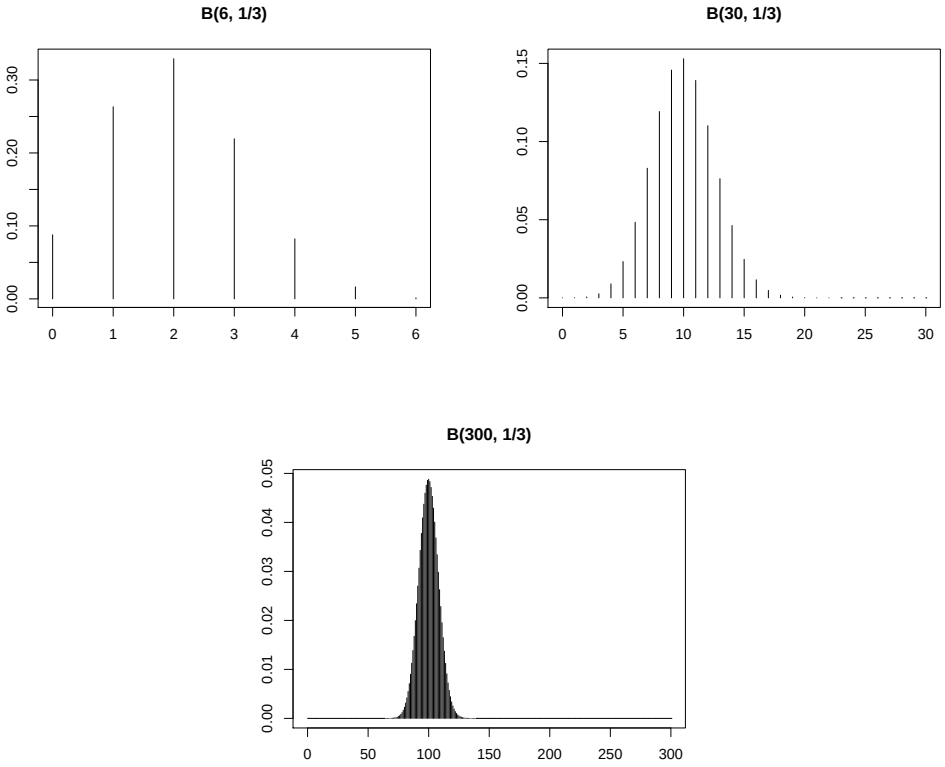


FIGURE 4.1 – Loi binomiale pour $n = 6$, $n = 30$ et $n = 300$ et $p = 1/3$.

Corollaire 4.10. *Soient $n_1, n_2 \geq 1$ des entiers et soit $p \in [0, 1]$. Soit S_1 une variable de loi $B(n_1, p)$ et S_2 une variable de loi $B(n_2, p)$ indépendante de X_1. La loi de $S_1 + S_2$ est $B(n_1 + n_2, p)$.*

Démonstration. Par la proposition précédente, on peut constuire S_1 et S_2 à l'aide de n_1 et n_2 variables aléatoires indépendantes de loi $\mathrm{Ber}(p)$. On voit alors que $S_1 + S_2$ est la somme de $n_1 + n_2$ variables indépendantes de loi $\mathrm{Ber}(p)$ ce qui suffit à conclure. □

Loi de Poisson

Définition 4.11. *La loi de Poisson de paramètre $\lambda > 0$ est la mesure de probabilité sur $E = \mathbb{N}$ définie par*

$$\mu(\{n\}) = e^{-\lambda} \frac{\lambda^n}{n!} , \quad n \geq 0 .$$

On la note $\mathcal{P}(\lambda)$ ou $Po(\lambda)$.

Le fait que la formule définisse bien une probabilité vient du développement en série de la fonction exponentielle :

$$e^{\lambda} = \sum_{n \geq 0} \frac{\lambda^n}{n!} \, .$$

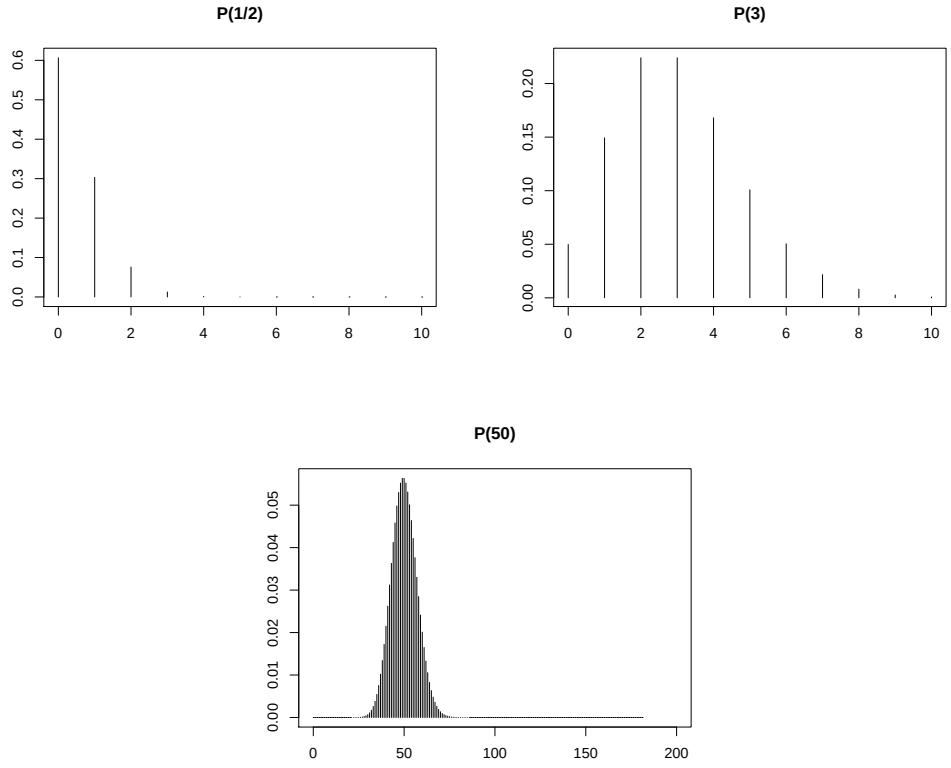


FIGURE 4.2 – Loi de Poisson pour des paramètres 1/2, 3 et 50.

La loi de Poisson est parfois appelée loi des événements rares : elle apparait comme une bonne approximation de la loi binomiale $B(n, \frac{\lambda}{n})$ pour n très grand. Par exemple, si un composant électronique a une chance sur 1000 d'être défectueux (indépendante pour chaque composant) et qu'une usine fabrique 1500 composants par jour, alors la loi du nombre de composants défectueux est donnée par une binomiale $\mathrm{B}(1500, 1/1000)$. Cette loi est bien approximée par une loi de Poisson de paramètre $\lambda = 3/2$. On notera qu'il est plus pratique de manipuler la loi de Poisson que la loi binomiale :

$$e^{-3/2}\frac{(3/2)^3}{6} \quad \text{vs} \quad \frac{1500 \times 1499 \times 1498}{6} \left(\frac{1}{1000}\right)^3 \left(\frac{999}{1000}\right)^{1497} .$$

Le résultat suivant "justifie" l'approximation énoncée ci-dessus.

Proposition 4.12. *Pour tout k entier et pour tout $\lambda > 0$,*

$$\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \to_{n\to\infty} e^{-\lambda}\frac{\lambda^k}{k!}.$$

Démonstration. On note $u_n = \binom{n}{k}(\frac{\lambda}{n})^k\left(1-\frac{\lambda}{n}\right)^{n-k}$. Lorsque $n \to \infty$, on observe que

$$u_n = \frac{\lambda^k}{k!}\underbrace{\frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{n^k}}_{\to 1}\left(1-\frac{\lambda}{n}\right)^n\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\to 1}$$

donc il suffit de montrer que $(1-\frac{\lambda}{n})^n \to e^{-\lambda}$. Le développement limité du logarithme donne

$$\begin{aligned}\left(1-\frac{\lambda}{n}\right)^n &= \exp\left(n\ln\left(1-\frac{\lambda}{n}\right)\right) = \exp\Big(-n\frac{\lambda}{n}\left(1+o(1)\right)\Big)\\ &\to e^{-\lambda}.\end{aligned}$$

□

Les variables de Poisson vérifient la propriété remarquable suivante.

Proposition 4.13. *Si X et Y sont deux variables aléatoires indépendantes de lois de Poisson de paramètres respectifs λ et μ, alors $X+Y$ suit une loi de Poisson de paramètre $\lambda+\mu$.*

Démonstration. Il est clair que $X+Y$ est à valeurs dans $\mathbb{N}$. On fixe $k \in \mathbb{N}$, et l'on calcule $\mathbb{P}(X+Y=k)$. Par la caractérisation simple des probabilités de la proposition 2.6 cela suffit à conclure. On a

$$\begin{aligned}\mathbb{P}(X+Y=k) &= \sum_{i=0}^{k}\mathbb{P}(X=i)\mathbb{P}(Y=k-i)\\ &= \sum_{i=0}^{k} e^{-\lambda}\frac{\lambda^i}{i!}e^{-\mu}\frac{\mu^{k-i}}{(k-i)!}\\ &= e^{-(\lambda+\mu)}\frac{1}{k!}\sum_{i=0}^{k}\binom{k}{i}\lambda^i\mu^{k-i}\\ &= e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^k}{k!}.\end{aligned}$$

□

Loi géométrique

Définition 4.14. *La loi géométrique de paramètre $p \in [0,1]$ est la mesure de probabilité sur $E = \mathbb{N}$ donnée par*

$$\mu(\{n\}) = p(1-p)^n.$$

On la note $\mathrm{Geom}(p)$.

Le fait que l'expression définisse bien une probabilité vient la formule pour la somme des termes d'une suite géométrique :

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}\ , \quad \forall x \in]-1,1[\ .$$

Une loi géométrique représente un temps d'attente aléatoire si à chaque instant un événement peut se produire indépendamment du passé. Plus précisément on a la proposition suivante.

Proposition 4.15. *Soit $(X_n)_{n\in\mathbb{N}}$ une suite infinie de variables aléatoires indépendantes de loi $B(p)$ pour un certain $p \in]0,1]$* [1]*. Soit $N = \inf\{n \geq 0 : X_n = 1\}$, c'est-à-dire que N est le "temps" où le premier "1" apparaît. N est une variable aléatoire de loi* $\mathrm{Geom}(p)$.

Démonstration. Il est clair que N prend ses valeurs dans $\mathbb{N}$. Soit $k \in \mathbb{N}$, on a

$$\begin{aligned}\mathbb{P}(N = k) &= \mathbb{P}(X_0 = 0, \dots, X_{k-1} = 0, X_k = 1)\\ &= \mathbb{P}(X_k = 1) \prod_{n=0}^{k-1} \mathbb{P}(X_i = 0)\\ &= p(1-p)^k.\end{aligned}$$

□

Remarque 4.15.1. Malheureusement, la convention de notation pour la loi géométrique n'est pas complètement fixée. Vous pourrez aussi voir les conventions suivantes :

- $E = \mathbb{N}$, $\mu(n) = (1-p)p^n$;
- $E = \mathbb{N}^*, \mu(n) = p(1-p)^{n-1} = \frac{p}{1-p}(1-p)^n$;
- $E = \mathbb{N}^*, \mu(n) = (1-p)p^{n-1} = \frac{1-p}{p}p^n$;

c'est-à-dire un échange de p et $1-p$ et/ou un décalage de 1 dans l'indiçage de la suite.

Autres lois classiques Nous présentons à présent des lois qui ne sont pas à connaître pour ce cours, mais qui reviennent régulièrement en probabilités.

1. Rigoureusement, nos définitions ne s'appliquent pas à ce cadre. Comme mentionné plus haut, on suppose que l'on peut définir proprement une telle suite et que l'on peut calculer des probabilités sur cette suite comme si l'espace était dénombrable.

Loi de Rademacher

Définition 4.16. *La loi de Rademacher de paramètre p est la mesure de probabilité sur $E = \{-1, 1\}$ donnée par*

$$\mu(\{-1\}) = 1 - p, \quad \mu(\{1\}) = p.$$

Si p n'est pas précisé, on prendra $p = \frac{1}{2}$.

Loi hypergéométrique

Définition 4.17. *La loi hypergéométrique de paramètres $N \geq K \geq 1$ et $N \geq n \geq 1$ est la mesure de probabilité sur $E = \{0, 1, \dots, , K\}$ définie par*

$$\mu(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.$$

Proposition 4.18. *On considère une urne contenant K boules blanches et $N - K$ boules noires. On tire successivement et **sans remise** n boules dans l'urne. Le nombre de boules blanches obtenues suit une loi hypergéométrique de paramètres n, K, N.*

Loi multinomiale

Définition 4.19. *Soient $p_1, \dots, p_k \in [0, 1]$ tels que $\sum_i p_i = 1$. Soit $N \geq 1$ un entier. La loi multinomiale de paramètres $p_1, \dots, p_k, N$ est la mesure de probabilité sur*

$$E = \{(x_1, \dots, x_k) \in \{0, \dots, N\}^k : \sum_i x_i = N\},$$

définie par

$$\mu(\{(x_1, \dots, x_k)\}) = \frac{N!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}.$$

C'est notre premier exemple de loi qui ne soit pas celle d'un nombre entier aléatoire mais d'un vecteur aléatoire de dimension k dont chaque composante est un entier.

Proposition 4.20. *Soit $(Y_n)_{1 \leq n \leq N}$ des variables indépendantes et de même loi donnée par $\mathbb{P}(Y_n = i) = p_i$ pour tout i entre 1 et k. Soit X_i le nombre de fois que la valeur i apparaît dans la suite Y_n, c'est-à-dire $X_i = \sum_n 1_{Y_n = i}$. Le vecteur $(X_1, \dots, X_k)$ suit la loi multinomiale de paramètre N et $p_1, \dots, p_k$.*

Démonstration. Il faut dénombrer le nombre de suites possibles contenant exactement x_1 fois la valeur 1, x_2 fois la valeur 2, ... C'est un problème d'anagramme que l'on a déjà abordé au Chapitre 1. □

La proposition montre que c'est une sorte de généralisation de la loi binomiale. Plus précisément, si on considère N variables qui prennent deux valeurs possibles, 1 avec probabilité p et 2 avec probabilité $1 - p$, alors le nombre X_1 de 1 est par définition une variable $\mathrm{B}(n, p)$ et la paire $(X_1, N - X_1)$ est un vecteur aléatoire de loi multinomiale de paramètres $p, 1 - p$ et N.

Loi de Pareto discrète

Définition 4.21. *La loi de Pareto discrète de paramètre $\alpha > 1$ est la mesure de probabilité sur $E = \mathbb{N}^*$ définie par*

$$\mu(n) = \frac{1}{\zeta(\alpha)} \frac{1}{n^\alpha},$$

où l'on a posé $\zeta(\alpha) = \sum_{n=1}^{\infty} \frac{1}{n^\alpha}$.

4.4 Fonction d'une variable aléatoire

Une façon d'obtenir de nouvelles variables aléatoires est de prendre des fonctions de variables aléatoires existantes.

Soit $(\Omega, \mathbb{P})$ un espace probabilisé. Soit X une variable aléatoire à valeurs dans un ensemble dénombrable E. Soit $g : E \to F$ une fonction, avec F un ensemble dénombrable. On peut définir une variable aléatoire $Y = g(X)$.

Proposition 4.22. *La variable aléatoire Y est bien définie et sa loi μ_Y sur F est donnée, pour tout $A \subset F$, par*

$$\mu_Y(\{A\}) = \mu_X(\{x \in E : g(x) \in A\}).$$

Démonstration. On a bien que Y est une fonction de Ω dans F, donnée par $Y(\omega) = g(X(\omega))$ pour tout $\omega \in \Omega$, donc c'est une variable aléatoire. On a alors par définition

$$\begin{aligned}
\mu_Y(A) &= \mathbb{P}(Y \in A) = \mathbb{P}(\{\omega \in \Omega : g(X(\omega)) \in A\}) \\
&= \mathbb{P}\left(\bigcup_{x \in E : g(x) \in A} \{\omega \in \Omega : X(\omega) = x\} \right) \\
&= \sum_{x \in E : g(x) \in A} \mu_X(\{x\}) = \mu_X(\{x \in E : g(x) \in A\}),
\end{aligned}$$

où l'on a utilisé le fait que l'union est disjointe. □

En appliquant la proposition ci-dessus à $A = \{y\}$ avec $y \in F$, on trouve une formule utile en pratique :

$$\mu_Y(\{y\}) = \mu_X(\{x \in E : g(x) = y\}) = \sum_{x \in E : g(x) = y} \mu_X(\{x\}). \tag{4.4.1}$$

Exemple 4.22.1. Soit X une variable aléatoire à valeurs dans $E = \{-1, 0, 1\}$ dont la loi est donnée par

$$\mu_X(\{-1\}) = \frac{1}{6}, \quad \mu_X(\{0\}) = \frac{1}{2}, \quad \mu_X(\{1\}) = \frac{2}{6}.$$

Soit Y la variable aléatoire à valeurs dans $F = \{0, 1\}$ donnée par $Y = X^2$. On trouve alors

$$\mu_Y(\{0\}) = \sum_{x \in E : x^2 = 0} \mu_X(\{x\}) = \mu_X(\{0\}) = \frac{1}{2}$$

et

$$\mu_Y(\{1\}) = \sum_{x\in E: x^2=1} \mu_X(\{x\}) = \mu_X(\{-1\}) + \mu_X(\{1\}) = \frac{1}{6} + \frac{2}{6} = \frac{1}{2}.$$

Chapitre 5

Espérance

Dans ce chapitre nous introduisons la notion d'espérance, qui est absolument centrale en théorie des probabilités. L'espérance donne un sens rigoureux au concept de moyenne d'une variable aléatoire. Sa définition est assez élémentaire dans le cas particulier où la variable aléatoire prend un nombre fini de valeurs. Par contre, la définition est plus délicate lorsque l'ensemble d'arrivée est infini dénombrable.

5.1 Définition et exemples

Commençons par quelques notions préliminaires. On rappelle la propriété suivante vérifiée par les sommes de termes positifs : la somme d'une collection finie ou infinie dénombrable de termes positifs est toujours bien définie, et l'ordre dans lequel l'on somme ces termes n'influe pas sur la finitude ni sur la valeur de cette somme. Par ailleurs ces sommes vérifient une propriété de sommation par paquets.

Soit I un ensemble fini ou infini dénombrable, et soit $(x_i)_{i\in I}$ une collection de réels indicée par I. On dira que $\sum_{i\in I} x_i$ est une somme bien définie si **au moins** une des deux sommes suivantes est finie :

$$\sum_{i\in I : x_i \in]0,\infty[} x_i < \infty \ , \qquad \sum_{i\in I : x_i \in]-\infty,0[} x_i > -\infty \ .$$

Dans ce cas, on pose

$$\sum_{i\in I} x_i := \sum_{i\in I : x_i \in]0,\infty[} x_i + \sum_{i\in I : x_i \in]-\infty,0[} x_i \ .$$

Cette quantité est ainsi un élément de $[-\infty, +\infty]$.

Quelques cas particuliers :

1. Si tous les x_i, $i \in I$ sont positifs alors la somme est bien définie même si la somme des x_i vaut $+\infty$: en effet, la somme portant sur les valeurs négatives est vide donc finie.

2. De même si tous les termes sont négatifs, alors la somme est bien définie même si cette somme peut valoir $-\infty$.
3. Si les x_i, $i \in I$ sont tous positifs (ou tous négatifs) sauf pour un nombre fini d'entre eux, alors la somme est bien définie.
4. Si $I = \mathbb{N}$, et $x_i = (-i)^n$, alors $\sum_{i\in I} x_i$ n'est pas définie car les deux sommes valent respectivement $-\infty$ et $+\infty$.

On notera que cette notion de somme bien définie n'implique pas la convergence de la série, et ne couvre pas toutes les séries alternées : par exemple $(-1)^n/n$, $n \geq 1$ ne donne pas lieu à une somme bien définie.
On notera enfin que les sommes bien définies héritent d'une propriété importante des séries à termes positifs : l'ordre dans lequel les éléments sont sommés n'influe pas sur la convergence ni sur la valeur de la somme.

Dans la suite, on manipulera des variables aléatoires réelles discrètes, c'est-à-dire, des variables aléatoires à valeurs dans un ensemble $E \subset \mathbb{R}$ fini ou infini dénombrable.

Définition 5.1. *Soit X une variable aléatoire discrète définie sur un espace probabilisé $(\Omega, \mathbb{P})$ et à valeurs dans un ensemble (dénombrable) $E \subset \mathbb{R}$. Soit μ_X la loi de X. On dit que X admet une espérance si la somme*

$$\sum_{x\in E} x\mu_X(\{x\})$$

est bien définie, c'est-à-dire si au moins l'une des deux sommes suivantes est finie

$$\sum_{x\in E\cap]0,\infty[} x\mu_X(\{x\}) < \infty \ , \qquad \sum_{x\in E\cap]-\infty,0[} x\mu_X(\{x\}) > -\infty \ .$$

Si X admet une espérance, on pose alors

$$\mathbb{E}[X] := \sum_{x\in E} x\mu_X(\{x\}) \in [-\infty, +\infty] \ .$$

A ce stade, on peut observer que l'espérance d'une variable aléatoire ne dépend que de sa loi. En d'autres termes, deux variables aléatoires distinctes mais dont les lois coïncident ont même espérance.

On notera que si X est une variable aléatoire réelle discrète, alors $X_+ := \max(X, 0)$, $X_- := \max(-X, 0)$ et $|X|$ sont également des variables aléatoires réelles discrètes. Comme ces variables aléatoires sont de signe constant (en fait, positif), leur espérance est bien définie dans tous les cas (mais cette espérance peut être infinie). De plus, par définition,

$$X = X_+ - X_- \ , \qquad |X| = X_+ + X_- \ .$$

On a alors le résultat suivant :

Lemme 5.2. *Une variable aléatoire réelle discrète X admet une espérance si et seulement si au moins l'un des deux termes $\mathbb{E}[X_+]$, $\mathbb{E}[X_-]$ est fini et dans ce cas on a*

$$\mathbb{E}[X] = \mathbb{E}[X_+] - \mathbb{E}[X_-] \, .$$

Démonstration. Il s'agit d'une conséquence immédiate de la définition. □

Les deux termes $\mathbb{E}[X_+]$ et $\mathbb{E}[X_-]$ sont toujours bien définis, en revanche leur différence $\mathbb{E}[X_+] - \mathbb{E}[X_-]$ est indéterminée si tous deux valent l'infini et c'est pourquoi l'on exige qu'au moins l'un des deux termes soit fini pour que l'espérance de X existe !

Définition 5.3. *On dit qu'une variable aléatoire réelle discrète X est intégrable si $\mathbb{E}[|X|] < \infty$, c'est-à-dire, si $\mathbb{E}[X_+] < \infty$ et $\mathbb{E}[X_-] < \infty$.*

On notera qu'une v.a. intégrable admet toujours une espérance. En revanche, une v.a. peut admettre une espérance sans être intégrable.

En résumé :

	$\mathbb{E}[X_+] < \infty$	$\mathbb{E}[X_+] = \infty$
$\mathbb{E}[X_-] < \infty$	$-\infty < \mathbb{E}[X] < \infty$ X intégrable	$\mathbb{E}[X] = \infty$ X pas intég.
$\mathbb{E}[X_-] = \infty$	$\mathbb{E}[X] = -\infty$ X pas intég.	$\mathbb{E}[X]$ indéfinie X pas intég.

Exemple 5.3.1. On tire un dé équilibré. On pose $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathbb{P}$ uniforme et X la variable donnant le résultat du dé (c.à.d $X(\omega) = \omega$). La variable aléatoire prend des valeurs positives donc elle admet une espérance :

$$\mathbb{E}[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

Exemple 5.3.2. On tire deux fois de suite une pièce équilibrée et on compte le nombre de "pile". $\Omega = \{P, F\}^2$, $\mathbb{P}$ uniforme et $X(\omega)$ le nombre de P dans la paire ω. Dans ce cas :

$$\begin{aligned} \mathbb{E}[X] &= 0 \cdot \mathbb{P}(\{FF\}) + 1 \cdot \mathbb{P}(\{PF; FP\}) + 2 \cdot \mathbb{P}(\{PP\}) \\ &= 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1. \end{aligned}$$

Exemple 5.3.3. On considère une variable aléatoire X de loi de Poisson de paramètre $\lambda > 0$. Comme X est positive, elle admet une espérance et l'on a

$$\begin{aligned} \mathbb{E}[X] &= \sum_{n=0}^{\infty} n \mu_X(\{n\}) = \sum_{n=0}^{\infty} n \frac{\lambda^n e^{-\lambda}}{n!} = \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{(n-1)!} = \lambda \sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} \\ &= \lambda. \end{aligned}$$

Exemple 5.3.4. On considère une variable de Pareto discrète de paramètre 2. Comme X est positive, elle admet une espérance et l'on a

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} n \frac{\zeta^{-1}(2)}{n^2} = \sum_{n=1}^{\infty} \frac{\zeta^{-1}(2)}{n} = +\infty.$$

Cette variable aléatoire n'est donc pas intégrable.

Exemple 5.3.5. Si $X(\omega)$ est une variable aléatoire constante égale à x alors $\mathbb{E}[X] = x \cdot \mathbb{P}(\Omega) = x$.

Exemple 5.3.6. On considère un espace probabilisé $(\Omega, \mathbb{P})$, un événement A et la variable aléatoire $X = 1_A$, c'est-à-dire $X(\omega) = 1$ si $\omega \in A$ et $X(\omega) = 0$ si $\omega \notin A$. On a alors

$$\mathbb{E}[X] = 1 \cdot \mathbb{P}(A) = \mathbb{P}(A) \, .$$

5.2 Formule de transfert et quelques propriétés

Nous abordons à présent une formule très utile dans le calcul des espérances. Soient g une fonction de E dans $\mathbb{R}$ et X une variable aléatoire à valeurs dans un ensemble dénombrable E. Si la variable aléatoire réelle discrète $g(X)$ admet une espérance, alors cette espérance s'exprime en fonction de la loi de $g(X)$. La formule de transfert énoncée ci-dessous montre qu'on peut en fait exprimer cette espérance à l'aide de la loi de X !

Théoreme 5.4 (Formule de transfert)**.** *Soit X une variable aléatoire sur un espace probabilisé $(\Omega, \mathbb{P})$ à valeurs dans un espace dénombrable E. Soit $g : E \to F$ une fonction, avec $F \subset \mathbb{R}$ un ensemble dénombrable. La variable aléatoire réelle discrète $g(X)$ admet une espérance si et seulement si la somme $\sum_{x \in E} g(x)\mu_X(\{x\})$ est bien définie et dans ce cas*

$$\mathbb{E}[g(X)] = \sum_{x \in E} g(x)\mu_X(\{x\}) \, .$$

On remarquera que

$$\sum_{x \in E} g(x)\mu_X(\{x\}) = \sum_{x \in X(\Omega)} g(x)\mu_X(\{x\})$$

où $X(\Omega) \subset E$ désigne l'image par X de l'ensemble Ω.

Démonstration. On a montré dans (4.4.1) que pour tout $y \in \mathbb{R}$

$$\mu_{g(X)}(\{y\}) = \sum_{x \in E : g(x) = y} \mu_X(\{x\}) \, .$$

On suppose à présent que g est positive (donc $F \subset [0, \infty[$). Dans ce cas, $g(X)$ admet bien une espérance et la somme $\sum_{x\in E} g(x)\mu_X(\{x\})$ est bien définie. Les propriétés de sommation par paquets des séries positives assurent que

$$\begin{aligned}\mathbb{E}[g(X)] &= \sum_{x\in F} y\mu_{g(X)}(\{y\}) \\ &= \sum_{x\in F} y \sum_{x\in E:g(x)=y} \mu_X(\{x\}) \\ &= \sum_{x\in F} \sum_{x\in E:g(x)=y} g(x)\mu_X(\{x\}) \\ &= \sum_{x\in E} g(x)\mu_X(\{x\}) \ ,\end{aligned}$$

ce qui prouve dans le cas positif la formule énoncée.
Passons au cas où g est de signe quelconque. On sait que $g(X)$ admet une espérance si et seulement si au moins l'une des deux espérances de $g_+(X)$ et $g_-(X)$ est finie, et dans ce cas $\mathbb{E}[g(X)] = \mathbb{E}[g_+(X)] + \mathbb{E}[g_-(X)]$. On peut appliquer le résultat déjà prouvé à $g_+(X)$ et $-g_-(X)$ et en déduire que $g(X)$ admet une espérance si et seulement si $\sum_{x\in E} g_+(x)\mu_X(\{x\})$ ou $\sum_{x\in E} g_-(x)\mu_X(\{x\})$ est finie. Or cette dernière condition est équivalente au fait que la somme $\sum_{x\in E} g(x)\mu_X(\{x\})$ est bien définie. On a donc établi l'équivalence entre "$g(X)$ admet une espérance" et "la somme $\sum_{x\in E} g(x)\mu_X(\{x\})$ est bien définie". Supposons à présent que cette somme est bien définie, et prouvons la formule de l'énoncé :

$$\begin{aligned}\mathbb{E}[g(X)] &= \mathbb{E}[g_+(X)] + \mathbb{E}[g_-(X)] \\ &= \sum_{x\in E} g_+(x)\mu_X(\{x\}) + \sum_{x\in E} g_-(x)\mu_X(\{x\}) \\ &= \sum_{x\in E} g(x)\mu_X(\{x\}) \ .\end{aligned}$$

□

La formule de transfert permet de ré-exprimer l'espérance comme une somme sur l'espace de départ Ω :

Corollaire 5.5. *Une v.a. réelle discrète X à valeurs dans un ensemble E admet une espérance si et seulement la somme*

$$\sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\})$$

est bien définie et dans ce cas

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\}) = \sum_{x\in E} x\mu_X(\{x\}) \ .$$

Démonstration. Il suffit d'appliquer la formule de transfert à la variable aléatoire $Y : \Omega \to \Omega$ définie par $Y(\omega) = \omega$, avec la fonction $g : \Omega \to \mathbb{R}$ définie par $g(\omega) = X(\omega)$. □

Proposition 5.6. *L'espérance hérite des propriétés habituelles des séries :*

1. *Multiplication par un scalaire : pour tout $a \in \mathbb{R}$ et toute v.a. X admettant une espérance, $\mathbb{E}[aX] = a\mathbb{E}[X]$.*
2. *Linéarité : $\mathbb{E}[X+Y] = \mathbb{E}[X]+\mathbb{E}[Y]$ pour toutes v.a. X, Y admettant une espérance telles que $X + Y$ admet une espérance (par ex. si au moins l'une des deux est intégrable, ou si les deux sont positives).*
3. *Monotonie : pour toutes v.a. X, Y admettant une espérance, si $X(\omega) \leq Y(\omega)$ pour tout $\omega \in \Omega$ alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$.*
4. *Valeur absolue : $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$ pour toute v.a. X admettant une espérance.*

Les preuves sont laissées en exercice.

Proposition 5.7. *Soit X une variable aléatoire réelle discrète. On suppose que $X(\omega) \geq 0$ pour tout $\omega \in \Omega$. On a alors équivalence entre $\mathbb{E}[X] = 0$ et $\mathbb{P}(X = 0) = 1$.*

Si X est de signe quelconque, on perd l'implication : $\mathbb{E}[X] = 0 \Rightarrow \mathbb{P}(X = 0) = 1$.

Démonstration. Comme X est positive, elle admet une espérance, et l'on a

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x\mathbb{P}(X = x) \, .$$

Si $\mathbb{P}(X = 0) = 1$ alors pour tout $x \neq 0$, on a $\mathbb{P}(X = x) = 0$ et l'on en déduit que

$$\mathbb{E}[X] = 0\mathbb{P}(X = 0) + \sum_{x \in X(\Omega)\setminus\{0\}} x\mathbb{P}(X = x) = 0 \, .$$

Réciproquement, supposons que $\mathbb{P}(X = 0) < 1$. Nécessairement il existe $x > 0$ tel que $\mathbb{P}(X = x) > 0$ et ainsi, par propriété des séries à termes positifs

$$\mathbb{E}[X] \geq x\mathbb{P}(X = x) > 0 \, .$$

On a donc prouvé l'équivalence de l'énoncé. □

Nous énonçons à présent une identité sur l'espérance de variables aléatoires à valeurs dans les entiers positifs, qui sera utile dans certains calculs.

Proposition 5.8. *Soit X une variable aléatoire à valeur dans $\mathbb{N}$, on a*

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k).$$

Démonstration. Tout d'abord, on note que tout entier $n \geq 0$ vérifie l'identité

$$n = \sum_{k=1}^{\infty} \mathbf{1}_{k \leq n} \, .$$

Si X est une variable entière, on a alors

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} n\mathbb{P}(X = n) = \sum_{n=1}^{\infty} \Big(\sum_{k=1}^{\infty} \mathbf{1}_{k \leq n} \Big) \mathbb{P}(X = n) \, .$$

Comme tous les termes sont positifs, on peut échanger les deux sommes pour obtenir

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} \mathbf{1}_{k \leq n} \mathbb{P}(X = n) \, .$$

Or

$$\sum_{n=1}^{\infty} \mathbf{1}_{k \leq n} \mathbb{P}(X = n) = \sum_{n=k}^{\infty} \mathbb{P}(X = n) = \mathbb{P}\big(\cup_{n \geq k} \{X = n\} \big) = \mathbb{P}(X \geq k) \, ,$$

et ainsi

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k) \, .$$

□

5.3 Variance

Nous introduisons à présent un concept important en probabilité : celui de variance d'une variable aléatoire. Avant cela, commençons par observer le fait suivant.

Définition 5.9. *Soit X une variable aléatoire réelle discrète. On dit que X est* de carré intégrable *si X^2 est intégrable, c'est-à-dire si $\mathbb{E}[X^2] < \infty$.*

Comme le lemme ci-dessous le montre, l'intégrabilité du carré implique l'intégrabilité. Attention, la réciproque est fausse ; il y a des variables aléatoires intégrables qui ne sont pas de carré intégrable.

Lemme 5.10. *Soit X une variable aléatoire réelle discrète de carré intégrable. Alors nécessairement X est intégrable.*

Démonstration. Tout d'abord, $1 + X^2$ est intégrable car c'est la somme de deux variables aléatoires intégrables. Par ailleurs, on observe que $|X| \leq 1 + X^2$. Comme $|X|$ est positive, elle admet une espérance et l'on déduit de la propriété de monotonie de la Proposition 5.6 que $\mathbb{E}[|X|] \leq \mathbb{E}[1 + X^2] < \infty$. Ainsi X est intégrable. □

Définition 5.11. *Soit X une variable aléatoire réelle discrète de carré intégrable. On appelle* ***variance de*** *X la quantité*

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

La variance de X est aussi égale à $\mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Démonstration. Pour prouver l'égalité, on utilise le fait que X et X^2 sont intégrables et la linéarité de l'espérance :

$$\begin{aligned}\mathbb{E}[(X - \mathbb{E}[X])^2] &= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] \\ &= \mathbb{E}[X^2] + \mathbb{E}[-2X\mathbb{E}[X]] + \mathbb{E}[\mathbb{E}[X]^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2.\end{aligned}$$

□

La variance est une mesure de l'incertitude ou des fluctuations d'une variable aléatoire : si la variance est petite alors la variable aléatoire a tendance à être proche de son espérance. Si l'on pousse cette idée à l'extrême on obtient :

Corollaire 5.12. *Si X est une variable aléatoire réelle discrète de carré intégrable. Si $\mathrm{Var}(X) = 0$, alors X est constante, c'est-à-dire, $\mathbb{P}(X = \mathbb{E}[X]) = 1$.*

Démonstration. Si $\mathrm{Var}(X) = 0$ alors par la Proposition 5.7, $\mathbb{P}((X - \mathbb{E}[X])^2 = 0) = 1$, ce qui est équivalent à $\mathbb{P}(X = \mathbb{E}[X]) = 1$. □

La variance pose une difficulté d'interprétation à cause du carré dans sa définition. Si une variable aléatoire mesure une quantité physique dans une certaine unité de mesure, par exemple une distance en mètre, alors la variance s'exprime dans le carré de l'unité choisie, ici des m^2.
Pour contourner ce problème, on introduit **l'écart type** :

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}\,,$$

qui s'exprime alors dans la même unité que la variable aléatoire de départ.

5.4 Calculs pour des lois classiques

Dans cette section, on calcule l'espérance et la variance d'une variable aléatoire X qui suit une loi classique.

Bernoulli de paramètre p. $\mathbb{E}[X] = p$ et $\mathrm{Var}(X) = p(1-p)$.
On a

$$\mathbb{E}[X] = 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) = p \, .$$

Concernant la variance, on peut fait le calcul de deux manières. Premièrement on peut commencer par calculer

$$\mathbb{E}[X^2] = 0^2 \cdot \mathbb{P}(X = 0) + 1^2 \cdot \mathbb{P}(X = 1) = p \, .$$

On en déduit alors que

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p - p^2 = p(1-p) \, .$$

L'autre méthode consiste à calculer directement

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] = (0-p)^2 \cdot \mathbb{P}(X = 0) + (1-p)^2 \cdot \mathbb{P}(X = 1) \\ &= p^2(1-p) + (1-p)^2 p = p(1-p) \, . \end{aligned}$$

Uniforme sur $\{1, \ldots, n\}$. $\mathbb{E}[X] = \frac{n+1}{2}$ et $\mathrm{Var}(X) = \frac{n^2-1}{12}$.
On a

$$\mathbb{E}[X] = \sum_{i=i}^{n} i\mathbb{P}(X = i) = \frac{1}{n} \sum_{i=1}^{n} i = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2} \, .$$

Pour la variance, on trouve

$$\mathbb{E}[X^2] = \frac{1}{n} \sum_{i=1}^{n} i^2 = \frac{1}{n} \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6} \, .$$

On en déduit que

$$\mathrm{Var}(X) = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2} \right)^2 = \frac{n^2-1}{12} \, .$$

Binomiale de paramètres n et p. $\mathbb{E}[X] = np$ et $\mathrm{Var}(X) = np(1-p)$.

Concernant l'espérance :

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!}p^k(1-p)^{n-k} \\
&= \sum_{i=0}^{n-1} \frac{n(n-1)!}{i!(n-i-1)!}pp^i(1-p)^{n-1-i} \\
&= np\sum_{i=0}^{n-1} \frac{(n-1)!}{i!(n-i-1)!}p^i(1-p)^{n-1-i} \\
&= np(p+(1-p))^{n-1} \\
&= np.
\end{aligned}$$

Concernant la variance, il est plus simple de commencer par calculer $\mathbb{E}[X(X-1)]$:

$$\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{k=0}^{n} k(k-1)\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \\
&= n(n-1)p^2\sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!}p^{k-2}(1-p)^{n-k} \\
&= n(n-1)p^2\sum_{i=0}^{n-2} \frac{(n-2)!}{i!(n-2-i)!}p^i(1-p)^{n-2-i} \\
&= n(n-1)p^2.
\end{aligned}$$

On en déduit que

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 \\
&= n(n-1)p^2 + np - (np)^2 = np(1-p)\ .
\end{aligned}$$

On remarque que l'espérance et la variance d'une binomiale de paramètres n et p coïncident avec n fois l'espérance et la variance d'une Bernoulli de paramètre p. On rappelle que si $Y_1, \ldots, Y_n$ sont des v.a. indépendantes de loi de Bernoulli de paramètre p, alors $X := Y_1 + \ldots + Y_n$ suit une loi binomiale de paramètres n et p. La linéarité de l'espérance assure alors que $\mathbb{E}[X] = \mathbb{E}[Y_1] + \ldots + \mathbb{E}[Y_n] = n\mathbb{E}[Y_1] = np$. Le lien entre variance de la somme des Bernoulli et variance de la binomiale sera expliqué un peu plus loin.

Géométrique de paramètre p. $\mathbb{E}[X] = \frac{1-p}{p}$ et $\mathrm{Var}(X) = \frac{1-p}{p^2}$.

Espérance : première méthode. On utilise la formule $\mathbb{E}[X] = \sum_{n=1}^{\infty} \mathbb{P}(X \geq n)$. On calcule

$$\begin{aligned}\mathbb{P}(X \geq n) &= \sum_{k=n}^{\infty} \mathbb{P}(X = k) = \sum_{k=n}^{\infty} p(1-p)^k \\ &= p(1-p)^n \sum_{i=0}^{\infty} (1-p)^i \\ &= p(1-p)^n \frac{1}{1-(1-p)} = (1-p)^n.\end{aligned}$$

Et on en déduit

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} (1-p)^n = (1-p)\frac{1}{1-(1-p)} = \frac{1-p}{p}.$$

Pour la suite, on rappelle que pour tout $x \in]-1,1[$

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}, \qquad \sum_{n=1}^{\infty} n x^{n-1} = \frac{1}{(1-x)^2}, \qquad \sum_{n=2}^{\infty} n(n-1)x^{n-2} = \frac{2}{(1-x)^3}. \quad (5.4.1)$$

La première est la série géométrique bien connue, et les deux suivantes se retrouvent en dérivant de part et d'autre, et en ôtant de la somme les termes nuls. Attention, il n'est pas évident que la dérivée et la somme infinie commutent, il s'agit d'un échange de limite, et nous prenons pour acquis qu'une telle commutation est possible (c'est bien le cas ici).

Espérance : deuxième méthode.

$$\begin{aligned}\mathbb{E}[X] &= \sum_{n=0}^{\infty} n\mathbb{P}(X = n) = \sum_{n=1}^{\infty} np(1-p)^n \\ &= p(1-p)\sum_{n=1}^{\infty} n(1-p)^{n-1} \overset{x=1-p}{=} \frac{p(1-p)}{(1-(1-p))^2} = \frac{1-p}{p}.\end{aligned}$$

Variance : Comme pour la binomiale, calculons d'abord

$$\begin{aligned}\mathbb{E}[X(X-1)] &= \sum_{n=0}^{\infty} n(n-1)\mathbb{P}(X = n) = \sum_{n=2}^{\infty} n(n-1)p(1-p)^n \\ &= p(1-p)^2 \sum_{n=2}^{\infty} n(n-1)(1-p)^{n-2} \overset{x=1-p}{=} p(1-p)^2 \frac{2}{(1-(1-p))^3} \\ &= \frac{2(1-p)^2}{p^2}\end{aligned}$$

puis

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 \\ &= \frac{2(1-p)^2}{p^2} + \frac{1-p}{p} - \left(\frac{1-p}{p}\right)^2 = \frac{1-p}{p^2}\,.\end{aligned}$$

Loi de Poisson de paramètre λ. $\mathbb{E}[X] = \lambda$ et $\mathrm{Var}(X) = \lambda$.
On rappelle que pour tout réel x

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Concernant l'espérance, on a

$$\begin{aligned}\mathbb{E}[X] &= \sum_{n=0}^{\infty} n e^{-\lambda} \frac{\lambda^n}{n!} = \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} \\ &= \lambda e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{(i)!} = \lambda \ .\end{aligned}$$

Concernant la variance, on peut utiliser la même astuce que précédemment et calculer $\mathbb{E}[X(X-1)]$:

$$\begin{aligned}\mathbb{E}[X(X-1)] &= \sum_{n=0}^{\infty} n(n-1) e^{-\lambda} \frac{\lambda^n}{n!} \\ &= \lambda^2 \sum_{n=2}^{\infty} e^{-\lambda} \frac{\lambda^{n-2}}{(n-2)!} \\ &= \lambda^2 \sum_{i=0}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!} = \lambda^2.\end{aligned}$$

On en déduit que

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

5.5 Espérance et indépendance

Dans cette partie, nous examinons les liens entre l'indépendance de variables aléatoires et l'espérance de certaines quantités construites à partir de ces variables.

Tout d'abord nous énonçons une conséquence de l'indépendance en termes d'espérance.

Proposition 5.13. *Soit $N \geq 1$ un entier et $(X_k)_{k=1,\dots,N}$ une suite de variables aléatoires indépendantes. On suppose que chaque variable X_k est à valeurs dans un ensemble fini ou infini dénombrable E_k. Pour tout entier $1 \leq n \leq N$, pour toutes fonctions $f_1, \dots, f_n$ de $E_1, \dots, E_n$ dans $\mathbb{R}$ telles que $f_k(X_k)$ est intégrable pour tout $k \leq n$, on a*

$$\mathbb{E}\big[\prod_{k=1}^{n} f_k(X_k)\big] = \prod_{k=1}^{n} \mathbb{E}\big[f_k(X_k)\big] \ .$$

En particulier si X et Y sont deux variables intégrables indépendantes,

$$\mathbb{E}\big[XY\big] = \mathbb{E}\big[X\big]\mathbb{E}\big[Y\big] \ .$$

La condition d'intégrabilité des $f_k(X_k)$ peut être remplacée par une condition de positivité, sans changer le résultat.

Démonstration. On commence par le cas de deux variables. Soient X et Y des variables aléatoires indépendantes à valeurs dans E et F respectivement. Soient $f : E \to \mathbb{R}$ et $g : F \to \mathbb{R}$ des fonctions que l'on suppose, pour l'instant, positives. En considérant (X, Y) comme une variable aléatoire à valeurs dans l'ensemble dénombrable $E \times F$, on obtient

$$\begin{aligned}
\mathbb{E}\big[f(X)g(Y)\big] &= \sum_{(x,y)\in E\times F} f(x)g(y)\mathbb{P}(\{X=x\}\cap\{Y=y\}) \\
&= \sum_{(x,y)\in E\times F} f(x)g(y)\mathbb{P}(X=x)\mathbb{P}(Y=y) \\
&= \sum_{x\in E}\sum_{y\in F} f(x)g(y)\mathbb{P}(X=x)\mathbb{P}(Y=y) \\
&= \left(\sum_{x\in E} f(x)\mathbb{P}(X=x)\right)\left(\sum_{y\in F} g(y)\mathbb{P}(Y=y)\right) \\
&= \mathbb{E}[f(X)]\mathbb{E}[g(Y)] \,.
\end{aligned}$$

Les opérations qu'on a effectuées sur les sommes sont justifiées par le fait que tous les termes sont positifs.

Maintenant dans le cas de deux fonctions telles que $f(X)$ et $g(Y)$ sont intégrables, le calcul précédent montre que $\mathbb{E}[|f(X)g(Y)|] = \mathbb{E}[|f(X)|]\mathbb{E}[|g(Y)|] < \infty$ et ainsi $f(X)g(Y)$ est intégrable. En décomposant f et g en leurs parties positives et négatives $f = f_+ - f_-$ et $g = g_+ - g_-$, on obtient en utilisant l'identité déjà prouvée dans le cas positif

$$\begin{aligned}
&\mathbb{E}\big[f(X)g(Y)\big] = \mathbb{E}\big[(f_+(X) - f_-(X))(g_+(Y) - g_-(Y))\big] \\
&\quad = \mathbb{E}\big[f_+(X)g_+(Y)\big] - \mathbb{E}\big[f_-(X)g_+(Y)\big] - \mathbb{E}\big[f_+(X)g_-(Y)\big] + \mathbb{E}\big[f_-(X)g_-(Y)\big] \\
&\quad = \mathbb{E}[f_+(X)]\mathbb{E}[g_+(Y)] - \mathbb{E}[f_-(X)]\mathbb{E}[g_+(Y)] - \mathbb{E}[f_+(X)]\mathbb{E}[g_-(Y)] + \mathbb{E}[f_-(X)]\mathbb{E}[g_-(Y)] \\
&\quad = \mathbb{E}[f(X)]\mathbb{E}[g_+(Y)] - \mathbb{E}[f(X)]\mathbb{E}[g_-(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)] \,.
\end{aligned}$$

Le cas d'une famille finie de variables aléatoires se déduit du cas de deux variables par récurrence. □

Proposition 5.14. *(Caractérisation de l'indépendance) Soit $N \geq 1$ un entier et $(X_k)_{k=1,\dots,N}$ une suite de variables aléatoires indépendantes. On suppose que chaque variable X_k est à valeurs dans un ensemble fini ou infini dénombrable E_k. Ces variables sont indépendantes si et seulement si pour tout entier $1 \leq n \leq N$, pour toutes fonctions $f_1, \dots, f_n$* ***bornées*** *de $E_1, \dots, E_n$ dans $\mathbb{R}$*

$$\mathbb{E}\big[\prod_{k=1}^{n} f_k(X_k)\big] = \prod_{k=1}^{n} \mathbb{E}\big[f_k(X_k)\big] \,.$$

Il est très important de distinguer les deux propositions précédentes. La première proposition tire des conséquences de l'indépendance et énonce un résultat pour une

"très grande classe" de fonctions f_k. La second proposition établit une caractérisation de l'indépendance et met en jeu une "petite classe" de fonctions f_k.

La deuxième proposition fournit également une nouvelle définition de l'indépendance de variables aléatoires.

Démonstration. On a déjà montré le sens direct dans la proposition précédente. On ne montre donc que la réciproque. Soit un entier $1 \leq n \leq N$. Si l'on se donne $x_1, \ldots, x_n$ dans $E_1, \ldots, E_n$ et qu'on pose $f_k(x) = \mathbf{1}_{\{x=x_k\}}$, on obtient

$$\begin{aligned}\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) &= \mathbb{E}[\mathbf{1}_{\{X_1=x_1,\ldots,X_n=x_n\}}] \\ &= \mathbb{E}[\prod_{k=1}^{n} f_k(X_k)] = \prod_{k=1}^{n} \mathbb{E}[f_k(X_k)] = \prod_{k=1}^{n} \mathbb{P}(X_k = x_k).\end{aligned}$$

On retrouve ainsi directement la définition de l'indépendance. □

Chapitre 6

Vecteurs aléatoires

6.1 Covariance

Avant d'introduire la notion de covariance, commençons par un résultat préliminaire.

Lemme 6.1. *Soient X et Y deux v.a. réelles discrètes de carrés intégrables. La variable aléatoire XY est alors intégrable.*

Démonstration. Cela découle de l'inégalité : $|XY| \leq X^2+Y^2$, de l'intégrabilité de X^2+Y^2 et de la monotonie de l'espérance énoncée dans la Proposition 5.6. □

Définition 6.2. *Soient X et Y deux v.a. réelles discrètes de carrés intégrables. La covariance de X et Y est définie par*

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

Elle est aussi égale à $\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]$.

Démonstration. On a

$$\begin{aligned}\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] &= \mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\ .\end{aligned}$$

□

La covariance souffre d'un défaut d'inhomogénéité similaire à celui de la variance : si X s'exprime en mètres et Y en inverses de secondes, la covariance s'exprime en mètres par secondes. Pour corriger ce défaut, on introduit (pour des variables X et Y de carrés intégrables) le **coefficient de corrélation**

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}. \tag{6.1.1}$$

On verra plus tard, au Corollaire 7.4 qu'on a toujours $-1 \leq \rho(X,Y) \leq 1$.

Exemple 6.2.1. On tire un dé standard avec des faces numérotées de 1 à 6. On pose $X = 1$ si le résultat est pair et 0 sinon, $Y = 1$ si on a tiré un 6 et 0 sinon et $Z = 1$ si on a tiré un 1 et 0 sinon. Ces trois variables sont toutes différentes. Intuitivement, X et Y ont tendance à valoir 1 en même temps (positivement corrélées) alors que Z a tendance à valoir 1 lorsque les autres sont égales à 0 (négativement corrélées). Un calcul simple donne

$$\mathbb{E}[X] = \frac{1}{2}, \quad \mathbb{E}[Y] = \frac{1}{6}, \quad \mathbb{E}[Z] = \frac{1}{6}.$$

D'un autre coté, on remarque que $XY = Y$ puisque si $Y = 1$ c'est que le résultat est 6 et donc qu'il est pair. On remarque aussi que $XZ = 0$ et $XY = 0$ puisque les variables ne peuvent pas valoir 1 en même temps. On a donc

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] = \frac{1}{12}\ ,$$

ainsi que

$$\mathrm{Cov}(X, Z) = -\frac{1}{12}, \quad \mathrm{Cov}(Y, Z) = -\frac{1}{36}\ .$$

Exemple 6.2.2. On prend X et Y indépendantes uniformes sur $\{1, 2, 3\}$ et on pose $Z = X - Y$. On a

$$\mathbb{E}[X] = \mathbb{E}[Y] = 2, \quad \mathbb{E}[Z] = \mathbb{E}[X] - \mathbb{E}[Y] = 0$$

et

$$\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 1 \cdot \frac{1}{3} + 4 \cdot \frac{1}{3} + 9 \cdot \frac{1}{3} = \frac{14}{3}\ .$$

On en déduit que

$$\begin{aligned}
\mathrm{Cov}(X, Z) &= \mathbb{E}[XZ] - \mathbb{E}[X]\mathbb{E}[Z] \\
&= \mathbb{E}[X^2 - XY] - 2 \times 0 \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]\mathbb{E}[Y] = \frac{2}{3} \\
\mathrm{Cov}(Y, Z) &= \mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z] \\
&= \mathbb{E}[XY - Y^2] - 2 \times 0 \\
&= \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y^2] = -\frac{2}{3}
\end{aligned}$$

La variance et de la covariance vérifient des identités remarquables :

Proposition 6.3. *Soient X, X', Y, Y' des variables aléatoires réelles de carrés intégrables et a, b des réels. On a*

1. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$,
2. $\mathrm{Cov}(X + a, Y + b) = \mathrm{Cov}(X, Y)$,
3. $\mathrm{Cov}(aX, bY) = ab\,\mathrm{Cov}(X, Y)$,

4. $\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X)$,
5. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$,
6. $\mathrm{Cov}(X+X', Y+Y') = \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Y') + \mathrm{Cov}(X',Y) + \mathrm{Cov}(X',Y')$, $\mathrm{Var}(X+X') = \mathrm{Var}(X) + 2\,\mathrm{Cov}(X,X') + \mathrm{Var}(X')$.

Démonstration. Laissé en exercice, c'est une application directe des définitions et de la linéarité de l'espérance. □

Examinons à présent le lien entre indépendance et covariance.

Proposition 6.4. *Soit X et Y des variables aléatoires de carrés intégrables. Si X et Y sont indépendantes alors* $\mathrm{Cov}(X,Y) = 0$.

Attention, la réciproque est fausse !

Démonstration. On écrit

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

Or l'indépendance de X et Y combinées à l'intégrabilité des variables aléatoires X, Y et XY permet d'appliquer la Proposition 5.13 et de déduire que $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ ce qui permet de conclure. □

Corollaire 6.5. *Soient $X_1, \ldots, X_n$ des v.a. réelles discrètes de carrés intégrables. On a*

$$\mathrm{Var}\Big(\sum_{i=1}^n X_i\Big) = \sum_{i=1}^n \mathrm{Var}(X_i) + \sum_{i\neq j} \mathrm{Cov}(X_i, X_j) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j).$$

Si de plus les variables sont indépendantes alors

$$\mathrm{Var}\Big(\sum_{i=1}^n X_i\Big) = \sum_{i=1}^n \mathrm{Var}(X_i).$$

Démonstration. La première expression s'obtient en utilisant la définition de la variance sous forme de l'espérance d'un carré. La deuxième s'en déduit alors en utilisant le fait que la covariance entre deux variables indépendantes est nulle. □

Remarque 6.5.1. Revenons sur la valeur de la variance d'une loi binomiale. Soient $Y_1, \ldots, Y_n$ des v.a. indépendantes de loi de Bernoulli de paramètre p, alors $X := Y_1 + \ldots + Y_n$ suit une loi binomiale de paramètres n et p. En appliquant le corollaire, on en déduit que $\mathrm{Var}(X) = \mathrm{Var}(Y_1) + \ldots + \mathrm{Var}(Y_n) = n\,\mathrm{Var}(Y_1) = np(1-p)$.

6.2 Loi jointe, loi marginale

Soient $X_1, \ldots, X_d$ des variables aléatoires à valeurs dans des espaces (finis ou infinis dénombrables) $E_1, \ldots, E_d$. Notons que le vecteur $(X_1, \ldots, X_d)$ peut être vu comme une variable aléatoire à valeurs dans l'espace produit $E_1 \times \ldots \times E_d$.

Définition 6.6. *Soit $X_1, \ldots, X_d$ des variables aléatoires, on appelle **loi jointe** des variables $X_1, \ldots, X_d$ la loi du vecteur $(X_1, \ldots, X_d)$.*

Définition 6.7. *Soit $(X_1, \ldots, X_d)$ un vecteur aléatoire, on appelle **loi marginale** de X_i (ou loi marginale de la coordonnée i du vecteur) la loi de la variable aléatoire X_i.*

La loi marginale de X_i se calcule à partir de la loi jointe de la façon suivante :

$$\mathbb{P}(X_i = x_i) = \sum_{x_1 \in E_1} \cdots \sum_{x_{i-1} \in E_{i-1}} \sum_{x_{i+1} \in E_{i+1}} \cdots \sum_{x_n \in E_n} \mathbb{P}((X_1, \ldots, X_d) = (x_1, \ldots, x_n)) \,. \tag{6.2.1}$$

On notera que la loi jointe d'un vecteur contient **beaucoup** plus d'information que les lois marginales. En particulier, la connaissance de la loi jointe suffit à retrouver les lois marginales, alors que les lois marginales ne suffisent pas en général à déterminer la loi jointe.

Pour illustrer ce dernier point, considérons le cas d'un vecteur (X, Y) pour lequel X et Y sont uniformes sur $\{1, \ldots, 6\}$. Il existe de nombreuses lois jointes possibles pour (X, Y). Par exemple, il se pourrait que $X = Y$ auquel cas la loi jointe donne une probabilité nulle à toutes les paires (x, y) qui sont telles que $x \neq y$. Il se pourrait aussi que les variables X et Y soient indépendantes auquel cas la loi jointe est en fait la mesure uniforme sur toutes les paires $(x, y) \in \{1, \ldots, 6\}^2$.

6.3 Indépendance et loi conditionnelle

La loi jointe de deux variables X et Y contient toute l'information sur les variables et sur leurs corrélations : en particulier, on peut y lire si les deux variables sont indépendantes.

Proposition 6.8. *Soient X et Y deux variables aléatoires à valeurs dans des ensembles (finis ou infinis dénombrables) E et F. Ces variables sont indépendantes si et seulement si on peut trouver deux fonctions positives f et g et une constante $c > 0$ telles que pour tout $x \in E$ et pour tout $y \in F$,*

$$\mathbb{P}(X = x, Y = y) = c f(x) g(y).$$

Dans ce cas on a de plus

$$\mathbb{P}(X = x) = \frac{f(x)}{\sum_{x' \in E} f(x')}, \quad \mathbb{P}(Y = y) = \frac{g(y)}{\sum_{y' \in F} g(y')}.$$

Démonstration. Si X et Y sont indépendantes, alors on a $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ donc on peut prendre $f(x) = \mathbb{P}(X = x)$, $g(y) = \mathbb{P}(Y = y)$ et $c = 1$ dans l'énoncé. Le deuxième énoncé est alors immédiatement vrai car les sommes au dénominateur valent 1.
Pour la réciproque, d'après la définition de l'indépendance et l'expression (6.2.1) des lois marginales, on doit montrer que pour tout x et y

$$\mathbb{P}(X = x, Y = y) = \left(\sum_{y' \in F} \mathbb{P}(X = x, Y = y')\right)\left(\sum_{x' \in E} \mathbb{P}(X = x', Y = y)\right).$$

En réécrivant ces expressions en termes de f, g et c, cela signifie qu'on doit montrer que

$$cf(x)g(y) = \left(\sum_{y'} cf(x)g(y')\right)\left(\sum_{x'} cf(x')g(y)\right). \tag{6.3.1}$$

Pour cela, on part de l'égalité $\sum_{x'} \mathbb{P}(X = x') = 1$. En la réécrivant en terme de f et de g, elle devient

$$\sum_{x'}\left(\sum_{y'} cf(x')g(y')\right) = 1,$$

soit en réordonnant les termes

$$c\left(\sum_{x'} f(x')\right)\left(\sum_{y'} g(y')\right) = 1$$

En multipliant de part et d'autre par $cf(x)g(y)$ on a alors

$$cf(x)g(y) = c^2 f(x)g(y)\left(\sum_{y'} g(y')\right)\left(\sum_{x'} f(x')\right),$$

et en réorganisant à nouveau les termes on obtient (6.3.1).

Pour le dernier résultat on repart de l'égalité

$$c\left(\sum_{x'} f(x')\right)\left(\sum_{y'} g(y')\right) = 1.$$

En divisant par $\sum_{x'} f(x')$ et en multipliant par $f(x)$, elle devient

$$\sum_{y'} cf(x)g(y') = \frac{f(x)}{\sum_{x'} f(x')}$$

et dans le terme de gauche on reconnait $\mathbb{P}(X = x)$. Même chose pour Y. □

La proposition reste vraie quand on a plus de deux variables.

Proposition 6.9. *Soit $X_1, \dots, X_d$ des variables aléatoires à valeurs dans des ensembles (finis ou infinis dénombrables) $E_1, \dots, E_d$. Ces variables sont indépendantes si et seulement s'il existe des fonctions positives $f_1, \dots, f_d$ et une constante $c > 0$ telles que pour tout $x_i \in E_i$*

$$\mathbb{P}(X_1 = x_1, \dots, X_d = x_d) = c \prod_{i=1}^{d} f_i(x_i).$$

Dans ce cas on a pour tout i et pour tout $x \in E_i$

$$\mathbb{P}(X_i = x) = \frac{f_i(x)}{\sum_{x' \in E_i} f_i(x')}.$$

La preuve est similaire au cas de deux variables.

Remarque 6.9.1. Comme dans le cas de variables indépendantes, on peut "lire" facilement dans une loi jointe des lois conditionnelles à constante multiplicative près.

Plus précisément, si on a deux variables X et Y, alors pour tout y et pour tout x tel que $\mathbb{P}(X = x) > 0$,

$$\mathbb{P}(Y = y | X = x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)} = \frac{\mathbb{P}(X = x, Y = y)}{\sum_{y'} \mathbb{P}(X = x, Y = y')}.$$

On remarque que pour un x fixé, les valeurs $\mathbb{P}(X = x, Y = y)$ pour les différents y donnent la loi conditionnelle de Y à une constante près.

6.4 Tableau de loi jointe

Il est souvent pratique de représenter la loi jointe de deux variables à valeurs dans des espaces **finis** par un tableau à double entrée. Nous allons illustrer cela sur des exemples.

Exemple 6.9.2. On considère X et Y de loi jointe donnée par le tableau suivant :

$X \setminus Y$	0	1	2	3
0	$\frac{0}{48}$	$\frac{1}{48}$	$\frac{2}{48}$	$\frac{3}{48}$
1	$\frac{1}{48}$	$\frac{2}{48}$	$\frac{3}{48}$	$\frac{4}{48}$
2	$\frac{2}{48}$	$\frac{3}{48}$	$\frac{4}{48}$	$\frac{5}{48}$
3	$\frac{3}{48}$	$\frac{4}{48}$	$\frac{5}{48}$	$\frac{6}{48}$

L'entrée (i, j) du tableau donne la valeur de $\mathbb{P}(X = i, Y = j)$. Ici on peut se convaincre que $\mathbb{P}(X = i, Y = j) = \frac{i+j}{48}$ si i et j sont dans $\{0, 1, 2, 3\}$.
En général, pour qu'un tableau définisse bien une loi jointe, il suffit de vérifier que la somme de toutes les entrées vaut 1 et que toutes les entrées sont positives ou nulles.

On peut obtenir les lois marginales de X et Y de façon très simple. Il suffit d'ajouter une colonne et une ligne, et d'y inscrire la somme des valeurs du tableau dans la ligne et la colonne correspondantes :

$X \setminus Y$	0	1	2	3	
0	$\frac{0}{48}$	$\frac{1}{48}$	$\frac{2}{48}$	$\frac{3}{48}$	$\frac{6}{48}$
1	$\frac{1}{48}$	$\frac{2}{48}$	$\frac{3}{48}$	$\frac{4}{48}$	$\frac{10}{48}$
2	$\frac{2}{48}$	$\frac{3}{48}$	$\frac{4}{48}$	$\frac{5}{48}$	$\frac{14}{48}$
3	$\frac{3}{48}$	$\frac{4}{48}$	$\frac{5}{48}$	$\frac{6}{48}$	$\frac{18}{48}$
	$\frac{6}{48}$	$\frac{10}{48}$	$\frac{14}{48}$	$\frac{18}{48}$	$\frac{48}{48}$

Dans cet exemple, le tableau est symétrique entre X et Y ce qui signifie que (X, Y) et (Y, X) ont la même loi.

Exemple 6.9.3. On considère X et Y de loi jointe donnée par le tableau suivant, où l'on a directement inclus les sommes.

$X \setminus Y$	1	2	3	
1	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{6}{36}$
2	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{6}{36}$	$\frac{12}{36}$
3	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{9}{36}$	$\frac{18}{36}$
	$\frac{6}{36}$	$\frac{12}{36}$	$\frac{18}{36}$	$\frac{36}{36}$

On peut se convaincre que $\mathbb{P}(X = i, Y = y) = \frac{xy}{36}$ pour x et y dans $\{1, 2, 4\}$. Ainsi X et Y sont indépendantes.

Exemple 6.9.4. On considère X et Y de loi jointe donnée par

$X \setminus Y$	0	1	4	
-2	0	0	$\frac{1}{5}$	$\frac{1}{5}$
-1	0	$\frac{1}{5}$	0	$\frac{1}{5}$
0	$\frac{1}{5}$	0	0	$\frac{1}{5}$
1	0	$\frac{1}{5}$	0	$\frac{1}{5}$
2	0	0	$\frac{1}{5}$	$\frac{1}{5}$
	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{5}{5}$

Dans cet exemple, on remarque que chaque ligne ne contient qu'un seul terme non nul. Or on a vu dans la deuxième partie que les termes dans la ligne $X = x$ donnent (à une constante près) la loi conditionnelle de Y sachant $X = x$. Si dans cette loi toutes les probabilités sont nulles sauf une, c'est que sachant $X = x$, Y n'est plus aléatoire du tout. Comme c'est le cas pour toutes les lignes, cela signifie que Y est une fonction de X. Ici, on peut constater que $Y = X^2$.

À l'inverse, il y a plusieurs termes non nuls dans certaines colonnes car si $Y = X^2 = 4$, alors $X = 2$ ou $X = -2$.

Chapitre 7

Quelques outils

7.1 Inégalités sur l'espérance

7.1.1 Inégalité de Markov

Proposition 7.1 (Inégalité de Markov)**.** *Soit X une variable aléatoire réelle positive discrète. Pour tout $x > 0$, on a*

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}[X]}{x}.$$

Démonstration. On fixe $x > 0$. On pose $Y = x\mathbf{1}_{\{X \geq x\}}$, c'est-à-dire $Y = 0$ quand $X < x$ et $Y = x$ quand $X \geq x$. Puisque que X est positive ou nulle, on en déduit que $Y \leq X$ et par conséquent

$$\mathbb{E}[X] \geq \mathbb{E}[Y] = x\mathbb{P}(X \geq x) \ .$$

□

Remarque 7.1.1. L'inégalité ne dit rien d'intéressant si $x \leq \mathbb{E}[X]$ puisque dans ce cas la borne est supérieure à 1.

7.1.2 Inégalité de Bienaymé-Tchebychev

Proposition 7.2 (Inégalité de Bienaymé-Tchebychev)**.** *Soit X une variable aléatoire réelle discrète de carré intégrable. Pour tout x réel positif*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq x) \leq \frac{\mathrm{Var}(X)}{x^2},$$

ou en d'autres termes, pour tout $c > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq c\sqrt{\mathrm{Var}(X)}) \leq \frac{1}{c^2}.$$

Démonstration. On pose $Y = (X - \mathbb{E}[X])^2$. Ainsi $\mathbb{E}[Y] = \mathrm{Var}(X)$ et $\mathbb{P}(|X - \mathbb{E}[X]| \geq x) = \mathbb{P}(Y \geq x^2)$. Il suffit alors d'appliquer l'inégalité de Markov à cette dernière quantité pour obtenir

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq x) \leq \frac{\mathrm{Var}(X)}{x^2} .$$

La deuxième expression de l'énoncé est une conséquence immédiate de la première. □

Cette inégalité est une nouvelle justification de l'interprétation de l'écart type comme une mesure de l'ordre de grandeur des déviations entre une variable et sa moyenne : la probabilité d'observer une déviation de 10 fois l'écart type ne sera jamais plus grande que 1/100.

Exemple 7.2.1. Une application fondamentale de l'inégalité de Bienaymé-Tchebychev est le contrôle des intervalles de confiance en statistiques. Considérons un sondage lors d'un référendum. On suppose qu'il y a une proportion p de gens qui veulent voter "oui" et $(1 - p)$ qui veulent voter "non". On interroge n personnes choisies uniformément au hasard. Si l'on note S_n le nombre de réponses "oui" obtenues, il est naturel de supposer que S_n suit une loi binomiale de paramètres n et p. On veut alors estimer la différence entre le résultat du sondage S_n/n et la proportion p.
On a

$$\mathbb{E}[S_n/n] = p , \quad \mathrm{Var}(S_n/n) = \frac{1}{n^2}\mathrm{Var}(S_n) = \frac{p(1-p)}{n} .$$

Ainsi par l'inégalité de Bienaymé-Tchebychev on trouve

$$\mathbb{P}\Big(\Big|\frac{S_n}{n} - p\Big| \geq x\Big) \leq \frac{p(1-p)}{nx^2}.$$

On voit donc qu'en sondant un nombre n suffisamment grand de personnes, le résultat du sondage sera proche de la proportion effective avec grande probabilité.

7.1.3 Inégalité de Cauchy-Schwarz

Proposition 7.3 (Inégalité de Cauchy-Schwarz)**.** *Soient X et Y des variables aléatoires réelles discrètes de carrés intégrables. La variable XY est intégrable et vérifie l'inégalité*

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

On a égalité dans l'expression ci-dessus si et seulement si il existe $c \in \mathbb{R}$ tel que $\mathbb{P}(X = cY) = 1$.

Démonstration. L'intégrabilité a déjà été prouvée au Lemme 6.1. On remarque à présent que pour tout $\lambda \in \mathbb{R}$

$$0 \leq \mathbb{E}[(\lambda X + Y)^2] = \lambda^2\mathbb{E}[X^2] + 2\lambda\mathbb{E}[XY] + \mathbb{E}[Y^2] .$$

Il s'agit là d'un trinôme en λ, qui est toujours positif. Cela implique que son discriminant doit être négatif, c'est-à-dire, que

$$\Delta = (2\mathbb{E}[XY])^2 - 4\mathbb{E}[X^2]\mathbb{E}[Y^2] \leq 0 .$$

En prenant la racine carrée et en divisant par 4 on obtient

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

Le cas d'égalité survient exactement quand le discriminant est nul ce qui est équivalent à l'existence d'un $\lambda \in \mathbb{R}$ tel que $\mathbb{E}[(\lambda X + Y)^2] = 0$. On conclut par la positivité de l'espérance. □

Corollaire 7.4. *Soient X, Y de carrés intégrables. Le coefficient de corrélation défini en (6.1.1) satisfait $-1 \leq \rho(X,Y) \leq 1$.*

Démonstration. On a vu que $\mathrm{Cov}(X,Y) = \mathbb{E}[X'Y']$ avec $X' = X - \mathbb{E}[X]$ et $Y' = Y - \mathbb{E}[Y]$. En appliquant Cauchy-Schwartz au couple (X', Y'), on trouve

$$|\,\mathrm{Cov}(X,Y)| \leq \mathbb{E}[X'^2]\mathbb{E}[Y'^2] = \mathrm{Var}(X)\,\mathrm{Var}(Y).$$

Le résultat suit en prenant la racine. □

7.1.4 Inégalité de Jensen

Proposition 7.5 (Inégalité de Jensen)**.** *Soit X une variable aléatoire réelle discrète et $f : \mathbb{R} \to \mathbb{R}_+$ une fonction convexe. Si X est intégrable alors*

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]\,.$$

On rappelle qu'une fonction réelle $f : \mathbb{R} \to \mathbb{R}$ est convexe si pour tous $x, y \in \mathbb{R}$ et pour tout $p \in]0,1[$, on a

$$f(px + (1-p)y) \leq pf(x) + (1-p)f(y)\,.$$

On rappelle également que si f est convexe alors pour tout point $x_0 \in \mathbb{R}$ il existe $a, b \in \mathbb{R}$ tels que

$$f(x) \geq ax + b \forall x \in \mathbb{R}\,, \quad f(x_0) = ax_0 + b\,.$$

Preuve de l'inégalité de Jensen. On pose $x_0 := \mathbb{E}[X]$. La propriété rappelée ci-dessus assure qu'il existe $a, b \in \mathbb{R}$ tels que

$$f(X) \geq aX + b\,, \quad f(\mathbb{E}[X]) = a\mathbb{E}[X] + b\,.$$

Comme $f(X) \geq 0$, cette variable aléatoire admet une espérance et l'on obtient par monotonie de l'espérance

$$\mathbb{E}[f(X)] \geq \mathbb{E}[aX + b] = a\mathbb{E}[X] + b = f(\mathbb{E}[X])\,.$$

□

7.2 Caractérisation d'une loi

On rappelle qu'une loi de probabilité μ sur un ensemble (fini ou infini dénombrable) E est caractérisée par la donnée de $\mu(\{a\})$ pour tout $a \in E$. Dans cette partie, nous allons nous intéresser à des caractérisations alternatives.

7.2.1 Fonction de répartition

Définition 7.6. *Soit X une variable aléatoire réelle discrète. La fonction de répartition de la variable X est la fonction*

$$F_X : \begin{cases} \mathbb{R} \to \mathbb{R} \\ x \mapsto \mathbb{P}(X \leq x). \end{cases}$$

Il est assez facile de vérifier qu'une telle fonction est positive, croissante et tend vers 0 en $-\infty$ et vers 1 en $+\infty$.

Proposition 7.7. *Soit X une variable aléatoire réelle discrète. Pour tout $a \in \mathbb{R}$*

$$\mathbb{P}(X = a) = F(a) - F(a-) \ ,$$

où $F(a-) := \lim_{x \uparrow a} F(x)$ est la limite à gauche de F en a.
En conséquence, la fonction de répartition caractérise la loi. Plus précisément si X et Y sont des variables aléatoires réelles discrètes, alors X et Y ont même loi si et seulement si $F_X = F_Y$.

Démonstration. On constate que pour tout $a \in \mathbb{R}$, la suite d'événements $\{a - \frac{1}{n} < X \leq a\}$, $n \geq 1$ est décroissante pour l'inclusion et l'intersection de tous ces événements coïncide avec l'événement $\{X = a\}$. Ainsi, par la Proposition 2.11,

$$\mathbb{P}(X = a) = \lim_{n \to \infty} \mathbb{P}\left(a - \frac{1}{n} < X \leq a\right) \ .$$

Or pour tout $n \geq 1$,

$$\begin{aligned} F(a) - F\left(a - \frac{1}{n}\right) &= \mathbb{P}\left(X \leq a\right) - \mathbb{P}\left(X \leq a - \frac{1}{n}\right) = \mathbb{P}\Big(\{X \leq a\} \backslash \Big\{X \leq a - \frac{1}{n}\Big\}\Big) \\ &= \mathbb{P}\left(a - \frac{1}{n} < X \leq a\right) \ , \end{aligned}$$

et l'on peut conclure.
Passons à la preuve de la deuxième partie de l'énoncé. Si X et Y ont même loi alors $\mathbb{P}(X \leq a) = \mathbb{P}(Y \leq a)$ pour tout $a \in \mathbb{R}$, et l'on en déduit que $F_X = F_Y$. Réciproquement si $F_X = F_Y$ alors par l'identité que l'on vient de prouver, on déduit que $\mathbb{P}(X = a) = \mathbb{P}(Y = a)$ pour tout $a \in \mathbb{R}$ et cela suffit à conclure. □

Exemple 7.7.1. $X \sim B(p)$, alors

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1-p & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1. \end{cases}$$

Exemple 7.7.2. $X \sim U(\{1, \ldots, n\})$, alors

$$F_X(x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{\lfloor x \rfloor}{n} & \text{si } 1 \leq x < n \\ 1 & \text{si } x \geq n. \end{cases}$$

où on rappelle que $\lfloor x \rfloor$ est la partie entière de x.

7.2.2 Fonction génératrice

On introduit à présent une caractérisation de la loi qui est spécifique aux variables aléatoires à valeurs dans $\mathbb{N}$.

Définition 7.8. *Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. La fonction génératrice de X est la fonction :*

$$G_X : \begin{cases} [-1,1] & \to \mathbb{R} \\ z & \mapsto \sum_{n=0}^{\infty} z^n \mathbb{P}(X = n) \end{cases}$$

On notera que $G_X(z) = \mathbb{E}[z^X]$ pour tout $z \in [-1, 1]$.

Remarque 7.8.1. Du fait que $\sum_n \mathbb{P}(X = n) = 1$, la série apparaissant dans la fonction génératrice est absolument convergente sur $[-1, 1]$.

Proposition 7.9. *La fonction génératrice caractérise la loi. Plus précisément, soient X et Y deux variables aléatoires à valeurs dans $\mathbb{N}$. X et Y ont même loi si et seulement si leurs fonctions génératrices coïncident.*

Démonstration. Le sens direct de l'équivalence est simple : si X et Y ont même loi, alors $\mathbb{P}(X = n) = \mathbb{P}(Y = n)$ pour tout $n \in \mathbb{N}$ et ainsi $G_X = G_Y$.
On passe à l'implication réciproque. La théorie générale des séries entières assure qu'à l'intérieur du disque de convergence, on peut dériver sous le symbole de sommation. En particulier on voit que pour tout $n \in \mathbb{N}$,

$$G_X^{(n)}(0) = n! \mathbb{P}(X = n) \, .$$

Ainsi l'égalité $G_X = G_Y$ assure l'égalité des lois. □

Un des intérêts de la fonction génératrice est qu'elle permet de faire de nombreux calculs très explicitement.

Proposition 7.10. *Pour toute variable aléatoire X, G_X prend ses valeurs dans $[-1,1]$, vérifie $G_X(1)=1$ et est une fonction convexe sur $[0,1]$.*

Démonstration. Pour le premier point, pour tout $s\in[-1,1]$, $|\mathbb{E}[s^X]|\leq\mathbb{E}[|s^X|]\leq\mathbb{E}[1]=1$ car $|s^X|\leq 1$ avec probabilité 1. Pour le second point, on calcule $G_X(1)=\mathbb{E}[1^X]=\mathbb{E}[1]=1$. Avant de prouver la troisième propriété, on commence par noter que pour tout $n\in\mathbb{N}$, la fonction puissance $y\mapsto y^n$ est convexe sur $[0,\infty[$. Ainsi pour tous $a,b\in[0,1]$ et $p\in[0,1]$, $pa^n+(1-p)b^n\leq(pa+(1-p)b)^n$. On en déduit alors que

$$\begin{aligned} pG_X(a)+(1-p)G_X(b) &= \mathbb{E}[pa^X+(1-p)b^X]\leq\mathbb{E}[(pa+(1-p)b)^X] \\ &= G_X(pa+(1-p)b)\ , \end{aligned}$$

ce qui assure la convexité de G_X. □

Proposition 7.11. *Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. Pour tout n, la quantité suivante est bien définie :*

$$G^{(n)}(1-):=\lim_{z\uparrow 1}G^{(n)}(z)=\mathbb{E}\Big[X\times(X-1)\times\ldots\times(X-n+1)\Big]\in[0,+\infty]\ .$$

Par ailleurs, $\mathbb{E}[X^n]<\infty$ si et seulement si $G^{(n)}(1-)<\infty$.

Démonstration. Les propriétés des séries entières assure que la fonction $z\mapsto G(z)$ est de classe $\mathcal{C}^\infty$ sur le disque ouvert $\{z\in\mathbb{C}:|z|<1\}$ et l'on a

$$G^{(n)}(z)=\sum_{k\geq n}z^{k-n}\frac{k!}{(k-n)!}\mathbb{P}(X=k)\ .$$

Par ailleurs, pour tout $z\in[0,\infty)$ la variable aléatoire $X\times(X-1)\times\ldots\times(X-n+1)z^{X-n}$ est positive donc admet une espérance (possiblement infinie). La formule de transfert assure alors que

$$\mathbb{E}[X\times(X-1)\times\ldots\times(X-n+1)z^{X-n}]=\sum_{k\geq n}z^{k-n}\frac{k!}{(k-n)!}\mathbb{P}(X=k)=G^{(n)}(z)\ .$$

On peut alors vérifier que $z\in\sum_{k\geq n}z^{k-n}\frac{k!}{(k-n)!}\mathbb{P}(X=k)$ est une fonction continue et croissante de $[0,1]$ dans $[0,\infty]$. Ceci assure l'existence de $G^{(n)}(1-)$ comme limite croissante ainsi que l'équivalence entre la finitude de $G^{(n)}(1-)$ et celle de $\mathbb{E}[X\times(X-1)\times\ldots\times(X-n+1)]$. Pour finir, on peut montrer par récurrence qu'il y a équivalence entre la finitude de $\mathbb{E}[X\times(X-1)\times\ldots\times(X-n+1)]$ et celle de $\mathbb{E}[X^n]$. □

La fonction génératrice est particulièrement adaptée lorsque l'on manipule des variables aléatoires indépendantes.

Proposition 7.12. *Soient X et Y deux variables aléatoires indépendantes à valeurs dans $\mathbb{N}$. Pour tout z dans $[-1,1]$, on a*

$$G_{X+Y}(z)=G_X(z)\cdot G_Y(z)\ .$$

Démonstration. C'est une conséquence de la Proposition 5.13. En effet, cette proposition assure que pour tout z,

$$\mathbb{E}[z^{X+Y}] = \mathbb{E}[z^X z^Y] = \mathbb{E}[z^X]\mathbb{E}[z^Y] \, .$$

□

7.2.3 Fonction caractéristique

La partie précédente s'est concentrée sur les variables aléatoires à valeurs dans $\mathbb{N}$. On peut souhaiter étendre cette caractérisation de la loi à toutes les variables aléatoires réelles discrètes. Il se trouve que le fait que l'espace d'arrivée était l'ensemble $\mathbb{N}$ jouait un rôle crucial dans la définition de la fonction génératrice et qu'il n'est pas possible d'introduire une telle extension. Cependant, il existe une autre caractérisation de la loi dont l'expression est très semblable à celle de la fonction génératrice, mais dont la portée est beaucoup plus générale.

Définition 7.13. *Soit X une variable aléatoire réelle discrète. La fonction caractéristique de X est définie par*

$$\varphi_X : \begin{cases} \mathbb{R} & \to \mathbb{C} \\ t & \mapsto \mathbb{E}[e^{itX}] \, . \end{cases}$$

Notons que l'espérance est bien définie car $|e^{itX}| = 1$ et ainsi la variable aléatoire e^{itX} est intégrable.

Par des arguments semblables à ceux présentés dans le cas de la fonction génératrice, on peut prouver le résultat suivant (nous n'en donnerons pas de preuve)

Proposition 7.14. *La fonction caractéristique caractérise la loi. Plus précisément, soient X et Y deux variables aléatoires réelles discrètes. $\varphi_X = \varphi_Y$ si et seulement si X et Y ont même loi.*

7.3 Convolution discrète

Dans cette section, on souhaite déterminer la loi de la somme de deux variables aléatoires indépendantes à valeurs dans $\mathbb{Z}$. Pour cela, nous allons introduire un outil très pratique :

Définition 7.15. *Soient μ et ν deux lois de probabilités sur $\mathbb{Z}$. On définit la convolée $\mu * \nu$ de μ et ν comme la mesure de probabilité sur $\mathbb{Z}$ vérifiant*

$$\mu * \nu(\{k\}) = \sum_{m \in \mathbb{Z}} \mu(\{m\})\nu(\{k-m\}) \, , \quad k \in \mathbb{Z} \, .$$

A l'aide de la Proposition 2.6, on vérifie qu'il s'agit bien d'une mesure de probabilité sur $\mathbb{Z}$. Tout d'abord, il est clair que tous les termes $\mu * \nu(\{k\})$ sont positifs. Par ailleurs, les propriétés des séries à termes positifs assurent que

$$\begin{aligned}\sum_{k\in\mathbb{Z}} \mu * \nu(\{k\}) &= \sum_{k\in\mathbb{Z}} \sum_{m\in\mathbb{Z}} \mu(\{m\})\nu(\{k-m\}) \\ &= \sum_{m\in\mathbb{Z}} \mu(\{m\}) \sum_{k\in\mathbb{Z}} \nu(\{k-m\}) \\ &= \sum_{m\in\mathbb{Z}} \mu(\{m\}) \\ &= 1\,.\end{aligned}$$

Proposition 7.16. *Soient X et Y deux variables aléatoires indépendantes à valeurs dans $\mathbb{Z}$. La loi de $X+Y$ est donnée par la convolée des lois de X et Y, c'est-à-dire*

$$\mathbb{P}(X+Y=k) = \sum_{m\in\mathbb{Z}} \mathbb{P}(X=m)\mathbb{P}(Y=k-m)\,, \quad k\in\mathbb{Z}\,.$$

Démonstration. Tout d'abord, la variable aléatoire $X+Y$ prend ses valeurs dans $\mathbb{Z}$. Pour caractériser sa loi, il suffit de déterminer $\mathbb{P}(X+Y=k)$ pour tout $k\in\mathbb{Z}$. Or

$$\begin{aligned}\{X+Y=k\} &= \bigcup_{m\in\mathbb{Z}} \Big(\{X=k-Y\}\cap\{X=m\}\Big) \\ &= \bigcup_{m\in\mathbb{Z}} \Big(\{X=m\}\cap\{Y=k-m\}\Big)\,.\end{aligned}$$

Les événements apparaissant dans l'union sur m sont disjoints et ainsi, par indépendance de X et Y on obtient

$$\mathbb{P}(X+Y=k) = \sum_{m\in\mathbb{Z}} \mathbb{P}(\{X=m\}\cap\{Y=k-m\}) = \sum_{m\in\mathbb{Z}} \mathbb{P}(X=m)\mathbb{P}(Y=k-m)\,.$$

□

Chapitre 8

La marche aléatoire simple

Le but de ce chapitre est d'introduire et d'étudier le modèle de la marche aléatoire simple sur $\mathbb{Z}$. Nous commencerons par des considérations purement combinatoires. Puis nous nous intéresserons à la question suivante : avec quelle probabilité une marche aléatoire simple revient-elle à son point de départ ? Enfin, nous considérerons le modèle de la marche aléatoire simple biaisée et le problème classique de la ruine du joueur.

8.1 Aspects combinatoires

On commence par introduire le modèle de la marche aléatoire simple. Informellement : partant d'un point donné, disons $0 \in \mathbb{Z}$, à chaque pas on tire à pile ou face pour décider si l'on avance ou recule d'une unité. Plus formellement, on se donne un entier $n \geq 1$ et une suite $(X_k)_{1 \leq k \leq n}$ de variables aléatoires indépendantes de loi de Rademacher,

$$\mathbb{P}(X_k = 1) = \mathbb{P}(X_k = -1) = \frac{1}{2} \, , \quad \forall k \in \{1, \ldots, n\} \, .$$

On pose alors

$$S_0 := 0 \, , \quad S_k := X_1 + \ldots + X_k \, , \quad \forall k \in \{1, \ldots, n\} \, .$$

A ce stade, il convient de formuler quelques observations simples. Le vecteur $(S_1, \ldots, S_n)$ est une transformation linéaire (inversible) du vecteur $(X_1, \ldots, X_n)$. Ainsi, la connaissance de l'un de ces deux vecteurs suffit pour déterminer l'autre vecteur. Dans la suite, on manipulera la loi de $(X_1, \ldots, X_n)$ ou celle de $(S_1, \ldots, S_n)$ en fonction des besoins. Notons que $(X_1, \ldots, X_n)$ prend ses valeurs dans l'ensemble

$$\tilde{\Omega}_n := \{(x_1, \ldots, x_n) : x_k \in \{-1, 1\} \quad \forall k \in \{1, \ldots, n\}\} \, ,$$

et que sa loi est la loi uniforme sur $\tilde{\Omega}_n$. De même, $(S_0, S_1, \ldots, S_n)$ prend ses valeurs dans l'ensemble

$$\Omega_n := \{(s_0, s_1, \ldots, s_n) : s_0 = 0 \, , s_{k+1} - s_k \in \{-1, +1\} \quad \forall k \in \{0, \ldots, n-1\}\} \, ,$$

et sa loi est uniforme sur cet ensemble.

On notera également que pour tout $0 \leq n \leq n'$, si l'on part de la marche aléatoire simple à n' pas $(S_0, S_1, \ldots, S_{n'})$, alors sa restriction à ses $(n+1)$ premières coordonnées a la loi de la marche aléatoire simple à n pas. En fait, il est possible de construire la marche aléatoire de longueur infinie (c'est-à-dire $n = \infty$), mais cela dépasse le cadre des probabilités discrètes (l'ensemble des configurations est alors infini non dénombrable).

Lemme 8.1. *Pour tout $n \geq 1$, la variable aléatoire S_n prend ses valeurs dans $\{-n, -n+2, -n+4, \ldots, n-2, n\}$, et pour tout ℓ dans cet ensemble on a*

$$\mathbb{P}(S_n = \ell) = 2^{-n} \binom{n}{(n+\ell)/2} .$$

Démonstration. La première propriété se prouve par récurrence en notant que l'ensemble des valeurs possibles pour S_{n+1} est obtenu en ajoutant ± 1 aux valeurs possibles pour S_n. Concernant la seconde propriété, on note que

$$S_n = |\{k : X_k = 1\}| - |\{k : X_k = -1\}| = 2|\{k : X_k = 1\}| - n .$$

Ainsi $S_n = \ell$ est équivalent à $|\{k : X_k = 1\}| = (n+\ell)/2$. Or il existe $\binom{n}{(n+\ell)/2}$ incréments $x \in \tilde{\Omega}_n$ vérifiant $|\{k : x_k = 1\}| = (n+\ell)/2$. Par ailleurs, le cardinal de $\tilde{\Omega}_n$ vaut 2^n. Comme la loi sur cet ensemble est uniforme, on obtient

$$\mathbb{P}(S_n = \ell) = \frac{|\{S_n = \ell\}|}{|\tilde{\Omega}_n|} = 2^{-n} \binom{n}{(n+\ell)/2} .$$

□

Introduisons à présent le maximum de la marche aléatoire

$$M_n := \max(S_0, \ldots, S_n) .$$

Proposition 8.2 (Principe de réflexion)**.** *Pour tout $k \geq 0$ et tout $\ell \leq k$ on a l'égalité*

$$\mathbb{P}(M_n \geq k, S_n \leq \ell) = \mathbb{P}(S_n \geq 2k - \ell) .$$

On notera que cette identité est **tout à fait remarquable** : à gauche, on s'intéresse à un événement concernant le couple (M_n, S_n) alors qu'à droite, seule la v.a. S_n est mise en jeu.

Démonstration. Soit $s \in \Omega_n$ un chemin tel que $s_n \leq \ell$ et $\max(s_0, \ldots, s_n) \geq k$. On note $i \in \{0, \ldots, n\}$ le premier indice tel que $s_i = k$. On introduit alors le chemin s' qui coïncide avec s jusqu'au temps i, et qui coïncide avec la trajectoire **réfléchie** par rapport à l'axe $y = k$:

$$s'_j = s_j \ , \forall j \in \{0, \ldots, i\} \ , \quad s'_j = 2k - s_j \ , \forall j \in \{i+1, \ldots, n\} .$$

On pourra se convaincre que cela revient à considérer le chemin construit à partir des incréments

$$x'_j = x_j \ , \forall j \in \{0, \dots, i\} \ , \quad x'_j = -x_j \ , \forall j \in \{i+1, \dots, n\} \ .$$

Clairement $s'_n \geq 2k - \ell$. En fait la transformation $s \mapsto s'$ est une bijection entre l'ensemble des chemins vérifiant

$$s_n \leq \ell \ , \quad \max(s_0, \dots, s_n) \geq k \ ,$$

et l'ensemble des chemins vérifiant

$$s_n \geq 2k - \ell \ .$$

Le caractère injectif est facile à vérifier. Concernant le caractère surjectif, étant donné un chemin s' appartenant au deuxième ensemble, il suffit d'appliquer la **même** transformation et l'on obtient un chemin du premier ensemble.
On a donc trouvé une bijection entre les deux ensembles sur lesquels portent les deux probabilités de l'égalité de l'énoncé. La loi de la marche aléatoire simple étant uniforme, cela suffit à conclure. □

Ceci permet alors de déterminer la loi du maximum de la marche aléatoire simple.

Proposition 8.3. *Pour tout $k \in \{0, \dots, n\}$, on a*

$$\begin{aligned} \mathbb{P}(M_n = k) &= \mathbb{P}(S_n = k) + \mathbb{P}(S_n = k+1) \\ &= \frac{1}{2^n} \begin{cases} \binom{n}{(n+k)/2} & \text{si } k \text{ et } n \text{ ont la même parité} \\ \binom{n}{(n+k+1)/2} & \text{sinon.} \end{cases} \end{aligned}$$

Démonstration. On remarque que $\mathbb{P}(M_n = k) = \mathbb{P}(M_n = k \ , S_n \leq k)$ et l'on calcule

$$\begin{aligned} \mathbb{P}(M_n = k \ , S_n \leq k) &= \mathbb{P}(M_n \geq k \ , S_n \leq k) - \mathbb{P}(M_n \geq k+1 \ , S_n \leq k) \\ &= \mathbb{P}(S_n \geq 2k - k) - \mathbb{P}(S_n \geq 2k + 2 - k) \\ &= \mathbb{P}(S_n = k) + \mathbb{P}(S_n = k+1) \ . \end{aligned}$$

Le lemme précédent permet alors de conclure. □

8.2 Récurrence de la marche aléatoire simple

Nous souhaitons répondre à la question suivante : avec quelle probabilité la marche aléatoire simple revient-elle en son point de départ ?

Tout d'abord, il nous faut donner un sens précis à cette question. On introduit alors

$$r_n := \mathbb{P}(\exists k \in \{1, \dots, n\} : S_k = 0) \ ,$$

et l'on remarque que la suite $(r_n)_{n \geq 1}$ est croissante car les événements correspondants sont emboîtés :

$$\{\exists k \in \{1, \dots, n\} : S_k = 0\} \subset \{\exists k \in \{1, \dots, n+1\} : S_k = 0\} \ .$$

Il existe donc $r := \lim_{n\to\infty} r_n$, et l'on a

$$r = \mathbb{P}(\exists k \geq 1 : S_k = 0) \ .$$

Notons cependant que le cadre présenté dans ce cours ne permet pas de donner un sens précis à la quantité de droite. En effet, on manipule ici la marche aléatoire simple de longueur infinie pour laquelle l'espace des configurations est infini non dénombrable. Toujours est-il que la quantité r est bien définie comme limite croissante.

Théoreme 8.4. *La marche aléatoire simple est récurrente, c'est-à-dire, $r = 1$.*

Ce résultat affirme donc qu'avec probabilité 1, la marche aléatoire simple revient en son point de départ. Bien sûr, ce premier temps de retour est aléatoire, et l'on en étudiera des propriétés dans la section suivante.

La démonstration de ce résultat nécessite l'introduction de quantités et résultats intermédiaires. Tout d'abord, on observe que la marche ne peut valoir 0 qu'à des temps pairs (S_n a la même parité que n). On pose alors

$$u_{2n} := \mathbb{P}(S_{2n} = 0) \ , \quad f_{2n} := \mathbb{P}(S_2 \neq 0 \ , \dots \ , S_{2(n-1)} \neq 0 \ , S_{2n} = 0) \ .$$

On voit alors que

$$\begin{aligned} \{\exists k \in \{1, \dots, 2n\} : S_k = 0\} &= \{\exists k \in \{1, \dots, n\} : S_{2k} = 0\} \\ &= \bigcup_{k\in\{1,\dots,n\}} \{S_2 \neq 0 \ , \dots \ , S_{2(k-1)} \neq 0 \ , S_{2k} = 0\} \ , \end{aligned}$$

et que les événements de droite sont deux-à-deux disjoints. Ainsi

$$r_{2n} = \sum_{k=1}^{n} f_{2k} \ .$$

Par ailleurs $r_{2n} = r_{2n+1}$ car la marche ne peut atteindre 0 aux temps impairs.

Lemme 8.5. *Pour tout $n \geq 1$, on a*

$$u_{2n} = \sum_{k=1}^{n} f_{2k} u_{2(n-k)} \ .$$

Démonstration. Soit $C_{2n} \subset \Omega_{2n}$ l'ensemble des chemins de longueur $2n$ arrivant en 0 au temps $2n$. Pour chaque chemin $s \in C_{2n}$, on note 2ℓ le premier temps strictement positif auquel le chemin touche 0 : on note que 2ℓ peut valoir $2, 4, \dots, 2n$. On note $C_{2n,2\ell}$ l'ensemble des chemins correspondants. On voit alors que

$$C_{2n} = \cup_{\ell=1}^{n} C_{2n,2\ell} \ ,$$

et que l'union est disjointe. On observe alors que le cardinal de $C_{2n,2\ell}$ est : le nombre de chemins de longueur 2ℓ revenant en 0 pour la première fois au temps 2ℓ, c'est-à-dire

$2^{2\ell} f_{2\ell}$, multiplié par le nombre de chemins de longueur $2(n-\ell)$ qui se terminent en 0, c'est-à-dire $2^{2(n-\ell)} u_{2(n-\ell)}$:

$$|C_{2n,2\ell}| = 2^{2n} f_\ell u_{n-\ell} \, .$$

On voit donc que

$$u_{2n} = 2^{-2n}|C_{2n}| = \sum_{\ell=1}^{n} 2^{-2n}|C_{2n,2\ell}| = \sum_{\ell=1}^{n} f_{2\ell} u_{2(n-\ell)} \, .$$

□

En utilisant ce lemme et le fait que $u_0 = 1$, on obtient

$$\begin{aligned}\sum_{n=1}^{N} u_{2n} &= \sum_{n=1}^{N}\sum_{k=1}^{n} f_{2k} u_{2(n-k)} = \sum_{k=1}^{N}\sum_{n=k}^{N} f_{2k} u_{2(n-k)} = \sum_{k=1}^{N} f_{2k} \sum_{n=0}^{N-k} u_{2n} \\ &\leq \sum_{k=1}^{N} f_{2k}(1 + \sum_{n=1}^{N} u_{2n}) \, ,\end{aligned}$$

de sorte que

$$r_{2N} = \sum_{k=1}^{N} f_{2k} \geq \frac{\sum_{n=1}^{N} u_{2n}}{1 + \sum_{n=1}^{N} u_{2n}} \, .$$

En passant à la limite $N \to \infty$, et en utilisant le fait que $r \leq 1$, on obtient l'implication

$$\sum_{n=1}^{+\infty} u_{2n} = +\infty \Rightarrow r = 1 \, .$$

Pour prouver le théorème, il nous suffit donc de montrer que la série de gauche diverge. Pour cela, on utilise le Lemme 8.1 pour trouver

$$u_{2n} = \mathbb{P}(S_{2n} = 0) = \frac{1}{2^{2n}} \binom{2n}{n} \, .$$

On rappelle alors la formule de Stirling :

$$n! \sim \frac{1}{\sqrt{2\pi n}} \Big(\frac{n}{e}\Big)^n \, , \quad n \to \infty \, .$$

Un calcul permet de montrer que

$$u_{2n} \sim \frac{1}{\sqrt{\pi n}} \, , \quad n \to \infty \, .$$

Par un critère de comparaison entre séries, combiné au critère de Riemann, on en déduit que $\sum_{n=1}^{+\infty} u_{2n} = +\infty$, ce qui termine la preuve du théorème.

8.3 Temps de retour

On introduit à présent la variable aléatoire

$$T_0 := \inf\{n \geq 1 : S_n = 0\} ,$$

sous la convention $\inf \emptyset = +\infty$. (On notera que cette variable aléatoire nécessite l'introduction de la marche aléatoire simple de longueur infinie, et sort donc du cadre strict de ce cours). La partie précédente a montré que $T_0 < \infty$ avec probabilité 1. On souhaite calculer la série génératrice de T_0

$$F_0(z) := \mathbb{E}[z^{T_0}] = \sum_{n\geq 0} z^{2n}\mathbb{P}(T_0 = 2n) = \sum_{n\geq 0} z^{2n} f_{2n} .$$

Pour ce faire, on introduit la série

$$Q_0(z) := \sum_{n\geq 0} z^{2n} u_{2n} .$$

Proposition 8.6. *Pour tout $z \in (-1, 1)$*
- $Q_0(z) = 1 + Q_0(z)F_0(z)$,
- $Q_0(z) = (1 - z^2)^{-1/2}$,
- $F_0(z) = 1 - (1 - z^2)^{1/2}$.

Démonstration. On calcule

$$\begin{aligned} Q_0(z) &= 1 + \sum_{n\geq 1} z^{2n} u_{2n} = 1 + \sum_{n\geq 1} z^{2n} \sum_{k=1}^{n} f_{2k} u_{2(n-k)} \\ &= 1 + \sum_{k\geq 1} \sum_{n\geq k} z^{2n} f_{2k} u_{2(n-k)} \\ &= 1 + \sum_{k\geq 1} z^{2k} f_{2k} \sum_{n\geq 0} z^{2n} u_{2n} \\ &= 1 + F_0(z)Q_0(z) . \end{aligned}$$

Grâce aux calculs de la section précédente, on sait que

$$Q_0(z) := \sum_{n\geq 0} z^{2n} \frac{1}{2^{2n}} \binom{2n}{n} ,$$

et l'on "reconnaît" le développement en série entière de la fonction $(1 - z^2)^{-1/2}$. Cela permet alors de déduire que

$$F_0(z) = \frac{Q_0(z) - 1}{Q_0(z)} = 1 - Q_0(z)^{-1} = 1 - (1 - z^2)^{1/2} .$$

□

Cela permet d'obtenir pour $z \in (-1,1)$

$$\mathbb{E}[T_0 z^{T_0-1}] = \sum_{n\geq 1}(2n)z^{2n-1}f_{2n} = \frac{d}{dz}F_0(z) = \frac{z}{(1-z^2)^{1/2}} \ .$$

Or le terme de droite tend vers $+\infty$ quand $z \uparrow 1$. Ainsi, par la Proposition 7.11, $\mathbb{E}[T_0] = +\infty$. On en déduit donc que le premier temps de retour, bien que fini avec probabilité 1, est de moyenne infinie.

8.4 Ruine du joueur

On s'intéresse au problème suivant. Deux personnes jouent à un jeu d'argent. On suppose que la fortune initiale du premier joueur est un entier $k \geq 1$ et l'on note $m-k \geq 1$ la fortune initiale du second joueur de sorte que la mise totale est égale à m. A chaque étape, on tire à pile ou face (avec probabilité p et $1-p$ pour $p \in]0,1[$) pour déterminer si le premier joueur récupère une unité d'argent du second joueur, ou l'inverse. Le jeu s'arrête dès qu'un des deux joueurs est ruiné.

On peut modéliser la fortune du premier joueur à l'aide de la marche aléatoire simple suivante (partant de k et possiblement biaisée)

$$S_n := k + \sum_{i=1}^{n} X_i \ ,$$

où $(X_i)_{i\geq 1}$ est une suite de v.a. indépendantes et de même loi de Rademacher de probabilité p

$$\mathbb{P}(X_i = 1) = p \ , \quad \mathbb{P}(X_i = -1) = 1-p \ .$$

On note

$$R := \{\exists n \geq 1 : S_n = 0\} \ ,$$

l'événement sur lequel le premier joueur est ruiné, ainsi que la variable aléatoire

$$\tau := \min\{n \geq 1 : S_n \in \{0, m\}\} \ ,$$

qui modélise le temps auquel la partie s'achève.

Théoreme 8.7 (Ruine du joueur)**.** *Si $p = 1/2$ alors*

$$\mathbb{P}(R) = 1 - \frac{k}{m} \ , \quad \mathbb{E}[\tau] = k(m-k) \ .$$

Si $p \neq 1/2$ alors

$$\mathbb{P}(R) = \frac{\left(\frac{1-p}{p}\right)^m - \left(\frac{1-p}{p}\right)^k}{\left(\frac{1-p}{p}\right)^m - 1} \ , \quad \mathbb{E}[\tau] = \frac{1}{1-2p}\left(k - m\frac{1-\left(\frac{1-p}{p}\right)^k}{1-\left(\frac{1-p}{p}\right)^m}\right) \ .$$

On notera que si $p > 1/2$, alors dans la limite $m \to \infty$, la probabilité de R est typiquement très faible : plus précisément, la solution de $\mathbb{P}(R) = 1/2$ est donnée par un entier k de l'ordre de $\ln 2 / \ln(p/1-p)$.

Démonstration. Il sera commode de noter $\mathbb{P}_k$ la probabilité sous laquelle la marche aléatoire $(S_n)_{n\geq 0}$ part de k au temps $n = 0$. On introduit $t_k := \mathbb{E}_k[\tau]$ et $r_k := \mathbb{P}_k(R)$. On obtient alors pour tout $k \in \{1, \ldots, m-1\}$

$$\begin{aligned} r_k &= \mathbb{P}_k(R) = \mathbb{P}_k(R \mid X_1 = 1)\mathbb{P}_k(X_1 = 1) + \mathbb{P}_k(R \mid X_1 = -1)\mathbb{P}_k(X_1 = -1) \\ &= \mathbb{P}_k(R \mid X_1 = 1)p + \mathbb{P}_k(R \mid X_1 = -1)(1-p) \\ &= r_{k+1}p + r_{k-1}(1-p) \ . \end{aligned}$$

En effet, conditionnellement à l'événement $\{X_1 = 1\}$, la probabilité que le premier joueur soit ruiné est la même que celle partant d'une fortune initiale égale à $k+1$. Par ailleurs on a clairement

$$r_0 = 1 \ , \quad r_m = 0 \ .$$

On résout alors cette suite récurrente linéaire d'ordre 2. Le polynôme caractéristique est donné par

$$pX^2 - X + (1-p) \ ,$$

dont le discriminant est $\frac{1-4(1-p)p}{2}$. Si $p = 1/2$ alors ce polynôme admet 1 comme racine double, et la solution est donnée par

$$r_k = a + bk \ .$$

Les conditions limites impliquent

$$a = 1 \ , \quad 1 + bm = 0 \ ,$$

d'où

$$r_k = 1 - \frac{k}{m} \ .$$

Si $p \neq 1/2$ on obtient deux racines 1 et $(1-p)/p$. Ainsi

$$r_k = a + b\Big(\frac{1-p}{p}\Big)^k \ ,$$

et les conditions limites impliquent

$$a + b = 1 \ , \quad a + b\Big(\frac{1-p}{p}\Big)^m = 0 \ ,$$

d'où

$$r_k = \frac{\left(\frac{1-p}{p}\right)^k - \left(\frac{1-p}{p}\right)^m}{1 - \left(\frac{1-p}{p}\right)^m} \ .$$

Passons au calcul de t_k. Par un raisonnement analogue, on voit que $(t_k)_k$ satisfait

$$\begin{aligned} t_k &= \mathbb{E}_k[\tau] = \mathbb{E}_k[\tau \mid X_1 = 1]\mathbb{P}_k(X_1 = 1) + \mathbb{E}_k[\tau \mid X_1 = -1]\mathbb{P}_k(X_1 = -1) \\ &= \mathbb{E}_k[\tau \mid X_1 = 1]p + \mathbb{E}_k[\tau \mid X_1 = -1](1-p) \\ &= (1 + t_{k+1})p + (1 + t_{k-1})(1-p) \, . \end{aligned}$$

Par ailleurs les conditions limites sont données par

$$t_0 = t_m = 0 \, .$$

La partie homogène de l'équation est la même que précédemment. Il existe une unique solution à ce problème, donnée par une solution particulière de l'équation inhomogène (ne tenant pas compte des conditions limites) plus une solution de l'équation homogène (choisie de sorte que les conditions limites soient vérifiées). On peut vérifier que $-k^2$ et $\frac{k}{1-2p}$ sont des solutions particulières pour $p = 1/2$ et $p \neq 1/2$ respectivement. Un calcul montre alors que les formules de l'énoncé du théorème conviennent. □

8.5 La marche aléatoire simple réfléchie

On s'intéresse à présent à une marche aléatoire simple réfléchie. Il s'agit de la même suite que précédemment, excepté que lorsque la marche arrive en 0 elle ne peut qu'augmenter de 1. En revanche, la marche est toujours "absorbée" en m. Cela correspond à la situation où le premier joueur est renfloué chaque fois qu'il est ruiné.

Dans cette situation, le premier joueur finira toujours par l'emporter. On peut s'intéresser au temps auquel cela se produit :

$$\tau := \inf\{n \geq 1 : S_n = m\} \, .$$

Le calcul développé dans la preuve précédente reste toujours valide, à savoir que $t_k := \mathbb{E}_k[\tau]$ vérifie

$$t_k = (1 + t_{k+1})p + (1 + t_{k-1})(1-p) \, , \quad k \in \{1, \ldots, m-1\} \, ,$$

en revanche les conditions limites sont données par

$$t_0 = 1 + t_1 \, , \quad t_m = 0 \, .$$

On peut conserver la solution particulière de l'équation inhomogène, à savoir $-k^2$ et $\frac{k}{1-2p}$. En revanche, il faut ajuster la solution de l'équation homogène afin de satisfaire les nouvelles conditions limites. Pour $p = 1/2$ on trouve

$$t_k = m^2 - k^2 \, ,$$

alors que pour $p \neq 1/2$ on obtient

$$t_k = \frac{k-m}{1-2p} - \frac{2(1-p)p}{(1-2p)^2}\Big(\Big(\frac{1-p}{p}\Big)^k - \Big(\frac{1-p}{p}\Big)^m\Big) \, .$$

Annexe A

Dénombrabilité

La notion de dénombrabilité est issue de la problématique de comparer la "taille" d'ensembles infinis, dont l'un des premiers résultats marquants a été de formaliser la notion qu'il y a "infiniment plus de nombres réels que de nombres rationnels". En particulier dans ce cadre, les ensembles infinis dénombrables apparaissent comme les plus "petits" des ensembles infinis. Au-delà de ces considérations théoriques, il s'est avéré plus tard que les ensembles dénombrables jouent un rôle particulier dans différents domaines des mathématiques et en particulier vis-à-vis des probabilités. Ce chapitre présente donc quelques résultats et exemples de base pour déterminer si un ensemble est dénombrable ou pas.

A.1 Définition et propriétés

Définition A.1. *Un ensemble E est dit infini dénombrable si et seulement s'il existe une bijection entre E et $\mathbb{N}$.*

Remarque A.1.1. Nous verrons qu'il existe des ensembles infinis non dénombrables et on peut plus généralement définir une notion de "taille infinie" en disant que E et F ont la même taille s'il existe une bijection entre E et F.

Nous commençons par un résultat fondamental qui permet de donner du sens à des "inégalités" sur la notion de taille d'ensemble.

Théoreme A.2 (Cantor-Bernstein)**.** *Soit E et F deux ensembles. S'il existe une fonction injective $f : E \to F$ et une fonction injective $g : F \to E$, alors il existe une bijection entre E et F.*

Démonstration. Soit $\tilde{F} = g(F)$ et soit $A = \cup_{n\in\mathbb{N}}(g \circ f)^n(E \setminus \tilde{F})$ avec la convention que $(g \circ f)^0 = Id$. On pose $\phi(x) = f(x)$ si $x \in A$ et $\phi(x) = g^{-1}(x)$ si $x \notin A$. Montrons que ϕ est une bijection de E dans F.

Tout d'abord, notons que si $x \notin A$, alors en particulier $x \notin E\setminus\tilde{F}$, ou en d'autres termes $x \in g(F)$. La fonction ϕ est donc bien définie. Pour l'injectivité, clairement le seul cas non trivial est pour $x \in A$ et $y \notin A$. Dans ce cas, on remarque que $g(\phi(x)) = g \circ f(x) \in A$

et $g \circ \phi(y) = g \circ g^{-1}(y) = y \notin A$ donc $\phi(x) \neq \phi(y)$. Enfin pour la surjectivité, si $y \in F$ et $g(y) \notin A$, alors par définition $y = \phi(g(y))$. Si $g(y) \in A$, alors par construction $g(y) \in \cup_{n\geq 1}(g \circ f)^n(E \setminus \tilde{F})$ puisque le terme d'ordre 0 n'intersecte pas $g(F) = \tilde{F}$. En particulier $f^{-1}(y)$ est bien défini et $y = \phi(f^{-1})(y)$. □

L'intérêt de ce résultat est qu'il est en pratique plus facile de construire séparément deux fonctions injectives que de construire une bijection. Par ailleurs, l'existence d'une injection de E dans F peut s'interpréter comme une inégalité sur les cardinaux : le cardinal de E est inférieur ou égal à celui de F. Cette affirmation est rigoureuse lorsque les ensembles sont finis, mais n'a pas de sens rigoureux dans le cas contraire.

Le résultat suivant assure que les ensembles infinis dénombrables sont les "plus petits" ensembles infinis.

Proposition A.3. *Pour tout ensemble infini E, il existe une injection de $\mathbb{N}$ dans E.*

Démonstration. On remarque que pour tout $x \in E$, $E \setminus \{x\}$ est toujours infini. On peut construire[1] une suite $(x_n)_n$ d'éléments deux-à-deux disjoints en choisissant d'abord $x_0 \in E$, puis $x_1 \in E \setminus \{x_0\}$, $x_2 \in E \setminus \{x_0, x_1\}$ et ainsi de suite. □

En combinant ce résultat avec le théorème de Cantor Bernstein, on voit qu'il suffit, pour montrer qu'un ensemble infini E est dénombrable, de trouver une application injective de E dans n'importe quel ensemble qu'on sait être dénombrable. C'est ce qu'on fera en pratique.

A.2 Exemples et contre-exemples

Dans cette section, on donne des exemples d'ensembles dénombrables et non dénombrables en tentant de couvrir tous les cas classiques.

Proposition A.4. *Les ensembles $\mathbb{Z}$ et $\mathbb{Q}$ sont dénombrables. Pour tout $d \in \mathbb{N}$, les ensembles $\mathbb{N}^d, \mathbb{Z}^d$ et $\mathbb{Q}^d$ sont dénombrables.*

Démonstration. Pour $\mathbb{Z}$, on vérifie facilement que la fonction $f(z) = 2z$ si $z \geq 0$ et $f(z) = -2z - 1$ si $z < 0$ est une bijection de $\mathbb{Z}$ dans $\mathbb{N}$.

Pour $\mathbb{Q}$, l'écriture d'une fraction sous forme irréductible est clairement une injection de $\mathbb{Q}$ dans $\mathbb{Z} \times \mathbb{N}$.

Enfin les puissances s'obtiennent directement en appliquant plusieurs fois la proposition sur le produit cartésien (Proposition 2.2). □

Cette proposition couvre déjà beaucoup de cas, mais il y a une dernière catégorie notable d'ensembles dénombrables qu'on peut décrire vaguement comme les suites finies d'éléments dénombrables.

1. Cela requiert l'axiome du choix

Proposition A.5. *L'ensemble des parties de $\mathbb{N}$ de cardinaux finis est dénombrable. L'ensemble des polynômes dont les coefficients sont des nombres rationnels est dénombrable. L'ensemble des suites de rationnels qui sont constantes à partir d'un certain rang est dénombrable.*

Démonstration. On commence par le cas des polynômes. Clairement l'ensemble P_d des polynômes à coefficients rationnels de degré au plus d est en bijection avec $\mathbb{Q}^{d+1}$, il suffit de lister les coefficients. L'ensemble de tous les polynômes est donc de la forme $\cup_{d\geq 0} P_d$ avec $\mathbb{P}_d$ dénombrable pour tout d, ce qui est bien dénombrable.

La même preuve s'applique exactement au cas des suites qui sont constantes à partir d'un certain rang. Enfin, on peut construire une injection de l'ensemble des parties finies de $\mathbb{N}$ dans les suites constantes à partir d'un certain rang en énumérant les éléments d'une partie donnée puis en ajoutant une infinité de -1. □

On passe maintenant à des contre-exemples.

Proposition A.6. *L'ensemble des parties de $\mathbb{N}$ n'est pas dénombrable.*

Démonstration. Supposons par l'absurde que $\mathcal{P}(\mathbb{N})$ soit dénombrable et soit f une bijection entre $\mathbb{N}$ et $\mathcal{P}(\mathbb{N})$. On pose $E = \{n : n \notin f(n)\}$ et $x = f^{-1}(E)$. Si $x \in E$ alors $x \in f(x)$ et par définition $x \notin E$. D'un autre coté si $x \notin E$, alors $x \notin f(x)$ et $x \in E$. On a donc une contradiction dans tous les cas. □

Corollaire A.7. *L'ensemble des suites à valeurs dans $\{0,1\}$ n'est pas dénombrable, $\mathbb{R}$ n'est pas dénombrable, $\mathbb{C}$ n'est pas dénombrable.*

Démonstration. Tout d'abord on remarque que si E n'est pas dénombrable et que s'il existe une injection de E dans F alors F ne peut pas être dénombrable non plus.

Pour les suites, on remarque que $E \to 1_E$ définit une bijection entre $\mathcal{P}(\mathbb{N})$ et les suites de 0 et de 1. Pour les réels, on remarque que pour $x = (x_n)_{n\in\mathbb{N}}$ une suite de 0 et de 1, la fonction $f(x) = \sum_{n=0}^{\infty} \frac{x_n}{10^n}$ est bien injective de $\{0,1\}^{\mathbb{N}}$ dans $\mathbb{R}$. L'injection de $\mathbb{R}$ dans $\mathbb{C}$ est triviale. □

En fait les ensembles de ce premier corollaire ont la même "taille" que $\mathcal{P}(\mathbb{N})$. On sort du cadre du chapitre mais par contre les ensembles du prochain corollaire sont vraiment plus grands que $\mathcal{P}(\mathbb{N})$ et $\mathbb{R}$.

Corollaire A.8. *L'ensemble des parties de $\mathbb{R}$, l'ensemble des fonctions réelles ne sont pas dénombrables.*

Annexe B

Sommes et séries classiques

Proposition B.1 (Sommes des n premiers entiers)**.** *Pour tout n dans $\mathbb{N}_*$, on a*

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

Proposition B.2 (Sommes des n premiers carrés d'entiers)**.** *Pour tout n dans $\mathbb{N}_*$, on a*

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

Proposition B.3 (Série exponentielle)**.** *Pour tout z dans $\mathbb{C}$, on a*

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

Proposition B.4 (Série logarithmique)**.** *Pour tout $x \in]-1, 1[$, on a*

$$-\ln(1-x) = \sum_{n=1}^{\infty} \frac{x^n}{n}.$$

Proposition B.5 (Série géométrique)**.** *Pour tout z dans $\mathbb{C}$ tel que $|z| < 1$, on a*

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n.$$

Rappel : en dérivant, on trouve $\sum_{n=1}^{\infty} n x^{n-1} = \frac{1}{(1-x)^2}$ et $\sum_{n=2}^{\infty} n(n-1)x^{n-2} = \frac{2}{(1-x)^3}$, voir (5.4.1).

Proposition B.6. *Quand N tend vers l'infini, on a*

$$\sum_{n=1}^{N} \frac{1}{n} = \ln(N) + O(1)\ .$$

En particulier la série de terme général $(\frac{1}{n})_{n\geq 1}$ n'est pas sommable.