



**UNIVERSITÉ
BOURGOGNE
EUROPE**



**INSTITUT DE MATHÉMATIQUES
DE BOURGOGNE**

COURS DE PROBABILITÉS

Yoann Offret 2025-2026

Table des matières

1	Modélisation probabiliste	7
1.1	Différentes interprétations	8
1.1.1	Une proportion (approche classique de Laplace)	8
1.1.2	Une fréquence (approche fréquentiste)	9
1.1.3	Une crédence (approche bayésienne)	9
1.2	Triplet fondamental	10
1.2.1	Ensemble des issues possibles	10
1.2.2	Tribu des événements	11
1.2.3	Mesure de probabilité	14
2	Indépendance et probabilités conditionnelles	31
2.1	Indépendance d'événements et de tribus	33
2.1.1	Indépendance de deux événements	33
2.1.2	Indépendance de deux tribus	35
2.1.3	Indépendance d'une famille finie d'événements	36
2.1.4	Indépendance d'une famille quelconque d'événements	38
2.2	Probabilités conditionnelles	39
3	Les variables aléatoires et leurs lois	47
3.1	Définitions générales	49
3.1.1	Définition d'une variable aléatoire	49
3.1.2	Loi d'une variable aléatoire	50
3.1.3	Indépendance de variables aléatoires	51
3.2	Variables aléatoires discrètes	53
3.2.1	Définitions et caractérisations	53
3.2.2	Lois discrètes classiques	53
3.2.3	Indépendance	56
3.2.4	Lois conditionnelles	58
3.3	Variables aléatoires réelles	59
3.3.1	Définition en termes d'intervalles	60

3.3.2	Tribu borélienne de \mathbb{R}	60
3.3.3	Fonctions de Répartition	61
3.3.4	Indépendance	65
3.3.5	Variables Aléatoires Continues	66
3.3.6	Lois continues classiques	69
4	Espérance d'une variable aléatoire réelle	75
4.1	Cas des variables aléatoires réelles discrètes	76
4.1.1	Définition	77
4.1.2	Théorème de Transfert	77
4.1.3	Croissance et linéarité de l'espérance	79
4.1.4	Fonction génératrice d'une variable aléatoire discrète	80
4.1.5	Indépendance et espérance	81
4.1.6	Variance et covariance	83
4.1.7	Moments d'une variable aléatoire	85
4.1.8	Cas des lois discrètes classiques	87
4.2	Cas des variables aléatoires réelles continues	90
4.2.1	Définition et propriétés	90
4.2.2	Théorème de transfert	90
4.2.3	Variance	91
4.2.4	Indépendance et espérance	91
4.2.5	Lois continues classiques	91
4.2.6	Un mot sur la fonction caractéristique	92
5	Théorèmes limites et inégalités de concentrations	93
5.1	Inégalités de concentration	93
5.2	Loi des Grands Nombres	94
5.3	Théorème Central Limite	95
6	Introduction aux graphes et aux chaînes de Markov finies	97
6.1	Notions de base sur les graphes	97
6.1.1	Définition d'un graphe	97
6.1.2	Sommets adjacents et degré	98
6.1.3	Ordre et taille d'un graphe	100
6.1.4	Chaîne et longueur d'une chaîne	101
6.1.5	Graphe connexe	102
6.1.6	Matrice d'adjacence d'un graphe	103
6.2	Chaînes de Markov	104
6.2.1	Matrices stochastiques et graphes de transition	104
6.2.2	Définition d'une chaîne de Markov finie	105
6.2.3	Lois invariantes	107
6.2.4	Chaînes à deux états : analyse complète	109

7 Exercices	111
7.1 Espaces probabilisés	112
7.2 Indépendance et probabilités conditionnelles	115
7.3 Les variables aléatoires et leurs lois	120
7.4 Espérance des variables aléatoires	125
7.5 Inégalités de concentration	129
7.6 Graphes et chaînes de Markov	130

Modélisation probabiliste

Une grande partie de notre vie repose sur la conviction que l'avenir n'est pas prévisible. Par exemple, les jeux de hasard tels que les dés ou la roulette auraient peu d'amateurs si leurs résultats étaient connus à l'avance. Nous exprimons cette croyance à l'aide de mots tels que « hasard », « chance », « aléatoire », « stochastique » ou encore « probabilité », et nous cherchons à attribuer à ces termes un sens à la fois qualitatif et quantitatif.

La théorie mathématique des probabilités incorpore ces notions, qui existent de manière implicite dans le langage courant. Elle les formalise au moyen d'un système d'axiomes conduisant à des conclusions cohérentes avec l'expérience et la pratique. Évidemment, il n'y a pas de « hasard » au sens strict dans cette théorie : elle repose entièrement sur la **théorie des ensembles**.

D'ailleurs, il n'est pas certain qu'il existe un hasard fondamental dans les lois régissant notre univers, au sens où il ne résulterait pas simplement d'un manque d'information. Un tel hasard est parfois postulé à très petite échelle en mécanique quantique. Mais dans la plupart des situations, le « hasard », la « chance » ou l'« aléa » découlent avant tout de notre ignorance : si l'information initiale était connue avec précision, le résultat pourrait être déterminé. Par exemple, si l'on connaissait exactement la position, la vitesse, le moment cinétique et l'inclinaison d'une pièce de monnaie au moment du lancer, les lois de la physique permettraient de prévoir avec certitude le résultat. De même, le nombre de personnes présentes dans une file d'attente à la caisse d'un supermarché peut être modélisé par la théorie des probabilités ; cependant, si l'on connaissait parfaitement les emplois du temps et les intentions de chaque client, il n'y aurait plus rien de véritablement aléatoire.

Ainsi, tout comme la théorie des équations différentielles sert d'outil pour décrire la mécanique newtonienne, la théorie des probabilités fournit un cadre mathématique pour modéliser des situations où intervient le « hasard », qu'il résulte d'un manque d'information ou d'un mécanisme plus fondamental.

1.1 Différentes interprétations

La théorie des probabilités nous servira à modéliser une expérience aléatoire, c'est-à-dire un phénomène dont on ne peut pas prédire l'issue avec certitude, que ce soit en raison d'un manque d'information ou d'un véritable aléa sous-jacent. Par exemple :

1. « Le dé lancé tombera sur un multiple de 3. »
2. « L'enfant à naître sera une fille. »
3. « L'équipe de foot de Dijon sera reléguée en seconde ligue cette année. »

La première étape consiste à décrire l'ensemble des différentes issues possibles de l'expérience — très souvent noté Ω et appelé univers — puis à définir l'ensemble de tous les événements qui peuvent en découler — noté \mathcal{F} . On cherche ensuite à associer à chacune de ces assertions une probabilité, c'est-à-dire un nombre compris entre 0 et 1 qui mesure la chance qu'elle se réalise. Autrement dit, on définit une fonction

$$\mathbb{P} : \mathcal{F} \subset \mathcal{P}(\Omega) \longrightarrow [0, 1].$$

Étant donné un événement A , le nombre $\mathbb{P}(A)$ peut recevoir diverses interprétations. Historiquement, trois grandes approches se sont imposées : l'interprétation classique (ou laplacienne), l'interprétation fréquentiste (ou objective), et l'interprétation bayésienne (ou subjective).

1.1.1 Une proportion (approche classique de Laplace)

Dans le premier exemple, l'expérience consiste à lancer un dé. Il y a 6 issues possibles : $\Omega = \{1, 2, 3, 4, 5, 6\}$. Comme il n'y a aucune raison de privilégier une face plutôt qu'une autre, on suppose que la situation est équiprobable et l'on associe naturellement la probabilité $1/6$ à chacun des événements élémentaires $\{1\}, \dots, \{6\}$.

Tout événement correspond alors à une partie de Ω . Par exemple, l'événement A : « le dé lancé tombera sur un multiple de 3 » correspond à $A = \{3, 6\}$. On pourra se référer à la Figure 1.1.

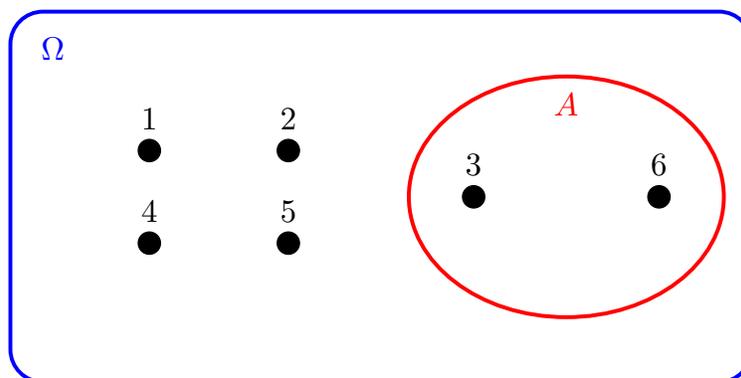


FIGURE 1.1 – Issues possibles et événements.

Ainsi, $\mathcal{F} = \mathcal{P}(\Omega)$ et, pour calculer la probabilité d'un événement A , il suffit d'évaluer la proportion de A dans Ω , i.e.

$$\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}.$$

En particulier, la probabilité de l'événement A est $2/6$, ce que l'on pressentait déjà.

Cette interprétation, héritée de Laplace, est adaptée lorsque le nombre d'issues est fini et que toutes sont symétriques. Elle montre vite ses limites lorsque les situations ne sont pas équiprobables ou lorsqu'on a affaire à un nombre infini de possibilités.

1.1.2 Une fréquence (approche fréquentiste)

Pour l'exemple d'un enfant à naître, on estime intuitivement qu'il y a une chance sur deux environ que ce soit une fille. On peut interpréter cette probabilité comme une fréquence, c'est-à-dire comme une proportion asymptotique. Plus précisément, si F_n désigne le nombre de filles parmi n naissances, on imagine que

$$\lim_{n \rightarrow \infty} \frac{F_n}{n} = \frac{1}{2}.$$

Cependant, il n'est pas évident que cette limite existe toujours. Sous certaines hypothèses, ce résultat est garanti par la **loi forte des grands nombres**.

Cette approche, défendue notamment par von Mises, fonde la statistique fréquentiste et les méthodes d'estimation ou de tests d'hypothèses. Elle repose sur l'idée qu'une probabilité n'a de sens que par répétition d'une expérience. Elle est parfois critiquée lorsqu'il s'agit d'événements uniques, qui ne peuvent être répétés.

1.1.3 Une crédence (approche bayésienne)

Une crédence est le degré de confiance que l'on accorde à une hypothèse ou une théorie. C'est une croyance quantifiée par une probabilité. Par exemple, on peut estimer que le DFCO a une chance sur dix de monter en Ligue 1 cette année. Dans le monde des paris, on parle plutôt de cote : on dirait ici une cote de un contre neuf.

On ne peut pas identifier directement cette probabilité à une proportion ou à une fréquence, car d'une part le nombre de saisons n'est pas fixé, et d'autre part elles ne sont ni indépendantes ni identiques.

Cette approche, dite bayésienne, considère la probabilité comme un degré de croyance rationnelle, qui doit être actualisé par de nouvelles informations grâce à la **formule de Bayes**. Elle met en jeu des lois a priori (avant observation des données) et des lois a posteriori (après observation).

Bien sûr, toutes les crédences ne se valent pas : celles des bookmakers professionnels, fondées sur de nombreuses données, sont généralement plus fiables que celles de votre professeur de probabilités. De plus, ce ne sont que des croyances a priori, qu'il faut toujours actualiser a posteriori en fonction des observations.

1.2 Triplet fondamental

Toute expérience impliquant le hasard peut être modélisée par un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Un tel espace se compose de trois objets :

- un ensemble Ω , appelé univers ou ensemble des issues possibles, qui regroupe tous les résultats élémentaires $\omega \in \Omega$ de l'expérience ;
- une collection \mathcal{F} d'événements, c'est-à-dire de sous-ensembles de Ω , représentant les assertions que l'on souhaite modéliser ;
- une mesure de probabilité \mathbb{P} définie sur \mathcal{F} , qui associe à chaque événement $A \in \mathcal{F}$ un nombre $\mathbb{P}(A) \in [0, 1]$ interprété comme sa probabilité.

Nous allons définir plus précisément ces trois objets et en donner des exemples. Les notations d'un espace de probabilité ne sont pas figées : $(\Omega, \mathcal{F}, \mathbb{P})$ est la plus courante, mais $(E, \mathcal{A}, \mathbb{Q})$ ou d'autres variantes peuvent être rencontrées.

1.2.1 Ensemble des issues possibles

Lorsque l'on modélise une expérience aléatoire, la première étape consiste à identifier toutes les situations que l'on considère comme des résultats possibles. Cet ensemble joue un rôle fondamental : il constitue l'*univers* de l'expérience.

Définition 1 (Univers). *L'ensemble Ω des issues possibles d'une expérience est un ensemble non vide. On l'appelle également univers de l'expérience.*

Le choix de Ω n'est pas toujours évident, et surtout il n'est pas unique : il dépend de ce que l'on souhaite modéliser et du degré de précision recherché.



Le choix de l'univers n'est donc pas canonique, comme l'illustre l'exemple suivant.

★ Exemple d'un lancer de dé

Pour un lancer de dé, l'univers le plus naturel est évidemment $\Omega = \{1, 2, 3, 4, 5, 6\}$. Cependant, on peut vouloir raffiner la description. Par exemple, on pourrait inclure des situations atypiques telles que : « le dé glisse sous le canapé et n'est retrouvé qu'au prochain déménagement », ou encore « le dé reste sur la tranche ». En notant ω_1 et ω_2 ces deux issues particulières, on peut alors définir $\Omega = \{1, 2, 3, 4, 5, 6, \omega_1, \omega_2\}$.

Il est important de remarquer qu'à ce stade, nous ne faisons aucune hypothèse sur les probabilités associées : certaines de ces éventualités (comme le dé restant sur la tranche) pourront recevoir une probabilité nulle par la suite.

1.2.2 Tribu des événements

Étant donné un univers Ω décrivant l'expérience aléatoire, il faut préciser quels sont les sous-ensembles de Ω que l'on décide de considérer comme des *événements*. Autrement dit, il s'agit de choisir une collection \mathcal{F} de parties de Ω sur laquelle nous pourrions définir des probabilités. On a déjà remarqué que, par nature, \mathcal{F} doit être un sous-ensemble de l'ensemble des parties de Ω , c'est-à-dire

$$\mathcal{F} \subset \mathcal{P}(\Omega).$$

Quelle structure est-il naturel de supposer sur \mathcal{F} ?

i) Stabilité par passage au complémentaire

Si A est un événement, i.e. $A \in \mathcal{F}$, il est naturel de demander que son *complémentaire*, noté A^c (ou parfois \bar{A}), soit lui aussi un événement. On l'appelle l'*événement contraire* de A .

★ Exemple de trois lancers d'une pièce de monnaie

On modélise l'expérience avec $\Omega = \{P, F\}^3$. Si $A =$ « le second lancer tombe sur pile », alors $A^c =$ « le second lancer tombe sur face » et on a :

$$A = \{(P, P, P), (F, P, P), (P, P, F), (F, P, F)\} = \{P, F\} \times \{P\} \times \{P, F\},$$

et

$$A^c = \{(P, F, P), (F, F, P), (P, F, F), (F, F, F)\} = \{P, F\} \times \{F\} \times \{P, F\}.$$

ii) Stabilité par réunion

Si A et B sont deux événements de \mathcal{F} , il est naturel de supposer que leur réunion $A \cup B$ en soit également un. On appelle $A \cup B$ l'événement « A ou B », le « ou » étant inclusif (les deux événements peuvent donc se réaliser simultanément).

★ Exemple d'un lancer de dé

On modélise l'expérience en prenant $\Omega = \{1, \dots, 6\}$. Si $A =$ « le résultat du dé est un multiple de 3 » et $B =$ « le résultat du dé est inférieur ou égal à 3 », alors on a $A = \{3, 6\}$, $B = \{1, 2, 3\}$ et $A \cup B = \{1, 2, 3, 6\}$.

iii) Stabilité par intersection

De la même manière, si $A, B \in \mathcal{F}$, il est naturel que $A \cap B$ appartienne à \mathcal{F} . On appelle cet événement l'événement « A et B ».

★ Exemple de trois lancers d'une pièce de monnaie

On prend une nouvelle fois $\Omega = \{P, F\}^3$. Si $A =$ « on obtient exactement une fois pile et deux fois face » et $B =$ « le second lancer tombe sur face », alors on a :

$$A = \{(P, F, F), (F, P, F), (F, F, P)\}, \quad B = \{(P, F, P), (P, F, F), (F, F, P), (F, F, F)\},$$

et

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

Ces trois conditions de stabilité (complémentaire, réunion et intersection) paraissent naturelles. Reste alors à se demander ce qu'il en est pour les unions et intersections d'un nombre fini, dénombrable ou même indénombrable d'événements.

On peut montrer que la stabilité par réunion ou intersection de deux événements entraîne automatiquement la stabilité par unions et intersections finies. En effet, pour tout $n \geq 1$, si $A_1, \dots, A_n \in \mathcal{F}$, alors

$$\bigcup_{i=1}^n A_i \in \mathcal{F} \quad \text{et} \quad \bigcap_{i=1}^n A_i \in \mathcal{F}.$$

La preuve se fait facilement par récurrence en utilisant l'associativité de \cup et \cap .

Néanmoins, restreindre \mathcal{F} aux seules unions et intersections finies s'avère trop limitatif pour de nombreuses situations probabilistes. En pratique, on a besoin de manipuler des unions ou intersections infinies, en particulier dénombrables, comme on le verra plus loin.



Exiger la stabilité par unions ou intersections d'un nombre infini **non dénombrable** d'événements peut mener à de sérieuses difficultés. C'est pourquoi on impose uniquement la stabilité par opérations *dénombrables*.

Il est temps maintenant de rassembler ces propriétés dans une définition générale, qui jouera un rôle central dans toute la théorie des probabilités.

Définition 2 (Tribu des événements). *L'ensemble des événements \mathcal{F} d'une expérience aléatoire décrite par un univers Ω est une collection non vide de parties de Ω telle que :*

1. $\emptyset, \Omega \in \mathcal{F}$;
2. $\forall A \in \mathcal{F}, A^c \in \mathcal{F}$;
3. $\forall (A_n)_{n \geq 1} \subset \mathcal{F}, \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$;
4. $\forall (A_n)_{n \geq 1} \subset \mathcal{F}, \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.

*On dit que \mathcal{F} est une **tribu** (ou σ -algèbre) sur Ω , et (Ω, \mathcal{F}) est alors appelé un **espace probabilisable**.*

Cette définition peut paraître un peu lourde au premier abord, mais elle s'allège grâce à une caractérisation équivalente :

Propriété 1. *Soit $\mathcal{F} \subset \mathcal{P}(\Omega)$. Alors \mathcal{F} est une tribu sur Ω si et seulement si :*

1. $\Omega \in \mathcal{F}$;
2. $\forall A \in \mathcal{F}, A^c \in \mathcal{F}$;
3. $\forall (A_n)_{n \geq 1} \subset \mathcal{F}, \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Démonstration. La preuve élémentaire est laissée au lecteur. □



Dans la suite, on supposera toujours que \mathcal{F} contient tous les « événements élémentaires » $\{\omega\}$ avec $\omega \in \Omega$, ce qui n'est pas strictement obligatoire en toute généralité mais simplifie grandement la présentation.

iv) Un peu de vocabulaire

En probabilités, on utilise un vocabulaire spécifique pour parler des ensembles et des relations entre eux. Ce langage, différent du vocabulaire ensembliste usuel, met davantage l'accent sur l'interprétation aléatoire des objets manipulés. Le tableau ci-dessous (Figure 1.2) présente les principales correspondances.

Notations	Vocabulaire ensembliste	Vocabulaire probabiliste
Ω	ensemble total / plein	univers, totalité des issues
\emptyset	ensemble vide	événement impossible
ω	élément de Ω	issue de l'expérience
A	sous-ensemble de Ω	événement
$\omega \in A$	ω appartient à A	l'événement A est réalisé
$\{\omega\}$	singleton	événement élémentaire
$A \subset B$	A inclus dans B	si A est réalisé alors B aussi
A^c (ou \bar{A})	complémentaire	événement contraire
$A \cup B$	union de A et B	A ou B (ou inclusif)
$A \cap B$	intersection de A et B	A et B
$B \setminus A$	B privé de A	B mais pas A
$B \Delta A$	différence symétrique	A ou B , mais pas les deux (exclusif)
$A \cap B = \emptyset$	A et B disjoints	A et B incompatibles
$\bigcup_{n \geq 1} A_n$	union dénombrable	il existe au moins un A_n qui se réalise
$\bigcap_{n \geq 1} A_n$	intersection dénombrable	tous les A_n se réalisent simultanément
$\bigcup_{k \in \mathbb{N}} \bigcap_{n \geq k} A_n$	limite inférieure	tous les A_n se réalisent à partir d'un certain rang
$\bigcap_{k \in \mathbb{N}} \bigcup_{n \geq k} A_n$	limite supérieure	les A_n se réalisent une infinité de fois

FIGURE 1.2 – Correspondance entre vocabulaire ensembliste et probabiliste.



Si $A, B \subset \Omega$, on n'écrira pas $A \implies B$ (qui n'a pas de sens ensembliste) mais bien $A \subset B$. L'interprétation probabiliste consiste à dire que si l'événement A est réalisé alors nécessairement B aussi.

1.2.3 Mesure de probabilité

Soit (Ω, \mathcal{F}) où Ω représente un univers et \mathcal{F} une tribu d'événements décrivant une expérience aléatoire : on dit alors que (Ω, \mathcal{F}) est un espace probabilisable. Dans le langage de la théorie de la mesure, on dit aussi que c'est un espace mesurable.

À tout événement $A \in \mathcal{F}$, on cherche à associer une probabilité $\mathbb{P}(A)$ qui peut représenter une fréquence, une proportion ou une crédence. Dans tous les cas, on a $\mathbb{P}(A) \in [0, 1]$. Une telle fonction $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ sera appelée une *mesure de probabilité*.

Cependant, on ne peut pas assigner arbitrairement des probabilités. En premier lieu, il est clair que $\mathbb{P}(\emptyset) = 0$ et $\mathbb{P}(\Omega) = 1$. En effet, \emptyset représente l'événement impossible (aucun résultat) tandis que Ω représente l'événement certain (tous les résultats possibles).

De plus, considérons un exemple simple : si la probabilité d'« être né un samedi » est de $13/100$ et celle d'« être né un dimanche » est de $12/100$, il est naturel d'affirmer que la probabilité d'« être né un week-end » est de $25/100$. On voit ici apparaître la règle d'additivité pour des événements disjoints : pour tout $A, B \in \mathcal{F}$ tels que $A \cap B = \emptyset$, on a

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

Plus généralement, par récurrence, pour tout $n \geq 1$ et $A_1, \dots, A_n \in \mathcal{F}$ disjoints,

$$\mathbb{P}(A_1 \sqcup \dots \sqcup A_n) = \mathbb{P}(A_1) + \dots + \mathbb{P}(A_n).$$

On dit alors qu'une mesure de probabilité est **additive**. Or, une tribu est stable par réunion dénombrable. Il est donc naturel d'exiger que la mesure soit non seulement finiment additive, mais aussi **σ -additive**, c'est-à-dire additive sur toute famille dénombrable d'événements disjoints : pour toute suite $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ telle que $A_i \cap A_j = \emptyset$ pour $i \neq j$,

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

Remarquons que l'événement $\bigcup_{n=0}^{\infty} A_n$ appartient bien à \mathcal{F} , donc sa probabilité est bien définie. Par ailleurs, pour une suite de nombres positifs $(\mathbb{P}(A_n))_{n \geq 0}$, la série

$$\sum_{n=0}^{\infty} \mathbb{P}(A_n) \stackrel{\text{def}}{=} \lim_{k \rightarrow \infty} \sum_{n=0}^k \mathbb{P}(A_n)$$

existe toujours, car la suite des sommes partielles est croissante et majorée par 1.

Enfin, on rappelle un fait classique de la théorie des séries : dès qu'une série est à termes positifs (ou plus généralement absolument convergente), sa valeur ne dépend ni de l'ordre de sommation, ni d'éventuels regroupements de termes (sommation par paquets).

i) Définition d'une mesure de probabilité

Après avoir discuté des propriétés souhaitables d'une probabilité, nous pouvons maintenant en donner la définition rigoureuse.

Définition 3. Soit (Ω, \mathcal{F}) un espace probabilisable. On appelle **mesure de probabilité** (ou loi de probabilité) sur (Ω, \mathcal{F}) une application $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ telle que :

1. $\mathbb{P}(\Omega) = 1$;
2. pour toute suite $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ d'événements deux à deux disjoints,

$$\mathbb{P}\left(\bigsqcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n) \quad (\sigma\text{-additivité}).$$

Le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ est alors appelé un **espace de probabilité**.

Pour mieux comprendre la définition, examinons un exemple classique.

★ Exemple : lancer d'une pièce équilibrée ou biaisée

On pose $\Omega = \{P, F\}$ et $\mathcal{F} = \mathcal{P}(\Omega) = \{\emptyset, \{P\}, \{F\}, \Omega\}$. Supposons que la pièce puisse être biaisée : on note $p \in [0, 1]$ la probabilité d'obtenir pile. Lorsque $p = 1/2$, la pièce est équilibrée. On définit alors

$$\mathbb{P}(\{P\}) = p, \quad \mathbb{P}(\{F\}) = 1 - p, \quad \mathbb{P}(\Omega) = 1, \quad \mathbb{P}(\emptyset) = 0.$$

Autrement dit :

$$\begin{array}{rcl} \mathbb{P} : & \mathcal{F} & \longrightarrow [0, 1] \\ & \emptyset & \longmapsto 0 \\ & \{P\} & \longmapsto p \\ & \{F\} & \longmapsto 1 - p \\ & \Omega & \longmapsto 1 \end{array}$$

On vérifie facilement que \mathbb{P} est bien une mesure de probabilité.

À partir de la définition d'une mesure de probabilité, on peut établir un certain nombre de propriétés fondamentales. Elles constituent la « boîte à outils » de base que nous utiliserons tout au long du cours.

Propriété 2 (Propriétés élémentaires). Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Alors :

1. $\mathbb{P}(\emptyset) = 0$.
2. Pour tout $A \in \mathcal{F}$, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
3. Si $A, B \in \mathcal{F}$ et $A \subset B$, alors $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$.
4. Si $A, B \in \mathcal{F}$ et $A \subset B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$.
5. Pour tous $A, B \in \mathcal{F}$, $\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c)$.
6. Pour tous $A, B \in \mathcal{F}$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Démonstration. Ces résultats se démontrent directement à partir des axiomes et constituent d'excellents exercices d'entraînement. \square

Parmi les applications immédiates, citons deux formules très classiques.

Propriété 3 (Formule des probabilités totales 1). Soit $(A_n)_{n \geq 0} \subset \mathcal{F}$ une partition d'événements de Ω . Alors, pour tout $B \in \mathcal{F}$, on a

$$\mathbb{P}(B) = \sum_{n=0}^{\infty} \mathbb{P}(B \cap A_n).$$

On rappelle qu'une famille $(A_n)_{n \geq 0}$ est une **partition** de Ω si

- les A_n sont deux à deux disjoints : $A_i \cap A_j = \emptyset$ si $i \neq j$;
- leur réunion recouvre tout l'univers : $\bigcup_{n \geq 0} A_n = \Omega$.

★ Exemple : Tirage de deux dés

On lance deux dés équilibrés à 6 faces. Soit $B = \ll$ la somme des deux dés vaut 7 \gg . On prend comme système complet $(A_i)_{i=1}^6$ où $A_i = \ll$ le premier dé vaut i \gg . Alors

$$\mathbb{P}(B) = \sum_{i=1}^6 \mathbb{P}(B \cap A_i).$$

Or $B \cap A_i$ signifie « premier dé = i et second dé = $7 - i$ », ce qui donne exactement une issue favorable parmi les 36 équiprobables, donc $\mathbb{P}(B \cap A_i) = \frac{1}{36}$ pour tout $i = 1, \dots, 6$. Ainsi

$$\mathbb{P}(B) = \sum_{i=1}^6 \frac{1}{36} = \frac{6}{36} = \frac{1}{6}.$$

Propriété 4 (Formule du crible). Soient $n \geq 1$ et A_1, \dots, A_n des événements. Alors

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathbb{P}(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} \mathbb{P}(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} \mathbb{P}(A_1 \cap \dots \cap A_n).$$

C'est un outil classique de dénombrement, très utilisé en combinatoire.

★ Exemple : multiples de 2, 3 et 5

On choisit un entier au hasard entre 1 et 30. Soit $A_1 =$ « divisible par 2 », $A_2 =$ « divisible par 3 » et $A_3 =$ « divisible par 5 ». La formule du crible donne :

$$\mathbb{P}(A_1 \cup A_2 \cup A_3) = \sum_{i=1}^3 \mathbb{P}(A_i) - \sum_{1 \leq i < j \leq 3} \mathbb{P}(A_i \cap A_j) + \mathbb{P}(A_1 \cap A_2 \cap A_3).$$

On calcule les différentes probabilités :

$$\mathbb{P}(A_1) = \frac{15}{30}, \quad \mathbb{P}(A_2) = \frac{10}{30}, \quad \mathbb{P}(A_3) = \frac{6}{30},$$

$$\mathbb{P}(A_1 \cap A_2) = \frac{5}{30}, \quad \mathbb{P}(A_1 \cap A_3) = \frac{3}{30}, \quad \mathbb{P}(A_2 \cap A_3) = \frac{2}{30},$$

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{1}{30}.$$

On obtient ainsi :

$$\mathbb{P}(A_1 \cup A_2 \cup A_3) = \frac{31}{30} - \frac{10}{30} + \frac{1}{30} = \frac{22}{30}.$$

La probabilité qu'un entier choisi au hasard et uniformément entre 1 et 30 soit divisible par 2, 3 ou 5 est de $22/30$.

D'autres résultats concernent les limites de suites d'événements. Ils illustrent l'importance des propriétés de monotonie et de σ -additivité.

Propriété 5 (Probabilité d'une union croissante). Soit $(A_n)_{n \geq 0}$ une suite croissante d'événements (pour l'inclusion), i.e. $A_n \subset A_{n+1}$ quelque soit $n \geq 0$. Alors

$$\mathbb{P}\left(\bigcup_{n \geq 0} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n),$$

la limite étant celle d'une suite croissante.

★ Exemple : Apparition d'un 6 au lancer de dés

On répète à l'infini le lancer d'un dé équilibré. Soit $A_n =$ « un 6 apparaît au plus tard au n -ème lancer ». Alors (A_n) est une suite croissante et :

$$\bigcup_{n \geq 1} A_n = \text{« un 6 apparaît un jour »}.$$

On a alors

$$\mathbb{P}(A_n) = 1 - \left(\frac{5}{6}\right)^n.$$

On en déduit donc que

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = \lim_{n \rightarrow \infty} \left(1 - \left(\frac{5}{6}\right)^n\right) = 1.$$

Propriété 6 (Probabilité d'une intersection décroissante). Soit $(A_n)_{n \geq 0}$ une suite décroissante d'événements (pour l'inclusion). Alors

$$\mathbb{P}\left(\bigcap_{n \geq 0} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n),$$

la limite étant celle d'une suite décroissante.

★ Exemple : suites décroissantes d'événements

On lance une pièce équilibrée. Pour chaque $n \geq 1$, on note

$$A_n = \text{« les } n \text{ premiers lancers donnent tous pile »}.$$

Alors (A_n) est une suite décroissante et

$$\mathbb{P}(A_n) = \left(\frac{1}{2}\right)^n.$$

On remarque également que

$$\bigcap_{n \geq 1} A_n = \text{« tous les lancers sont pile »}.$$

Cet événement est donc de probabilité 0 car

$$\mathbb{P}\left(\bigcap_{n \geq 1} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n) = 0.$$

Enfin, rappelons une propriété très utile, conséquence directe de la σ -additivité.

Propriété 7 (σ -sous-additivité). Soit $(A_n)_{n \geq 0}$ une famille dénombrable d'événements. Alors

$$\mathbb{P}\left(\bigcup_{n \geq 0} A_n\right) \leq \sum_{n=0}^{\infty} \mathbb{P}(A_n).$$

★ Exemple : suite de piles d'au moins 5 en 10 lancers

On lance une pièce équilibrée 10 fois. On s'intéresse à l'événement

$B = \text{« la plus longue suite de piles est d'au moins 5 »}.$

Pour $i = 1, \dots, 6$, définissons

$E_i = \text{« les lancers } i, i+1, \dots, i+4 \text{ donnent tous pile »}.$

On a alors $B = \bigcup_{i=1}^6 E_i$. Cependant, les événements $(E_i)_{i=1}^6$ se chevauchent (ils ne sont pas disjoints), de sorte qu'on ne peut pas simplement additionner leurs probabilités. On utilise la propriété de σ -sous-additivité :

$$\mathbb{P}(B) = \mathbb{P}\left(\bigcup_{i=1}^6 E_i\right) \leq \sum_{i=1}^6 \mathbb{P}(E_i).$$

Or, pour chaque i , on a

$$\mathbb{P}(E_i) = \left(\frac{1}{2}\right)^5 = \frac{1}{32}.$$

Ainsi :

$$\mathbb{P}(B) \leq 6 \times \frac{1}{32} = \frac{6}{32} = 0,1875.$$

La probabilité d'observer au moins une suite de 5 piles consécutives en 10 lancers est donc inférieure ou égale à 0,1875.

Démonstration. Les preuves de la Propriété 2 et de la formule des probabilités totales sont élémentaires et constituent d'excellents exercices. La formule du crible se démontre par récurrence. La Propriété 5 découle de la correspondance entre unions disjointes et unions croissantes, et 6 s'en déduit par passage au complémentaire. Enfin, la σ -sous-additivité s'obtient en réécrivant une union comme une union disjointe, puis en appliquant le point 4. de la Propriété 2. \square

ii) Probabilité uniforme sur un ensemble fini

Commençons par un exemple très classique et élémentaire.

★ Exemple : lancer d'un dé

Considérons $\Omega = \{1, 2, 3, 4, 5, 6\}$ et $\mathcal{F} = \mathcal{P}(\Omega)$. Ici, le nombre d'événements est déjà important ($\text{card}(\mathcal{F}) = 2^6 = 64$) et il serait fastidieux de tous les décrire.

On procède donc autrement : en supposant que chaque face est équiprobable, la probabilité d'un événement $A \subset \Omega$ est

$$\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}.$$

Par exemple, si $A = \{2, 4, 6\}$ (événement « le résultat est pair »), on obtient

$$\mathbb{P}(A) = \frac{3}{6} = \frac{1}{6} + \frac{1}{6} + \frac{1}{6}.$$

On appelle cette loi la **probabilité uniforme** sur Ω .

La notion d'équiprobabilité rencontrée dans l'exemple du dé se généralise naturellement au cadre suivant.

Définition 4 (Probabilité uniforme ou équiprobabilité). Soit (Ω, \mathcal{F}) un espace probabilisable et $D \subset \Omega$ un ensemble fini. On appelle **probabilité uniforme sur D** la mesure de probabilité définie par

$$\mathbb{P}(A) = \frac{\text{card}(A \cap D)}{\text{card}(D)}, \quad \forall A \in \mathcal{F}.$$

Tous les événements élémentaires de D sont alors équiprobables.

Dans la plupart des cas, on prend $\Omega = D$: on parle alors simplement de probabilité uniforme. Quitte à restreindre \mathbb{P} à $\mathcal{P}(D)$, on peut toujours se ramener à cette situation. Cette généralisation est surtout d'ordre pratique et technique, car elle permet de traiter des cas plus complexes.

Pourquoi n'existe-t-il pas de probabilité uniforme sur un ensemble infini dénombrable ?

iii) Espace de probabilité discret

Encore une fois, commençons par traiter un exemple simple.

★ Exemple d'un univers infini dénombrable

On lance une pièce de monnaie biaisée jusqu'au moment où elle tombe sur pile. Une manière de modéliser cette expérience est de considérer $\Omega = B \sqcup \bigsqcup_{n=1}^{\infty} A_n$ où $B = \{b\}$ est l'ensemble constitué de la suite $b = (F, F, \dots)$ d'une infinité de faces et $A_n = \{a_n\}$ l'ensemble constitué du n -uplet $a_n = (F, \dots, F, P)$ comportant $n - 1$ faces et un pile en dernière position. Il y a donc une infinité dénombrable d'événements élémentaires.

On peut remarquer que si la tribu des événements \mathcal{F} les contient tous alors nécessairement $\mathcal{F} = \mathcal{P}(\Omega)$. Il nous reste donc à définir une probabilité sur \mathcal{F} .

Nous allons tout d'abord associer une probabilité à chaque événement élémentaire. Supposons que la pièce soit biaisée et tombe sur pile avec une probabilité $p \in]0, 1[$. Il est alors raisonnable de poser

$$\mathbb{P}(A_n) = \underbrace{(1-p) \cdots (1-p)}_{n-1} p = (1-p)^{n-1} p.$$

On obtient alors que

$$\mathbb{P}(B) = \mathbb{P} \left(\left(\bigsqcup_{n=0}^{\infty} A_n \right)^c \right) = 1 - \mathbb{P} \left(\bigsqcup_{n=0}^{\infty} A_n \right) = 1 - \sum_{n=1}^{\infty} (1-p)^{n-1} p = 0.$$

Il nous reste donc maintenant à associer une probabilité à l'infinité (non-dénombrable) des événements, pas seulement les événements élémentaires. Pour ce faire, on pose

$$\begin{aligned} \mathbb{P} : \mathcal{F} &\longrightarrow [0, 1] \\ A &\longmapsto \sum_{\omega \in A} \mathbb{P}(\{\omega\}) . \end{aligned}$$

Pour être clair, l'événement $\{\omega\}$ est un événement élémentaire, i.e. qu'il correspond à l'un des A_n ou à B , on connaît donc sa probabilité. Par ailleurs, l'événement A peut être constitué d'une infinité dénombrable d'issues possibles de l'expérience, la somme est donc à prendre au sens des familles sommables.

Détaillons un exemple. Soit C l'événement correspondant à faire pile pour la première fois après un nombre paire de lancers. On a $C = \{a_2, a_4, \dots\}$. Ainsi on obtient

$$\mathbb{P}(C) = \sum_{\omega \in C} \mathbb{P}(\{\omega\}) = \sum_{p=1}^{\infty} \mathbb{P}(\{a_{2p}\}) = \sum_{p=1}^{\infty} (1-p)^{2p-1} p = \frac{(1-p)p}{1-(1-p)^2}.$$

La deuxième égalité est directement issue de la théorie des familles sommables car on a choisit un ordre de sommation particulier.

Le fait que \mathbb{P} ainsi définie soit une probabilité est une conséquence directe du théorème de sommation par paquets.

La construction d'une mesure de probabilité vue dans cet exemple s'étend en réalité de manière beaucoup plus générale. On obtient alors la notion d'espace de probabilités discret.

Propriété 8 (Espace de probabilités discret). Soit (Ω, \mathcal{F}) un espace probabilisable et $D \subset \Omega$ une partie finie ou infinie dénombrable. Soit $(p_\omega)_{\omega \in D}$ une famille de réels positifs ou nuls telle que

$$\sum_{\omega \in D} p_\omega = 1.$$

Alors il existe une unique mesure de probabilité \mathbb{P} sur \mathcal{F} telle que, pour tout $\omega \in D$,

$$\mathbb{P}(\{\omega\}) = p_\omega.$$

Elle est donnée par la formule explicite :

$$\begin{aligned} \mathbb{P} : \mathcal{F} &\longrightarrow [0, 1] \\ A &\longmapsto \sum_{\omega \in A \cap D} p_\omega . \end{aligned}$$

Démonstration. Seule la σ -additivité mérite d'être vérifiée. Elle découle directement du théorème de sommation par paquets pour les familles sommables à termes positifs. \square



La famille $(p_\omega)_{\omega \in D}$ n'est pas une mesure de probabilité en elle-même. Elle ne fournit que les *poids* attribués aux événements élémentaires $\{\omega\}$. Autrement dit, $p_\omega \in [0, 1]$ représente la probabilité que le résultat de l'expérience soit exactement ω .

Une mesure de probabilité discrète peut toujours être définie sur la tribu des parties $\mathcal{F} = \mathcal{P}(\Omega)$. Lorsque $\Omega = D$ est fini ou dénombrable, on retrouve la construction des exemples précédents (pièce, dé, etc.). On peut donc toujours se ramener à cette situation, mais il est commode, pour des raisons pratiques et techniques, d'énoncer directement le résultat dans ce cadre plus général.

Définition 5 (Support d'une loi discrète). Soit \mathbb{P} une mesure de probabilité discrète. On appelle **support** de \mathbb{P} le plus petit événement D tel que $\mathbb{P}(D) = 1$. On le note $\text{supp}(\mathbb{P})$.

iv) Représentation par des mesures de Dirac

La construction des probabilités discrètes peut être exprimée de manière encore plus compacte à l'aide des *mesures de Dirac*.

Définition 6 (Mesure de Dirac). Soit (Ω, \mathcal{F}) un espace probablisable et $\omega \in \Omega$. On appelle **mesure de Dirac en ω** , notée δ_ω , la loi de probabilité définie par :

$$\begin{aligned} \delta_\omega : \mathcal{F} &\longrightarrow [0, 1] \\ A &\longmapsto \mathbb{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{sinon.} \end{cases} \end{aligned}$$



Attention à ne pas confondre mesure de Dirac et fonction indicatrice :

- δ_ω prend en argument un *événement* A et renvoie 1 si $\omega \in A$, 0 sinon ;
- $\mathbb{1}_A$ prend en argument une *issue* ω et renvoie 1 si $\omega \in A$, 0 sinon.

La mesure de Dirac δ_ω modélise le fait que, avec probabilité 1, l'expérience produit toujours l'issue ω . Son intérêt est fondamental : toute loi discrète peut être vue comme une combinaison convexe (finie ou dénombrable) de mesures de Dirac.

Plus généralement, une combinaison convexe de mesures de probabilités reste une mesure de probabilité : Soient (Ω, \mathcal{F}) un espace probabilisable, I un ensemble fini ou dénombrable, $(\mathbb{P}_i)_{i \in I}$ une famille de mesures de probabilités sur (Ω, \mathcal{F}) , et $(p_i)_{i \in I}$ une famille de poids positifs tels que

$$\sum_{i \in I} p_i = 1.$$

On définit

$$\mathbb{P}(A) = \sum_{i \in I} p_i \mathbb{P}_i(A), \quad \forall A \in \mathcal{F}.$$

Lemme 1 (Combinaisons convexes de mesures de probabilité). *Toute combinaison convexe comme en (1.2.3) définit encore une mesure de probabilité.*

Démonstration. Ce résultat découle de la théorie des familles sommables, notamment du théorème de sommation par paquet. \square

On en déduit que toute probabilité discrète peut s'exprimer comme une combinaison de mesures de Dirac.

Propriété 9 (Représentation discrète par Dirac). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité discret défini comme dans la Propriété 8. Alors*

$$\mathbb{P} = \sum_{\omega \in D} p_\omega \delta_\omega.$$

★ Exemple : lancer d'un dé équilibré et probabilité uniforme

Dans ce cas, la probabilité s'écrit

$$\mathbb{P} = \sum_{i=1}^6 \frac{1}{6} \delta_i.$$

Par exemple, la probabilité d'obtenir un multiple de 3 est

$$\mathbb{P}(\{3, 6\}) = \sum_{i=1}^6 \frac{1}{6} \delta_i(\{3, 6\}) = \sum_{i=1}^6 \frac{1}{6} \mathbb{1}_{\{3,6\}}(i) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6}.$$

Plus généralement, pour tout ensemble fini D , la probabilité uniforme sur D s'écrit

$$\mathbb{P} = \sum_{\omega \in D} \frac{1}{|D|} \delta_{\omega}.$$

★ Exemple : lancer d'une pièce équilibrée ou biaisée

Pour une seule pièce, on a $\mathbb{P} = p\delta_P + (1-p)\delta_F$. Si l'on s'intéresse au rang du premier succès (apparition de pile), on obtient la loi géométrique :

$$\mathbb{P} = \sum_{n=1}^{\infty} (1-p)^{n-1} p \delta_n.$$

Vocabulaire et notations complémentaires

Nous introduisons ici quelques notations et termes supplémentaires, utiles pour manipuler les probabilités. Étant donnés deux événements A et B , on notera indifféremment

$$\mathbb{P}(A \cap B) \equiv \mathbb{P}(A, B) \equiv \mathbb{P}(A \text{ et } B), \quad \mathbb{P}(A \cup B) \equiv \mathbb{P}(A \text{ ou } B).$$

De plus, lorsqu'un événement A est tel que $\mathbb{P}(A) = 1$, on dira que A a lieu **presque sûrement** (p.s.), ou encore qu'il est **de probabilité totale** ou que c'est un événement **presque certain** ou **presque sûr**. De même, si $\mathbb{P}(A) = 0$, on parle d'un événement **presque impossible** ou **négligeable**. Ces notions sont résumées dans la Figure 1.3.

Notations	Vocabulaire ensembliste	Vocabulaire probabiliste
$\mathbb{P}(A) = 0$	ensemble négligeable	événement négligeable
$\mathbb{P}(A) = 1$	ensemble de mesure totale	événement presque sûr

FIGURE 1.3 – Vocabulaire probabiliste (suite).

Propriété 10 (Manipulation des événements négligeables ou presque sûrs). Soient A, B deux événements. Alors :

$$\mathbb{P}(B) = \mathbb{P}(B \cap A) \quad \text{si } \mathbb{P}(A) = 1, \quad \mathbb{P}(B \cup A) = \mathbb{P}(B) \quad \text{si } \mathbb{P}(A) = 0.$$

De plus, une réunion dénombrable d'événements négligeables est encore négligeable, et une intersection dénombrable d'événements presque sûrs est encore un événement presque sûr. En particulier, si pour tout $n \geq 0$, $\mathbb{P}(A_n) = 0$ et $\mathbb{P}(B_n) = 1$, alors

$$\mathbb{P}\left(\bigcup_{n \geq 0} A_n\right) = 0, \quad \mathbb{P}\left(\bigcap_{n \geq 0} B_n\right) = 1.$$

Démonstration. Ces égalités se justifient rapidement à partir des propriétés élémentaires des probabilités et de la σ -sous-additivité. \square



Il faut distinguer un événement négligeable de l'événement vide \emptyset , et un événement presque sûr de l'événement certain Ω . En effet, on peut avoir $\mathbb{P}(A) = 0$ sans que $A = \emptyset$, et $\mathbb{P}(A) = 1$ sans que $A = \Omega$.

Ces notions sont essentielles : dans de nombreux contextes, des événements de probabilité nulle surviennent malgré tout. Par exemple, la probabilité que vous dormiez *exactement* le même temps (avec une précision infinie) que la nuit précédente est nulle, et pourtant cela s'est déjà produit. De même, la probabilité de lancer une fléchette *exactement* au même endroit que le tir précédent est nulle, mais pas impossible.

Un exemple plus concret provient du lancer d'une infinité de pièces. On avait introduit l'événement B : « on ne fait jamais pile », et les événements A_n : « pile apparaît pour la première fois au n -ème lancer ». On obtenait

$$\mathbb{P}(B) = 0 \quad \text{avec } B \neq \emptyset, \quad \mathbb{P}\left(\bigsqcup_{n \geq 1} A_n\right) = 1 \quad \text{avec } \bigsqcup_{n \geq 1} A_n \neq \Omega.$$

Ici, l'événement B n'est pas vide, mais reste de probabilité nulle. Ce phénomène devient inévitable dès que l'on manipule des univers infinis non dénombrables ou lorsque l'on fait du conditionnement.

Par exemple, nous avons déjà rencontré une première version de la formule des probabilités totales dans le cas d'une partition exacte d'événements. Dans de nombreuses situations, il est utile de disposer d'une extension plus souple, lorsque les événements forment seulement une partition « au sens probabiliste ».

Définition 7 (Système complet d'événements). Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. On dit qu'une famille dénombrable d'événements $(A_n)_{n \geq 1}$ forme une **partition au sens probabiliste**, ou encore un **système complet d'événements**, si

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = 1, \quad \forall i \neq j, \mathbb{P}(A_i \cap A_j) = 0.$$

Autrement dit, les événements (A_n) recouvrent presque sûrement l'univers et sont « presque » disjoints deux à deux (ils peuvent éventuellement s'intersecter, mais seulement sur un ensemble de probabilité nulle). Une telle notion permet de formuler une version généralisée de la formule des probabilités totales.

Propriété 11 (Formule des probabilités totales 2). Soit $(A_n)_{n \geq 0} \subset \mathcal{F}$ un système complet d'événements. Alors, pour tout événement $B \in \mathcal{F}$,

$$\mathbb{P}(B) = \sum_{n=0}^{\infty} \mathbb{P}(B \cap A_n).$$

Propriété 12 (Formule des probabilités totales 2). Soit $(A_n)_{n \geq 0} \subset \mathcal{F}$ un système complet d'événements. Alors, pour tout événement $B \in \mathcal{F}$,

$$\mathbb{P}(B) = \sum_{n=0}^{\infty} \mathbb{P}(B \cap A_n).$$

Démonstration. Comme $(A_n)_{n \geq 0}$ est un système complet, leur réunion recouvre presque sûrement tout Ω et ils sont deux à deux incompatibles à une probabilité nulle près. Pour formaliser cela, posons

$$K = \left(\bigcup_{n \geq 0} A_n\right)^c \quad \text{et} \quad D = \bigcup_{i < j} (A_i \cap A_j).$$

Ici, K représente l'ensemble des issues non couvertes par la réunion des A_n , et D les recouvrements éventuels entre deux A_i . Par définition d'un système complet, on a donc $\mathbb{P}(K) = 0$ et $\mathbb{P}(D) = 0$.

On peut alors construire une véritable partition de l'univers en considérant

$$\{K\} \sqcup \{A_n \cap D^c : n \geq 0\} \sqcup \{D\}.$$

Cette famille est bien une partition, et la première formule des probabilités totales s'y applique. On obtient

$$\mathbb{P}(B) = \mathbb{P}(B \cap K) + \mathbb{P}(B \cap D) + \sum_{n=0}^{\infty} \mathbb{P}(B \cap A_n \cap D^c).$$

Or, comme K et D sont négligeables, $\mathbb{P}(B \cap K) = 0$ et $\mathbb{P}(B \cap D) = 0$. De plus, l'intersection avec D^c ne change rien pour les A_n , car leurs recouvrements ont probabilité nulle. Ainsi, $\mathbb{P}(B \cap A_n \cap D^c) = \mathbb{P}(B \cap A_n)$ pour tout n . En reportant ces simplifications, on obtient bien

$$\mathbb{P}(B) = \sum_{n=0}^{\infty} \mathbb{P}(B \cap A_n),$$

ce qui conclut la démonstration. □

Indépendance et probabilités conditionnelles

La notion d'indépendance en probabilité est assez naturelle et intuitive. Elle a d'ailleurs déjà été évoquée implicitement dans ce polycopié. Dans la plupart des situations simples, on fait cette hypothèse sans même s'en rendre compte. Chacun comprend vaguement ce que signifie l'indépendance pour deux événements A et B : il s'agirait simplement de dire que la réalisation de A n'influe pas sur celle de B , et réciproquement.

Cependant, cette intuition n'est pas entièrement satisfaisante. Par exemple, si l'on lance deux dés avec la même main, est-il vraiment raisonnable de dire que leurs résultats sont indépendants ? Et qu'en est-il si une même personne lance deux fois le même dé ?

Pour s'en convaincre davantage, considérons l'expérience suivante. Vous demandez à votre meilleur ami de produire deux tirages de pile ou face indépendants. Celui-ci, un peu espiègle, applique sans vous le dire l'algorithme suivant :

- il lance simultanément une pièce avec sa main gauche, et deux autres pièces (numérotées 1 et 2) avec sa main droite ;
- il choisit comme premier résultat celui de la pièce lancée par sa main gauche ;
- pour le second résultat, il choisit celui de la pièce 1 si la première est tombée sur pile, ou celui de la pièce 2 si la première est tombée sur face.

Considérons les événements $A =$ « le premier résultat donné par votre ami est pile » et $B =$ « le second résultat donné par votre ami est face ». À première vue, on ne peut pas dire que A n'influence pas B . En effet, supposons que les résultats des deux pièces lancées par la main droite soient (P, F) . Alors :

- si A est réalisé (la main gauche donne pile), votre ami choisit la pièce 1, donc B n'est pas réalisé ;
- si A n'est pas réalisé (la main gauche donne face), votre ami choisit la pièce 2, et cette fois B est réalisé.

Pourtant, en termes de probabilités, il n'y a *aucune influence*. De votre point de vue, les deux résultats donnés par votre ami sont équiprobables et satisfont donc

$$\mathbb{P}(\{(P, P)\}) = \mathbb{P}(\{(P, F)\}) = \mathbb{P}(\{(F, P)\}) = \mathbb{P}(\{(F, F)\}) = \frac{1}{4}.$$

On peut aussi vérifier que cela revient à écrire

$$\mathbb{P}(A) = \frac{1}{2}, \quad \mathbb{P}(B) = \frac{1}{2}, \quad \text{et} \quad \mathbb{P}(A \cap B) = \frac{1}{4}.$$

Ces relations sont exactement celles qui définissent l'indépendance. Votre ami ne vous a donc pas piégé : il vous a bien fourni deux tirages indépendants au sens probabiliste. Si l'expérience était répétée un grand nombre de fois, vous constateriez que la proportion de cas où B est réalisé parmi ceux où A l'est est d'environ 50%, tout comme dans les cas où A ne l'est pas.

2.1 Indépendance d'événements et de tribus

Commençons par le cas simple de deux événements, avant de généraliser la notion à des familles plus larges.

2.1.1 Indépendance de deux événements

Un ingrédient essentiel manquait dans notre discussion heuristique précédente : la *mesure de probabilité*. En effet, ce qui caractérise réellement l'indépendance est le fait que

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \quad (1)$$

Pour s'en convaincre, revenons à une interprétation concrète en termes de proportions. Imaginons que l'on s'intéresse à la proportion de consommateurs réguliers de chocolat chaud le matin, et à la proportion de joueurs de Zelda. À priori, les événements « consommer du chocolat chaud » et « jouer à Zelda » n'ont rien à voir entre eux. Notons :

- Ω la population considérée,
- A l'ensemble des personnes qui consomment du chocolat chaud le matin,
- B l'ensemble des personnes qui jouent à Zelda.

Intuitivement, si ces deux événements sont indépendants, alors :

- la proportion de buveurs de chocolat chaud parmi les joueurs de Zelda doit être la même que dans la population globale,
- de même, la proportion de joueurs de Zelda parmi les buveurs de chocolat chaud doit coïncider avec celle observée dans l'ensemble de la population.

Cela se traduit par

$$\frac{\text{card}(A \cap B)}{\text{card}(B)} = \frac{\text{card}(A)}{\text{card}(\Omega)}, \quad \frac{\text{card}(A \cap B)}{\text{card}(A)} = \frac{\text{card}(B)}{\text{card}(\Omega)}. \quad (2)$$

Ces relations sont équivalentes à

$$\frac{\text{card}(A \cap B)}{\text{card}(\Omega)} = \frac{\text{card}(A)}{\text{card}(\Omega)} \cdot \frac{\text{card}(B)}{\text{card}(\Omega)}.$$

Autrement dit, la proportion de personnes qui *à la fois* consomment du chocolat chaud et jouent à Zelda est exactement le produit des proportions de chacune des deux caractéristiques. La Figure 2.1 illustre un tel exemple numérique : $\text{card}(A) = 16$, $\text{card}(B) = 18$, $\text{card}(A \cap B) = 4$, $\text{card}(\Omega) = 72$. La proportion de B dans A est $4/16 = 1/4$, identique à la proportion de B dans Ω , qui vaut $18/72 = 1/4$.

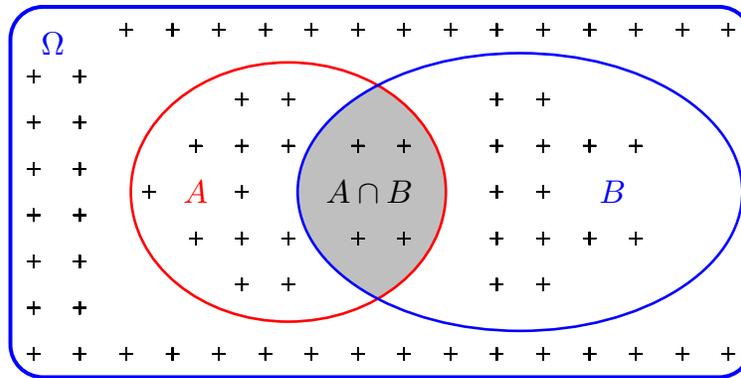


FIGURE 2.1 – Illustration d'événements indépendants.

En termes de probabilités, si l'on choisit uniformément une personne au hasard dans la population, les événements A et B sont indépendants au sens de la formule (1).

Définition 8 (Indépendance de deux événements). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $A, B \in \mathcal{F}$ deux événements. On dit que A et B sont **indépendants** lorsque*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \quad (3)$$

On remarque aussitôt que tout événement de probabilité totale ($\mathbb{P}(A) = 1$) ou négligeable ($\mathbb{P}(A) = 0$) est indépendant de n'importe quel autre.

Propriété 13 (Indépendance et complémentaires). *Si A et B sont deux événements indépendants, alors les paires (A, B^c) , (A^c, B) et (A^c, B^c) sont également indépendantes.*

Démonstration. On écrit simplement

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

et on utilise l'égalité (3). □

Cette propriété est cohérente avec l'intuition : si le fait que A se réalise ne change pas la probabilité de B , alors le fait que A ne se réalise pas ne la change pas davantage.

★ Exemple : loi zêta et événements arithmétiques

On considère la loi de probabilité \mathbb{P} sur $\Omega = \mathbb{N}^*$ définie par

$$\mathbb{P}(\{n\}) = \frac{1}{\zeta(2)} \frac{1}{n^2}, \quad \text{avec} \quad \zeta(2) = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

Soit $A = \ll \text{l'entier est pair} \gg$ et $B = \ll \text{l'entier est divisible par 7} \gg$. On a

$$\mathbb{P}(A) = \frac{1}{\zeta(2)} \sum_{k=1}^{\infty} \frac{1}{(2k)^2} = \frac{1}{4}, \quad \mathbb{P}(B) = \frac{1}{\zeta(2)} \sum_{k=1}^{\infty} \frac{1}{(7k)^2} = \frac{1}{49}.$$

De même, comme 2 et 7 sont premiers entre eux, on a

$$\mathbb{P}(A \cap B) = \frac{1}{\zeta(2)} \sum_{k=1}^{\infty} \frac{1}{(14k)^2} = \frac{1}{196}.$$

On vérifie donc que $\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B)$. Ainsi, sous la loi zêta de paramètre 2, les événements « être pair » et « être divisible par 7 » sont indépendants.

2.1.2 Indépendance de deux tribus

Jusqu'ici, nous avons étudié l'indépendance de deux événements. Il est toutefois souvent nécessaire de raisonner sur des collections entières d'événements : par exemple, lorsqu'on considère l'évolution d'un processus aléatoire dans le temps, ou bien lorsqu'on compare deux systèmes de variables aléatoires.

Définition 9 (Indépendance de deux tribus d'événements). Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$ deux sous-tribus. On dit que \mathcal{A} et \mathcal{B} sont **indépendantes** si, pour tout $A \in \mathcal{A}$ et tout $B \in \mathcal{B}$, on a

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

Autrement dit, l'indépendance d'événements se propage naturellement à l'indépendance des familles d'événements générées par ces événements. Par exemple, à partir d'un événement A , on peut former la plus petite tribu qui le contient, notée $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$, et de même pour un événement B : $\sigma(B) = \{\emptyset, B, B^c, \Omega\}$.

On obtient ainsi le résultat suivant, qui relie les deux notions.

Propriété 14. Deux événements A et B sont indépendants si et seulement si les tribus $\sigma(A)$ et $\sigma(B)$ sont indépendantes.

2.1.3 Indépendance d'une famille finie d'événements



Lorsqu'on cherche à généraliser la notion d'indépendance à plusieurs événements, une erreur fréquente consiste à exiger seulement l'indépendance deux à deux. L'exemple suivant montre que ce n'est pas suffisant.

★ Exemple : deux lancers de pièce

On lance deux fois une pièce équilibrée et on considère les événements :

- A : « pile au premier lancer »,
- B : « face au second lancer »,
- C : « les deux lancers donnent le même résultat ».

Les événements A et B sont indépendants, de même que A et C , ainsi que B et C .

En effet

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}, \quad \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{4}.$$

Cependant, les trois événements ne sont pas mutuellement indépendants, car

$$\mathbb{P}(A \cap B \cap C) = 0 \neq \frac{1}{8} = \mathbb{P}(A) \mathbb{P}(B) \mathbb{P}(C).$$



Une autre erreur consiste à demander uniquement que la probabilité de l'intersection de tous les événements soit égale au produit des probabilités individuelles. L'exemple suivant montre que ce n'est pas suffisant non plus.

★ Exemple : lancer d'un dé

On lance un dé équilibré et on considère :

- A : « le résultat est un multiple de 2 »,
- B : « le résultat est supérieur ou égal à 4 »,
- C : « le résultat n'est pas un multiple de 3 ».

On a alors

$$\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}, \quad \mathbb{P}(C) = \frac{2}{3},$$

et

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \mathbb{P}(B) \mathbb{P}(C) = \frac{1}{6}.$$

Pourtant, A et B ne sont pas indépendants deux à deux, car

$$\mathbb{P}(A \cap B) = \frac{2}{6} \neq \frac{1}{4} = \mathbb{P}(A) \mathbb{P}(B).$$

Ces exemples montrent que la bonne généralisation est la suivante.

Définition 10 (Indépendance mutuelle). Une famille finie d'événements A_1, \dots, A_n est dite **mutuellement indépendante** (ou parfois simplement indépendante) si, pour tout sous-ensemble d'indices $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$, on a

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \prod_{j=1}^k \mathbb{P}(A_{i_j}). \quad (4)$$

Autrement dit, l'égalité (4) doit être vérifiée pour toutes les intersections non triviales, et pas seulement pour les paires ou l'intersection complète.

On dit alors que les événements A_1, \dots, A_n sont **mutuellement indépendants**, par opposition à l'indépendance seulement deux à deux.

Définition 11 (Indépendance mutuelle de tribus). Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $(\mathcal{A}_i)_{i \in I}$ une famille finie de sous-tribus. On dit que ces tribus sont **indépendantes** si, pour tout choix d'événements $A_i \in \mathcal{A}_i$ ($i \in I$), on a

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i). \quad (5)$$

On pourrait craindre que cette définition soit mal posée puisqu'elle ne demande pas explicitement l'égalité (5) pour toute sous-famille. Mais ce n'est pas nécessaire : chaque tribu contient Ω , ce qui permet de compléter une intersection. Par exemple,

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

Ainsi, l'indépendance de tribus se formule plus simplement que celle d'événements.

Propriété 15. Une famille finie d'événements $(A_i)_{i \in I}$ est mutuellement indépendante si et seulement si les tribus $(\sigma(A_i))_{i \in I}$ sont indépendantes.

Démonstration. Par récurrence sur le cardinal de I . □

2.1.4 Indépendance d'une famille quelconque d'événements

Jusqu'ici, nous avons défini l'indépendance pour une famille finie d'événements. Il est naturel de vouloir l'étendre à une famille arbitraire, éventuellement infinie. L'idée clé est que l'indépendance ne se teste qu'à travers les sous-familles finies.

Définition 12. Soit $(A_i)_{i \in I}$ une famille (finie ou infinie dénombrable) d'événements. On dit que cette famille est **indépendante** si toute sous-famille finie l'est. De même, une famille $(\mathcal{A}_i)_{i \in I}$ de sous-tribus d'événements est dite **indépendante** si toute sous-famille finie l'est.

Ainsi, l'indépendance d'une famille quelconque d'événements est toujours équivalente à l'indépendance des tribus engendrées par ces événements.

2.2 Probabilités conditionnelles

De nombreuses déclarations sur le hasard prennent la forme « si A se produit, alors la probabilité de B est p », où A et B sont des événements. Nous avons vu que, si A et B sont indépendants, la probabilité de B reste égale à p que A soit réalisé ou non. Or l'indépendance est une hypothèse forte, rarement vérifiée en pratique.

Pour introduire la probabilité conditionnelle, repartons de l'angle *proportions* et de la discussion sur l'indépendance (Figure 2.1). Considérons cette fois deux événements clairement dépendants : « mesurer plus de 2 mètres » et « jouer au basket-ball ». Notons A l'ensemble des individus mesurant plus de 2 mètres et B l'ensemble des basketteurs dans une population Ω . On s'attend à ce que la proportion d'individus très grands parmi les basketteurs soit plus élevée que dans la population globale. Autrement dit, contrairement à (2), on a l'inégalité

$$\frac{\text{card}(A \cap B)}{\text{card}(B)} \neq \frac{\text{card}(A)}{\text{card}(\Omega)}.$$

Interprétons les deux termes. Celui de droite est la proportion de grands dans toute la population (i.e. $\mathbb{P}(A)$); celui de gauche est la proportion de grands parmi les basketteurs, que l'on appelle **probabilité de A sachant B** et que l'on note $\mathbb{P}(A|B)$. Dire que A et B ne sont pas indépendants revient donc à écrire

$$\mathbb{P}(A|B) \neq \mathbb{P}(A),$$

ou, de manière équivalente, $\mathbb{P}(B|A) \neq \mathbb{P}(B)$. Par exemple, sur la Figure 2.2 (variation numérique de la précédente), on lit

$$\mathbb{P}(A|B) = \frac{4}{18} = \frac{2}{9} \quad \text{et} \quad \mathbb{P}(A) = \frac{16}{60} = \frac{4}{15}.$$

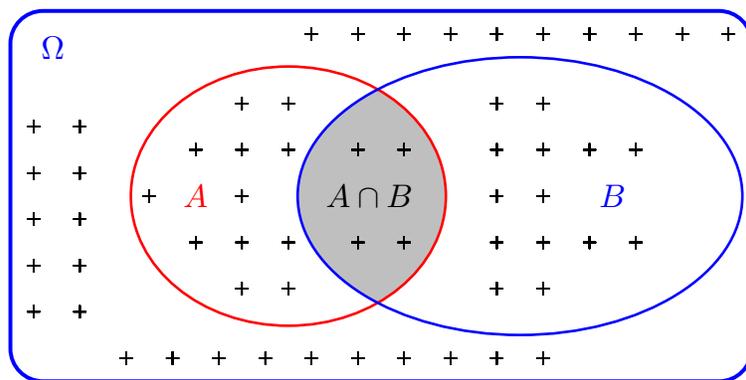


FIGURE 2.2 – Probabilités conditionnelles.

Comment définir cette probabilité conditionnelle de manière générale? Observons tout d'abord que

$$\frac{\text{card}(A \cap B)}{\text{card}(B)} = \frac{\frac{\text{card}(A \cap B)}{\text{card}(\Omega)}}{\frac{\text{card}(B)}{\text{card}(\Omega)}}.$$

On obtient ainsi la définition standard suivante.

Définition 13 (Probabilité conditionnelle). Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $B \in \mathcal{F}$ tel que $\mathbb{P}(B) > 0$. La probabilité de A **sachant** B est définie par

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

Par convention (facultative), si $\mathbb{P}(B) = 0$, on pose parfois $\mathbb{P}(A|B) = \mathbb{P}(A)$. On dira pour abréger « probabilité de A sachant B » ou « probabilité de A conditionnellement à B ». D'autres notations existent, par exemple $\mathbb{P}_B(A)$.

Propriété 16. Si A et B sont indépendants, alors

$$\mathbb{P}(B|A) = \mathbb{P}(B) \quad \text{si} \quad \mathbb{P}(A) > 0.$$

Démonstration. Immédiat, car $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. □

Les probabilités conditionnelles donnent des formulations commodes des *probabilités totales*.

Propriété 17 (Formule des probabilités totales 3). Soit $(A_n)_{n \geq 0}$ un système complet d'événements. Pour tout événement B ,

$$\mathbb{P}(B) = \sum_{n=0}^{\infty} \mathbb{P}(A_n) \mathbb{P}(B|A_n).$$

Démonstration. Conséquence directe de la Propriété 12. □

Cette formule est souvent présentée pour une partition finie et visualisée via un **arbre de probabilités** : les nœuds représentent des événements, les arêtes portent les probabilités conditionnelles (Figure 2.3, partition en A et A^c).

En général, au bout d'une branche de longueur n , on lit la probabilité d'une intersection de n événements A_1, \dots, A_n . À l'étape k de la branche, on lit $\mathbb{P}(A_k | A_1 \cap \dots \cap A_{k-1})$.

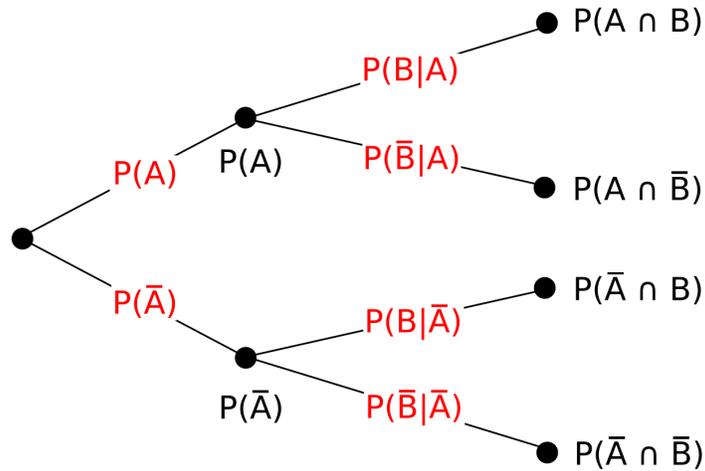


FIGURE 2.3 – Arbre de probabilités.

★ Exemple : probabilité d'un échec dans une urne choisie au hasard

On dispose d'une infinité d'urnes $(U_n)_{n \geq 1}$. Dans l'urne U_n , on place une boule rouge et n boules noires, soit un total de $n + 1$ boules. On choisit d'abord une urne U_n au hasard, avec la probabilité

$$\mathbb{P}(A_n) = \frac{1}{\zeta(2)} \frac{1}{n^2}, \quad n \geq 1,$$

où $\zeta(2) = \pi^2/6$. Puis on tire une boule au hasard dans l'urne choisie. Ici on note $A_n =$ « on a choisi l'urne U_n ». Soit $B =$ « la boule tirée est noire ». Les événements $(A_n)_{n \geq 1}$ forment une partition dénombrable de Ω .

Par la formule des probabilités totales,

$$\mathbb{P}(B) = \sum_{n=1}^{\infty} \mathbb{P}(A_n) \mathbb{P}(B|A_n).$$

Or, $\mathbb{P}(B|A_n) = \frac{n}{n+1}$. Donc

$$\mathbb{P}(B) = \sum_{n=1}^{\infty} \frac{1}{\zeta(2)} \cdot \frac{1}{n^2} \cdot \frac{n}{n+1} = \frac{1}{\zeta(2)} \sum_{n=1}^{\infty} \frac{1}{n(n+1)}.$$

La série est télescopique :

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \sum_{n=1}^{\infty} \left(\frac{1}{n} - \frac{1}{n+1} \right) = 1.$$

Ainsi

$$\mathbb{P}(B) = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}.$$

Propriété 18 (Formule des probabilités composées). Pour des événements A_1, \dots, A_n ,

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \mathbb{P}(A_2|A_1) \mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}).$$

★ Exemple : tirer trois as sans remise

On tire successivement trois cartes d'un jeu de 52, sans remise. On veut calculer la probabilité que les trois cartes soient des as. Soient :

— A_1 : « la première carte est un as »,

— A_2 : « la deuxième carte est un as »,

— A_3 : « la troisième carte est un as ».

D'après la formule des probabilités composées :

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1) \mathbb{P}(A_2|A_1) \mathbb{P}(A_3|A_1 \cap A_2).$$

On calcule :

$$\mathbb{P}(A_1) = \frac{4}{52}, \quad \mathbb{P}(A_2|A_1) = \frac{3}{51}, \quad \mathbb{P}(A_3|A_1 \cap A_2) = \frac{2}{50}.$$

Ainsi :

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} = \frac{1}{5525}.$$

La probabilité de tirer trois as consécutivement sans remise est donc environ 0,018%.

Passons à la **formule de Bayes**, simple réécriture de la définition combinée aux probabilités totales. Pour A, B de probabilité strictement positive,

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

On a utilisé $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A) = \mathbb{P}(B)\mathbb{P}(A|B)$ et $\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c)$. Plus généralement, si $(A_i)_{i \in I}$ est un système complet, on a la propriété suivante.

Propriété 19 (Formule de Bayes généralisée). Soit $(A_i)_{i \in I}$ une partition probabiliste et B un événement. Pour tout $i_0 \in I$,

$$\mathbb{P}(A_{i_0}|B) = \frac{\mathbb{P}(A_{i_0}) \mathbb{P}(B|A_{i_0})}{\sum_{i \in I} \mathbb{P}(A_i) \mathbb{P}(B|A_i)}.$$

★ Exemple : test de dépistage et paradoxe des faux positifs

Un reportage vante les mérites d'un nouveau test RT-PCR. On apprend que sa **sensibilité** est de 80% : sur 100 personnes infectées, 80 sont détectées positives (et 20 sont des faux négatifs). Sa **spécificité** est de 98% : sur 100 personnes non infectées, 98 sont correctement négatives (et 2 sont des faux positifs).

Par ailleurs, le taux de létalité du virus est estimé à 0.7%. Comme 84 décès ont été constatés en 24 heures dans un pays de 60 millions d'habitants, on en déduit qu'il y avait environ

$$\frac{84}{0.007} \simeq 12\,000$$

personnes infectées un mois auparavant.

Une campagne de dépistage massive est alors menée : sur 1 million de tests aléatoires, 20 136 reviennent positifs, soit environ 2%. À première vue, on pourrait conclure qu'il y a 1.2 million de personnes infectées dans le pays. Faut-il pour autant céder à la panique générale ?

★ Exemple : un paradoxe familial

Cet exemple volontairement déstabilisant illustre que la formule de Bayes n'est pas toujours intuitive. Une camarade vous confie que ses parents ont eu deux enfants (elle comprise). Quelle est la probabilité qu'elle ait une sœur plutôt qu'un frère ?

La première réaction, même d'un professeur de mathématiques, est de répondre 1/2. Après tout, chaque naissance est une chance sur deux d'avoir une fille ou un garçon. Pourtant, le calcul montre que ce n'est pas la bonne réponse.

Modélisons soigneusement. L'univers des possibles est

$$\Omega = \{FF, FG, GF, GG\},$$

où la première lettre désigne le sexe du premier enfant, la seconde celui du second (F = fille, G = garçon). La probabilité est uniforme :

$$\mathbb{P}(\{FF\}) = \mathbb{P}(\{FG\}) = \mathbb{P}(\{GF\}) = \mathbb{P}(\{GG\}) = \frac{1}{4}.$$

L'information dont nous disposons est que notre amie est l'un des deux enfants, et que c'est une fille. Cela revient à savoir qu'il y a au moins une fille dans la fratrie, c'est-à-dire

$$B = \{FF, FG, GF\}.$$

La question devient : quelle est la probabilité que l'autre enfant soit aussi une fille, i.e.

$$A = \{FF\}.$$

On calcule alors

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

Ainsi, la probabilité qu'elle ait une sœur est de $1/3$, et non $1/2$. Mais où est l'erreur avec l'argument du $1/2$?

Tout est une question d'information. Si votre amie avait précisé : « je suis l'aînée », alors la probabilité serait bien $1/2$, car

$$\mathbb{P}(\{FF\}|\{FF, GF\}) = \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2}.$$

L'ambiguïté vient donc de la condition initiale : « il y a au moins une fille » n'est pas la même information que « je suis l'aînée et je suis une fille ».

Pour s'en convaincre expérimentalement, on peut simuler : prenez un jeu de cartes, formez des paires au hasard et conservez seulement celles contenant au moins une carte rouge. Environ un tiers de ces paires contiendront deux rouges, confirmant le résultat.

★ *Le « paradoxe de Simpson »* ©David Louapre

Pas de chance : on vient de vous découvrir des calculs rénaux. Deux traitements existent : le « Traitement A », une chirurgie ouverte, et le « Traitement B », une chirurgie par petites incisions. Vous demandez au médecin lequel choisir. Il vous montre les statistiques globales : sur 350 patients pour chaque traitement, 273 ont guéri avec A et 289 avec B. « Facile, conclut le médecin : 79% de réussite pour A, contre 83% pour B. Choisissez B. »

En sortant, vous interrogez un autre praticien, qui précise que les patients étaient de deux types : porteurs de « petits » calculs et de « gros » calculs. Voici son tableau :

	« petits » calculs		« gros » calculs	
	A	B	A	B
Réussites	81 (93%)	234 (87%)	192 (73%)	55 (69%)
Échecs	6	36	71	25

Surprise : quel que soit le type de calculs, A est meilleur que B. Et pourtant, globalement, c'est B qui semblait meilleur ! Comment est-ce possible ?

Traduction probabiliste :

1. PC : « le patient a un petit calcul »,
2. $GC = PC^c$: « le patient a un gros calcul »,
3. A : « traitement A »,
4. $B = A^c$: « traitement B »,
5. S : « succès du traitement ».

On observe

$$\mathbb{P}(S|A \cap PC) > \mathbb{P}(S|B \cap PC), \quad \mathbb{P}(S|A \cap GC) > \mathbb{P}(S|B \cap GC),$$

mais malgré tout

$$\mathbb{P}(S|B) > \mathbb{P}(S|A).$$

Où est le « paradoxe » ? En réalité, il n'y a pas contradiction :

$$\begin{aligned} \mathbb{P}(S|B) &= \mathbb{P}(GC|B) \mathbb{P}(S|B \cap GC) + \mathbb{P}(PC|B) \mathbb{P}(S|B \cap PC), \\ \mathbb{P}(S|A) &= \mathbb{P}(GC|A) \mathbb{P}(S|A \cap GC) + \mathbb{P}(PC|A) \mathbb{P}(S|A \cap PC). \end{aligned}$$

Tout dépend donc de la répartition des cas difficiles : si A est utilisé plus souvent pour les gros calculs, sa performance globale paraît inférieure alors qu'il est supérieur dans chaque sous-groupe. C'est l'illustration typique du **paradoxe de Simpson**.

« Il y a trois sortes de mensonges : les petits mensonges, les gros mensonges, et les statistiques. » — Mark Twain

Les variables aléatoires et leurs lois

Certaines quantités dépendent de phénomènes aléatoires. Par exemple : le gain à un jeu de hasard, l'évolution du cours de la bourse, ou encore le nombre de voitures rouges observées à un carrefour entre 17h et 19h. Dans ces situations, ce n'est pas tant le phénomène aléatoire lui-même qui nous intéresse (le trajet de la bille de roulette, le passage de chaque voiture), mais plutôt une *conséquence* de ce hasard (le gain, le nombre de voitures rouges). On appelle de telles fonctions des **variables aléatoires**.

À toute variable aléatoire, on peut associer une probabilité qui décrit la façon dont sont distribuées ses valeurs : c'est sa **loi** ou sa **distribution**. Nous distinguerons principalement deux classes : les variables aléatoires **discrètes**, qui prennent au plus un nombre dénombrable de valeurs avec probabilité 1, et les variables aléatoires **continues**, que nous évoquerons plus brièvement.

Commençons par un exemple introductif. On vous propose le jeu suivant : une pièce équilibrée est lancée deux fois. Si les deux lancers donnent pile, vous gagnez 2 euros ; si un seul lancer donne pile, vous gagnez 1 euro ; si aucun ne donne pile, vous ne gagnez rien. Une manière de modéliser cette expérience est de prendre

$$\Omega = \{P, F\} \times \{P, F\} = \{(P, P), (P, F), (F, P), (F, F)\}, \quad \mathcal{F} = \mathcal{P}(\Omega),$$

et \mathbb{P} la probabilité uniforme sur Ω . Le gain obtenu est alors une fonction $X : \Omega \rightarrow \{0, 1, 2\}$ définie par

$$X((P, P)) = 2, \quad X((P, F)) = X((F, P)) = 1, \quad X((F, F)) = 0.$$

Comment alors décrire la distribution de ces gains ? Ici, l'exemple est suffisamment simple pour donner la réponse intuitivement : on a

$$\mathbb{P}(X = 1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2},$$

car X prend la valeur 1 uniquement pour (P, F) et (F, P) , chacune ayant une probabilité $1/4$. Mais, rigoureusement, que signifie la notation $\mathbb{P}(X = 1)$? La condition « $X = 1$ » doit

correspondre à un événement de Ω . Plus précisément, « $X = 1$ » représente l'ensemble des issues pour lesquelles X vaut 1, c'est à dire $\{(P, F), (F, P)\}$. Autrement dit, « $X = 1$ » est l'**image réciproque** de $\{1\}$ par X , notée $X^{-1}(\{1\})$.

En général, si $f : E \rightarrow F$ est une fonction et $A \subset F$, on définit

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

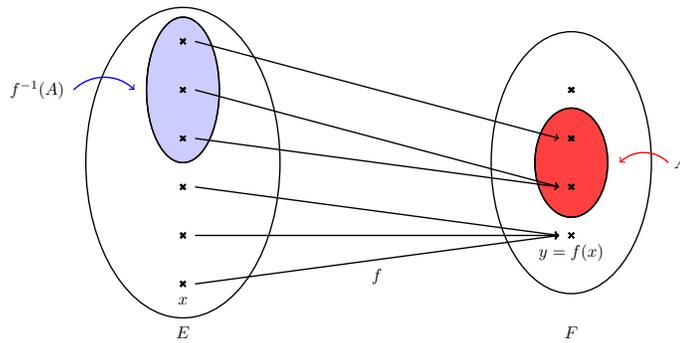


FIGURE 3.1 – Illustration de l'image réciproque.

De plus, l'image réciproque est compatible avec les opérations ensemblistes usuelles. On a notamment, pour toute famille $(A_i)_{i \in I}$ d'ensembles de F :

- $f^{-1}(\bigcap_{i \in I} A_i) = \bigcap_{i \in I} f^{-1}(A_i)$,
- $f^{-1}(\bigcup_{i \in I} A_i) = \bigcup_{i \in I} f^{-1}(A_i)$,
- $f^{-1}(A^c) = (f^{-1}(A))^c$.

3.1 Définitions générales

Nous introduisons maintenant la notion générale de variable aléatoire, sa loi et l'indépendance dans un cadre abstrait.

3.1.1 Définition d'une variable aléatoire

Définition 14 (Variable aléatoire générale). Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et (E, \mathcal{A}) un espace probabilisable. Une fonction $X : \Omega \rightarrow E$ est appelée **variable aléatoire** à valeurs dans E lorsque

$$\forall A \in \mathcal{A}, \quad X^{-1}(A) \in \mathcal{F}.$$

Autrement dit, l'image réciproque par X de tout événement de l'espace d'arrivée est un événement de l'espace de départ.

On note aussi

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

Ainsi, les événements $\{X = x\}$ et $\{X \leq x\}$ correspondent respectivement aux antécédents de $\{x\}$ et de $] -\infty, x]$ par X . Par abus, on écrira directement $\mathbb{P}(X \in A)$ pour désigner

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

De même, on notera $\mathbb{P}(X = x)$ et $\mathbb{P}(X \leq x)$ les probabilités des événements correspondants.

★ Exemple : somme de deux dés

On lance deux dés équilibrés de manière indépendante. L'univers est $\Omega = \{1, \dots, 6\}^2$, muni de la tribu $\mathcal{F} = \mathcal{P}(\Omega)$ et de la probabilité uniforme \mathbb{P} . On définit la variable aléatoire $X : \Omega \rightarrow \mathbb{N}$ par

$$X(\omega_1, \omega_2) = \omega_1 + \omega_2,$$

c'est-à-dire la somme des deux résultats. L'espace d'arrivée est $E = \mathbb{N}$ muni de la tribu $\mathcal{A} = \mathcal{P}(\mathbb{N})$. Pour tout $A \in \mathcal{A}$, on a bien $X^{-1}(A) \in \mathcal{F}$.

Par exemple :

$$\{X = 7\} = X^{-1}(\{7\}) = \{(\omega_1, \omega_2) \in \Omega : \omega_1 + \omega_2 = 7\},$$

qui contient 6 couples possibles : $(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)$. De même,

$$\{X \leq 4\} = X^{-1}(\{1, 2, 3, 4\}),$$

qui contient également 6 couples possibles.

3.1.2 Loi d'une variable aléatoire

Définition 15 (Loi d'une variable aléatoire). Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et (E, \mathcal{A}) un espace probabilisable. La **loi** (ou **distribution**) d'une variable aléatoire $X : \Omega \rightarrow E$ est la fonction \mathbb{P}_X définie par

$$\begin{aligned} \mathbb{P}_X : \mathcal{A} &\longrightarrow [0, 1] \\ A &\longmapsto \mathbb{P}(X \in A). \end{aligned}$$

On vérifie que \mathbb{P}_X est une mesure de probabilité sur (E, \mathcal{A}) grâce aux propriétés de l'image réciproque. Remarquons que si μ est une loi de probabilité sur (E, \mathcal{A}) , il existe toujours une variable aléatoire de loi μ . Il suffit de prendre $(\Omega, \mathcal{F}, \mathbb{P}) = (E, \mathcal{A}, \mu)$ et $X(\omega) = \omega$. C'est ce que l'on appelle la **construction canonique**.

De plus, lorsque $E = E_1 \times E_2$ et que

$$\mathcal{A} \supset \mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(\{A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}),$$

on peut écrire $X = (X_1, X_2)$. Les lois de X_1 et X_2 sont appelées **lois marginales** de X . Ici, $\sigma(F)$ désigne la plus petite tribu contenant la famille d'ensembles F .

★ Exemple : un couple de dés et ses lois marginales

On lance deux dés équilibrés de manière indépendante. L'univers est

$$\Omega = \{1, \dots, 6\}^2,$$

avec la tribu $\mathcal{F} = \mathcal{P}(\Omega)$ et la probabilité uniforme \mathbb{P} . On définit le couple de variables aléatoires $X = (X_1, X_2)$ par

$$X_1(\omega_1, \omega_2) = \omega_1, \quad X_2(\omega_1, \omega_2) = \omega_2,$$

où $(\omega_1, \omega_2) \in \Omega$. L'espace d'arrivée est $E = \{1, \dots, 6\}^2$, muni de la tribu $\mathcal{A} = \mathcal{P}(E)$.

La loi de X est donnée par

$$\mathbb{P}_X(\{(i, j)\}) = \mathbb{P}(X = (i, j)) = \frac{1}{36}, \quad (i, j) \in E.$$

C'est donc la loi uniforme sur E .

Les lois marginales correspondent aux lois de X_1 et X_2 . Par exemple, pour tout $k \in \{1, \dots, 6\}$,

$$\mathbb{P}_{X_1}(\{k\}) = \mathbb{P}(X_1 = k) = \sum_{j=1}^6 \mathbb{P}_X(\{(k, j)\}) = \sum_{j=1}^6 \frac{1}{36} = \frac{6}{36} = \frac{1}{6}.$$

De même,

$$\mathbb{P}_{X_2}(\{k\}) = \mathbb{P}(X_2 = k) = \frac{1}{6}.$$

Autrement dit, X_1 et X_2 suivent chacun la loi uniforme sur $\{1, \dots, 6\}$.

3.1.3 Indépendance de variables aléatoires

Définition 16 (Deux variables aléatoires). Soient X et Y deux variables aléatoires définies sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, à valeurs respectivement dans les espaces probabilisables (E, \mathcal{A}) et (F, \mathcal{G}) . On dit que X et Y sont **indépendantes** si

$$\forall A \in \mathcal{A}, \forall B \in \mathcal{G}, \quad \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B). \quad (1)$$

Autrement dit, connaître la réalisation de X ne donne aucune information supplémentaire sur la probabilité des événements liés à Y , et réciproquement.

★ Exemple : un couple non indépendant

On considère l'univers $\Omega = \{(i, j) \in \{1, \dots, n\}^2 : i \leq j\}$, muni de la probabilité uniforme. Autrement dit, toutes les paires (i, j) avec $1 \leq i \leq j \leq n$ sont équiprobables. On définit

$$X(i, j) = i, \quad Y(i, j) = j.$$

Alors X et Y ne sont pas indépendantes :

$$\mathbb{P}(X = 1, Y = 3) = \mathbb{P}(\{(1, 3)\}) = \frac{1}{6} \neq \mathbb{P}(X = 1) \mathbb{P}(Y = 3) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

On peut réécrire (1) sous la forme

$$\mathbb{P}(X^{-1}(A) \cap Y^{-1}(B)) = \mathbb{P}(X^{-1}(A)) \mathbb{P}(Y^{-1}(B)),$$

ce qui souligne que l'indépendance est avant tout une propriété des *événements engendrés* par les variables.

Définition 17 (Tribu engendrée par une variable aléatoire). Soit (E, \mathcal{A}) un espace probabilisable et $X : \Omega \rightarrow E$ une fonction. La **tribu engendrée** par X , notée $\sigma(X)$, est la plus petite tribu sur Ω telle que X soit une variable aléatoire. On a

$$\sigma(X) = \{X^{-1}(A) : A \in \mathcal{A}\}.$$

Propriété 20. Deux variables aléatoires X et Y sont indépendantes si et seulement si les tribus $\sigma(X)$ et $\sigma(Y)$ sont indépendantes.

Cette reformulation met en évidence que l'indépendance n'est pas propre aux variables elles-mêmes, mais à l'*information* qu'elles contiennent.

Définition 18 (Famille de variables aléatoires). Une famille $(X_i)_{i \in I}$ de variables aléatoires définies sur un même espace de probabilité est dite **indépendante** si les tribus $(\sigma(X_i))_{i \in I}$ le sont. Autrement dit, pour toute partie finie $J \subset I$ et pour tout choix d'événements $(A_j)_{j \in J} \in \prod_{j \in J} \mathcal{A}_j$,

$$\mathbb{P}\left(\bigcap_{j \in J} \{X_j \in A_j\}\right) = \prod_{j \in J} \mathbb{P}(X_j \in A_j).$$

En particulier, si $I = \{1, \dots, n\}$ est fini, les variables X_1, \dots, X_n sont indépendantes si et seulement si

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i),$$

pour tout choix de parties $A_i \in \mathcal{A}_i$.

L'indépendance joue un rôle central en probabilité. En pratique, on travaille souvent avec des suites de variables aléatoires indépendantes, qui modélisent la répétition d'une même expérience (par exemple, lancer plusieurs fois une pièce ou un dé). Dans ce cadre, on introduit la notion fondamentale de suites **i.i.d.**, c'est-à-dire indépendantes et de même loi.

Définition 19 (Suite i.i.d.). Une famille de variables aléatoires $(X_i)_{i \in I}$ est dite **i.i.d.** lorsqu'elles sont à la fois indépendantes et identiquement distribuées, autrement dit de même loi. De nombreux modèles probabilistes reposent sur cette hypothèse, qui formalise l'idée d'une répétition d'expériences identiques réalisées dans les mêmes conditions.

3.2 Variables aléatoires discrètes

La première grande famille de variables aléatoires — et sans doute la plus simple à manipuler — est constituée de variables aléatoires prenant leurs valeurs dans un ensemble discret, c'est-à-dire fini ou dénombrable. Ce cadre est particulièrement adapté aux situations de type « expériences répétées », comme le lancer de dés, de pièces de monnaie ou les tirages dans une urne.

3.2.1 Définitions et caractérisations

Définition 20. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et (E, \mathcal{A}) un espace probablisable. Une variable aléatoire $X : \Omega \rightarrow E$ est dite **discrète** si sa loi \mathbb{P}_X est une probabilité discrète (voir Définition 8 et Propriété 9). On appelle **support** de X le support de sa distribution (voir Définition 5), et on le note $\text{supp}(X)$.

Autrement dit, il existe un sous-ensemble $D \subset E$ au plus dénombrable tel que

$$\mathbb{P}(X \in D) = 1.$$

Plus précisément, il existe une unique famille de réels $(p_x)_{x \in D}$ positifs ou nuls telle que

$$\mathbb{P}_X = \sum_{x \in D} p_x \delta_x \quad \iff \quad \forall x \in D, \mathbb{P}(X = x) = p_x \quad \text{et} \quad \sum_{x \in D} p_x = 1.$$

De plus, $D = \text{supp}(X)$ si et seulement si $p_x > 0$ pour tout $x \in D$.



Il est important de distinguer l'image $X(\Omega)$ et le support $\text{supp}(X)$. Si l'image $X(\Omega)$ est dénombrable, alors X est nécessairement discrète. Mais la réciproque est fautive : une variable aléatoire peut être discrète sans que son image soit strictement dénombrable. En pratique, on a toujours $\text{supp}(X) \subset X(\Omega)$, mais pas nécessairement l'égalité.

3.2.2 Lois discrètes classiques

Après avoir défini ce qu'est une variable aléatoire discrète, intéressons-nous maintenant aux principales lois de probabilité qui apparaissent dans ce cadre. Ces lois, dites **classiques**, interviennent dans de nombreuses situations de modélisation et constituent des points de repère incontournables. Nous allons en présenter les définitions et donner, pour chacune, une interprétation probabiliste.

i) Lois de Bernoulli $\mathcal{B}(p)$

On dit que X est une variable aléatoire de Bernoulli de paramètre $p \in [0, 1]$, et on note $X \sim \mathcal{B}(p)$, lorsque

$$\mathbb{P}_X = p\delta_1 + (1 - p)\delta_0,$$

c'est-à-dire, de manière équivalente,

$$\mathbb{P}(X = 1) = p \quad \text{et} \quad \mathbb{P}(X = 0) = 1 - p.$$

Cette loi décrit la situation la plus simple possible : une expérience aléatoire à deux issues.

Modélisation 1. *La loi de Bernoulli modélise par exemple le lancer d'une pièce de monnaie, équilibrée ou non, ou plus généralement toute expérience aléatoire à deux résultats possibles (succès/échec, oui/non, etc.).*

ii) Lois Binomiales $\mathcal{B}(n, p)$

On dit que X suit une loi binomiale de paramètres $n \in \mathbb{N}^*$ et $p \in [0, 1]$, et on note $X \sim \mathcal{B}(n, p)$, lorsque

$$\mathbb{P}_X = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} \delta_k,$$

c'est-à-dire, de façon équivalente, pour tout entier $0 \leq k \leq n$:

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

Il est facile de vérifier, grâce à la formule du binôme de Newton, que cette expression définit bien une loi de probabilité.

Modélisation 2. *La loi binomiale modélise le nombre de succès obtenus après n expériences indépendantes, chaque expérience ayant une probabilité de succès p .*

Propriété 21 (Somme de Bernoulli = Binomiale). *Soient X_1, \dots, X_n i.i.d. de loi $\mathcal{B}(p)$. Posons $S_n = \sum_{i=1}^n X_i$. Alors $S_n \sim \mathcal{B}(n, p)$.*

iii) Lois géométriques $\mathcal{G}(p)$

On dit que T suit une loi géométrique de paramètre $p \in (0, 1]$, et on note $T \sim \mathcal{G}(p)$, lorsque

$$\mathbb{P}_T = \sum_{n=1}^{\infty} (1-p)^{n-1} p \delta_n,$$

c'est-à-dire encore, pour tout entier $n \geq 1$:

$$\mathbb{P}(T = n) = (1-p)^{n-1} p.$$

Modélisation 3. *La loi géométrique modélise le nombre d'expériences indépendantes à réaliser jusqu'au premier succès (de probabilité p). Elle est associée à des phénomènes sans mémoire, ou sans vieillissement.*

Propriété 22 (Temps d'attente géométrique). *Soient $(X_k)_{k \geq 1}$ i.i.d. de loi $\mathcal{B}(p)$ et $T = \inf\{k \geq 1 : X_k = 1\}$. Alors $T \sim \mathcal{G}(p)$.*

iv) Lois de Poisson $\mathcal{P}(\lambda)$

On dit que Y suit une loi de Poisson de paramètre $\lambda > 0$, et on note $Y \sim \mathcal{P}(\lambda)$, lorsque

$$\mathbb{P}_Y = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \delta_n,$$

c'est-à-dire encore, pour tout entier $n \geq 0$:

$$\mathbb{P}(Y = n) = e^{-\lambda} \frac{\lambda^n}{n!}.$$

Modélisation 4. *La loi de Poisson modélise un nombre d'événements rares (nombre de gagnants au loto, nombre d'accidents sur un tronçon de route, etc.). Elle est également centrale dans la théorie des files d'attente (nombre de connexions simultanées à un serveur, par exemple).*

Propriété 23 (Loi des événements rares). Soit $\lambda > 0$ et $Y_n \sim \mathcal{B}(n, \lambda/n)$. Alors, pour tout $k \in \mathbb{N}$,

$$\mathbb{P}(Y_n = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \xrightarrow{n \rightarrow \infty} e^{-\lambda} \frac{\lambda^k}{k!}.$$

v) Autres lois

Pour finir, mentionnons rapidement d'autres lois discrètes classiques, sans donner ici leurs formules explicites, mais seulement leur interprétation. Le lecteur est invité à en retrouver les expressions exactes.

Modélisation 5 (Lois hypergéométriques). Nombre de truites obtenues en pêchant (sans remise) n poissons dans un lac contenant N poissons dont K truites. On parle alors de tirage sans remise.

Modélisation 6 (Lois multinomiales). Une urne contient des billes de k couleurs différentes en proportions respectives p_1, \dots, p_k . On tire avec remise n billes uniformément au hasard et l'on note (N_1, \dots, N_k) le nombre de billes de chaque couleur obtenues. Ce vecteur aléatoire suit alors une loi multinomiale.

Modélisation 7 (Lois binomiales négatives). Nombre d'échecs observés avant d'obtenir n succès lors d'expériences indépendantes de probabilité de succès p .

3.2.3 Indépendance

Nous nous plaçons ici dans le cas d'une famille finie de variables aléatoires, le cas général s'en déduisant sans difficulté. La caractérisation que nous allons donner est fondamentale et servira de critère pratique pour vérifier l'indépendance de variables discrètes. La démonstration complète est donnée à la fin de la section.

Propriété 24. Soient X_1, \dots, X_n des variables aléatoires discrètes définies sur le même espace de probabilité et respectivement presque sûrement à valeurs dans les ensembles dénombrables D_1, \dots, D_n . Alors ces variables sont indépendantes si et seulement si

$$\forall x_1 \in D_1, \dots, x_n \in D_n, \quad \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n). \quad (2)$$



C'est sous cette forme que l'on caractérisera en pratique l'indépendance de variables discrètes.

Démonstration. Supposons (2) vérifiée et montrons l'indépendance des variables. Soient A_1, \dots, A_n des événements des espaces d'arrivée, c'est-à-dire tels que $\{X_i \in A_i\}$ soient des événements de l'espace de probabilité. Comme les événements $\{X_i \in D_i\}$ sont presque sûrs, leurs complémentaires $\{X_i \notin D_i\}$ sont négligeables. Ainsi, pour tout événement A :

$$\mathbb{P}(A \cap \{X_i \in D_i\}) = \mathbb{P}(A), \quad \mathbb{P}(A \cap \{X_i \notin D_i\}) = 0.$$

On en déduit

$$\begin{aligned} \mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) &= \mathbb{P}(X_1 \in A_1 \cap D_1, \dots, X_n \in A_n \cap D_n) \\ &= \mathbb{P}\left(\bigsqcup_{x_1 \in A_1 \cap D_1, \dots, x_n \in A_n \cap D_n} \{X_1 = x_1, \dots, X_n = x_n\}\right) \\ &= \sum_{x_1 \in A_1 \cap D_1, \dots, x_n \in A_n \cap D_n} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\ &= \sum_{x_1 \in A_1 \cap D_1, \dots, x_n \in A_n \cap D_n} \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n) \\ &= \mathbb{P}(X_1 \in A_1 \cap D_1) \cdots \mathbb{P}(X_n \in A_n \cap D_n) \\ &= \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n). \end{aligned}$$

On a donc bien l'indépendance des variables. La réciproque est immédiate. \square

Propriété 25. Soient X_1, \dots, X_n des variables aléatoires discrètes indépendantes. Alors, pour toutes fonctions $f_1, \dots, f_n : \mathbb{R} \rightarrow \mathbb{R}$, les variables $f_1(X_1), \dots, f_n(X_n)$ sont indépendantes.

Démonstration. Il suffit d'appliquer la caractérisation précédente. \square

3.2.4 Lois conditionnelles

Les lois conditionnelles permettent de décrire la distribution d'une variable aléatoire lorsque l'on impose une information supplémentaire, par exemple la valeur prise par une autre variable. C'est un outil central en probabilités, en particulier lorsqu'on cherche à affiner une modélisation à partir d'observations partielles.

Définition 21 (Lois conditionnelles). *Soient Y une variable aléatoire et X une variable aléatoire discrète définies sur le même espace de probabilité, à valeurs respectivement dans E et F . On note $D \subset E$ un support dénombrable de X . La **loi conditionnelle de Y sachant X** est la famille de mesures de probabilité discrètes $(\mu_x)_{x \in D}$ définies, pour tout $x \in D$, par*

$$\mu_x : B \mapsto \mathbb{P}(Y \in B \mid X = x).$$

On note aussi

$$\mu_x = \mathcal{L}(Y \mid X = x) \quad \text{ou} \quad \mu_x = \mathbb{P}_{Y|X=x}.$$

★ Exemple : lancer de deux dés

Deux dés sont lancés simultanément, l'un avec la main gauche et l'autre avec la main droite. On vous informe que la somme des deux dés vaut 6. Quelle est alors la probabilité que le dé gauche ait donné k (pour $k = 1, \dots, 6$) ? Plus généralement, quelle est la probabilité qu'au moins l'un des deux dés ait donné k ?

Modélisons l'expérience par deux variables aléatoires indépendantes X_1 et X_2 , représentant respectivement les résultats des deux dés. On cherche à calculer

$$\mathbb{P}(X_1 = k \mid X_1 + X_2 = 6), \quad k \in \{1, \dots, 6\}.$$

Par définition,

$$\mathbb{P}(X_1 = k \mid X_1 + X_2 = 6) = \frac{\mathbb{P}(X_1 = k, X_1 + X_2 = 6)}{\mathbb{P}(X_1 + X_2 = 6)}.$$

Mais $\{X_1 = k\} \cap \{X_1 + X_2 = 6\} = \{X_1 = k, X_2 = 6 - k\}$. Par indépendance :

$$\mathbb{P}(X_1 = k \mid X_1 + X_2 = 6) = \frac{\mathbb{P}(X_1 = k) \mathbb{P}(X_2 = 6 - k)}{\mathbb{P}(X_1 + X_2 = 6)}.$$

Or $\mathbb{P}(X_1 + X_2 = 6) = 5/36$. De plus, pour $1 \leq k \leq 5$, on a $\mathbb{P}(X_1 = k) = \mathbb{P}(X_2 = 6 - k) = 1/6$, et pour $k = 6$ le terme disparaît. On obtient donc

$$\mathbb{P}(X_1 = k \mid X_1 + X_2 = 6) = \frac{(1/6)(1/6)}{5/36} = \frac{1}{5}, \quad 1 \leq k \leq 5,$$

et 0 si $k = 6$. Ainsi, sachant que la somme vaut 6, la loi de X_1 est uniforme sur $\{1, \dots, 5\}$.

Pour la seconde question, on calcule

$$\mathbb{P}(\{X_1 = k\} \cup \{X_2 = k\} \mid X_1 + X_2 = 6).$$

Si $k = 6$, cette probabilité est nulle. Pour $1 \leq k \leq 5$, on obtient

$$\mathbb{P}(X_1 = k \text{ ou } X_2 = k \mid X_1 + X_2 = 6) = \begin{cases} \frac{2}{5}, & k \in \{1, 2, 4, 5\}, \\ \frac{1}{5}, & k = 3. \end{cases}$$

3.3 Variables aléatoires réelles

Toutes les variables aléatoires étudiées jusqu'ici prennent leurs valeurs dans \mathbb{R} (ou dans \mathbb{R}^k pour la loi multinomiale). Pourtant, il est plus délicat d'aborder ce cas de manière générale. La difficulté provient de la structure de la tribu associée aux événements.

En effet, lorsqu'il s'agit d'une probabilité discrète μ sur un ensemble E , la tribu \mathcal{A} importe peu. On peut toujours supposer que $\mathcal{A} = \mathcal{P}(E)$. La raison est simple : on demande seulement que \mathcal{A} contienne tous les événements élémentaires associés au support dénombrable $D = \{x_i : i \in I\}$. Or, comme I est dénombrable, il suffit que \mathcal{A} contienne toutes les parties de D . Cela permet de calculer toutes les probabilités pertinentes, car les événements en dehors de D ont de toute façon une probabilité nulle. On peut donc, sans perte de généralité, travailler avec $E = D$ et $\mathcal{A} = \mathcal{P}(D)$.

La situation change radicalement lorsque l'on considère un espace non dénombrable, comme \mathbb{R} . En effet, demander que tous les événements élémentaires $\{x\}$, $x \in \mathbb{R}$, appartiennent à la tribu ne suffit pas. Prenons un exemple concret : supposons que X représente la taille d'un individu, mesurée avec une précision arbitrairement grande. On souhaite pouvoir calculer la probabilité que X soit inférieure ou égale à 1,81 mètre, c'est-à-dire $\mathbb{P}(X \leq 1,81)$. Pour cela, il faut que l'ensemble $\{X \leq 1,81\}$ appartienne à la tribu. Or cet ensemble ne peut pas s'exprimer comme une combinaison dénombrable d'événements élémentaires de la forme $\{X = x\}$, $x \in \mathbb{R}$.

C'est précisément ce qui rend nécessaire l'introduction de tribus plus riches (comme la tribu borélienne) et de mesures adaptées (comme les mesures absolument continues) pour traiter rigoureusement les variables aléatoires réelles.

3.3.1 Définition en termes d'intervalles

Définition 22 (Variables aléatoires réelles). Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $X : \Omega \rightarrow \mathbb{R}$. On dit que X est une variable aléatoire réelle si

$$\forall x \in \mathbb{R}, \{X \leq x\} \in \mathcal{F}.$$

Cela peut sembler perturbant car on ne parle nul part ici de tribu d'événements sur \mathbb{R} . On sous-entend simplement que $] - \infty, x]$ est un événement de \mathbb{R} puisque son image réciproque est un événement de Ω . Mettons cela de côté pour l'instant et poursuivons. Remarquons que le complémentaire $\{X > x\}$ de $\{X \leq x\}$ est également un événement d'après les propriétés élémentaires des tribus. Plus généralement, on peut remarquer que

$$\{X < x\} = \bigcup_{n \geq 1} \left\{ X \leq x - \frac{1}{n} \right\} \quad \text{et} \quad \{X \geq x\} = \bigcap_{n \geq 1} \left\{ X > x - \frac{1}{n} \right\},$$

sont également des événements, tout comme $\{a \leq X \leq b\}$, $\{a \leq X < b\}$, $\{a < X \leq b\}$ et enfin $\{a < X < b\}$. Ceci peut se résumer par la propriété suivante.

Propriété 26 (Variables aléatoires réelles suite). X est une variable aléatoire réelle si et seulement si

$$\forall I \subset \mathbb{R} \text{ intervalle, } \{X \in I\} \in \mathcal{F}.$$

3.3.2 Tribu borélienne de \mathbb{R}

Pour travailler rigoureusement avec des variables aléatoires réelles, il faut préciser la collection d'ensembles de \mathbb{R} que l'on autorise comme événements. Nous avons vu que se contenter des singletons $\{x\}$, $x \in \mathbb{R}$, ne suffit pas. Il est donc naturel d'exiger que la tribu contienne au moins tous les intervalles de \mathbb{R} , puisque ceux-ci représentent les événements les plus élémentaires que l'on souhaite mesurer.

Définition 23 (Tribu borélienne de \mathbb{R}). On appelle **tribu borélienne** de \mathbb{R} , notée $\mathcal{B}(\mathbb{R})$, la plus petite tribu sur \mathbb{R} contenant tous les intervalles.

Propriété 27. On a $\mathcal{B}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})$. De plus, une application $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire réelle si et seulement si

$$\forall A \in \mathcal{B}(\mathbb{R}), \quad \{X \in A\} \in \mathcal{F}.$$

Démonstration. La démonstration de la première assertion dépasse le cadre de ce cours (nous y reviendrons plus tard). Pour la seconde, considérons

$$\mathcal{T} = \{A \subset \mathbb{R} : X^{-1}(A) \in \mathcal{F}\}.$$

On remarque que \mathcal{T} est une tribu contenant tous les intervalles de \mathbb{R} . Par définition de $\mathcal{B}(\mathbb{R})$, on a donc $\mathcal{B}(\mathbb{R}) \subset \mathcal{T}$, ce qui donne le résultat. \square

La tribu borélienne est donc la structure de référence pour définir les lois de probabilité sur \mathbb{R} , et elle jouera un rôle central dans l'étude des variables aléatoires continues

3.3.3 Fonctions de Répartition

La fonction de répartition est un outil fondamental en probabilités : elle permet de décrire de manière unifiée la loi d'une variable aléatoire réelle, qu'elle soit discrète, continue ou plus générale.

Définition 24 (Fonction de répartition). Soit $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire réelle. On appelle **fonction de répartition** de X l'application $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par

$$F_X(x) = \mathbb{P}(X \leq x).$$

Propriété 28 (Propriétés caractéristiques). Toute fonction de répartition F satisfait :

1. F est croissante et continue à droite, i.e. $\forall x \in \mathbb{R}, \lim_{h \rightarrow 0^+} F(x+h) = F(x)$.
2. $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow \infty} F(x) = 1$.

Démonstration. La croissance provient de la monotonie d'une mesure de probabilité. Pour les limites, remarquons que

$$\lim_{n \rightarrow \infty} \mathbb{P}(X \leq -n) = \mathbb{P} \left(\bigcap_{n \geq 0} \{X \leq -n\} \right) = 0,$$

et

$$\lim_{n \rightarrow \infty} \mathbb{P}(X \leq n) = \mathbb{P} \left(\bigcup_{n \geq 0} \{X \leq n\} \right) = 1.$$

Enfin, la continuité à droite découle de la propriété de continuité décroissante des probabilités appliquée aux ensembles $\{X \leq x + 1/n\}$. \square

Comme F est croissante, elle admet une limite à gauche en tout point. On dit qu'elle est « càdlàg » (continue à droite, limite à gauche). On note

$$F(x-) = \lim_{h \rightarrow 0^-} F(x+h).$$

Propriété 29 (Propriétés élémentaires). Soit F_X la fonction de répartition d'une variable aléatoire réelle X . Alors, pour tous $x, y \in \mathbb{R}$ avec $x \leq y$:

$$\mathbb{P}(x < X \leq y) = F_X(y) - F_X(x), \quad \mathbb{P}(x \leq X \leq y) = F_X(y) - F_X(x-).$$

En particulier,

$$\mathbb{P}(X = x) = F_X(x) - F_X(x-),$$

et F_X est continue en x si et seulement si $\mathbb{P}(X = x) = 0$.

Démonstration. La première égalité résulte de l'égalité ensembliste $\{X \leq x\} \sqcup \{x < X \leq y\} = \{X \leq y\}$. Pour la seconde, on écrit

$$\{x \leq X \leq y\} = \bigcap_{n \geq 1} \{x - 1/n < X \leq y\},$$

et on applique la continuité décroissante des probabilités. \square

Propriété 30 (Fonction de répartition d'une variable discrète). Si X est une variable aléatoire discrète, l'ensemble D des points de discontinuité de F_X coïncide avec le support de X . De plus, si $x < y$ sont deux points de D et si $]x, y[\cap D = \emptyset$, alors F_X est constante sur $[x, y[$. On dit que F_X est **constante par morceaux**.

★ Exemple d'un jeu de hasard (suite)

Reprenons l'exemple présenté au début du chapitre 3. On obtient la fonction de répartition suivante :

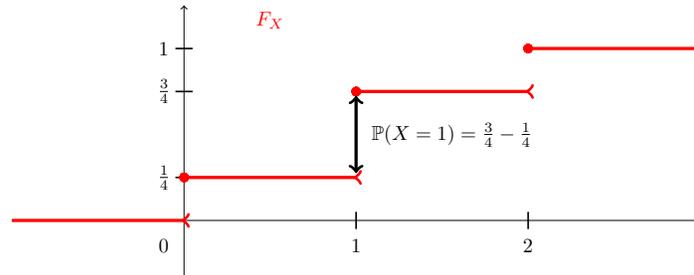


FIGURE 3.2 – Fonction de répartition associée à un jeu de hasard.

On peut de la même manière tracer les fonctions de répartition des lois discrètes classiques (Bernoulli, Binomiale, Poisson, Géométrique...). Toutes présentent un aspect constant par morceaux.

Propriété 31 (La fonction de répartition caractérise la loi). *Pour deux variables aléatoires réelles X et Y , on a*

$$\mathbb{P}_X = \mathbb{P}_Y \iff F_X = F_Y.$$

Démonstration. L'implication $\mathbb{P}_X = \mathbb{P}_Y \Rightarrow F_X = F_Y$ est immédiate. La réciproque repose sur le lemme des classes monotones et dépasse le cadre de ce cours. Cependant, une preuve simple peut-être faite dans le cadre des variables aléatoires discrètes. \square

Ainsi, on peut également parler de fonction de répartition d'une **mesure de probabilité** ou d'une **loi**.

Théorème 1 (Caractérisation des fonctions de répartition). *Soit $F : \mathbb{R} \rightarrow [0, 1]$ une fonction telle que :*

1. F est croissante et continue à droite.
2. $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow \infty} F(x) = 1$.

Alors il existe une variable aléatoire réelle X dont F est la fonction de répartition.

Démonstration. Hors programme. Ce résultat repose sur le théorème d'extension de Carathéodory. \square

★ Deux fonctions de répartition particulières

La fonction de Cantor (Figure 3.3). On définit par récurrence une suite de fonctions $(f_n)_{n \geq 0}$ sur $[0, 1]$ par $f_0(x) = x$ et

$$f_{n+1}(x) = \begin{cases} \frac{1}{2}f_n(3x) & \text{si } 0 \leq x \leq \frac{1}{3}, \\ \frac{1}{2} & \text{si } \frac{1}{3} \leq x \leq \frac{2}{3}, \\ \frac{1}{2} + \frac{1}{2}f_n(3x - 2) & \text{si } \frac{2}{3} \leq x \leq 1. \end{cases}$$

Ces fonctions sont prolongées par 0 pour $x \leq 0$ et 1 pour $x \geq 1$. On vérifie que chaque f_n est une fonction de répartition continue. La suite (f_n) converge uniformément vers une fonction f , continue et croissante : la **fonction de Cantor**. Celle-ci est constante sur les intervalles complémentaires de l'ensemble de Cantor K , et l'on a $\mathbb{P}(X \in K) = 1$ mais $\mathbb{P}(X = x) = 0$ pour tout x .

Un mélange discret-continu (Figure 3.4). Cette figure présente une fonction de répartition qui n'est ni constante par morceaux (donc pas purement discrète), ni continue (donc pas purement continue). Il s'agit d'un exemple de loi **mixte**, combinant à la fois une partie discrète et une partie continue.

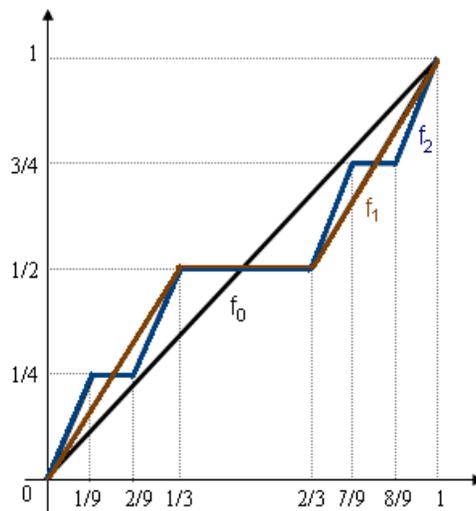


FIGURE 3.3 – Trois premières itérations de la fonction de Cantor.

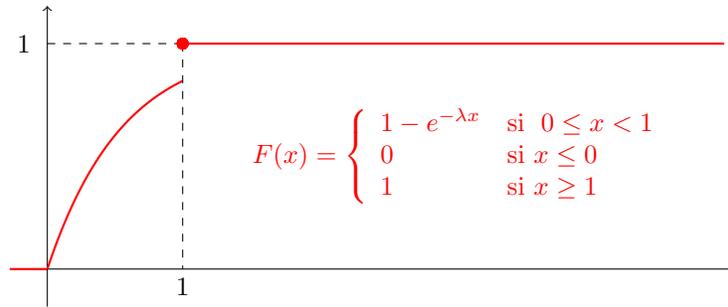


FIGURE 3.4 – Exemple de fonction de répartition ni discrète ni continue.

Ces propriétés de F_X nous mèneront naturellement, dans la suite, à relier la fonction de répartition aux densités lorsqu’elles existent et au calcul de l’espérance par intégration par rapport à la loi de X .

3.3.4 Indépendance

La notion d’indépendance des variables aléatoires est centrale : elle formalise l’idée que la connaissance de la valeur prise par certaines variables n’apporte aucune information supplémentaire sur les autres. Les résultats qui suivent sont admis, mais ils constituent des outils fondamentaux.

Propriété 32. Soient X_1, \dots, X_n des variables aléatoires réelles définies sur le même espace de probabilité. Alors ces variables sont indépendantes si et seulement si, pour tout choix d’intervalles $I_1, \dots, I_n \subset \mathbb{R}$, on a

$$\mathbb{P}(X_1 \in I_1, \dots, X_n \in I_n) = \mathbb{P}(X_1 \in I_1) \cdots \mathbb{P}(X_n \in I_n).$$

Il suffit même de vérifier cette égalité lorsque pour les intervalles $] -\infty, x]$, $x \in \mathbb{R}$.

Cette caractérisation montre que l’indépendance n’est pas seulement une intuition (« l’un n’influence pas l’autre »), mais bien une propriété précise des lois de probabilité.

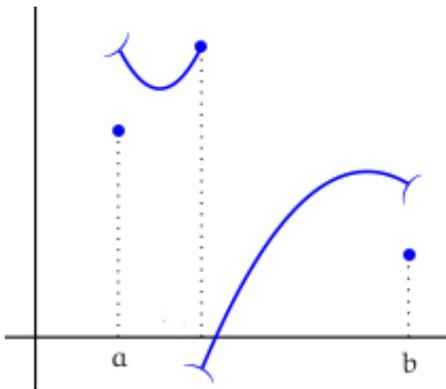
Propriété 33. Soient X_1, \dots, X_n des variables aléatoires réelles indépendantes. Alors, pour toutes fonctions $f_1, \dots, f_n : \mathbb{R} \rightarrow \mathbb{R}$ continues par morceaux, les variables $f_1(X_1), \dots, f_n(X_n)$ sont encore indépendantes.

Ce dernier résultat exprime une idée simple mais puissante : l’indépendance se conserve par

transformations séparées des variables. Par exemple, si l'on jette deux dés indépendants, alors leurs carrés, leurs doubles ou encore leurs valeurs « réduites modulo 3 » restent indépendants.

3.3.5 Variables Aléatoires Continues

Nous abordons maintenant la deuxième grande famille de variables aléatoires : les variables continues. Contrairement au cas discret, elles nécessitent l'usage de l'intégration et d'outils d'analyse plus fins. Cette section suppose donc connus les résultats de base sur l'intégration (au sens de Riemann) des fonctions continues par morceaux sur un segment.



Une fonction f est dite continue par morceaux sur $[a, b]$ s'il existe une subdivision $a = x_0 < x_1 < \dots < x_{n-1} < x_n = b$ telle que la restriction de f à $]x_i, x_{i+1}[$ admette un prolongement continu sur $[x_i, x_{i+1}]$, pour tout $1 \leq i \leq n - 1$.

Plus généralement, f est continue par morceaux sur un intervalle quelconque I si elle l'est sur tout segment $[a, b] \subset I$.

Définition 25 (Fonctions intégrables). Soit f une fonction continue par morceaux sur un intervalle ouvert $I =]a, b[$ avec $-\infty \leq a < b \leq \infty$. On dit que f est **intégrable** lorsque

$$\lim_{(u,v) \rightarrow (a+, b-)} \int_u^v |f(x)| dx < \infty.$$

Cette limite existe toujours dans $[0, \infty]$ par croissance de l'intégrale. On définit de manière analogue l'intégrabilité sur les intervalles semi-ouverts $]a, b]$ ou $[a, b[$.

Propriété 34. Si f est intégrable sur $I =]a, b[$, alors la limite

$$\lim_{(u,v) \rightarrow (a+, b-)} \int_u^v f(x) dx$$

existe et est finie. On la note

$$\int_a^b f(x) dx \quad \text{ou encore} \quad \int_I f(x) dx.$$

De plus, f est intégrable sur I si et seulement si $f \mathbf{1}_I$ est intégrable sur \mathbb{R} , ce qui permet d'écrire

$$\int_{-\infty}^{\infty} f(x) \mathbf{1}_I(x) dx = \int_{\mathbb{R}} f(x) \mathbf{1}_I(x) dx.$$

Démonstration. Il suffit de décomposer f en $f = f^+ - f^-$ où f^+ et f^- désignent respectivement les parties positive et négative de f , de sorte que $|f| = f^+ + f^-$. \square

★ Exemple : une exponentielle complexe

Soit $f(x) = \lambda e^{(-\lambda+i)x} \mathbf{1}_{\{x>0\}}$ avec $\lambda > 0$. On a $|f(x)| = \lambda e^{-\lambda x} \mathbf{1}_{\{x>0\}}$ et si $a \leq 0 \leq b$,

$$\int_a^b |f(x)| dx = \int_0^b \lambda e^{-\lambda y} dy = 1 - e^{-\lambda b}.$$

La limite vaut 1, donc f est intégrable. De plus,

$$\int_a^b f(x) dx = \int_0^b \lambda e^{(-\lambda+i)z} dz = \frac{\lambda}{\lambda - i} (1 - e^{(-\lambda+i)b}),$$

et finalement

$$\int_{-\infty}^{\infty} f(s) ds = \frac{\lambda}{\lambda - i}.$$

★ Exemple : le critère de Riemann

Soit $f(x) = \frac{1}{x^\alpha} \mathbf{1}_{\{x>0\}}$ avec $\alpha \in \mathbb{R}$. Alors :

- f est intégrable sur $]0, 1]$ si et seulement si $\alpha < 1$,
- f est intégrable sur $[1, \infty[$ si et seulement si $\alpha > 1$.

Ces conditions se vérifient facilement en calculant des primitives.

Définition 26 (Densité de probabilité). On appelle **densité de probabilité** toute fonction f continue par morceaux sur \mathbb{R} telle que :

1. $f(x) \geq 0$ pour tout $x \in \mathbb{R}$ (sauf éventuellement en un nombre dénombrable de points),
2. f est intégrable sur \mathbb{R} et

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

Définition 27 (Variable aléatoire continue). Une variable aléatoire réelle X est dite **continue**, de densité f_X , si pour tout intervalle $I \subset \mathbb{R}$,

$$\mathbb{P}(X \in I) = \int_I f_X(t) dt.$$

En particulier, pour une variable continue on a nécessairement $\mathbb{P}(X = x) = 0$ pour tout $x \in \mathbb{R}$, même si $\mathbb{P}(X \in \mathbb{R}) = 1$. Cela illustre que la décomposition $\mathbb{R} = \bigsqcup_{x \in \mathbb{R}} \{x\}$ n'est pas exploitable, car elle fait intervenir une partition non dénombrable. La fonction de répartition s'écrit alors

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(t) dt,$$

et elle est nécessairement continue.

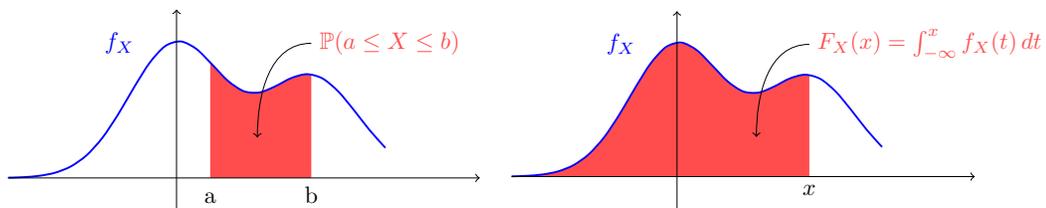


FIGURE 3.5 – Exemple de densité de probabilité et de sa fonction de répartition.

Propriété 35. Si X est une variable aléatoire réelle continue de densité f_X , alors sa fonction de répartition F_X est continue et C^1 par morceaux. En tout point où F_X est dérivable, on a

$$F'_X(x) = f_X(x).$$

Réciproquement, si F_X est continue et C^1 par morceaux, alors X admet une densité donnée par cette dérivée.

En particulier, pour une variable continue on a

$$\mathbb{P}(a < X < b) = 0 \iff F_X \text{ est constante sur }]a, b[\iff f_X = 0 \text{ sur }]a, b[,$$

sauf peut-être sur un ensemble dénombrable. La première équivalence reste valable pour toute variable réelle.

Théorème 2 (Existence et unicité des lois absolument continues). Soit f une densité de probabilité sur \mathbb{R} . Alors il existe une unique mesure de probabilité μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que, pour tout intervalle I ,

$$\mu(I) = \int_I f(t) dt.$$

Ainsi, la densité caractérise entièrement la loi. En particulier, si X et Y sont deux variables continues de densités f_X et f_Y , alors

$$\mathbb{P}_X = \mathbb{P}_Y \iff f_X = f_Y \quad (\text{sauf éventuellement sur un ensemble dénombrable}).$$

Démonstration. C'est une conséquence directe du Théorème 1. □

Ainsi, de la même façon que la fonction de répartition, la densité d'une variable aléatoire continue ne dépend que de sa loi. On peut donc parler indifféremment de la densité d'une variable, ou de la densité d'une mesure de probabilité.

3.3.6 Lois continues classiques

i) Loi uniformes sur des intervalles

La loi continue de référence est la loi uniforme sur $[0, 1]$. Plus précisément, on dit qu'une variable aléatoire U est de loi uniforme sur $[0, 1]$, et on note $U \sim \mathcal{U}([0, 1])$, lorsque U a pour densité de probabilité la fonction $f_U = \mathbb{1}_{[0,1]}$. En particulier, pour tout $0 \leq a \leq b \leq 1$,

$$\mathbb{P}(a \leq U \leq b) = \int_a^b \mathbb{1}_{[0,1]}(x) dx = \int_a^b 1 dx = b - a.$$

Plus généralement, quelque soit l'intervalle $I \subset [0, 1]$ on a

$$\mathbb{P}(U \in I) = \text{longueur}(I).$$

On appelle souvent la loi uniforme sur $[0, 1]$ la mesure de Lebesgue sur $[0, 1]$ et on la note λ ou $\lambda_{[0,1]}$. La fonction de répartition de U est donnée (voir Figure 3.6) par

$$F_U(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

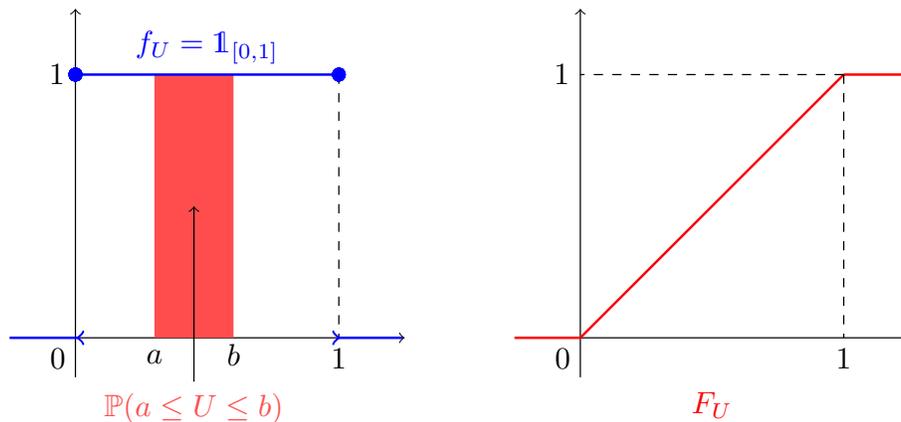


FIGURE 3.6 – Densité et fonction de répartition d'une loi uniforme sur $[0, 1]$.

Modélisation 8. *La loi uniforme sur $[0, 1]$ est fondamentale en informatique et en simulation. En effet, on peut montrer que toutes les variables aléatoires continues ou discrètes (et bien plus encore) peuvent se construire à partir de variables aléatoires uniformes sur $[0, 1]$. C'est le fameux générateur aléatoire `rand()` souvent présent sur vos machines. Cependant, il bon de savoir que ces générateurs ne sont que des versions approchées des variables aléatoires uniformes.*

Plus généralement, on dit qu'une variable aléatoire U est de loi uniforme sur un intervalle borné J de \mathbb{R} si U a pour densité de probabilité la fonction $\frac{1}{\text{longueur}(J)} \mathbb{1}_J$.

En particulier, quelque soit l'intervalle $I \subset J$ on a

$$\mathbb{P}(U \in I) = \frac{\text{longueur}(I)}{\text{longueur}(J)}.$$

De même que précédemment, la fonction de répartition est continue, affine sur J , et constante égale à 0 et à 1 à gauche et à droite de cet intervalle. On pourra voir la Figure 3.7.

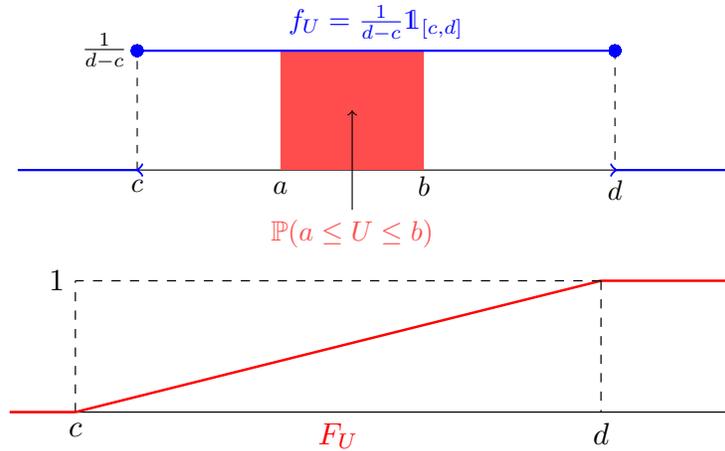


FIGURE 3.7 – Densité et fonction de répartition d’une loi uniforme sur $[c, d]$.

Modélisation 9. Par exemple, en l’absence de réveil, lorsque vous ouvrez les yeux le matin (ou l’après-midi) et que regardez la trotteuse de votre montre, il est raisonnable de penser que la position de la trotteuse est distribuée de manière uniforme sur $[0, 60]$, si vous vous intéressez aux secondes, ou encore sur $[0, 2\pi]$ si vous regardez l’angle qu’elle fait avec le nord.

ii) Lois exponentielles $\mathcal{E}(\lambda)$

On dit qu’une variable aléatoire continue T est de loi exponentielle de paramètre $\lambda > 0$, et on note $T \sim \mathcal{E}(\lambda)$, lorsque T a pour densité de probabilité la fonction $f_T(x) = \lambda e^{-\lambda x} \mathbb{1}_{\{x > 0\}}$. En particulier, quelque soit l’intervalle $I \subset [0, \infty[$,

$$\mathbb{P}(T \in I) = \int_I \lambda e^{-\lambda x} dx.$$

De plus, la fonction de répartition de T est donnée par

$$F_T(y) = \begin{cases} 0 & \text{si } y \leq 0 \\ 1 - e^{-\lambda y} & \text{si } y \geq 0 \end{cases}$$

On pourra consulter la Figure 3.8.

Modélisation 10. Les lois exponentielles modélisent des temps de vie de phénomènes sans mémoire ou sans vieillissement. Par exemple le temps de vie d’une ampoule ou d’un isotope radioactif.

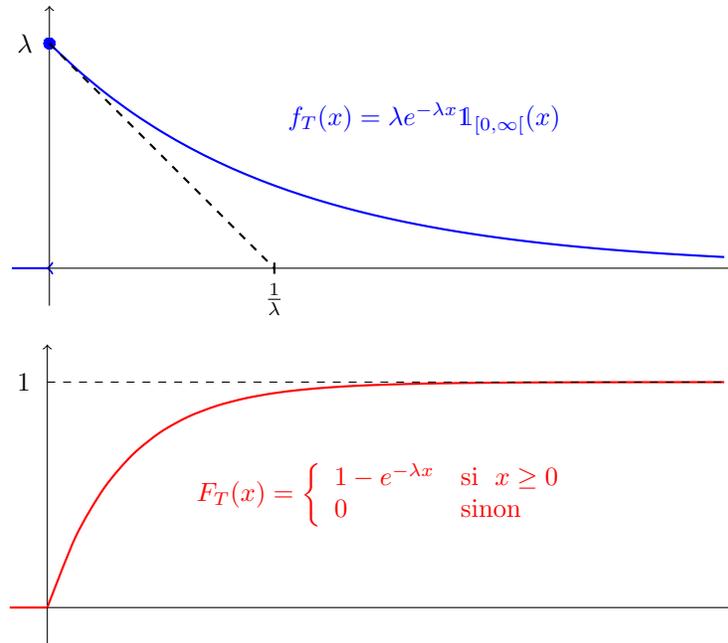


FIGURE 3.8 – Densité et fonction de répartition d’une loi exponentielle.

iii) Lois normales (ou gaussiennes) $\mathcal{N}(m, \sigma^2)$

On dit qu’une variable aléatoire réelle continue Y est de loi normale centrée réduite, et on note $Y \sim \mathcal{N}(0, 1)$, lorsque qu’elle a pour densité de probabilité la fonction

$$f_Y(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

En particulier, quelque soit l’intervalle I de \mathbb{R} on a

$$\mathbb{P}(Y \in I) = \int_I \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Il n’existe cependant pas d’expression de la fonction de répartition d’une telle loi à partir des fonctions usuelles.

Plus généralement, on dit qu’une variable aléatoire Y est de loi normale de moyenne m et de variance $\sigma^2 > 0$, et on note $Y \sim \mathcal{N}(m, \sigma^2)$, lorsque qu’elle a pour densité de probabilité la fonction

$$f_Y(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

En particulier, quelque soit l’intervalle I on a

$$\mathbb{P}(Y \in I) = \int_I \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx.$$

De même, il n’existe pas d’expression explicite de sa fonction de répartition. On appelle souvent le graphe de ces densités des courbes de Gauss ou des courbes en cloche, et ces

variables aléatoires des variables aléatoires gaussiennes. On pourra consulter la Figure 3.9 pour une illustration graphique de ces densités.

Modélisation 11. *Ces lois sont extrêmement fréquentes dans la nature comme dans les applications statistiques, du fait du théorème central limite : tout phénomène modélisable comme une somme de nombreuses variables indépendantes, de moyenne et variance finies, a une distribution asymptotiquement normale.*

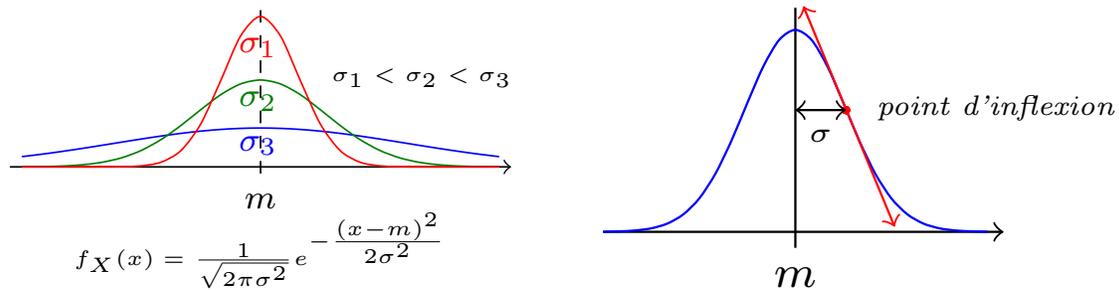


FIGURE 3.9 – Densités de plusieurs lois gaussiennes.



Toutes les lois ne sont pas nécessairement discrètes ou continues. Par exemple la loi de Cantor dont la fonction de répartition est construite dans la Figure 3.3 n'est ni discrète ni continue. Quant à la loi présentée Figure 3.4 c'est un mélange d'une loi continue et d'une loi discrète.



La définition des variables aléatoires réelles (et continues) se généralise aisément aux cas des variables aléatoires vectorielles ou complexes à valeurs dans \mathbb{R}^n ou \mathbb{C}^n .

Espérance d'une variable aléatoire réelle

Une variable aléatoire réelle X représente une quantité soumise à l'aléa. Par exemple, le nombre de gagnants au loto lors d'un tirage, ou encore la durée de vie d'un atome de césium 137. Face à une telle variable, une question naturelle se pose : *quelle est sa valeur moyenne ?* Cette moyenne, appelée **espérance** et notée $\mathbb{E}[X]$, est la version probabiliste de la notion classique de moyenne.

L'intuition est simple : si l'on observe un grand nombre de réalisations indépendantes de X , notées $X_1(\omega), \dots, X_n(\omega)$, la moyenne empirique

$$\frac{X_1(\omega) + \dots + X_n(\omega)}{n} \tag{1}$$

se rapproche de la valeur théorique $\mathbb{E}[X]$. Cette idée est formalisée par la **loi forte des grands nombres**, que nous verrons plus tard. Le terme de gauche s'appelle la **moyenne empirique (ou expérimentale)**, tandis que $\mathbb{E}[X]$ est la **moyenne théorique**.

Dans cette comparaison, les variables X_1, \dots, X_n sont supposées indépendantes et de même loi que X . Par exemple $X_1(\omega)$ peut représenter le nombre de gagnants au loto lors du premier tirage, $X_2(\omega)$ celui du second, etc. ou encore $X_1(\omega)$ la durée de vie d'un premier atome de césium 137, $X_2(\omega)$ celle d'un second, etc.

Un exemple particulièrement parlant est celui d'une pièce de monnaie équilibrée. Posons

$$X_k(\omega) = \begin{cases} 1 & \text{si le } k\text{-ème lancer donne pile,} \\ 0 & \text{sinon.} \end{cases}$$

Alors (1) traduit simplement que la proportion de piles après un grand nombre n de lancers est proche de $1/2$. La fraction $1/2$ est ici l'espérance de X , qui suit une loi de Bernoulli de paramètre $1/2$. Comment passer de cette intuition à une définition générale de l'espérance, valable pour toutes les lois de probabilité ?

4.1 Cas des variables aléatoires réelles discrètes

Reprenons l'exemple précédent en gardant à l'esprit la correspondance « probabilités - proportions ». On a en effet :

$$\mathbb{E}[X] = \frac{1}{2} \times 1 + \frac{1}{2} \times 0 = (\text{proportion de } 1) \times 1 + (\text{proportion de } 0) \times 0.$$

Les proportions correspondent ici aux probabilités $\mathbb{P}(X = 1)$ et $\mathbb{P}(X = 0)$. On voit donc clairement que l'espérance d'une variable aléatoire discrète n'est rien d'autre qu'une moyenne pondérée des différentes valeurs possibles, les poids étant les probabilités. Illustrons cela par un exemple concret.

Imaginons que vous souhaitiez calculer votre moyenne de semestre. Supposons que vos copies notées sur 20 se répartissent de la manière suivante : 30% de notes égales à 12, 50% de notes égales à 15 et 20% de notes égales à 6. Comment calculer votre moyenne ? La réponse est immédiate : il suffit de faire la moyenne pondérée des notes par leurs fréquences :

$$m = 0.3 \times 12 + 0.5 \times 15 + 0.2 \times 6 = 12.3.$$

Une autre façon de voir la chose est de supposer que vous avez 10 copies : 12, 12, 12, 15, 15, 15, 15, 6, 6. La moyenne arithmétique classique de ces 10 valeurs est encore 12.3. En langage probabiliste, cette moyenne correspond à l'espérance de la variable aléatoire « note », si l'on choisit au hasard et uniformément l'une de vos copies.

Enfin, une troisième interprétation consiste à dire que vous n'avez en réalité que trois notes distinctes : 12, 15 et 6, mais affectées de coefficients différents : 12 coefficient 3, 15 coefficient 5, 6 coefficient 2. Le calcul de la moyenne devient alors

$$m = \frac{3 \times 12 + 5 \times 15 + 2 \times 6}{10} = 12.3.$$

En termes de probabilités, cela revient à choisir une note parmi $\{12, 15, 6\}$ avec des probabilités respectives 0.3, 0.5 et 0.2.

Dans tous les cas, on obtient la même valeur : cette moyenne pondérée correspond bien à l'**espérance** de la variable aléatoire « note ».

4.1.1 Définition

Après avoir vu plusieurs exemples concrets, donnons maintenant une définition générale de l'espérance dans le cas discret.

Définition 28. Soit X une variable aléatoire réelle discrète telle que $\mathbb{P}(X \in D) = 1$, où D est un ensemble au plus dénombrable. On appelle **espérance** de X la quantité, lorsqu'elle existe,

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

Cette somme doit être comprise au sens des familles sommables, et elle est bien définie dans les deux cas suivants :

1. si $\mathbb{P}(X \geq 0) = 1$, auquel cas $\mathbb{E}[X] \in [0, +\infty[$;
2. si $\sum_{x \in D} |x| \mathbb{P}(X = x) < \infty$, auquel cas $\mathbb{E}[X] \in \mathbb{R}$.

Cette définition générale recouvre naturellement les cas les plus simples. Par exemple, si X est une variable constante presque sûrement, c'est-à-dire s'il existe $c \in \mathbb{R}$ tel que $\mathbb{P}(X = c) = 1$, alors $\mathbb{E}[X] = c$. De même, si A est un événement, la variable aléatoire indicatrice $\mathbf{1}_A$ suit une loi de Bernoulli, et l'on obtient

$$\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A).$$

Il est aussi courant d'utiliser l'expression $X \geq 0$ p.s. (presque sûrement) pour indiquer que $\mathbb{P}(X \geq 0) = 1$. Autrement dit, on peut choisir le support de X inclus dans $[0, +\infty[$.

4.1.2 Théorème de Transfert

Dans la pratique, les variables aléatoires que l'on manipule ne sont pas toujours données directement : elles apparaissent souvent comme des *transformations* de variables que l'on connaît déjà. Un exemple classique est celui de la variance d'une variable aléatoire, qui fait intervenir l'espérance de X^2 , c'est-à-dire la fonction $f(X) = X^2$. Pour calculer ce type d'espérances, il est indispensable de disposer d'un outil général : le **théorème de transfert**.

Ce résultat fondamental permet de ramener le calcul de l'espérance d'une fonction d'une variable aléatoire à une formule explicite en termes de sa loi.

Théorème 3 (Théorème de transfert). Soit X une variable aléatoire réelle discrète, presque sûrement à valeurs dans D , et soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction. Alors

$$\mathbb{E}[f(X)] \text{ est bien définie} \iff \sum_{x \in D} f(x) \mathbb{P}(X = x) \text{ est bien définie.}$$

Dans ce cas, on a l'égalité

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

Démonstration. Ce théorème n'est rien d'autre, dans le cas discret, qu'une application du principe de **sommation par paquets**. En effet, soit $K = f(D)$ l'image de D par f . Pour tout $y \in K$, posons $K_y = f^{-1}(\{y\}) \cap D$. Alors $\bigsqcup_{y \in K} K_y = D$.

On peut écrire :

$$\begin{aligned} \mathbb{E}[f(X)] &= \sum_{y \in K} y \mathbb{P}(f(X) = y) \\ &= \sum_{y \in K} y \mathbb{P}\left(\bigsqcup_{x \in K_y} \{X = x\}\right) \\ &= \sum_{y \in K} \sum_{x \in K_y} f(x) \mathbb{P}(X = x) \\ &= \sum_{x \in D} f(x) \mathbb{P}(X = x). \end{aligned}$$

□

Ainsi, la preuve dans le cadre discret n'est rien d'autre qu'une mise en œuvre du regroupement des termes « par paquets ». Le théorème de transfert sera utilisé constamment par la suite, car il permet de calculer efficacement l'espérance de n'importe quelle fonction d'une variable aléatoire. Enfin, la notion d'intégrabilité formalise la condition nécessaire pour que l'espérance soit bien définie.

Définition 29. On dit qu'une variable aléatoire réelle discrète, à valeurs dans D , est **intégrable** lorsque son espérance absolue est finie, c'est-à-dire d'après le théorème de transfert lorsque

$$\mathbb{E}[|X|] = \sum_{x \in D} |x| \mathbb{P}(X = x) < \infty.$$

4.1.3 Croissance et linéarité de l'espérance

L'espérance se comporte de manière très proche de l'intégrale : elle est croissante et linéaire. Ces deux propriétés sont fondamentales car elles permettent de manipuler l'espérance comme une moyenne généralisée, et elles serviront de base à tous les calculs qui suivront.

Propriété 36 (Croissance de l'espérance). *Soient X et Y deux variables aléatoires discrètes, positives ou nulles, ou bien intégrables. Si $X \leq Y$ p.s., c'est-à-dire lorsque $\mathbb{P}(X \leq Y) = 1$, alors*

$$\mathbb{E}[X] \leq \mathbb{E}[Y].$$

Cette propriété découle naturellement de la linéarité de l'espérance, que nous présentons maintenant.

Propriété 37 (Linéarité de l'espérance). *Soient X et Y deux variables aléatoires définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ et $\lambda, \mu \in \mathbb{R}$. Alors $\lambda X + \mu Y$ est une variable aléatoire discrète et*

$$\mathbb{E}[\lambda X + \mu Y] = \lambda \mathbb{E}[X] + \mu \mathbb{E}[Y],$$

dans chacun des deux cas suivants :

- *lorsque $X, Y \geq 0$ p.s. et $\lambda, \mu \geq 0$;*
- *lorsque X et Y sont intégrables.*

Démonstration. Commençons par le premier cas. Soient A et B des ensembles dénombrables tels que $\mathbb{P}(X \in A) = \mathbb{P}(Y \in B) = 1$. Par hypothèse, $A, B \subset [0, \infty[$. Alors $\lambda X + \mu Y$ est une variable aléatoire discrète presque sûrement positive ou nulle, et son ensemble de valeurs possibles est

$$K = \{z \in \mathbb{R} : \exists (x, y) \in A \times B, z = \lambda x + \mu y\}.$$

Le produit $A \times B$ étant dénombrable, on définit pour tout $z \in K$

$$K_z = \{(x, y) \in A \times B : \lambda x + \mu y = z\}, \quad A \times B = \bigsqcup_{z \in K} K_z.$$

On obtient alors :

$$\begin{aligned} \mathbb{E}[\lambda X + \mu Y] &= \sum_{z \in K} z \mathbb{P}(\lambda X + \mu Y = z) \\ &= \sum_{z \in K} \sum_{(x, y) \in K_z} (\lambda x + \mu y) \mathbb{P}(X = x, Y = y) \\ &= \sum_{(x, y) \in A \times B} (\lambda x + \mu y) \mathbb{P}(X = x, Y = y) \end{aligned}$$

$$\begin{aligned}
&= \lambda \sum_{x \in A} x \mathbb{P}(X = x) + \mu \sum_{y \in B} y \mathbb{P}(Y = y) \\
&= \lambda \mathbb{E}[X] + \mu \mathbb{E}[Y].
\end{aligned}$$

Le cas intégrable se traite en vérifiant que $\lambda X + \mu Y$ reste intégrable (par inégalité triangulaire). \square



Il faut être capable de justifier chacune des égalités de la preuve : les arguments essentiels sont la σ -additivité des probabilités, le regroupement des sommes (*sommation par paquets*) et la formule des probabilités totales.

Ces propriétés illustrent bien que l'espérance partage les mêmes caractéristiques que l'intégrale. En réalité, l'espérance est un cas particulier d'une intégrale plus générale : l'**intégrale de Lebesgue**. Celle-ci fournit un cadre unifié pour traiter aussi bien les variables discrètes que continues, mais son étude dépasse le cadre de ce cours.

4.1.4 Fonction génératrice d'une variable aléatoire discrète

Un outil très puissant pour l'étude des lois discrètes est la **fonction génératrice**. Elle encode la loi d'une variable aléatoire sous forme de série entière et permet notamment de calculer facilement espérances, variances et moments.

Définition 30 (Fonction génératrice). *Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N} = \{0, 1, 2, \dots\}$. On appelle **fonction génératrice** de X la fonction*

$$G_X(t) = \mathbb{E}[t^X].$$

D'après le Théorème de Transfert, on obtient immédiatement le développement en série :

$$G_X(t) = \sum_{k=0}^{\infty} t^k \mathbb{P}(X = k).$$

Les coefficients de cette série entière sont exactement les probabilités $p_k = \mathbb{P}(X = k)$. Comme $\sum_{k \geq 0} p_k = 1$, on voit que la série converge absolument pour tout $|t| \leq 1$. On en déduit que le rayon de convergence de G_X est toujours supérieur ou égal à 1.

Propriété 38 (Caractérisation de la loi). *La fonction génératrice G_X caractérise entièrement la loi de X . Plus précisément, pour tout $k \in \mathbb{N}$, on a*

$$\mathbb{P}(X = k) = \frac{G_X^{(k)}(0)}{k!}.$$

Démonstration. On part de l'expression

$$G_X(t) = \sum_{k=0}^{\infty} \mathbb{P}(X = k) t^k.$$

Il s'agit d'une série entière en t . Par unicité du développement en série entière, le coefficient d'ordre k est donné par

$$\mathbb{P}(X = k) = \frac{1}{k!} G_X^{(k)}(0).$$

Autrement dit, les dérivées successives en 0 de G_X permettent de retrouver la loi de X terme par terme. \square

Ainsi, deux variables aléatoires X et Y à valeurs dans \mathbb{N} ont la même loi si et seulement si elles possèdent la même fonction génératrice. La fonction génératrice agit donc comme une véritable « empreinte » de la loi.

Dans la suite, nous verrons qu'il est souvent plus simple de manipuler directement G_X , car ses dérivées permettent d'obtenir immédiatement l'espérance, la variance et plus généralement les moments de X .

4.1.5 Indépendance et espérance

Avant de donner le résultat général, commençons par considérer le cas de deux variables aléatoires X et Y . Si X et Y sont indépendantes, on a pour tous événements A, B :

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

En particulier, si f et g sont deux fonctions positives (ou bornées), alors la variable aléatoire $f(X)g(Y)$ vérifie

$$\mathbb{E}[f(X)g(Y)] = \sum_{x,y} f(x)g(y) \mathbb{P}(X = x, Y = y).$$

Or, par indépendance, $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$, et par le théorème de Fubini on obtient

$$\mathbb{E}[f(X)g(Y)] = \left(\sum_x f(x) \mathbb{P}(X = x) \right) \left(\sum_y g(y) \mathbb{P}(Y = y) \right) = \mathbb{E}[f(X)] \mathbb{E}[g(Y)].$$

Autrement dit, **l'espérance d'un produit se factorise en produit des espérances** lorsque les variables sont indépendantes.

Ce résultat se généralise sans difficulté à un nombre quelconque de variables.

Propriété 39. Soient X_1, \dots, X_n des variables aléatoires discrètes définies sur le même espace de probabilité. Les variables sont indépendantes si et seulement si

$$\mathbb{E}[f_1(X_1) \cdots f_n(X_n)] = \mathbb{E}[f_1(X_1)] \cdots \mathbb{E}[f_n(X_n)],$$

pour toutes fonctions $f_1, \dots, f_n : \mathbb{R} \rightarrow \mathbb{R}$ telles que $f_i(X_i)$ soit p.s. positive ou intégrable.

Démonstration. Il suffit d'appliquer le Théorème de Transfert : dans le cas discret, les espérances se traduisent par des sommes sur les supports dénombrables des X_i . L'indépendance assure la factorisation des probabilités en produit, ce qui entraîne immédiatement la factorisation des espérances. \square

Voici une application importante de ce principe.

Propriété 40 (Fonction génératrice d'une somme de variables indépendantes). Soient X_1, \dots, X_n des variables aléatoires indépendantes, à valeurs dans \mathbb{N} . Alors la fonction génératrice de leur somme $S_n = X_1 + \dots + X_n$ est donnée par

$$G_{S_n}(t) = G_{X_1}(t) \cdots G_{X_n}(t).$$

Démonstration. Par définition,

$$G_{S_n}(t) = \mathbb{E}[t^{S_n}] = \mathbb{E}[t^{X_1 + \dots + X_n}] = \mathbb{E}[t^{X_1} \cdots t^{X_n}].$$

Comme les X_i sont indépendantes, l'espérance se factorise :

$$\mathbb{E}[t^{X_1} \cdots t^{X_n}] = \mathbb{E}[t^{X_1}] \cdots \mathbb{E}[t^{X_n}] = G_{X_1}(t) \cdots G_{X_n}(t).$$

\square

Nous généralisons maintenant la notion de fonction génératrice à plusieurs variables aléatoires simultanément.

Définition 31 (Fonction génératrice conjointe). Soient X et Y deux variables aléatoires discrètes à valeurs dans \mathbb{N} . On appelle **fonction génératrice conjointe** de (X, Y) la fonction

$$G_{(X,Y)}(s, t) = \mathbb{E}[s^X t^Y] = \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} s^x t^y \mathbb{P}(X = x, Y = y).$$

Cette fonction est bien définie au moins pour $|s| \leq 1$ et $|t| \leq 1$, car la série converge absolument dans ce domaine.

Propriété 41 (Critère d'indépendance via la fonction génératrice conjointe). *Soient X et Y deux variables aléatoires discrètes à valeurs dans \mathbb{N} . Alors*

$$G_{(X,Y)}(s, t) = G_X(s) G_Y(t).$$

si et seulement si X et Y sont indépendantes.

Démonstration. Si X et Y sont indépendantes, alors $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$. En remplaçant dans la définition de $G_{(X,Y)}$, on obtient

$$G_{(X,Y)}(s, t) = \sum_{x,y} s^x t^y \mathbb{P}(X = x)\mathbb{P}(Y = y) = \left(\sum_x s^x \mathbb{P}(X = x) \right) \left(\sum_y t^y \mathbb{P}(Y = y) \right),$$

c'est-à-dire $G_X(s)G_Y(t)$.

Réciproquement, si $G_{(X,Y)}(s, t) = G_X(s)G_Y(t)$, les coefficients du développement en série double sont les probabilités $\mathbb{P}(X = x, Y = y)$. Par unicité du développement en série entière, on retrouve nécessairement que $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ pour tout (x, y) , ce qui signifie que X et Y sont indépendantes. \square

Ainsi, les fonctions génératrices (simples ou conjointes) fournissent un outil particulièrement efficace pour caractériser l'indépendance et calculer les lois de sommes de variables aléatoires.

4.1.6 Variance et covariance

Définition 32 (Variance et écart-type). *Soit X une variable aléatoire réelle discrète. On appelle **variance** de X la quantité (quand elle existe)*

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

Si D est un support dénombrable de X (i.e. $\mathbb{P}(X \in D) = 1$), alors

$$\mathbb{V}(X) = \sum_{x \in D} (x - \mathbb{E}[X])^2 \mathbb{P}(X = x).$$

La variance est bien définie dès que X est **de carré intégrable** (voir ci-dessous). On appelle **écart-type** la racine carrée de la variance :

$$\sigma_X = \sqrt{\mathbb{V}(X)}.$$

La variance mesure l'**étalement** des valeurs de X autour de sa moyenne $\mathbb{E}[X]$. L'écart-type, homogène à X , fournit une échelle de fluctuation typique. On a $\mathbb{V}(X) = 0$ si et seulement si $X = \mathbb{E}[X]$ p.s.

Définition 33 (Variables de carré intégrable). *On dit qu'une variable aléatoire réelle discrète X est **de carré intégrable** lorsque $\mathbb{E}[X^2] < \infty$.*

Dans ce cas, $\mathbb{E}[X]$ est fini, $\mathbb{V}(X) < \infty$, et l'on dispose de la formule suivante.

Propriété 42 (Formule de la variance). *Si X est de carré intégrable, alors*

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

Démonstration. Développer $(X - \mathbb{E}[X])^2 = X^2 - 2X\mathbb{E}[X] + (\mathbb{E}[X])^2$ puis utiliser la linéarité de l'espérance. \square

En particulier, on déduit de la positivité de la variance que

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

ce qui n'est rien d'autre qu'un cas particulier de l'inégalité de Cauchy-Schwarz.

Définition 34 (Covariance). *Pour deux variables aléatoires discrètes X, Y de carré intégrable, on définit la **covariance** par*

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

Si D_X, D_Y sont des supports dénombrables de X, Y , alors, *par transfert*,

$$\text{Cov}(X, Y) = \sum_{x \in D_X} \sum_{y \in D_Y} (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) \mathbb{P}(X = x, Y = y).$$

Propriété 43 (Propriétés fondamentales). Pour X, Y de carré intégrable :

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ et $\text{Cov}(X, X) = \mathbb{V}(X)$;
2. bilinéarité en chaque argument : $\text{Cov}(aX + bZ, Y) = a \text{Cov}(X, Y) + b \text{Cov}(Z, Y)$;
3. Cauchy-Schwarz : $|\text{Cov}(X, Y)| \leq \sqrt{\mathbb{V}(X) \mathbb{V}(Y)}$;
4. variance d'une somme :

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2 \text{Cov}(X, Y).$$

En particulier, si X et Y sont indépendantes, alors $\text{Cov}(X, Y) = 0$ et $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$.



Mises en garde et règles de base. En général, $\mathbb{V}(X + Y) \neq \mathbb{V}(X) + \mathbb{V}(Y)$. On a toutefois

$$\mathbb{V}(\lambda X) = \lambda^2 \mathbb{V}(X), \quad \mathbb{V}(X + b) = \mathbb{V}(X).$$

On définit (lorsque $\sigma_X, \sigma_Y > 0$) le **coefficient de corrélation** par

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \in [-1, 1].$$

Il mesure l'intensité de la dépendance linéaire :

- $\rho \approx 1$: variables fortement corrélées positivement ;
- $\rho \approx -1$: variables fortement corrélées négativement ;
- $\rho \approx 0$: variables non corrélées (mais pas forcément indépendantes).

Les cas extrêmes $\rho = \pm 1$ correspondent à une relation affine parfaite p.s. entre X et Y .

4.1.7 Moments d'une variable aléatoire

Jusqu'à présent, nous avons introduit deux quantités fondamentales associées à une variable aléatoire X : son **espérance** $\mathbb{E}[X]$, qui décrit la tendance centrale, et sa **variance** $\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$, qui mesure la dispersion autour de la moyenne. Ces deux notions s'inscrivent en fait dans un cadre plus général : celui des **moments**.

Définition 35 (Moments). Soit X une variable aléatoire réelle discrète et $p \geq 0$. On dit que X **admet un moment d'ordre p** si

$$\mathbb{E}[|X|^p] < \infty.$$

Le **moment d'ordre p** est alors la quantité $\mathbb{E}[X^p]$.

Les moments d'une variable aléatoire sont des espérances de puissances de X (éventuellement prises en valeur absolue), et permettent de caractériser de manière fine la distribution. Par exemple : le moment d'ordre 1 est l'espérance ; le moment d'ordre 2 intervient dans la variance ; les moments d'ordre supérieur donnent des informations supplémentaires sur la forme de la loi (asymétrie, aplatissement, etc.).

Propriété 44 (Hiérarchie des moments). Si X admet un moment d'ordre $p > 0$, alors, pour tout $q \in [0, p]$, X admet aussi un moment d'ordre q . Autrement dit,

$$\mathbb{E}[|X|^p] < \infty \implies \forall q \in [0, p], \mathbb{E}[|X|^q] < \infty.$$

Démonstration. Pour tout $x \in \mathbb{R}$ et $0 \leq q \leq p$, on a $|x|^q \leq 1 + |x|^p$ (car $|x|^q \leq |x|^p$ si $|x| \geq 1$, et $|x|^q \leq 1$ sinon). Donc $\mathbb{E}[|X|^q] \leq \mathbb{E}[1 + |X|^p] = 1 + \mathbb{E}[|X|^p] < \infty$. \square

Les fonctions génératrices fournissent un outil très pratique pour calculer les moments. En effet, en dérivant G_X et en évaluant en $t = 1$, on obtient directement l'espérance et les moments d'ordre 2.

Propriété 45 (Moments via la fonction génératrice). Soit X une variable aléatoire discrète à valeurs dans \mathbb{N} de fonction génératrice G_X . Alors :

$$\mathbb{E}[X] = G'_X(1^-), \quad \mathbb{E}[X(X-1)] = G''_X(1^-).$$

En particulier,

$$\mathbb{E}[X^2] = G''_X(1^-) + G'_X(1^-).$$

La variance, si les quantités précédentes sont finies, est alors égale à

$$\mathbb{V}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = G''_X(1^-) + G'_X(1^-) - (G'_X(1^-))^2.$$

Démonstration. On part de l'expression

$$G_X(t) = \sum_{k=0}^{\infty} p_k t^k, \quad p_k = \mathbb{P}(X = k).$$

Alors

$$G'_X(t) = \sum_{k=1}^{\infty} k p_k t^{k-1}, \quad G''_X(t) = \sum_{k=2}^{\infty} k(k-1) p_k t^{k-2}.$$

En passant à la limite $t \rightarrow 1^-$, on retrouve

$$G'_X(1^-) = \sum_{k=1}^{\infty} k p_k = \mathbb{E}[X], \quad G''_X(1^-) = \sum_{k=2}^{\infty} k(k-1) p_k = \mathbb{E}[X(X-1)].$$

Enfin, $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X]$, ce qui donne les formules annoncées. \square

Ainsi, la fonction génératrice permet non seulement de caractériser la loi, mais aussi de calculer très simplement les moments d'ordre 1 et 2, qui jouent un rôle central en probabilité.

4.1.8 Cas des lois discrètes classiques

Nous présentons ici les principales lois discrètes usuelles. Pour chacune, nous calculons l'espérance, la variance et nous donnons également la fonction génératrice. Les preuves sont données de manière élémentaire afin d'illustrer les méthodes classiques (dérivation de séries entières, décomposition en variables de Bernoulli, etc.).

i) Loi de Bernoulli

Si $X \sim \mathcal{B}(p)$, alors

$$\mathbb{E}[X] = 1 \times p + 0 \times (1-p) = p.$$

Pour la variance, on applique la formule $\mathbb{V}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. Comme $X^2 = X$ (puisque $X \in \{0, 1\}$), on a $\mathbb{E}[X^2] = \mathbb{E}[X] = p$. D'où

$$\mathbb{V}(X) = p - p^2 = p(1-p).$$

On vérifie facilement que sa fonction génératrice est

$$G_X(t) = \mathbb{E}[t^X] = (1-p) + pt.$$

ii) Loi Binomiale

Si $X \sim \mathcal{B}(n, p)$, alors

$$\mathbb{E}[X] = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k}$$

$$\begin{aligned}
&= \sum_{k=1}^n n \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np.
\end{aligned}$$

On a utilisé la relation $k \binom{n}{k} = n \binom{n-1}{k-1}$ ainsi que la formule du binôme de Newton.

Pour la variance, il est commode de voir X comme une somme de n variables de Bernoulli indépendantes $X = X_1 + \dots + X_n$, chacune de paramètre p . Alors

$$\mathbb{V}(X) = \mathbb{V}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{V}(X_i) = np(1-p).$$

En utilisant que X est somme de n Bernoulli indépendantes, sa fonction génératrice est donnée par

$$G_X(t) = ((1-p) + pt)^n.$$

iii) Loi Géométrique

Si $T \sim \mathcal{G}(p)$, on a

$$\mathbb{E}[T] = \sum_{k=1}^{\infty} k (1-p)^{k-1} p = pf'(1-p),$$

où $f(x) = \sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ pour $0 \leq x < 1$. On obtient donc

$$\mathbb{E}[T] = \frac{p}{(1-(1-p))^2} = \frac{1}{p}.$$

Pour la variance, on calcule de même $\sum k^2 q^{k-1}$ avec $q = 1-p$. On utilise

$$f(x) = \frac{1}{1-x}, \quad f'(x) = \frac{1}{(1-x)^2}, \quad f''(x) = \frac{2}{(1-x)^3}.$$

Ainsi

$$\sum_{k=1}^{\infty} k^2 q^{k-1} = f'(q) + qf''(q) = \frac{1}{(1-q)^2} + \frac{2q}{(1-q)^3} = \frac{1+q}{(1-q)^3}.$$

D'où

$$\mathbb{E}[T^2] = p \cdot \frac{1+q}{(1-q)^3} = \frac{2-p}{p^2}.$$

Finalement

$$\mathbb{V}(T) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Sa fonction génératrice est quant à elle donnée par :

$$G_T(t) = \sum_{k=1}^{\infty} p(1-p)^{k-1} t^k = \frac{pt}{1-(1-p)t}, \quad |t| < \frac{1}{1-p}.$$

iv) **Loi de Poisson**

Si $Y \sim \mathcal{P}(\lambda)$, alors

$$\mathbb{E}[Y] = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = \lambda.$$

On a utilisé le développement en série entière de exp.

Pour la variance, on reprend avec k^2 :

$$\mathbb{E}[Y^2] = \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!}.$$

On écrit $k^2 = k(k-1) + k$:

$$\mathbb{E}[Y^2] = \sum_{k=0}^{\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} + \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!}.$$

Le second terme vaut $\mathbb{E}[Y] = \lambda$. Pour le premier, poser $m = k - 2$ donne

$$\sum_{k=2}^{\infty} k(k-1) \frac{\lambda^k}{k!} = \lambda^2 \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = \lambda^2 e^\lambda.$$

Donc

$$\mathbb{E}[Y^2] = e^{-\lambda}(\lambda^2 e^\lambda) + \lambda = \lambda^2 + \lambda,$$

et par suite

$$\mathbb{V}(Y) = \lambda.$$

La fonction génératrice vérifie

$$G_Y(t) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{(\lambda t)^k}{k!} = e^{\lambda(t-1)}.$$

Tableau récapitulatif

Loi	Espérance	Variance	Fonction génératrice $G_X(t)$
Bernoulli $\mathcal{B}(p)$	p	$p(1-p)$	$1-p+pt$
Binomiale $\mathcal{B}(n, p)$	np	$np(1-p)$	$(1-p+pt)^n$
Géométrique $\mathcal{G}(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pt}{1-(1-p)t}$
Poisson $\mathcal{P}(\lambda)$	λ	λ	$e^{\lambda(t-1)}$

4.2 Cas des variables aléatoires réelles continues

Jusqu'ici nous avons travaillé essentiellement avec des variables aléatoires discrètes. Mais de nombreuses situations pratiques font intervenir des variables aléatoires continues : par exemple, la taille d'un individu, le temps d'attente avant un événement, ou encore la mesure d'une grandeur physique. Dans ce cas, la loi n'est plus décrite par une probabilité affectée à chaque valeur isolée, mais par une **densité de probabilité**.

L'espérance et la variance se définissent alors à l'aide d'intégrales, ce qui généralise naturellement les sommes que nous avons rencontrées dans le cas discret. Les grandes propriétés de l'espérance (linéarité, positivité, croissance) restent valables, mais nous n'insisterons pas sur les démonstrations techniques.

4.2.1 Définition et propriétés

L'espérance d'une variable continue se calcule de manière similaire au cas discret : il suffit de remplacer la somme pondérée par une intégrale pondérée par la densité.

Théorème 4 (Espérance d'une variable continue). *Soit X une variable aléatoire réelle continue de densité h , continue par morceaux. Alors*

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x h(x) dx,$$

dès que cette intégrale est définie (par exemple si $xh(x) \geq 0$ ou si $xh(x)$ est intégrable).

4.2.2 Théorème de transfert

Comme dans le cas discret, il est souvent nécessaire de calculer l'espérance d'une fonction d'une variable aléatoire. Le principe reste le même : remplacer la somme par une intégrale.

Théorème 5 (Théorème de transfert pour les lois continues). *Soit X une variable aléatoire continue de densité h , et $f : \mathbb{R} \rightarrow \mathbb{R}$ continue par morceaux. Alors*

$$\mathbb{E}[f(X)] = \int_{-\infty}^{\infty} f(x) h(x) dx,$$

dès que l'intégrale est définie.

Autrement dit, pour passer d'une loi discrète à une loi continue, on remplace la somme pondérée par les probabilités par une intégrale pondérée par la densité.

4.2.3 Variance

La variance se définit de la même manière que dans le cas discret, mais via une intégrale. Elle mesure toujours la dispersion autour de la moyenne :

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

En développant, on retrouve la formule pratique :

$$\mathbb{V}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2, \quad \mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 h(x) dx.$$

4.2.4 Indépendance et espérance

Le principe d'indépendance est identique à celui du cas discret : les espérances se factorisent lorsque les variables sont indépendantes.

Propriété 46. Soient X_1, \dots, X_n des variables aléatoires réelles continues. Elles sont indépendantes si et seulement si, pour toute famille de fonctions continues bornées (ou positives) f_1, \dots, f_n on a

$$\mathbb{E}[f_1(X_1) \cdots f_n(X_n)] = \mathbb{E}[f_1(X_1)] \cdots \mathbb{E}[f_n(X_n)].$$

4.2.5 Lois continues classiques

Voici trois exemples fondamentaux de lois continues, que l'on retrouve dans de nombreuses applications : la loi uniforme, la loi exponentielle et la loi normale. Chacune illustre un type de situation aléatoire.

i) Loi uniforme sur $[0, 1]$.

Si $X \sim \mathcal{U}([0, 1])$, alors $h(x) = 1$ pour $x \in [0, 1]$. On calcule

$$\mathbb{E}[X] = \int_0^1 x dx = \frac{1}{2}, \quad \mathbb{V}(X) = \int_0^1 x^2 dx - \frac{1}{4} = \frac{1}{12}.$$

ii) Exponentielle de paramètre $\lambda > 0$.

Si $X \sim \mathcal{E}(\lambda)$, alors $h(x) = \lambda e^{-\lambda x} \mathbf{1}_{\{x \geq 0\}}$. On obtient

$$\mathbb{E}[X] = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}, \quad \mathbb{V}(X) = \frac{1}{\lambda^2}.$$

iii) Normale centrée réduite.

Si $X \sim \mathcal{N}(0, 1)$, alors $h(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. On a

$$\mathbb{E}[X] = 0, \quad \mathbb{V}(X) = 1.$$

Plus généralement, si $X \sim \mathcal{N}(m, \sigma^2)$, alors $\mathbb{E}[X] = m$ et $\mathbb{V}(X) = \sigma^2$.

iv) Tableau récapitulatif.

Loi	Densité $h(x)$	Espérance $\mathbb{E}[X]$	Variance $\mathbb{V}(X)$
Uniforme $\mathcal{U}([0, 1])$	$h(x) = 1$ pour $x \in [0, 1]$	$\frac{1}{2}$	$\frac{1}{12}$
Exponentielle $\mathcal{E}(\lambda)$	$h(x) = \lambda e^{-\lambda x}$, $x \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normale $\mathcal{N}(m, \sigma^2)$	$h(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-m)^2/2\sigma^2}$	m	σ^2

4.2.6 Un mot sur la fonction caractéristique

Dans le cas discret, nous avons vu que la fonction génératrice est un outil très pratique pour caractériser une loi et calculer les moments. Dans le cas continu (et même en général), l'analogue est la **fonction caractéristique** :

$$\varphi_X(t) = \mathbb{E}[e^{itX}], \quad t \in \mathbb{R}.$$

Nous n'entrons pas dans les détails techniques, mais il faut retenir qu'elle joue le même rôle : elle caractérise entièrement la loi et permet de retrouver les moments par dérivation.

Théorèmes limites et inégalités de concentrations

Nous regroupons ici trois résultats structurants qui expliquent à la fois la stabilité des moyennes et la forme de leurs fluctuations. Ils constituent un socle minimal pour l'analyse probabiliste et statistique :

- **Inégalités de concentration** (Markov, Tchebychev) : premiers outils non asymptotiques pour borner le risque de grandes déviations autour de la moyenne ;
- **Loi forte des grands nombres** : formalise la stabilisation de la moyenne empirique lorsque l'on répète indépendamment une même expérience ;
- **Théorème central limite** : montre que, après recentrage et renormalisation, la distribution des moyennes prend une forme gaussienne universelle.

Une présentation pleinement rigoureuse nécessiterait d'introduire les différents modes de convergence (en probabilité, presque sûre, en loi), que nous n'aborderons ici qu'à titre indicatif.

5.1 Inégalités de concentration

Les inégalités de Markov et de Tchebychev permettent d'encadrer la probabilité qu'une variable aléatoire s'éloigne trop de sa moyenne. Elles sont les premiers outils pour quantifier le fait que « la masse de probabilité se concentre autour de l'espérance ».

Propriété 47 (Inégalité de Markov). *Soit X une variable aléatoire positive, et $a > 0$. Alors*

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

Démonstration. On note $A = \{X \geq a\}$. Alors $X \geq a \mathbf{1}_A$, puisque si $X \geq a$, le côté droit vaut a , et sinon $0 \leq X$. En prenant l'espérance, on obtient

$$\mathbb{E}[X] \geq a \mathbb{E}[\mathbf{1}_A] = a \mathbb{P}(X \geq a).$$

En divisant par a , il vient bien

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

□

Propriété 48 (Inégalité de Tchebychev). *Soit X une variable aléatoire d'espérance μ et de variance σ^2 . Alors, pour tout $\varepsilon > 0$,*

$$\mathbb{P}(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Démonstration. On applique l'inégalité de Markov à la variable positive $(X - \mu)^2$ avec $a = \varepsilon^2$:

$$\mathbb{P}((X - \mu)^2 \geq \varepsilon^2) \leq \frac{\mathbb{E}[(X - \mu)^2]}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}.$$

C'est exactement l'inégalité voulue.

□

5.2 Loi des Grands Nombres

Ces inégalités montrent que plus la variance est petite, plus la variable X est concentrée autour de sa moyenne. Elles permettent aussi de comprendre pourquoi les moyennes d'un grand nombre d'observations deviennent stables.

En effet, si l'on considère X_1, \dots, X_n indépendantes de même espérance μ et variance σ^2 , alors la moyenne empirique

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$$

a encore pour espérance μ , mais une variance plus petite σ^2/n . L'inégalité de Tchebychev donne donc que pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

On voit ainsi que lorsque n croît, la probabilité que la moyenne empirique s'éloigne de μ devient de plus en plus petite : c'est la première forme de la **Loi des Grands Nombres**, la **Loi Faible des Grands Nombres**.

La version forte de la Loi des Grands Nombres affirme que la convergence des moyennes vers l'espérance a lieu avec probabilité 1 (on dit aussi « presque sûrement »).

Théorème 6 (Loi Forte des Grands Nombres). Soient $(X_n)_{n \geq 1}$ des variables aléatoires indépendantes i.i.d., d'espérance finie $\mu = \mathbb{E}[X_1]$. Alors, avec probabilité 1,

$$\frac{1}{n} \sum_{k=1}^n X_k \longrightarrow \mu \quad \text{quand } n \rightarrow \infty.$$

5.3 Théorème Central Limite

La Loi des Grands Nombres assure que la moyenne empirique se stabilise vers l'espérance. Reste à décrire la *forme* des fluctuations autour de cette moyenne : c'est l'objet du Théorème Central Limite (TCL).

Théorème 7 (Théorème Central Limite). Soient X_1, X_2, \dots i.i.d., d'espérance μ et de variance $\sigma^2 > 0$, et posons $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$. Alors, lorsque n est grand,

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \text{ a une loi proche de la loi normale } \mathcal{N}(0, 1).$$

Autrement dit, la fonction de répartition du terme de gauche se rapproche de celle d'une normale centrée réduite (et, de manière équivalente, sa fonction caractéristique aussi). En pratique, pour des calculs de probabilités ou d'espérances « raisonnables » impliquant cette quantité, on peut remplacer sa loi par $\mathcal{N}(0, 1)$: l'approximation s'améliore quand n croît. La loi normale apparaît ainsi comme la loi *limite* des fluctuations de la moyenne.

Le TCL a diverse applications :

- **Sondages et intervalles de confiance.** Le TCL justifie l'estimation d'une marge d'erreur pour la moyenne (ou une proportion). Lorsque σ est connue (ou remplacée par un estimateur $\hat{\sigma}$), un intervalle de confiance à 95% s'écrit usuellement $\bar{X}_n \pm 1,96 \hat{\sigma}/\sqrt{n}$.
- **Mesures physiques.** Il explique l'usage de la normale pour modéliser les erreurs expérimentales issues d'agrégations de petites perturbations indépendantes.
- **Statistique inférentielle.** Il constitue le socle de nombreux tests et méthodes d'estimation via des approximations normales.

Une démonstration rigoureuse s'appuie sur les fonctions caractéristiques. Des versions plus générales (hypothèses d'indépendance affaiblies, variances non identiques) existent, mais dépassent le cadre de ce cours.

Introduction aux graphes et aux chaînes de Markov finies

Les graphes sont un outil central des mathématiques discrètes. Ils offrent un langage simple et puissant pour représenter des réseaux de relations entre objets (villes et routes, personnes d'un réseau social, états et transitions d'un processus aléatoire), et pour raisonner sur ces structures de manière unifiée.

6.1 Notions de base sur les graphes

Nous introduisons ici les notions élémentaires nécessaires : les constituants (sommets et arêtes) et la manière dont ils encodent des relations.

6.1.1 Définition d'un graphe

Un **graphe orienté** (ou **digraphe**) est un couple $G = (V, E)$ où :

- V est un ensemble fini non vide, dont les éléments sont les **sommets** (**nœuds**) ;
- E est un ensemble de **paires ordonnées** (u, v) avec $u, v \in V$, appelées **arcs**.

Un arc $(u, v) \in E$ relie le sommet u au sommet v : on dit que l'arc *sort* de u et *entre* dans v , ou encore que u est **adjacent à** v . Un graphe est dit **non orienté** si, pour chaque paire non ordonnée $\{u, v\}$, la présence d'un arc (u, v) entraîne aussi celle de (v, u) ; on les représente alors par une arête non orientée unique. Les graphes peuvent également être **pondérés** : à chaque arc (u, v) est associé un poids réel (longueur, coût, probabilité de transition, etc.).

★ Exemple : cycles orienté et non orienté

Considérons $V = \{1, 2, 3, 4\}$. Posons d'abord

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

Le graphe $G = (V, E)$ est alors un **cycle orienté** à 4 sommets.

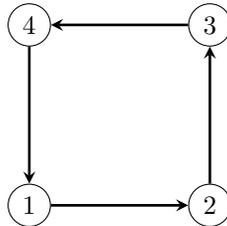


FIGURE 6.1 – Cycle orienté à 4 sommets.

Si l'on ajoute les arcs inverses $(2, 1), (3, 2), (4, 3), (1, 4)$, on obtient le **cycle non orienté** à 4 sommets, que l'on représente par des arêtes non orientées.

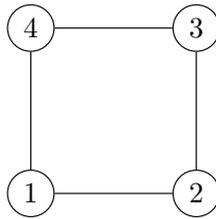


FIGURE 6.2 – Cycle non orienté à 4 sommets.

6.1.2 Sommets adjacents et degré

La notion d'adjacence et celle de degré précisent la manière dont les sommets d'un graphe sont reliés.

- Deux sommets $u, v \in V$ sont **adjacents** s'il existe au moins un arc $(u, v) \in E$ ou $(v, u) \in E$ qui les relie.
- Dans un graphe **orienté**, on distingue :
 - le **degré sortant** de v , noté $\deg^+(v)$, égal au nombre d'arcs partant de v ;
 - le **degré entrant** de v , noté $\deg^-(v)$, égal au nombre d'arcs arrivant en v .
- Dans un graphe **non orienté**, chaque arête $\{u, v\}$ correspond aux deux arcs (u, v) et (v, u) . On définit alors le **degré** de v , noté $\deg(v)$, comme le nombre de sommets adjacents à v (ou, ce qui revient au même, le nombre d'arêtes incidentes à v).

★ Exemple : degrés sur un cycle à 4 sommets

Dans le cycle **non orienté** à 4 sommets, chaque sommet est relié à deux voisins :

$$\deg(1) = \deg(2) = \deg(3) = \deg(4) = 2.$$

Dans la version **orientée**, chaque sommet possède un degré entrant et un degré sortant égaux à 1 :

$$\deg^+(v) = \deg^-(v) = 1 \quad \text{pour tout } v \in \{1, 2, 3, 4\}.$$

Relation des poignées de main. Dans tout graphe non orienté,

$$\sum_{v \in V} \deg(v) = 2|E|.$$

Démonstration. En effet, chaque arête relie deux sommets et contribue pour 1 au degré de chacun d'eux. □

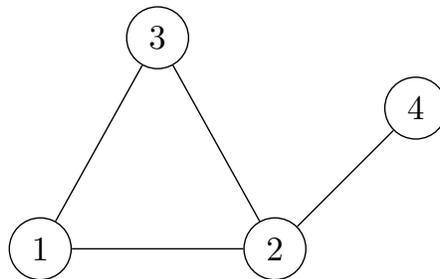


FIGURE 6.3 – On constate que $\deg(1) + \deg(2) + \deg(3) + \deg(4) = 8 = 2|E|$.

6.1.3 Ordre et taille d'un graphe

Deux paramètres fondamentaux décrivent un graphe : son **ordre** et sa **taille**.

- L'**ordre** de $G = (V, E)$ est le nombre de sommets, noté $|V|$.
- La **taille** de G est le nombre d'arcs (orienté) ou d'arêtes (non orienté), noté $|E|$.

Ainsi, l'ordre compte les sommets et la taille mesure combien de connexions les relie.

★ Exemple : graphe complet K_n

Un **graphe complet** à n sommets, noté K_n , relie chaque paire de sommets par une arête. Il est d'ordre n et de taille

$$|E| = \binom{n}{2} = \frac{n(n-1)}{2},$$

car chaque arête correspond à une paire de sommets. De plus, chaque sommet est adjacent à tous les autres, donc

$$\deg(v) = n - 1 \quad \text{pour tout } v.$$

Par exemple, K_3 est un triangle. K_5 est illustré Figure 6.4.

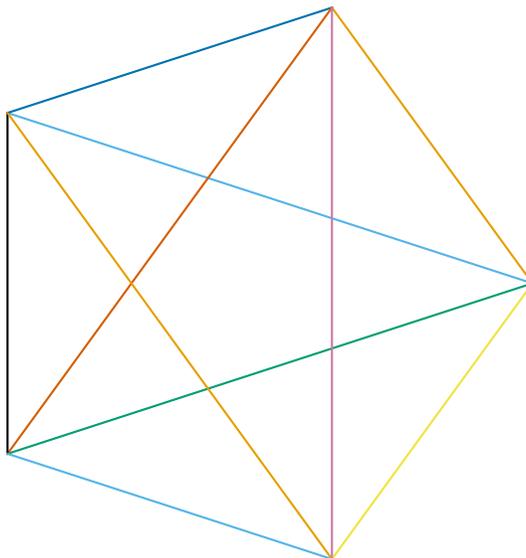


FIGURE 6.4 – Graphe complet K_5 : chaque paire de sommets est reliée par une arête.

6.1.4 Chaîne et longueur d'une chaîne

Un concept fondamental est celui de **chaîne** (souvent appelée **chemin**). Il s'agit d'une suite ordonnée de sommets telle que chaque paire consécutive soit reliée par un arc (ou une *arête* dans le cas non orienté).

- Une **chaîne** (ou **chemin**) de longueur k dans $G = (V, E)$ est une suite v_0, v_1, \dots, v_k telle que, pour tout $i = 0, \dots, k - 1$, on ait $(v_i, v_{i+1}) \in E$ (ou $\{v_i, v_{i+1}\} \in E$ en non orienté). On dit qu'elle **relie** v_0 à v_k .
- La **longueur** d'une chaîne est le nombre d'arcs (ou d'arêtes) parcourus, ici k .

Nous adopterons les conventions suivantes :

- une chaîne est dite **simple** si elle ne visite pas deux fois le même sommet ;
- une chaîne est **fermée** si $v_0 = v_k$; si elle est simple et de longueur au moins 3, on parle de **cycle** ;
- la **distance** entre deux sommets u, v est la longueur minimale d'une chaîne (chemin) reliant u à v ; on la note $d(u, v)$, avec la convention $d(u, v) = +\infty$ s'il n'existe aucune chaîne reliant u à v .

Définition 36 (Diamètre d'un graphe). Soit $G = (V, E)$ et $d(u, v)$ la distance définie ci-dessus. Le **diamètre** de G est

$$\text{diam}(G) = \sup\{d(u, v) : u, v \in V \text{ et } d(u, v) < +\infty\}.$$

Autrement dit, le diamètre est la plus grande distance entre deux sommets reliés par une chaîne ; s'il existe des paires de sommets non reliées, on peut convenir que $\text{diam}(G) = +\infty$.

★ Exemple : diamètres extrêmes

1) **Graphe complet** K_n . Toute paire de sommets est adjacente, donc $d(u, v) = 1$ pour $u \neq v$ et

$$\text{diam}(K_n) = 1,$$

qui est le diamètre minimal possible pour un graphe connexe non trivial.

2) **Graphe en ligne** L_n . Posons $V = \{1, \dots, n\}$ et $E = \{\{i, i+1\} : 1 \leq i \leq n-1\}$. On a $d(i, j) = |i - j|$, d'où

$$\text{diam}(L_n) = n - 1.$$

Parmi les graphes connexes à n sommets (voir section suivante) L_n réalise le diamètre maximal (non infini).

6.1.5 Graphe connexe

La connexité mesure à quel point un graphe est « relié ».

- Un graphe **non orienté** est **connexe** si, pour toute paire $u, v \in V$, il existe une chaîne (chemin) reliant u à v .
- Pour un graphe **orienté**, on distingue :
 - la **forte connexité** : pour tous $u, v \in V$, il existe une chaîne orientée de u vers v et de v vers u ;
 - la **faible connexité** : en oubliant l'orientation des arcs, le graphe sous-jacent est connexe.

★ Exemple : connexité et décomposition

Dans le cycle C_4 (non orienté), n'importe quel sommet est atteignable depuis n'importe quel autre : le graphe est connexe.

À l'inverse, si l'on prend deux cycles disjoints (par exemple deux triangles non reliés), il n'existe aucune chaîne entre un sommet du premier et un sommet du second : le graphe n'est pas connexe.

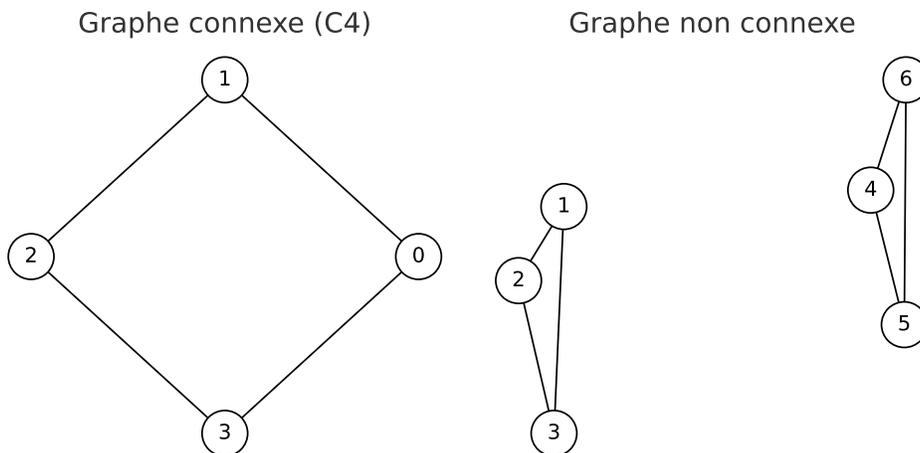


FIGURE 6.5 – À gauche : un graphe connexe (C_4). À droite : un graphe non connexe (deux composantes disjointes).

Dans le cadre des chaînes de Markov (voir sections suivantes), la forte connexité correspond à l'**irréductibilité** : chaque état est accessible depuis n'importe quel autre.

6.1.6 Matrice d'adjacence d'un graphe

Soit $G = (V, E)$ un graphe à n sommets. En notant $V = \{1, 2, \dots, n\}$, on définit la **matrice d'adjacence** de G comme la matrice carrée $A = (a_{ij})_{1 \leq i, j \leq n}$ donnée par

$$a_{ij} = \begin{cases} 1, & \text{si } (i, j) \in E \text{ (ou } \{i, j\} \in E \text{ en non orienté),} \\ 0, & \text{sinon.} \end{cases}$$

★ Exemple : lecture d'une matrice d'adjacence

Considérons le graphe non orienté de la Figure 6.3, de sommets $V = \{1, 2, 3, 4\}$ et d'arêtes

$$E = \{\{1, 2\}, \{2, 3\}, \{3, 1\}, \{2, 4\}\}.$$

Sa matrice d'adjacence est

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Propriété. Pour tout $k \geq 1$, l'entrée $(A^k)_{ij}$ est le **nombre de chaînes de longueur k reliant i à j .**

Démonstration. Pour $k \geq 2$, la formule du produit matriciel donne

$$(A^k)_{ij} = \sum_{\ell=1}^n (A^{k-1})_{i\ell} a_{\ell j}.$$

Or $a_{\ell j} = 1$ s'il existe une arête (ou un arc) de ℓ vers j . Chaque chaîne de longueur $k - 1$ reliant i à ℓ , prolongée par cette arête, produit une chaîne de longueur k reliant i à j . La somme ci-dessus compte exactement toutes ces prolongations. Le cas $k = 1$ est immédiat, d'où le résultat par récurrence. □

★ Exemple : puissances de A et comptage de chaînes

Pour le graphe de la Figure 6.3 on a en particulier

$$A^2 = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 3 & 1 & 0 \\ 1 & 1 & 2 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}.$$

Quelques points à garder en tête :

- dans un graphe non orienté, la matrice d'adjacence A est **symétrique** ;
- la correspondance graphe–matrice est centrale : les puissances de A donnent le nombre de chaînes de toute longueur.

6.2 Chaînes de Markov

Les chaînes de Markov constituent un outil fondamental en probabilités appliquées. Elles permettent de modéliser des systèmes dynamiques où l'évolution se fait par étapes successives, avec un caractère aléatoire mais gouverné par des règles simples.

On les retrouve dans de nombreux domaines : en physique (mouvements aléatoires de particules, diffusion), en biologie (modélisation des séquences d'ADN, de la propagation d'une maladie ou de populations), en informatique (algorithmes de type PageRank, réseaux de communication, apprentissage automatique), en économie (évolution de marchés, files d'attente, chaînes logistiques), etc.

Leur intérêt réside dans la simplicité de la règle de transition (qui ne dépend que de l'état présent) et dans la richesse des comportements qui en résultent : états récurrents ou transitoires, convergence vers une loi stationnaire, propriétés de mélange, etc.

6.2.1 Matrices stochastiques et graphes de transition

Définition 37 (Matrice stochastique). *Une matrice carrée $P = (p_{ij})_{1 \leq i, j \leq N}$ est dite **stochastique** si*

$$\forall 1 \leq i, j \leq N, p_{ij} \geq 0 \quad \text{et} \quad \forall 1 \leq i \leq N, \sum_{j=1}^N p_{ij} = 1.$$

Autrement dit, chaque ligne de P définit une loi de probabilité décrivant les transitions possibles à partir de l'état i .

Graphe de transition associé

À toute matrice stochastique P on associe un graphe orienté et pondéré : les sommets $V = \{1, \dots, N\}$ représentent les états, et pour chaque $p_{ij} > 0$ on trace un arc de i vers j de poids p_{ij} . Intuitivement, p_{ij} est la probabilité, en partant de i , de transiter vers j au pas de temps suivant.

★ Exemple : un graphe de transition à 3 états

Considérons $V = \{1, 2, 3\}$ et

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Le graphe de transition associé à la matrice stochastique P est représenté dans la Figure 6.6 ci-dessous.

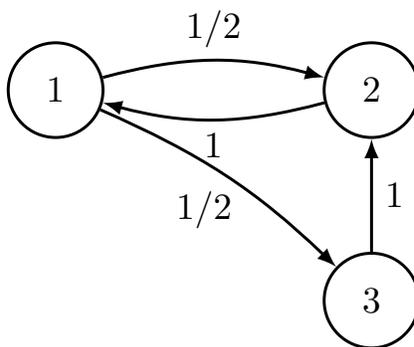


FIGURE 6.6 – Graphe de transition associé à la matrice stochastique P .

6.2.2 Définition d'une chaîne de Markov finie

Après avoir associé à une matrice stochastique P un graphe de transition, nous pouvons modéliser l'évolution aléatoire en temps discret : on observe une suite $(X_n)_{n \geq 0}$ d'états dans $E = \{1, \dots, N\}$, et à chaque pas on passe de i à j avec probabilité p_{ij} . Le point clé est que la loi du prochain état ne dépend que de l'état courant : c'est la *propriété de Markov*.

Définition 38 (Chaîne de Markov finie). Une **chaîne de Markov** de matrice de transition $P = (p_{ij})_{1 \leq i, j \leq N}$ est un processus aléatoire $(X_n)_{n \geq 0}$ à valeurs dans un ensemble fini d'états $E = \{1, \dots, N\}$ tel que, pour tout $n \geq 0$, pour tous $i_0, \dots, i_{n+1} \in E$, tels que

$$\mathbb{P}(X_n = i_n, \dots, X_0 = i_0) > 0,$$

on ait

$$\mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n) = p_{i_n, i_{n+1}}.$$

On dit que μ est la loi initiale de la chaîne lorsque la loi de X_0 est égale à μ .

Autrement dit, l'évolution future dépend uniquement de l'état présent, et non de l'histoire passée. On se représentera la chaîne comme un parcours aléatoire sur le graphe de transition : les états sont les sommets et les poids des arcs donnent les probabilités de passage.

Propriété. Soit $(X_n)_{n \geq 0}$ une chaîne de Markov de matrice de transition $P = (p_{ij})_{1 \leq i, j \leq N}$. Alors, pour tout $n \geq 1$ et tous i, j ,

$$(P^n)_{ij} = \mathbb{P}(X_n = j \mid X_0 = i).$$

En particulier P^n est encore une matrice stochastique.

Démonstration. Cas $n = 1$: par définition de P , $(P)_{ij} = p_{ij} = \mathbb{P}(X_1 = j \mid X_0 = i)$.

Supposons l'égalité vraie à l'ordre n et montrons-la à l'ordre $n + 1$. Par la formule des probabilités totales puis la propriété de Markov,

$$\mathbb{P}(X_{n+1} = j \mid X_0 = i) = \sum_{k=1}^N \mathbb{P}(X_n = k \mid X_0 = i) \mathbb{P}(X_{n+1} = j \mid X_n = k).$$

Par l'hypothèse de récurrence, $\mathbb{P}(X_n = k \mid X_0 = i) = (P^n)_{ik}$, d'où

$$\mathbb{P}(X_{n+1} = j \mid X_0 = i) = \sum_{k=1}^N (P^n)_{ik} p_{kj} = (P^{n+1})_{ij}.$$

La conclusion suit par récurrence. □

Ceci permet de calculer efficacement les probabilités après plusieurs pas par simple opérations matricielles.

★ Exemple (suite) : transitions en deux pas

Reprenons la matrice

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

On obtient

$$P^2 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix}.$$

Ainsi, P^n fournit directement les probabilités des transitions en n étapes.

On s'intéresse au **comportement asymptotique** de la loi de X_n quand $n \rightarrow \infty$. Questions : la loi $\mathbb{P}(X_n \in \cdot)$ se stabilise-t-elle ? Existe-t-il une *loi invariante* (stationnaire) vers laquelle on converge ?

6.2.3 Lois invariantes

On identifie toute probabilité μ sur $E = \{1, \dots, N\}$ à un *vecteur ligne* $\mu = (\mu_1, \dots, \mu_N)$ avec $\mu_i \geq 0$ et $\sum_{i=1}^N \mu_i = 1$. Pour $P = (p_{ij})_{1 \leq i, j \leq N}$, matrice stochastique, on définit le vecteur ligne $\mu P = ((\mu P)_1, \dots, (\mu P)_N)$ par

$$(\mu P)_j = \sum_{i=1}^N \mu_i p_{ij}.$$

On vérifie aisément que μP est encore une probabilité. En effet, on a $(\mu P)_j \geq 0$ puisque $\mu_i \geq 0$ et $p_{ij} \geq 0$. De plus,

$$\sum_{j=1}^N (\mu P)_j = \sum_{j=1}^N \sum_{i=1}^N \mu_i p_{ij} = \sum_{i=1}^N \mu_i \left(\sum_{j=1}^N p_{ij} \right) = \sum_{i=1}^N \mu_i = 1,$$

car chaque ligne de P somme à 1.

Propriété. Si $X_0 \sim \mu$, alors pour tout $n \geq 1$ et tout $1 \leq i \leq N$,

$$\mathbb{P}(X_n = i) = (\mu P^n)_i.$$

Démonstration. Cas $n = 1$: pour tout j , $\mathbb{P}(X_1 = j) = \sum_{i=1}^N \mu_i p_{ij} = (\mu P)_j$. Supposons l'assertion vraie à l'ordre n : $\mathbb{P}(X_n = \cdot) = \mu P^n$. Alors, pour tout j ,

$$\mathbb{P}(X_{n+1} = j) = \sum_{k=1}^N \mathbb{P}(X_n = k) p_{kj} = \sum_{k=1}^N (\mu P^n)_k p_{kj} = (\mu P^n P)_j = (\mu P^{n+1})_j,$$

ce qui conclut par récurrence. □

En particulier, le calcul de la loi de X_n se ramène à un *calcul matriciel*, et le comportement asymptotique de $(X_n)_{n \geq 0}$ est étroitement lié au *spectre* de P .

Définition 39 (Loi invariante). Une probabilité π sur E est dite **invariante** si

$$\pi P = \pi.$$

Ainsi, si $X_0 \sim \pi$, alors $X_n \sim \pi$ pour tout $n \geq 1$: la loi est stationnaire. Notons que π est un *vecteur propre à gauche* de P associé à la valeur propre 1.

La proposition suivante montre que si la loi de X_n se stabilise à long terme (on parlera alors de *loi stationnaire*), la limite doit nécessairement être invariante.

Propriété. Si $\mu P^n \rightarrow \pi$ quand $n \rightarrow \infty$, alors π est invariante.

Démonstration. On a

$$\pi = \lim_{n \rightarrow \infty} \mu P^n = \lim_{n \rightarrow \infty} (\mu P^n) P = \pi P,$$

d'où $\pi P = \pi$. □

★ Exemple : loi invariante pour une chaîne à 3 états

Reprenons l'exemple (voir Figure 6.6) de matrice de transition

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Cherchons $\pi = (\pi_1, \pi_2, \pi_3)$ telle que $\pi P = \pi$ et $\sum_{i=1}^3 \pi_i = 1$. Le système donne

$$\pi_1 = \pi_2, \quad \pi_3 = \frac{1}{2}\pi_1, \quad \pi_1 + \pi_2 + \pi_3 = \frac{5}{2}\pi_1 = 1,$$

d'où

$$\pi = \left(\frac{2}{5}, \frac{2}{5}, \frac{1}{5} \right).$$

★ Exemple : cas bistochastique et loi uniforme

Une matrice P est **bistochastique** si chaque colonne somme à 1 (en plus de chaque ligne). Dans ce cas, la loi uniforme $\mu = (1/N, \dots, 1/N)$ est invariante :

$$(\mu P)_j = \sum_{i=1}^N \frac{1}{N} p_{ij} = \frac{1}{N} \sum_{i=1}^N p_{ij} = \frac{1}{N} = \mu_j.$$

Par exemple,

$$P = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

est bistochastique, et $\mu = (1/3, 1/3, 1/3)$ vérifie $\mu P = \mu$.

6.2.4 Chaînes à deux états : analyse complète

On considère une chaîne sur $E = \{1, 2\}$ de matrice de transition

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}, \quad 0 \leq \alpha, \beta \leq 1.$$

Écartons d'abord les cas dégénérés :

- Cas $(\alpha, \beta) = (0, 0)$. Les deux états sont absorbants.
- Cas $\alpha = 0, \beta > 0$. 1 est absorbant ; en partant de 2, on finit presque sûrement en 1.
- Cas $\beta = 0, \alpha > 0$. Situation symétrique : 2 est absorbant ; en partant de 1, on finit presque sûrement en 2.

Dans la suite, on suppose $\alpha > 0$ et $\beta > 0$ (chaîne irréductible) et l'on détermine d'abord la loi invariante avant d'établir une formule explicite pour P^n .

Propriété. Si $\alpha, \beta > 0$, il existe une unique probabilité invariante, donnée par

$$\pi = (\pi_1, \pi_2) = \left(\frac{\beta}{\alpha + \beta}, \frac{\alpha}{\alpha + \beta} \right).$$

Démonstration. L'égalité $\pi P = \pi$ équivaut à

$$\pi_1 \alpha = \pi_2 \beta \quad \text{et} \quad \pi_1 + \pi_2 = 1,$$

d'où la formule annoncée (unicité par irréductibilité sur un espace fini). □

On peut ensuite calculer explicitement P^n , ce qui fournit une expression fermée de la loi au temps n et met en évidence la convergence.

Propriété. Posons $\lambda = 1 - \alpha - \beta$. Alors, pour tout $n \geq 0$,

$$P^n = \underbrace{\begin{pmatrix} \pi_1 & \pi_2 \\ \pi_1 & \pi_2 \end{pmatrix}}_{\Pi} + \lambda^n \underbrace{\begin{pmatrix} 1 - \pi_1 & -\pi_2 \\ -\pi_1 & 1 - \pi_2 \end{pmatrix}}_{I - \Pi}.$$

En particulier, pour toute loi initiale $\mu = (\mu_1, \mu_2)$,

$$(\mu P^n)_1 = \pi_1 + \lambda^n(\mu_1 - \pi_1), \quad (\mu P^n)_2 = \pi_2 + \lambda^n(\mu_2 - \pi_2).$$

Démonstration. La matrice Π a deux lignes égales à π et vérifie $\Pi^2 = \Pi$. De plus,

$$I - \Pi = \begin{pmatrix} 1 - \pi_1 & -\pi_2 \\ -\pi_1 & 1 - \pi_2 \end{pmatrix}, \quad (I - \Pi)^2 = I - \Pi, \quad \Pi(I - \Pi) = (I - \Pi)\Pi = 0.$$

On vérifie l'identité de décomposition $P = \Pi + \lambda(I - \Pi)$, où $\lambda = 1 - \alpha - \beta$ est la seconde valeur propre de P . Le binôme matriciel donne alors

$$P^n = (\Pi + \lambda(I - \Pi))^n = \Pi + \lambda^n(I - \Pi),$$

puis la formule pour μP^n en multipliant à gauche par μ . □

Cette écriture met en évidence la convergence : si $\alpha > 0$ et $\beta > 0$ et si $(\alpha, \beta) \neq (1, 1)$, alors $|\lambda| < 1$ et la loi μP^n tend vers π . Dans le cas particulier $\alpha = \beta = 1$, on a $\lambda = -1$ et la chaîne alterne $1 \leftrightarrow 2$: la loi ne converge pas point par point.

Théorème 8 (Convergence exponentielle et oubli de la condition initiale). *Supposons $\alpha > 0$, $\beta > 0$ et $(\alpha, \beta) \neq (1, 1)$. Alors, pour toute loi initiale μ ,*

$$\mathbb{P}(X_n = 1) \xrightarrow[n \rightarrow \infty]{} \pi_1, \quad \mathbb{P}(X_n = 2) \xrightarrow[n \rightarrow \infty]{} \pi_2,$$

et la convergence est exponentielle :

$$\left| \mathbb{P}(X_n = 1) - \pi_1 \right| = \left| \mathbb{P}(X_n = 2) - \pi_2 \right| = |\lambda|^n |\mu_1 - \pi_1|.$$

Autrement dit, la composante portée par la seconde valeur propre $\lambda = 1 - \alpha - \beta$ décroît comme $|\lambda|^n$: la chaîne oublie sa condition initiale à vitesse exponentielle.

Ce résultat est un cas particulier (dimension 2) du théorème de **Perron–Frobenius** pour les matrices stochastiques irréductibles.

Chapitre **7**

Exercices



7.1 Espaces probabilisés

Exercice 1. Soit A, B, C trois événements d'un univers Ω . Exprimer en fonction de A, B, C et des opérations ensemblistes usuelles les événements suivants :

1. A seul se produit ;
2. les trois événements se produisent ;
3. l'un au moins des événements se produit ;
4. deux au moins des événements se produisent ;
5. un événement au plus se produit ;
6. aucun des événements ne se produit ;
7. exactement deux événements se produisent.

Exercice 2. Soient \mathcal{F} et \mathcal{G} deux tribus d'événements sur un ensemble Ω . Montrer que $\mathcal{F} \cap \mathcal{G}$ est encore une tribu. Trouver un exemple pour lequel $\mathcal{F} \cup \mathcal{G}$ n'est pas une tribu.

Exercice 3. On introduit

$$\mathcal{T} = \{ A \subset \Omega ; A \text{ est dénombrable ou } A^c \text{ est dénombrable} \}.$$

1. Vérifier que \mathcal{T} est une tribu sur Ω .
2. Justifier que \mathcal{T} est la plus petite tribu contenant les singletons $\{\omega\}$ pour $\omega \in \Omega$.
3. Vérifier que si Ω est dénombrable alors $\mathcal{T} = \mathcal{P}(\Omega)$.

Exercice 4. Soit Ω un ensemble et $(A_n)_{n \in \mathbb{N}}$ une suite de parties de Ω . Déterminer les ensembles $\limsup_n A_n$ et $\liminf_n A_n$ dans les cas suivants :

- $A_n =]-\infty, n]$;
- $A_n =]-\infty, -n]$;
- $A_n =]-\infty, (-1)^n]$.

Exercice 5 (Lemme de Borel-Cantelli). Soient $(A_n)_{n \geq 0}$ des événements. Montrer que

$$\sum_{n \geq 0} \mathbb{P}(A_n) < \infty \implies \mathbb{P} \left(\bigcap_{k \geq 0} \bigcup_{n \geq k} A_n \right) = 0.$$

Interpréter ce résultat. On considère l'application suivante : le jour n , on lance une pièce équilibrée n fois et l'on note

$$A_n = \text{« les } n \text{ lancers du jour } n \text{ sont tous pile »}.$$

Que peut-on dire sur le nombre de jours n où A_n est réalisé ?

Exercice 6. Soient A_1, \dots, A_n des événements. Montrer, en utilisant la σ -sous-additivité d'une probabilité, que

$$\mathbb{P}\left(\bigcap_{k=1}^n A_k\right) \geq 1 - \sum_{k=1}^n (1 - \mathbb{P}(A_k)).$$

Exercice 7. Une boîte contient 10 boules numérotées de 1 à 10. On tire au hasard une boule. On considère les événements suivants :

- A : « le numéro de la boule est pair » ;
- B : « le numéro de la boule est multiple de 3 ».

1. Expliciter les ensembles A , B , et $A \cap B$.
2. Calculer $\mathbb{P}(A)$, $\mathbb{P}(B)$ et $\mathbb{P}(A \cap B)$ et en déduire $\mathbb{P}(A \cup B)$.

Exercice 8. Soit \mathbb{P} une probabilité discrète sur $\{1, \dots, n\}$ telle qu'il existe $\alpha > 0$ tel que pour tout $1 \leq k \leq n$ on ait $\mathbb{P}(\{1, \dots, k\}) = \alpha k$. Déterminer α et $\mathbb{P}(\{k\})$ pour tout $1 \leq k \leq n$.

Exercice 9. Déterminer $\lambda \in \mathbb{R}$ tel qu'il existe une probabilité \mathbb{P} sur $(\mathbb{N}^*, \mathcal{P}(\mathbb{N}^*))$ vérifiant

$$\mathbb{P}(\{n, n+1, \dots\}) = \frac{\lambda}{n}.$$

Exercice 10. On choisit au hasard et uniformément une application $f : \{1, 2, 3\} \rightarrow \{1, 2\}$.

1. Quel est le nombre d'applications possibles ? Quel est le nombre de surjections possibles ?
2. En déduire la probabilité d'obtenir une surjection.

Exercice 11. On procède en deux étapes :

- On tire au hasard un entier $n \geq 1$ selon la loi

$$\mathbb{P}(\{n\}) = \frac{1}{2^n}.$$

- Une fois n tiré, on lance une pièce équilibrée n fois.

On s'intéresse à l'événement B : « le nombre total de piles obtenus est égal à 0 ». On notera également $A_n = \{n\}$ l'événement « le nombre tiré est n ».

1. Justifier que \mathbb{P} est bien une probabilité sur \mathbb{N}^* .
2. Justifier que la famille $(A_n)_{n \geq 1}$ forme une partition de l'espace probabilisé.
3. Que vaut $\mathbb{P}(B \cap A_n)$? En déduire $\mathbb{P}(B)$.

Exercice 12. Soient $1 \leq p \leq n$. On choisit au hasard et uniformément une application $f : \{1, \dots, n\} \rightarrow \{1, \dots, p\}$.

1. Quel est le nombre d'applications possibles ?
2. On note A_i l'événement « la fonction tirée au sort ne prend pas la valeur i ».
 - (a) Calculer $\mathbb{P}(A_i)$.
 - (b) Plus généralement, calculer $\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k})$ lorsque $1 \leq i_1 < \dots < i_k \leq p$.
 - (c) En déduire $\mathbb{P}(\bigcup_{i=1}^p A_i)$ puis la probabilité de tirer au sort une surjection.

Exercice 13 (Attrapez-les tous!). Dans chaque paquet de Corn Flakes se trouve une figurine de Pokémon parmi Pikachu, Dracaufeu, Mewtwo, Lucario et Évoli, la probabilité qu'un paquet donné contienne un Pokémon en particulier étant $1/5$, indépendamment des autres paquets. Vous avez déjà Lucario et Évoli et vous achetez 6 paquets de Corn Flakes. Montrer que la probabilité de compléter votre collection est

$$p = 1 - 3 \left(\frac{4}{5}\right)^6 + 3 \left(\frac{3}{5}\right)^6 - \left(\frac{2}{5}\right)^6.$$

On pourra faire intervenir les événements A_α : « le Pokémon α appartient à au moins l'un des 6 paquets » pour $\alpha \in \{\text{Pikachu}, \text{Dracaufeu}, \text{Mewtwo}\}$.

Exercice 14. Votre tante Quechua et votre cousin Germain sont invités à un repas de famille. Cependant ces deux personnes sont en froid depuis quelque temps et il est préférable qu'elles ne soient pas assises côte à côte à la table. Malheureusement vous avez placé les convives au hasard (uniformément).

1. Quelle est la probabilité que Quechua et Germain soient voisins lorsque la table est ronde et qu'il y a 4 invités ?
2. Même question pour une table ronde et n convives.
3. Même question si les n personnes sont alignées sur un même banc.

7.2 Indépendance et probabilités conditionnelles

Exercice 15. Soit \mathbb{P} la probabilité uniforme sur $\Omega = \{1, \dots, p\}$.

1. On suppose $p = 6$. Montrer que $A = \{1, 2, 3\}$ et $B = \{2, 3, 5, 6\}$ sont indépendants.
2. On suppose $p = 12$. Montrer que A et B ne sont pas indépendants.
3. Supposons p premier. Montrer que si C et D sont deux événements indépendants, alors l'un des deux événements est nécessairement \emptyset ou Ω .

Exercice 16. Soient A_1, \dots, A_n des événements d'un espace probabilisé (Ω, \mathbb{P}) . On les suppose mutuellement indépendants. Montrer que

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = 1 - \prod_{i=1}^n (1 - \mathbb{P}(A_i)).$$

Application. On suppose qu'une personne est soumise à n expériences indépendantes les unes des autres et qu'à chaque expérience, elle ait une probabilité p d'avoir un accident. Quelle est la probabilité qu'elle ait au moins un accident ?

Exercice 17. Soit $n \geq 1$ un entier fixé. On choisit de manière équiprobable un entier x dans $\{1, \dots, n\}$. Pour tout entier $m \leq n$, on note A_m l'événement « m divise x ». On note également B l'événement « x est premier avec n ». Enfin, on note p_1, \dots, p_r les diviseurs premiers de n .

1. Exprimer B en fonction des A_{p_k} .
2. Pour tout entier naturel m qui divise n , calculer la probabilité de A_m .
3. Montrer que les événements A_{p_1}, \dots, A_{p_r} sont mutuellement indépendants.
4. En déduire la probabilité de B .
5. On note $\varphi(n)$ le nombre d'entiers compris entre 1 et n qui sont premiers avec n . Déduire des questions précédentes que

$$\varphi(n) = n \prod_{k=1}^r \left(1 - \frac{1}{p_k}\right).$$

Exercice 18. On tire au hasard un nombre entier strictement positif de telle façon que la probabilité d'obtenir n soit 2^{-n} . On note A_k l'événement « l'entier obtenu est multiple de k ».

1. Vérifier que $\mathbb{P}(\Omega) = 1$.
2. Calculer $\mathbb{P}(A_k)$, $k \in \mathbb{N}^*$.
3. Montrer que A_2 et A_3 ne sont pas des événements indépendants.

Exercice 19. Soit $\alpha > 1$ et \mathbb{P} une probabilité sur \mathbb{N}^* telle que

$$\mathbb{P}(\{n\}) = \frac{1}{\zeta(\alpha)} \cdot \frac{1}{n^\alpha}, \quad \text{où } \zeta(\alpha) = \sum_{n \geq 1} \frac{1}{n^\alpha}.$$

On notera $\{p_1, p_2, \dots\}$ l'ensemble des nombres premiers et $A_k = k\mathbb{N}^* = \{k, 2k, \dots\}$, $k \in \mathbb{N}^*$.

1. Calculer $\mathbb{P}(A_k)$.
2. Montrer que les événements $(A_{p_i})_{i \geq 1}$ sont indépendants.
3. Exprimer plus simplement $\bigcap_{i=1}^{+\infty} A_{p_i}^c$.
4. Justifier que

$$\mathbb{P}\left(\bigcap_{i=1}^{+\infty} A_{p_i}^c\right) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcap_{i=1}^n A_{p_i}^c\right)$$

5. En déduire

$$\zeta(\alpha) = \prod_{i=1}^{\infty} \frac{1}{1 - p_i^{-\alpha}}.$$

Exercice 20 (Lemme de Borel-Cantelli, suite). Soient $(A_n)_{n \geq 0}$ des événements indépendants.

1. Montrer que pour tout $j \geq i \geq 0$, on a :

$$\mathbb{P}\left(\bigcap_{k=i}^j A_k^c\right) \leq e^{-\sum_{k=i}^j \mathbb{P}(A_k)}.$$

On pourra utiliser l'inégalité $1 - x \leq e^{-x}$.

2. En déduire que

$$\mathbb{P}\left(\bigcap_{k=i}^{\infty} A_k^c\right) \leq e^{-\sum_{k=i}^{\infty} \mathbb{P}(A_k)}.$$

3. Montrer que

$$\sum_{k=0}^{\infty} \mathbb{P}(A_k) = \infty \implies \mathbb{P}\left(\bigcap_{i \geq 0} \bigcup_{k \geq i} A_k\right) = 1.$$

Exercice 21. Une urne contient n boules numérotées de 1 à n . On tire successivement et indépendamment, avec remise et au hasard, n boules dans l'urne. Si, au i -ème tirage, on obtient la boule numéro i , on dit qu'il y a concordance. On note $A_{n,i}$ cet événement, et

$$A_n = \bigcup_{i=1}^n A_{n,i}.$$

1. Interpréter A_n .
2. Calculer $\mathbb{P}(A_{n,i})$ et en déduire que

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = 1 - e^{-1}.$$

Exercice 22. Une urne contient 2 boules rouges et 6 boules bleues. Trois joueurs numérotés 1, 2, 3 tirent successivement, dans l'ordre de leur numéro, une boule dans l'urne, sans remise. Si un joueur tire une boule rouge, la partie s'arrête et il a gagné. Sinon, c'est au tour du joueur suivant. Si personne ne gagne, la partie est déclarée nulle.

1. Quel joueur a le plus de chances de gagner ? On formalisera soigneusement la solution en faisant intervenir les événements A_i : « le joueur i gagne » et en utilisant la formule des probabilités composées.
2. Même question pour une partie avec 7 joueurs.

Exercice 23. Une information de type vrai/faux est transmise à l'intérieur d'une population. À chaque transmission, avec probabilité p , l'information reçue est transmise telle quelle à la personne suivante ; avec probabilité $1 - p$, elle est transmise de façon contraire. On note I_n l'événement : « l'information après n transmissions est correcte », et p_n sa probabilité.

1. Établir une relation de récurrence entre p_{n+1} et p_n en utilisant la formule des probabilités totales.
2. En déduire l'expression fermée de p_n en fonction de p et de n .
3. Calculer $\lim_{n \rightarrow \infty} p_n$ et commenter le résultat.

Exercice 24. On joue à pile ou face avec une pièce non équilibrée. À chaque lancer, la probabilité d'obtenir pile est $2/3$. Les lancers sont supposés indépendants. On note A_n l'événement « au n -ième lancer, on a obtenu pour la première fois deux piles consécutives », et $p_n = \mathbb{P}(A_n)$.

On introduit également P_k et F_k les événements correspondant respectivement à obtenir pile et face au k -ième lancer. On fait l'hypothèse habituelle d'indépendance de la famille $(P_k)_{k \geq 1}$, donc aussi de $(F_k)_{k \geq 1}$.

1. Exprimer A_n en fonction des événements $(P_k)_{k \geq 1}$ et $(F_k)_{k \geq 1}$.
2. Déterminer les valeurs de p_2 et p_3 . On utilisera les événements précédents et on justifiera les calculs.
3. Montrer que, pour tout $n \geq 4$, on a

$$p_n = \frac{2}{9} p_{n-2} + \frac{1}{3} p_{n-1}.$$

On fera intervenir le système complet d'événements constitué de $P_1 \cap P_2$, $P_1 \cap F_2$, F_1 , et on utilisera la formule des probabilités totales. On donnera en particulier les probabilités conditionnelles de A_n sachant ces 3 événements.

4. En déduire une expression explicite de p_n pour tout $n \geq 2$.
5. Calculer $\sum_{n=2}^{+\infty} p_n$ et interpréter le résultat.

Exercice 25. On estime que 40% des e-mails sont des spams. Un logiciel se charge de filtrer vos messages. Celui-ci détecte correctement 99% des spams. Cependant, il y a 5% de chance qu'un message non spam soit classé à tort comme spam.

Vous recevez un message. Considérons les événements suivants :

- A : « le message reçu est classé comme spam par votre messagerie » ;
- B : « le message reçu est vraiment un spam ».

1. Donner les probabilités $\mathbb{P}(B)$, $\mathbb{P}(A|B)$, $\mathbb{P}(B^c)$ et $\mathbb{P}(A|B^c)$.
2. Calculer $\mathbb{P}(A)$. On justifiera soigneusement.
3. Quelle est la probabilité qu'un message classé comme spam le soit à tort ?

Exercice 26 (© Lê Nguyễn Hoàng). Un ami vous informe qu'il fait partie d'une fratrie de deux enfants et qu'il est né un mardi. Quelle est la probabilité qu'il ait un frère ? On modélise l'expérience de la manière suivante. On note G_1 un garçon né un mardi, G_2 un garçon né un autre jour, et F une fille. On pose alors :

$$\Omega = \{G_1G_1, G_1G_2, G_2G_1, G_2G_2, G_1F, G_2F, FG_1, FG_2, FF\}.$$

Par exemple, $\{G_1G_2\}$ représente l'événement « le premier enfant est un garçon né un mardi et le second enfant est un garçon né un autre jour », et $\{FG_2\}$ l'événement « le premier enfant est une fille et le second enfant est un garçon qui n'est pas né un mardi ».

1. Donner les probabilités de chaque événement élémentaire. Par exemple, on vérifiera que $\mathbb{P}(\{G_2F\}) = 3/14$. On fera bien apparaître les hypothèses d'indépendance et d'uniformité dans les calculs.
2. Soit K : « le premier enfant est un garçon », L : « le second enfant est une fille », et M : « le premier enfant est un garçon né un mardi ». Expliciter K , L , M comme sous-ensembles de Ω et vérifier que :

$$\mathbb{P}(K) = \frac{1}{2}, \quad \mathbb{P}(L) = \frac{1}{2}, \quad \mathbb{P}(K \cap L) = \frac{1}{4}, \quad \mathbb{P}(M|K) = \frac{1}{7}.$$

3. Soient A : « votre ami a un frère » et B : « dans sa famille, il y a au moins un garçon né un mardi ».

(a) Expliciter A , B et $A \cap B$.

(b) Calculer les probabilités de ces événements.

(c) En déduire que la probabilité cherchée est $\frac{13}{27} \simeq 48\%$.

Exercice 27. Une maladie rare touche 1 personne sur 100 000. Un test permet de détecter 99% des personnes malades, mais il est aussi positif dans 1% des cas pour des personnes saines. Quelle est la probabilité que vous soyez malade sachant que votre test est positif ? On modélisera soigneusement les données dans le langage des probabilités conditionnelles.

Exercice 28. Une compagnie d'assurance répartit ses clients en trois classes R_1 , R_2 et R_3 : les bons risques, les risques moyens, et les mauvais risques. Ces trois classes représentent respectivement 20%, 50% et 30% de la population totale. Les probabilités d'avoir un accident au cours de l'année dans chacune de ces classes sont respectivement 0,05, 0,15 et 0,30.

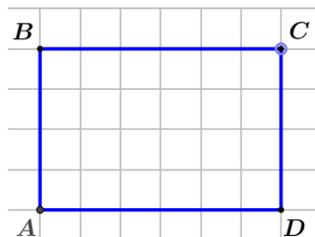
1. Quelle est la probabilité qu'une personne choisie au hasard dans la population ait un accident dans l'année ?
2. Si M. Martin n'a pas eu d'accident cette année, quelle est la probabilité qu'il soit un bon risque ?

Exercice 29. Deux tireurs, Alice et Bob, tirent une fois au but chacun, indépendamment. La probabilité pour Alice de marquer est égale à 0,8 contre 0,4 pour Bob. On note A l'événement « Alice marque », B l'événement « Bob marque » et S l'événement « un seul but est marqué ».

1. Calculer $\mathbb{P}(S)$. On justifiera soigneusement.
2. Vous êtes allé aux toilettes pendant cette séance de tirs. À votre retour, vous constatez qu'il n'y a eu qu'un seul but. Avec quelle probabilité a-t-il été marqué par Alice ?

7.3 Les variables aléatoires et leurs lois

Exercice 30. Vous décidez de tracer un rectangle $ABCD$ au hasard sur votre feuille. On note X la longueur AB , Y la longueur AD et A l'aire du rectangle. On suppose que X est une variable aléatoire de loi uniforme sur $\{1, \dots, 10\}$ et que Y est une v.a. de loi uniforme sur $\{1, \dots, 20\}$ indépendante de X .



1. Quel était la probabilité d'obtenir le rectangle dessiner ci dessus ?
2. Calculer $\mathbb{P}(A = 72)$ et $\mathbb{P}(X = 9|A = 72)$.
3. Décrire l'événement E : « le rectangle est carré » en fonction des variables aléatoires X et Y et calculer $\mathbb{P}(E)$.

Exercice 31. Soient $\Omega = \{0, 1\} \times \{0, 1\}$ munit de l'ensemble de ses parties $\mathcal{F} = \mathcal{P}(\Omega)$ et de la probabilité uniforme \mathbb{P} . On pose pour tout $\omega = (\omega_1, \omega_2) \in \Omega$, $X_1(\omega) = \omega_1$ et $X_2(\omega) = \omega_2$.

1. On suppose que \mathbb{P} est la loi uniforme.
 - (a) Montrer que X_1 et X_2 sont de loi de Bernoulli de paramètre $1/2$.
 - (b) Montrer que X_1 et X_2 sont indépendantes.
2. Soit $-1 \leq \rho \leq 1$. On suppose maintenant que

$$\mathbb{P} = \frac{1+\rho}{4}\delta_{(1,1)} + \frac{1-\rho}{4}\delta_{(1,0)} + \frac{1-\rho}{4}\delta_{(0,1)} + \frac{1+\rho}{4}\delta_{(0,0)}.$$

- (a) Montrer que X_1 et X_2 sont toujours de loi de Bernoulli de paramètre $1/2$.
- (b) Montrer que X_1 et X_2 ne sont plus indépendantes sauf si $\rho = 0$.
- (c) Calculer alors $\mathbb{P}(X_2 = 1|X_1 = 1)$.

Exercice 32. Soit $\alpha > 0$ et X une variable aléatoire à valeur dans $\{-(n-1), \dots, n-1\}$ telle que

$$\mathbb{P}(X = k) = \begin{cases} \alpha(n-k) & \text{si } k \in \{0, \dots, n-1\}, \\ \alpha(n+k) & \text{si } k \in \{-(n-1), \dots, 0\}. \end{cases}$$

Déterminer α .

Exercice 33. Soient X_1 et X_2 deux variables aléatoires indépendantes toutes deux de loi uniforme sur $\{0, 1, \dots, n-1\}$. Déterminer la loi de $T = X_1 - X_2$. On justifiera soigneusement chaque étape des calculs.

Exercice 34. Déterminer la loi de la somme de deux variables aléatoires indépendantes de loi uniforme sur $\{1, \dots, n\}$. On fera une application au lancer de dés ($n = 6$) où l'on représentera la loi par un diagramme en batons.

Exercice 35. Soit X, Y iid de loi uniforme sur $0, \dots, n$. Calculer $P(X=Y)$ et $P(X<Y)$.

Exercice 36. Soit X une v.a.r. suivant la loi $\mathcal{B}(n, p)$. On pose $Y = n - X$.

1. Déterminer la loi de Y .
2. Les variables X et Y sont-elles indépendantes ?

Exercice 37. Montrer que $X_1 + X_2 \sim \mathcal{B}(n_1 + n_2, p)$ si $X_1 \sim \mathcal{B}(n_1, p)$ et $X_2 \sim \mathcal{B}(n_2, p)$ avec X_1 et X_2 indépendantes.

Exercice 38. Soit $X \sim \mathcal{G}(p)$.

1. Calculer $\mathbb{P}(X > n)$ pour tout $n \in \mathbb{N}$.
2. Montrer si Z est une v.a. à valeurs dans \mathbb{N}^* telle que pour tout $n \in \mathbb{N}$ on a $P(Y > n) = \alpha^n$ alors Y est une v.a. de loi géométrique dont on précisera le paramètre.
3. Soit Y une v.a. geo de param q indépendante de X et $Z = \min(X, Y)$. Montrer que Y est une v.a. de loi geo et préciser le param.
4. Soit X, Y iid de loi $G(p)$. calculer $P(X=Y)$ et $P(X<Y)$. On pourra dans le deuxième cas utiliser la formule de prob totale appliqué au sys complet $X=n, n \text{ in } \mathbb{N}^*$.
5. Montrer que pour tout $n, m > 0$ on a

$$\mathbb{P}(X > n + m | X > m) = \mathbb{P}(X > n).$$

On appelle cette propriété l'absence de mémoire.

Exercice 39. Soient X, Y deux variables aléatoires indépendantes de loi géométrique de paramètre p . On pose $S = X + Y$.

1. Donner le support de S (le plus petit ensemble E tel que $\mathbb{P}(S \in E) = 1$).
2. Déterminer la loi de $S = X + Y$.
3. Application : vous lancez un dé jusqu'à ce que vous ayez fait deux 6 en tout pour la première fois. Avec quelle probabilité vous arrêtez vous après 2 lancers ? 3 lancers ?

Exercice 40. Soit $Y \sim \mathcal{P}(\lambda)$ et $X_n \sim \mathcal{B}(n, p_n)$ tel que $np_n \rightarrow \lambda$ lorsque $n \rightarrow \infty$. Montrer que pour tout $k \in \mathbb{N}$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = \mathbb{P}(Y = k).$$

Exercice 41. Montrer que $X_1 + X_2 \sim \mathcal{P}(\lambda + \mu)$ si $X_1 \sim \mathcal{P}(\lambda)$ et $X_2 \sim \mathcal{P}(\mu)$ avec X_1 et X_2 indépendantes.

Exercice 42. Soit X une va de poisson de param λ et Y une autre va de poisson indépendante de X de paramètre μ . Calculer la loi de X sachant $X + Y$ ie $P(X=k | X+Y=n)$ pour tout $k, n \geq 0$.

Exercice 43. Un insecte pond des œufs. Le nombre d'œufs pondus est une variable aléatoire X suivant une loi de Poisson $\mathcal{P}(\lambda)$.

Chaque œuf a une probabilité p d'éclore, indépendamment des autres. Soit Z le nombre d'œufs qui ont éclos.

1. Pour $(k, n) \in \mathbb{N}^2$, calculer $\mathbb{P}(Z = k \mid X = n)$.
2. En déduire la loi de Z .
3. Quelle est l'espérance de Z ?

Exercice 44. Soit $(X_k)_{k \geq 1}$ une suite de variable aléatoire indépendantes de loi de Bernoulli de paramètre p et N une variable aléatoire de loi de Poisson de paramètre λ indépendante de $(X_k)_{k \geq 1}$. On pose alors

$$S = \sum_{k=1}^N X_k \quad (S = 0 \text{ si } N = 0).$$

Par exemple si $N = 3$ et $(X_1, X_2, X_3) = (1, 0, 1)$ alors $S = 1 + 0 + 1 = 2$.

1. Justifier que le support de (S, N) est $\{(s, n) \in \mathbb{N} \times \mathbb{N} : s \leq n\}$.
2. Justifier que $\{S = s, N = n\} = \{\sum_{k=1}^n X_k = s, N = n\}$.
3. En déduire la loi de (S, N) .
4. Puis la loi de S . On utilisera le système complet d'événements $(\{N = n\})_{n \geq 0}$.
5. Donner la loi conditionnelle de S sachant $N = n$.

Exercice 45. Soit a et $\lambda > 0$ et X et Y des v.a. à valeurs dans \mathbb{N} tels que pour tout i, j dans \mathbb{N} on a

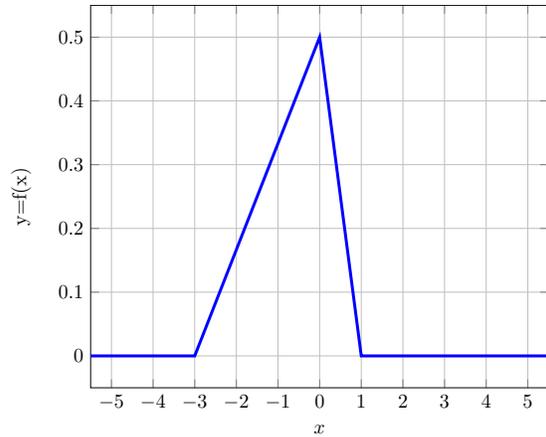
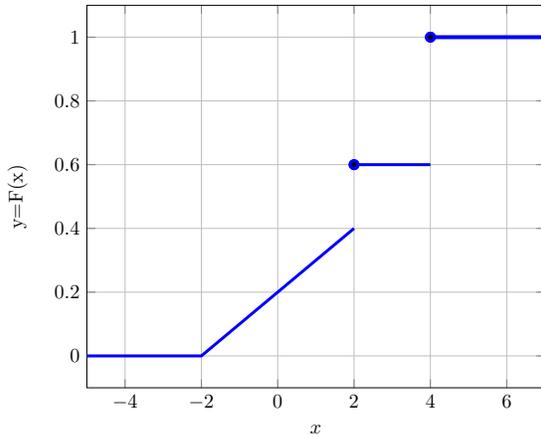
$$\mathbb{P}(X = i, Y = j) = a \frac{(i + j)\lambda^{i+j}}{i!j!}$$

1. Déterminer a en fonction de λ
2. Déterminer les lois marginales de X et Y .
3. Les variables sont elles indépendantes ?

Exercice 46. Soit X une variable aléatoire réelle discrète de fonction de répartition $F(x) = 0$ si $x < -1$, $F(x) = 1/3$ si $-1 \leq x < 1$ et $F(x) = 1$ si $x \geq 1$.

1. Tracer F et déterminer la loi de X .
2. Montrer que $2Y - 1$ a la même loi que X où Y est une variable aléatoire de loi de Bernoulli de paramètre $2/3$.

Exercice 47. Les figures représentent le graphe de la fonction de répartition F d'une variable aléatoire X (à gauche) et le graphe de la densité f d'une variable aléatoire Y (à droite).



Déterminer les probabilités suivantes : **i)** $\mathbb{P}(X = 0)$ **ii)** $\mathbb{P}(X \geq 0)$ **iii)** $\mathbb{P}(-1 < X \leq 2)$
iv) $\mathbb{P}(-1 < Y < 2)$ **v)** $\mathbb{P}(Y \geq 0)$.

Exercice 48. Soit F la fonction définie par

$$F(x) = \begin{cases} 0 & \text{si } x < -1, \\ \frac{x+1}{4} & \text{si } -1 \leq x < 1, \\ \frac{2}{3} & \text{si } 1 \leq x < 2, \\ 1 & \text{si } x \geq 2. \end{cases}$$

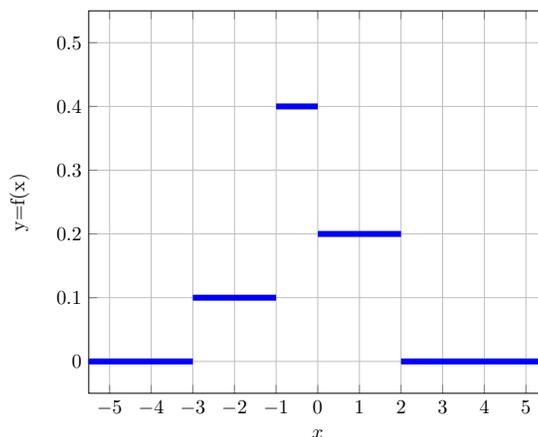
1. Justifier que F est une fonction de répartition.

2. Soit Z une v.a.r. de fonction de répartition F . Calculer les probabilités suivantes :

a) $\mathbb{P}(Z \leq 0)$ b) $\mathbb{P}(Z = 0)$ c) $\mathbb{P}\left(Z > \frac{3}{2}\right)$ d) $\mathbb{P}(0 \leq Z < 2)$ et e) $\mathbb{P}(Z = 1)$.

Exercice 49. Soit f une densité de probabilité de la forme $f(x) = \alpha(R - |x|)\mathbf{1}_{-R < x < R}$. Dessiner f et déterminer la constante de proportionnalité α .

Exercice 50. Soit X une v.a. réelle de densité f donnée graphiquement ci-dessous.



Calculer $\mathbb{P}(X \leq -1)$, $\mathbb{P}(X > 1)$, $\mathbb{P}(-2 \leq X < 0.5)$, $\mathbb{P}(X = 2)$ et tracer sa fonction de répartition.

Exercice 51. On pose $F(x) = 1 - e^{-\lambda x} - e^{-\mu x} + e^{-(\lambda+\mu)x}$ pour tout $x \geq 0$ et 0 sinon. Les paramètres λ, μ sont deux réels strictement positifs.

1. Montrer que F est une fonction de répartition.
2. Justifier que la loi associée admet bien une densité et la calculer.

Exercice 52. On pose $F(x) = 0$ pour tout $x < 0$, $F(x) = x/4$ si $0 \leq x < 1$, $F(x) = 1/2 - (x-1)/4$ si $1 \leq x < 2$, $F(x) = 11/12$ si $2 \leq x < 3$ et $F(x) = 1$ sinon.

1. Montrer que F est une fonction de répartition.
2. Tracer F . La loi a-t-elle une densité ? Justifier.
3. Soit X de fonction de répartition F .
 - (a) Calculer $\mathbb{P}(X = 1)$, $\mathbb{P}(X = 2)$ et $\mathbb{P}(X = 3)$.
 - (b) Calculer $\mathbb{P}(1/2 < X < 3/2)$

Exercice 53. Soit T une variable aléatoire de loi exponentielle de paramètre λ . On pose $X = \lfloor T \rfloor + 1$ où $\lfloor \cdot \rfloor$ est la partie entière.

1. Justifier que X est une variable aléatoire discrète et préciser son support.
2. Déterminer la loi de X .

Exercice 54. Soit U de loi $\mathcal{U}([0, 1])$. On pose $T = \ln(U)$.

1. Justifier que T est bien définie presque sûrement.
2. Déterminer sa fonction de répartition.
3. En déduire sa loi.

Exercice 55. Soit $T \sim \mathcal{E}(\lambda)$. Montrer que pour tout $t, s > 0$ on a

$$\mathbb{P}(T > s + t | T > s) = \mathbb{P}(T > t).$$

Exercice 56. Soit X et Y deux variables aléatoires telles que $X \sim \mathcal{G}(p)$ et pour tout intervalle $I \subset \mathbb{R}$ et $n \in \mathbb{N}^*$,

$$\mathbb{P}(Y \in I | X = n) = \int_I n e^{-nt} \mathbb{1}_{\{t>0\}} dt$$

On dit que la loi de Y sachant $X = n$ est $\mathcal{E}(n)$.

1. Que vaut $\mathbb{P}(Y > y | X = n)$ lorsque $y > 0$?
2. En déduire que

$$\mathbb{P}(Y > y) = \frac{p}{e^y - (1 - p)},$$

puis que Y a pour densité de probabilité

$$f_Y(y) = \frac{pe^y}{(e^y - (1 - p))^2} \mathbb{1}_{\{y>0\}}.$$

3. Procéder de même lorsque $X \sim \mathcal{P}(\lambda)$ et montrer que la fonction de répartition et la densité sont données par

$$F_Y(y) = 1 - e^{\lambda(e^{-y}-1)} \mathbb{1}_{\{y>0\}} \quad \text{et} \quad f_Y(y) = \lambda e^{-y} e^{\lambda(e^{-y}-1)} \mathbb{1}_{y>0}.$$

Exercice 57. Exo montrant que si $X \sim N(0, 1)$ alors $Y = \sigma X + m$ suit $N(m, \sigma^2)$ via fonction de répartition et unicité de celle-ci

7.4 Espérance des variables aléatoires

Exercice 58. Soit X une variable aléatoire réelle de fonction de répartition F définie par :

$$F(x) = \begin{cases} 0 & \text{si } x < -1, \\ \frac{1}{2} & \text{si } -1 \leq x < 0, \\ \frac{2}{3} & \text{si } 0 \leq x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

1. Représenter graphiquement F et déterminer la loi de X .
2. Calculer $\mathbb{E}(X)$ et $\mathbb{V}(X)$.

Exercice 59. Soit X une variable aléatoire discrète à valeurs dans $\{0, \dots, N\}$.

1. Montrer que

$$\mathbb{E}[X] = \sum_{n=0}^{N-1} \mathbb{P}(X > n).$$

On pourra commencer par montrer que $X = \sum_{k=0}^{N-1} \mathbb{1}_{\{X > k\}}$.

2. Dans une urne contenant N boules numérotées de 1 à N , on tire avec remise n boules successivement. On note X le numéro de la plus grande boule tirée.
 - (a) Calculer $\mathbb{P}(X > k)$.
 - (b) En déduire $\mathbb{E}[X]$.

Exercice 60. Soit X une variable aléatoire uniforme sur $\{0, \dots, n-1\}$. Calculer $\mathbb{E}[X]$ et $\mathbb{V}(X)$. On rappelle la formule suivante :

$$\sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}.$$

Exercice 61. Soient X et Y deux variables aléatoires à valeurs dans \mathbb{N} . On suppose que la loi conjointe de (X, Y) est donnée par :

$$\forall (j, k) \in \mathbb{N}^2, \quad \mathbb{P}(X = j, Y = k) = a \cdot \frac{j+k}{2^{j+k}} \quad \text{avec } a \in \mathbb{R}.$$

1. Déterminer la valeur de a .
2. Déterminer les lois marginales de X et Y .
3. Les variables X et Y sont-elles indépendantes ?
4. Calculer $\mathbb{P}(X = Y)$.

Exercice 62. Soit X une variable aléatoire suivant une loi binomiale de paramètres $n \geq 1$ et $p \in]0, 1[$. Calculer l'espérance de la variable aléatoire

$$Y = \frac{1}{X + 1}.$$

Exercice 63. On réalise une suite de lancers indépendants d'une pièce ayant une probabilité $p \in]0, 1[$ de tomber côté « Pile ». On note X la longueur de la première série de lancers identiques, et Y la longueur de la seconde série. Par exemple, les successions de lancers « PPFPP... » et « FFPPPF... » correspondent à $X = 2$ et $Y = 3$.

1. Déterminer la loi conjointe du couple (X, Y) .
2. Déterminer les lois marginales.
3. Calculer les espérances de X et Y .

Exercice 64. Vous lancez deux dés de manière répétée. En moyenne, combien de lancers faites-vous avant de voir un double six pour la première fois ? Même question pour l'observation de n'importe quel double.

Exercice 65. Soient U et V les scores de deux dés indépendants. On note $X = \min(U, V)$ et $Y = \max(U, V)$, respectivement le minimum et le maximum des scores.

1. Déterminer les lois de X et Y .
2. Calculer $\mathbb{E}[X]$ par calcul direct en utilisant la loi de X .
3. Calculer $\mathbb{E}[\min(U, V)]$ en appliquant le théorème de transfert à la variable aléatoire discrète (U, V) dont la loi est donnée, pour tout $(i, j) \in \{1, \dots, 6\}^2$, par :

$$\mathbb{P}((U, V) = (i, j)) = \frac{1}{36}.$$

4. En remarquant que $X + Y = U + V$, en déduire l'espérance de Y .

Exercice 66. À la sortie d'un concert, le portier remet à n personnes leurs chapeaux pris au hasard. Les clients sont numérotés de 1 à n : pour le client i , on désigne par X_i la variable aléatoire valant 1 si le chapeau distribué est le bon, et 0 sinon. Le nombre de chapeaux correctement distribués est donc $S = X_1 + \dots + X_n$.

1. Donner la loi de chaque X_i .
2. En déduire l'espérance de S .
3. Calculer $\mathbb{P}(X_i = 1, X_j = 1)$ puis $\mathbb{E}[X_i X_j]$ pour $i \neq j$.
4. En déduire la variance de S . Est-elle égale à la somme des variances des X_i ?

Exercice 67. Soit X une variable aléatoire admettant un moment d'ordre 2.

Démontrer que $a \mapsto \mathbb{E}[(X - a)^2]$ est minimale pour $a = \mathbb{E}[X]$.

Exercice 68. Soient X_1, \dots, X_n des variables aléatoires discrètes. On note C la matrice carrée de taille n définie par $C_{i,j} = \text{Cov}(X_i, X_j)$ pour tout $1 \leq i, j \leq n$.

1. Montrer que $C \in \mathcal{S}_n^+$, c'est-à-dire que C est une matrice symétrique définie positive.
2. Que peut-on dire si C n'est pas inversible ?

Exercice 69. Soit M une matrice aléatoire de taille $n \times n$ dont les coefficients $M_{i,j}$ sont des variables aléatoires discrètes, indépendantes, identiquement distribuées, centrées et admettant un moment d'ordre 2. On pose $Y = \det(M)$.

1. Montrer, en utilisant le développement (formule de Leibniz), que $\mathbb{E}[Y] = 0$.
2. Soient Z_1, \dots, Z_N , $N \in \mathbb{N}^*$, des variables aléatoires discrètes admettant un moment d'ordre 2. Exprimer $\mathbb{V}(Z_1 + \dots + Z_N)$ en fonction des $\mathbb{E}[Z_i Z_j]$, pour $1 \leq i, j \leq N$.
3. Toujours en utilisant la formule de Leibniz, calculer la variance de Y .

Exercice 70. On considère une expérience aléatoire ayant une probabilité $p > 0$ de succès et $1 - p$ d'échec. On répète l'expérience indépendamment jusqu'à l'obtention de m succès, et on note T_m le nombre d'essais nécessaires à l'obtention de ces m succès.

1. Reconnaître la loi de T_1 et donner sa fonction génératrice.
2. Justifier que $T_m = \tau_1 + \dots + \tau_m$ avec $(\tau_i)_{1 \leq i \leq m}$ i.i.d. dont on précisera la loi. Que représente τ_k ?
3. En déduire la fonction génératrice de T_m et son espérance.
4. Exprimer le développement en série entière de $\frac{1}{(1-t)^m}$.
5. En déduire la loi de T_m .

Exercice 71. Montrer, à l'aide des fonctions génératrices, qu'il est impossible de « truquer » deux dés cubiques et indépendants pour que la somme d'un lancer suive une loi uniforme sur $\{2, \dots, 12\}$. On pourra raisonner par l'absurde et étudier les racines des fonctions génératrices.

Exercice 72. Soit f une densité de probabilité définie par :

$$f(x) = \begin{cases} \alpha |x|^\beta & \text{si } -1 \leq x \leq 1, \\ 0 & \text{sinon,} \end{cases} \quad \text{avec } \alpha > 0 \text{ et } \beta \geq 0.$$

1. Exprimer α en fonction de β .
2. Calculer et tracer la fonction de répartition associée.

Exercice 73. Calculer $\mathbb{E}[e^{itX}]$ lorsque :

- $X \sim \mathcal{U}([0, 1])$;
- $X \sim \mathcal{E}(\lambda)$.

Exercice 74. Soit $T \sim \mathcal{E}(\lambda)$.

1. Que vaut $\mathbb{P}(T > t)$? Quel est le lien avec la fonction de répartition ?
2. Calculer $\mathbb{E}[T]$ et $\mathbb{V}[T]$.
3. Soient T_1 et T_2 deux variables aléatoires indépendantes, de lois respectives $\mathcal{E}(\lambda_1)$ et $\mathcal{E}(\lambda_2)$. Montrer, en calculant sa fonction de répartition, que le minimum $T_1 \wedge T_2$ suit encore une loi exponentielle (dont on précisera le paramètre).

Exercice 75. Soit $X \sim \mathcal{E}(\lambda)$. Calculer le moment d'ordre n de X , c'est-à-dire $\mathbb{E}[X^n]$, pour tout $n \in \mathbb{N}^*$. On pourra procéder par récurrence en utilisant une intégration par parties.

Exercice 76. Soit f la fonction définie sur \mathbb{R} par $f(x) = \frac{a}{1+x^2}$.

1. Déterminer a pour que f soit la densité de probabilité d'une variable X .
2. Déterminer la fonction de répartition de X .
3. X admet-elle une espérance ?
4. Quelle est la loi de $Y = 1/X$?

7.5 Inégalités de concentration

Exercice 77. Soit X une variable aléatoire.

1. Montrer que tout $a > 0$ et tout $p \geq 0$,

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|^p]}{a^p}.$$

2. Montrer que pour tout $\lambda > 0$,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda a}}.$$

Exercice 78. Soit $(X_k)_{k \geq 1}$ une suite de variables aléatoires mutuellement indépendantes telles que, pour tout $k \in \mathbb{N}^*$,

$$\mathbb{P}(X_k = 1) = \mathbb{P}(X_k = -1) = \frac{1}{2}.$$

On pose :

$$S_n = \sum_{k=1}^n X_k, \quad n \in \mathbb{N}^*.$$

1. Pour tout $t \in \mathbb{R}$, calculer $\mathbb{E}(e^{tS_n})$.
2. En déduire que, pour tout $n \in \mathbb{N}^*$ et tout $\lambda \in \mathbb{R}_+$,

$$\mathbb{P}(|S_n| \geq \lambda) \leq 2 \exp\left(-\frac{\lambda^2}{2n}\right).$$

On pourra utiliser, sans la démontrer, l'inégalité :

$$\cosh(u) \leq \exp\left(\frac{u^2}{2}\right).$$

3. Soit $c > 1$. Pour tout $n \in \mathbb{N}^*$, on note : $A_n = \left\{ |S_n| \leq \sqrt{2cn \ln n} \right\}$.
Démontrer que :

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

Exercice 79. On jette 3600 fois un dé équilibré à 6 faces. Minorer la probabilité que le nombre d'apparitions du numéro 1 soit compris entre 480 et 720. On pourra appliquer l'inégalité de Bienaymé–Tchebychev.

Exercice 80. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires deux à deux indépendantes, avec X_n suivant une loi de Bernoulli de paramètre p_n . Montrer que, pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n p_i\right| < \varepsilon\right) = 1.$$

7.6 Graphes et chaînes de Markov

Exercice 81. Une puce se déplace sur un triangle de la façon suivante : si elle est sur un sommet, elle se déplace de façon équiprobable vers l'un des deux autres sommets (elle ne peut rester sur place). On note les sommets 1, 2 et 3.

1. Donner la matrice de transition A associée à cette marche aléatoire.
2. Montrer que, pour tout entier $n \geq 0$, on a :

$$A^n = \begin{pmatrix} u_n & v_n & v_n \\ v_n & u_n & v_n \\ v_n & v_n & u_n \end{pmatrix} \quad \text{avec} \quad u_n = \frac{1}{3} \left(1 - \left(-\frac{1}{2} \right)^{n-1} \right), \quad v_n = \frac{1}{3} \left(1 - \left(-\frac{1}{2} \right)^n \right).$$

3. En déduire que, quel que soit l'état initial, la suite

$$\lim_{n \rightarrow \infty} (\mathbb{P}(X_n = 1), \mathbb{P}(X_n = 2), \mathbb{P}(X_n = 3)) = (1/3, 1/3, 1/3).$$

4. Vérifier que $\mu = (1/3, 1/3, 1/3)$ est une probabilité invariante de la chaîne.

Exercice 82. On considère une chaîne de Markov à trois états, dont la matrice de transition est donnée par

$$P = \begin{pmatrix} 2/5 & 3/5 & 0 \\ 1/5 & 1/2 & 3/10 \\ 0 & 2/5 & 3/5 \end{pmatrix}.$$

1. Donner une représentation graphique de l'espace d'états et des transitions.
2. Vérifier que la chaîne est irréductible.
3. Calculer sa probabilité invariante.
4. Vérifier que 0, 1 et 1/2 sont des valeurs propres de la matrice de transition, et déterminer les espaces propres associés.
5. En déduire que la chaîne converge vers sa probabilité invariante.

Exercice 83 (© Thomas Cabaret). Montrer que si l'on numérote les cases de l'échiquier en Figure 7.1 et que l'on dit que deux cases sont adjacentes si et seulement si on peut passer de l'une à l'autre avec un cavalier (déplacement en L), alors on obtient le graphe non orienté en Figure 7.2. **Application.** Montrer qu'en partant de la configuration des cavaliers en Figure 7.1, on peut bouger les cavaliers un par un, de sorte qu'à la fin les cavaliers blancs occupent les places des cavaliers noirs, et inversement.

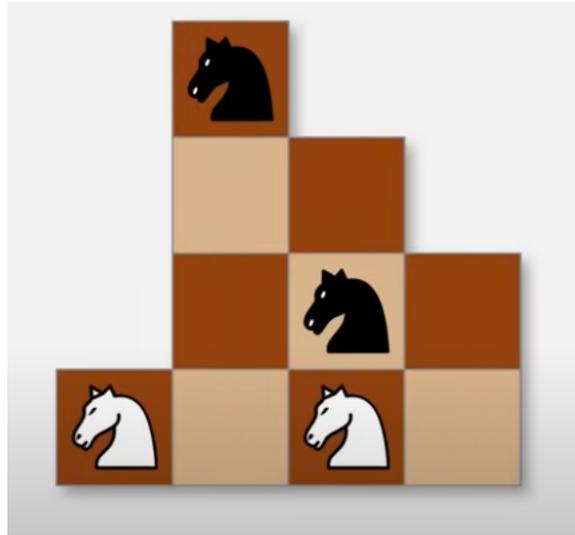


FIGURE 7.1 – Position de départ sur l'échiquier



FIGURE 7.2 – Graphe associé aux mouvements du cavalier