ENSEIGNEMENT SUPÉRIEUR ET UNIVERSITAIRE

UNIVERSITÉ OFFICIELLE DE BUKAVU

ÉCOLE DES MINES

Cours de Probabilité

Notes de cours destinées aux étudiants de deuxième année de licence (bachelier) en mines, pétrole et génie civil

AMBO AMANDURE Jean-Médard, Chef de Travaux

« *On ne peut plus expliquer le monde, faire ressentir sa beauté à ceux qui n'ont aucune connaissance profonde des mathématiques* » (Richard Feynman)

Année académique 2022-2023

DESCRIPTION DE L'UNITÉ D'ENSEIGNEMENT (Année Académique 2022-2023)		
Unité de rattachement de l'UE (Faculté)	ÉCOLE DES MINES (Université Officielle de Bukavu)	
Code de l'UE		
Titre de l'UE	Probabilités et Statistique (L2 LMD Mines, Pétrole et Génie Civil)	
Nombre de crédits	4	60h=30h (Théorie)+ 30h (Pratique)
Semestre	S1	
Éléments constitutifs (EC) de l'UE :	1. Probabilités (2 Crédits) 2. Statistique inférentielle (2 Crédits)	
Enseignants	1. CT. AMBO AMANDURE Jean-Médard pour la partie probabilité 2. Un autre enseignant (à fixer par l'école des mines) pour la partie statistique inférentielle	
Prérequis	Mathématiques générales (analyse, algèbre, ...)	
Objectif(s) du cours	Cette UE permet à l'étudiant de s'approprier des notions, principes et techniques de probabilités et statistique afin de les appliquer en ingénierie.	
Contenu (partie : Probabilité)	1. Notions de base 1.1. Introduction : Qu'est-ce que la probabilité ? 1.2. Rappel de quelques notions de la théorie des ensembles 1.3. Expériences aléatoires et probabilités 1.4. Probabilité conditionnelle 2. Variables aléatoires 2.1. Notions, définitions et exemples 2.2. Variables aléatoires discrètes 2.2. Variables aléatoires continues 3. Théorèmes limites 3.1. Loi des grands nombres 3.2. Théorème central limite	
Compétences professionnelles et disciplinaires visées	1. Résoudre des problèmes (MA1) 2. Modéliser des situations (MA2) 3. Calculer dans des contextes divers (MA3) 4. Communiquer à l'aide du langage mathématique (MA4) 5. Raisonner (MA5)	
Approches pédagogiques	Cours magistraux interactifs, Utilisation pédagogique des TIC (logiciels de présentation, tableurs, logiciels de simulation, etc.), TP et TD.	
Modalités d'évaluation	Contrôle continu (exercices de synthèse, de démonstration, de calcul, de raisonnement, problèmes, exposés individuels ou en équipe), TP, TD, Interrogations, Examen écrit.	

Chapitre 1

Notions de base

Dans ce chapitre, nous présentons quelques concepts et définitions de base. Nous commençons par une brève discussion sur ce qu'est la probabilité. Puis nous passons en revue certaines bases mathématiques nécessaires au développement de la théorie des probabilités. Nous abordons ensuite les notions d'expériences aléatoires et les axiomes de la probabilité. Nous introduisons ensuite les modèles de probabilité discrets et continus. Enfin, nous discutons de la probabilité conditionnelle.

1.1 Introduction : Qu'est-ce que la probabilité ?

L'aléatoire et l'incertitude existent dans nos vies quotidiennes ainsi que dans toutes les disciplines de la science, de l'ingénierie et de la technologie. La théorie des probabilités est un cadre mathématique qui nous permet de décrire et d'analyser des phénomènes aléatoires dans le monde qui nous entoure. Par phénomènes aléatoires, nous entendons des événements ou des expériences dont nous ne pouvons prédire avec certitude les résultats.

Considérons quelques applications spécifiques de la probabilité afin d'obtenir une certaine intuition. Tout d'abord, réfléchissons plus attentivement à ce que nous entendons par les termes "aléatoire" et "probabilité" dans le contexte de l'une des expériences aléatoires les plus simples possibles : *lancer d'une pièce de monnaie équilibrée*.

Une façon de penser à l'« aléatoire » est que c'est une façon d'exprimer ce que nous ne savons pas. Peut-être que si nous en savions plus sur la force avec laquelle j'ai lancé la pièce, l'orientation initiale de la pièce, le point d'impact entre mon doigt et la pièce, la turbulence dans l'air, la douceur de la surface de la table sur laquelle la pièce atterrit, les matériaux caractéristiques de la pièce et de la table, etc., nous serions en mesure de dire avec certitude si la pièce tomberait pile ou face. Cependant, en l'absence de toutes ces informations, nous ne pouvons pas prédire le résultat du tirage au sort. Lorsque nous disons que quelque chose est aléatoire, nous disons que notre connaissance du résultat est limitée, nous ne pouvons donc pas être certains de ce qui se passera.

Étant donné que la pièce est équilibrée, si nous ne savons rien sur la façon dont elle a été lancée, la probabilité qu'elle tombe face est de 50%, ou $\frac{1}{2}$. Qu'entend-on exactement par là ? Il existe deux interprétations courantes du mot « probabilité ». L'un est en termes de **fréquence relative**. En d'autres termes, si nous lançons la pièce un très grand nombre de fois, elle tombera face pour environ $\frac{1}{2}$ de fois. Au fur et à mesure que le nombre de lancers de pièces augmente, la proportion de piles (ou faces) aura tendance à se rapprocher de plus en plus de $\frac{1}{2}$. En fait, cette compréhension intuitive de la probabilité est un cas particulier de la **loi des grands nombres**, que nous énoncerons formellement dans les chapitres ultérieurs.

Une deuxième interprétation de la probabilité est qu'il s'agit d'une quantification de notre degré de **croyance personnelle subjective** que quelque chose va se passer. Pour avoir une idée de ce que nous entendons par là, il peut être utile de considérer un deuxième exemple : la prévision du temps. Lorsque nous pensons aux chances qu'il pleuve aujourd'hui, nous considérons des choses comme s'il y a des nuages dans le ciel et l'humidité. Cependant, les croyances que nous formons sur la base de ces facteurs peuvent varier d'une personne à l'autre – différentes personnes peuvent faire des estimations différentes de la probabilité qu'il pleuve. Souvent, ces deux interprétations de la probabilité coïncident – par exemple, nous pouvons fonder nos croyances personnelles sur la probabilité qu'il pleuve sur une évaluation de la fréquence relative de la pluie les jours avec des conditions comme aujourd'hui. La probabilité est utilisée comme un outil majeur dans les sciences de l'ingénieur et dans beaucoup d'autres disciplines.

La beauté de la théorie des probabilités est qu'elle est applicable quelle que soit l'interprétation de la probabilité que nous utilisons (c'est-à-dire en termes de fréquence à long terme ou de degré de croyance). La théorie des probabilités fournit un cadre solide pour étudier les phénomènes aléatoires. Elle commence par supposer des axiomes de probabilité, puis construit toute la théorie à l'aide d'arguments mathématiques.

1.2 Rappel de quelques notions de la théorie des ensembles

La théorie des probabilités utilise le langage des ensembles. Comme nous le verrons plus tard, la probabilité est définie et calculée pour les ensembles. Ainsi, nous passons brièvement en revue ici quelques concepts de base de la théorie des ensembles qui sont utilisés dans ce cours. Nous discutons des notations d'ensemble, des définitions et des opérations (telles que les intersections et les unions). Nous introduisons ensuite les ensembles dénombrables et indénombrables. Enfin, nous discutons brièvement des fonctions. Cette section peut sembler quelque peu théorique et donc moins intéressante que le reste du cours, mais elle pose les bases de ce qui va suivre.

Un **ensemble** est une collection d' objets (éléments). On utilise souvent des majuscules pour désigner un ensemble. Pour définir un ensemble on peut simplement lister tous les éléments entre accolades, par exemple pour définir un ensemble A qui se compose de deux éléments $\clubsuit$ et $\diamondsuit$, nous écrivons $A = \{\clubsuit, \diamondsuit\}$. Dire que $\diamondsuit$ appartient à A, nous écrivons $\diamondsuit \in A$, où "$\in$" se prononce "appartient à". Pour dire qu'un élément n'appartient pas à un ensemble, on utilise $\notin$. Par exemple, nous pouvons écrire $\heartsuit \notin A$.

$$\boxed{\text{Un } \textbf{ensemble} \text{ est une collection d'objets (éléments).}}$$

Notez que l' *ordre n'a pas d'importance* dans un ensemble. Par exemple, les deux ensembles $\{\clubsuit, \diamondsuit\}$ et $\{\diamondsuit, \clubsuit\}$ sont égaux. Souvent on travaille avec des ensembles des nombres dont quelques uns sont listés dans l'exemple (1.1).

Exemple 1.1. Quelques ensembles utilisés dans ce cours :

- L'ensemble des nombres naturels, $\mathbb{N} = \{0, 1, 2, 3, \cdots\}$.
- L'ensemble des entiers, $\mathbb{Z} = \{\cdots, -3, -2, -1, 0, 1, 2, 3, \cdots\}$.
- L'ensemble des nombres rationnels $\mathbb{Q}$.
- L'ensemble des nombres réels $\mathbb{R}$.
- Intervalles fermés sur la ligne réelle. Par exemple, $[2, 3]$ est l'ensemble de tous les nombres réels x tel que $2 \leqslant x \leqslant 3$.
- Intervalles ouverts sur la ligne réelle. Par exemple $]-1, 3[$ est l'ensemble de tous les nombres réels x tel que $-1 < x < 3$.
- De même, $[1, 2[$ est l'ensemble de tous les nombres réels x tel que $1 \leqslant x < 2$.
- L'ensemble des nombres complexes $\mathbb{C}$ est l'ensemble des nombres sous la forme $a + b\mathrm{i}$, où $a, b \in \mathbb{R}$, et $\mathrm{i} = \sqrt{-1}$.

Nous pouvons également définir un ensemble en énonçant mathématiquement les propriétés satisfaites par les éléments de l'ensemble. En particulier, on peut écrire :

$A = \{x | x \text{ satisfait une propriété}\}$ ou $A = \{x : x \text{ satisfait une propriété}\}$

Les symboles "$|$" et " :" se prononcent "**tel que**".

Exemple 1.2. Voici quelques exemples d'ensembles définis par l' énoncé de leurs propriétés :

- Si l'ensemble C est défini comme $C = \{x | x \in \mathbb{Z}, -2 \leqslant x < 10\}$, alors $C = \{-2, -1, 0, \cdots, 9\}$.
- Si l'ensemble D est défini comme $D = \{x^2 | x \in \mathbb{N}\}$, alors $D = \{1, 4, 9, 16, \cdots\}$.
- L'ensemble des nombres rationnels peut être défini comme $\mathbb{Q} = \{\frac{a}{b} | a, b \in \mathbb{Z}, b \neq 0\}$.
- Pour les nombres réels a et b, où $a < b$, nous pouvons écrire $]a, b] = \{x \in \mathbb{R} \mid a < x \leqslant b\}$.
- $\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}, i = \sqrt{-1}\}$.

Un ensemble A est un **sous-ensemble** de l'ensemble B si chaque élément de A est aussi un élément de B. Nous écrivons $A \subset B$, où "$\subset$" indique "sous-ensemble". De manière équivalente, nous disons B est un **sur-ensemble** de A, ou $B \supset A$.

Exemple 1.3. Voici quelques exemples d'ensembles et de leurs sous-ensembles :

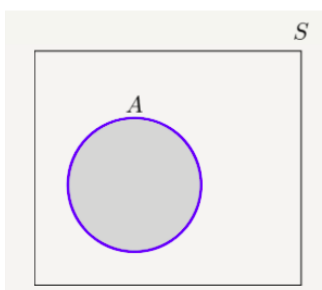
- Si $E = \{1,4\}$ et $C = \{1,4,9\}$, alors $E \subset C$.
- $\mathbb{N} \subset \mathbb{Z}$.
- $\mathbb{Q} \subset \mathbb{R}$.

Deux ensembles sont **égaux** s'ils ont exactement les mêmes éléments. Ainsi, $A = B$ si et seulement si $A \subset B$ et $B \subset A$. Par exemple, $\{1,2,3\} = \{3,2,1\}$, et $\{a,a,b\} = \{a,b\}$. L'ensemble sans élément, noté $\emptyset = \{\}$ est l' **ensemble vide**. Pour tout ensemble A, $\emptyset \subset A$.

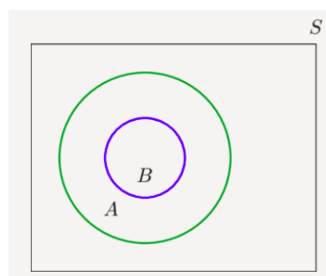
L' **ensemble fondamental** ou l' **univers** est l'ensemble de toutes les choses que nous pourrions envisager dans le contexte que nous étudions. Ainsi chaque ensemble A est un sous-ensemble de l' univers. Dans ce cours, nous désignons l'ensemble fondamental par S ou Ω. Par exemple, si nous lançons un dé, l' ensemble fondamental peut être défini comme $S = \{1,2,3,4,5,6\}$, ou si nous lançons une pièce de monnaie une fois, l' ensemble fondamental est $S = \{H,T\}$ (H pour face et T pour pile).

1.2.1 Diagrammes de Venn

Les diagrammes de Venn sont très utiles pour visualiser les relations entre les ensembles. Dans un diagramme de Venn, tout ensemble est représenté par une région fermée. La figure (1.1a) montre un exemple de diagramme de Venn. Dans cette figure, le grand rectangle montre l'univers S. La zone ombrée montre un autre ensemble A. La figure (1.1b) montre deux ensembles A et B, où $B \subset A$.



(a) Diagrammes de Venn.



(b) Diagramme de Venn pour deux ensembles A et B, où $B \subset A$.

FIGURE 1.1

1.2.2 Opérations sur les ensembles

Union

L' **union** de deux ensembles est un ensemble contenant tous les éléments qui sont dans A ou dans B (éventuellement les deux). Par exemple, $\{1,2\} \cup \{2,3\} = \{1,2,3\}$. Ainsi, nous pouvons écrire $x \in (A \cup B)$ si et seulement si $(x \in A)$ ou $(x \in B)$. Notez que $A \cup B = B \cup A$. Dans la figure (1.2), l'union des ensembles A et B est représenté par la zone ombrée dans le diagramme de Venn.

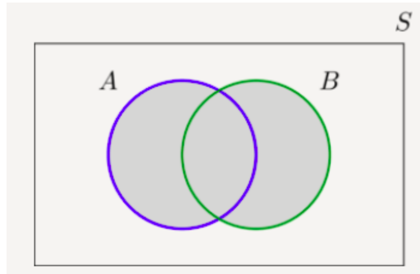


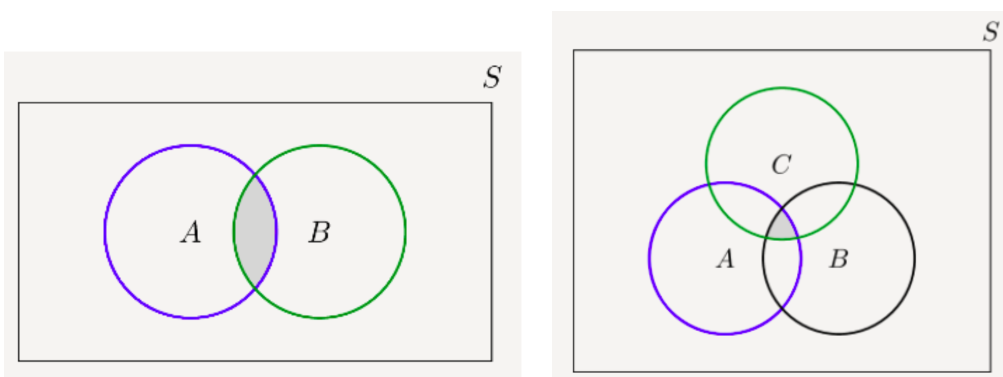
FIGURE 1.2 – Ensemble $A \cup B$.

De même, nous pouvons définir l'union de trois ensembles ou plus. En particulier, si A_1, A_2, A_3,$\cdots$, A_n sont n ensembles, leur union $A_1 \cup A_2 \cup A_3 \cdots \cup A_n$ est un ensemble contenant tous les éléments qui sont dans au moins un des ensembles. On peut écrire cette union de manière plus compacte par $\bigcup_{i=1}^{n} A_i$.

Par exemple, si $A_1 = \{a, b, c\}$, $A_2 = \{c, h\}$, $A_3 = \{a, d\}$, alors $\bigcup_i A_i = A_1 \cup A_2 \cup A_3 = \{a, b, c, h, d\}$. On peut de même définir l'union d'une infinité d'ensembles $A_1 \cup A_2 \cup A_3 \cup \cdots$.

Intersection

L' **intersection** de deux ensembles A et B, notée par $A \cap B$, se compose de tous les éléments qui sont à la fois dans A <u>et</u> B. Par exemple, $\{1, 2\} \cap \{2, 3\} = \{2\}$. Dans la figure (1.3a), l'intersection des ensembles A et B est représentée par la zone ombrée à l'aide d'un diagramme de Venn.



(a) Ensemble $A \cap B$.

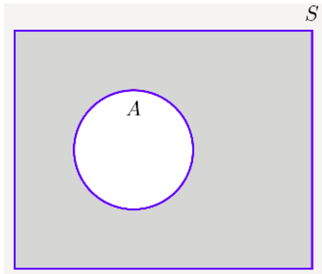
(b) Ensemble $A \cap B \cap C$.

FIGURE 1.3

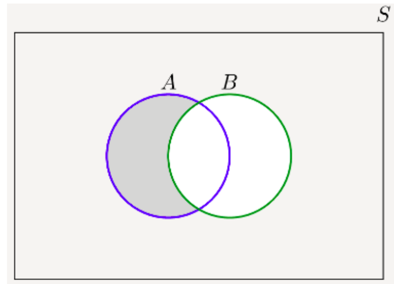
Plus généralement, pour les ensembles A_1, A_2, A_3, $\cdots$, leur intersection $\bigcap_i A_i$ est défini comme l'ensemble constitué des éléments qui sont dans tous A_i. La figure (1.3b) montre l'intersection de trois ensembles.

Complément

Le **complément** d'un ensemble A, noté par A^c ou $\bar{A}$, est l'ensemble de tous les éléments qui sont dans l'univers S mais ne sont pas dans A. Dans la Figure (1.4a), $\bar{A}$ est représenté par la zone ombrée à l'aide d'un diagramme de Venn.



(a) La zone grise montre l'ensemble $\bar{A} = A^c$.



(b) La zone grise montre l'ensemble $A - B$.

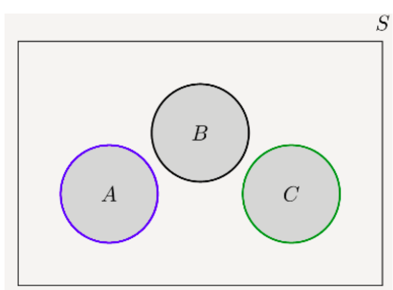
FIGURE 1.4

Différence

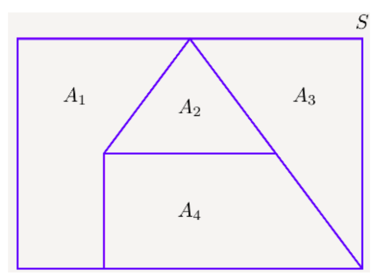
La **différence** (**soustraction**) est définie comme suit. L'ensemble $A - B$ $(A \setminus B)$ se compose d'éléments qui se trouvent dans A mais pas dans B. Par exemple si $A = \{1,2,3\}$ et $B = \{3,5\}$, alors $A - B = \{1,2\}$. Dans la Figure (1.4b), $A - B$ est représenté par la zone ombrée à l'aide d'un diagramme de Venn. Notez que $A - B = A \cap B^c$.

Ensembles disjoints et partition

Deux ensembles A et B sont mutuellement exclusifs ou disjoints s'ils n'ont pas d'éléments communs ; c'est-à-dire que leur intersection est l'ensemble vide, $A \cap B = \emptyset$. Plus généralement, plusieurs ensembles sont dits disjoints s'ils sont deux à deux disjoints, c'est-à-dire que deux d'entre eux ne partagent pas un élément commun. La figure (1.5a) montre trois ensembles disjoints.



(a) Les ensembles A, B, et C sont disjoints.



(b) La collection d'ensembles A_1, A_2, A_3 et A_4 est une partition de S.

FIGURE 1.5

La surface de la Terre (ensemble fondamental) est divisée en différents continents. De même, un pays est divisé en différentes provinces ou régions. En général, une collection d'ensembles non vides A_1, A_2,$\cdot$, $\cdots$, A_n est une **partition** d'un ensemble A s'ils sont disjoints et que leur union est A. Dans la figure (1.5b), les ensembles A_1, A_2, A_3 et A_4 forment une partition de l'univers S.

Voici quelques règles qui sont souvent utiles lorsque vous travaillez avec des ensembles. Nous verrons bientôt des exemples de leur utilisation.

Théorème 1.1 (Loi de De Morgan)**.** *Pour tous les ensembles A_1, A_2,$\cdots$, A_n, on a*

- $(A_1 \cup A_2 \cup A_3 \cup \cdots A_n)^c = A_1^c \cap A_2^c \cap A_3^c \cdots \cap A_n^c$;
- $(A_1 \cap A_2 \cap A_3 \cap \cdots A_n)^c = A_1^c \cup A_2^c \cup A_3^c \cdots \cup A_n^c$.

Théorème 1.2 (Loi distributive)**.** *Pour tous les ensembles A, B, et C on a*

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

Exemple 1.4. Si l'univers est donné par $S = \{1,2,3,4,5,6\}$, et $A = \{1,2\}$, $B = \{2,4,5\}$, $C = \{1,5,6\}$ sont trois ensembles, trouver les ensembles suivants :

a) $A \cup B$

b) $A \cap B$

c) $\overline{A}$

d) $\overline{B}$

e) Vérifiez la loi de De Morgan en trouvant $(A \cup B)^c$ et $A^c \cap B^c$.

f) Vérifiez la loi distributive en trouvant $A \cap (B \cup C)$ et $(A \cap B) \cup (A \cap C)$.

Solution

a) $A \cup B = \{1,2,4,5\}$.

b) $A \cap B = \{2\}$.

c) $\overline{A} = \{3,4,5,6\}$($\overline{A}$ se compose d'éléments qui se trouvent dans S mais pas dans A).

d) $\overline{B} = \{1,3,6\}$.

e) Nous avons $(A \cup B)^c = \{1,2,4,5\}^c = \{3,6\}$, qui est le même que $A^c \cap B^c = \{3,4,5,6\} \cap \{1,3,6\} = \{3,6\}$.

f) Nous avons $A \cap (B \cup C) = \{1,2\} \cap \{1,2,4,5,6\} = \{1,2\}$, qui est le même que $(A \cap B) \cup (A \cap C) = \{2\} \cup \{1\} = \{1,2\}$.

Produit cartésien

Un **produit cartésien** de deux ensembles A et B, écrit comme $A \times B$, est l'ensemble contenant les paires ordonnées de A et B. C'est-à-dire si $C = A \times B$, alors chaque élément de C est de la forme (x, y), où $x \in A$ et $y \in B$:

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

Par exemple, si $A = \{1, 2, 3\}$ et $B = \{H, T\}$, alors

$$A \times B = \{(1, H), (1, T), (2, H), (2, T), (3, H), (3, T)\}.$$

Notez qu'ici les paires sont ordonnées, donc par exemple,$(1, H) \neq (H, 1)$. Ainsi $A \times B$ **n'est pas** le même que $B \times A$.

Si vous avez deux ensembles finis A et B, où A a M éléments et B a N éléments, alors $A \times B$ a MN éléments. Cette règle s'appelle le principe de multiplication et est très utile pour compter le nombre d'éléments dans des ensembles. Le nombre d'éléments d'un ensemble est noté $|A|$, donc ici on écrit $|A| = M$, $|B| = N$, et $|A \times B| = MN$. Dans l'exemple (1.4), $|A| = 3$, $|B| = 2$, Donc $|A \times B| = 3 \times 2 = 6$. On peut de même définir le produit cartésien de n ensembles A_1, A_2, $\cdots$, A_n comme

$$A_1 \times A_2 \times A_3 \times \cdots \times A_n = \{(x_1, x_2, \cdots, x_n) | x_1 \in A_1 \text{ et } x_2 \in A_2 \text{ et } \cdots x_n \in A_n\}.$$

Le principe de multiplication stipule que pour des ensembles finis A_1, A_2, $\cdots$, A_n, si

$$|A_1| = M_1,\ |A_2| = M_2,\ \cdots, |A_n| = M_n,$$

alors

$$| A_1 \times A_2 \times A_3 \times \cdots \times A_n |= M_1 \times M_2 \times M_3 \times \cdots \times M_n.$$

Un exemple important d'ensembles obtenus à l'aide d'un produit cartésien est $\mathbb{R}^n$, où n est un nombre naturel. Pour $n = 2$, on a

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y) | x \in \mathbb{R}, y \in \mathbb{R}\}.$$

Ainsi, $\mathbb{R}^2$ est l'ensemble constitué de tous les points du plan. De même, $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$, etc.

1.2.3 Cardinalité : ensembles dénombrables et indénombrables

La **cardinalité** d'un ensemble est le nombre d'éléments de cet ensemble. La cardinalité d'un ensemble est notée $|A|$. Nous discutons d'abord de la cardinalité des ensembles finis, puis nous parlons des ensembles infinis.

1.2.4 Ensembles finis

Considérons un ensemble A. Si A n'a qu'un nombre fini d'éléments, sa cardinalité est simplement le nombre d'éléments dans A. Par exemple, si $A = \{2, 4, 6, 8, 10\}$, alors $|A| = 5$. Avant de discuter des ensembles infinis, parlons d'une règle très utile : le **principe d'inclusion-exclusion**. Pour deux ensembles finis A et B, on a

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

Pour voir cela, notez que lorsque nous ajoutons $|A|$ et $|B|$, nous comptons les éléments dans $|A \cap B|$ deux fois, donc en le soustrayant de $|A| + |B|$, on obtient le nombre d'éléments dans $|A \cup B|$. Nous pouvons étendre la même idée à trois ensembles ou plus.

Principe d'inclusion-exclusion :

1. $|A \cup B| = |A| + |B| - |A \cap B|$,
2. $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$.

Généralement, pour n ensembles finis A_1, A_2, A_3,$\cdots$, A_n, nous pouvons écrire

$$\boxed{\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{n} |A_i| - \sum_{i<j} \left|A_i \cap A_j\right| + \sum_{i<j<k} \left|A_i \cap A_j \cap A_k\right| - \ \cdots\ + (-1)^{n+1} |A_1 \cap \cdots \cap A_n|.}$$

Exemple 1.5. Dans une fête,

- 10 personnes ont porté des chemises blanches ;
- 8 personnes ont porté des chemises rouges ;
- 4 personnes ont porté des chaussures noires et des chemises blanches ;
- 3 personnes ont porté des chaussures noires et des chemises rouges ;
- le nombre total de personnes portant des chemises blanches ou rouges ou des chaussures noires est 21.

Combien de personnes ont des chaussures noires ?

Solution

Soient W, R, et B le nombre de personnes portant respectivement des chemises blanches, des chemises rouges et des chaussures noires. Ensuite, voici le récapitulatif des informations disponibles :

$$|W| = 10; \ |R| = 8; \ |W \cap B| = 4; \ |R \cap B| = 3; \ |W \cup B \cup R| = 21.$$

Aussi, il est raisonnable de supposer que W et R sont disjoints, $|W \cap R| = 0$. Ainsi, en appliquant le principe d'inclusion-exclusion, nous obtenons

$$
\begin{aligned}
|W \cup R \cup B| =& 21 \\
=& |W| + |R| + |B| - |W \cap R| - |W \cap B| - |R \cap B| + |W \cap R \cap B| \\
=& 10 + 8 + |B| - 0 - 4 - 3 + 0.
\end{aligned}
$$

Ainsi

$$|B| = 10.$$

Notez qu'une autre façon de résoudre ce problème consiste à utiliser un diagramme de Venn, comme illustré à la figure (1.6).

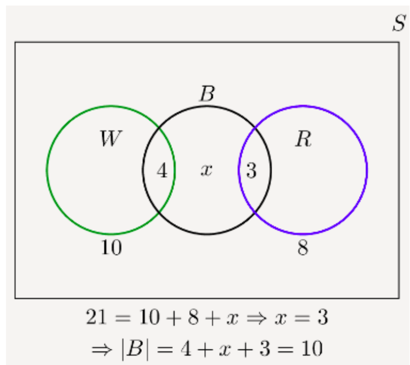


FIGURE 1.6 – Diagramme de Venn inclusion-exclusion.

1.2.5 Ensembles infinis

Et qu'est-ce qui se passerait si A est un ensemble infini ? Il s'avère que nous devons faire la distinction entre deux types d'ensembles infinis, où un type est significativement "plus grand" que l'autre. En particulier, le premier type d'ensemble est dit **dénombrable**, tandis que l'autre est dit **indénombrable**. Des ensembles tels que $\mathbb{N}$ et $\mathbb{Z}$ sont dénombrables, mais des ensembles "plus grands" tels que $\mathbb{R}$ sont dits indénombrables. La différence entre les deux types est que vous pouvez lister les éléments d'un ensemble dénombrable A, c'est-à-dire que vous pouvez écrire $A = \{a_1, a_2, \cdots\}$, mais vous ne pouvez pas répertorier les éléments d'un ensemble indénombrable. Par exemple, vous pouvez écrire

- $\mathbb{N} = \{1, 2, 3, \cdots\}$,
- $\mathbb{Z} = \{0, 1, -1, 2, -2, 3, -3, \cdots\}$.

Le fait que vous puissiez lister les éléments d'un ensemble dénombrable infini signifie que l'ensemble peut être mis en correspondance biunivoque avec des nombres naturels $\mathbb{N}$. En revanche, vous ne pouvez pas lister les éléments dans $\mathbb{R}$, c'est donc un ensemble indénombrable. Pour être précis, voici la définition.

Définition 1.1. L'ensemble est dénombrable si l'une des conditions suivantes est vraie

a) si c'est un ensemble fini, $|A| < \infty$; ou

b) il peut être mis en correspondance biunivoque avec des nombres naturels $\mathbb{N}$, auquel cas l'ensemble est dit dénombrable infini.
Un ensemble est dit indénombrable s'il n'est pas dénombrable.

Voici une règle simple pour décider si un ensemble est dénombrable ou non. En ce qui concerne la probabilité appliquée, cette ligne directrice devrait suffire dans la plupart des cas.

- $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, et tous leurs sous-ensembles sont dénombrables.
- Tout ensemble contenant un intervalle sur la ligne réelle tel que $[a,b]$, $]a,b]$, $[a,b[$, ou $]a,b[$, où $a < b$ est indénombrable.

1.2.6 Fonctions

Nous avons souvent besoin du concept de fonctions en probabilité. Une fonction f prend une entrée d'un ensemble spécifique, appelé **domaine**, et produit une sortie d'un autre ensemble, appelé **co-domaine**. Ainsi, une fonction associe les éléments de l'ensemble de domaines aux éléments du co-domaine avec la propriété que chaque entrée est associée à exactement une sortie. Pour une fonction f, si x est un élément du domaine, alors la valeur de la fonction (la sortie de la fonction) est représentée par $f(x)$. Si A est le domaine et B est le co-domaine de la fonction f, nous utilisons la notation suivante :

$$f : A \to B.$$

Exemple 1.6.

- Considérons la fonction $f : \mathbb{R} \to \mathbb{R}$, défini comme $f(x) = x^2$. Cette fonction prend n'importe quel nombre réel x et sorties x^2. Par exemple, $f(2) = 4$.
- Considérons la fonction $g : \{H,T\} \to \{0,1\}$, défini comme $g(H) = 0$ et $g(T) = 1$. Cette fonction ne peut prendre que deux entrées possibles H ou T, où H est associé à 0 et T est mappé à 1.

La sortie d'une fonction $f : A \to B$ appartient toujours au co-domaine B. Cependant, toutes les valeurs du co-domaine ne sont pas toujours couvertes par la fonction. Dans l'exemple ci-dessus, $f : \mathbb{R} \to \mathbb{R}$, la valeur de la fonction est toujours un nombre positif $f(x) = x^2 \geqslant 0$.

Nous définissons l'image d'une fonction comme l'ensemble contenant toutes les valeurs possibles de $f(x)$. Ainsi, l'image d'une fonction est toujours un sous-ensemble de son co-domaine. Pour la fonction ci-dessus $f(x) = x^2$, l'image de f est donnée par

$$\mathrm{Im}(f) = \mathbb{R}^+ = \{x \in \mathbb{R} | x \geqslant 0\}.$$

La figure (1.7) montre graphiquement une fonction, son domaine, son co-domaine et son image. La figure montre qu'un élément x dans le domaine est mappé à $f(x)$ dans le périmètre.

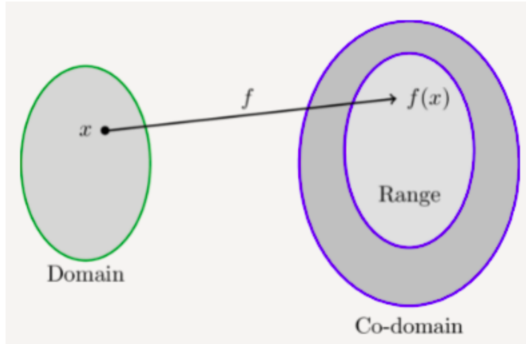


FIGURE 1.7 – Fonction $f : A \to B$, l'image est toujours un sous-ensemble du co-domaine.

Exercices

1. Soient A, B et C trois ensembles. Pour chacun des ensembles suivants, dessinez un diagramme de Venn et ombrez la zone représentant l'ensemble donné.

   a) $A \cup B \cup C$
   b) $A \cap B \cap C$
   c) $A \cup (B \cap C)$
   d) $A - (B \cap C)$
   e) $A \cup (B \cap C)^c$

2. À l'aide des diagrammes de Venn, vérifiez les identités suivantes.

   (a) $A = (A \cap B) \cup (A - B)$
   (b) Si A et B sont des ensembles finis, on a

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

3. Soit $S = \{1, 2, 3\}$. Écrivez toutes les partitions possibles de S.

4. Déterminez si chacun des ensembles suivants est dénombrable ou indénombrable.

   (a) $A = \{x \in \mathbb{Q} | -100 \leqslant x \leqslant 100\}$
   (b) $B = \{(x, y) | x \in \mathbb{N}, y \in \mathbb{Z}\}$
   (c) $C =]0, 0.1]$
   (d) $D = \{\frac{1}{n} | n \in \mathbb{N}^*\}$

5. Trouver l'image de la fonction $f : \mathbb{R} \to \mathbb{R}$ défini comme $f(x) = \sin(x)$.

1.3 Expériences aléatoires et probabilités

1.3.1 Expériences aléatoires

Avant de lancer un dé, vous ne connaissez pas le résultat. Ceci est un exemple d'**expérience aléatoire**. En particulier, une expérience aléatoire est un processus par lequel nous observons quelque chose d'incertain. Après l'expérience, le résultat de l'expérience aléatoire est connu. L'ensemble de tous les résultats possibles est appelé **ensemble fondamental** (ou **univers**). Voici quelques exemples d'expériences aléatoires et leurs univers :

- Expérience aléatoire : lancer d'une pièce de monnaie ; univers : $S = \{face, pile\}$ ou comme nous l'écrivons habituellement, $\{H, T\}$ (H pour heads (face) et T pour tails (pile) en anglais)
- Expérience aléatoire : lancer d'un dé ; univers : $S = \{1, 2, 3, 4, 5, 6\}$.
- Expérience aléatoire : observez le nombre des téléphones TECNO vendus par un magasin en 2022 ; univers : $S = \{0, 1, 2, 3, \cdots\}$.
- Expérience aléatoire : observez le nombre de buts dans un match de football ; univers : $S = \{0, 1, 2, 3, \cdots\}$.

Lorsque nous répétons plusieurs fois une expérience aléatoire, nous appelons chacune d'elles un essai. Ainsi, un essai est une performance particulière d'une expérience aléatoire. Dans l'exemple du lancer d'une pièce de monnaie, chaque essai se traduira par des faces ou des piles. Notez que l'univers est défini en fonction de la façon dont vous définissez votre expérience aléatoire. Par exemple,

Exemple 1.7. Nous lançons une pièce trois fois et nous observons la séquence face/pile. L'univers ici peut être défini comme

$$S = \{(H, H, H), (H, H, T), (H, T, H), (T, H, H), (H, T, T), (T, H, T), (T, T, H), (T, T, T)\}.$$

Notre objectif est d'attribuer une probabilité à certains événements. Par exemple, supposons que nous voudrions connaître la probabilité que le résultat du lancer d'un dé équilibré soit un nombre pair. Dans ce cas, notre événement est l'ensemble $E = \{2, 4, 6\}$. Si le résultat de notre expérience aléatoire appartient à l'ensemble E, on dit que l'événement E s'est produit. Ainsi, un événement est un ensemble de résultats possibles. En d'autres termes, un événement est un sous-ensemble de l'univers auquel nous attribuons une probabilité. Bien que nous n'ayons pas encore discuté de la façon de trouver la probabilité d'un événement, vous pourriez être en mesure de deviner que la probabilité de $\{2, 4, 6\}$ est 50% qui est le même que $\frac{1}{2}$ dans la convention de la théorie des probabilités.

Résultat :	Résultat d'une expérience aléatoire.
Univers :	Ensemble de tous les résultats possibles.
Événement :	Un sous-ensemble de l'univers.

Union et **intersection** : si A et B sont des événements, alors $A \cup B$ et $A \cap B$ sont aussi des événements. En se souvenant de la définition de l'union et de l'intersection, on observe que $A \cup B$ se produit si A ou B se produit. De même, $A \cap B$ se produit si les deux A et B se produisent. De même, si A_1, A_2,$\cdots$, A_n sont des événements, alors l'événement $A_1 \cup A_2 \cup A_3 \cdots \cup A_n$ survient si au moins un des $A_1, A_2, \cdots, A_n$ se produit. L'événement $A_1 \cap A_2 \cap A_3 \cdots \cap A_n$ se produit si tous $A_1, A_2, \cdots, A_n$ se produisent. Il peut être utile de rappeler que les mots clés ≪ **ou** ≫ et ≪ **au moins** ≫ correspondent à des **unions** et les mots clés ≪ **et** ≫ et ≪ **tous** ≫ correspondent à des **intersections**.

1.3.2 Probabilité

Nous attribuons une mesure de **probabilité** $\mathbb{P}(A)$ à un événement A. C'est une valeur entre 0 et 1 qui montre la probabilité de l'événement. Si $\mathbb{P}(A)$ est près de 0, il est très peu probable que l'événement A se produise. D'autre part, si $\mathbb{P}(A)$ est près de 1, A est très susceptible de se produire. Le sujet principal de la théorie des probabilités est de développer des outils et des techniques pour calculer les probabilités de différents événements. La théorie des probabilités est basée sur certains axiomes qui servent de fondement à la théorie, alors énonçons et expliquons ces axiomes.

Axiomes de probabilité :

- **Axiome 1 :** Pour tout événement A, $\mathbb{P}(A) \geqslant 0$.
- **Axiome 2 :** Probabilité de l'univers S est $\mathbb{P}(S) = 1$.
- **Axiome 3 :** Si A_1, A_2, A_3, $\cdots$ sont des événements disjoints, alors

$$\mathbb{P}(A_1 \cup A_2 \cup A_3 \cdots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) + \cdots$$

Prenons quelques instants et assurons-nous de bien comprendre chaque axiome. Le premier axiome stipule que la probabilité ne peut pas être négative. La plus petite valeur pour $\mathbb{P}(A)$ est nulle et si $\mathbb{P}(A) = 0$, alors l'événement A ne se produira jamais. Le deuxième axiome stipule que la probabilité de l'ensemble de l'univers est égale à un, c'est-à-dire, 100 pour cent. La raison en est que l'univers S contient tous les résultats possibles de notre expérience aléatoire. Ainsi, le résultat de chaque essai appartient toujours à S, c'est-à-dire l'événement S se produit toujours et $\mathbb{P}(S) = 1$. Dans l'exemple du lancer de dé, $S = \{1, 2, 3, 4, 5, 6\}$, et puisque le résultat est toujours parmi les nombres 1 à 6, $\mathbb{P}(S) = 1$.

Le troisième axiome est probablement le plus intéressant. L'idée de base est que si certains événements sont disjoints (c'est-à-dire il n'y a pas de chevauchement entre eux), alors la probabilité de leur union doit être la somme de leurs probabilités. Une autre façon de penser à cela est d'imaginer la probabilité d'un ensemble comme l'aire de cet ensemble dans le diagramme de Venn. Si plusieurs ensembles sont disjoints comme ceux de la figure (1.5a) , alors l'aire totale de leur union est la somme des aires individuelles. L'exemple (1.8) suivant illustre l'idée derrière le troisième axiome.

Exemple 1.8. À une élection présidentielle, il y a quatre candidats : A, B, C et D. Sur la base d'une analyse de sondage, on estime que A a 20% de chance de gagner l'élection, tandis que B a 40% de chance de gagner. Quelle est la probabilité que A ou B remporte l'élection ?

Solution

Notez que les événements $\{\text{A gagne}\}$, $\{\text{B gagne}\}$, $\{\text{C gagne}\}$, et $\{\text{D gagne}\}$ sont disjoints puisque plus d'un d'entre eux ne peut pas se produire en même temps. Par exemple, si A gagne, alors B ne peut pas gagner. D'après le troisième axiome de probabilité, la probabilité de l'union de deux événements disjoints est la somme des probabilités individuelles. Donc,

$$\begin{aligned}\mathbb{P}(\text{A gagne ou B gagne}) =& \mathbb{P}\big(\{\text{A gagne}\} \cup \{\text{B gagne}\}\big)\\ =& \mathbb{P}(\{\text{A gagne}\}) + \mathbb{P}(\{\text{B gagne}\})\\ =& 0,2 + 0,4\\ =& 0,6.\end{aligned}$$

En résumé, si A_1 et A_2 sont des événements disjoints, alors $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$. Le même argument est vrai lorsque vous avez n événements disjoints $A_1, A_2, \cdots, A_n$:

$$\mathbb{P}(A_1 \cup A_2 \cup A_3 \cdots \cup A_n) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_n), \text{ si } A_1, A_2, \cdots, A_n \text{ sont disjoints.}$$

En fait, le troisième axiome va au-delà de cela et déclare que la même chose est vraie même pour un nombre dénombrable infini d'événements disjoints.

Comme nous l'avons vu, lorsque vous travaillez avec des événements, intersection signifie "et" et union signifie "ou" . La probabilité d'intersection de A et B, $\mathbb{P}(A \cap B)$, se traduit parfois par $\mathbb{P}(A, B)$ ou $\mathbb{P}(A \text{ et } B)$.

Notation :

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

Parallélisme entre la terminologie ensembliste et la terminologie probabiliste

Un événement étant un élément de $\mathcal{P}(S)$ (ensemble des parties) obéit à la théorie des ensembles. Nous allons indiquer dans le tableau (1.1) ci-après comment certaines notions ensemblistes s'expriment, ou se traduisent, en termes d'événements.

TABLE 1.1 – Parallélisme entre la terminologie ensembliste et la terminologie probabiliste

Ensemble	**Événement**
On a observé le résultat ω et $\omega \in A$	L'événement A est réalisé.
$A = B$	Les événements A et B sont identiques.
$A \subset B$	L'événement A implique l'événement B.
$\emptyset$	Événement impossible.
S	Événement certain.
$A \cup B$	Au moins un des deux événements est réalisé.
$A \cap B$	Les deux événements A et B sont réalisés.
$A \cap B = \emptyset$	Les événements A et B sont incompatibles.
$\overline{A} = S - A$ ou A^c	L'événement A n'est pas réalisé.

Le couple $(S, \mathcal{P}(S))$ s'appelle un **espace probabilisable**.

1.3.3 Détermination des probabilités

Supposons qu'on nous donne une expérience aléatoire avec un univers S. Pour trouver la probabilité d'un événement, il y a généralement deux étapes : premièrement, nous utilisons les informations spécifiques dont nous disposons sur l'expérience aléatoire. Deuxièmement, nous utilisons les axiomes de probabilité. Prenons un exemple. Bien qu'il s'agisse d'un exemple simple et que vous puissiez être tenté d'écrire la réponse sans suivre les étapes, nous vous encourageons à suivre les étapes.

Exemple 1.9. Vous lancez un dé équilibré. Quelle est la probabilité de l'événement $E = \{1,5\}$?

Solution
Utilisons d'abord les informations spécifiques dont nous disposons sur l'expérience aléatoire. Le problème indique que le dé est parfaitement équilibré, ce qui signifie que les six résultats possibles sont également probables, c'est-à-dire, $\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \cdots = \mathbb{P}(\{6\})$. Maintenant, nous pouvons utiliser les axiomes de probabilité. En particulier, comme les événements $\{1\}$, $\{2\}$, $\cdots$, $\{6\}$ sont disjoints on peut écrire

$$1 = \mathbb{P}(S) = \mathbb{P}\Big(\{1\} \cup \{2\} \cup \cdots \cup \{6\}\Big) = \mathbb{P}(\{1\}) + \mathbb{P}(\{2\}) + \cdots + \mathbb{P}(\{6\}) = 6\mathbb{P}(\{1\}).$$

Ainsi,

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \cdots = \mathbb{P}(\{6\}) = \frac{1}{6}.$$

Encore comme $\{1\}$ et $\{5\}$ sont disjoints, nous avons

$$\mathbb{P}(E) = \mathbb{P}(\{1,5\}) = \mathbb{P}(\{1\}) + \mathbb{P}(\{5\}) = \frac{2}{6} = \frac{1}{3}.$$

Il est à noter que nous allons souvent écrire $\mathbb{P}(1)$ au lieu de $\mathbb{P}(\{1\})$ pour simplifier la notation, mais nous devons souligner que la probabilité est définie pour des ensembles (événements) et non pour des résultats individuels. Ainsi, lorsque nous écrivons $\mathbb{P}(2) = \frac{1}{6}$, nous voulons vraiment dire, $\mathbb{P}(\{2\}) = \frac{1}{6}$.

Nous verrons que les deux étapes expliquées ci-dessus peuvent être utilisées pour trouver des probabilités pour des événements beaucoup plus compliqués et des expériences aléatoires. Entraînons-nous maintenant à utiliser les axiomes en prouvant quelques faits utiles.

Exemple 1.10. En utilisant les axiomes de probabilité, prouver ce qui suit :

a) Pour tout événement A, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

b) La probabilité de l'ensemble vide est nulle, c'est-à-dire $\mathbb{P}(\emptyset) = 0$.

c) Pour tout événement A, $\mathbb{P}(A) \leqslant 1$.

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

e) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$, (principe d'inclusion-exclusion pour $n = 2$).

f) Si $A \subset B$, alors $\mathbb{P}(A) \leqslant \mathbb{P}(B)$.

Solution

a) Cela indique que la probabilité que A ne se produise pas est $1 - \mathbb{P}(A)$. Pour le prouver en utilisant les axiomes, on peut écrire

$$\begin{aligned} 1 =& \mathbb{P}(S) \text{ (axiome 2)} \\ =& \mathbb{P}(A \cup A^c) \text{ (définition du complément)} \\ =& \mathbb{P}(A) + \mathbb{P}(A^c) \text{ (puisque } A \text{ et } A^c \text{ sont disjoints)} \end{aligned}$$

b) Comme $\emptyset = S^c$, nous utilisons le résultat trouvé à (a). Ce qui donne $\mathbb{P}(\emptyset) = 1 - \mathbb{P}(S) = 0$. Notez que cela a du sens par définition : un événement se produit si le résultat de l'expérience aléatoire appartient à cet événement. Puisque l'ensemble vide n'a aucun élément, le résultat de l'expérience n'appartient jamais à l'ensemble vide.

c) De (a), $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$ et comme $\mathbb{P}(A^c) \geqslant 0$ (le premier axiome), on a $\mathbb{P}(A) \leqslant 1$.

d) Montrons que $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A - B)$. A noter que les deux ensembles $A \cap B$ et $A - B$ sont disjoints et leur union est A (Figure (1.8)). Ainsi, par le troisième axiome de probabilité

$$\begin{aligned} \mathbb{P}(A) &= P\big((A \cap B) \cup (A - B)\big) & \text{(puisque } A = (A \cap B) \cup (A - B)) \\ &= \mathbb{P}(A \cap B) + \mathbb{P}(A - B) & \text{(puisque } A \cap B \text{ et } A - B \text{ sont disjoints).} \end{aligned}$$

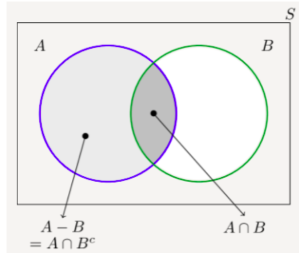


FIGURE 1.8 – $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A - B)$.

A noter que comme $A - B = A \cap B^c$, nous avons

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

Notez également que les deux ensembles B et B^c forment une partition de l'univers (puisqu'ils sont disjoints et que leur union est tout l'univers). Il s'agit d'une forme simple de loi de probabilité totale dont nous parlerons bientôt et qui est une règle très utile pour déterminer la probabilité de certains événements.

e) Notez que A et $B - A$ sont des ensembles disjoints et leur réunion est $A \cup B$. Ainsi,

$$\begin{aligned}\mathbb{P}(A \cup B) =& \mathbb{P}(A \cup (B - A)) \quad (\text{car } A \cup B = A \cup (B - A))\\ =& \mathbb{P}(A) + \mathbb{P}(B - A) \ (\text{puisque } A \text{ et } B - A \text{ sont disjoints})\\ =& \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \qquad (\text{partie (d)})\end{aligned}$$

f) Notez que $A \subset B$ signifie que chaque fois A se produit B se produit aussi. Ainsi intuitivement on s'attend à ce que $\mathbb{P}(A) \leqslant \mathbb{P}(B)$. Encore une fois, la preuve est similaire à la précédente. Si $A \subset B$, alors $A \cap B = A$. Ainsi,

$$\begin{aligned}\mathbb{P}(B) =& \mathbb{P}(A \cap B) + \mathbb{P}(B - A) \qquad (\text{par partie (d)})\\ =& \mathbb{P}(A) + \mathbb{P}(B - A) \qquad (\text{puisque } A = A \cap B)\\ \geqslant& P(A) \qquad (\text{par l'axiome 1})\end{aligned}$$

Exemple 1.11. Supposons que nous ayons les informations suivantes :

1. Il y a 60% de chance qu'il pleuve aujourd'hui.
2. Il y a 50% de chance qu'il pleuve demain.
3. Il y a 30% de chance qu'il ne pleuve pas un jour ou l'autre.

Trouvez les probabilités suivantes :

a) La probabilité qu'il pleuve aujourd'hui ou demain.

b) La probabilité qu'il pleuve aujourd'hui et demain.

c) La probabilité qu'il pleuve aujourd'hui mais pas demain.

d) La probabilité qu'il pleuve aujourd'hui ou demain, mais pas les deux.

Solution
Une étape importante dans la résolution de problèmes comme celui-ci consiste à les convertir correctement en langage de probabilité. Ceci est particulièrement utile lorsque les problèmes deviennent complexes. Pour ce problème, définissons A comme l'événement "il va pleuvoir aujourd'hui", et B comme l'événement "il pleuvra demain". Ensuite, résumons les informations disponibles :

1. $\mathbb{P}(A) = 0,6$,

2. $\mathbb{P}(B) = 0,5$,

3. $\mathbb{P}(A^c \cap B^c) = 0,3$.

Maintenant que nous avons résumé les informations, nous devrions pouvoir les utiliser avec les règles de probabilité pour trouver les probabilités demandées :

a) La probabilité qu'il pleuve aujourd'hui ou demain est $\mathbb{P}(A \cup B)$. Pour trouver cela, nous remarquons que

$$\begin{aligned}
\mathbb{P}(A \cup B) =& 1 - P\bigg((A \cup B)^c\bigg) && \text{par l'exemple 1.10} \\
=& 1 - \mathbb{P}(A^c \cap B^c) && \text{par la loi de De Morgan} \\
=& 1 - 0,3 \\
=& 0,7
\end{aligned}$$

b) La probabilité qu'il pleuve aujourd'hui et demain : c'est $\mathbb{P}(A \cap B)$. Pour la trouver, notons que

$$\begin{aligned}
\mathbb{P}(A \cap B) =& \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) && \text{par l'exemple 1.10} \\
=& 0,6 + 0,5 - 0,7 \\
=& 0,4
\end{aligned}$$

c) La probabilité qu'il pleuve aujourd'hui mais pas demain : c'est $\mathbb{P}(A \cap B^c)$.

$$\begin{aligned}
\mathbb{P}(A \cap B^c) =& \mathbb{P}(A - B) \\
=& \mathbb{P}(A) - \mathbb{P}(A \cap B) && \text{par l'exemple 1.10} \\
=& 0,6 - 0,4 \\
=& 0,2
\end{aligned}$$

d) La probabilité qu'il pleuve aujourd'hui ou demain mais pas les deux est $\mathbb{P}(A-B)+\mathbb{P}(B-A)$. Nous avons déjà trouvé $\mathbb{P}(A-B)=0,2$. De même, on peut trouver $\mathbb{P}(B-A)$:

$$\begin{aligned}\mathbb{P}(B-A) =&\mathbb{P}(B)-\mathbb{P}(B\cap A) && \text{par l'exemple 1.10}\\ =&0,5-0,4\\ =&0,1\end{aligned}$$

Ainsi,

$$\mathbb{P}(A-B)+\mathbb{P}(B-A)=0,2+0,1=0,3$$

Dans ce problème, il est dit qu'il existe un 50% de chance qu'il pleuve demain. Vous avez peut-être entendu cette information dans les nouvelles à la télévision. Une question plus intéressante est de savoir comment le nombre 50 est obtenu. Ceci est un exemple d'un problème réel dans lequel des outils de probabilité et de statistiques sont utilisés.

Principe d'inclusion-exclusion :
La formule $\mathbb{P}(A\cup B)=\mathbb{P}(A)+\mathbb{P}(B)-\mathbb{P}(A\cap B)$ que nous avons prouvé dans l'exemple 1.10 est une forme simple du principe d'inclusion-exclusion. On peut l'étendre à l'union de trois ensembles ou plus.

Principe d'inclusion-exclusion :

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

Généralement pour n événements A_1, $A_2,\cdots$, A_n, on a

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

1.3.4 Modèles de probabilité discrète

Ici, nous distinguerons deux types différents d'univers, discret et continu. Nous discuterons de la différence plus en détail plus tard, lorsque nous discuterons des variables aléatoires. L'idée de base est que dans les modèles de probabilité discrète, nous pouvons calculer la probabilité des événements en ajoutant tous les résultats correspondants, tandis que dans les modèles de probabilité continue, nous devons utiliser l'intégration au lieu de la sommation.

Considérez un univers S. Si S est un ensemble dénombrable, cela fait référence à un modèle de probabilité **discrète**. Dans ce cas, puisque S est dénombrable, on peut lister tous les éléments de S :

$$S=\{s_1,s_2,s_3,\cdots\}.$$

Si $A \subset B$ est un événement, alors A est aussi dénombrable, et par le troisième axiome de probabilité on peut écrire

$$\mathbb{P}(A) = \mathbb{P}(\bigcup_{s_j \in A} \{s_j\}) = \sum_{s_j \in A} \mathbb{P}(s_j).$$

Ainsi, dans un univers dénombrable, pour trouver la probabilité d'un événement, tout ce que nous devons faire est d'additionner la probabilité des éléments individuels de cet ensemble.

Exemple 1.12. Je joue à un jeu de hasard dans lequel je vais gagner $k-2$ dollars avec une probabilité de $\frac{1}{2^k}$ pour tout $k \in \mathbb{N}_0$, c'est-à-dire,

- avec une probabilité de $\frac{1}{2}$, je perds 1 dollar ;
- avec une probabilité de $\frac{1}{4}$, je gagne 0 dollar ;
- avec une probabilité de $\frac{1}{8}$, je gagne 1 dollar ;
- avec une probabilité de $\frac{1}{16}$, je gagne 2 dollars ;
- avec une probabilité de $\frac{1}{32}$, je gagne 3 dollars ;
- $\cdots$

Quelle est la probabilité que je gagne au moins 1 dollar et moins de 4 dollars ? Quelle est la probabilité que je gagne plus de 2 dollars ?

Solution
Dans ce problème, l'expérience aléatoire est le jeu de hasard et les résultats sont le montant en dollars que je gagne (perds). Ainsi on peut écrire

$$S = \{-1, 0, 1, 2, 3, 4, 5, \cdots\}.$$

Comme nous le voyons, il s'agit d'un ensemble infini mais dénombrable. Le problème indique également que

$$\mathbb{P}(k) = \mathbb{P}(\{k\}) = \frac{1}{2^{k+2}} \text{ pour } k \in S.$$

Tout d'abord, vérifions qu'il s'agit d'une mesure de probabilité valide. Pour ce faire, nous devons vérifier si toutes les probabilités s'additionnent à un, c'est-à-dire, $\mathbb{P}(S) = 1$. Nous avons

$$\begin{aligned}
\mathbb{P}(S) &= \sum_{k=-1}^{\infty} \mathbb{P}(k) \\
&= \sum_{k=-1}^{\infty} \frac{1}{2^{k+2}} \\
&= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots \qquad \text{(somme géométrique)} \\
&= 1.
\end{aligned}$$

Résolvons maintenant le problème. Définissons A comme l'événement que je gagne au moins 1 dollar et moins de 4 dollars, et B comme l'événement que je gagne plus de 2 dollars. Ainsi,

$$A = \{1, 2, 3\}, \ B = \{3, 4, 5, \cdots\}.$$

Puis

$$\begin{aligned}\mathbb{P}(A) =& \mathbb{P}(1) + \mathbb{P}(2) + \mathbb{P}(3) \\ =& \frac{1}{8} + \frac{1}{16} + \frac{1}{32} \\ =& \frac{7}{32} \\ \approx& 0,219\end{aligned}$$

De même,

$$\begin{aligned}\mathbb{P}(B) =& \mathbb{P}(3) + \mathbb{P}(4) + \mathbb{P}(5) + \mathbb{P}(6) + \cdots \\ =& \frac{1}{32} + \frac{1}{64} + \frac{1}{128} + \frac{1}{256} + \cdots \qquad \text{(somme géométrique)} \\ =& \frac{1}{16} \\ =& 0,0625\end{aligned}$$

Notez qu'une autre façon de trouver $\mathbb{P}(B)$ est d'écrire

$$\begin{aligned}\mathbb{P}(B) =& 1 - \mathbb{P}(B^c) \\ =& 1 - \mathbb{P}(\{-1, 0, 1, 2\}) \\ =& 1 - \Big(\mathbb{P}(-1) + \mathbb{P}(0) + \mathbb{P}(1) + \mathbb{P}(2)\Big) \\ =& 1 - \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16}\right) \\ =& 1 - \frac{15}{16} \\ =& \frac{1}{16} \\ =& 0,0625\end{aligned}$$

Remarque. Ici, nous avons utilisé la formule de somme des séries géométriques. En particulier, pour tout $a, x \in \mathbb{R}$, on a

$$a + ax + ax^2 + ax^3 + \cdots + ax^{n-1} = \sum_{k=0}^{n-1} ax^k = a\frac{1 - x^n}{1 - x}$$

De plus, si $|x| < 1$, alors nous avons

$$a + ax + ax^2 + ax^3 + \cdots = \sum_{k=0}^{\infty} ax^k = a\frac{1}{1 - x}.$$

Univers finis avec des résultats également probables :

Un cas particulier important des modèles de probabilités discrètes est lorsque nous avons un univers fini S, où chaque résultat est également probable, c'est-à-dire,

$$S = \{s_1, s_2, \cdots, s_N\}, \text{ où } \mathbb{P}(s_i) = \mathbb{P}(s_j) \text{ pour tout } i, j \in \{1, 2, \cdots, N\}.$$

Lancer d' un dé équilibré est un exemple d'un tel modèle de probabilité. Puisque tous les résultats sont également probables, nous devons avoir

$$\mathbb{P}(s_i) = \frac{1}{N}, \text{ pour tout } i \in \{1, 2, \cdots, N\}.$$

Dans un tel modèle, si A est tout événement de cardinalité $|A| = M$, nous pouvons écrire

$$\mathbb{P}(A) = \sum_{s_j \in A} \mathbb{P}(s_j) = \sum_{s_j \in A} \frac{1}{N} = \frac{M}{N} = \frac{|A|}{|S|}.$$

Ainsi, trouver la probabilité de A se réduit à un problème de comptage dans lequel nous devons compter combien d'éléments sont dans A et S.

Exemple 1.13. Je lance deux fois un dé équilibré et j'obtiens deux nombres : X_1 = résultat du premier lancer, et X_2 = résultat du deuxième lancer. Notons l'univers S, et en supposant que tous les résultats sont également probables (parce que le dé est équilibré), trouver la probabilité de l'événement A défini comme l'événement $X_1 + X_2 = 8$.

Solution
L'univers S peut être écrit comme

$$\begin{aligned} S =& \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), \\ & (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), \\ & (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}. \end{aligned}$$

Comme on le voit il y a $|S| = 36$ éléments dans S. Pour trouver la probabilité de A, il ne nous reste plus qu'à trouver $M = |A|$. En particulier, A est défini comme

$$\begin{aligned} A =& \{(X_1, X_2) | X_1 + X_2 = 8, \ X_1, \ X_2 \in \{1, 2, \cdots, 6\}\} \\ =& \{(2,6), (3,5), (4,4), (5,3), (6,2)\}. \end{aligned}$$

Ainsi, $|A| = 5$, ce qui signifie que

$$\mathbb{P}(A) = \frac{|A|}{|S|} = \frac{5}{36}.$$

Une erreur très courante consiste à ne pas faire la distinction entre, disons $(2,6)$ et $(6,2)$. Il est important de noter qu'il s'agit de deux résultats différents : $(2,6)$ signifie que le premier lancer est un 2 et le deuxième lancer est un 6, tandis que (6,2) signifie que le premier lancer est un 6 et le deuxième lancer est un 2. Notez qu'il est très courant d'écrire $\mathbb{P}(X_1 + X_2 = 8)$ en se référant à

$\mathbb{P}(A)$ tel que défini ci-dessus. En réalité, X_1 et X_2 sont des exemples de variables aléatoires qui seront discutées en détail plus loin.

Dans un univers fini S, où tous les résultats sont également probables, la probabilité de tout événement A peut être trouvée par

$$\boxed{\mathbb{P}(\mathbf{A}) = \frac{|\mathbf{A}|}{|\mathbf{S}|}.}$$

La formule $\mathbb{P}(A) = \frac{|A|}{|S|}$ suggère qu'il est important de pouvoir compter les éléments dans les ensembles. Si les ensembles sont petits, c'est une tâche facile ; cependant, si les ensembles sont grands et définis implicitement, cela pourrait être une tâche difficile.

Exemple 1.14. Une urne contient 30 boules rouges et 70 boules vertes. Quelle est la probabilité d'obtenir exactement k boules rouges dans un échantillon de taille 20 si le prélèvement se fait sans remise (répétition non autorisée) ?

Solution
Soit A l'événement (l'ensemble) d'obtenir exactement k boules rouges. Trouver $\mathbb{P}(A) = \frac{|A|}{|S|}$, Nous devons trouver $|A|$ et $|S|$. Tout d'abord, notez que $|S| = \binom{100}{20}$. Ensuite, pour trouver $|A|$, nous devons découvrir de combien de manières nous pouvons choisir k boules rouges et $20 - k$ boules vertes. En utilisant le principe de multiplication, on a

$$|A| = \binom{30}{k}\binom{70}{20-k}.$$

Ainsi, nous avons

$$\mathbb{P}(A) = \frac{\binom{30}{k}\binom{70}{20-k}}{\binom{100}{20}}.$$

Exemple 1.15. Si k personnes sont dans une fête, quelle est la probabilité qu'au moins deux d'entre elles aient le même anniversaire ? Supposons qu'il y a $n = 365$ jours contenus dans une année et tous les jours sont également susceptibles d'être l'anniversaire d'une personne spécifique.

Solution
Soit A l'événement qu' au moins deux personnes ont le même anniversaire. A noter d'abord que si $k > n$, alors $\mathbb{P}(A) = 1$; alors, concentrons-nous sur le cas le plus intéressant où $k \leqslant n$. Encore une fois, l'expression "au moins" suggère qu'il pourrait être plus facile de trouver la probabilité de l'événement de complément, $\mathbb{P}(A^c)$. C'est l'événement où deux personnes n'ont pas le même anniversaire, et nous avons

$$\mathbb{P}(A) = 1 - \frac{|A^c|}{|S|}.$$

Ainsi, pour résoudre le problème, il suffit de trouver $|A^c|$ et $|S|$. Trouvons d'abord $|S|$. Quel est le nombre total de séquences possibles d'anniversaires de k personnes ? Eh bien, il y a $n = 365$ choix pour la première personne, $n = 365$ choix pour la deuxième personne, ... $n = 365$ choix pour la $k-$ième personne. Ainsi il y a

$$n^k$$

possibilités. Il s'agit, en fait, d'un échantillonnage ordonné avec problème de remplacement, et comme nous l'avons vu, la réponse devrait être n^k (ici on tire k échantillons, anniversaires, de l'ensemble $\{1, 2, \ldots, n = 365\}$). Trouvons maintenant $|A^c|$. Si aucun anniversaire n'est identique, cela revient à trouver $|S|$ à la différence que la répétition n'est pas autorisée, nous avons donc

$$|A^c| = P_n^k = n \times (n-1) \times \ldots \times (n-k+1).$$

Vous pouvez le voir directement en notant qu'il y a $n = 365$ choix pour la première personne, $n - 1 = 364$ choix pour la deuxième personne, ..., $n - k + 1$ choix pour la $k-$ième personne. Ainsi la probabilité de A peut être trouvée comme

$$\mathbb{P}(A) = 1 - \frac{|A^c|}{|S|} = 1 - \frac{P_n^k}{n^k}.$$

Discussion : La raison pour laquelle on appelle cela un paradoxe est que $\mathbb{P}(A)$ est numériquement différent de ce à quoi la plupart des gens s'attendent. Par exemple, s'il y a $k = 23$ personnes dans la fête, quelle est la probabilité qu'au moins deux d'entre elles aient le même anniversaire, $\mathbb{P}(A)$? La réponse est $0,5073$, ce qui est beaucoup plus élevé que ce que la plupart des gens pensent. La probabilité franchit 99% lorsque le nombre de personnes atteint 57. Mais pourquoi la probabilité est-elle plus élevée que ce à quoi nous nous attendons ?

Il est important de noter que dans le problème de l'anniversaire, aucune de deux personnes n'est choisie au préalable. Pour mieux répondre à cette question, abordons un autre problème : je suis dans une soirée avec $k - 1$ personnes. Quelle est la probabilité qu'au moins une personne du groupe ait le même anniversaire que le mien ? Eh bien, nous devons choisir les anniversaires de $k - 1$ personnes, le nombre total de façons de le faire est n^{k-1}. Le nombre total de façons de choisir les anniversaires pour que personne n'ait mon anniversaire est $(n-1)^{k-1}$. Ainsi, la probabilité qu'au moins une personne ait le même anniversaire que le mien est

$$\mathbb{P}(B) = 1 - (\frac{n-1}{n})^{k-1}.$$

Maintenant si $k = 23$, cette probabilité est seulement $\mathbb{P}(B) = 0,0586$, ce qui est beaucoup plus petit que le correspondant $\mathbb{P}(A) = 0,5073$. La raison en est que cet événement B ne regarde que le cas où une personne du groupe a le même anniversaire que moi. Il s'agit d'un événement beaucoup plus petit que l'événement A qui examine toutes les paires de personnes possibles. Ainsi, $\mathbb{P}(A)$ est beaucoup plus grand que $\mathbb{P}(B)$. On peut supposer que la valeur de $\mathbb{P}(A)$ est beaucoup plus faible qu'il ne l'est en réalité, car on pourrait le confondre avec $\mathbb{P}(B)$.

Exemple 1.16. Combien de séquences distinctes peut-on faire en utilisant 3 lettres "A" et 5 lettres "B" ? (AAABBBBB, AABABBBB, etc.)

Solution

Vous pouvez penser à ce problème de la manière suivante. Vous avez $3 + 5 = 8$ postes à pourvoir avec les lettres A ou B. De ces 8 positions, vous devez choisir 3 d'entre elles pour A. Tout ce qui reste sera rempli de B. Ainsi, le nombre total de possibilités est

$$\binom{8}{3}.$$

Maintenant, vous auriez pu choisir de manière équivalente les emplacements pour B, donc la réponse aurait été

$$\binom{8}{5}.$$

Ainsi, nous concluons que

$$\binom{8}{3} = \binom{8}{5}.$$

Le même argument peut être répété pour le cas général n et k :

$$\binom{n}{k} = \binom{n}{n-k}.$$

Exemple 1.17. Dix personnes ont un repas-partage. Cinq personnes sont sélectionnées pour apporter un plat principal, trois personnes pour apporter des boissons et deux personnes pour un dessert. De combien de manières peuvent-elles être divisées en ces trois groupes ?

Solution
Nous pouvons résoudre ce problème de la manière suivante. Premièrement, nous pouvons choisir 5 personnes pour le plat principal. Cela peut se faire en $\binom{10}{5}$ façons. Des 5 personnes restant, nous choisissons alors 3 personnes pour les boissons, et enfin les 2 personnes restant apporteront le désert. Ainsi, par le principe de multiplication, le nombre total de possibilités est donné par

$$\binom{10}{5}\binom{5}{3}\binom{2}{2} = \frac{10!}{5!5!} \cdot \frac{5!}{3!2!} \cdot \frac{2!}{2!0!} = \frac{10!}{5!3!2!}.$$

1.3.5 Modèles de probabilité continue

Considérez un scénario où votre univers S est, par exemple,$[0,1]$. C'est un ensemble indénombrable ; nous ne pouvons pas lister les éléments de l'ensemble. À l'heure actuelle, nous n'avons pas encore développé les outils nécessaires pour traiter les modèles de probabilité continue, mais nous pouvons donner une intuition en examinant un exemple simple.

Exemple 1.18. Votre ami vous dit qu'il s'arrêtera chez vous quelque temps à 1pm (après-midi c'est-à-dire 13h) ou après et avant 2pm (après-midi c'est-à-dire 14h), mais il ne peut pas vous donner plus d'informations car son emploi du temps est assez chargé. Votre ami est très fiable, vous êtes donc sûr qu'il s'arrêtera chez vous, mais à part cela, nous n'avons aucune information sur l'heure d'arrivée. Ainsi, nous supposons que l'heure d'arrivée est complètement aléatoire dans l'intervalle de 1pm et 2pm (de l'après-midi). (Comme nous le verrons, dans le langage de la théorie des probabilités, on dit que le temps d'arrivée est "uniformément" distribué sur l' intervalle $[1,2[$). Soit T l'heure d'arrivée.

1. Quel est l'univers S ?

2. Quelle est la probabilité $\mathbb{P}(1.5)$? Pourquoi ?

3. Quelle est la probabilité de $T \in [1;1.5[$

4. Pour $1 \leqslant a \leqslant b \leqslant 2$, que vaut $\mathbb{P}(a \leqslant T \leqslant b) = \mathbb{P}([a,b])$?

Solution

1. Tout nombre réel dans $[1,2[$ est un résultat possible, l'univers est en effet $S = [1,2[$.

2. Maintenant, regardons $\mathbb{P}(1.5)$. Une supposition raisonnable serait $\mathbb{P}(1,5) = 0$. Mais pouvons-nous donner une raison à cela ? Divisons l'intervalle $[1,2[$ à $2N+1$ intervalles égaux et disjoints,

$$[1,1+\frac{1}{2N+1}[,[1+\frac{1}{2N+1},1+\frac{2}{2N+1}[,\cdots,[1+\frac{N}{2N+1},1+\frac{N+1}{2N+1}[,\cdots,[1+\frac{2N}{2N+1},2[.$$

Voir Figure (1.9). Ici, N peut être n'importe quel entier positif.

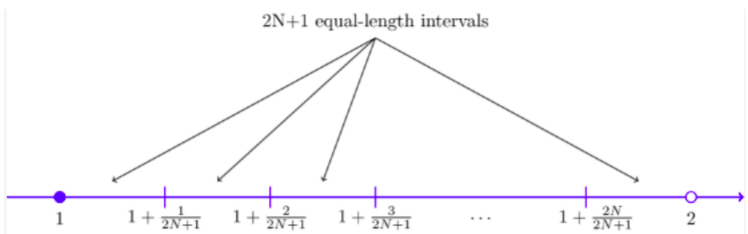


FIGURE 1.9 – Division de l'intervalle $[1,2[$ à $2N+1$ intervalles de même longueur.

La seule information dont nous disposons est que l'heure d'arrivée est "uniforme" sur l' intervalle $[1,2[$. Par conséquent, tous les intervalles ci-dessus devraient avoir la même probabilité, et puisque leur union est S nous concluons que

$$\begin{aligned}\mathbb{P}\left(\left[1,1+\frac{1}{2N+1}\right[\right) =& \mathbb{P}\left(\left[1+\frac{1}{2N+1},1+\frac{2}{2N+1}\right[\right) = \cdots \\ \cdots =& \mathbb{P}\left(\left[1+\frac{N}{2N+1},1+\frac{N+1}{2N+1}\right[\right) = \cdots \\ \cdots =& \mathbb{P}\left(\left[1+\frac{2N}{2N+1},2\right[\right) \\ =& \frac{1}{2N+1}.\end{aligned}$$

En particulier, en définissant $A_N = \left[1+\frac{N}{2N+1},1+\frac{N+1}{2N+1}\right[$, nous concluons que

$$\mathbb{P}(A_N) = \mathbb{P}\left(\left[1+\frac{N}{2N+1},1+\frac{N+1}{2N+1}\right[\right) = \frac{1}{2N+1}.$$

Notez maintenant que pour tout entier positif N, $1.5 \in A_N$. Ainsi, $\{1.5\} \in A_N$, alors

$$\mathbb{P}(1.5) \leqslant \mathbb{P}(A_N) = \frac{1}{2N+1}, \qquad \text{pour tout } N \in \mathbb{N}.$$

Notez que comme N devient grand, $\mathbb{P}(A)$ approche 0. Comme $\mathbb{P}(1.5)$ ne peut pas être négatif, on en déduit que $\mathbb{P}(1,5) = 0$. De même, nous pouvons affirmer que $\mathbb{P}(x) = 0$ pour tout $x \in [1,2[$.

3. Ensuite, on trouve $\mathbb{P}([1;1.5[)$. C'est la première moitié de l'ensemble de l'univers $S = [1,2[$ et à cause de l'uniformité, sa probabilité doit être 0.5. Autrement dit,

$$\mathbb{P}([1;1.5[) = \mathbb{P}([1.5,2[) \qquad \text{(par uniformité)}, \mathbb{P}([1;1.5[) + \mathbb{P}([1.5;2[) = \mathbb{P}(S) = 1.$$

Ainsi

$$\mathbb{P}([1;1.5[) = \mathbb{P}([1.5,2[) = \frac{1}{2}.$$

4. Le même argument d'uniformité suggère que tous les intervalles de $[1,2[$ de même longueur doit avoir la même probabilité. En particulier, la probabilité d'un intervalle est proportionnelle à sa longueur. Par exemple, comme

$$[1;1.5[= [1;1.25[\cup[1.25;1.5[.$$

Ainsi, nous concluons

$$\begin{aligned} P\big([1;1.5[\big) =& P\big([1;1.25[\big) + P\big([1.25;1.5[\big) \\ =& 2P\big([1;1.25[\big). \end{aligned}$$

Et enfin, puisque $P\big([1,2[\big) = 1$, nous concluons

$$\mathbb{P}([a,b]) = b - a, \qquad \text{pour } 1 \leqslant a \leqslant b \leqslant 2.$$

L'exemple ci-dessus était une situation assez simple dans laquelle nous avons un univers continu. En réalité, la probabilité peut ne pas être uniforme, nous devons donc développer des outils qui nous aident à gérer les distributions générales de probabilités. Ces outils seront présentés dans les chapitres suivants.

Discussion : Vous pourriez vous demander pourquoi $\mathbb{P}(x) = 0$ pour tout $x \in [1,2[$, mais en même temps, le résultat de l'expérience est toujours un nombre dans $[1,2[$? Nous pouvons répondre à cette question de différents points de vue. D'un point de vue mathématique, nous pouvons expliquer ce problème en utilisant l'analogie suivante : considérons un segment de droite de longueur un. Ce segment de droite est constitué de points de longueur nulle. Néanmoins, ces points de longueur nulle constituent dans leur ensemble un segment de droite de longueur un. D'un point de vue pratique, nous pouvons fournir l'explication suivante : nos résultats observés ne sont pas toutes des valeurs réelles dans $[1,2[$. Autrement dit, si nous observons le temps, notre mesure peut être précise jusqu'à des minutes, des secondes ou des millisecondes, etc. Notre modèle de probabilité continue est une limite d'un modèle de probabilité discrète, lorsque la précision devient infiniment précise. Ainsi, en réalité on s'intéresse toujours à la probabilité de certains intervalles plutôt qu'à un point spécifique x. Par exemple, lorsque nous disons : "quelle est la probabilité que votre ami se présente à $1:32pm$?", ce que nous voulons dire est : "quelle est la probabilité que votre ami se présente entre $1:32:00pm$ et $1:32:59pm$?" Cette probabilité est non nulle car elle fait référence à un intervalle d'une durée d'une minute. Ainsi, dans un certain sens, un modèle de probabilité continue peut être considéré comme la "limite" d'un espace discret. En se souvenant du calcul, nous notons que les intégrales sont définies comme les limites des sommes. C'est pourquoi nous utilisons les intégrales pour trouver des probabilités pour les modèles de probabilité continue, comme nous le verrons plus tard.

1.3.6 Problèmes résolus : expériences aléatoires et probabilités

1. Considérons un univers S et trois événements A, B, et C. Pour chacun des événements suivants, dessinez une représentation en diagramme de Venn ainsi qu'une expression d'ensemble.

   (a) Parmi A, B, et C, seulement A se produit.

   (b) Au moins un des événements A, B, ou C se produit.

   (c) A ou C se produit, mais pas B.

   (d) Au plus deux des événements A, B, ou C se produisent.

   Solution

   (a) Parmi A, B, et C, seulement A se produit : $A - B - C = A - (B \cup C)$.

   (b) Au moins un des événements A, B, ou C se produisent : $A \cup B \cup C$.

   (c) A ou C se produit, mais pas B : $(A \cup C) - B$.

   (d) Au plus deux des événements A, B, ou C se produisent : $(A \cap B \cap C)^c = A^c \cup B^c \cup C^c$. Les diagrammes de Venn sont illustrés à la Figure (1.10).

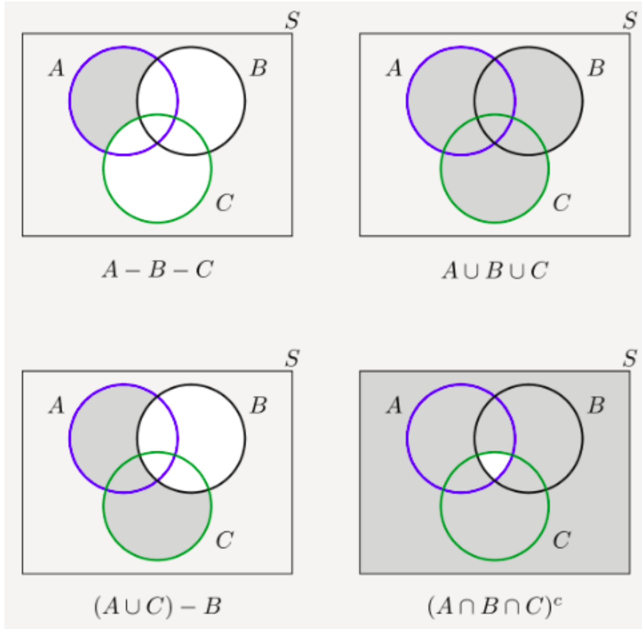


FIGURE 1.10 – Diagrammes de Venn pour le premier problème résolu.

2. Écrire l'univers S pour les expériences aléatoires suivantes.

   (a) Nous lançons une pièce jusqu'à ce que nous voyions deux piles consécutives. Nous enregistrons le nombre total de lancers de pièces.

   (b) Un sachet contient 4 boules : une rouge, une bleue, une blanche et une verte. Nous choisissons deux boules distinctes et enregistrons leur couleur dans l'ordre.

   (c) Un client arrive dans une banque et fait la queue. On observe T, qui est le temps total (en heures) pendant lequel le client attend dans la file. La banque a une politique stricte selon laquelle aucun client n'attend plus de 20 minutes en toutes circonstances.

Solution
Rappelez-vous que l'univers est l'ensemble de tous les résultats possibles. Habituellement, lorsque vous avez une expérience aléatoire, il existe différentes façons de définir l'univers S en fonction de ce que vous observez comme résultat. Dans ce problème, pour chaque expérience, il est indiqué quels résultats nous observons afin de vous aider à écrire l'univers S.

   (a) Nous lançons une pièce jusqu'à ce que nous voyions deux piles consécutives. Nous enregistrons le nombre total de lancers de pièces : ici, le nombre total de lancers de pièces est un entier naturel supérieur ou égal à 2. L'univers est $S = \{2, 3, 4, \cdots\}$.

   (b) Un sachet contient 4 boules : une rouge, une bleue, une blanche et une verte. Nous choisissons deux boules distinctes et enregistrons leur couleur dans l'ordre : l'univers peut être écrit comme

$$S = \{(R,B),(B,R),(R,W),(W,R),(R,G),(G,R),(B,W),(W,B),(B,G),$$
$$(G,B),(W,G),(G,W)\}.$$

(c) Un client arrive dans une banque et fait la queue. On observe $T \cdots$: En théorie T peut être n'importe quel nombre réel entre 0 et $\frac{1}{3} = 20$ minutes. Ainsi,

$$S = \left[0, \frac{1}{3}\right] = \left\{x \in \mathbb{R} | 0 \leqslant x \leqslant \frac{1}{3}\right\}.$$

3. Soient A, B, et C trois événements dans l'univers S. Supposons que nous sachions
   - $A \cup B \cup C = S$,
   - $\mathbb{P}(A) = \frac{1}{2}$,
   - $\mathbb{P}(B) = \frac{2}{3}$,
   - $\mathbb{P}(A \cup B) = \frac{5}{6}$.

   Répondre aux questions suivantes :

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

   (b) A, B, et C forment-ils une partition de S ?

   (c) Trouver $P\big(C - (A \cup B)\big)$.

   (d) Si $\mathbb{P}(C \cap (A \cup B)) = \frac{5}{12}$, trouver $\mathbb{P}(C)$.

   Solution
   Comme précédemment, il est toujours utile de tracer un diagramme de Venn ; cependant, nous fournissons ici la solution sans utiliser de diagramme de Venn.

   (a) En utilisant le principe d'inclusion-exclusion, nous avons $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
   Ainsi,

$$\begin{aligned} \mathbb{P}(A \cap B) = & \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) \\ = & \frac{1}{2} + \frac{2}{3} - \frac{5}{6} \\ = & \frac{1}{3}. \end{aligned}$$

   (b) Non, puisque $A \cap B \neq \emptyset$.

   (c) Nous pouvons écrire

$$\begin{aligned} C - (A \cup B) = & \Big(C \cup (A \cup B)\Big) - (A \cup B) \\ = & S - (A \cup B) (\text{puisque } A \cup B \cup C = S) \\ = & (A \cup B)^c. \end{aligned}$$

   Ainsi,

$$\begin{aligned} \mathbb{P}\big(C - (A \cup B)\big) = & \mathbb{P}\big((A \cup B)^c\big) \\ = & 1 - \mathbb{P}(A \cup B) \\ = & \frac{1}{6}. \end{aligned}$$

(d) Nous avons

$$\mathbb{P}(C) = \mathbb{P}(C \cap (A \cup B)) + \mathbb{P}(C - (A \cup B)) = \frac{5}{12} + \frac{1}{6} = \frac{7}{12}.$$

4. Je lance deux fois un dé et j'obtiens deux nombres $X_1 =$ résultat du premier lancer, et $X_2 =$ résultat du deuxième lancer. Trouvez la probabilité des événements suivants :

   (a) A défini comme "$X_1 < X_2$" ;

   (b) B défini comme "Vous observez un 6 au moins une fois".

   Solution
   Comme nous l'avons vu précédemment, l'univers S a 36 éléments.

   (a) Nous avons

   $$A = \{(1,2),(1,3),(1,4),(1,5),(1,6),(2,3),(2,4),(2,5),(2,6),(3,4),(3,5),(3,6),(4,5),(4,6),(5,6)\}.$$

   Ensuite, on obtient

   $$\mathbb{P}(A) = \frac{|A|}{|S|} = \frac{15}{36} = \frac{5}{12}.$$

   (b) Nous avons

   $$B = \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6),(1,6),(2,6),(3,6),(4,6),(5,6)\}.$$

   On obtient

   $$\mathbb{P}(B) = \frac{|B|}{|S|} = \frac{11}{36}.$$

5. Vous achetez un certain produit. Le manuel indique que la durée de vie T du produit, défini comme la durée (en années) pendant laquelle le produit fonctionne correctement jusqu'à ce qu'il tombe en panne, satisfait

   $$\mathbb{P}(T \geqslant t) = e^{-\frac{t}{5}} \text{ pour tout } t \geqslant 0.$$

   Par exemple, la probabilité que le produit dure 2 ans ou plus est

   $$\mathbb{P}(T \geqslant 2) = e^{-\frac{2}{5}} = 0,6703.$$

   (a) Ceci est un exemple de modèle de probabilité continue. Notez l'espace de l'échantillon S.

   (b) Vérifiez que la déclaration dans le manuel a du sens en trouvant $\mathbb{P}(T \geqslant 0)$ et $\lim_{t\to\infty} \mathbb{P}(T \geqslant t)$.

   (c) Vérifiez également que si $t_1 < t_2$, alors $\mathbb{P}(T \geqslant t_1) \geqslant \mathbb{P}(T \geqslant t_2)$. Pourquoi cela doit-il être vrai ?

(d) Trouvez la probabilité que le produit tombe en panne dans les trois ans suivant l'achat.

(e) Trouvez la probabilité que le produit tombe en panne au cours de la deuxième année, c'est-à-dire $\mathbb{P}(1 \leqslant T < 2)$.

Solution

(a) L'univers S est l'ensemble de tous les résultats possibles. Ici, les résultats possibles sont les valeurs possibles pour T qui peut être n'importe quel nombre réel supérieur ou égal à zéro. Ainsi

$$S = [0, +\infty[.$$

(b) Nous avons

$$\begin{aligned} P(T \geqslant 0) =& e^{-\frac{0}{5}} = 1, \\ \lim_{t\to\infty} \mathbb{P}(T \geqslant t) =& e^{-\infty} = 0, \end{aligned}$$

c'est ce que nous attendons. En particulier, T est toujours supérieur ou égal à zéro, on s'attend donc à $\mathbb{P}(T \geqslant 0) = 1$. De plus, étant donné que le produit finira par tomber en panne à un moment donné, nous nous attendons à ce que $\mathbb{P}(T \geqslant t)$ s'approche de zéro lorsque t va à l'infini.

(c) A noter d'abord que si $t_1 < t_2$, alors $\mathbb{P}(T \geqslant t_1) = e^{-\frac{t_1}{5}} > e^{-\frac{t_2}{5}} = \mathbb{P}(T \geqslant t_2)$ (puisque $f(x) = e^x$ est une fonction croissante). Ici, nous avons deux événements, A est l'événement que $T \geqslant t_1$ et B est l'événement que $T \geqslant t_2$. C'est-à-dire,

$$A = [t_1, \infty[, B = [t_2, \infty[.$$

Comme B est un sous-ensemble de A, i.e $B \subset A$, nous devons avoir $\mathbb{P}(B) \leqslant \mathbb{P}(A)$, donc

$$\mathbb{P}(A) = \mathbb{P}(T \geqslant t_1) \geqslant \mathbb{P}(T \geqslant t_2) = \mathbb{P}(B).$$

(d) La probabilité que le produit tombe en panne dans les trois ans suivant l'achat est

$$\mathbb{P}(T < 3) = 1 - \mathbb{P}(T \geqslant 3) = 1 - e^{-\frac{3}{5}} \approx 0.4512$$

(e) Notez que si $A \subset B$, alors

$$\begin{aligned} \mathbb{P}(B - A) =& \mathbb{P}(B) - \mathbb{P}(B \cap A) \\ =& \mathbb{P}(B) - \mathbb{P}(A) \qquad \text{(puisque } A \subset B). \end{aligned}$$

Choisir $A = [2, \infty[$ et $B = [1, \infty[$, nous pouvons écrire

$$\begin{aligned} \mathbb{P}(1 \leqslant T < 2) =& \mathbb{P}(T \geqslant 1) - \mathbb{P}(T \geqslant 2) \\ =& e^{-\frac{1}{5}} - e^{-\frac{2}{5}} \\ =& 0.1484 \end{aligned}$$

6. Vous cassez un bâton en trois morceaux au hasard. Quelle est la probabilité que vous constituez un triangle en utilisant les trois pièces ? Supposez que les points sont choisis complètement au hasard, c'est-à-dire si la longueur du bâton d'origine est 1 unité, et x, y, z sont les longueurs des trois pièces, alors (x, y, z) sont uniformément choisis dans l'ensemble

$$\{(x, y, z) \in \mathbb{R}^3 | x + y + z = 1, x, y, z \geqslant 0\}.$$

Solution
C'est encore un problème sur un espace de probabilité continu. L'idée de base est assez simple. Premièrement, nous devons identifier l'univers S. Dans ce cas, l'univers va être un ensemble à deux dimensions. Deuxièmement, nous devons identifier l'ensemble A qui contient les résultats favorables (l'ensemble des (x, y, z) dans S qui forment un triangle). Et enfin, puisque l'espace est uniforme, nous diviserons l'aire de l'ensemble A par la zone de S pour obtenir $\mathbb{P}(A)$.

Tout d'abord, nous devons trouver les ensembles S et A. Il s'agit essentiellement d'un problème de géométrie. Les deux ensembles, S et A, sont illustrés à la Figure (1.11).

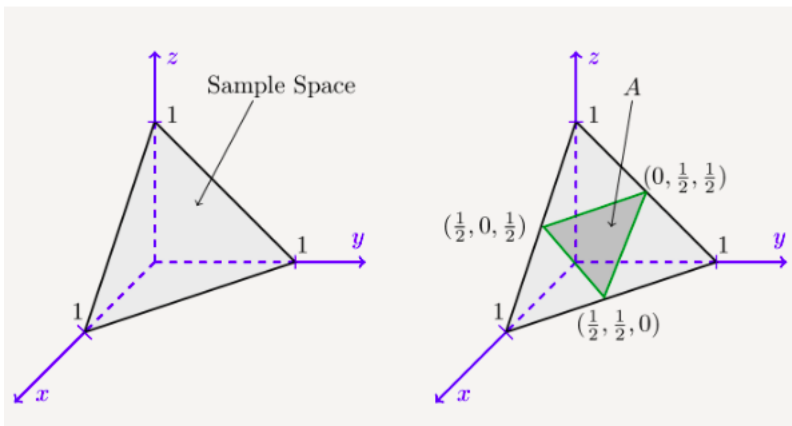


FIGURE 1.11 – L'univers et l'ensemble A

Notez que dans $\mathbb{R}^3$, $x + y + z = 1$ représente un plan qui passe par les points $(1, 0, 0), (0, 1, 0), (0, 0, 1)$. Pour trouver l'univers S, notez que $S = \{(x, y, z) \in \mathbb{R}^3 | x + y + z = 1, x, y, z \geqslant 0\}$, donc S est la partie du plan représentée sur la Figure (1.11).

Pour trouver l'ensemble A, notez que nous avons besoin de (x, y, z) pour satisfaire l'inégalité triangulaire

$$\begin{aligned} x + y &> z, \\ y + z &> x, \\ x + z &> y. \end{aligned}$$

A noter que comme $x + y + z = 1$, nous pouvons écrire de manière équivalente les trois équations comme

$$x < \frac{1}{2},$$
$$y < \frac{1}{2},$$
$$z < \frac{1}{2}.$$

Ainsi, nous concluons que l'ensemble A est la zone représentée sur la figure (1.11). En particulier, on note que l'ensemble S est composé de quatre triangles d'aires égales. Son aire est donc quatre fois supérieure à celle de A, et nous avons

$$\mathbb{P}(A) = \frac{\text{Zone de } A}{\text{Zone de } S} = \frac{1}{4}.$$

1.4 Probabilité conditionnelle

Dans cette section, nous discutons de l'un des concepts les plus fondamentaux de la théorie des probabilités. Voici la question : au fur et à mesure que vous obtenez des informations supplémentaires, comment devez-vous mettre à jour les probabilités d'événements ? Par exemple, supposons que dans une certaine ville, 23% des jours sont pluvieux. Ainsi, si vous choisissez un jour au hasard, la probabilité qu'il pleuve ce jour-là est 23% :

$$\mathbb{P}(R) = 0,23, \text{où } R \text{ est l'événement pour lequel il pleut le jour choisi au hasard.}$$

Supposons maintenant que je choisisse un jour au hasard, mais que je vous dise également qu'il fait nuageux le jour choisi. Maintenant que vous avez cette information supplémentaire, comment mettez-vous à jour la probabilité qu'il pleuve ce jour-là ? En d'autres termes, quelle est la probabilité qu'il pleuve **étant donné** qu'il fait nuageux ? Si C est l'événement que le jour choisi soit nuageux, alors nous écrivons ceci comme $\mathbb{P}(R|C)$, la *probabilité conditionnelle de R étant donné que C s'est produit*. Il est raisonnable de supposer que dans cet exemple, $\mathbb{P}(R|C)$ devrait être plus grand que l'original $\mathbb{P}(R)$, qui est appelée la probabilité a priori de R. Mais que doit exactement être $\mathbb{P}(R|C)$? Avant de fournir une formule générale, regardons un exemple simple.

Exemple 1.19. Je lance un dé équilibré. Soit A l'événement où le résultat est un nombre impair, c'est-à-dire, $A = \{1,3,5\}$. Soit B l'événement où le résultat est inférieur ou égal à 3, c'est-à-dire, $B = \{1,2,3\}$. Quelle est la probabilité de A, i.e $\mathbb{P}(A)$? Quelle est la probabilité de A étant donné B, i.e $\mathbb{P}(A|B)$?

Solution
Il s'agit d'un univers fini, donc

$$\mathbb{P}(A) = \frac{|A|}{|S|} = \frac{|\{1,3,5\}|}{6} = \frac{1}{2}.$$

Maintenant, trouvons la probabilité conditionnelle de A étant donné que B eût lieu. Si nous savons B s'est produit, le résultat doit être parmi $\{1,2,3\}$. Pourque que A aussi se réalise, le résultat doit être dans $A \cap B = \{1,3\}$. Puisque tous les jets de dé sont également probables, nous soutenons que $\mathbb{P}(A|B)$ doit être égal à $\mathbb{P}(A|B) = \frac{|A\cap B|}{|B|} = \frac{2}{3}$.

Voyons maintenant comment nous pouvons généraliser l'exemple ci-dessus. On peut réécrire le calcul en divisant le numérateur et le dénominateur par $|S|$ de la manière suivante

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

Bien que le calcul ci-dessus ait été effectué pour un univers fini avec des résultats également probables, il s'avère que la formule résultante est assez générale et peut être appliquée dans n'importe quel contexte. Ci-dessous, nous fournissons formellement la formule, puis expliquons l'intuition qui la sous-tend.

Si A et B sont deux événements dans un univers S, alors la probabilité conditionnelle de A étant donné B est définie comme

$$\mathbb{P}(\mathbf{A}|\mathbf{B}) = \frac{\mathbb{P}(\mathbf{A} \cap \mathbf{B})}{\mathbb{P}(\mathbf{B})}, \text{ quand } \mathbb{P}(\mathbf{B}) > \mathbf{0}.$$

Voici l'intuition derrière la formule. Quand on sait que B s'est produit, chaque résultat qui est en dehors de B doit être jeté. Ainsi, notre univers est réduit à l'ensemble B, Figure (1.12). Maintenant, la seule façon que A puisse arriver, c'est quand le résultat appartient à l'ensemble $A \cap B$. Nous divisons $\mathbb{P}(A \cap B)$ par $\mathbb{P}(B)$, de sorte que la probabilité conditionnelle du nouvel univers devienne 1, c'est-à-dire, $\mathbb{P}(B|B) = \frac{\mathbb{P}(B\cap B)}{\mathbb{P}(B)} = 1$.

Notez que la probabilité conditionnelle de $\mathbb{P}(A|B)$ est indéfinie lorsque $\mathbb{P}(B) = 0$. C'est bien parce que si $\mathbb{P}(B) = 0$, cela signifie que l'événement B ne se produit jamais, il n'est donc pas logique de parler de la probabilité de A étant donné B.

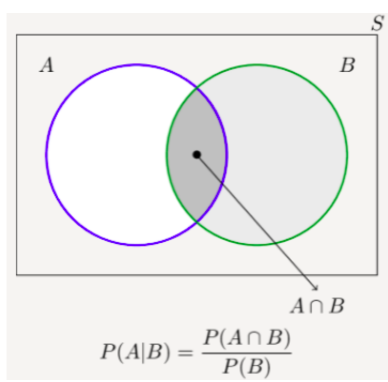


FIGURE 1.12 – Diagramme de Venn pour la probabilité conditionnelle, $\mathbb{P}(A|B)$.

Il est important de noter que la probabilité conditionnelle elle-même est une mesure de probabilité, elle satisfait donc les axiomes de probabilité. En particulier,

- **Axiome 1 :** Pour tout événement A, $\mathbb{P}(A|B) \geqslant 0$.

- **Axiome 2 :** Probabilité conditionnelle de B étant donné B est 1, c'est-à-dire, $\mathbb{P}(B|B) = 1$.
- **Axiome 3 :** Si A_1, A_2, A_3, $\cdots$ sont des événements disjoints, alors

$$\mathbb{P}(A_1 \cup A_2 \cup A_3 \cdots |B) = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B) + \mathbb{P}(A_3|B) + \cdots .$$

En fait, toutes les règles que nous avons apprises jusqu'à présent peuvent être étendues à la probabilité conditionnelle. Par exemple, les formules données dans l' exemple 1.10 peuvent être réécrites :

Exemple 1.20. Pour trois événements, A, B, et C, avec $\mathbb{P}(C) > 0$, on a :

- $\mathbb{P}(A^c|C) = 1 - \mathbb{P}(A|C)$;
- $\mathbb{P}(\emptyset|C) = 0$;
- $\mathbb{P}(A|C) \leqslant 1$;
- $\mathbb{P}(A - B|C) = \mathbb{P}(A|C) - \mathbb{P}(A \cap B|C)$;
- $\mathbb{P}(A \cup B|C) = \mathbb{P}(A|C) + \mathbb{P}(B|C) - \mathbb{P}(A \cap B|C)$;
- Si $A \subset B$ alors $\mathbb{P}(A|C) \leqslant \mathbb{P}(B|C)$.

Examinons quelques cas particuliers de probabilité conditionnelle :

Lorsque A et B sont disjoints : dans ce cas $A \cap B = \emptyset$, alors

$$\begin{aligned}\mathbb{P}(A|B) =& \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}\\ =& \frac{\mathbb{P}(\emptyset)}{\mathbb{P}(B)}\\ =& 0.\end{aligned}$$

C'est logique. En particulier, comme A et B sont disjoints, ils ne peuvent pas se produire tous les deux en même temps. Ainsi, étant donné que B s'est produit, la probabilité de A doit être nulle.

Lorsque B est un sous-ensemble de A : Si $B \subset A$, alors chaque fois que B arrive, A arrive aussi. Ainsi, étant donné que B s'est produit, nous nous attendons à ce que la probabilité de A soit un. Dans ce cas $A \cap B = B$, alors

$$\begin{aligned}\mathbb{P}(A|B) =& \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}\\ =& \frac{\mathbb{P}(B)}{\mathbb{P}(B)}\\ =& 1.\end{aligned}$$

Lorsque A est un sous-ensemble de B : dans ce cas $A \cap B = A$, alors

$$\begin{aligned}\mathbb{P}(A|B) =& \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \\ =& \frac{\mathbb{P}(A)}{\mathbb{P}(B)}.\end{aligned}$$

Exemple 1.21. Je lance un dé équilibré deux fois et j'obtiens deux nombres X_1 = résultat du premier lancer et X_2 = résultat du deuxième lancer. Étant donné que je sais $X_1 + X_2 = 7$, quelle est la probabilité que $X_1 = 4$ ou $X_2 = 4$?

Solution
Soit A l'événement $X_1 = 4$ ou $X_2 = 4$ et B l'événement $X_1 + X_2 = 7$. Nous sommes intéressés par $\mathbb{P}(A|B)$, nous pouvons donc utiliser

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

On remarque que

$$\begin{aligned} A =& \{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),(1,4),(2,4),(3,4),(5,4),(6,4)\}, \\ B =& \{(6,1),(5,2),(4,3),(3,4),(2,5),(1,6)\}, \\ A \cap B =& \{(4,3),(3,4)\}. \end{aligned}$$

Nous concluons

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\frac{2}{36}}{\frac{6}{36}} = \frac{1}{3}.$$

Examinons un célèbre problème de probabilité, appelé le problème des deux enfants. De nombreuses versions de ce problème existent.

Exemple 1.22. Considérons une famille qui a deux enfants. Nous nous intéressons aux sexes des enfants. Notre univers est $S = \{(G,G),(G,B),(B,G),(B,B)\}$. Supposons également que les quatre résultats possibles sont également probables.

1. Quelle est la probabilité que les deux enfants soient des filles sachant que le premier enfant est une fille ?

2. Nous demandons au père : « as-tu au moins une fille ? Il répond "oui !" Compte tenu de ces informations supplémentaires, quelle est la probabilité que les deux enfants soient des filles ? En d'autres termes, quelle est la probabilité que les deux enfants soient des filles sachant qu'au moins l'un d'entre eux est une fille ?

Solution
Soit A l'événement pour lequel les deux enfants sont des filles, c'est-à-dire $A = \{(G,G)\}$. Soit B le cas où le premier enfant est une fille, c'est-à-dire $B = \{(G,G),(G,B)\}$. Enfin, Soit C le cas où au moins un des enfants est une fille, c'est-à-dire $C = \{(G,G),(G,B),(B,G)\}$. Puisque les résultats sont également probables, nous pouvons écrire

$$\mathbb{P}(A) = \frac{1}{4}, \quad \mathbb{P}(B) = \frac{2}{4} = \frac{1}{2}, \quad \mathbb{P}(C) = \frac{3}{4}.$$

1. Quelle est la probabilité que les deux enfants soient des filles sachant que le premier enfant est une fille ? C'est $\mathbb{P}(A|B)$, on peut donc écrire

$$\begin{aligned}\mathbb{P}(A|B) =& \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)}{\mathbb{P}(B)} \quad \text{(puisque } A \subset B\text{)} \\ =& \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}.\end{aligned}$$

2. Quelle est la probabilité que les deux enfants soient des filles sachant qu'au moins l'un d'entre eux est une fille ? C'est $\mathbb{P}(A|C)$, on peut donc écrire

$$\begin{aligned}\mathbb{P}(A|C) =& \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)} = \frac{\mathbb{P}(A)}{\mathbb{P}(C)} \quad \text{(puisque } A \subset C\text{)} \\ =& \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.\end{aligned}$$

Discussion : Lorsqu'on leur a demandé de deviner les réponses dans l'exemple ci-dessus, beaucoup de gens devineraient que les deux $\mathbb{P}(A|B)$ et $\mathbb{P}(A|C)$ devrait être 50%. Cependant, comme on le voit $\mathbb{P}(A|B)$ est 50%, tandis que $\mathbb{P}(A|C)$ est seulement 33 pour cent. Ceci est un exemple où les réponses peuvent sembler contre-intuitives. Pour comprendre les résultats de ce problème, il est utile de noter que l'événement B est un sous-ensemble de l'événement C. En fait, il est strictement plus petit : il ne comprend pas l'élément (B, G), tandis que C possède cet élément. Ainsi l'ensemble C a plus de résultats qui ne sont pas dans A que B, ce qui signifie que $\mathbb{P}(A|C)$ devrait être plus petit que $\mathbb{P}(A|B)$.

Règle en chaîne pour la probabilité conditionnelle :

Écrivons la formule de la probabilité conditionnelle dans le format suivant

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

Ce format est particulièrement utile dans les situations où nous connaissons la probabilité conditionnelle, mais nous nous intéressons à la probabilité de l'intersection. Nous pouvons interpréter cette formule à l'aide d'un diagramme en arbre tel que celui présenté à la figure (1.13). Dans cette figure, nous obtenons la probabilité à chaque point en multipliant les probabilités sur les branches menant à ce point. Ce type de diagramme peut être très utile pour certains problèmes.

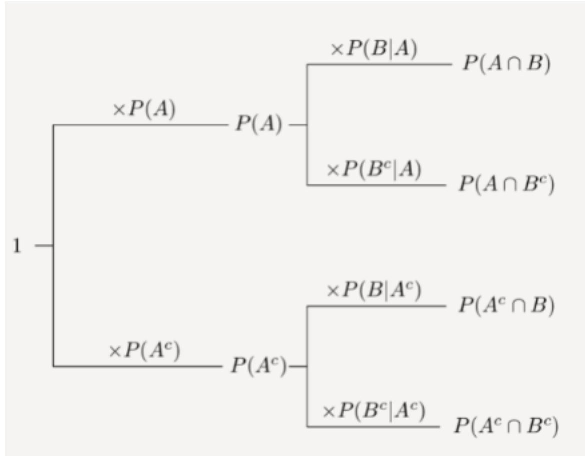


FIGURE 1.13 – Un diagramme en arbre.

Nous pouvons maintenant étendre cette formule à trois événements ou plus :

$$\begin{aligned}
\mathbb{P}(A \cap B \cap C) =& P\big(A \cap (B \cap C)\big) \\
=& \mathbb{P}(A)\mathbb{P}(B \cap C|A) \\
=& \mathbb{P}(A)\mathbb{P}(B|A)\mathbb{P}(C|A,B) \quad \text{(en conditionnant les deux côtés par } A) \\
=& \mathbb{P}(A)\mathbb{P}(B|A)\mathbb{P}(C|A,B)
\end{aligned}$$

Le point ici est de comprendre comment vous pouvez dériver ces formules et d'essayer d'avoir une intuition à leur sujet plutôt que de les mémoriser. Vous pouvez étendre l'arborescence de la Figure (1.13) à ce cas. Ici, l'arbre aura huit feuilles. Un énoncé général de la règle de la chaîne pour n événements est la suivante :

Règle en chaîne pour la probabilité conditionnelle :

$$\mathbb{P}(\mathbf{A_1} \cap \mathbf{A_2} \cap \cdots \cap \mathbf{A_n}) = \mathbb{P}(\mathbf{A_1})\mathbb{P}(\mathbf{A_2}|\mathbf{A_1})\mathbb{P}(\mathbf{A_3}|\mathbf{A_2}, \mathbf{A_1}) \cdots \mathbb{P}(\mathbf{A_n}|\mathbf{A_{n-1}}, \mathbf{A_{n-2}}, \cdots, \mathbf{A_1})$$

Exemple 1.23. Dans une usine, il y a 100 unités d'un produit, 5 sont défectueuses. Nous choisissons trois unités parmi 100 unités au hasard. Quelle est la probabilité qu'aucune d'elles ne soit défectueuse ?

Solution
Définissons A_i comme l'événement que la i-eme unité choisie n'est pas défectueuse, pour $j = 1, 2, 3$. Nous sommes intéressés par $\mathbb{P}(A_1 \cap A_2 \cap A_3)$. Notez que

$$\mathbb{P}(A_1) = \frac{95}{100}.$$

Étant donné que le premier élément choisi était bon, le deuxième élément sera choisi parmi 94 bonnes unités et 5 unités défectueuses, donc

$$\mathbb{P}(A_2|A_1) = \frac{94}{99}.$$

Étant donné que les premier et deuxième éléments choisis étaient corrects, le troisième élément sera choisi parmi 93 bonnes unités et 5 unités défectueuses, donc

$$\mathbb{P}(A_3|A_2, A_1) = \frac{93}{98}.$$

Ainsi, nous avons

$$\begin{aligned}\mathbb{P}(A_1 \cap A_2 \cap A_3) =& \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_2, A_1)\\ =& \frac{95}{100}\frac{94}{99}\frac{93}{98}\\ =& 0,8560\end{aligned}$$

Comme nous le verrons plus loin, une autre façon de résoudre ce problème est d'utiliser des arguments de comptage.

1.4.1 Indépendance

Soit A l'événement "il pleut demain", et supposons que $\mathbb{P}(A) = \frac{1}{3}$. Supposons également que nous lançons une pièce de monnaie. Soit B l'événement qu'il atterrit sur pile en haut. Nous avons $\mathbb{P}(B) = \frac{1}{2}$.

Quelle est la probabilité $\mathbb{P}(A|B)$? $\mathbb{P}(A|B) = \mathbb{P}(A) = \frac{1}{3}$. Le résultat de mon tirage au sort n'a rien à voir avec la météo de demain. Ainsi, peu importe si B arrive ou non, la probabilité de A ne devrait pas changer. Ceci est un exemple de deux événements indépendants. Deux événements sont indépendants si l'un ne véhicule aucune information sur l'autre. Donnons maintenant une définition formelle de l'indépendance.

Deux événements A et B sont indépendants si et seulement si $\mathbb{P}(\mathbf{A} \cap \mathbf{B}) = \mathbb{P}(\mathbf{A})\mathbb{P}(\mathbf{B})$.

Maintenant, réconcilions d'abord cette définition avec ce que nous avons mentionné plus tôt, $\mathbb{P}(A|B) = \mathbb{P}(A)$. Si deux événements sont indépendants, alors $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, alors

$$\begin{aligned}\mathbb{P}(A|B) =& \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}\\ =& \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)}\\ =& \mathbb{P}(A).\end{aligned}$$

Ainsi, si deux événements A et B sont indépendants et $\mathbb{P}(B) \neq 0$, alors $\mathbb{P}(A|B) = \mathbb{P}(A)$. Pour résumer, nous pouvons dire "l'indépendance signifie que nous pouvons multiplier les probabilités des événements pour obtenir la probabilité de leur intersection", ou de manière équivalente, "l'indépendance signifie que la probabilité conditionnelle d'un événement donné à un autre est la même que la probabilité (antérieure) d'origine". Parfois, l'indépendance de deux événements est assez claire parce que les deux événements semblent n'avoir aucune interaction physique l'un avec l'autre (comme les deux événements discutés ci-dessus). À d'autres moments, ce n'est pas aussi clair et nous devons vérifier s'ils satisfont à la condition d'indépendance. Prenons un exemple.

Exemple 1.24. On choisit un nombre au hasard parmi $\{1,2,3,\cdots,10\}$, et on le note N. on suppose que tous les résultats soient également probables. Soit A l'événement "N soit inférieur à 7", et soit B l'événement "N est un nombre pair". A et B sont-ils indépendants ?

Solution
Nous avons $A = \{1,2,3,4,5,6\}$, $B = \{2,4,6,8,10\}$, et $A \cap B = \{2,4,6\}$. Puis $\mathbb{P}(A) = 0,6, \quad \mathbb{P}(B) = 0,5, \quad \mathbb{P}(A \cap B) = 0,3.$
Donc, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, alors A et B sont indépendants. Cela signifie que sachant que B s'est produit ne change pas notre croyance quant à la probabilité de A. Dans ce problème, les deux événements ont à peu près le même nombre aléatoire, mais ils sont toujours indépendants car ils satisfont à la définition.

La définition de l'indépendance peut être étendue au cas de trois événements ou plus.
Trois événements A, B, et C sont indépendants si toutes les conditions suivantes sont remplies

$$\begin{aligned}\mathbb{P}(A \cap B) &= \mathbb{P}(A)\mathbb{P}(B),\\ \mathbb{P}(A \cap C) &= \mathbb{P}(A)\mathbb{P}(C),\\ \mathbb{P}(B \cap C) &= \mathbb{P}(B)\mathbb{P}(C),\\ \mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).\end{aligned}$$

Notez que les quatre conditions énoncées doivent être remplies pour que trois événements soient indépendants. En particulier, vous pouvez trouver des situations dans lesquelles trois d'entre elles tiennent, mais pas la quatrième. En général, pour que n événements A_1, A_2,$\cdots$,A_n soient indépendants il faut avoir

$$\begin{aligned}\mathbb{P}(A_i \cap A_j) &= \mathbb{P}(A_i)\mathbb{P}(A_j), \text{pour tout } i,j \in \{1,2,\cdots,n\} \text{ distinct;}\\ \mathbb{P}(A_i \cap A_j \cap A_k) &= \mathbb{P}(A_i)\mathbb{P}(A_j)\mathbb{P}(A_k), \text{pour tout distinct } i,j,k \in \{1,2,\cdots,n\};\\ &\quad . \qquad\qquad .\\ &\quad . \qquad\qquad .\\ &\quad . \qquad\qquad .\\ \mathbb{P}(A_1 \cap A_2 \cap A_3 \cdots \cap A_n) &= \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)\cdots\mathbb{P}(A_n).\end{aligned}$$

Cela peut sembler difficile, mais nous pouvons généralement affirmer que les événements sont indépendants d'une manière beaucoup plus simple. Par exemple, nous pourrions être en mesure de justifier l'indépendance en examinant la manière dont l'expérience aléatoire est réalisée. Un exemple simple d'événement indépendant est lorsque vous lancez une pièce à plusieurs reprises. Dans une telle expérience, les résultats de n'importe quel sous-ensemble de lancers de pièces n'ont aucun impact sur les autres.

Exemple 1.25. Je lance une pièce de monnaie à plusieurs reprises jusqu'à ce que j'observe les premières piles auquel point je m'arrête. Soit X le nombre total de lancers de pièces. Trouver $\mathbb{P}(X = 5)$.

Solution
Ici, le résultat de l'expérience aléatoire est un nombre X. Le but est de trouver $\mathbb{P}(A) = \mathbb{P}(5)$. Mais que signifie $X = 5$? Cela signifie que les 4 premier lancers de pièces donnent face et le cinquième donne pile. Le problème est donc de trouver la probabilité de la suite $HHHHT$ en lançant une pièce cinq fois. Notez que $HHHHT$ est un raccourci pour l'événement "(le premier lancer donne face) et (le deuxième lancer donne face) et (le troisième lancer donne face) et (le quatrième lancer donne face) et (le cinquième lancer donne pile)." Puisque tous les lancers de pièces sont indépendants, on peut écrire

$$\begin{aligned}\mathbb{P}(HHHHT) =& \mathbb{P}(H)\mathbb{P}(H)\mathbb{P}(H)\mathbb{P}(H)\mathbb{P}(T)\\ =& \frac{1}{2}.\frac{1}{2}.\frac{1}{2}.\frac{1}{2}.\frac{1}{2}\\ =& \frac{1}{32}.\end{aligned}$$

Exemple 1.26. Je dois choisir un mot de passe pour un compte d'ordinateur. La règle est que le mot de passe doit être composé de deux lettres minuscules (a à z) suivies d'une lettre majuscule (A à Z) suivies de quatre chiffres $(0, 1, \cdots, 9)$. Par exemple, ce qui suit est un mot de passe valide

$$ejT3018$$

- Trouver le nombre total de mots de passe possibles, N.
- Un hacker a été capable d'écrire un programme qui génère de manière aléatoire et indépendante 10^8 mots de passe selon la règle ci-dessus. Notez que le même mot de passe peut être généré plus d'une fois. Si l'un des mots de passe choisis au hasard correspond à mon mot de passe, il peut alors accéder aux informations de mon compte. Quelle est la probabilité qu'il réussisse à accéder aux informations de mon compte ?

Solution
Pour choisir un mot de passe, je dois d'abord choisir une lettre minuscule, puis une autre lettre minuscule, puis une lettre majuscule, et enfin 4 chiffres. Il y a 26 minuscules, 26 majuscules, et dix chiffres. Ainsi, par le principe de multiplication, le nombre total de mots de passe valides possibles est

$$N = 26 \times 26 \times 26 \times 10 \times 10 \times 10 \times 10 = 26^3 \times 10^4$$

Soit G_i l'événement que la conjecture (supposition ou jugement) du pirate informatique i corresponde à la mienne, pour $i = 1, 2, \cdots, 10^8$. La probabilité que le $i-$ème mot de passe choisi au hasard corresponde au mien est

$$\mathbb{P}(G_i) = \frac{1}{N}$$

Maintenant, soit $\mathbb{P}_{pirate}$ la probabilité que le pirate réussisse, c'est-à-dire qu'au moins un des mots de passe choisis au hasard corresponde au mien. Rappelons que ≪ au moins ≫ signifie union :

$$\mathbb{P}_{pirate} = \mathbb{P}\left(\bigcup_i G_i\right)$$

A noter que les événements G_i sont indépendants puisque les suppositions sont générées indépendamment, mais ils ne sont pas disjoints puisque plusieurs suppositions pourraient être correctes si le programme du pirate génère le même mot de passe. Par conséquent, dans ce cas, il est plus facile de travailler avec des intersections qu'avec des unions, nous trouverons donc d'abord la probabilité de l'événement complémentaire :

$$\begin{aligned}
\mathbb{P}\left(\bigcup_i G_i\right)^c =& \mathbb{P}\left(\bigcap_i G_i^c\right) \\
=& \prod_{i=1}^{N} \mathbb{P}(G_i^c) \qquad \text{(par indépendance)} \\
=& \left(1 - \frac{1}{N}\right)^{10^8}
\end{aligned}$$

Donc,

$$\begin{aligned}
\mathbb{P}_{pirate} =& 1 - \left(1 - \frac{1}{N}\right)^{10^8} \\
=& 1 - \left(1 - \frac{1}{26^3 \times 10^4}\right)^{10^8} \\
=& 0,4339
\end{aligned}$$

Nous avons vu que deux événements A et B sont indépendants si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Dans les deux résultats suivants, nous examinons ce que l'indépendance peut nous dire sur d'autres opérations ensemblistes telles que les compléments et les unions.

Lemme 1.1. *Si A et B sont indépendants, alors*

- *A et B^c sont indépendants,*
- *A^c et B sont indépendants,*
- *A^c et B^c sont indépendants.*

Démonstration. Nous démontrons la première car les autres peuvent être immédiatement déduites de la première. Nous avons

$$\begin{aligned}
\mathbb{P}(A \cap B^c) =& \mathbb{P}(A - B) \\
=& \mathbb{P}(A) - \mathbb{P}(A \cap B) \\
=& \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) \qquad \text{puisque } A \text{ et } B \text{ sont indépendants} \\
=& \mathbb{P}(A)(1 - \mathbb{P}(B)) \\
=& \mathbb{P}(A)\mathbb{P}(B^c).
\end{aligned}$$

Ainsi, A et B^c sont indépendants. □

Attention ! Une erreur courante consiste à confondre indépendance et disjonction. Ce sont des concepts complètement différents. Quand deux événements A et B sont disjoints, cela signifie que si l'un d'eux se produit, l'autre ne peut pas se produire, c'est-à-dire $A \cap B = \emptyset$. Ainsi, l'événement A donne généralement beaucoup d'informations sur l'événement B ce qui signifie qu'ils ne peuvent pas être indépendants. Précisons.

Lemme 1.2. *Considérez deux événements A et B, avec $\mathbb{P}(A) \neq 0$ et $\mathbb{P}(B) \neq 0$. Si A et B sont disjoints, alors ils ne sont pas indépendants.*

Démonstration. Comme A et B sont disjoints, nous avons

$$\mathbb{P}(A \cap B) = 0 \neq \mathbb{P}(A)\mathbb{P}(B).$$

Ainsi, A et B ne sont pas indépendants. □

Le tableau (1.2) résume les deux concepts de disjonction et d'indépendance.

TABLE 1.2 – Différences entre disjonction et indépendance.

Concept	**Signification**	**Formules**
Disjoint	A et B ne peuvent pas se produire en même temps	$A \cap B = \emptyset$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$
Indépendant	A ne donne aucune information sur B	$\mathbb{P}(A\|B) = \mathbb{P}(A)$, $\mathbb{P}(B\|A) = \mathbb{P}(B)$, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

1.4.2 Loi de probabilité totale

Commençons cette section en posant une question très simple : dans un certain pays, il y a trois provinces, appelons-les B_1, B_2, et B_3 (c'est-à-dire que le pays est divisé en trois ensembles disjoints B_1, B_2, et B_3). Nous nous intéressons à la superficie forestière totale du pays. Supposons que nous sachions que la superficie forestière de B_1, B_2, et B_3 sont $100\ km^2$, $50\ km^2$, et $150\ km^2$, respectivement. La superficie forestière totale du pays est

$$100\ km^2 + 50\ km^2 + 150\ km^2 = 300\ km^2.$$

Autrement dit, vous pouvez simplement ajouter des zones forestières dans chaque province (partition) pour obtenir la superficie forestière de tout le pays. C'est l'idée derrière la loi de probabilité totale, dans laquelle la superficie de la forêt est remplacée par la probabilité d'un événement A. En particulier, si vous voulez trouver $\mathbb{P}(A)$, vous pouvez regarder une partition de S, et ajouter la quantité de probabilité de A qui tombe dans chaque partition. Nous avons déjà vu le cas particulier où la partition est B et B^c : nous avons vu que pour deux événements quelconques A et B,

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

et en utilisant la définition de la probabilité conditionnelle, $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$, nous pouvons écrire

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

Nous pouvons énoncer une version plus générale de cette formule qui s'applique à une partition générale de l'univers S.

Loi de probabilité totale : Si B_1, B_2, B_3,$\cdots$ est une partition de l'univers S, alors pour tout événement A on a

$$\mathbb{P}(\mathbf{A}) = \sum_{\mathbf{i}} \mathbb{P}(\mathbf{A} \cap \mathbf{B_i}) = \sum_{\mathbf{i}} \mathbb{P}(\mathbf{A}|\mathbf{B_i})\mathbb{P}(\mathbf{B_i}).$$

À l'aide d'un diagramme de Venn, nous pouvons voir de manière imagée l'idée derrière la loi de probabilité totale. Dans la figure (1.14), nous avons

$$A_1 = A \cap B_1,\ A_2 = A \cap B_2,\ A_3 = A \cap B_3.$$

Comme on peut le voir sur la figure, A_1, A_2, et A_3 forment une partition de l'ensemble A, et donc par le troisième axiome de probabilité

$$\mathbb{P}(A) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3).$$

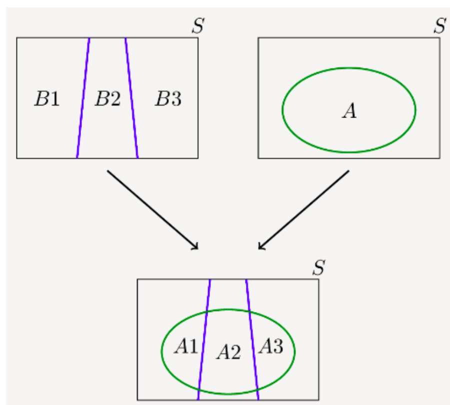


FIGURE 1.14 – Loi de probabilité totale.

Voici une preuve de la loi de probabilité totale en utilisant des axiomes de probabilité :
Voici un scénario typique dans lequel nous utilisons la loi de probabilité totale. On s'intéresse à la probabilité d'un événement A, mais nous ne savons pas comment trouver $\mathbb{P}(A)$ directement. Au lieu de cela, nous connaissons la probabilité conditionnelle de A étant donné certains événements B_i, où les B_i forment une partition de l'univers. Ainsi, nous pourrons trouver $\mathbb{P}(A)$ en utilisant la loi de probabilité totale, $\mathbb{P}(A) = \sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i)$.

Exemple 1.27. J'ai trois sacs qui contiennent chacun 100 billes :

- Le sac 1 a 75 billes rouges et 25 billes bleues ;
- Le sac 2 a 60 billes rouges et 40 billes bleues ;
- Le sac 3 a 45 billes rouges et 55 billes bleues.

Je choisis un des sacs au hasard, puis je choisis une bille dans le sac choisi, également au hasard. Quelle est la probabilité que la bille choisie soit rouge ?

Solution

Soit R l'événement où la bille choisie est rouge. Soit B_i l'événement que je choisisse le sac i. Nous savons déjà que

$$\begin{aligned}\mathbb{P}(R|B_1) =&0,75,\\ \mathbb{P}(R|B_2) =&0,60,\\ \mathbb{P}(R|B_3) =&0,45\end{aligned}$$

Nous choisissons notre partition comme B_1, B_2, B_3. Notez qu'il s'agit d'une partition valide car, premièrement, les B_i sont disjoints (un seul d'entre eux peut se produire), et deuxièmement, parce que leur union est l'ensemble fondamental (univers) car l'un des sacs sera choisi à coup sûr, c'est-à-dire, $\mathbb{P}(B_1 \cup B_2 \cup B_3) = 1$. En utilisant la loi de probabilité totale, on peut écrire

$$\begin{aligned}\mathbb{P}(R) =&\mathbb{P}(R|B_1)\mathbb{P}(B_1) + \mathbb{P}(R|B_2)\mathbb{P}(B_2) + \mathbb{P}(R|B_3)\mathbb{P}(B_3)\\ =&(0,75)\frac{1}{3} + (0,60)\frac{1}{3} + (0,45)\frac{1}{3}\\ = 0,60\end{aligned}$$

1.4.3 Règle de Bayes

Nous sommes maintenant prêts à énoncer l'un des résultats les plus utiles en probabilité conditionnelle : la règle de Bayes. Supposons que nous sachions $\mathbb{P}(A|B)$, mais on s'intéresse à la probabilité $\mathbb{P}(B|A)$. En utilisant la définition de la probabilité conditionnelle, nous avons

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

En divisant par $\mathbb{P}(A)$, on obtient

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

qui est la fameuse règle de Bayes. Souvent, pour trouver $\mathbb{P}(A)$ dans la formule de Bayes, nous devons utiliser la loi de probabilité totale, donc parfois la règle de Bayes est énoncée comme

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i)},$$

où B_1, B_2,$\cdots$, B_n forment une partition de l'univers.

Règle de Bayes

- Pour deux événements A et B, où $\mathbb{P}(A) \neq 0$, on a

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

- Si B_1, B_2, B_3,$\cdots$ forment une partition de l'univers S, et A est un événement avec $\mathbb{P}(A) \neq 0$, on a
$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i)}.$$

Exemple 1.28. Dans l' exemple (1.27) , supposons que l'on observe que la bille choisie est rouge. Quelle est la probabilité que le sac 1 ait été choisi ?

Solution
Ici, nous savons $\mathbb{P}(R|B_i)$ mais nous nous intéressons $\mathbb{P}(B_1|R)$, c'est donc un scénario dans lequel nous pouvons utiliser la règle de Bayes. Nous avons

$$\begin{aligned}\mathbb{P}(B_1|R) =& \frac{\mathbb{P}(R|B_1)\mathbb{P}(B_1)}{\mathbb{P}(R)}\\ =& \frac{0.75 \times \frac{1}{3}}{0.6}\\ =& \frac{5}{12}.\end{aligned}$$

$\mathbb{P}(R)$ a été obtenue en utilisant la loi de probabilité totale de l'exemple 1.24, nous n'avons donc pas eu à la recalculer ici. Aussi, notez que $\mathbb{P}(B_1|R) = \frac{5}{12} > \frac{1}{3}$. Cela a du sens intuitivement car le sac 1 est le sac avec le plus grand nombre de billes rouges. Ainsi si la bille choisie est rouge, il est plus probable que le sac 1 ait été choisi.

Exemple 1.29 (Paradoxe du faux positif)**.** Une certaine maladie affecte environ 1 personne sur 10 000. Il y a un test pour vérifier si la personne a la maladie. Le test est assez précis. En particulier, nous savons que

- la probabilité que le résultat du test soit positif (suggérant que la personne a la maladie), étant donné que la personne n'a pas la maladie, n'est que de 2%;
- la probabilité que le résultat du test soit négatif (suggérant que la personne n'a pas la maladie), étant donné que la personne a la maladie, n'est que de 1%.

Une personne au hasard est testée pour la maladie et le résultat est positif. Quelle est la probabilité que la personne soit atteinte de la maladie ?

Solution
Soit D l'événement que la personne a la maladie, et soit T l'événement pour le cas où le résultat du test est positif. Nous savons

$$\begin{aligned}&\mathbb{P}(D) = \frac{1}{10000},\\ &\mathbb{P}(T|D^c) = 0,02,\\ &\mathbb{P}(T^c|D) = 0,01\end{aligned}$$

Ce que nous voulons calculer est $\mathbb{P}(D|T)$. Encore une fois, nous utilisons la règle de Bayes :

$$
\begin{aligned}
\mathbb{P}(D|T) =&\frac{\mathbb{P}(T|D)\mathbb{P}(D)}{\mathbb{P}(T|D)\mathbb{P}(D)+\mathbb{P}(T|D^c)\mathbb{P}(D^c)}\\
=&\frac{(1-0,01)\times 0,0001}{(1-0,01)\times 0,0001+0,02\times(1-0,0001)}\\
=&0,0049
\end{aligned}
$$

Cela signifie qu'il y a moins d'un demi pour cent de chances que la personne soit atteinte de la maladie.

Discussion : Cela peut sembler quelque peu contre-intuitif car nous savons que le test est assez précis. Le fait est que la maladie est également très rare. Ainsi, il y a ici deux forces concurrentes, et puisque la rareté de la maladie (1 sur 10 000) est plus forte que la précision du test (98 ou 99%), il y a encore de bonnes chances que la personne n'ait pas la maladie.

1.4.4 Indépendance conditionnelle

Comme nous l'avons mentionné précédemment, presque tous les concepts définis pour la probabilité peuvent également être étendus à la probabilité conditionnelle. Rappelez-vous que deux événements A et B sont indépendants si

$$\mathbb{P}(A\cap B)=\mathbb{P}(A)\mathbb{P}(B), \quad \text{ou de manière équivalente, } \mathbb{P}(A|B)=\mathbb{P}(A).$$

On peut étendre ce concept à des événements conditionnellement indépendants. En particulier,

Définition 1.2. Deux événements A et B sont conditionnellement indépendants étant donné un événement C avec $\mathbb{P}(C)>0$ si

$$\mathbb{P}(A\cap B|C)=\mathbb{P}(A|C)\mathbb{P}(B|C) \tag{1.8}$$

1.4.5 Problèmes résolus : probabilité conditionnelle

Dans les problèmes de dés et de pièces, sauf indication contraire, il est supposé que les pièces et les dés sont équilibrés et que les essais répétés sont indépendants.

1. Vous achetez un certain produit. Le manuel indique que la durée de vie T du produit, défini comme la durée (en années) pendant laquelle le produit fonctionne correctement jusqu'à ce qu'il tombe en panne, satisfait

   $$\mathbb{P}(T\geqslant t)=e^{-\frac{t}{5}}, \text{ pour tout } t\geqslant 0.$$

   Par exemple, la probabilité que le produit dure plus de (ou égale à) 2 ans est $\mathbb{P}(T\geqslant 2)=e^{-\frac{2}{5}}=0,6703$. J'achète le produit et l'utilise depuis deux ans sans aucun problème. Quelle est la probabilité qu'il tombe en panne la troisième année ?

   Solution
   Soit A l'événement qu'un produit acheté tombe en panne au cours de la troisième année.

Aussi, soit B l'événement qu'un produit acheté ne tombe pas en panne au cours des deux premières années. Nous sommes intéressés par $\mathbb{P}(A|B)$. Nous avons

$$\mathbb{P}(B) = \mathbb{P}(T \geqslant 2) = e^{-\frac{2}{5}}.$$

Nous avons également

$$\mathbb{P}(A) = P(2 \leqslant T \leqslant 3) = \mathbb{P}(T \geqslant 2) - \mathbb{P}(T \geqslant 3) = e^{-\frac{2}{5}} - e^{-\frac{3}{5}}.$$

Enfin, comme $A \subset B$, on a $A \cap B = A$. Donc,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{e^{-\frac{2}{5}} - e^{-\frac{3}{5}}}{e^{-\frac{2}{5}}} = 0,1813$$

2. Vous lancez une pièce de monnaie équilibrée trois fois :
   (a) Quelle est la probabilité de trois faces, HHH?
   (b) Quelle est la probabilité que vous observiez exactement une face ?
   (c) Sachant que vous avez observé au moins une face, quelle est la probabilité que vous observiez au moins deux faces ?

   Solution
   Nous supposons que les tirages au sort sont indépendants.
   (a) $\mathbb{P}(HHH) = \mathbb{P}(H) \cdot \mathbb{P}(H) \cdot \mathbb{P}(H) = 0,5^3 = \frac{1}{8}$.
   (b) Pour trouver la probabilité d'exactement une face, on peut écrire

$$\begin{aligned}\mathbb{P}(\text{Une face}) =& \mathbb{P}(HTT \cup THT \cup TTH)\\ =& \mathbb{P}(HTT) + \mathbb{P}(THT) + \mathbb{P}(TTH)\\ =& \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\\ =& \frac{3}{8}.\end{aligned}$$

   (c) Sachant que vous avez observé au moins une face, quelle est la probabilité que vous observiez au moins deux faces ? Soit A_1 l'événement que vous observez au moins une face, et A_2 l'événement que vous observez au moins deux faces. Puis

$$\begin{aligned}A_1 = S - \{TTT\}, \text{ et } \mathbb{P}(A_1) = \frac{7}{8};\\ A_2 = \{HHT, HTH, THH, HHH\}, \text{ et } \mathbb{P}(A_2) = \frac{4}{8}.\end{aligned}$$

   Ainsi, nous pouvons écrire

$$\mathbb{P}(A_2|A_1) = \frac{\mathbb{P}(A_2 \cap A_1)}{\mathbb{P}(A_1)} = \frac{\mathbb{P}(A_2)}{\mathbb{P}(A_1)} = \frac{4}{8}.\frac{8}{7} = \frac{4}{7}.$$

3. Dans une ville, il pleut un tiers des jours. Étant donné qu'il pleut, il y aura un trafic important avec probabilité de $\frac{1}{2}$, et étant donné qu'il ne pleut pas, il y aura un trafic important avec probabilité de $\frac{1}{4}$. S'il pleut et qu'il y a beaucoup de circulation, vous arrivez en retard au travail avec probabilité de $\frac{1}{2}$. En revanche, la probabilité d'être en retard est réduite à $\frac{1}{8}$ s'il ne pleut pas et qu'il n'y a pas de circulation dense. Dans d'autres situations (pluvieux et pas de circulation, pas de pluie et circulation), la probabilité d'être en retard est $0,25$. Vous choisissez un jour au hasard.

   (a) Quelle est la probabilité qu'il ne pleuve pas et que la circulation soit dense et que vous ne soyez pas en retard ?

   (b) Quelle est la probabilité que vous soyez en retard ?

   (c) Étant donné que vous êtes arrivé en retard au travail, quelle est la probabilité qu'il ait plu ce jour-là ?

   Solution

   Soit R l'événement qu'il pleut, T l'événement qu'il y a un trafic lourd, et L l'événement que je suis en retard au travail. Comme on le voit dans l'énoncé du problème, on nous donne des probabilités conditionnelles sous forme de chaîne. Ainsi, il est utile de dessiner un diagramme en arbre. La figure (1.15) montre un diagramme en arbre pour ce problème. Dans cette figure, chaque feuille de l'arbre correspond à un seul résultat dans l'univers. Nous pouvons calculer les probabilités de chaque résultat dans l'espace échantillon en multipliant les probabilités sur les bords de l'arbre qui mènent au résultat correspondant.

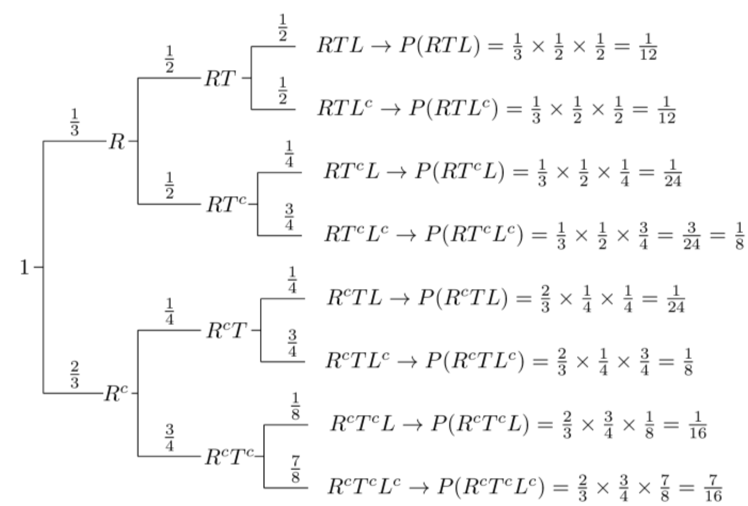


FIGURE 1.15 – Diagramme en arbre.

   (a) La probabilité qu'il ne pleuve pas et qu'il y ait beaucoup de circulation et qu' on ne soit pas en retard peut être trouvée à l'aide de l'arbre qui applique en fait la règle de la chaîne :

$$\mathbb{P}(R^c \cap T \cap L^c) = \mathbb{P}(R^c)\mathbb{P}(T|R^c)\mathbb{P}(L^c|R^c \cap T) = \frac{2}{3} \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{1}{8}.$$

(b) La probabilité qu'on soit en retard peut être trouvée à partir de l'arbre. Tout ce que nous avons à faire est d'additionner les probabilités des résultats qui correspondent au retard. En fait, nous utilisons ici la loi de la probabilité totale.

$$\begin{aligned}\mathbb{P}(L) =&\mathbb{P}(R,T,L)+\mathbb{P}(R,T^c,L)+\mathbb{P}(R^c,T,L)+\mathbb{P}(R^c,T^c,L)\\ =&\frac{1}{12}+\frac{1}{24}+\frac{1}{24}+\frac{1}{16}\\ =&\frac{11}{48}.\end{aligned}$$

(c) Nous pouvons trouver $\mathbb{P}(R|L)$ en utilisant $\mathbb{P}(R|L)=\frac{\mathbb{P}(R\cap L)}{\mathbb{P}(L)}$. Nous avons déjà trouvé $\mathbb{P}(L)=\frac{11}{48}$, et on peut trouver $\mathbb{P}(R\cap L)$ de même en ajoutant les probabilités des résultats qui appartiennent à $R\cap L$. En particulier,

$$\mathbb{P}(R\cap L)=\mathbb{P}(R,T,L)+\mathbb{P}(R,T^c,L)=\frac{1}{12}+\frac{1}{24}=\frac{1}{8}.$$

Ainsi, on obtient

$$\mathbb{P}(R|L)=\frac{\mathbb{P}(R\cap L)}{\mathbb{P}(L)}=\frac{1}{8}.\frac{48}{11}=\frac{6}{11}.$$

1.4.6 Exercices

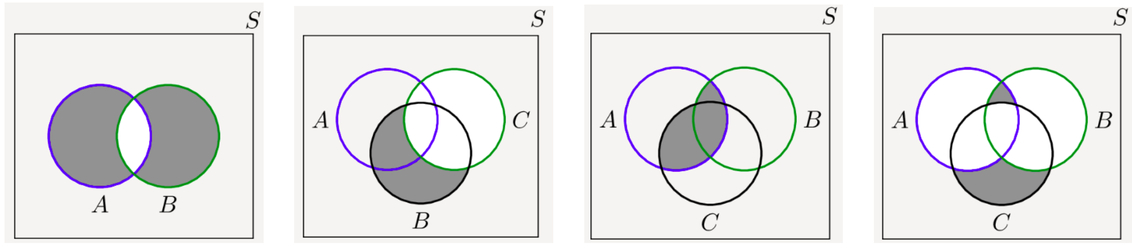
1. Supposons que l'univers S est défini comme $S=\{1,2,\cdots,10\}$ et $A=\{1,2,3\}$, $B=\{X\in S:2\leqslant X\leqslant 7\}$, et $C=\{7,8,9,10\}$.

   (a) Trouver $A\cup B$.
   (b) Trouver $(A\cup C)-B$.
   (c) Trouver $\bar{A}\cup(B-C)$.
   (d) Les ensembles A, B ,et C forment-ils une partition de S?

2. En travaillant avec des nombres réels, l'univers est $\mathbb{R}$. Trouvez chacun des ensembles suivants.

   (a) $[6,8]\cup[2,7[$
   (b) $[6,8]\cap[2,7[$
   (c) $[0,1]^c$
   (d) $[6,8]-]2,7[$

3. Pour chacun des diagrammes de Venn dans la figure (1.16), écrivez l'ensemble indiqué par la zone ombrée.



(a) (b) (c) (d)

FIGURE 1.16

4. Une pièce est lancée deux fois. Soit S l'ensemble de toutes les paires possibles qui peuvent être observées, c'est-à-dire $S = \{H, T\} \times \{H, T\} = \{(H, H), (H, T), (T, H), (T, T)\}$. Écrire les ensembles suivants en listant leurs éléments.

   (a) A : Le premier lancer donne face.
   (b) B : Au moins une pile est observée.
   (c) C : Les deux lancers de pièces donnent des résultats différents.

5. Soit $A = \{1, 2, \cdots, 100\}$. Pour tout $i \in \mathbb{N}$, Définir A_i comme l'ensemble des nombres dans A qui sont divisibles par i. Par exemple :

$$A_2 = \{2, 4, 6, \cdots, 100\}, \quad A_3 = \{3, 6, 9, \cdots, 99\}.$$

   (a) Trouver $|A_2|$, $|A_3|$, $|A_4|$, $|A_5|$.
   (b) Trouver $|A_2 \cup A_3 \cup A_5|$.

6. Supposer que A_1, A_2, A_3 forment une partition de l'univers S. Soit B un ensemble arbitraire. Supposons que nous sachions

$$\begin{aligned} |B \cap A_1| &= 10, \\ |B \cap A_2| &= 20, \\ |B \cap A_3| &= 15. \end{aligned}$$

   Trouver $|B|$.

7. Soient A et B deux événements tels que

$$\mathbb{P}(A) = 0,4, \ \mathbb{P}(B) = 0,7, \ \mathbb{P}(A \cup B) = 0,9$$

   (a) Trouver $\mathbb{P}(A \cap B)$.
   (b) Trouver $\mathbb{P}(A^c \cap B)$.
   (c) Trouver $\mathbb{P}(A - B)$.
   (d) Trouver $\mathbb{P}(A^c - B)$.
   (e) Trouver $\mathbb{P}(A^c \cup B)$.

(f) Trouver $\mathbb{P}(A \cap (B \cup A^c))$.

8. Je lance deux fois un dé et j'obtiens deux nombres : $X_1 =$ résultat du premier lancer, $X_2 =$ résultat du deuxième lancer.

   (a) Trouvez la probabilité que $X_2 = 4$.

   (b) Trouvez la probabilité que $X_1 + X_2 = 7$.

   (c) Trouvez la probabilité que $X_1 \neq 2$ et $X_2 \geqslant 4$.

9. Dans une usine, il y a 100 unités d'un certain produit, dont 5 sont défectueuses. Nous choisissons trois unités parmi les 100 unités au hasard. Quelle est la probabilité qu'exactement l'une d'entre elles soit défectueuse ?

10. Je lance deux fois un dé et j'obtiens deux nombres X et Y. Soit A l'événement $X = 2$, B l'événement $X + Y = 7$, et C l'événement $Y = 3$.

    (a) A et B sont-ils indépendant ?

    (b) A et C sont-ils indépendant ?

    (c) B et C sont-ils indépendant ?

    (d) A, B et C sont-ils indépendant ?

Chapitre 2

Variables aléatoires

2.1 Notions, définitions et exemples

En général, pour analyser des expériences aléatoires, nous nous concentrons généralement sur certains aspects numériques de l'expérience. Par exemple, dans un match de football, nous pouvons être intéressés par le nombre de buts, de tirs, de tirs au but, de coups de pied de coin, de fautes, etc. Si nous considérons un match de football entier comme une expérience aléatoire, alors chacun de ces résultats numériques donne une certaine information sur le résultat de l'expérience aléatoire. Ce sont des exemples de **variables aléatoires**. En un mot, une variable aléatoire est une variable à valeur réelle dont la valeur est déterminée par une expérience aléatoire sous-jacente.

Exemple 2.1. Je lance une pièce cinq fois. Il s'agit d'une expérience aléatoire et l'univers peut être écrit comme

$$S = \{TTTTT, TTTTH, \cdots, HHHHH\}.$$

Notez qu'ici l'univers S possède $2^5 = 32$ éléments. Supposons que dans cette expérience, nous nous intéressons au nombre de faces. On peut définir une variable aléatoire X dont la valeur est le nombre de faces observées. La valeur de X sera l'une des 0, 1, 2, 3, 4 ou alors 5 en fonction du résultat de l'expérience aléatoire.

Essentiellement, une variable aléatoire est une fonction à valeur réelle qui attribue une valeur numérique à chaque résultat possible de l'expérience aléatoire. Par exemple, la variable aléatoire X définie ci-dessus attribue la valeur 0 au résultat $TTTTT$, la valeur 2 au résultat $THHT$, etc. D'où la variable aléatoire X est une fonction de l'univers $S = \{TTTTT, TTTTH, \cdots, HHHHH\}$ aux nombres réels (pour cette variable aléatoire particulière, les valeurs sont toujours des nombres entiers entre 0 et 5).

Définition 2.1 (Variables aléatoires). Une variable aléatoire X est une fonction de l'univers aux nombres réels.

$$X : S \longrightarrow \mathbb{R}$$

Nous notons généralement les variables aléatoires par des lettres majuscules telles que X, Y, et Z. Puisqu'une variable aléatoire est une fonction, on peut parler de son image. L'image d'une variable aléatoire X, notée $Im(X)$ ou alors R_X, est l'ensemble des valeurs possibles de X. Dans l'exemple ci-dessus, $Im(X) = R_X = \{0, 1, 2, 3, 4, 5\}$.

L'image d'une variable aléatoire X, notée $Im(X)$ ou alors R_X, est l'ensemble des valeurs possibles de X.

Exemple 2.2. Trouvez l'image de chacune des variables aléatoires suivantes.

1. Je lance une pièce 100 fois. Soit X le nombre de faces que j'observe.
2. Je lance une pièce jusqu'à ce que les premières piles apparaissent. Soit Y le nombre total de lancers de pièces.
3. La variable aléatoire T est définie comme le temps (en heures) à partir de maintenant jusqu'à ce que le prochain tremblement de terre se produise dans une certaine ville.

Solution

1. La variable aléatoire X peut prendre n'importe quel entier de 0 à 100, alors $R_X = \{0, 1, 2, ..., 100\}$.
2. La variable aléatoire Y peut prendre n'importe quel entier positif, donc $R_Y = \{1, 2, 3, ...\} = \mathbb{N}$.
3. La variable aléatoire T peut en théorie prendre n'importe quel nombre réel positif, donc $R_T = [0, \infty[$.

Il existe deux classes importantes de variables aléatoires : les **variables aléatoires discrètes** et les **variables aléatoires continues**.

Définition 2.2. X est une variable aléatoire **discrète** si son image est dénombrable. Si X est une fonction continue, on dit que X est une variable aléatoire **continue**.

N'oubliez pas qu'un ensemble A est dénombrable s'il vérifie l'une des conditions suivantes :

1. A est un ensemble fini tel que $\{1, 2, 3, 4\}$, ou alors
2. il peut être mis en correspondance un à un (bijection) avec des nombres naturels (dans ce cas, l'ensemble est dit dénombrable infini)

En particulier, comme nous l'avons vu au chapitre 1, des ensembles tels que $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ et leurs sous-ensembles sont dénombrables, tandis que des ensembles tels que des intervalles non vides $[a, b]$ de $\mathbb{R}$ sont indénombrables. Une variable aléatoire est discrète si son domaine est un ensemble dénombrable. Dans l' exemple (2.2), les variables aléatoires X et Y sont discrètes, tandis que la variable aléatoire T n'est pas discrète.

2.2 Variables aléatoires discrètes

2.2.1 Fonction de masse de probabilité (PMF)

Si X est une variable aléatoire discrète alors son image R_X est un ensemble dénombrable, nous pouvons donc lister les éléments de R_X. En d'autres termes, nous pouvons écrire

$$R_X = \{x_1, x_2, x_3, ...\}.$$

Notez qu'ici x_1, x_2, x_3, ... sont les valeurs possibles de la variable aléatoire X. Alors que les variables aléatoires sont généralement désignées par des lettres majuscules, pour représenter les nombres dans l'image, nous utilisons généralement des lettres minuscules telles que x, x_1, y, z, etc. Pour une variable aléatoire discrète X, nous sommes intéressés à connaître les probabilités de $X = x_k$. Notez qu'ici, l'événement $A = \{X = x_k\}$ est défini comme l'ensemble des résultats s de l'univers S pour lequel la valeur correspondante de X est égale à x_k. En particulier,

$$A = \{s \in S | X(s) = x_k\}.$$

Les probabilités des événements $\{X = x_k\}$ sont formellement représentées par la **fonction de masse de probabilité (pmf)** de X.

Définition 2.3. Soit X une variable aléatoire discrète avec un co-domaine $R_X = \{x_1, x_2, x_3, ...\}$ (fini ou dénombrable infini). La fonction

$$\mathbb{P}_X(x_k) = \mathbb{P}(X = x_k), \text{ pour } k = 1, 2, 3, ...,$$

est appelée la *fonction de masse de probabilité (PMF)* de X.

Ainsi, le PMF est une mesure de probabilité qui nous donne des probabilités des valeurs possibles pour une variable aléatoire. Alors que la notation ci-dessus est la notation standard pour le PMF de X, cela peut sembler déroutant au premier abord. L'indice X indique ici qu'il s'agit du PMF de la variable aléatoire X. Ainsi, par exemple, $\mathbb{P}_X(1)$ montre la probabilité que $X = 1$. Pour mieux comprendre tous les concepts ci-dessus, examinons quelques exemples.

Exemple 2.3. Je lance deux fois une pièce équilibrée de monnaie. Soit X la variable définie comme le nombre de faces que j'observe. Trouvez l'image de X, i.e R_X, ainsi que sa fonction de masse de probabilité $\mathbb{P}_X$.

Solution
Ici, notre univers est donné par

$$S = \{HH, HT, TH, TT\}.$$

Le nombre de faces sera 0,1 ou 2. Ainsi

$$R_X = \{0, 1, 2\}.$$

Puisqu'il s'agit d'un ensemble fini (et donc dénombrable), la variable aléatoire X est une variable aléatoire discrète. Ensuite, nous devons trouver le PMF de X. Le PMF est défini comme

$$\mathbb{P}_X(k) = \mathbb{P}(X = k) \text{ pour } k = 0, 1, 2.$$

Nous avons

$$\mathbb{P}_X(0) = \mathbb{P}(X = 0) = \mathbb{P}(TT) = \frac{1}{4},$$
$$\mathbb{P}_X(1) = \mathbb{P}(X = 1) = \mathbb{P}(\{HT, TH\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2},$$
$$\mathbb{P}_X(2) = \mathbb{P}(X = 2) = \mathbb{P}(HH) = \frac{1}{4}.$$

Bien que le PMF soit généralement défini pour des valeurs comprises dans le co-domaine, il est parfois pratique d'étendre le PMF de X à tous les nombres réels. Si $x \notin R_X$, on peut simplement écrire $\mathbb{P}_X(x) = \mathbb{P}(X = x) = 0$. Ainsi, en général on peut écrire

$$\mathbb{P}_X(x) = \begin{cases} \mathbb{P}(X = x) & \text{si } x \text{ est dans } R_X \\ 0 & \text{sinon} \end{cases}$$

Pour mieux visualiser le PMF, nous pouvons le tracer. La figure (2.1) montre le PMF de la variable aléatoire X ci-dessus. Comme on le voit, la variable aléatoire peut prendre trois valeurs possibles 0, 1 et 2. La figure indique également clairement que l'événement $X = 1$ est deux fois plus probable que les deux autres valeurs possibles. La figure peut être interprétée de la manière suivante : si nous répétons l'expérience aléatoire (lancer une pièce deux fois) un grand nombre de fois, alors environ la moitié des fois nous observons $X = 1$, environ un quart des fois on observe $X = 0$, et environ un quart des fois on observe $X = 2$.

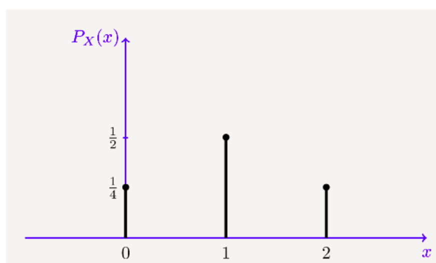


FIGURE 2.1 – PMF pour variable aléatoire X dans l'exemple 3.3.

Pour les variables aléatoires discrètes, le PMF est également appelé **distribution de probabilité**. Ainsi, lorsqu'on demande de trouver la distribution de probabilité d'une variable aléatoire discrète X, nous pouvons le faire en trouvant son PMF. L'expression *fonction de distribution* est généralement réservée exclusivement à la fonction de distribution cumulative CDF (telle que définie plus loin dans ce cours). Le mot *distribution* dans ce cours est utilisé dans un sens plus large et pourrait faire référence à PMF, fonction de densité de probabilité (PDF) ou CDF.

Exemple 2.4. J'ai une pièce non équilibrée pour laquelle $\mathbb{P}(H) = p$, où $0 < p < 1$. Je lance la pièce à plusieurs reprises jusqu'à ce que j'observe une face pour la première fois. Soit Y le nombre total de lancers de pièces. Trouver la distribution de Y.

Solution
Notons tout d'abord que la variable aléatoire Y peut potentiellement prendre n'importe quel entier

positif, nous avons donc $R_Y = \mathbb{N} = \{1, 2, 3, ...\}$. Pour trouver la distribution de Y, nous devons trouver $\mathbb{P}_Y(k) = \mathbb{P}(Y = k)$ pour $k = 1, 2, 3, ...$ Nous avons

$$\begin{aligned}
\mathbb{P}_Y(1) &= \mathbb{P}(Y = 1) = \mathbb{P}(H) = p,\\
\mathbb{P}_Y(2) &= \mathbb{P}(Y = 2) = \mathbb{P}(TH) = (1-p)p,\\
\mathbb{P}_Y(3) &= \mathbb{P}(Y = 3) = \mathbb{P}(TTH) = (1-p)^2 p,\\
&\cdot \quad\quad \cdot \quad\quad \cdot \quad\quad \cdot\\
&\cdot \quad\quad \cdot \quad\quad \cdot \quad\quad \cdot\\
&\cdot \quad\quad \cdot \quad\quad \cdot \quad\quad \cdot\\
\mathbb{P}_Y(k) &= \mathbb{P}(Y = k) = \mathbb{P}(TT...TH) = (1-p)^{k-1} p.
\end{aligned}$$

Ainsi, on peut écrire le PMF de Y de la manière suivante

$$\mathbb{P}_Y(y) = \begin{cases} (1-p)^{y-1}p & \text{pour } y = 1, 2, 3, ...\\ 0 & \text{sinon} \end{cases}$$

Considérons une variable aléatoire discrète X avec $Im(X) = R_X$. Notez que par définition, le PMF est une mesure de probabilité, il satisfait donc toutes les propriétés d'une mesure de probabilité. En particulier, nous avons

1. $0 \leqslant \mathbb{P}_X(x) \leqslant 1$ pour tout X, et
2. $\sum_{x \in R_X} \mathbb{P}_X(x) = 1$.

Notez également que pour tout ensemble $A \subset R_X$, on peut trouver la probabilité que $X \in A$ en utilisant le PMF

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

2.2.2 Distributions spéciales issues des variables aléatoires discrètes

Il s'avère que certaines distributions spécifiques sont utilisées à maintes reprises dans la pratique, elles ont donc reçu des noms spéciaux. Il y a une expérience aléatoire derrière chacune de ces distributions. Étant donné que ces expériences aléatoires modélisent de nombreux phénomènes réels, ces distributions spéciales sont fréquemment utilisées dans différentes applications. C'est pourquoi on leur a donné un nom et nous consacrons à leur étude. Nous fournirons des PMF pour toutes ces variables aléatoires spéciales, mais plutôt que d'essayer de mémoriser le PMF, vous devriez comprendre l'expérience aléatoire derrière chacune d'elles. Si vous comprenez les expériences aléatoires, vous pouvez simplement dériver les PMF lorsque vous en avez besoin. Bien qu'il puisse sembler qu'il y ait beaucoup de formules dans cette section, il y a en fait très peu de nouveaux concepts.

2.2.3 Distribution de Bernoulli

Quelle est la variable aléatoire discrète la plus simple (c'est-à-dire le PMF le plus simple) que vous puissiez imaginer ? Notre réponse à cette question est un PMF différent de zéro en un seul point.

Par exemple, si vous définissez

$$\mathbb{P}_X(x) = \begin{cases} 1 & \text{pour } x = 1 \\ 0 & \text{sinon} \end{cases}$$

alors X est une variable aléatoire discrète qui ne peut prendre qu'une seule valeur, c'est-à-dire $X = 1$ avec une probabilité de un. Mais ce n'est pas une distribution très intéressante car elle n'est pas réellement aléatoire. Ensuite, vous pourriez vous demander quelle est la prochaine distribution discrète la plus simple. Et notre réponse à cela est la **distribution de Bernoulli**. Une variable aléatoire de Bernoulli est une variable aléatoire qui ne peut prendre que deux valeurs possibles, généralement 0 et 1. Cette variable aléatoire modélise des expériences aléatoires qui ont deux résultats possibles, parfois appelés "succès" et "échec". Voici quelques exemples :

- Vous passez un examen de réussite ou d'échec. Soit vous réussissez (résultant en $X = 1$) ou échec (résultant en $X = 0$).
- Vous lancez une pièce de monnaie. Le résultat est pile ou face.
- Un enfant est né. Le sexe est soit masculin soit féminin.

On note généralement la probabilité de succès par p et la probabilité d'échec par $q = 1 - p$.

Formellement, la distribution de Bernoulli est définie comme suit :

Définition 2.4. Une variable aléatoire X est dite une variable aléatoire de *Bernoulli* de paramètre p, représentée par $X \sim Bernouilli(p)$ ou $X \sim \mathcal{B}(1, p)$, si son PMF est donné par

$$\mathbb{P}_X(x) = \begin{cases} p & \text{pour } x = 1 \\ 1 - p & \text{pour } x = 0 \\ 0 & \text{sinon} \end{cases}$$

où $0 < p < 1$.

La figure (2.2) montre le PMF d'une variable aléatoire de $Bernoulli(p)$.

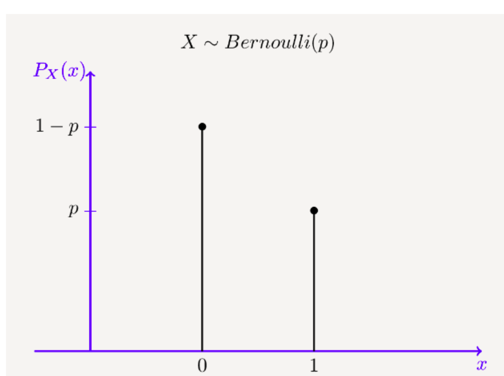


FIGURE 2.2 – PMF d'une variable aléatoire de $Bernoulli(p)$.

Une variable aléatoire de Bernoulli est associée à un certain événement A. Si événement A se produit (par exemple, si vous réussissez le test), alors $X = 1$; autrement $X = 0$. Pour cette raison, la variable aléatoire de Bernoulli est également appelée variable aléatoire **indicatrice**. En particulier, la variable aléatoire indicatrice I_A pour un événement A est définie par

$$I_A = \begin{cases} 1 & \text{si l'événement } A \text{ se produit} \\ 0 & \text{sinon} \end{cases}$$

La variable aléatoire indicatrice d'un événement A a une distribution de Bernoulli de paramètre $p = \mathbb{P}(A)$, on peut donc écrire

$$I_A \sim Bernoulli\big(\mathbb{P}(A)\big).$$

2.2.4 Distribution géométrique

L'expérience aléatoire derrière la distribution géométrique est la suivante. Supposons que j'ai une pièce avec $\mathbb{P}(H) = p$. Je lance la pièce jusqu'à ce que j'observe les premières faces. Nous définissons X comme le nombre total de lancers de pièces dans cette expérience. Alors X est dit avoir une distribution géométrique avec le paramètre p. En d'autres termes, vous pouvez considérer cette expérience comme la répétition d'essais de Bernoulli indépendants jusqu'à l'observation du premier succès. C'est exactement la même distribution que celle que nous avons vue dans l' exemple (2.4). Le co-domaine de X est $R_X = \{1, 2, 3, ...\}$. Dans l'exemple (2.4), nous avons obtenu

$$\mathbb{P}_X(k) = \mathbb{P}(X = k) = (1-p)^{k-1}p, \text{ pour } k = 1, 2, 3, ...$$

On note généralement $q = 1 - p$, on peut donc écrire $\mathbb{P}_X(k) = pq^{k-1}$, pour $k = 1, 2, 3, ...$ Dire qu'une variable aléatoire a une distribution géométrique de paramètre p, nous écrivons $X \sim Géométrique(p)$. Plus formellement, nous avons la définition suivante :

Définition 2.5. Une variable aléatoire X est dite une variable aléatoire géométrique de paramètre p, représentée comme $X \sim Géométrique(p)$, si son PMF est donné par

$$\mathbb{P}_X(k) = \begin{cases} \mathbb{P}(1-p)^{k-1} & \text{pour } k = 1, 2, 3, ... \\ 0 & \text{sinon} \end{cases}$$

où $0 < p < 1$.

La figure (2.3) montre le PMF d'une variable aléatoire $Géométrique(0, 3)$.

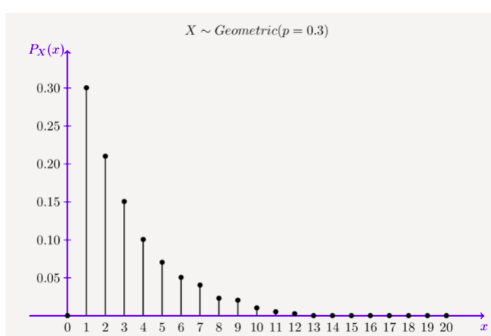


FIGURE 2.3 – PMF d'une variable aléatoire $Géométrique(0, 3)$.

Il convient de noter que certains livres définissent les variables aléatoires géométriques légèrement différemment. Ils définissent la variable aléatoire géométrique X comme le nombre total d'échecs avant d'observer le premier succès. Selon cette définition, l'image de X est $R_X = \{0, 1, 2, ...\}$ et le PMF est donné par

$$\mathbb{P}_X(k) = \begin{cases} \mathbb{P}(1-p)^k & \text{for } k = 0, 1, 2, 3, ... \\ 0 & \text{sinon} \end{cases}$$

Dans ce cours, chaque fois que nous écrivons $X \sim Géométrique(p)$, on veut toujours dire X comme le nombre total d'essais tel que défini dans la définition (2.5). Notez que tant que vous êtes cohérent dans votre analyse, peu importe la définition que vous utilisez. C'est pourquoi nous insistons sur le fait que vous devez comprendre comment dériver les PMF pour ces variables aléatoires plutôt que de les mémoriser.

2.2.5 Distribution binomiale

L'expérience aléatoire derrière la distribution binomiale est la suivante. Supposons que j'ai une pièce avec $\mathbb{P}(H) = p$. Je lance la pièce n fois et je définis X comme le nombre total de faces que j'observe avec pour image de X dans ce cas $R_X = \{0, 1, 2, ..., n\}$. C'est-à-dire, si nous avons une expérience dans laquelle nous effectuons n essais de Bernoulli indépendants et compter le nombre total de succès, nous l'appelons une **expérience binomiale**. Prenons un exemple afin de déterminer la fonction de masse de probabilité d'une distribution binomiale.

Exemple 2.5. Supposons que j'ai une pièce pour laquelle $\mathbb{P}(H) = p$ et $\mathbb{P}(T) = 1 - p$. Je lance la pièce 5 fois .

1. Quelle est la probabilité que le résultat soit THHHH ?
2. Quelle est la probabilité que le résultat soit HTHHH ?
3. Quelle est la probabilité que le résultat soit HHTHH ?
4. Quelle est la probabilité que j'observe exactement quatre faces et une pile ?
5. Quelle est la probabilité que j'observe exactement trois faces et deux piles ?
6. Si je lance la pièce n fois, quelle est la probabilité que j'observe exactement k faces et $n - k$ piles ?

Solution

1. Pour trouver la probabilité de l'événement $A = \{THHHH\}$, on remarque que A est l'intersection de 5 événements indépendants : $A \equiv$ le premier tirage au sort est pile, et les quatre prochains tirages au sort donnent face. Puisque les lancers individuels de pièces sont indépendants, nous obtenons

$$\mathbb{P}(THHHH) = \mathbb{P}(T) \times \mathbb{P}(H) \times \mathbb{P}(H) \times \mathbb{P}(H) \times \mathbb{P}(H) = (1-p)p^4.$$

2. De même,

$$\mathbb{P}(HTHHH) = \mathbb{P}(H) \times \mathbb{P}(T) \times \mathbb{P}(H) \times \mathbb{P}(H) \times \mathbb{P}(H) = (1-p)p^4.$$

3. De même,

$$\mathbb{P}(HHTHH) = \mathbb{P}(H) \times \mathbb{P}(H) \times \mathbb{P}(T) \times \mathbb{P}(H) \times \mathbb{P}(H) = (1-p)p^4.$$

4. Soit B l'événement que j'observe exactement une pile et quatre faces. Puis

$$B = \{THHHH, HTHHH, HHTHH, HHHTH, HHHHT\}.$$

Ainsi

$$\begin{aligned}\mathbb{P}(B) =& \mathbb{P}(THHHH) + \mathbb{P}(HTHHH) + \mathbb{P}(HHTHH) + \mathbb{P}(HHHTH) + \mathbb{P}(HHHHT)\\ =&(1-p)p^4 + (1-p)p^4 + (1-p)p^4 + (1-p)p^4 + (1-p)p^4\\ =&5p^4(1-p).\end{aligned}$$

5. Soit C l'événement que j'observe exactement trois faces et deux piles. Alors,

$$C = \{TTHHH, THTHH, THHHTH, \dots, HHHTT\}.$$

Ainsi

$$\begin{aligned}\mathbb{P}(C) =& \mathbb{P}(TTHHH) + \mathbb{P}(THTHH) + \mathbb{P}(THHTH) + ... + \mathbb{P}(HHHTT)\\ =&(1-p)^2p^3 + (1-p)^2p^3 + (1-p)^2p^3 + ... + (1-p)^2p^3\\ =&|C|p^3(1-p)^2.\end{aligned}$$

Mais $|C|$ vaut combien ? C'est le nombre total de séquences distinctes que vous pouvez créer en utilisant deux piles et trois faces. Ainsi, le nombre total d'éléments dans C est $\binom{5}{3}$, et

$$\mathbb{P}(C) = \binom{5}{3} p^3(1-p)^2.$$

6. Enfin, nous pouvons répéter le même argument lorsque nous lançons la pièce n fois et obtenir

$$\mathbb{P}(k \text{ faces et } n-k \text{ piles}) = \binom{n}{k} p^k(1-p)^{n-k}.$$

Notez qu'ici, au lieu d'écrire $\mathbb{P}(k \text{ faces et } n-k \text{ piles})$, nous pouvons simplement écrire $\mathbb{P}(k \text{ faces})$.

Formule binomiale :
Pour n essais de Bernoulli indépendants où chaque essai a une probabilité de succès p, la probabilité de k succès sont donnés par

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

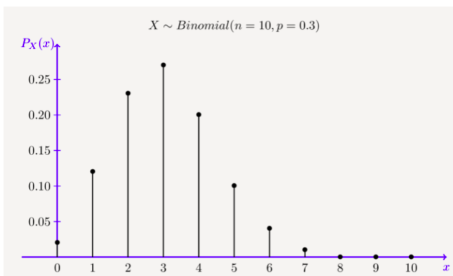
Nous avons la définition suivante :

Définition 2.6. Une variable aléatoire X est dite une variable aléatoire binomiale de paramètres n et p, représentée par $X \sim Binomiale(n,p)$ ou $X \sim \mathcal{B}(n,p)$, si son PMF est donné par

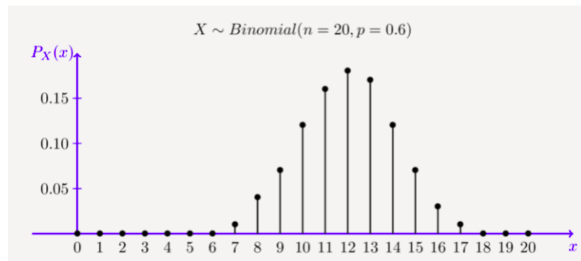
$$\mathbb{P}_X(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{pour } k = 0,1,2,\cdots,n \\ 0 & \text{sinon} \end{cases}$$

où $0 < p < 1$.

Les figures (2.4a) et (2.4b) montrent la loi $Binomiale(n,p)$ pour $n = 10$, $p = 0,3$ et $n = 20$, $p = 0,6$ respectivement.



(a) PMF d'une v.a. $\mathcal{B}(10; 0,3)$.



(b) PMF d'une v.a. $\mathcal{B}(20; 0.6)$.

FIGURE 2.4

Variable aléatoire binomiale comme somme de variables aléatoires de Bernoulli

Voici une façon utile de penser à une variable aléatoire binomiale. Notez qu'une v.a $\mathcal{B}(n,p)$ peut être obtenue par n lancers de pièces indépendants. Si nous considérons chaque pile ou face comme une v.a de $Bernoulli(p)$, la v.a $\mathcal{B}(n,p)$ est une somme de n v.a indépendante de $Bernoulli(p)$. Ceci est énoncé plus précisément dans le lemme suivant.

Lemme 2.1. *Si X_1, X_2, ..., X_n sont des v.a indépendantes de $Bernoulli(p)$, alors la variable aléatoire X définie par $X = X_1 + X_2 + ... + X_n$ a une distribution $Binomiale(n,p)$.*

Pour générer une variable aléatoire $X \sim \mathcal{B}(n,p)$, on peut lancer une pièce n fois et compter le nombre de faces. Compter le nombre de faces revient exactement à trouver $X_1 + X_2 + ... + X_n$, où chaque X_i est égal à 1 si le tirage au sort correspondant donne face et zéro dans le cas contraire. Cette interprétation des variables aléatoires binomiales est parfois très utile. Prenons un exemple.

Exemple 2.6. Soit $X \sim \mathcal{B}(n,p)$ et $Y \sim \mathcal{B}(m,p)$ deux variables aléatoires indépendantes. Définissons une nouvelle variable aléatoire $Z = X + Y$. Trouver le PMF de Z.

Solution

Comme $X \sim \mathcal{B}(n,p)$, on peut penser à X comme le nombre de faces dans n lancers de pièces indépendants, c'est-à-dire que nous pouvons écrire

$$X = X_1 + X_2 + ... + X_n,$$

où les X_i sont des v.a indépendantes de $Bernoulli(p)$. De même, Comme $Y \sim \mathcal{B}(m,p)$, on peut penser à Y comme le nombre de faces dans m lancers de pièces indépendants, c'est-à-dire que nous pouvons écrire

$$Y = Y_1 + Y_2 + ... + Y_m,$$

où les Y_j sont des v.a indépendantes de $Bernoulli(p)$. Ainsi, la variable aléatoire $Z = X + Y$ sera le nombre total de faces dans $n + m$ lancers de pièces indépendants :

$$Z = X + Y = X_1 + X_2 + ... + X_n + Y_1 + Y_2 + ... + Y_m,$$

où les X_i et Y_j sont des v.a indépendantes de $Bernoulli(p)$. Ainsi, d'après le lemme (2.1), Z est une variable aléatoire binomiale avec des paramètres $m + n$ et p, c'est-à- dire, $\mathcal{B}(m + n, p)$. Par conséquent, le PMF de Z est

$$\mathbb{P}_Z(k) = \begin{cases} \binom{m+n}{k} p^k (1-p)^{m+n-k} & \text{pour } k = 0, 1, 2, 3, ..., m+n \\ 0 & \text{sinon} \end{cases}$$

La solution ci-dessus est élégante et simple, mais nous pouvons également vouloir obtenir directement le PMF de Z en utilisant des règles de probabilité. Voici une autre méthode pour résoudre ce problème. Tout d'abord, nous remarquons que $R_Z = \{0, 1, 2, ..., m + n\}$. Pour $k \in R_Z$, nous pouvons écrire

$$\mathbb{P}_Z(k) = \mathbb{P}(Z = k) = \mathbb{P}(X + Y = k).$$

Nous allons trouver $\mathbb{P}(X + Y = k)$ en utilisant le la loi de probabilité totale. En particulier, on peut écrire

$$\begin{aligned}
\mathbb{P}_Z(k) =& \mathbb{P}(X + Y = k) \\
=& \sum_{i=0}^{n} \mathbb{P}(X + Y = k | X = i) \mathbb{P}(X = i) \text{ (loi de probabilité totale)} \\
=& \sum_{i=0}^{n} \mathbb{P}(Y = k - i | X = i) \mathbb{P}(X = i) \\
=& \sum_{i=0}^{n} \mathbb{P}(Y = k - i) \mathbb{P}(X = i) \text{ (puisque } X \text{ et } Y \text{ sont indépendants)} \\
=& \sum_{i=0}^{n} \binom{m}{k-i} p^{k-i} (1-p)^{m-k+i} \binom{n}{i} p^i (1-p)^{n-i} \text{ (puisque } X \text{ et } Y \text{ sont binomiaux)} \\
=& \sum_{i=0}^{n} \binom{m}{k-i} \binom{n}{i} p^k (1-p)^{m+n-k} \\
=& p^k (1-p)^{m+n-k} \sum_{i=0}^{n} \binom{m}{k-i} \binom{n}{i} \\
=& \binom{m+n}{k} p^k (1-p)^{m+n-k} \text{ (par l'exemple 2.8 (partie 3)).}
\end{aligned}$$

Ainsi, nous avons prouvé $Z \sim \mathcal{B}(m + n, p)$ en trouvant directement le PMF de Z.

2.2.6 Distribution binomiale négative (Pascal)

La distribution binomiale négative ou distribution de Pascal est une généralisation de la distribution géométrique. Elle concerne l'expérience aléatoire d'essais indépendants répétés jusqu'à observer m succès. Encore une fois, différents auteurs définissent la distribution de Pascal légèrement différemment, si vous comprenez l'une d'elles, vous pouvez facilement dériver les autres. Voici comment nous définissons la distribution de Pascal. Supposons que j'ai une pièce avec $\mathbb{P}(H) = p$. Je lance la pièce jusqu'à ce que j'observe m faces, où $m \in \mathbb{N}$. Soit X le nombre total de lancers de pièces dans cette expérience. Alors X est dit avoir une distribution de Pascal avec le paramètre m et p. Nous écrivons $X \sim Pascal(m, p)$. Noter que $Pascal(1, p) = Géométrique(p)$. Notez que selon notre définition, l'image de X est donnée par $R_X = \{m, m+1, m+2, m+3, \cdots\}$.

Dérivons le PMF d'une variable aléatoire X suivant la loi de $Pascal(m, p)$. Supposons que je lance la pièce jusqu'à ce que j'observe m faces, et X est défini comme le nombre total de lancers de pièces dans cette expérience. Pour trouver la probabilité de l'événement $A = \{X = k\}$, nous argumentons comme suit. Par définition, l'événement A peut être écrit comme $A = B \cap C$, où

- B est l'événement que nous observons $m-1$ faces (succès) au premier $k-1$ essais, et
- C est l'événement que nous observons une face dans le $k$ème essai.

Noter que B et C sont des événements indépendants parce qu'ils sont liés à différentes épreuves indépendantes (jeu de pièces). Ainsi on peut écrire

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

Maintenant nous avons $\mathbb{P}(C) = p$. Notez également que $\mathbb{P}(B)$ est la probabilité que j'observe $m-1$ faces dans le $k-1$ lancers de pièces. Cette probabilité est donnée par la formule binomiale, en particulier

$$\mathbb{P}(B) = \binom{k-1}{m-1} p^{m-1} (1-p)^{\big((k-1)-(m-1)\big)} = \binom{k-1}{m-1} p^{m-1} (1-p)^{k-m}.$$

Ainsi, on obtient

$$\mathbb{P}(A) = \mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C) = \binom{k-1}{m-1} p^{m} (1-p)^{k-m}.$$

Pour résumer, nous avons la définition suivante pour la variable aléatoire de Pascal.

Définition 2.7. Une variable aléatoire X suit une distribution de Pascal de paramètres m et p, représentée comme $X \sim Pascal(m, p)$, si son PMF est donné par

$$\mathbb{P}_X(k) = \begin{cases} \binom{k-1}{m-1} p^m (1-p)^{k-m} & \text{pour } k = m, m+1, m+2, m+3, \ldots \\ 0 & \text{sinon} \end{cases}$$

où $0 < p < 1$.

La figure (2.5) montre le PMF d'une loi de $Pascal(m, p)$ avec $m = 3$ et $p = 0,5$.

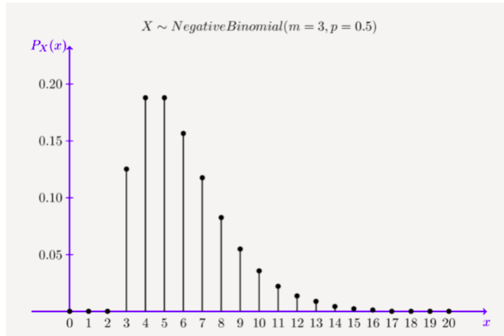


FIGURE 2.5 – PMF d'une distribution de $Pascal(3, 0.5)$(binomiale négative).

2.2.7 Distribution hypergéométrique

Voici l'expérience aléatoire derrière la distribution hypergéométrique. Vous avez un sac qui contient b billes bleues et r billes rouges. Tu choisis $k \leqslant b + r$ billes au hasard (sans remplacement). Soit X le nombre de billes bleues dans votre échantillon. Par cette définition, nous avons $X \leqslant \min(k, b)$. De plus, le nombre de billes rouges dans votre échantillon doit être inférieur ou égal à r, nous concluons donc $X \geqslant \max(0, k - r)$. Par conséquent, l'image de X est donnée par $R_X = \{\max(0, k - r), \max(0, k - r) + 1, \max(0, k - r) + 2, ..., \min(k, b)\}$.

Trouver $\mathbb{P}_X(x)$, notez que le nombre total de façons de choisir k billes de $b + r$ billes $\binom{b+r}{k}$. Le nombre total de façons de choisir x billes bleues et $k - x$ billes rouges est $\binom{b}{x}\binom{r}{k-x}$. Ainsi, nous avons

$$\mathbb{P}_X(x) = \frac{\binom{b}{x}\binom{r}{k-x}}{\binom{b+r}{k}}, \qquad \text{pour } x \in R_X.$$

La définition suivante résume la discussion ci-dessus.

Définition 2.8. Une variable aléatoire X suit une distribution hypergéométrique de paramètres b, r et k, représentée comme $X \sim Hypergéométrique(b, r, k)$, si son image est

$$R_X = \{\max(0, k - r), \max(0, k - r) + 1, \max(0, k - r) + 2, ..., \min(k, b)\},$$

et son PMF est donné par

$$\mathbb{P}_X(x) = \begin{cases} \frac{\binom{b}{x}\binom{r}{k-x}}{\binom{b+r}{k}} & \text{pour } x \in R_X \\ 0 & \text{sinon} \end{cases}$$

Encore une fois, il ne sert à rien de mémoriser le PMF. Tout ce que vous devez savoir, c'est comment résoudre des problèmes qui peuvent être formulés comme une variable aléatoire hypergéométrique.

2.2.8 Distribution de Poisson

La distribution de Poisson est l'une des distributions de probabilité les plus utilisées. Elle est généralement utilisée dans des scénarios où nous comptons les occurrences de certains événements dans un intervalle de temps ou d'espace. En pratique, il s'agit souvent d'une approximation d'une variable aléatoire réelle. Voici un exemple de scénario dans lequel une variable aléatoire de Poisson pourrait être utilisée. Supposons que nous comptions le nombre de clients qui visitent un certain magasin de 13h à 14h. D'après les données des jours précédents, nous savons qu'en moyenne $\lambda = 15$ clients visitent le magasin. Bien sûr, il y aura plus de clients certains jours et moins d'autres. Ici, nous pouvons modéliser la variable aléatoire X montrant le nombre de clients comme une variable aléatoire de Poisson de paramètre $\lambda = 15$. Introduisons d'abord le PMF de Poisson, puis nous parlerons d'autres exemples et interprétations de cette distribution.

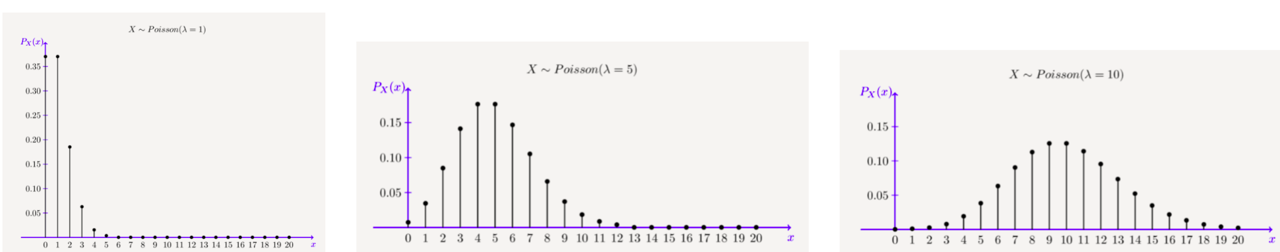
Définition 2.9. Une variable aléatoire X est dite une variable aléatoire de Poisson de paramètre λ, représentée comme $X \sim Poisson(\lambda)$, si son image est $R_X = \{0, 1, 2, 3, ...\}$, et son PMF est donné par

$$\mathbb{P}_X(k) = \begin{cases} \frac{e^{-\lambda}\lambda^k}{k!} & \text{pour } k \in R_X \\ 0 & \text{sinon} \end{cases}$$

Avant d'aller plus loin, vérifions qu'il s'agit d'un PMF valide. Tout d'abord, nous remarquons que $\mathbb{P}_X(k) \geqslant 0$ pour tout k. Ensuite, nous devons vérifier $\sum_{k \in R_X} \mathbb{P}_X(k) = 1$. Pour ce faire, rappelons d'abord la série de Taylor pour e^x, $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$. Maintenant on peut écrire

$$\begin{aligned}\sum_{k \in R_X} \mathbb{P}_X(k) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\ &= e^{-\lambda}e^{\lambda} (\text{par série de Taylor pour } e^{\lambda}) \\ &= 1.\end{aligned}$$

Les figures (2.6a), (2.6b) et (2.6c) montrent les PMF de $Poisson(\lambda)$ pour $\lambda = 1$, $\lambda = 5$, et $\lambda = 10$ respectivement.



(a) PMF d'une v.a. de Poisson pour $\lambda = 1$

(b) PMF d'une v.a. de Poisson pour $\lambda = 5$

(c) PMF d'une v.a. de Poisson pour $\lambda = 10$

FIGURE 2.6

Exemple 2.7. Le nombre d'e-mails que je reçois en un jour de semaine peut être modélisé par une distribution de Poisson avec une moyenne de $0,2$ e-mails par minute.

1. Quelle est la probabilité que je ne reçoive aucun e-mail dans un intervalle de longueur de 5 minutes ?

2. Quelle est la probabilité que j'obtienne plus de 3 e-mails dans un intervalle de longueur de 10 minutes ?

Solution
Soit X le nombre d'e-mails que je reçois dans les 5 minutes. Ensuite, par l'hypothèse X est une variable aléatoire de Poisson de paramètre $\lambda = 5(0,2) = 1$,

$$\mathbb{P}(X=0) = \mathbb{P}_X(0) = \frac{e^{-\lambda}\lambda^0}{0!} = \frac{e^{-1}\cdot 1}{1} = \frac{1}{e} \approx 0,3679$$

Soit Y le nombre d'e-mails que je reçois dans les 10 minutes. Puis par l'hypothèse Y est une variable aléatoire de Poisson de paramètre $\lambda = 10(0,2) = 2$,

$$\begin{aligned}
\mathbb{P}(Y>3) =& 1 - \mathbb{P}(Y \leqslant 3) \\
=& 1 - \big(\mathbb{P}_Y(0) + \mathbb{P}_Y(1) + \mathbb{P}_Y(2) + \mathbb{P}_Y(3)\big) \\
=& 1 - e^{-\lambda} - \frac{e^{-\lambda}\lambda}{1!} - \frac{e^{-\lambda}\lambda^2}{2!} - \frac{e^{-\lambda}\lambda^3}{3!} \\
=& 1 - e^{-2} - \frac{2e^{-2}}{1} - \frac{4e^{-2}}{2} - \frac{8e^{-2}}{6} \\
=& 1 - e^{-2}\left(1+2+2+\frac{8}{6}\right) \\
=& 1 - \frac{19}{3e^2} \approx 0,1429
\end{aligned}$$

Poisson comme approximation de la loi binomiale

La distribution de Poisson peut être considérée comme la limite de la distribution binomiale. Supposer $X \sim \mathcal{B}(n,p)$ où n est très grand et p est très petit. En particulier, supposons que $\lambda = np$ est une constante positive. Nous montrons que le PMF de X peut être approché par le PMF d'une v.a de $Poisson(\lambda)$. L'importance de ceci est que le PMF de Poisson est beaucoup plus facile à calculer que celui de la loi binomiale. Énonçons cela sous forme de théorème.

Théorème 2.1. *Soit $X \sim \mathcal{B}(n, p = \frac{\lambda}{n})$, où $\lambda > 0$ est fixé. Alors pour tout $k \in \{0,1,2,...\}$, on a*

$$\lim_{n\to\infty} \mathbb{P}_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

Démonstration. Nous avons

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}_X(k) =& \lim_{n\to\infty} \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\
=& \lambda^k \lim_{n\to\infty} \frac{n!}{k!(n-k)!}\left(\frac{1}{n^k}\right)\left(1-\frac{\lambda}{n}\right)^{n-k} \\
=& \frac{\lambda^k}{k!}.\lim_{n\to\infty}\left(\left[\frac{n(n-1)(n-2)...(n-k+1)}{n^k}\right]\left[\left(1-\frac{\lambda}{n}\right)^n\right]\left[\left(1-\frac{\lambda}{n}\right)^{-k}\right]\right).
\end{aligned}$$

A noter que pour un k fixé, on a

$$\lim_{n\to\infty} \frac{n(n-1)(n-2)...(n-k+1)}{n^k} = 1$$
$$\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^{-k} = 1$$
$$\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda}.$$

Ainsi, nous concluons

$$\lim_{n\to\infty} \mathbb{P}_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

□

2.2.9 Fonction de distribution cumulative ou fonction de répartition

Le PMF est une façon de décrire la distribution d'une variable aléatoire discrète. Comme nous le verrons un peu plus loin, le PMF ne peut pas être défini pour des variables aléatoires continues. La fonction de distribution cumulative (CDF) (ou fonction de répartition) d'une variable aléatoire est une autre méthode pour décrire la distribution des variables aléatoires. L'avantage du CDF est qu'il peut être défini pour tout type de variable aléatoire (discrète, continue et mixte).

Définition 2.10. La fonction de distribution cumulative (CDF) (ou fonction de répartition) de la variable aléatoire X est définie comme

$$F_X(x) = \mathbb{P}(X \leqslant x), \text{ pour tout } x \in \mathbb{R}.$$

Notez que l'indice X indique qu'il s'agit du CDF de la variable aléatoire X. Notez également que le CDF est défini pour tout $x \in \mathbb{R}$. Prenons un exemple.

Exemple 2.8. Je lance une pièce deux fois. Soit X le nombre de faces observées. Trouver le CDF de X.

Solution
Notez qu'ici $X \sim \mathcal{B}(2, \frac{1}{2})$. L'image de X est $R_X = \{0, 1, 2\}$ et son PMF est donné par

$$\mathbb{P}_X(0) = \mathbb{P}(X=0) = \frac{1}{4}, \qquad \mathbb{P}_X(1) = \mathbb{P}(X=1) = \frac{1}{2}, \qquad \mathbb{P}_X(2) = \mathbb{P}(X=2) = \frac{1}{4}.$$

Pour trouver le CDF, nous procédons comme suit. Tout d'abord, notez que si $x < 0$, alors

$$F_X(x) = \mathbb{P}(X \leqslant x) = 0, \text{ pour } x < 0.$$

Ensuite, si $x \geqslant 2$,

$$F_X(x) = \mathbb{P}(X \leqslant x) = 1, \text{ pour } x \geqslant 2.$$

Ensuite, si $0 \leqslant x < 1$,

$$F_X(x) = \mathbb{P}(X \leqslant x) = \mathbb{P}(X=0) = \frac{1}{4}, \text{ pour } 0 \leqslant x < 1.$$

Enfin, si $1 \leqslant x < 2$,

$$F_X(x) = \mathbb{P}(X \leqslant x) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}, \text{ pour } 1 \leqslant x < 2.$$

Ainsi, pour résumer, nous avons

$$F_X(x) = \begin{cases} 0 & \text{pour } x < 0 \\ \frac{1}{4} & \text{pour } 0 \leqslant x < 1 \\ \frac{3}{4} & \text{pour } 1 \leqslant x < 2 \\ 1 & \text{pour } x \geqslant 2 \end{cases}$$

Propriété. *Pour tout $a \leqslant b$, on a*

$$\mathbb{P}(a < X \leqslant b) = F_X(b) - F_X(a) \tag{2.1}$$

Démonstration. Notez que pour $a \leqslant b$ on a

$$\mathbb{P}(X \leqslant b) = \mathbb{P}(X \leqslant a) + \mathbb{P}(a < X \leqslant b).$$

Ainsi,

$$F_X(b) = F_X(a) + \mathbb{P}(a < X \leqslant b).$$

□

Encore une fois, faites attention à l'utilisation de "$<$" et "$\leqslant$" car ils pourraient faire une différence dans le cas de variables aléatoires discrètes. Nous verrons plus loin que l'équation (2.1) est vraie pour tous les types de variables aléatoires (discrètes, continues et mixtes). Notez que le CDF nous donne $\mathbb{P}(X \leqslant x)$. Trouver $\mathbb{P}(X < x)$, pour une variable aléatoire discrète, on peut simplement écrire

$$\mathbb{P}(X < x) = \mathbb{P}(X \leqslant x) - \mathbb{P}(X = x) = F_X(x) - \mathbb{P}_X(x).$$

Exemple 2.9. Soit X une variable aléatoire discrète avec image $R_X = \{1, 2, 3, ...\}$. Supposons que le PMF de X est donné par

$$\mathbb{P}_X(k) = \frac{1}{2^k} \text{ pour } k = 1, 2, 3, ...$$

1. Trouver et tracer le CDF de X, $F_X(x)$.
2. Trouver $\mathbb{P}(2 < X \leqslant 5)$.
3. Trouver $\mathbb{P}(X > 4)$.

Solution

Tout d'abord, notez qu'il s'agit d'un PMF valide. En particulier,

$$\sum_{k=1}^{\infty} \mathbb{P}_X(k) = \sum_{k=1}^{\infty} \frac{1}{2^k} = 1 \text{ (somme géométrique)}$$

1. Pour trouver le CDF, notez que

$$\begin{aligned}
\text{Pour } x < 1, \quad & F_X(x) = 0. \\
\text{Pour } 1 \leqslant x < 2, \quad & F_X(x) = \mathbb{P}_X(1) = \frac{1}{2}. \\
\text{Pour } 2 \leqslant x < 3, \quad & F_X(x) = \mathbb{P}_X(1) + \mathbb{P}_X(2) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}.
\end{aligned}$$

En général nous avons, Pour $0 < k \leqslant x < k+1$,

$$F_X(x) = \mathbb{P}_X(1) + \mathbb{P}_X(2) + \ldots + \mathbb{P}_X(k) = \frac{1}{2} + \frac{1}{4} + \ldots + \frac{1}{2^k} = \frac{2^k - 1}{2^k}.$$

La figure (2.7) montre le CDF de X.

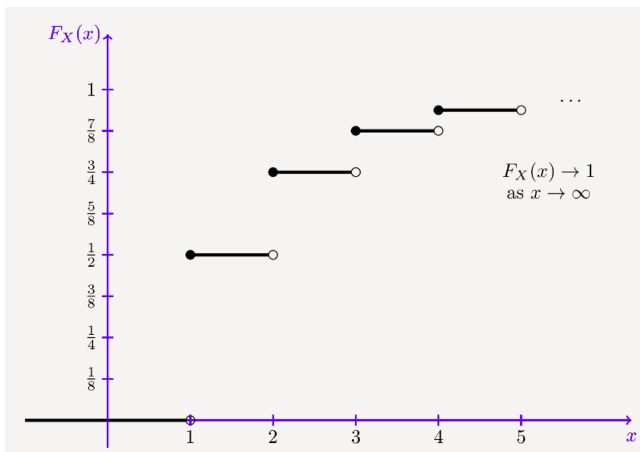


FIGURE 2.7 – CDF d'une variable aléatoire discrète donnée dans l'exemple (2.9).

2. Trouver $\mathbb{P}(2 < X \leqslant 5)$, nous pouvons écrire

$$\mathbb{P}(2 < X \leqslant 5) = F_X(5) - F_X(2) = \frac{31}{32} - \frac{3}{4} = \frac{7}{32}.$$

Ou de façon équivalente, on peut écrire

$$\mathbb{P}(2 < X \leqslant 5) = \mathbb{P}_X(3) + \mathbb{P}_X(4) + \mathbb{P}_X(5) = \frac{1}{8} + \frac{1}{16} + \frac{1}{32} = \frac{7}{32},$$

qui donne la même réponse.

3. Trouver $\mathbb{P}(X > 4)$, nous pouvons écrire

$$\mathbb{P}(X > 4) = 1 - \mathbb{P}(X \leqslant 4) = 1 - F_X(4) = 1 - \frac{15}{16} = \frac{1}{16}.$$

2.2.10 Espérance mathématique

Si vous avez une collection de nombres a_1, a_2, ..., a_N, leur moyenne est un nombre unique qui décrit l'ensemble de la collection. Considérons maintenant une variable aléatoire X. Nous voudrions définir sa moyenne, ou comme on l'appelle en probabilité, son espérance mathématique ou sa valeur espérée ou moyenne. L' espérance mathématique est définie comme la moyenne pondérée des valeurs de l'image.

Définition 2.11. Soit X une variable aléatoire discrète avec une image $R_X = \{x_1, x_2, x_3, ...\}$ (finie ou dénombrable infinie). L' espérance mathématique de X, notée par $E(X)$ est définie comme

$$E(X) = \sum_{x_k \in R_X} x_k \mathbb{P}(X = x_k) = \sum_{x_k \in R_X} x_k \mathbb{P}_X(x_k).$$

Pour comprendre le concept derrière $E(X)$, considérons une variable aléatoire discrète avec une image $R_X = \{x_1, x_2, x_3, ...\}$. Cette variable aléatoire est le résultat d'une expérience aléatoire. Supposons que nous répétions cette expérience un très grand nombre de fois soit N fois, et que les essais sont indépendants. Soit N_1 le nombre de fois que nous observons x_1, N_2 le nombre de fois que nous observons x_2, ..., N_k le nombre de fois que nous observons x_k, etc. Comme $\mathbb{P}(X = x_k) = \mathbb{P}_X(x_k)$, on s'attend à ce que

$$\mathbb{P}_X(x_1) \approx \frac{N_1}{N}, \quad \mathbb{P}_X(x_2) \approx \frac{N_2}{N}, \quad \cdots, \quad \mathbb{P}_X(x_k) \approx \frac{N_k}{N}, \quad \cdots$$

En d'autres termes, nous avons $N_k \approx N\mathbb{P}_X(x_k)$. Maintenant, si nous prenons la moyenne des valeurs observées de X, on obtient

$$\begin{aligned}
\text{Moyenne } =& \frac{N_1 x_1 + N_2 x_2 + N_3 x_3 + ...}{N} \\
\approx& \frac{x_1 N\mathbb{P}_X(x_1) + x_2 N\mathbb{P}_X(x_2) + x_3 N\mathbb{P}_X(x_3) + ...}{N} \\
=& x_1 \mathbb{P}_X(x_1) + x_2 \mathbb{P}_X(x_2) + x_3 \mathbb{P}_X(x_3) + ... \\
=& E(X).
\end{aligned}$$

Ainsi, l'intuition derrière $E(X)$ est que si vous répétez l'expérience aléatoire indépendamment N fois et prendre la moyenne des données observées, la moyenne se rapproche de plus en plus de $E(X)$ comme N devient de plus en plus grand. On note parfois $E(X)$ par μ_X.

Différentes notations pour l'espérance mathématique de X : $E(X) = \mu_X = EX = E[X]$.

Calculons les valeurs attendues de certaines distributions bien connues.

Exemple 2.10. Soit $\mathbf{X} \sim \mathbf{Bernouilli(p)}$. Trouver $E(X)$.

Solution
Pour la distribution de Bernoulli, l'image de X est $R_X = \{0, 1\}$, et $\mathbb{P}_X(1) = p$ et $\mathbb{P}_X(0) = 1 - p$. Ainsi,

$$E(X) = 0 \cdot \mathbb{P}_X(0) + 1 \cdot \mathbb{P}_X(1) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

Pour une variable aléatoire de Bernoulli, trouver l'espérance $E(X)$ était facile. Cependant, pour certaines variables aléatoires, pour trouver l'espérance mathématique, vous aurez peut-être besoin d'un peu d'algèbre. Prenons un autre exemple.

Exemple 2.11. Soit $\mathbf{X} \sim \mathbf{Géométrique(p)}$. Trouver $E(X)$.

Solution
Pour la distribution géométrique, l'image est $R_X = \{1, 2, 3, ...\}$ et le PMF est donné par

$$\mathbb{P}_X(k) = q^{k-1}p, \qquad \text{pour } k = 1, 2, ...$$

où, $0 < p < 1$ et $q = 1 - p$. Ainsi, nous pouvons écrire

$$E(X) = \sum_{x_k \in R_X} x_k \mathbb{P}_X(x_k) = \sum_{k=1}^{\infty} kq^{k-1}p = p\sum_{k=1}^{\infty} kq^{k-1}.$$

Maintenant, nous connaissons déjà la formule de la somme géométrique

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}, \qquad \text{pour } |x| < 1.$$

Mais il faut trouver une somme $\sum_{k=1}^{\infty} kq^{k-1}$. Heureusement, nous pouvons convertir la somme géométrique dans la forme que nous voulons en prenant la dérivée par rapport à X, c'est à dire,

$$\frac{d}{dx}\sum_{k=0}^{\infty} x^k = \frac{d}{dx}\frac{1}{1-x}, \qquad \text{pour } |x| < 1.$$

Ainsi, nous avons

$$\sum_{k=0}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}, \qquad \text{pour } |x| < 1.$$

Pour finir de trouver l'espérance, on peut écrire

$$E(X) = p\sum_{k=1}^{\infty} kq^{k-1} = p\frac{1}{(1-q)^2} = p\frac{1}{p^2} = \frac{1}{p}.$$

Donc pour $X \sim Géométrique(p)$, $E(X) = \frac{1}{p}$. Notez que cela a un sens intuitivement. L'expérience aléatoire derrière la distribution géométrique consistait à lancer une pièce de monnaie jusqu'à ce que nous observions les premières faces, où $\mathbb{P}(H) = p$. Ici, nous avons découvert qu'en moyenne, vous devez lancer la pièce $\frac{1}{p}$ fois dans cette expérience. En particulier, si p est petit (les faces sont peu probables), alors $\frac{1}{p}$ est grand, vous devez donc lancer la pièce un grand nombre de fois avant d'observer une face. A l'inverse, pour les grands p quelques lancers de pièces suffisent généralement.

Exemple 2.12. Soit $\mathbf{X} \sim \mathbf{Poisson}(\lambda)$. Trouver $E(X)$.

Solution
Avant de faire le calcul, nous vous suggérons d'essayer de deviner quelle serait l'espérance mathématique. Il peut être judicieux de réfléchir aux exemples où la distribution de Poisson est utilisée. Pour la distribution de Poisson, l'image est $R_X = \{0, 1, 2, \cdots\}$ et le PMF est donné par

$$\mathbb{P}_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad \text{for } k = 0, 1, 2, ...$$

Ainsi, nous pouvons écrire

$$E(X) = \sum_{x_k \in R_X} x_k \mathbb{P}_X(x_k) = \sum_{k=0}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \quad e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^{(j+1)}}{j!} (\text{ en laissant } j = k-1)$$
$$= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$$
$$= \lambda e^{-\lambda} e^{\lambda} (\text{ série de Taylor pour } e^{\lambda})$$
$$= \lambda.$$

L' espérance mathématique est donc λ. Rappelez-vous, lorsque nous avons parlé pour la première fois de la distribution de Poisson, nous avons introduit son paramètre λ comme le nombre moyen d'événements. Il n'est donc pas surprenant que l'espérance mathématique soit $E(X) = \lambda$.

Avant d'examiner d'autres exemples, nous aimerions parler d'une propriété importante de l'espérance, qui est la linéarité. Notez que si X est une variable aléatoire, toute fonction de X est également une variable aléatoire, nous pouvons donc parler de son espérance mathématique. Par exemple, si $Y = aX + b$, nous pouvons en parler de $E[Y] = E[aX + b]$. Ou si vous définissez $Y = X_1 + X_2 + \cdots + X_n$, où les X_i sont des variables aléatoires, nous pouvons parler de $E[Y] = E[X_1 + X_2 + \cdots + X_n]$. Le théorème suivant indique que l'espérance est linéaire, ce qui facilite le calcul de l'espérance mathématique des fonctions linéaires de variables aléatoires.

Théorème 2.2 (Linéarité de l'espérance)**.**

- $E[aX + b] = aE(X) + b$, *pour tout* $a, b \in \mathbb{R}$;
- $E[X_1 + X_2 + \cdots + X_n] = E(X_1) + E(X_2) + \cdots + E(X_n)$, *pour tout ensemble de variables aléatoires* X_1, X_2, $\cdots$, X_n.

Nous prouverons ce théorème plus loin, mais ici nous voudrions souligner son importance avec un exemple.

Exemple 2.13. Soit $\mathbf{X} \sim \mathcal{B}(\mathbf{n}, \mathbf{p})$. Trouver $E(X)$.

Solution
Nous proposons deux façons de résoudre ce problème. Une façon est comme avant : nous calculons $E(X) = \sum_{x_k \in R_X} x_k \mathbb{P}_X(x_k)$ ce qui sera un peu fastidieux. Un moyen beaucoup plus rapide serait d'utiliser la linéarité des espérances. En particulier, rappelez-vous que si $X_1, X_2, ..., X_n$ sont des v.a indépendantes de $Bernoulli(p)$, alors la variable aléatoire $X = X_1 + X_2 + ... + X_n$ a une distribution $\mathcal{B}(n, p)$. Ainsi, nous pouvons écrire

$$E(X) = E[X_1 + X_2 + \cdots + X_n] = E(X_1) + E(X_2) + \cdots + E(X_n) \qquad \text{par linéarité de l'espérance}$$
$$= p + p + \cdots + p = np.$$

Nous fournirons le calcul direct de $E(X) = \sum_{x_k \in R_X} x_k \mathbb{P}_X(x_k)$ dans la section Problèmes résolus et comme vous le verrez, il faut beaucoup plus d'algèbre que ci-dessus. L'essentiel est que la linéarité de l'espérance peut parfois rendre nos calculs beaucoup plus faciles. Prenons un autre exemple.

Exemple 2.14. Soit $\mathbf{X} \sim \mathbf{Pascal(m, p)}$. Trouver $E(X)$. (Astuce : Essayez d'écrire $X = X_1 + X_2 + \cdots + X_m$, de sorte que vous savez déjà $E(X_i.)$

Solution
Nous affirmons que si les X_i sont indépendants et $X_i \sim Géométrique(p)$, pour $i = 1, 2, \cdots, m$, alors la variable aléatoire $X = X_1 + X_2 + \cdots + X_m$ suit la loi de $Pascal(m, p)$.. Maintenant, puisque nous savons déjà $E(X_i = \frac{1}{p}$, nous concluons

$$\begin{aligned} E(X) =& E[X_1 + X_2 + \cdots + X_m] = E(X_1) + E(X_2) + \cdots + E(X_m) \qquad \text{par linéarité de l'espérance} \\ =& \frac{1}{p} + \frac{1}{p} + \cdots + \frac{1}{p} = \frac{m}{p}. \end{aligned}$$

Encore une fois, vous pouvez essayer de trouver $E(X)$ directement et comme vous le verrez, vous avez besoin de beaucoup plus d'algèbre par rapport à l'utilisation de la linéarité de l'espérance.

Fonctions des variables aléatoires

Si X est une variable aléatoire et $Y = g(X)$, alors Y est elle-même une variable aléatoire. Ainsi, nous pouvons parler de son PMF, de son CDF et de son espérance mathématique. Tout d'abord, notez que l'image de Y peut être écrite comme

$$R_Y = \{g(x) | x \in R_X\}.$$

Si nous connaissons déjà le PMF de X, pour trouver le PMF de $Y = g(X)$, nous pouvons écrire

$$\begin{aligned} P_Y(y) &= \mathbb{P}(Y = y) \\ &= \mathbb{P}(g(X) = y) \\ &= \sum_{x : g(x) = y} \mathbb{P}_X(x) \end{aligned}$$

Exemple 2.15. Soit X une variable aléatoire discrète avec $\mathbb{P}_X(k) = \frac{1}{5}$ pour $k = -1, 0, 1, 2, 3$. Soit $Y = 2|X|$. Trouver l'image et le PMF de Y.

Solution
Tout d'abord, notez que l'image de Yest

$$R_Y = \{2|x| \text{ où } x \in R_X\} = \{0, 2, 4, 6\}.$$

Pour trouver $\mathbb{P}_Y(y)$, nous devons trouver $\mathbb{P}(Y = y)$ pour $y = 0, 2, 4, 6$. Nous avons

$$\begin{aligned} \mathbb{P}_Y(0) =& \mathbb{P}(Y = 0) = \mathbb{P}(2|X| = 0) = \mathbb{P}(X = 0) = \frac{1}{5}; \\ \mathbb{P}_Y(2) =& \mathbb{P}(Y = 2) = \mathbb{P}(2|X| = 2) = \mathbb{P}\big((X = -1) \text{ ou } (X = 1)\big) = \mathbb{P}_X(-1) + \mathbb{P}_X(1) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}; \\ \mathbb{P}_Y(4) =& \mathbb{P}(Y = 4) = \mathbb{P}(2|X| = 4) = \mathbb{P}(X = 2) + \mathbb{P}(X = -2) = \frac{1}{5}; \\ \mathbb{P}_Y(6) =& \mathbb{P}(Y = 6) = \mathbb{P}(2|X| = 6) = \mathbb{P}(X = 3) + \mathbb{P}(X = -3) = \frac{1}{5}. \end{aligned}$$

Donc, pour résumer,

$$\mathbb{P}_Y(k) = \begin{cases} \frac{1}{5} & \text{pour } k = 0,4,6 \\ \frac{2}{5} & \text{pour } k = 2 \\ 0 & \text{sinon} \end{cases}$$

Espérance mathématique d'une fonction d'une variable aléatoire (LOTUS)

Soit X une variable aléatoire discrète avec PMF $\mathbb{P}_X(x)$, et soit $Y = g(X)$. Supposons que nous cherchions à trouver $E[Y]$. Une façon de trouver $E[Y]$ est de trouver d'abord le PMF de Y puis utilisez la formule d'espérance $E[Y] = E[g(X)] = \sum_{y \in R_Y}$ et $\mathbb{P}_Y(y)$. Mais il existe un autre moyen qui est généralement plus facile. C'est ce qu'on appelle la loi du statisticien inconscient (LOTUS).

Loi du statisticien inconscient (LOTUS) pour variables aléatoires discrètes :

$$\mathbf{E}[\mathbf{g}(\mathbf{X})] = \sum_{\mathbf{x_k} \in \mathbf{R_X}} \mathbf{g}(\mathbf{x_k}) \mathbb{P}_{\mathbf{X}}(\mathbf{x_k}) \tag{2.2}$$

Vous pouvez prouver cel en écrivant $E[Y] = E[g(X)] = \sum_{y \in R_Y}$ et $\mathbb{P}_Y(y)$ en terme de $\mathbb{P}_X(x)$. En pratique, il est généralement plus facile d'utiliser LOTUS que la définition directe lorsque nous avons besoin de $E[g(X)]$.

Exemple 2.16. Soit X une variable aléatoire discrète avec une image $R_X = \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi\}$, tel que $\mathbb{P}_X(0) = \mathbb{P}_X(\frac{\pi}{4}) = \mathbb{P}_X(\frac{\pi}{2}) = \mathbb{P}_X(\frac{3\pi}{4}) = \mathbb{P}_X(\pi) = \frac{1}{5}$. Trouver $E[\sin(X)]$.

Solution
En utilisant LOTUS, nous avons

$$\begin{aligned} E[g(X)] = \sum_{x_k \in R_X} g(x_k)\mathbb{P}_X(x_k) &= \sin(0) \cdot \frac{1}{5} + \sin(\frac{\pi}{4}) \cdot \frac{1}{5} + \sin(\frac{\pi}{2}) \cdot \frac{1}{5} + \sin(\frac{3\pi}{4}) \cdot \frac{1}{5} + \sin(\pi) \cdot \frac{1}{5} \\ &= 0 \cdot \frac{1}{5} + \frac{\sqrt{2}}{2} \cdot \frac{1}{5} + 1 \cdot \frac{1}{5} + \frac{\sqrt{2}}{2} \cdot \frac{1}{5} + 0 \cdot \frac{1}{5} = \frac{\sqrt{2}+1}{5}. \end{aligned}$$

Exemple 2.17. Prouver $E[aX + b] = aE(X) + b$ (linéarité de l'espérance).

Solution
Ici $g(X) = aX + b$, donc en utilisant LOTUS nous avons

$$\begin{aligned} E[aX + b] &= \sum_{x_k \in R_X} (ax_k + b)\mathbb{P}_X(x_k) \\ &= \sum_{x_k \in R_X} ax_k\mathbb{P}_X(x_k) + \sum_{x_k \in R_X} b\mathbb{P}_X(x_k) \\ &= a \sum_{x_k \in R_X} x_k\mathbb{P}_X(x_k) + b \sum_{x_k \in R_X} \mathbb{P}_X(x_k) \\ &= aE(X) + b. \end{aligned}$$

2.2.11 Variance

Considérons deux variables aléatoires X et Y avec les PMF suivants.

$$\mathbb{P}_X(x) = \begin{cases} 0,5 & \text{pour } x = -100 \\ 0,5 & \text{pour } x = 100 \\ 0 & \text{sinon} \end{cases} \tag{2.3}$$

$$\mathbb{P}_Y(y) = \begin{cases} 1 & \text{pour } y = 0 \\ 0 & \text{sinon} \end{cases} \tag{2.4}$$

Noter que $E(X) = E(Y) = 0$. Bien que les deux variables aléatoires aient la même valeur moyenne, leur distribution est complètement différente. Y est toujours égal à sa moyenne de 0, pendant que X est soit 100 ou alors -100, assez éloigné de sa valeur moyenne. La variance est une mesure de l'étalement de la distribution d'une variable aléatoire. Ici, la variance de Y est assez faible car sa distribution est concentrée sur une seule valeur, tandis que la variance de X sera plus importante puisque sa distribution est plus étalée.

Définition 2.12. La variance d'une variable aléatoire X, avec moyenne $E(X) = \mu_X$, est définie comme

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

Par définition, la variance de X est la valeur moyenne de $(X - \mu_X)^2$. Puisque $(X - \mu_X)^2 \geqslant 0$, la variance est toujours supérieure ou égale à zéro. Une grande valeur de la variance signifie que $(X - \mu_X)^2$ est souvent grand, donc X prend souvent des valeurs éloignées de sa moyenne. Cela signifie que la distribution est très étalée. En revanche, une faible variance signifie que la distribution est concentrée autour de sa moyenne.

Notez que si nous n'avons pas égalé la différence entre X et sa moyenne, le résultat serait 0. C'est

$$E[X - \mu_X] = E(X) - E[\mu_X] = \mu_X - \mu_X = 0.$$

X est parfois en dessous de sa moyenne et parfois au-dessus de sa moyenne. Ainsi, $X - \mu_X$ est parfois négatif et parfois positif, mais en moyenne il est nul.

Pour calculer $Var(X) = E\big[(X - \mu_X)^2\big]$, notez que nous devons trouver l'espérance mathématique de $g(X) = (X - \mu_X)^2$, nous pouvons donc utiliser LOTUS. En particulier, on peut écrire

$$\mathrm{Var}(X) = E\big[(X - \mu_X)^2\big] = \sum_{x_k \in R_X} (x_k - \mu_X)^2 \mathbb{P}_X(x_k).$$

Par exemple, pour X et Y défini dans les équations (2.3) et (2.4), nous avons

$$\begin{aligned} \mathrm{Var}(X) = (-100 - 0)^2(0,5) + (100 - 0)^2(0,5) = 10000 \\ \mathrm{Var}(Y) = (0 - 0)^2(1) = 0. \end{aligned}$$

Comme on s'y attend, X a une très grande variance tandis que $Var(Y) = 0$.

A noter que $Var(X)$ a une autre unité que X. Par exemple, si X se mesure en *mêtres* alors $Var(X)$ est en $mêtres^2$. Pour résoudre ce problème, nous définissons une autre mesure, appelée **écart-type**, généralement représentée par σ_X, qui est simplement la racine carrée de la variance.

Définition 2.13. L'**écart-type** d'une variable aléatoire X est défini comme

$$\text{SD}(\mathbf{X}) = \sigma_X = \sqrt{\text{Var}(\mathbf{X})}.$$

L'écart type de X a la même unité que X. Pour X et Y défini dans les équations (2.3) et (2.4), nous avons

$$\begin{aligned}\sigma_X &= \sqrt{10000} = 100\\ \sigma_Y &= \sqrt{0} = 0.\end{aligned}$$

Voici une formule utile pour calculer la variance.

Formule de calcul de la variance :

$$\text{Var}(\mathbf{X}) = \mathbf{E}\left[\mathbf{X^2}\right] - \left[\mathbf{E}(\mathbf{X})\right]^{\mathbf{2}} \tag{2.5}$$

Pour le prouver notez que

$$\begin{aligned}\text{Var}(X) &= E\left[(X-\mu_X)^2\right]\\ &= E\left[X^2 - 2\mu_X X + \mu_X^2\right]\\ &= E\left[X^2\right] - 2E\left[\mu_X X\right] + E\left[\mu_X^2\ big\right] \qquad \text{par linéarité de l'espérance.}\end{aligned}$$

Notez que pour une variable aléatoire donnée X, μ_X est juste un nombre réel constant. Ainsi, $E\left[\mu_X X\right] = \mu_X E[X] = \mu_X^2$, et $E[\mu_X^2] = \mu_X^2$, nous avons donc

$$\begin{aligned}\text{Var}(X) &= E\left[X^2\right] - 2\mu_X^2 + \mu_X^2\\ &= E\left[X^2\right] - \mu_X^2.\end{aligned}$$

Exemple 2.18. Je lance un dé équilibré et soit X le nombre résultant. Trouver $E(X)$, où $Var(X)$, et σ_X.

Solution
Nous avons $R_X = \{1, 2, 3, 4, 5, 6\}$ et $\mathbb{P}_X(k) = \frac{1}{6}$ pour $k = 1, 2, ..., 6$. Ainsi, nous avons

$$\begin{aligned}E(X) &= 1\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + 3\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + 5\cdot\frac{1}{6} + 6\cdot\frac{1}{6} = \frac{7}{2};\\ E(X)^2 &= 1\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + 9\cdot\frac{1}{6} + 16\cdot\frac{1}{6} + 25\cdot\frac{1}{6} + 36\cdot\frac{1}{6} = \frac{91}{6}.\end{aligned}$$

Ainsi

$$\begin{aligned}\mathrm{Var}(X) =& E\left[X^2\right] - \left(E(X)\right)^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} \approx 2,92,\\ \sigma_X =& \sqrt{\mathrm{Var}(X)} \approx \sqrt{2,92} \approx 1,71\end{aligned}$$

Notez que la variance n'est pas un opérateur linéaire. En particulier, on a le théorème suivant.

Pour une variable aléatoire X et les nombres réels a et b,

$$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X) \tag{2.6}$$

Si $Y = aX + b$, $E(Y) = aE(X) + b$. Ainsi,

$$\begin{aligned}\mathrm{Var}(Y) &= E[(Y-EY)^2]\\ &= E[(aX+b-aE(X)-b)^2]\\ &= E[a^2(X-\mu_X)^2]\\ &= a^2E[(X-\mu_X)^2]\\ &= a^2\mathrm{Var}(X)\end{aligned}$$

De l'équation (2.6), nous concluons que, pour l'écart type, $\mathrm{SD}(aX+b) = |a|\mathrm{SD}(X)$. Nous avons mentionné que la variance n'est PAS une opération linéaire. Mais il y a un cas très important, dans lequel la variance se comporte comme une opération linéaire et c'est quand on regarde la somme de variables aléatoires indépendantes.

Théorème 2.3. *Si $X_1, X_2, \cdots, X_n$ sont des variables aléatoires indépendantes et $X = X_1 + X_2 + \cdots + X_n$, alors*

$$Var(X) = Var(X_1) + Var(X_2) + \cdots + Var(X_n)$$

Nous prouverons ce théorème, mais pour l'instant nous pouvons regarder un exemple pour voir comment nous pouvons l'utiliser.

Exemple 2.19. Si $X \sim \mathcal{B}(n,p)$. Trouver $Var(X)$.

Solution
Nous savons que nous pouvons écrire une v.a de $\mathcal{B}(n,p)$ comme la somme de n indépendantes v.a de $Bernoulli(p)$, c'est-à-dire $X = X_1 + X_2 + \cdots + X_n$. Ainsi, nous concluons

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n).$$

Si $X_i \sim Bernoulli(p)$, alors sa variance est

$$\mathrm{Var}(X_i) = E[X_i^2] - (E(X)_i)^2 = 1^2 \cdot p + 0^2 \cdot (1-p) - p^2 = p(1-p).$$

Ainsi,

$$\mathrm{Var}(X) = p(1-p) + p(1-p) + \cdots + p(1-p) = np(1-p).$$

2.2.12 Problèmes résolus : à savoir plus sur les variables aléatoires discrètes

1. Soit X une variable aléatoire discrète avec le PMF suivant

$$\mathbb{P}_X(x) = \begin{cases} 0,3 & \text{pour } x = 3 \\ 0,2 & \text{pour } x = 5 \\ 0,3 & \text{pour } x = 8 \\ 0,2 & \text{pour } x = 10 \\ 0 & \text{sinon} \end{cases}$$

Trouver et tracer le CDF de X.

Solution
Le CDF est défini par $F_X(x) = \mathbb{P}(X \leqslant x)$. Nous avons

$$F_X(x) = \begin{cases} 0 & \text{pour } x < 3 \\ \mathbb{P}_X(3) = 0.3 & \text{pour } 3 \leqslant x < 5 \\ \mathbb{P}_X(3) + \mathbb{P}_X(5) = 0.5 & \text{pour } 5 \leqslant x < 8 \\ \mathbb{P}_X(3) + \mathbb{P}_X(5) + \mathbb{P}_X(8) = 0.8 & \text{for } 8 \leqslant x < 10 \\ 1 & \text{for } x \geqslant 10 \end{cases}$$

2. Soit X une variable aléatoire discrète avec le PMF suivant

$$\mathbb{P}_X(k) = \begin{cases} 0.1 & \text{pour } k = 0 \\ 0.4 & \text{pour } k = 1 \\ 0.3 & \text{pour } k = 2 \\ 0.2 & \text{pour } k = 3 \\ 0 & \text{sinon} \end{cases}$$

(a) Trouver $E(X)$.

(b) Trouver $Var(X)$.

(c) Si $Y = (X-2)^2$, trouver $E(Y)$.

Solution

(a) $E(X) = \sum_{x_k \in R_X} x_k \mathbb{P}_X(x_k) = 0(0,1) + 1(0,4) + 2(0,3) + 3(0,2) = 1,6$

(b) Nous pouvons utiliser $Var(X) = E(X)^2 - (E(X))^2 = E(X)^2 - (1.6)^2$. Ainsi nous devons trouver $E(X)^2$. En utilisant LOTUS, nous avons

$$E(X)^2 = 0^2(0.1) + 1^2(0.4) + 2^2(0.3) + 3^2(0.2) = 3.4$$

(c) Ainsi, nous avons

$$\mathrm{Var}(X) = (3,4) - (1,6)^2 = 0,84$$

(d) Encore une fois, en utilisant LOTUS, nous avons

$$E(X-2)^2 = (0-2)^2(0,1) + (1-2)^2(0,4) + (2-2)^2(0,3) + (3-2)^2(0,2) = 1.$$

3. Soit X une variable aléatoire discrète avec PMF

$$\mathbb{P}_X(k) = \begin{cases} 0.2 & \text{pour } k = 0 \\ 0.2 & \text{pour } k = 1 \\ 0.3 & \text{pour } k = 2 \\ 0.3 & \text{pour } k = 3 \\ 0 & \text{sinon} \end{cases}$$

Définir $Y = X(X-1)(X-2)$. Trouver le PMF de Y.

Solution
Tout d'abord, notez que $R_Y = \{x(x-1)(x-2) | x \in \{0,1,2,3\}\} = \{0,6\}$. Ainsi,

$$\begin{aligned} \mathbb{P}_Y(0) =& \mathbb{P}(Y=0) = \mathbb{P}\big((X=0) \text{ ou } (X=1) \text{ ou } (X=2)\big) \\ =& \mathbb{P}_X(0) + \mathbb{P}_X(1) + \mathbb{P}_X(2) \\ =& 0,7; \\ \mathbb{P}_Y(6) =& \mathbb{P}(X=3) = 0,3 \end{aligned}$$

Ainsi,

$$\mathbb{P}_Y(k) = \begin{cases} 0,7 & \text{pour } k = 0 \\ 0,3 & \text{pour } k = 6 \\ 0 & \text{sinon} \end{cases}$$

4. Soit $X \sim Géométrique(p)$. Trouver $E\left[\frac{1}{2^X}\right]$.

Solution
Le PMF de X est donné par

$$\mathbb{P}_X(k) = \begin{cases} pq^{k-1} & \text{pour } k = 1,2,3,... \\ 0 & \text{sinon} \end{cases}$$

où $q = 1 - p$. Ainsi,

$$E\left[\frac{1}{2^X}\right] = \sum_{k=1}^{\infty} \frac{1}{2^k}\mathbb{P}_X(k) = \sum_{k=1}^{\infty} \frac{1}{2^k} q^{k-1} p = \frac{p}{2} \sum_{k=1}^{\infty} \left(\frac{q}{2}\right)^{k-1} = \frac{p}{2} \frac{1}{1-\frac{q}{2}} = \frac{p}{1+p}.$$

5. Si $X \sim Hypergéométrique(b,r,k)$, trouver $E(X)$.

Solution
Le PMF de X est donné par

$$\mathbb{P}_X(x) = \begin{cases} \frac{\binom{b}{x}\binom{r}{kx}}{\binom{b+r}{k}} & \text{pour } x \in R_X \\ 0 & \text{sinon} \end{cases}$$

où $R_X = \{\max(0,kr), \max(0,kr)+1, \max(0,kr)+2, ..., \min(k,b)\}$. Trouver $E(X)$ directement semble être très compliqué. Essayons donc de voir si nous pouvons trouver un moyen plus facile de trouver $E(X)$. En particulier, un outil puissant dont nous disposons est la linéarité des espérances. Peut-on écrire X comme la somme de variables aléatoires plus simples X_i? Pour ce faire, rappelons l'expérience aléatoire derrière la distribution hypergéométrique. Vous avez un sac qui contient b billes bleues et r billes rouges. Vous choisissez $k \leqslant b+r$ billes au hasard (sans remplacement) et Soit X le nombre de billes bleues dans votre échantillon. En particulier, définissons les variables aléatoires indicatrices X_i comme suit :

$$X_i = \begin{cases} 1 & \text{si la } i\text{ème bille choisie est bleue} \\ 0 & \text{sinon} \end{cases}$$

Ensuite, on peut écrire

$$X = X_1 + X_2 + \cdots + X_k.$$

Ainsi,

$$E(X) = E(X_1) + E(X_2) + \cdots + E(X_k).$$

Pour trouver $\mathbb{P}(X_i = 1)$, on note que pour tout particulierX_i toutes les billes sont également susceptibles d'être choisies. C'est à cause de la symétrie : aucune bille n'est plus susceptible d'être choisie que la $i$ème bille comme n'importe quelle autre bille. Donc,

$$\mathbb{P}(X_i = 1) = \frac{b}{b+r} \text{ pour tout } i \in \{1, 2, \cdots, k\}.$$

Nous concluons

$$E(X_i) = 0 \cdot \mathbb{P}(X_i = 0) + 1 \cdot \mathbb{P}(X_i = 1) = \frac{b}{b+r}.$$

Ainsi, nous avons

$$E(X) = \frac{kb}{b+r}.$$

6. Si $X \sim \mathcal{B}(n,p)$, alors $E(X) = np$. Nous l'avons prouvé en écrivant X comme la somme de n v.a de $Bernoulli(p)$. Maintenant, prouver $E(X)$ en utilisant directement $E(X) = \sum_{x_k \in R_X} x_k \mathbb{P}_X(x_k)$. (Astuce : utiliser $k\binom{n}{k} = n\binom{n-1}{k-1}$).

   Solution

   Première note que nous pouvons prouver $k\binom{n}{k} = n\binom{n-1}{k-1}$ par l'interprétation combinatoire suivante : Supposons qu'à partir d'un groupe de n étudiants nous voudrions choisir un comité de k étudiants, dont l'un est choisi pour être le président du comité. Nous pouvons le faire

   (a) en choisissant k personnes d'abord (en $\binom{n}{k}$ façons), puis en choisissant l'une d'elles pour être le président (k façons), ou

   (b) en choisissant d'abord la chaise (n possibilités, puis choisir $k-1$ étudiants du reste $n-1$ étudiants (en $\binom{n-1}{k-1}$ façons)).

Ainsi, nous concluons

$$k\binom{n}{k} = n\binom{n-1}{k-1}.$$

Maintenant, trouvons $E(X)$ pour $X \sim \mathcal{B}(n,p)$.

$$\begin{aligned}
E(X) = \sum_{k=0}^{n} k\binom{n}{k}p^k q^{nk} = \sum_{k=1}^{n} k\binom{n}{k}p^k q^{nk} = \sum_{k=1}^{n} n\binom{n-1}{k-1}p^k q^{nk} \\
=np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}q^{nk} = np\sum_{l=0}^{n-1}\binom{n-1}{l}p^l q^{(n-1)-l} = np
\end{aligned}$$

Notez que la dernière ligne est vraie car $\sum_{l=0}^{n-1}\binom{n-1}{l}p^l q^{(n-1)-l} = \sum_{l=0}^{n-1}\mathbb{P}_Y(l)$ pour une variable aléatoire Y qui suit une distribution $\mathcal{B}(n-1,p)$, il est donc égal à 1.

7. Soit X une variable aléatoire discrète avec $R_X \subset \{0,1,2,...\}$. Prouver

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

Solution
Noter que

$$\begin{aligned}
\mathbb{P}(X > 0) =& \mathbb{P}_X(1) + \mathbb{P}_X(2) + \mathbb{P}_X(3) + \mathbb{P}_X(4) + \cdots, \\
\mathbb{P}(X > 1) =& \mathbb{P}_X(2) + \mathbb{P}_X(3) + \mathbb{P}_X(4) + \cdots, \\
\mathbb{P}(X > 2) =& \mathbb{P}_X(3) + \mathbb{P}_X(4) + \mathbb{P}_X(5) + \cdots.
\end{aligned}$$

Ainsi

$$\begin{aligned}
\sum_{k=0}^{\infty}\mathbb{P}(X > k) =& \mathbb{P}(X > 0) + \mathbb{P}(X > 1) + \mathbb{P}(X > 2) + ... \\
=& \mathbb{P}_X(1) + 2\mathbb{P}_X(2) + 3\mathbb{P}_X(3) + 4\mathbb{P}_X(4) + ... \\
=& E(X).
\end{aligned}$$

8. Si $X \sim Poisson(\lambda)$, trouver $Var(X)$.

Solution
Nous savons déjà $E(X) = \lambda$, donc $Var(X) = E(X)^2 - \lambda^2$. Tu peux trouver $E(X)^2$ directement en utilisant LOTUS ; cependant, c'est un peu plus facile à trouver $E[X(X-1)]$ en premier. En particulier, en utilisant LOTUS, nous avons

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

Nous avons donc $\lambda^2 = E[X(X-1)] = E(X)^2 - E(X) = E(X)^2 - \lambda$. Ainsi, $E(X)^2 = \lambda^2 + \lambda$ et nous concluons

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

9. Soit X et Y deux variables aléatoires indépendantes. Supposons que nous connaissions $Var(2X - Y) = 6$ et $Var(X + 2Y) = 9$. Trouver $Var(X)$ et $Var(Y)$.

Solution
Assurons-nous d'abord de bien comprendre ce que $Var(2X - Y)$ et $Var(X + 2A)$ signifient. Ils sont $Var(Z)$ et $Var(W)$, où les variables aléatoires Z et Y sont définies comme $Z = 2X - Y$ et $W = X + 2Y$. Comme X et Y sont des variables aléatoires indépendantes, alors $2X$ et $-Y$ sont des variables aléatoires indépendantes. Aussi, X et $2Y$ sont des variables aléatoires indépendantes. Ainsi, nous pouvons écrire

$$\begin{aligned}\mathrm{Var}(2X - Y) &= \mathrm{Var}(2X) + \mathrm{Var(-Y)} = 4\mathrm{Var}(X) + \mathrm{Var(Y)} = 6,\\ \mathrm{Var}(X + 2Y) &= \mathrm{Var}(X) + \mathrm{Var\ (2Y)} = \mathrm{Var}(X) + 4\mathrm{Var\ (Y)} = 9.\end{aligned}$$

2.3 Variables aléatoires continues

Rappelez-vous que les variables aléatoires discrètes ne peuvent prendre qu'un nombre dénombrable de valeurs possibles. D'autre part, une variable aléatoire continue X a une image sous la forme d'un intervalle ou d'une union d'intervalles non superposés sur la ligne réelle (éventuellement toute la ligne réelle). Aussi, pour tout $x \in \mathbb{R}$, $\mathbb{P}(X = x) = 0$, le PMF ne fonctionne pas pour les variables aléatoires continues. Ainsi, nous devons développer de nouveaux outils pour traiter les variables aléatoires continues. La bonne nouvelle est que la théorie des variables aléatoires continues est complètement analogue à la théorie des variables aléatoires discrètes. En effet, si nous voulons trop simplifier les choses, nous pourrions dire ce qui suit : prenez n'importe quelle formule sur des variables aléatoires discrètes, puis remplacez les sommes par des intégrales, et remplacez les PMF par des fonctions de densité de probabilité (PDF), et vous obtiendrez la formule correspondante pour les variables aléatoires continues.

Définition 2.14. Une variable aléatoire X avec CDF $F_X(x)$ est dite continue si $F_X(x)$ est une fonction continue pour tout $x \in \mathbb{R}$.

Nous supposerons également que le CDF d'une variable aléatoire continue est différentiable presque partout dans $\mathbb{R}$.

2.3.1 Fonction de densité de probabilité (PDF)

Pour déterminer la distribution d'une variable aléatoire discrète, nous pouvons soit fournir son PMF ou son CDF. Pour les variables aléatoires continues, le CDF est bien défini, nous pouvons donc fournir le CDF. Cependant, le PMF ne fonctionne pas pour les variables aléatoires continues, car pour une variable aléatoire continue $\mathbb{P}(X = x) = 0$ pour tous $x \in \mathbb{R}$. Au lieu de cela, nous pouvons généralement définir la fonction de densité de probabilité (PDF). Le PDF est la densité de probabilité plutôt que la fonction de masse de probabilité. Le concept est très similaire à la densité de masse en physique : son unité est la probabilité par unité de longueur. Pour avoir une idée du PDF, considérons une variable aléatoire continue X et définissons la fonction $f_X(x)$ comme suit (là où la limite existe) :

$$f_X(x) = \lim_{\Delta \to 0^+} \frac{\mathbb{P}(x < X \leqslant x + \Delta)}{\Delta}.$$

La fonction $f_X(x)$ nous donne la densité de probabilité au point X. C'est la limite de la probabilité de l'intervalle $]x, x+\Delta]$ divisé par la longueur de l'intervalle lorsque la longueur de l'intervalle tend vers 0. Rappelez-vous que

$$\mathbb{P}(x < X \leqslant x + \Delta) = F_X(x + \Delta) - F_X(x).$$

Donc, nous concluons que

$$f_X(x) = \lim_{\Delta \to 0} \frac{F_X(x+\Delta) - F_X(x)}{\Delta} = \frac{dF_X(x)}{dx} = F'_X(x), \qquad \text{si } F_X(x) \text{ est dérivable en } x.$$

Ainsi, nous avons la définition suivante pour le PDF des variables aléatoires continues :

Définition 2.15. Considérons une variable aléatoire continue X avec un CDF absolument continu $F_X(x)$. La fonction $f_X(x)$ définie par

$$f_X(x) = \frac{dF_X(x)}{dx} = F'_X(x), \qquad \text{si } F_X(x) \text{ est dérivable en } x$$

est appelée la fonction de densité de probabilité (PDF) de X.

Puisque le PDF est la dérivée du CDF, le CDF peut être obtenu à partir du PDF par intégration (en supposant une continuité absolue) :

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du.$$

Aussi, nous avons

$$\mathbb{P}(a < X \leqslant b) = F_X(b) - F_X(a) = \int_a^b f_X(u)du.$$

En particulier, si on intègre sur toute la droite réelle, on doit obtenir 1, i.e,

$$\int_{-\infty}^{\infty} f_X(u)du = 1.$$

Autrement dit, l'aire sous la courbe PDF doit être égale à un.

Propriété. *Considérons une variable aléatoire continue X avec PDF $f_X(x)$. Nous avons*

1. $f_X(x) \geqslant 0$ *pour tout* $x \in \mathbb{R}$.
2. $\int_{-\infty}^{\infty} f_X(u)du = 1$.

3. $\mathbb{P}(a < X \leqslant b) = F_X(b) - F_X(a) = \int_a^b f_X(u)du$.
4. *Plus généralement, pour un ensemble A,* $\mathbb{P}(X \in A) = \int_A f_X(u)du$.

Dans le dernier élément ci-dessus, l'ensemble A doit satisfaire à certaines conditions qui sont presque toujours satisfaites dans la pratique. Un exemple d'ensemble A pourrait être une réunion de certains intervalles disjoints. Par exemple, si vous voulez trouver $\mathbb{P}(X \in [0,1] \cup [3,4])$, vous pouvez écrire

$$\mathbb{P}(X \in [0,1] \cup [3,4]) = \int_0^1 f_X(u)du + \int_3^4 f_X(u)du.$$

Prenons un exemple pour mettre en pratique les concepts ci-dessus.

Exemple 2.20. Soit X une variable aléatoire continue avec le PDF suivant

$$f_X(x) = \begin{cases} ce^{-x} & x \geqslant 0 \\ 0 & \text{sinon} \end{cases}$$

où c est une constante positive.

1. Trouver c.
2. Trouver le CDF de X, $F_X(x)$.
3. Trouver $\mathbb{P}(1 < X < 3)$.

Solution

1. Trouver c, nous pouvons utiliser la propriété 2 ci-dessus, en particulier

$$1 = \int_{-\infty}^{\infty} f_X(u)du = \int_0^{\infty} ce^{-u}du = c\Big[-e^{-x}\Big]_0^{\infty} = c.$$

Ainsi, nous devons avoir $c = 1$.

2. Pour trouver le CDF de X, on utilise $F_X(x) = \int_{-\infty}^{x} f_X(u)du$, donc pour $x < 0$, on obtient $F_X(x) = 0$. Pour $x \geqslant 0$, on a

$$F_X(x) = \int_0^x e^{-u}du = 1 - e^{-x}.$$

Ainsi,

$$F_X(x) = \begin{cases} 1 - e^{-x} & x \geqslant 0 \\ 0 & \text{sinon} \end{cases}$$

3. Nous pouvons trouver $\mathbb{P}(1 < X < 3)$ en utilisant le CDF ou le PDF. Si nous utilisons le CDF, nous avons

$$\mathbb{P}(1 < X < 3) = F_X(3) - F_X(1) = \left[1 - e^{-3}\right] - \left[1 - e^{-1}\right] = e^{-1} - e^{-3}.$$

De manière équivalente, nous pouvons utiliser le PDF. Nous avons

$$\mathbb{P}(1 < X < 3) = \int_1^3 f_X(t)dt = \int_1^3 e^{-t}dt = e^{-1} - e^{-3}.$$

Image

L'image d'une variable aléatoire X est l'ensemble des valeurs possibles de la variable aléatoire. Si X est une variable aléatoire continue, nous pouvons définir l'image de X comme l'ensemble des nombres réels X pour lequel le PDF est supérieur à zéro, i.e,

$$R_X = \{x | f_X(x) > 0\}.$$

L'ensemble R_X défini ici peut ne pas afficher exactement toutes les valeurs possibles de X, mais la différence est pratiquement sans importance.

2.3.2 Espérance mathématique et variance

Comme nous l'avons mentionné précédemment, la théorie des variables aléatoires continues est très similaire à la théorie des variables aléatoires discrètes. En particulier, les sommations sont généralement remplacées par des intégrales et les PMF sont remplacés par des PDF. Les preuves et les idées sont très analogues au cas discret, donc parfois nous énonçons les résultats sans dérivations mathématiques dans un souci de brièveté.

Rappelez-vous que l'espérance mathématique d'une variable aléatoire discrète peut être obtenue comme

$$E(X) = \sum_{x_k \in R_X} x_k \mathbb{P}_X(x_k).$$

Maintenant, en remplaçant la somme par une intégrale et PMF par PDF, nous pouvons écrire la définition de l'espérance mathématique d'une variable aléatoire continue comme

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Exemple 2.21. Soit X une variable aléatoire continue avec PDF

$$f_X(x) = \begin{cases} 2x & 0 \leqslant x \leqslant 1 \\ 0 & \text{sinon} \end{cases}$$

Trouver l'espérance mathématique de X.

Solution
Nous avons

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x(2x) dx = \int_0^1 2x^2 dx = \frac{2}{3}.$$

Espérance mathématique d'une fonction d'une variable aléatoire continue

Rappelons la loi du statisticien inconscient (LOTUS) pour les variables aléatoires discrètes :

$$E[g(X)] = \sum_{x_k \in R_X} g(x_k) \mathbb{P}_X(x_k) \qquad (4.2)$$

Maintenant, en changeant la somme en intégrale et en changeant le PMF en PDF, nous obtiendrons la formule similaire pour les variables aléatoires continues.

Loi du statisticien inconscient (LOTUS) pour variables aléatoires continues :

$$\mathbf{E[g(X)]} = \int_{-\infty}^{\infty} \mathbf{g(x) f_X(x) dx} \tag{4.3}$$

Comme nous l'avons vu précédemment, l'espérance est une opération linéaire, donc nous avons toujours

- $E[aX+b] = aE(X)+b$, pour tout $a,b \in \mathbb{R}$, et
- $E[X_1+X_2+...+X_n] = E(X_1)+E(X_2)+...+E(X_n)$, pour tout ensemble de variables aléatoires $X_1, X_2, ..., X_n$.

Exemple 2.22. Soit X une variable aléatoire continue avec PDF

$$f_X(x) = \begin{cases} x+\frac{1}{2} & 0 \leqslant x \leqslant 1 \\ 0 & \text{sinon} \end{cases}$$

Trouver $E(X^n)$, où $n \in \mathbb{N}$.

Solution
En utilisant LOTUS, nous avons

$$E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)dx = \int_0^1 x^n(x+\frac{1}{2})dx = \left[\frac{1}{n+2}x^{n+2} + \frac{1}{2(n+1)}x^{n+1}\right]_0^1 = \frac{3n+4}{2(n+1)(n+2)}.$$

Variance

Rappelez-vous que la variance de toute variable aléatoire est définie comme

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

Ainsi, pour une variable aléatoire continue, on peut écrire

$$\begin{aligned}\mathrm{Var}(X) =& E\left[(X-\mu_X)^2\right] = \int_{-\infty}^{\infty}(x-\mu_X)^2 f_X(x)dx \\ =& E(X^2) - (E(X))^2 = \int_{-\infty}^{\infty} x^2 f_X(x)dx - \mu_X^2\end{aligned}$$

Rappelez-vous aussi que pour $a,b \in \mathbb{R}$, nous avons toujours

$$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X). \tag{4.4}$$

Exemple 2.23. Soit X une variable aléatoire continue avec PDF

$$f_X(x) = \begin{cases} \frac{3}{x^4} & x \geqslant 1 \\ 0 & \text{sinon} \end{cases}$$

Trouver la moyenne et la variance de X.

Solution

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_{1}^{\infty} \frac{3}{x^3} dx = \left[-\frac{3}{2} x^{-2} \right]_1^{\infty} = \frac{3}{2}.$$

Ensuite, on trouve $E(X^2)$ en utilisant LOTUS,

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_{1}^{\infty} \frac{3}{x^2} dx = \left[-3x^{-1} \right]_1^{\infty} = 3.$$

Ainsi, nous avons

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = 3 - \frac{9}{4} = \frac{3}{4}.$$

2.3.3 Problèmes résolus : variables aléatoires continues

1. Soit X une variable aléatoire avec PDF donné par

$$f_X(x) = \begin{cases} cx^2 & |x| \leqslant 1 \\ 0 & \text{sinon} \end{cases}$$

   (a) Trouver la constante c.
   (b) Trouver $E(X)$ et $Var(X)$.
   (c) Trouver $\mathbb{P}(X \geqslant \frac{1}{2})$.

   Solution

   (a) Pour trouver c, on peut utiliser $\int_{-\infty}^{\infty} f_X(u) du = 1$:

$$1 = \int_{-\infty}^{\infty} f_X(u) du = \int_{-1}^{1} cu^2 du = \frac{2}{3} c.$$

   Ainsi, nous devons avoir $c = \frac{3}{2}$.

   (b) Pour trouver $E(X)$, nous pouvons écrire

$$E(X) = \int_{-1}^{1} u f_X(u) du = \frac{3}{2} \int_{-1}^{1} u^3 du = 0.$$

   En fait, on aurait pu deviner $E(X) = 0$ parce que le PDF est symétrique autour $x = 0$. Pour trouver la $Var(X)$, on a $\mathrm{Var}(X) = E(X^2) - (E(X))^2 = E(X^2) = \int_{-1}^{1} u^2 f_X(u) du = \frac{3}{2} \int_{-1}^{1} u^4 du = \frac{3}{5}$.
   Pour trouver $\mathbb{P}(X \geqslant \frac{1}{2})$, nous pouvons écrire

$$\mathbb{P}(X \geqslant \frac{1}{2}) = \frac{3}{2} \int_{\frac{1}{2}}^{1} x^2 dx = \frac{7}{16}.$$

2. Soit X une variable aléatoire continue de PDF donnée par

$$f_X(x) = \frac{1}{2}e^{-|x|}, \qquad \text{pour tout } x \in \mathbb{R}.$$

Si $Y = X^2$, trouver le CDF de Y.

Solution
Tout d'abord, nous remarquons que $R_Y = [0, \infty[$. Pour $y \in [0, \infty[$, on a

$$\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \leqslant y) \\
&= \mathbb{P}(X^2 \leqslant y) \\
&= \mathbb{P}(-\sqrt{y} \leqslant X \leqslant \sqrt{y}) \\
&= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2}e^{-|x|}dx \\
&= \int_0^{\sqrt{y}} e^{-x}dx \\
&= 1 - e^{-\sqrt{y}}.
\end{aligned}$$

Ainsi,

$$F_Y(y) = \begin{cases} 1 - e^{-\sqrt{y}} & y \geqslant 0 \\ 0 & \text{sinon} \end{cases}$$

3. Soit X une variable aléatoire continue avec PDF

$$f_X(x) = \begin{cases} 4x^3 & 0 < x \leqslant 1 \\ 0 & \text{sinon} \end{cases}$$

Trouver $\mathbb{P}(X \leqslant \frac{2}{3} | X > \frac{1}{3})$.

Solution
Nous avons

$$\mathbb{P}(X \leqslant \frac{2}{3} | X > \frac{1}{3}) = \frac{\mathbb{P}(\frac{1}{3} < X \leqslant \frac{2}{3})}{\mathbb{P}(X > \frac{1}{3})} = \frac{\int_{\frac{1}{3}}^{\frac{2}{3}} 4x^3dx}{\int_{\frac{1}{3}}^{1} 4x^3dx} = \frac{3}{16}.$$

4. Soit X une variable aléatoire continue avec PDF

$$f_X(x) = \begin{cases} x^2\left(2x + \frac{3}{2}\right) & 0 < x \leqslant 1 \\ 0 & \text{sinon} \end{cases}$$

Si $Y = \frac{2}{X} + 3$, trouver $Var(Y)$.

Solution
Tout d'abord, notez que

$$\mathrm{Var}(Y) = \mathrm{Var}\left(\frac{2}{X}+3\right) = 4\mathrm{Var}\left(\frac{1}{X}\right), \quad \text{en utilisant l'équation 4.4}$$

Ainsi, il suffit de trouver $Var(\frac{1}{X}) = E[\frac{1}{X^2}] - (E[\frac{1}{X}])^2$. En utilisant LOTUS, nous avons

$$E\left[\frac{1}{X}\right] = \int_0^1 x\left(2x+\frac{3}{2}\right)dx = \frac{17}{12}$$
$$E\left[\frac{1}{X^2}\right] = \int_0^1 \left(2x+\frac{3}{2}\right)dx = \frac{5}{2}.$$

Ainsi, $Var\left(\frac{1}{X}\right) = E[\frac{1}{X^2}] - (E[\frac{1}{X}])^2 = \frac{71}{144}$. Alors, on obtient

$$\mathrm{Var}(Y) = 4\mathrm{Var}\left(\frac{1}{X}\right) = \frac{71}{36}.$$

2.3.4 Distributions spéciales issues des variables aléatoires continues

2.3.5 Distribution uniforme

Choisissons un nombre réel uniformément au hasard dans l'intervalle $[a,b]$, et appelons ce nombre X. L'uniformité implique que la probabilité d'un intervalle de longueur l dans $[a,b]$ doit être proportionnelle à sa longueur :

$$\mathbb{P}(X \in [x_1,x_2]) \propto (x_2 - x_1), \quad \text{où } a \leqslant x_1 \leqslant x_2 \leqslant b.$$

Puisque $\mathbb{P}(X \in [a,b]) = 1$, nous concluons

$$\mathbb{P}(X \in [x_1,x_2]) = \frac{x_2 - x_1}{b-a}, \quad \text{où } a \leqslant x_1 \leqslant x_2 \leqslant b.$$

Maintenant, trouvons le CDF. Par définition $F_X(x) = \mathbb{P}(X \leqslant x)$, on a donc immédiatement

$$F_X(x) = 0, \quad \text{pour } x < a,$$
$$F_X(x) = 1, \quad \text{pour } x \geqslant b.$$

Pour $a \leqslant x \leqslant b$, on a

$$F_X(x) = \mathbb{P}(X \leqslant x) = \mathbb{P}(X \in [a,x]) = \frac{x-a}{b-a}.$$

Ainsi, pour résumer

$$F_X(x) = \begin{cases} 0 & \text{pour } x < a \\ \frac{x-a}{b-a} & \text{pour } a \leqslant x \leqslant b \\ 1 & \text{pour } x > b \end{cases} \tag{2.7}$$

Notez qu'ici peu importe si nous utilisons "$<$" ou alors "$\leqslant$", comme chaque point individuel a une probabilité nulle, donc par exemple $\mathbb{P}(X < 2) = \mathbb{P}(X \leqslant 2)$. La figure (2.8) montre le CDF de X. Comme nous nous attendons à ce que le CDF commence à zéro et se termine à 1.

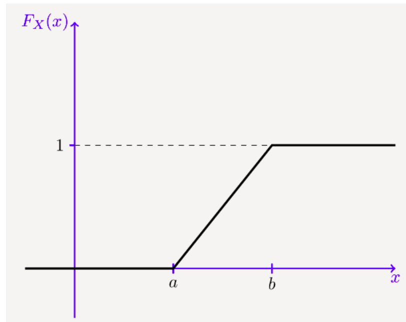


FIGURE 2.8 – CDF pour une variable aléatoire continue uniformément répartie sur $[a, b]$.

Définition 2.16. Une variable aléatoire continue X est dite avoir une distribution **uniforme** sur l'intervalle $[a, b]$, représentée comme $X \sim Uniforme(a, b)$, si son PDF est donné par

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & x < a \text{ ou } x > b \end{cases}$$

Notez que le CDF n'est pas différentiable aux points a et b. Néanmoins, comme nous le verrons plus loin, ce n'est pas important. La figure (2.9) montre le PDF de X. Comme on le voit, la valeur du PDF est constante dans l'intervalle de a à b. C'est pourquoi nous disons X est uniformément réparti sur $[a, b]$.

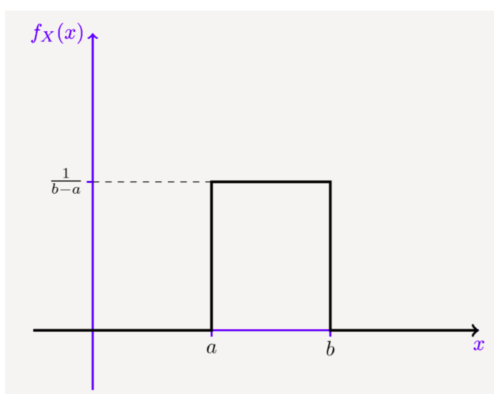


FIGURE 2.9 – PDF pour une variable aléatoire continue uniformément répartie sur $[a, b]$.

La distribution uniforme est la variable aléatoire continue la plus simple qu'on puisse imaginer. Pour les autres types de variables aléatoires continues, le PDF n'est pas uniforme.

Exemple 2.24. Soit $X \sim Uniforme(-1, 1)$ et $Y = X^2$. Trouver le CDF et le PDF de Y.

Solution
Tout d'abord, nous remarquons que $R_Y = [0,1]$. Comme d'habitude, nous commençons par le CDF. Pour $y \in [0,1]$, on a

$$\begin{aligned}
F_Y(y) =& \mathbb{P}(Y \leqslant y) = \mathbb{P}(X^2 \leqslant y) = \mathbb{P}(-\sqrt{y} \leqslant X \leqslant \sqrt{y}) \\
=& \frac{\sqrt{y} - (-\sqrt{y})}{1-(-1)} \qquad\qquad \text{puisque } X \sim Uniforme(-1,1) \\
=& \sqrt{y}.
\end{aligned}$$

Ainsi, le CDF de Y est donné par

$$F_Y(y) = \begin{cases} 0 & \text{pour } y < 0 \\ \sqrt{y} & \text{pour } 0 \leqslant y \leqslant 1 \\ 1 & \text{pour } y > 1 \end{cases}$$

Notez que le CDF est une fonction continue de Y, alors Y est une variable aléatoire continue. Ainsi, nous pouvons trouver le PDF de Y en différenciant $F_Y(y)$,

$$f_Y(y) = F'_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} & \text{pour } 0 \leqslant y \leqslant 1 \\ 0 & \text{sinon} \end{cases}$$

La moyenne de la distribution uniforme est donnée par

$$\boxed{E(X) = \frac{a+b}{2}}.$$

En effet, comme nous l'avons vu, le PDF de X est donné par

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & x < a \text{ ou } x > b \end{cases}$$

donc pour trouver son espérance mathématique, on peut écrire

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_a^b x(\frac{1}{b-a}) dx = \frac{1}{b-a}\left[\frac{1}{2}x^2\right]_a^b = \frac{a+b}{2}.$$

Ce résultat est intuitivement raisonnable : puisque X est uniformément réparti sur l'intervalle $[a,b]$, nous nous attendons à ce que sa moyenne soit le point médian, i.e, $E(X) = \frac{a+b}{2}$.

Exemple 2.25. Soit X une v.a $Uniforme(0,1)$, et soit $Y = e^X$.

1. Trouver le CDF de Y.
2. Trouver le PDF de Y.
3. Trouver $E(Y)$.

Solution
Tout d'abord, notez que nous connaissons déjà le CDF et le PDF de X. En particulier,

$$F_X(x) = \begin{cases} 0 & \text{pour } x < 0 \\ x & \text{pour } 0 \leqslant x \leqslant 1 \\ 1 & \text{pour } x > 1 \end{cases}$$

C'est une bonne idée de penser à l'image de Y avant de trouver la distribution. Comme e^x est une fonction croissante de X et $R_X = [0,1]$, nous concluons que $R_Y = [1,e]$. Donc on sait tout de suite que

$$F_Y(y) = \mathbb{P}(Y \leqslant y) = 0, \quad \text{pour } \ y < 1,$$
$$F_Y(y) = \mathbb{P}(Y \leqslant y) = 1, \quad \text{pour } \ y \geqslant e.$$

1. Pour trouver $F_Y(y)$ pour $y \in [1,e]$, nous pouvons écrire

$$\begin{aligned} F_Y(y) =& \mathbb{P}(Y \leqslant y) \\ =& \mathbb{P}(e^X \leqslant y) \\ =& \mathbb{P}(X \leqslant \ln y) \quad \text{puisque } e^x \text{ est une fonction croissante} \\ =& F_X(\ln y) \\ =& \ln y \qquad \text{puisque } \ 0 \leqslant \ln y \leqslant 1. \end{aligned}$$

   Pour résumer

$$F_Y(y) = \begin{cases} 0 & \text{pour } y < 1 \\ \ln y & \text{pour } \ 1 \leqslant y < e \\ 1 & \text{pour } \ y \geqslant e \end{cases}$$

2. Le CDF ci-dessus est une fonction continue, nous pouvons donc obtenir le PDF de Y en prenant sa dérivée. Nous avons

$$f_Y(y) = F'_Y(y) = \begin{cases} \frac{1}{y} & \text{pour } 1 \leqslant y \leqslant e \\ 0 & \text{sinon} \end{cases}$$

   Notez que le CDF n'est pas techniquement différentiable aux points 1 et e, mais comme nous l'avons mentionné précédemment, nous ne nous en soucions pas car il s'agit d'une variable aléatoire continue et la modification du PDF en un nombre fini de points ne modifie pas les probabilités.

3. Pour trouver $E(Y)$, nous pouvons appliquer directement LOTUS,

$$E[Y] = E[e^X] = \int_{-\infty}^{\infty} e^x f_X(x) dx = \int_0^1 e^x dx = e - 1.$$

   Pour ce problème, on pourrait aussi trouver $E(Y)$ en utilisant le PDF de Y,

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_1^e y \frac{1}{y} dy = e - 1.$$

   Notez que puisque nous avons déjà trouvé le PDF de Y peu importait la méthode que nous utilisions pour trouver $E[Y]$. Cependant, si le problème ne demandait que $E[Y]$ sans demander le PDF de Y, alors utiliser LOTUS serait beaucoup plus facile.

Et la variance de la distribution uniforme est

$$\boxed{\mathrm{Var}(X) = E(X^2) - (E(X))^2 = \frac{(b-a)^2}{12}}.$$

2.3.6 Distribution exponentielle

La distribution exponentielle est l'une des distributions continues les plus utilisées. Elle est souvent utilisée pour modéliser le temps écoulé entre les événements. Nous allons maintenant définir mathématiquement la distribution exponentielle et en déduire sa moyenne et son espérance mathématique. Ensuite, nous développerons l'intuition de la distribution et discuterons de plusieurs propriétés intéressantes qu'elle possède.

Définition 2.17. Une variable aléatoire continue X est dite avoir une distribution **exponentielle** de paramètre $\lambda > 0$, représentée comme $X \sim Exponentielle(\lambda)$, si son PDF est donné par

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{sinon} \end{cases}$$

La figure (2.10) montre le PDF de distribution exponentielle pour plusieurs valeurs de λ.

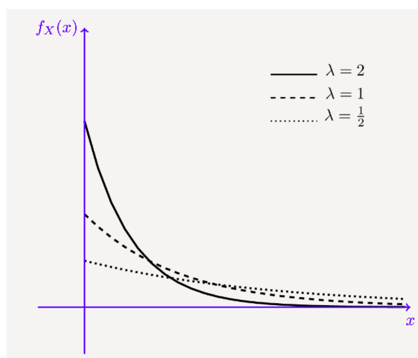


FIGURE 2.10 – PDF de la variable aléatoire exponentielle.

Si $X \sim Exponentielle(\lambda)$, alors sa moyenne est

$$E(X) = \frac{1}{\lambda}$$

et sa variance est

$$Var(X) = \frac{1}{\lambda^2}.$$

Une propriété intéressante de la distribution exponentielle est qu'elle peut être considérée comme un analogue continu de la distribution géométrique. Pour le voir, rappelez-vous l'expérience aléatoire derrière la distribution géométrique : vous lancez une pièce de monnaie (répétez une expérience de Bernoulli) jusqu'à ce que vous observiez les premières faces (succès). Maintenant, supposons que les lancers de pièces soient Δ secondes d'intervalle et à chaque lancer, la probabilité de succès est $p = \Delta\lambda$. Supposons également que Δ est très petit, donc les lancers de pièces sont très rapprochés dans le temps et la probabilité de succès à chaque essai est très faible. Soit X le moment où vous observez le premier succès.

Pour avoir une idée de cette interprétation de la distribution exponentielle, supposons que vous attendiez qu'un événement se produise. Par exemple, vous êtes dans un magasin et attendez le prochain client. À chaque milliseconde, la probabilité qu'un nouveau client entre dans le magasin est très faible. Vous pouvez imaginer que, à chaque milliseconde, une pièce (avec un tout petit $\mathbb{P}(H)$) est jetée, et si elle tombe sur la face, un nouveau client entre. Si vous lancez une pièce toutes les millisecondes, le temps jusqu'à ce qu'un nouveau client arrive suit approximativement une distribution exponentielle.

2.3.7 Distribution normale (gaussienne)

La distribution normale est de loin la distribution de probabilité la plus importante. L'une des principales raisons à cela est le *théorème central limite* (CLT) dont nous parlerons. Pour vous donner une idée, le CLT stipule que si vous ajoutez un grand nombre de variables aléatoires, la distribution de la somme sera approximativement normale sous certaines conditions. L'importance de ce résultat vient du fait que de nombreuses variables aléatoires dans la vie réelle peuvent être exprimées comme la somme d'un grand nombre de variables aléatoires et, par le CLT, nous pouvons affirmer que la distribution de la somme devrait être normale. Le CLT est l'un des résultats les plus importants en probabilité et nous en discuterons plus tard. Ici, nous allons introduire des variables aléatoires normales.

La distribution normale est de loin la distribution de probabilité la plus importante. Nous définissons d'abord la **variable aléatoire normale standard**.

Définition 2.18. Une variable aléatoire continue Z est dite une variable aléatoire normale standard (gaussienne standard), représentée par $Z \sim N(0,1)$, si son PDF est donné par

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\}, \qquad \text{pour tout } z \in \mathbb{R}.$$

Le $\frac{1}{\sqrt{2\pi}}$ est là pour s'assurer que la zone sous le PDF est égale à un. La figure (2.11) montre le PDF de la variable aléatoire normale standard.

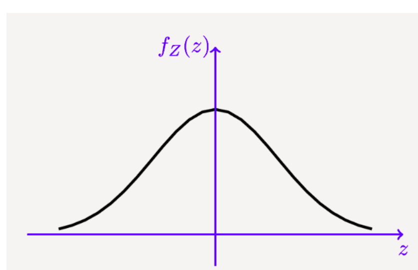


FIGURE 2.11 – PDF de la variable aléatoire normale standard.

Si $Z \sim N(0,1)$, alors sa moyenne est

$$E[Z] = 0$$

et sa variance est

$$Var(Z) = 1.$$

CDF de la distribution normale standard

Pour trouver le CDF de la distribution normale standard, nous devons intégrer la fonction PDF. En particulier, nous avons

$$F_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left\{-\frac{u^2}{2}\right\} du.$$

Cette intégrale n'a pas de solution de forme fermée. Néanmoins, en raison de l'importance de la distribution normale, les valeurs de $F_Z(z)$ ont été tabulées et de nombreuses calculatrices et progiciels ont cette fonction. Nous désignons généralement le CDF normal standard par Φ.

Définition 2.19. Le CDF de la distribution normale standard est désignée par la fonction Φ :

$$\Phi(x) = \mathbb{P}(Z \leqslant x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{u^2}{2}\right\} du.$$

Comme nous le verrons dans un instant, le CDF de toute variable aléatoire normale peut être écrite en termes de la fonction Φ, donc la fonction Φ est largement utilisée en probabilité. La figure (2.12) montre la fonction Φ.

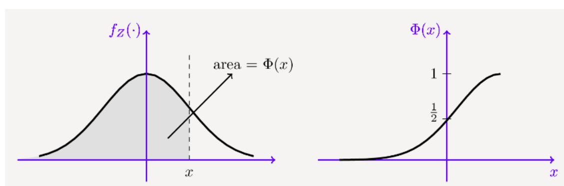


FIGURE 2.12 – La fonction Φ (CDF de normale standard).

Voici quelques propriétés de la fonction Φ qui peut être montrée à partir de sa définition.

1. $\lim_{x \to \infty} \Phi(x) = 1, \quad \lim_{x \to -\infty} \Phi(x) = 0;$
2. $\Phi(0) = \frac{1}{2};$
3. $\Phi(-x) = 1 - \Phi(x)$, pour tout $x \in \mathbb{R}$.

Aussi, puisque la fonction Φ n'a pas de forme fermée, il est parfois utile d'utiliser des bornes supérieures ou inférieures. En particulier, nous pouvons énoncer les bornes suivantes. Pour tout $x \geqslant 0$,

$$\frac{1}{\sqrt{2\pi}} \frac{x}{x^2+1} \exp\left\{-\frac{x^2}{2}\right\} \leqslant 1 - \Phi(x) \leqslant \frac{1}{\sqrt{2\pi}} \frac{1}{x} \exp\left\{-\frac{x^2}{2}\right\} \qquad (4.7)$$

Comme nous l'avons mentionné précédemment, en raison de l'importance de la distribution normale, les valeurs des Φ ont été tabulées et de nombreuses calculatrices et progiciels ont cette fonction. Par exemple, vous pouvez utiliser la commande $normcdf$ dans MATLAB (OCTAVE) pour calculer $\Phi(x)$ pour un nombre donné x. Plus précisément, $normcdf(x)$ donne $\Phi(x)$. Aussi, la fonction $norminv$ donne $\Phi^{-1}(x)$. C'est-à-dire que si vous compilez $x = norminv(y)$, alors x sera le nombre réel pour lequel $\Phi(x) = y$.

Variables aléatoires normales

Maintenant que nous avons vu la variable aléatoire normale standard, nous pouvons obtenir n'importe quelle variable aléatoire normale en déplaçant et en mettant à l'échelle une variable aléatoire normale standard. En particulier, définir

$$X = \sigma Z + \mu, \qquad \text{où } \sigma > 0.$$

Alors,

$$\begin{aligned} E(X) &= \sigma E(Z) + \mu = \mu, \\ \mathrm{Var}(X) &= \sigma^2 \mathrm{Var}(Z) = \sigma^2. \end{aligned}$$

Nous disons que X est une variable aléatoire normale de moyenne μ et variance σ^2. Nous écrivons $X \sim N(\mu, \sigma^2)$.

Définition 2.20. Si Z est une variable aléatoire normale standard et $X = \sigma Z + \mu$, alors X est une variable aléatoire normale de moyenne μ et de variance σ^2, i.e,

$$X \sim N(\mu, \sigma^2).$$

A l'inverse, si $X \sim N(\mu, \sigma^2)$, la variable aléatoire définie par $Z = \frac{X-\mu}{\sigma}$ est une variable aléatoire normale standard, i.e $Z \sim N(0,1)$. Pour trouver le CDF de $X \sim N(\mu, \sigma^2)$, nous pouvons écrire

$$\begin{aligned} F_X(x) =& \mathbb{P}(X \leqslant x) = \mathbb{P}(\sigma Z + \mu \leqslant x) \qquad \big(\text{où } Z \sim N(0,1)\big) \\ =& \mathbb{P}\left(Z \leqslant \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right). \end{aligned}$$

Pour trouver le PDF, on peut prendre la dérivée de F_X,

$$\begin{aligned} f_X(x) =& \frac{d}{dx} F_X(x) = \frac{d}{dx}\Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma}\Phi'\left(\frac{x-\mu}{\sigma}\right) \qquad \text{(règle de chaîne pour la dérivée)} \\ =& \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}. \end{aligned}$$

Si X est une variable aléatoire normale de moyenne μ et variance σ^2, i.e, $X \sim N(\mu, \sigma^2)$, alors

$$\begin{aligned} f_X(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \\ F_X(x) &= \mathbb{P}(X \leqslant x) = \Phi\left(\frac{x-\mu}{\sigma}\right), \\ \mathbb{P}(a < X \leqslant b) &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). \end{aligned}$$

La figure (2.13) montre le PDF de la distribution normale pour plusieurs valeurs de μ et σ.

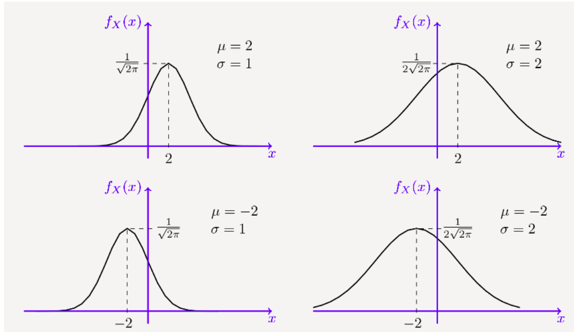


FIGURE 2.13 – PDF pour une distribution normale.

Exemple 2.26. Soit $X \sim N(-5,4)$.

1. Trouver $\mathbb{P}(X < 0)$.
2. Trouver $\mathbb{P}(-7 < X < -3)$.
3. Trouver $\mathbb{P}(X > -3|X > -5)$.

Solution
X est une variable aléatoire normale avec $\mu = -5$ et $\sigma = \sqrt{4} = 2$, on a donc

1. Trouver $\mathbb{P}(X < 0)$:

$$\mathbb{P}(X < 0) = F_X(0) = \Phi\left(\frac{0-(-5)}{2}\right) = \Phi(2,5) \approx 0,99$$

2. Trouver $\mathbb{P}(-7 < X < -3)$:

$$\begin{aligned}\mathbb{P}(-7 < X < -3) =& F_X(-3) - F_X(-7) = \Phi\left(\frac{(-3)-(-5)}{2}\right) - \Phi\left(\frac{(-7)-(-5)}{2}\right)\\ =&\Phi(1) - \Phi(-1) = 2\Phi(1) - 1 \qquad (\text{puisque } \Phi(-x) = 1 - \Phi(x))\\ \approx& 0,68\end{aligned}$$

3. Trouver $\mathbb{P}(X > -3|X > -5)$:

$$\begin{aligned}\mathbb{P}(X > -3|X > -5) =& \frac{\mathbb{P}(X > -3, X > -5)}{\mathbb{P}(X > -5)} = \frac{\mathbb{P}(X > -3)}{\mathbb{P}(X > -5)} = \frac{1 - \Phi\left(\frac{(-3)-(-5)}{2}\right)}{1 - \Phi\left(\frac{(-5)-(-5)}{2}\right)}\\ =& \frac{1-\Phi(1)}{1-\Phi(0)} \approx \frac{0.1587}{0.5} \approx 0.32\end{aligned}$$

2.3.8 Distribution gamma

La distribution gamma est une autre distribution largement utilisée. Son importance est largement due à sa relation avec les distributions exponentielles et normales. Ici, nous allons fournir une introduction à la distribution gamma. Avant d'introduire la variable aléatoire gamma, nous devons introduire la fonction gamma.

Fonction gamma : La fonction gamma, représentée par $\Gamma(x)$, est une extension de la fonction factorielle aux nombres réels (et complexes). Concrètement, si $n \in \{1, 2, 3, ...\}$, alors

$$\Gamma(n) = (n-1)!$$

Plus généralement, pour tout nombre réel positif α , $\Gamma(\alpha)$ est défini comme

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \mathrm{d}x, \qquad \text{pour } \alpha > 0.$$

La figure (2.14) montre la fonction gamma pour des valeurs réelles positives.

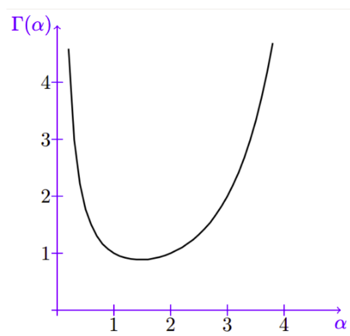


FIGURE 2.14 – La fonction Gamma pour certaines valeurs réelles de α.

Notez que pour $\alpha = 1$, nous pouvons écrire

$$\Gamma(1) = \int_0^\infty e^{-x} dx = 1.$$

Utiliser le changement de variable $x = \lambda y$, nous pouvons montrer l'équation suivante qui est souvent utile lorsque l'on travaille avec la distribution gamma :

$$\Gamma(\alpha) = \lambda^\alpha \int_0^\infty y^{\alpha-1} e^{-\lambda y} dy \qquad \text{pour } \alpha, \lambda > 0.$$

De plus, en utilisant l'intégration par parties, on peut montrer que

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha), \qquad \text{pour } \alpha > 0.$$

Notez que si $\alpha = n$, où n est un entier positif, l'équation ci-dessus se réduit à

$$n! = n \cdot (n-1)!$$

Propriété (Propriétés de la fonction gamma). *Pour tout nombre réel positif* α :

1. $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$;
2. $\int_0^\infty x^{\alpha-1}e^{-\lambda x}dx = \frac{\Gamma(\alpha)}{\lambda^\alpha}, \quad pour \ \ \lambda > 0$;
3. $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$;
4. $\Gamma(n) = (n-1)!$, *pour* $n = 1, 2, 3, \cdots$;
5. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Exemple 2.27.

1. Trouver $\Gamma(\frac{7}{2})$.
2. Trouver la valeur de l'intégrale suivante :

$$I = \int_0^\infty x^6 e^{-5x}dx.$$

Solution

1. Trouver $\Gamma(\frac{7}{2})$, nous pouvons écrire

$$\begin{aligned}
\Gamma(\frac{7}{2}) &= \frac{5}{2}\cdot\Gamma(\frac{5}{2}) \qquad \text{(en utilisant la propriété 3)} \\
&= \frac{5}{2}\cdot\frac{3}{2}\cdot\Gamma(\frac{3}{2}) \qquad \text{(en utilisant la propriété 3)} \\
&= \frac{5}{2}\cdot\frac{3}{2}\cdot\frac{1}{2}\cdot\Gamma(\frac{1}{2}) \qquad \text{(en utilisant la propriété 3)} \\
&= \frac{5}{2}\cdot\frac{3}{2}\cdot\frac{1}{2}\cdot\sqrt{\pi} \qquad \text{(en utilisant la propriété 5)} \\
&= \frac{15}{8}\sqrt{\pi}.
\end{aligned}$$

2. Utilisation de la propriété 2 avec $\alpha = 7$ et $\lambda = 5$, on obtient

$$\begin{aligned}
I &= \int_0^\infty x^6 e^{-5x}dx \\
&= \frac{\Gamma(7)}{5^7} \\
&= \frac{6!}{5^7} \qquad \text{(en utilisant la propriété 4)} \\
&\approx 0,0092
\end{aligned}$$

Distribution gamma :

Nous définissons maintenant la distribution gamma en fournissant son PDF :
Une variable aléatoire continue X est dite avoir une distribution gamma avec des paramètres $\alpha > 0$ et $\lambda > 0$, représentée comme $X \sim Gamma(\alpha, \lambda)$, si son PDF est donné par

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & x > 0 \\ 0 & \text{sinon} \end{cases}$$

Si nous posons $\alpha = 1$, on obtient

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{sinon} \end{cases}$$

Ainsi, nous concluons $Gamma(1, \lambda) = Exponentielle(\lambda)$. Plus généralement, si vous additionnez n variables aléatoires $Exponentiel(\lambda)$ indépendantes, alors vous obtiendrez une variable aléatoire $Gamma(n, \lambda)$. La figure (2.15) montre le PDF de la distribution gamma pour plusieurs valeurs de α.

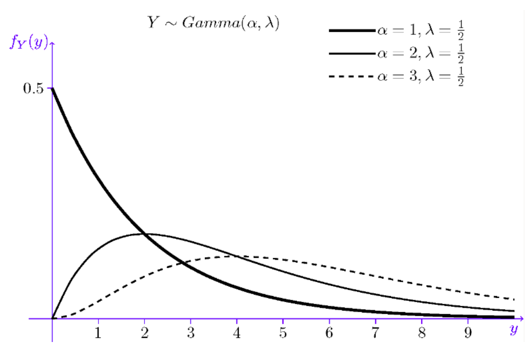


FIGURE 2.15 – PDF de la distribution gamma pour certaines valeurs de α et λ.

Exemple 2.28. À l'aide des propriétés de la fonction gamma, montrez que le PDF gamma s'intègre à 1, i.e que pour α, $\lambda > 0$, on a

$$\int_0^\infty \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} dx = 1.$$

Solution
Nous pouvons écrire

$$\begin{aligned} \int_0^\infty \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} dx &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-\lambda x} dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha)}{\lambda^\alpha} \qquad \text{(en utilisant la propriété 2 de la fonction gamma)} \\ &= 1. \end{aligned}$$

Si $X \sim Gamma(\alpha, \lambda)$, alors sa moyenne est

$$E(X) = \frac{\alpha}{\lambda},$$

et sa variance est

$$Var(X) = \frac{\alpha}{\lambda^2}.$$

2.3.9 Autres distributions

En plus des distributions spéciales dont nous avons discuté précédemment, il existe de nombreuses autres variables aléatoires continues qui sont utilisées dans la pratique. Selon les applications qui vous intéressent, vous devrez peut-être en traiter certaines dans vos cours comme les Lois du Ξ^2 , de Student et de Fisher.

2.3.10 Problèmes résolus : distributions continues spéciales

1. Supposons que le nombre de clients arrivant dans un magasin obéit à une distribution de Poisson avec une moyenne de λ clients par unité de temps. C'est-à-dire si Y est le nombre de clients arrivant dans un intervalle de longueur t, alors $Y \sim Poisson(\lambda t)$. Supposons que le magasin ouvre à l'heure $t = 0$. Soit X l'heure d'arrivée du premier client. Montrer que $X \sim Exponentielle(\lambda)$.

   Solution On trouve d'abord $\mathbb{P}(X > t)$:

   $$\mathbb{P}(X > t) = \mathbb{P}(\text{Pas d'arrivée dans } [0, t]) = e^{-\lambda t}\frac{(\lambda t)^0}{0!} = e^{-\lambda t}.$$

   Ainsi, le CDF de X pour $x > 0$ est donné par

   $$F_X(x) = 1 - \mathbb{P}(X > x) = 1 - e^{-\lambda x},$$

   qui est le CDF de $Exponentiel(\lambda)$. Notez que par le même argument, le temps entre le premier et le deuxième client a également une distribution $Exponentiel(\lambda)$. En général, le temps entre le k-ième et $k+1$-ième client est $Exponentiel(\lambda)$.

2. **(Exponentiel comme limite de la distribution géométrique)** Soit $Y \sim Géométrique(p)$, où $p = \lambda\Delta$. Définir $X = Y\Delta$, où λ, $\Delta > 0$. Prouver que pour tout $x \in (0, \infty)$, on a

   $$\lim_{\Delta \to 0} F_X(x) = 1 - e^{-\lambda x}.$$

   Solution
   Si $Y \sim Géométrique(p)$ et $q = 1 - p$, alors

   $$\mathbb{P}(Y \leqslant n) = \sum_{k=1}^{n} pq^{k-1} = p.\frac{1-q^n}{1-q} = 1 - (1-p)^n.$$

Alors pour tout $y \in]0, \infty[$, nous pouvons écrire

$$\mathbb{P}(Y \leqslant y) = 1 - (1-p)^{\lfloor y \rfloor},$$

où $\lfloor y \rfloor$ est le plus grand entier inférieur ou égal à y. Maintenant, puisque $X = Y\Delta$, on a

$$\begin{aligned} F_X(x) =& \mathbb{P}(X \leqslant x) \\ =& \mathbb{P}\left(Y \leqslant \frac{x}{\Delta}\right) \\ =& 1 - (1-p)^{\lfloor \frac{x}{\Delta} \rfloor} = 1 - (1-\lambda\Delta)^{\lfloor \frac{x}{\Delta} \rfloor}. \end{aligned}$$

Maintenant nous avons

$$\begin{aligned} \lim_{\Delta \to 0} F_X(x) = & \lim_{\Delta \to 0} 1 - (1-\lambda\Delta)^{\lfloor \frac{x}{\Delta} \rfloor} \\ =& 1 - \lim_{\Delta \to 0} (1-\lambda\Delta)^{\lfloor \frac{x}{\Delta} \rfloor} \\ =& 1 - e^{-\lambda x}. \end{aligned}$$

La dernière égalité tient car $\frac{x}{\Delta} - 1 \leqslant \lfloor \frac{x}{\Delta} \rfloor \leqslant \frac{x}{\Delta}$, et nous savons

$$\lim_{\Delta \to 0^+} (1-\lambda\Delta)^{\frac{1}{\Delta}} = e^{-\lambda}.$$

3. Soit $U \sim Uniforme(0,1)$ et $X = -\ln(1-U)$. Montrer que $X \sim Exponentiel(1)$.

   Solution
   A noter d'abord que puisque $R_U =]0,1[$, $R_X =]0, \infty[$. On trouvera le CDF de X. Pour $x \in]0, \infty[$, on a

$$F_X(x) = \mathbb{P}(X \leqslant x) = \mathbb{P}(-\ln(1-U) \leqslant x) = \mathbb{P}\left(\frac{1}{1-U} \leqslant e^x\right) = \mathbb{P}(U \leqslant 1 - e^{-x}) = 1 - e^{-x},$$

   qui est le CDF d'une variable aléatoire $Exponentielle(1)$.

4. Soit $X \sim N(2,4)$ et $Y = 3 - 2X$.

   (a) Trouver $\mathbb{P}(X > 1)$.
   (b) Trouver $\mathbb{P}(-2 < Y < 1)$.
   (c) Trouver $\mathbb{P}(X > 2 | Y < 1)$.

   Solution

(a) Trouver $\mathbb{P}(X > 1)$: Nous avons $\mu_X = 2$ et $\sigma_X = 2$. Ainsi,

$$\mathbb{P}(X > 1) = 1 - \Phi\left(\frac{1-2}{2}\right) = 1 - \Phi(-0.5) = \Phi(0.5) = 0.6915$$

(b) Trouver $\mathbb{P}(-2 < Y < 1)$: Comme $Y = 3 - 2X$, en utilisant le théorème 4.3, on a $Y \sim N(-1, 16)$. Donc,

$$\mathbb{P}(-2 < Y < 1) = \Phi\left(\frac{1-(-1)}{4}\right) - \Phi\left(\frac{(-2)-(-1)}{4}\right) = \Phi(0.5) - \Phi(-0.25) = 0.29$$

(c) Trouver $\mathbb{P}(X > 2|Y < 1)$:

$$\begin{aligned}\mathbb{P}(X > 2|Y < 1) =& \mathbb{P}(X > 2|3 - 2X < 1) = \mathbb{P}(X > 2|X > 1)\\ =& \frac{\mathbb{P}(X > 2, X > 1)}{\mathbb{P}(X > 1)} = \frac{\mathbb{P}(X > 2)}{\mathbb{P}(X > 1)}\\ =& \frac{1 - \Phi(\frac{2-2}{2})}{1 - \Phi(\frac{1-2}{2})} = \frac{1 - \Phi(0)}{1 - \Phi(-0.5)} \approx 0,72\end{aligned}$$

5. $X \sim N(0, \sigma^2)$. Trouver $E|X|$

Solution
Nous pouvons écrire $X = \sigma Z$, où $Z \sim N(0, 1)$. Ainsi, $E|X| = \sigma E|Z|$. Nous avons

$$\begin{aligned}E|Z| =& \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |t| e^{-\frac{t^2}{2}} dt = \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} |t| e^{-\frac{t^2}{2}} dt \qquad \text{(intégrale d'une fonction paire)}\\ =& \sqrt{\frac{2}{\pi}} \int_0^{\infty} t e^{-\frac{t^2}{2}} dt = \sqrt{\frac{2}{\pi}} \Big[- e^{-\frac{t^2}{2}} \Big]_0^{\infty} = \sqrt{\frac{2}{\pi}}\end{aligned}$$

Ainsi, nous concluons $E|X| = \sigma E|Z| = \sigma \sqrt{\frac{2}{\pi}}$.

6. Montrer que la constante de la distribution normale doit être $\frac{1}{\sqrt{2\pi}}$. C'est-à-dire montrer que

$$I = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}.$$

Indication : écrire I^2 comme une intégrale double en coordonnées polaires.

Solution
Soit $I = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx$. Nous montrons que $I^2 = 2\pi$. Pour voir cela, notez

$$I^2 = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx dy.$$

Pour évaluer cette double intégrale, nous pouvons passer aux coordonnées polaires. Cela peut être fait par changement de variables $x = r\cos\theta,\ y = r\sin\theta$, et $dxdy = rdrd\theta$. En particulier, nous avons

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}}dxdy = \int_0^{\infty}\int_0^{2\pi} e^{-\frac{r^2}{2}}rd\theta dr = 2\pi\int_0^{\infty} re^{-\frac{r^2}{2}}dr = 2\pi\Big[-e^{-\frac{r^2}{2}}\Big]_0^{\infty} = 2\pi.$$

7. $Z \sim N(0,1)$. Prouver pour tout $x \geqslant 0$,

$$\frac{1}{\sqrt{2\pi}}\frac{x}{x^2+1}e^{-\frac{x^2}{2}} \leqslant \mathbb{P}(Z \geqslant x) \leqslant \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-\frac{x^2}{2}}.$$

Solution

Pour montrer la borne supérieure, on peut écrire

$$\begin{aligned}\mathbb{P}(Z \geqslant x) =& \frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-\frac{u^2}{2}}du \leqslant \frac{1}{\sqrt{2\pi}}\int_x^{\infty}\frac{u}{x}e^{-\frac{u^2}{2}}du \quad (\text{puisque } u \geqslant x > 0)\\ =& \frac{1}{\sqrt{2\pi}}\frac{1}{x}\left[-e^{-\frac{u^2}{2}}\right]_X^{\infty} = \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-\frac{x^2}{2}}.\end{aligned}$$

Pour montrer la borne inférieure, soit $Q(x) = \mathbb{P}(Z \geqslant x)$, et

$$h(x) = Q(x) - \frac{1}{\sqrt{2\pi}}\frac{x}{x^2+1}e^{-\frac{x^2}{2}}, \quad \text{pour tout } x \geqslant 0.$$

Il suffit de montrer que $h(x) \geqslant 0, \quad \text{pour tout } x \geqslant 0$. Pour voir cela, notez que la fonction h a les propriétés suivantes

(a) $h(0) = \frac{1}{2}$;

(b) $\lim\limits_{x\to\infty} h(x) = 0$;

(c) $h'(x) = -\frac{2}{\sqrt{2\pi}}\left(\frac{e^{-\frac{x^2}{2}}}{(x^2+1)^2}\right) < 0, \ \text{pour tout } \ x\ geq0.$

Donc, $h(x)$ est une fonction strictement décroissante qui commence à $h(0) = \frac{1}{2}$ et diminue comme X augmente. Il s'approche de 0 comme X va à l'infini. Nous concluons que $h(x) \geqslant 0$, pour tout $x \geqslant 0$.

8. Soit $X \sim Gamma(\alpha,\lambda)$, où $\alpha,\ \lambda > 0$. Trouver $E(X)$, et $Var(X)$.

Solution
Trouver $E(X)$ nous pouvons écrire

$$
\begin{aligned}
E(X) = \int_0^\infty x f_X(x) dx &= \int_0^\infty x \cdot \frac{\lambda^\alpha}{\Gamma\alpha} x^{\alpha-1} e^{-\lambda x} \mathrm{d}x = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x \cdot x^{\alpha-1} e^{-\lambda x} \mathrm{d}x = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-\lambda x} \mathrm{d}x \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} \qquad \text{(en utilisant la propriété 2 de la fonction gamma)} \\
&= \frac{\alpha\Gamma(\alpha)}{\lambda\Gamma(\alpha)} \qquad \text{(en utilisant la propriété 3 de la fonction gamma)} \\
&= \frac{\alpha}{\lambda}.
\end{aligned}
$$

De même, on peut trouver $E(X^2)$:

$$
\begin{aligned}
E(X^2) &= \int_0^\infty x^2 \mathrm{d}x = \int_0^\infty x^2 \cdot \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathrm{d}x \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^2 \cdot x^{\alpha-1} e^{-\lambda x} \mathrm{d}x = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha+1} e^{-\lambda x} \mathrm{d}x \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+2)}{\lambda^{\alpha+2}} \qquad \text{(en utilisant la propriété 2 du gamma fonction)} \\
&= \frac{(\alpha+1)\Gamma(\alpha+1)}{\lambda^2\Gamma(\alpha)} \qquad \text{(en utilisant la propriété 3 de la fonction gamma)} \\
&= \frac{(\alpha+1)\alpha\Gamma(\alpha)}{\lambda^2\Gamma(\alpha)} \qquad \text{(en utilisant la propriété 3 de la fonction gamma)} \\
&= \frac{\alpha(\alpha+1)}{\lambda^2}.
\end{aligned}
$$

Ainsi, nous concluons

$$
Var(X) = E(X^2) - (E(X))^2 = \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}.
$$

2.4 Problèmes de fin de chapitre

2.4.1 Variables aléatoires discrètes

1. Soit X une variable aléatoire discrète avec le PMF suivant :

$$
\mathbb{P}_X(x) = \begin{cases} \frac{1}{2} & \text{pour } x = 0 \\ \frac{1}{3} & \text{pour } x = 1 \\ \frac{1}{6} & \text{pour } x = 2 \\ 0 & \text{sinon} \end{cases}
$$

   (a) Trouver R_X, l'image de la variable aléatoire X.
   (b) Trouver $\mathbb{P}(X \geqslant 1,5)$.

(c) Trouver $\mathbb{P}(0 < X < 2)$.

(d) Trouver $\mathbb{P}(X = 0 | X < 2)$.

2. Soit X le nombre de voitures réparées dans un atelier de réparation. Nous avons les informations suivantes :

   (a) A tout moment, il y a au plus 3 voitures en réparation.

   (b) La probabilité d'avoir 2 voitures au magasin est la même que la probabilité d'avoir une voiture.

   (c) La probabilité de ne pas avoir de voiture au magasin est la même que la probabilité d'avoir 3 voitures.

   (d) La probabilité d'avoir 1 ou alors 2 voitures est la moitié de la probabilité d'avoir 0 ou alors 3 voitures.

   Trouver le PMF de X.

3. Je lance deux dés et observe deux nombres X et Y. Si $Z = X - Y$, trouvez l'image et le PMF de Z.

4. Soit X et Y deux variables aléatoires discrètes indépendantes avec les PMF suivants :

$$\mathbb{P}_X(k) = \begin{cases} \frac{1}{4} & \text{pour } k = 1 \\ \frac{1}{8} & \text{pour } k = 2 \\ \frac{1}{8} & \text{pour } k = 3 \\ \frac{1}{2} & \text{pour } k = 4 \\ 0 & \text{sinon} \end{cases}$$

   et

$$\mathbb{P}_Y(k) = \begin{cases} \frac{1}{6} & \text{pour } k = 1 \\ \frac{1}{6} & \text{pour } k = 2 \\ \frac{1}{3} & \text{pour } k = 3 \\ \frac{1}{3} & \text{pour } k = 4 \\ 0 & \text{sinon} \end{cases}$$

   (a) Trouver $\mathbb{P}(X \leqslant 2 \text{ et } Y \leqslant 2)$.

   (b) Trouver $\mathbb{P}(X > 2 \text{ ou } Y > 2)$.

   (c) Trouver $\mathbb{P}(X > 2 | Y > 2)$.

   (d) Trouver $\mathbb{P}(X < Y)$.

5. 50 étudiants vivent dans un dortoir. Le parking a la capacité de 30 voitures. Si chaque élève a une voiture avec probabilité $\frac{1}{2}$ (indépendamment des autres élèves), quelle est la probabilité qu'il n'y ait pas assez de places de parking pour toutes les voitures ?

6. (le problème d'appariement)
   N invités arrivent à une fête. Chaque personne porte un chapeau. Nous collectons tous les chapeaux, puis redistribuons au hasard les chapeaux, en donnant à chacun l'un des N chapeaux

au hasard. Soit X_N le nombre de personnes qui reçoivent leurs propres chapeaux. Trouver le PMF de X_N.
Astuce :

$$\mathbb{P}(X_N = 0) = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots (-1)^N \frac{1}{N!},$$

pour $N = 1, 2, \cdots$. À l'aide de cela, trouver $\mathbb{P}(X_N = k)$ pour tout $k \in \{0, 1, \cdots N\}$.

7. Pour chacune des variables aléatoires suivantes, trouvez $\mathbb{P}(X > 5), \mathbb{P}(2 < X \leqslant 6)$ et $\mathbb{P}(X > 5 | X < 8)$.

   (a) $X \sim Géométrique(\frac{1}{5})$
   (b) $X \sim \mathcal{B}(10, \frac{1}{3})$
   (c) $X \sim Pascal(3, \frac{1}{2})$
   (d) $X \sim Hypergéométrique(10, 10, 12)$
   (e) $X \sim Poisson(5)$

8. Supposons que vous passiez un test de réussite à plusieurs reprises. Soit S_k l'événement que vous réussissez dans votre k^e essai et F_k l'événement où vous échouez au test dans votre k^e essai. Lors de votre premier essai, vous avez un 50 pourcentage de chance de réussir le test.

$$\mathbb{P}(S_1) = 1 - \mathbb{P}(F_1) = \frac{1}{2}.$$

   Supposons qu'à mesure que vous passez le test plus souvent, vos chances d'échouer au test diminuent. En particulier,

$$\mathbb{P}(F_k) = \frac{1}{2} \cdot \mathbb{P}(F_{k-1}), \qquad \text{pour } k = 2, 3, 4, \cdots$$

   Cependant, les résultats des différents examens sont indépendants. Supposons que vous passiez le test à plusieurs reprises jusqu'à ce que vous réussissiez le test pour la première fois. Soit X le nombre total de tests que vous passez, donc $Image(X) = \{1, 2, 3, \cdots\}$.

   (a) Trouver $\mathbb{P}(X = 1)$, $\mathbb{P}(X = 2)$, $\mathbb{P}(X = 3)$.
   (b) Trouver une formule générale pour $\mathbb{P}(X = k)$ pour $k = 1, 2, \cdots$.
   (c) Trouvez la probabilité que vous passiez le test plus de 2 fois.
   (d) Étant donné que vous passez le test plus d'une fois, trouvez la probabilité que vous le fassiez exactement deux fois.

9. Une urne se compose de 20 boules rouges et 30 boules vertes. Nous choisissons dix boules au hasard de l'urne. Le prélèvement se fait sans remise (répétition non autorisée).

   (a) Quelle est la probabilité qu'il y ait exactement 4 boules rouges parmi les boules choisies ?
   (b) Sachant qu'il y a au moins 3 boules rouges parmi les boules choisies, quelle est la probabilité qu'il y ait exactement 4 boules rouges ?

10. Le nombre d'e-mails que je reçois en semaine (du lundi au vendredi) peut être modélisé par une distribution de Poisson avec une moyenne de $\frac{1}{6}$ e-mails par minute. Le nombre d'e-mails que je reçois le week-end (samedi et dimanche) peut être modélisé par une distribution de Poisson avec une moyenne de $\frac{1}{30}$ e-mails par minute.

    (a) Quelle est la probabilité que je ne reçoive aucun e-mail dans un intervalle de longueur de 4 heures un dimanche ?

    (b) Un jour aléatoire est choisi (tous les jours de la semaine sont également susceptibles d'être sélectionnés), et un intervalle aléatoire d'une durée d'une heure est sélectionné le jour choisi. Il est observé que je n'ai reçu aucun e-mail dans cet intervalle. Quelle est la probabilité que le jour choisi soit un jour de semaine ?

11. Soit X une variable aléatoire discrète avec le PMF suivant :

$$\mathbb{P}_X(x) = \begin{cases} 0.2 & \text{pour } x = -2 \\ 0.3 & \text{pour } x = -1 \\ 0.2 & \text{pour } x = 0 \\ 0.2 & \text{pour } x = 1 \\ 0.1 & \text{pour } x = 2 \\ 0 & \text{sinon} \end{cases}$$

    Trouver et tracer le CDF de X.

12. Soit X une variable aléatoire discrète avec le CDF suivant :

$$F_X(x) = \begin{cases} 0 & \text{pour } x < 0 \\ \frac{1}{6} & \text{pour } 0 \leqslant x < 1 \\ \frac{1}{2} & \text{pour } 1 \leqslant x < 2 \\ \frac{3}{4} & \text{pour } 2 \leqslant x < 3 \\ 1 & \text{pour } x \geqslant 3 \end{cases}$$

    Trouver l'image et le PMF de X.

13. Soit X une variable aléatoire discrète avec le PMF suivant :

$$\mathbb{P}_X(k) = \begin{cases} 0,5 & \text{pour } k = 1 \\ 0,3 & \text{pour } k = 2 \\ 0.2 & \text{pour } k = 3 \\ 0 & \text{sinon} \end{cases}$$

    (a) Trouver $E(X)$.

    (b) Trouver $Var(X)$ et $ET(X)$ (écart-type).

    (c) Si $Y = \frac{2}{X}$, trouver $E[Y]$.

14. Soit $X \sim Géométrique(\frac{1}{3})$, et $Y = |X - 5|$. Trouver l'image et le PMF de Y.

15. Soit X une variable aléatoire discrète avec le PMF suivant

$$\mathbb{P}_X(k) = \begin{cases} \frac{1}{21} & \text{pour } k \in \{-10, -9, \cdots, -1, 0, 1, \cdots, 9, 10\} \\ \\ 0 & \text{sinon} \end{cases}$$

La variable aléatoire $Y = g(X)$ est définie comme

$$Y = g(X) = \begin{cases} 0 & \text{si } X \leqslant 0 \\ X & \text{si } 0 < X \leqslant 5 \\ 5 & \text{sinon} \end{cases}$$

Trouver le PMF de Y.

16. Soit $X \sim Géométrique(p)$. Trouver $Var(X)$.

17. $X \sim Pascal(m, p)$. Trouver $Var(X)$.

18. Supposer que $Y = -2X + 3$. Si nous savons $EY = 1$ et $EY^2 = 9$, trouver $E(X)$ et $Var(X)$.

19. Soit X une variable aléatoire de moyenne $E(X) = \mu$. Définir la fonction $f(\alpha)$ comme

$$f(\alpha) = E[(X - \alpha)^2].$$

Trouver la valeur de α qui minimise F.

20. On vous propose de jouer au jeu suivant. Vous lancez un dé équilibré une fois et observez le résultat qui est indiqué par la variable aléatoire X. À ce stade, vous pouvez arrêter le jeu et gagner X dollars. Vous pouvez également choisir de lancer le dé une deuxième fois pour observer la valeur Y. Dans ce cas, vous gagnerez Y dollars. Soit W la valeur que vous gagnez dans ce jeu. Quelle stratégie utilisez-vous pour maximiser $E(W)$? Quel est le maximum $E(W)$ vous pouvez atteindre en utilisant votre stratégie ?

2.4.2 Variables continues

1. Choisir un nombre réel uniformément au hasard dans l'intervalle $[2, 6]$ et appeler-le X. Trouver le CDF de X, $F_X(x)$. Trouver $E(X)$.

2. Soit X une variable aléatoire continue avec le PDF suivant

$$f_X(x) = \begin{cases} ce^{-4x} & x \geqslant 0 \\ 0 & sinon \end{cases}$$

où c est une constante positive.

   (a) Trouvez C.
   (b) Trouver le CDF de X, $F_X(x)$.
   (c) Trouver $\mathbb{P}(2 < X < 5)$.
   (d) Trouver $E(X)$.

3. Soit X une variable aléatoire continue avec PDF

$$f_X(x) = \begin{cases} x^2 + \frac{2}{3} & 0 \leqslant x \leqslant 1 \\ 0 & sinon \end{cases}$$

   (a) Trouver $E(X^n)$, pour $n = 1, 2, 3, \cdots$.
   (b) Trouver la variance de X.

4. Soit X une variable aléatoire $uniforme(0,1)$, et soit $Y = e^{-X}$.
   (a) Trouver le CDF de Y.
   (b) Trouver le PDF de Y.
   (c) Trouver EY.

5. Soit X une variable aléatoire continue avec PDF

$$f_X(x) = \begin{cases} \frac{5}{32}x^4 & 0 < x \leqslant 2 \\ 0 & sinon \end{cases}$$

   et $Y = X^2$.

   (a) Trouver le CDF de Y.
   (b) Trouver le PDF de Y.
   (c) Trouver EY.

6. Soit $X \sim Exponentielle(\lambda)$, et $Y = aX$, où a est un nombre réel positif. Montrer que

$$Y \sim Exponentielle\left(\frac{\lambda}{a}\right).$$

7. Soit $X \sim Exponentielle(\lambda)$. Montrer que
   (a) $E(X)^n = \frac{n}{\lambda}E(X)^{n-1}$, pour $n = 1, 2, 3, \cdots$;
   (b) $E(X)^n = \frac{n!}{\lambda^n}$, pour $n = 1, 2, 3, \cdots$.

8. Soit $X \sim N(3, 9)$.
   (a) Trouver $\mathbb{P}(X > 0)$.
   (b) Trouver $\mathbb{P}(-3 < X < 8)$.
   (c) Trouver $\mathbb{P}(X > 5 | X > 3)$.

9. Soit $X \sim N(3, 9)$ et $Y = 5 - X$.
   (a) Trouver $\mathbb{P}(X > 2)$.
   (b) Trouver $\mathbb{P}(-1 < Y < 3)$.
   (c) Trouver $\mathbb{P}(X > 4 | Y < 2)$.

10. Soit X une variable aléatoire continue avec PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad \text{pour tout } x \in \mathbb{R}.$$

et $Y = \sqrt{|X|}$. Trouver $f_Y(y)$.

11. Soit $X \sim Exponentiel(2)$ et $Y = 2 + 3X$.
    (a) Trouver $\mathbb{P}(X > 2)$.
    (b) Trouver EY et $Var(Y)$.
    (c) Trouver $\mathbb{P}(X > 2 | Y < 11)$.

12. La médiane d'une variable aléatoire continue X peut être définie comme le nombre réel unique m qui satisfait

$$\mathbb{P}(X \geqslant m) = \mathbb{P}(X < m) = \frac{1}{2}. \tag{2.8}$$

Trouver la médiane des variables aléatoires suivantes
    (a) $X \sim Uniforme(a, b)$.
    (b) $Y \sim Exponentiel(\lambda)$.
    (c) $W \sim N(\mu, \sigma^2)$.

13. Soit X une variable aléatoire avec le CDF suivant

$$F_X(x) = \begin{cases} 0 & \text{pour } : x < 0 \\ x & \text{pour } : 0 \leqslant x < \frac{1}{4} \\ x + \frac{1}{2} & \text{pour } : \frac{1}{4} \leqslant x < \frac{1}{2} \\ 1 & \text{pour } : x \geqslant \frac{1}{2} \end{cases}$$

Tracer le graphique de $F_X(x)$ et expliquer pourquoi X est une variable aléatoire mixte.
    (a) Trouver $\mathbb{P}(X \leqslant \frac{1}{3})$.
    (b) Trouver $\mathbb{P}(X \geqslant \frac{1}{4})$.
    (c) Ecrire le CDF de X sous la forme de

$$F_X(x) = C(x) + D(x),$$

    où $C(x)$ est une fonction continue et $D(x)$ se présente sous la forme d'une fonction en escalier, i.e

$$D(x) = \sum_k a_k u(x - x_k).$$

    (d) Trouver $c(x) = \frac{d}{dx} C(x)$.
    (e) Trouver $E(X)$ en utilisant $E(X) = \int_{-\infty}^{\infty} xc(x)dx + \sum_k x_k a_k$

14. Soit X une variable aléatoire avec le CDF suivant

$$F_X(x) = \begin{cases} 0 & \text{pour } : x < 0 \\ x & \text{pour } : 0 \leqslant x < \frac{1}{4} \\ x + \frac{1}{2} & \text{pour } : \frac{1}{4} \leqslant x < \frac{1}{2} \\ 1 & \text{pour } : x \geqslant \frac{1}{2} \end{cases}$$

   (a) Trouver le PDF généralisé de X, $f_X(x)$.
   (b) Trouver $E(X)$ en utilisant $f_X(x)$.
   (c) Trouver $Var(X)$ en utilisant $f_X(x)$.

15. Soit X une variable aléatoire mixte avec le PDF généralisé suivant

$$f_X(x) = \frac{1}{3}\delta(x+2) + \frac{1}{6}\delta(x-1) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

   (a) Trouver $\mathbb{P}(X = 1)$ et $\mathbb{P}(X = -2)$.
   (b) Trouver $\mathbb{P}(X \geqslant 1)$.
   (c) Trouver $\mathbb{P}(X = 1 | X \geqslant 1)$.
   (d) Trouver $E(X)$ et $Var(X)$.

16. Une entreprise fabrique un certain appareil. Nous nous intéressons à la durée de vie de l'appareil. On estime qu'environ 2% des appareils sont défectueux dès le départ donc ils ont une durée de vie de 0 ans. Si un appareil n'est pas défectueux, alors la durée de vie de l'appareil est distribuée de manière exponentielle avec un paramètre $\lambda = 2$ années. Soit X la durée de vie d'un appareil choisi au hasard.

   (a) Trouver le PDF généralisé de X.
   (b) Trouver $\mathbb{P}(X \geqslant 1)$.
   (c) Trouver $\mathbb{P}(X > 2 | X \geqslant 1)$.
   (d) Trouver $E(X)$ et $Var(X)$.

17. On dit qu'une variable aléatoire continue a une distribution de $Laplace(\mu, b)$ si son PDF est donné par

$$f_X(x) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right) = \begin{cases} \frac{1}{2b}\exp\left(\frac{x-\mu}{b}\right) & \text{si } : x < \mu \\ \frac{1}{2b}\mathbb{P}\left(-\frac{x-\mu}{b}\right) & \text{si } : x \geqslant \mu \end{cases}$$

   où $\mu \in \mathbb{R}$ et $b > 0$.

   (a) Si $X \sim Laplace(0,1)$, trouver $E(X)$ et $Var(X)$.
   (b) Si $X \sim Laplace(0,1)$ et $Y = bX + \mu$, montrer que $Y \sim Laplace(\mu, b)$.
   (c) Soit $Y \sim Laplace(\mu, b)$, où $\mu \in \mathbb{R}$ et $b > 0$. Trouver EY et $Var(Y)$.

18. Soit $X \sim Laplace(0, b)$, i.e,

$$f_X(x) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right),$$

où $b > 0$. Définir $Y = |X|$. Montrer que $\sim Exponentielle\left(\frac{1}{b}\right)$.

19. Une variable aléatoire continue est dite avoir la distribution de **Cauchy standard** si son PDF est donné par

$$f_X(x) = \frac{1}{\pi(1 + x^2)}.$$

Si X a une distribution de Cauchy standard, montrer que $E(X)$ n'est pas bien défini. Aussi, montrez $E(X)^2 = \infty$.

20. On dit qu'une variable aléatoire continue a une distribution de Rayleigh de paramètre σ si son PDF est donné par

$$f_X(x) = \frac{x}{\sigma^2} e^{-x^2/2\sigma^2} u(x) = \begin{cases} \frac{x}{\sigma^2} e^{-x^2/2\sigma^2} & \text{si } : x \geqslant 0 \\ 0 & \text{si } x < 0 \end{cases}$$

où $\sigma > 0$.

(a) Si $X \sim Rayleigh(\sigma)$, trouver $E(X)$.
(b) Si $X \sim Rayleigh(\sigma)$, trouver le CDF de X, $F_X(x)$.
(c) Si $X \sim Exponentiel(1)$ et $Y = \sqrt{2\sigma^2 X}$, montrer que $Y \sim Rayleigh(\sigma)$.

21. On dit qu'une variable aléatoire continue a une distribution $Pareto(x_m, \alpha)$ si son PDF est donné par

$$f_X(x) = \begin{cases} \alpha \frac{x_m^\alpha}{x^{\alpha+1}} & \text{pour } x \geqslant x_m \\ 0 & \text{pour } x < x_m \end{cases}$$

où x_m, $\alpha > 0$. Soit $X \sim Pareto(x_m, \alpha)$.

(a) Trouver le CDF de X, $F_X(x)$.
(b) Trouver $\mathbb{P}(X > 3x_m | X > 2x_m)$.
(c) Si $\alpha > 2$, trouver $E(X)$ et $Var(X)$.

22. Soit $Z \sim N(0, 1)$. Si nous définissons $X = e^{\sigma Z + \mu}$, alors on dit que X a une distribution log-normale avec des paramètres μ et σ, et on écrit $X \sim LogNormal(\mu, \sigma)$.

(a) Si $X \sim LogNormal(\mu, \sigma)$, trouver le CDF de X en termes de Φ une fonction. Trouver $E(X)$ et $Var(X)$.

23. Soit X_1, X_2, $\cdots$, X_n des variables aléatoires indépendantes avec $X_i \sim Exponentielle(\lambda)$. Définir

$$Y = X_1 + X_2 + \cdots + X_n.$$

Comme nous le verrons plus tard, Y a une distribution Gamma avec des paramètres n et λ, i.e, $Y \sim Gamma(n, \lambda)$. En utilisant ceci, montrez que si $Y \sim Gamma(n, \lambda)$, alors $EY = \frac{n}{\lambda}$ et $Var(Y) = \frac{n}{\lambda^2}$.

Chapitre 3

Théorèmes limites

Dans ce chapitre, nous discuterons des théorèmes limites Les théorèmes limites sont parmi les résultats les plus fondamentaux de la théorie des probabilités. Nous aborderons deux théorèmes limites importants dans la section suivante : la loi des grands nombres (LLN) et le théorème central limite (CLT). Nous parlerons également de l'importance de ces théorèmes appliqués en pratique. Le LLN stipule essentiellement que la moyenne d'un grand nombre de variables aléatoires indépendamment et identiquement distribuées (iid) converge vers l'espérance mathématique. Le CLT stipule que, sous certaines conditions, la somme d'un grand nombre de variables aléatoires a une distribution approximativement normale.

3.1 Loi des grands nombres

La **loi des grands nombres** joue un rôle très central dans les probabilités et les statistiques. Il stipule que si vous répétez une expérience indépendamment un grand nombre de fois et que vous faites la moyenne du résultat, ce que vous obtenez doit être proche de l'espérance mathématique. Il existe deux versions principales de la loi des grands nombres. On les appelle les lois **faibles** et **fortes** des grands nombres. La différence entre elles est surtout théorique. Dans ce chapitre, nous énonçons et prouvons la loi faible des grands nombres (WLLN). La loi forte des grands nombres n'est pas discutée ici. Avant de discuter du WLLN, définissons la moyenne de l'échantillon.

Définition 3.1. Pour des variables aléatoires iid $X_1, X_2, \ldots, X_n$, la moyenne de l'échantillon, notée par $\overline{X}$, est définie comme

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Une autre notation courante pour la moyenne de l'échantillon est M_n. Si les X_i ont des CDF $F_X(x)$, nous pourrions montrer la moyenne de l'échantillon par $M_n(X)$ pour indiquer la distribution des X_i.

A noter que comme les X_i sont des variables aléatoires, la moyenne de l'échantillon, $\overline{X} = M_n(X)$, est également une variable aléatoire. En particulier, nous avons

$$\begin{aligned} E[\overline{X}] &= \frac{EX_1 + EX_2 + \ldots + EX_n}{n} && \text{(par linéarité de l'espérance)} \\ &= \frac{nEX}{n} && \text{(puisque } EX_i = EX\text{)} \\ &= EX. \end{aligned}$$

Aussi, la variance de $\overline{X}$ est donnée par

$$\begin{aligned} \mathrm{Var}(\overline{X}) &= \frac{\mathrm{Var}(X_1 + X_2 + \ldots + X_n)}{n^2} && \text{(puisque } \mathrm{Var}(aX) = a^2\mathrm{Var}(X)\text{)} \\ &= \frac{\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \ldots + \mathrm{Var}(X_n)}{n^2} && \text{(car les } X_i \text{ sont indépendants)} \\ &= \frac{n\mathrm{Var}(X)}{n^2} && \text{(puisque } \mathrm{Var}(X_i) = \mathrm{Var}(X)\text{)} \\ &= \frac{\mathrm{Var}(X)}{n}. \end{aligned}$$

Énonçons et démontrons maintenant la loi faible des grands nombres (WLLN).

Théorème 3.1 (La loi faible des grands nombres (WLLN))**.** *Soient $X_1,\ X_2, \ldots, X_n$ des variables aléatoires iid avec une espérance mathématique finie $EX_i = \mu < \infty$. Ensuite, pour tout $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X} - \mu| \geqslant \varepsilon) = 0.$$

Démonstration. La démonstration de la loi faible des grands nombres est plus facile si l'on suppose $\mathrm{Var}(X) = \sigma^2$ est fini. Dans ce cas, nous pouvons utiliser l'inégalité de Chebyshev pour écrire

$$\begin{aligned} \mathbb{P}(|\overline{X} - \mu| \geqslant \varepsilon) &\leqslant \frac{\mathrm{Var}(\overline{X})}{\varepsilon^2} \\ &= \frac{\mathrm{Var}(X)}{n\varepsilon^2}, \end{aligned}$$

qui tend vers zéro comme $n \to \infty$. $\square$

3.2 Théorème central limite

Le théorème central limite (CLT) est l'un des résultats les plus importants de la théorie des probabilités. Il stipule que, sous certaines conditions, la somme d'un grand nombre de variables aléatoires est approximativement normale. Ici, nous énonçons une version du CLT qui s'applique aux variables aléatoires iid. Supposer que $X_1,\ X_2, \ldots, X_n$ sont des variables aléatoires iid avec des espérances mathématiques $EX_i = \mu < \infty$ et variance $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Ensuite, comme nous l'avons vu ci-dessus, la moyenne de l'échantillon $\overline{X} = \frac{X_1+X_2+\ldots+X_n}{n}$ a du sens $E\overline{X} = \mu$ et variance $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$. Ainsi, la variable aléatoire normalisée (standardisée)

$$Z_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sqrt{n}\sigma}$$

a pour moyenne $EZ_n = 0$ et variance $\mathrm{Var}(Z_n) = 1$. Le théorème central limite stipule que le CDF de Z_n converge vers le CDF de la distribution normale standard.

Théorème 3.2 (Le théorème central limite (CLT)). *Soient $X_1,\ X_2, \ldots, X_n$ des variables aléatoires iid avec une espérance mathématique $EX_i = \mu < \infty$ et variance $0 < \mathrm{Var}(X_i) = \sigma^2 < \infty$. Alors, la variable aléatoire*

$$Z_n = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sqrt{n}\sigma}$$

converge en distribution vers la variable aléatoire normale standard comme n tend vers l'infini, c'est-à-dire

$$\lim_{n \to \infty} \mathbb{P}(Z_n \leqslant x) = \Phi(x), \qquad \text{pour tout } x \in \mathbb{R},$$

où $\Phi(x)$ est le CDF de la distribution normale standard.

Une chose intéressante à propos du CLT est que peu importe la distribution des X_i. Les X_i peuvent être des variables aléatoires discrètes, continues ou mixtes. Pour avoir une idée du CLT, regardons quelques exemples. Supposons que les X_i sont des v.a de $Bernoulli(p)$. Alors $EX_i = p$, $\mathrm{Var}(X_i) = p(1-p)$. Aussi, $Y_n = X_1 + X_2 + \ldots + X_n$ possède une distribution $Binomiale(n, p)$. Ainsi,

$$Z_n = \frac{Y_n - np}{\sqrt{np(1-p)}},$$

où $Y_n \sim Binomiale(n, p)$. La figure (3.1) montre le PMF de Z_n pour différentes valeurs de n. Comme vous le voyez, la forme du PMF se rapproche d'une courbe PDF normale lorsque n augmente. Ici, Z_n est une variable aléatoire discrète, donc mathématiquement parlant, elle a un PMF et non un PDF. C'est pourquoi le CLT stipule que le CDF (et non le PDF) de Z_n converge vers le CDF de la distribution normale standard. Néanmoins, étant donné que PMF et PDF sont conceptuellement similaires, la figure est utile pour visualiser la convergence vers la distribution normale.

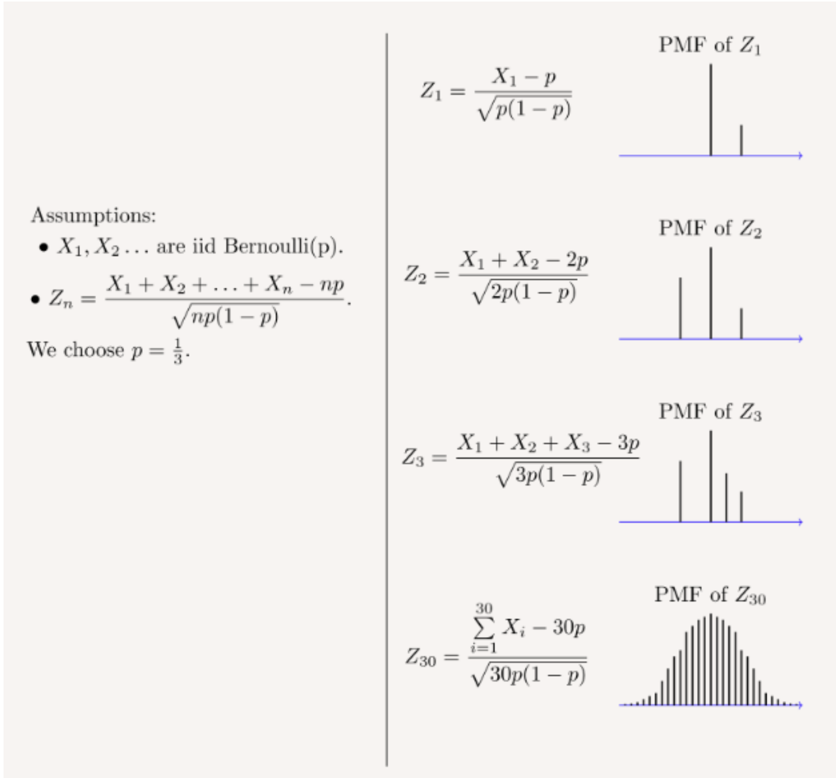


FIGURE 3.1 – Z_n est la somme normalisée de n indépendantes v.a de $Bernoulli(p)$. La forme de son PMF, $\mathbb{P}_{Z_n}(z)$, ressemble à la courbe de la distribution normale comme n augmente.

Comme autre exemple, supposons que X_i sont $Uniforme(0,1)$. Alors $EX_i = \frac{1}{2}$, $\text{Var}(X_i) = \frac{1}{12}$. Dans ce cas,

$$Z_n = \frac{X_1 + X_2 + \ldots + X_n - \frac{n}{2}}{\sqrt{n/12}}.$$

La figure (3.2) montre le PDF de Z_n pour différentes valeurs de n. Comme vous le voyez, la forme du PDF se rapproche du PDF normal à mesure que n augmente.

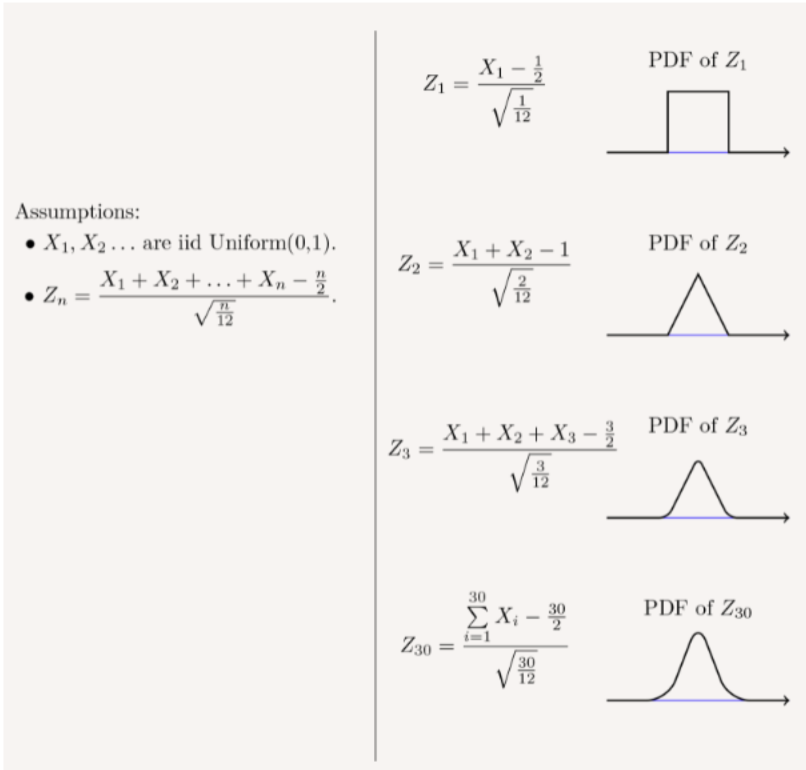


FIGURE 3.2 – Z_n est la somme normalisée de n indépendantes v.a de $Uniforme(0,1)$. La forme de son PDF, $\mathbb{P}_{Z_n}(z)$, se rapproche de la courbe de la distribution normale comme n augmente.

On aurait pu regarder directement $Y_n = X_1 + X_2 + \ldots + X_n$, alors pourquoi le normalisons-nous d'abord et disons que la version normalisée (Z_n) devient à peu près normal ? Ceci est dû au fait $EY_n = nEX_i$ et $\text{Var}(Y_n) = n\sigma^2$ tend vers l'infini comme n tend vers l'infini. Nous normalisons Y_n pour avoir une moyenne et une variance finies ($EZ_n = 0$, $\text{Var}(Z_n) = 1$). Néanmoins, pour tout n fixé, le CDF de Z_n est obtenu en mettant à l'échelle et en déplaçant le CDF de Y_n. Ainsi, les deux CDF ont des formes similaires.

L'importance du théorème central limite découle du fait que, dans de nombreuses applications réelles, une certaine variable aléatoire d'intérêt est la somme d'un grand nombre de variables aléatoires indépendantes. Dans ces situations, nous sommes souvent en mesure d'utiliser le CLT pour justifier l'utilisation de la distribution normale. On trouve des exemples de telles variables aléatoires dans presque toutes les disciplines. Voici quelques-uns :

- Les erreurs de mesure en laboratoire sont généralement modélisées par des variables aléatoires normales.
- Dans les communications et le traitement du signal, le bruit gaussien est le modèle de bruit le plus fréquemment utilisé.
- En finance, les variations en pourcentage des prix de certains actifs sont parfois modélisées par des variables aléatoires normales.

- Lorsque nous effectuons un échantillonnage aléatoire à partir d'une population pour obtenir des connaissances statistiques sur la population, nous modélisons souvent la quantité résultante comme une variable aléatoire normale.

Le CLT est également très utile dans le sens où il peut simplifier considérablement nos calculs. Si vous avez un problème dans lequel vous vous intéressez à une somme de mille variables aléatoires iid, il peut être extrêmement difficile, voire impossible, de trouver la distribution de la somme par calcul direct. En utilisant le CLT, nous pouvons immédiatement écrire la distribution, si nous connaissons la moyenne et la variance des X_i.

Une autre question qui vient à l'esprit est la taille que devrait avoir n pour que nous puissions utiliser l'approximation normale. La réponse dépend généralement de la distribution des X_i. Néanmoins, en règle générale, on dit souvent que si n est supérieur ou égal à 30, alors l'approximation normale est très bonne.

Résumons comment nous utilisons le CLT pour résoudre des problèmes :

Comment appliquer le théorème central limite (CLT).

Voici les étapes dont nous avons besoin pour appliquer le CLT :

1. Écrire la variable aléatoire d'intérêt, Y, comme la somme de n variables aléatoires iid X_i :
$$Y = X_1 + X_2 + \ldots + X_n.$$
2. Trouver EY et $\mathrm{Var}(Y)$ en notant que
$$EY = n\mu, \qquad \mathrm{Var}(Y) = n\sigma^2,$$
où $\mu = EX_i$ et $\sigma^2 = \mathrm{Var}(X_i)$.
3. Selon le CLT, conclure que $\frac{Y-EY}{\sqrt{\mathrm{Var}(Y)}} = \frac{Y-n\mu}{\sqrt{n}\sigma}$ est approximativement la distribution normale standard ; ainsi, trouver $\mathbb{P}(y_1 \leqslant Y \leqslant y_2)$, nous pouvons écrire
$$\begin{aligned}\mathbb{P}(y_1 \leqslant Y \leqslant y_2) &= \mathbb{P}\left(\frac{y_1 - n\mu}{\sqrt{n}\sigma} \leqslant \frac{Y - n\mu}{\sqrt{n}\sigma} \leqslant \frac{y_2 - n\mu}{\sqrt{n}\sigma}\right) \\ &\approx \Phi\left(\frac{y_2 - n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{y_1 - n\mu}{\sqrt{n}\sigma}\right).\end{aligned}$$

Examinons quelques exemples pour voir comment nous pouvons utiliser le théorème central limite.

Exemple 3.1. Un caissier de banque sert un par un des clients qui font la queue. Supposons que le temps de service X_i pour le client i a du sens $EX_i = 2$ (minutes) et $\mathrm{Var}(X_i) = 1$. Nous supposons que les temps de service pour les différents clients de la banque sont indépendants. Soit Y le temps total que le caissier de banque passe à servir 50 clients. Trouver $\mathbb{P}(90 < Y < 110)$.

Solution

$$Y = X_1 + X_2 + \ldots + X_n,$$

où $n = 50$, $EX_i = \mu = 2$, et $\mathrm{Var}(X_i) = \sigma^2 = 1$. Ainsi, nous pouvons écrire

$$\begin{aligned}
\mathbb{P}(90 < Y \leqslant 110) &= \mathbb{P}\left(\frac{90 - n\mu}{\sqrt{n}\sigma} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \frac{110 - n\mu}{\sqrt{n}\sigma}\right) \\
&= \mathbb{P}\left(\frac{90 - 100}{\sqrt{50}} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \frac{110 - 100}{\sqrt{50}}\right) \\
&= \mathbb{P}\left(-\sqrt{2} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \sqrt{2}\right).
\end{aligned}$$

Par le CLT, $\frac{Y-\mu}{\sqrt{n}\sigma}$ est approximativement la distribution normale standard, nous pouvons donc écrire

$$\begin{aligned}
\mathbb{P}(90 < Y \leqslant 110) &\approx \Phi(\sqrt{2}) - \Phi(-\sqrt{2}) \\
&= 0,8427
\end{aligned}$$

Exemple 3.2. Dans un système de communication, chaque paquet de données consiste en 1000 bits. En raison du bruit, chaque bit peut être reçu avec erreur de probabilité 0,1. Il est supposé que les erreurs sur les bits se produisent indépendamment. Trouver la probabilité qu'il y ait plus de 120 erreurs dans un certain paquet de données.

Solution
Définissons X_i comme variable aléatoire indicatrice pour le $i-$ème bit dans le paquet. C'est que, $X_i = 1$ si le $i-$ème bit est reçu par erreur, et $X_i = 0$ sinon. Alors les X_i sont iid et $X_i \sim Bernoulli(p = 0,1)$. Si Y est le nombre total d'erreurs sur les bits dans le paquet, nous avons

$$Y = X_1 + X_2 + \ldots + X_n.$$

Comme $X_i \sim Bernoulli(p = 0,1)$, on a

$$EX_i = \mu = p = 0.1, \qquad \mathrm{Var}(X_i) = \sigma^2 = p(1-p) = 0.09$$

En utilisant le CLT, nous avons

$$\begin{aligned}
\mathbb{P}(Y > 120) &= \mathbb{P}\left(\frac{Y - n\mu}{\sqrt{n}\sigma} > \frac{120 - n\mu}{\sqrt{n}\sigma}\right) \\
&= \mathbb{P}\left(\frac{Y - n\mu}{\sqrt{n}\sigma} > \frac{120 - 100}{\sqrt{90}}\right) \\
&\approx 1 - \Phi\left(\frac{20}{\sqrt{90}}\right) \\
&= 0.0175
\end{aligned}$$

Correction de continuité :

Supposons que $Y \sim Binomiale(n = 20, p = \frac{1}{2})$, et supposons que l'on s'intéresse à $\mathbb{P}(8 \leqslant Y \leqslant 10)$. Nous savons qu'une v.a $Binomiale(n = 20, p = \frac{1}{2})$ peut s'écrire comme la somme de n v.a iid de $Bernoulli(p)$:

$$Y = X_1 + X_2 + \ldots + X_n.$$

Comme $X_i \sim Bernoulli(p = \frac{1}{2})$, on a

$$EX_i = \mu = p = \frac{1}{2}, \qquad \mathrm{Var}(X_i) = \sigma^2 = p(1-p) = \frac{1}{4}.$$

Ainsi, nous pouvons vouloir appliquer le CLT pour écrire

$$\begin{aligned}
\mathbb{P}(8 \leqslant Y \leqslant 10) &= P\left(\frac{8 - n\mu}{\sqrt{n}\sigma} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \frac{10 - n\mu}{\sqrt{n}\sigma}\right) \\
&= P\left(\frac{8 - 10}{\sqrt{5}} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \frac{10 - 10}{\sqrt{5}}\right) \\
&\approx \Phi(0) - \Phi\left(\frac{-2}{\sqrt{5}}\right) \\
&= 0.3145
\end{aligned}$$

Puisque, ici, $n = 20$ est relativement petit, nous pouvons en fait trouver $\mathbb{P}(8 \leqslant Y \leqslant 10)$ avec précision. Nous avons

$$\mathbb{P}(8 \leqslant Y \leqslant 10) = \sum_{k=8}^{10} \binom{n}{k} p^k (1-p)^{n-k} = \left[\binom{20}{8} + \binom{20}{9} + \binom{20}{10}\right] \left(\frac{1}{2}\right)^{20} = 0,4565$$

Nous remarquons que notre approximation n'est pas très bonne. Une partie de l'erreur est due au fait que Y est une variable aléatoire discrète et nous utilisons une distribution continue pour trouver $\mathbb{P}(8 \leqslant Y \leqslant 10)$. Voici une astuce pour obtenir une meilleure approximation, appelée **correction de continuité**. Comme Y ne peut prendre que des valeurs entières, on peut écrire

$$\begin{aligned}
\mathbb{P}(8 \leqslant Y \leqslant 10) &= \mathbb{P}(7.5 < Y < 10.5) \\
&= P\left(\frac{7.5 - n\mu}{\sqrt{n}\sigma} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \frac{10.5 - n\mu}{\sqrt{n}\sigma}\right) \\
&= P\left(\frac{7.5 - 10}{\sqrt{5}} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \frac{10.5 - 10}{\sqrt{5}}\right) \\
&\approx \Phi\left(\frac{0.5}{\sqrt{5}}\right) - \Phi\left(\frac{-2.5}{\sqrt{5}}\right) \\
&= 0.4567
\end{aligned}$$

Comme nous le voyons, en utilisant la correction de continuité, notre approximation s'est considérablement améliorée. La correction de continuité est particulièrement utile lorsque l'on souhaite trouver $\mathbb{P}(y_1 \leqslant$

$Y \leqslant y_2)$, où Y est binomial et y_1 et y_2 sont proches les uns des autres.

Correction de continuité pour les variables aléatoires discrètes

Soient $X_1,\ X_2, \cdots,\ X_n$ des variables aléatoires discrètes indépendantes et soit

$$Y = X_1 + X_2 + \cdots + X_n.$$

Supposons que nous cherchions à trouver $\mathbb{P}(A) = \mathbb{P}(l \leqslant Y \leqslant u)$ en utilisant le CLT, où l et u sont des entiers. Comme Y est une variable aléatoire à valeur entière, on peut écrire

$$\mathbb{P}(A) = \mathbb{P}(l - \frac{1}{2} \leqslant Y \leqslant u + \frac{1}{2}).$$

Il s'avère que l'expression ci-dessus fournit parfois une meilleure approximation pour $\mathbb{P}(A)$ lors de l'application du CLT. C'est ce qu'on appelle la correction de continuité et il est particulièrement utile lorsque X_i sont Bernoulli (c'est-à-dire, Y est binomial).

3.3 Problèmes résolus

1. Il y a 100 hommes dans un avion. Soit X_i le poids (en kilogrammes) du $i-$ème homme dans l'avion. Supposons que les X_i sont iid, et $EX_i = \mu = 170$ et $\sigma_{X_i} = \sigma = 30$. Trouver la probabilité que le poids total des hommes dans l'avion dépasse 18 000 kg.

   Solution
   Si Y est le poids total, alors $W = X_1 + X_2 + \cdots + X_n$, où $n = 100$. Nous avons

   $$\begin{aligned} EW =& n\mu = (100)(170) = 17000, \\ \mathrm{Var}(W) =& 100\mathrm{Var}(X_i) = (100)(30)^2 = 90000. \end{aligned}$$

   Ainsi, $\sigma_W = 300$. Nous avons

   $$\begin{aligned} \mathbb{P}(W > 18000) &= \mathbb{P}\left(\frac{W-17000}{300} > \frac{18000-17000}{300}\right) = \mathbb{P}\left(\frac{W-17000}{300} > \frac{10}{3}\right) \\ &= 1 - \Phi\left(\frac{10}{3}\right) \quad \text{(par CLT)} \\ &\approx 4,3 \times 10^{-4}. \end{aligned}$$

2. Soit $X_1,\ X_2, \cdots,\ X_{25}$ des v.a iid avec le PMF suivant

   $$\mathbb{P}_X(k) = \begin{cases} 0.6 & k = 1 \\ 0.4 & k = -1 \\ 0 & \text{sinon} \end{cases}$$

Et soit

$$Y = X_1 + X_2 + \cdots + X_n.$$

En utilisant le CLT et la correction de continuité, estimer $\mathbb{P}(4 \leqslant Y \leqslant 6)$.

Solution
Nous avons

$$\begin{aligned} EX_i =& (0.6)(1) + (0.4)(-1) = \frac{1}{5}, \\ EX_i^2 =& 0.6 + 0.4 = 1. \end{aligned}$$

Donc,

$$\begin{aligned} \mathrm{Var}(X_i) &= 1 - \frac{1}{25} = \frac{24}{25}; \\ \text{ainsi,} \quad \sigma_{X_i} &= \frac{2\sqrt{6}}{5}. \end{aligned}$$

Donc,

$$EY = 25 \times \frac{1}{5} = 5,$$

$$\begin{aligned} \mathrm{Var}(Y) &= 25 \times \frac{24}{25} = 24; \\ \text{ainsi,} \quad \sigma_Y &= 2\sqrt{6}. \end{aligned}$$

$$\begin{aligned} \mathbb{P}(4 \leqslant Y \leqslant 6) &= \mathbb{P}(3.5 \leqslant Y \leqslant 6.5) \quad \text{(correction de continuité)} \\ &= \mathbb{P}\left(\frac{3.5-5}{2\sqrt{6}} \leqslant \frac{Y-5}{2\sqrt{6}} \leqslant \frac{6.5-5}{2\sqrt{6}}\right) \\ &= \mathbb{P}\left(-0.3062 \leqslant \frac{Y-5}{2\sqrt{6}} \leqslant +0.3062\right) \\ &\approx \Phi(0.3062) - \Phi(-0.3062) \quad \text{(par le CLT)} \\ &= 2\Phi(0.3062) - 1 \\ &\approx 0.2405 \end{aligned}$$

3. Vous avez invité 64 invités à une fête. Vous devez faire des sandwichs pour les invités. Vous pensez qu'un client pourrait avoir besoin de 0, 1 ou 2 sandwichs avec probabilités $\frac{1}{4}$, $\frac{1}{2}$, et $\frac{1}{4}$ respectivement. Vous partez du principe que le nombre de sandwichs dont chaque client a besoin est indépendant de celui des autres clients. Combien de sandwichs devriez-vous faire pour être à 95% sûr qu'il n'y a pas de pénurie ?

Solution
Soit X_i le nombre de sandwichs que la $i-$ème personne a besoin, et soit

$$Y = X_1 + X_2 + \cdots + X_{64}.$$

Le but est de trouver y tel que

$$\mathbb{P}(Y \leqslant y) \geqslant 0,95 \tag{3.1}$$

A noter d'abord que

$$EX_i = \frac{1}{4}(0) + \frac{1}{2}(1) + \frac{1}{4}(2) = 1,$$
$$EX_i^2 = \frac{1}{4}(0^2) + \frac{1}{2}(1^2) + \frac{1}{4}(2^2) = \frac{3}{2}.$$

Ainsi,

$$\mathrm{Var}(X_i) = EX_i^2 - (EX_i)^2 = \frac{3}{2} - 1 = \frac{1}{2} \quad \rightarrow \quad \sigma_{X_i} = \frac{1}{\sqrt{2}}.$$

Ainsi,

$$EY = 64 \times 1 = 64,$$
$$\mathrm{Var}(Y) = 64 \times \frac{1}{2} = 32 \rightarrow \sigma_Y = 4\sqrt{2}.$$

Maintenant, nous pouvons utiliser le CLT pour trouver y

$$\mathbb{P}(Y \leqslant y) = \mathbb{P}\left(\frac{Y-64}{4\sqrt{2}} \leqslant \frac{y-64}{4\sqrt{2}}\right) = \Phi\left(\frac{y-64}{4\sqrt{2}}\right) \quad \text{(par CLT)}.$$

Nous pouvons écrire

$$\Phi\left(\frac{y-64}{4\sqrt{2}}\right) = 0.95$$

Donc,

$$\frac{y-64}{4\sqrt{2}} = \Phi^{-1}(0.95) \approx 1.6449$$

Ainsi, $y = 73,3$. Par conséquent, si vous faites 74 sandwichs, vous êtes à 95% sûr qu'il n'y a pas de pénurie. Notez que vous pouvez trouver la valeur numérique de $\Phi^{-1}(0,95)$ en exécutant la commande *norminv(0.95)* dans MATLAB ou OCTAVE.

4. Soit $X_1,\ X_2, \cdots, X_n$ des variables aléatoires iid $Exponentiel(\lambda)$ avec $\lambda = 1$. Soit

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Quelle devrait être la taille de n pour que

$$\mathbb{P}\left(0.9 \leqslant \overline{X} \leqslant 1.1\right) \geqslant 0.95\,?$$

Solution
Soit $Y = X_1 + X_2 + \cdots + X_n$, alors $\overline{X} = \frac{Y}{n}$. Comme $X_i \sim Exponentielle(1)$, on a

$$E(X_i) = \frac{1}{\lambda} = 1, \qquad \mathrm{Var}(X_i) = \frac{1}{\lambda^2} = 1.$$

Donc,

$$E(Y) = nEX_i = n, \qquad \mathrm{Var}(Y) = n\mathrm{Var}(X_i) = n,$$

$$\begin{aligned}
\mathbb{P}(0.9 \leqslant \overline{X} \leqslant 1.1) &= \mathbb{P}\left(0.9 \leqslant \frac{Y}{n} \leqslant 1.1\right) = \mathbb{P}\big(0.9n \leqslant Y \leqslant 1.1n\big) \\
&= \mathbb{P}\left(\frac{0.9n - n}{\sqrt{n}} \leqslant \frac{Y-n}{\sqrt{n}} \leqslant \frac{1.1n - n}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(-0.1\sqrt{n} \leqslant \frac{Y-n}{\sqrt{n}} \leqslant 0.1\sqrt{n}\right).
\end{aligned}$$

Par le CLT $\frac{Y-n}{\sqrt{n}}$ est approximativement $N(0,1)$, alors

$$\begin{aligned}
\mathbb{P}(0.9 \leqslant \overline{X} \leqslant 1.1) &\approx \Phi\left(0.1\sqrt{n}\right) - \Phi\left(-0.1\sqrt{n}\right) \\
&= 2\Phi\left(0.1\sqrt{n}\right) - 1 \quad (\text{puisque} \quad \Phi(-x) = 1 - \Phi(x)).
\end{aligned}$$

Nous devons avoir

$$2\Phi\left(0.1\sqrt{n}\right) - 1 \geqslant 0.95, \qquad \text{donc} \quad \Phi\left(0.1\sqrt{n}\right) \geqslant 0,975.$$

Ainsi,

$$0.1\sqrt{n} \geqslant \Phi^{-1}(0.975) = 1.96 \tag{3.2}$$
$$\sqrt{n} \geqslant 19.6 \tag{3.3}$$
$$n \geqslant 384.16 \tag{3.4}$$

Comme n est un entier, on en déduit $n \geqslant 385$.

3.4 Exercices

1. Soit X_i une v.a $Uniforme(0,1)$ iid. Nous définissons la moyenne de l'échantillon comme

$$M_n = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

   (a) Trouver $E[M_n]$ et $\mathrm{Var}(M_n)$ en tant que fonction de n.
   (b) En utilisant l'inégalité de Chebyshev, trouvez une borne supérieure sur

$$\mathbb{P}\left(\left|M_n - \frac{1}{2}\right| \geqslant \frac{1}{100}\right).$$

(c) À l'aide de votre limite, montrez que

$$\lim_{n\to\infty} \mathbb{P}\left(\left|M_n - \frac{1}{2}\right| \geqslant \frac{1}{100}\right) = 0.$$

2. Le nombre d'accidents dans une certaine ville est modélisé par une variable aléatoire de Poisson avec un taux moyen de dix accidents par jour. Supposons que le nombre d'accidents sur différents jours soit indépendant. Utilisez le théorème central limite pour trouver la probabilité qu'il y ait plus de 3800 accidents au cours d'une certaine année. Supposons qu'il existe 365 jours dans une année.

3. Dans un système de communication, chaque mot de code se compose de 1000 bits. En raison du bruit, chaque bit peut être reçu avec erreur de probabilité 0,1. Il est supposé que les erreurs sur les bits se produisent indépendamment. Étant donné que des codes de correction d'erreurs sont utilisés dans ce système, chaque mot de code peut être décodé de manière fiable s'il y a moins ou égal à 125 erreurs dans le mot de code reçu, sinon le décodage échoue. À l'aide du CLT, trouvez la probabilité d'échec du décodage.

4. 50 étudiants vivent dans un dortoir. Le parking a la capacité de 30 voitures. Chaque élève a une voiture avec probabilité $\frac{1}{2}$, indépendamment des autres étudiants. Utilisez le CLT (avec correction de continuité) pour trouver la probabilité qu'il n'y ait pas assez de places de stationnement pour toutes les voitures.

5. Le temps nécessaire à une certaine machine pour traiter une tâche est une variable aléatoire dont la moyenne $EX_i = 10$ minutes et $\mathrm{Var}(X_i) = 2 \text{ minutes}^2$. Les temps nécessaires aux différentes tâches sont indépendants les uns des autres. Trouvez la probabilité que la machine traite au moins 40 tâches dans les 7 heures.

6. Vous avez une pièce équitable. Vous lancez la pièce n fois. Soit X la portion de fois que vous observez les têtes. Que doit être la valeur de n pour que vous soyez à 95% sûr que $0,45 \leqslant X \leqslant 0,55$? En d'autres termes, quelle est la taille que doit avoir n pour que

$$\mathbb{P}(0.45 \leqslant X \leqslant 0.55) \geqslant .95 \qquad ? \tag{3.5}$$

7. Un ingénieur mesure une quantité q. On suppose qu'il y a une erreur aléatoire dans chaque mesure, donc l'ingénieur prendra n mesures et rapporte la moyenne des mesures comme valeur estimée de q. Concrètement, si Y_i est la valeur obtenue dans la $i-$ème mesure, nous supposons que

$$Y_i = q + X_i,$$

où X_i est l'erreur dans la $i-$ème mesure. Nous supposons que les X_i sont iid avec $EX_i = 0$ et $\mathrm{Var}(X_i) = 4$ unités. L'ingénieur rapporte la moyenne des mesures

$$M_n = \frac{Y_1 + Y_2 + \ldots + Y_n}{n}.$$

Combien de mesures l'ingénieur doit-il effectuer jusqu'à ce qu'il soit 95% sûr que l'erreur finale est inférieure à 0,1 unité ? En d'autres termes, quelle devrait être la valeur de n pour que

$$\mathbb{P}(q - 0.1 \leqslant M_n \leqslant q + 0.1) \geqslant 0.95 \qquad ? \tag{3.6}$$