

Statistique des données expérimentales de la physique

Notes de cours, Partie “Probabilités”

Céline Esser

Année académique 2026–2027

Table des matières

Introduction	iii
1 Événements et probabilités	1
1.1 Expérience aléatoire et événements	1
1.2 Approche fréquentiste des probabilités	4
1.3 Mesures de probabilité	5
1.4 Mesures de probabilités dans un cas équiprobable	10
1.5 Mesures de probabilité sur un ensemble fini ou dénombrable	15
1.6 Mesures de probabilité sur $\mathbb{R}$	16
2 Probabilités conditionnelles et indépendance	17
2.1 Probabilités conditionnelles	17
2.2 Probabilités composées	20
2.3 Loi des probabilités totales	20
2.4 Formule de Bayes	22
2.5 Indépendance	24
3 Variables aléatoires	28
3.1 Variables aléatoires	28
3.2 Variables aléatoires discrètes et continues	31
3.3 Fonction de répartition	35
3.4 Espérance et moments d'une variable aléatoire	40
3.5 Lois discrètes usuelles	48
3.5.1 Loi discrète uniforme	48
3.5.2 Loi de Bernoulli et loi binomiale	49
3.5.3 Loi géométrique	51
3.5.4 Loi de Poisson	54
3.6 Lois continues usuelles	55
3.6.1 Loi continue uniforme	55
3.6.2 Loi exponentielle	56
3.6.3 Loi normale	59
3.6.4 Deux lois dérivées de la Gaussienne	62

3.7	Transformation de variables aléatoires	63
4	Lois jointes	68
4.1	Loi jointe et fonction de répartition	68
4.2	Lois bivariées discrètes	69
4.3	Lois bivariées continues	77
4.4	Lois bivariées mixtes	83
4.5	Indépendance de variables aléatoires	83
4.6	Covariance et corrélation	91
4.6.1	Propagation des incertitudes : approximation linéaire	96
5	Théorèmes limites	97
5.1	Loi des grands nombres	97
5.2	Théorème central limite	99
5.3	Une application : méthode de Monte-Carlo	104
5.4	Extension au cas multivarié	105
	Références	108

Introduction

Ce document constitue la première partie du cours de *Statistique des données expérimentales de la physique*. Elle est consacrée aux bases de la théorie des probabilités qui seront nécessaires dans la seconde partie du cours pour aborder la statistique inférentielle.

L'objectif est de disposer d'un cadre mathématique permettant de modéliser le hasard et l'incertitude, puis de décrire quantitativement les résultats d'expériences aléatoires. En statistique, ces outils permettront ensuite de raisonner à partir d'un nombre limité d'observations afin d'obtenir de l'information sur une population, un paramètre ou un modèle sous-jacent. Les notions de loi de probabilité, d'espérance, de variance, de loi jointe, d'indépendance et de convergence jouent ainsi un rôle central dans la suite du cours.

L'approche adoptée ici est volontairement progressive. Nous partons d'abord de l'interprétation fréquentiste des probabilités : lorsqu'une expérience est répétée un grand nombre de fois dans des conditions comparables, la fréquence d'apparition d'un événement tend à se stabiliser. Cette idée, essentielle dans les sciences expérimentales, fournit une intuition naturelle de la notion de probabilité. Elle ne constitue cependant pas à elle seule une définition mathématique satisfaisante. Nous introduisons donc ensuite le cadre axiomatique de Kolmogorov, dans lequel une probabilité est définie comme une mesure sur une famille d'événements.

Le but n'est pas de développer la théorie de la mesure dans toute sa généralité, mais d'utiliser le formalisme nécessaire pour pouvoir manipuler les probabilités de manière rigoureuse. Une attention particulière sera portée à l'interprétation des objets introduits, à leur utilisation concrète et aux liens avec les situations rencontrées en physique expérimentale.

Le cours est organisé de la manière suivante. Le premier chapitre introduit les événements, les espaces probabilisés et les principales règles de calcul des probabilités. Le deuxième chapitre est consacré aux probabilités conditionnelles, à la formule de Bayes et à l'indépendance. Nous introduisons ensuite les variables aléatoires, leurs lois, leurs fonctions de répartition, leurs moments et les principales lois usuelles. Le chapitre suivant étudie les couples de variables aléatoires, les lois jointes, les lois marginales et conditionnelles, ainsi que la covariance et la corrélation. Enfin, le dernier chapitre présente les principaux théorèmes limites, notamment la loi des grands nombres et le théorème central limite, qui fournissent le lien naturel entre théorie des probabilités et statistique.

Ces notes s'appuient notamment sur les cours de probabilités et de statistique de D. Paindaveine et Y. Swan, dont plusieurs exemples, idées pédagogiques et illustrations ont été repris ou adaptés. D'autres références sont indiquées dans la bibliographie en fin de document.

1 Événements et probabilités

La théorie des probabilités cherche à formaliser et à quantifier notre compréhension du hasard et de l'incertitude. Étant donnée une expérience aléatoire, on cherche à donner une mesure de la confiance que l'on a en le fait qu'un événement lié à cette expérience se produise ou non. Une approche fondamentale pour appréhender les probabilités est l'approche fréquentiste, qui repose sur la répétition des expériences et l'observation des fréquences de réalisation des événements. Cette approche se révèle vite limitée, c'est pourquoi nous présentons une approche abstraite et théorique, basée sur les axiomes de Kolmogorov. Ces axiomes fournissent un cadre rigoureux pour attribuer des probabilités aux événements tout en garantissant la cohérence au travers des opérations ensemblistes.

1.1 Expérience aléatoire et événements

Considérons une *expérience aléatoire*, c'est-à-dire une expérience qui a plusieurs résultats possibles dont un seul se produit à chaque réalisation de l'expérience, et dont l'issue n'est pas connue avec certitude à l'avance. Nous supposons que, même si le résultat de l'expérience n'est pas connu, l'ensemble des résultats possibles est connu. On le note Ω et on l'appelle *l'univers* ou *l'espace fondamental* de l'expérience. On peut penser à une expérience aléatoire comme l'acte de “choisir un résultat $\omega \in \Omega$ au hasard”.

Exemple 1.1.1.

▷ Si on lance une pièce de monnaie, $\Omega = \{\text{Pile}, \text{Face}\}$.

▷ Si on lance une pièce de monnaie deux fois d'affilée,

$$\Omega = \{(\text{Pile}, \text{Pile}), (\text{Face}, \text{Face}), (\text{Pile}, \text{Face}), (\text{Face}, \text{Pile})\}.$$

▷ Si on lance deux dés distinguables, $\Omega = \{(j, k) : j, k \in \{1, \dots, 6\}\}$.

▷ Si on lance deux dés indistinguables, $\Omega = \{\{j, k\} : j, k \in \{1, \dots, 6\}\}$.

▷ Si on s'intéresse au gagnant d'un tournoi de tennis masculin, Ω est l'ensemble des joueurs inscrits au tournoi.

▷ Si on mesure le temps d'attente en minutes à un arrêt de bus, $\Omega = [0, +\infty[$.

Un *événement* lié à cette expérience est une proposition liée au résultat de cette expérience, qui est soit vraie soit fausse à l'issue de l'expérience. On peut le voir comme une question que l'on peut se poser par rapport à cette expérience. Tout événement peut s'identifier à l'ensemble des résultats $\omega \in \Omega$ pour lesquels il se réalise. Tout événement est donc un sous-ensemble de Ω . On dira que l'événement A a lieu si le résultat de l'expérience est $\omega \in A$.

Exemple 1.1.2.

- ▷ Si on lance une pièce de monnaie deux fois d'affilée, on peut s'intéresser à l'événement "*Le deuxième lancer est Pile*" qui correspond au sous-ensemble $\{(Pile, Pile), (Face, Pile)\}$.
- ▷ Si on lance deux dés distinguables, on peut s'intéresser à l'événement "*la somme des deux dés est égale à 8*", qui correspond au sous-ensemble

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

- ▷ Lors d'un tournoi de tennis, on peut s'intéresser à l'événement "*le gagnant appartient au Big Four*" qui correspond au sous-ensemble

$$\{\text{Nadal, Federer, Djokovic, Murray}\}.$$

- ▷ Lorsqu'on attend le bus, on peut s'intéresser à l'événement "*le bus arrive dans 10 minutes maximum*", qui correspond au sous-ensemble $[0, 10]$.

Les opérations ensemblistes permettent de construire de nouveaux événements. Pour rappel, on considère les définitions suivantes.

Définition 1.1.3. Soient A et B deux sous-ensembles de Ω .

- ▷ $A \cap B = \{\omega \in \Omega : \omega \in A \text{ et } \omega \in B\}$ correspond à l'événement "*A et B ont lieu*".
- ▷ $A \cup B = \{\omega \in \Omega : \omega \in A \text{ ou } \omega \in B\}$ correspond à l'événement "*A ou B ont lieu*".
- ▷ $A^c = \Omega \setminus A = \{\omega \in \Omega : \omega \notin A\}$ correspond à l'événement "*A n'a pas lieu*".
- ▷ $A \setminus B = \{\omega \in \Omega : \omega \in A \text{ et } \omega \notin B\} = A \cap B^c$ correspond à l'événement "*A a lieu mais pas B*".

Exemple 1.1.4. Considérons l'expérience qui consiste à lancer un dé. On a

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

On s'intéresse aux événements "*le résultat est pair*" et "*le résultat est strictement plus grand que 4*". On a donc $A = \{2, 4, 6\}$ et $B = \{5, 6\}$. Ainsi, on a $A \cap B = \{6\}$, $A \cup B = \{2, 4, 5, 6\}$ et $A^c = \{1, 3, 5\}$.

Rappelons que les opérations d'union et d'intersection ont les propriétés suivantes. Tout est intuitif si on se rappelle que " $\cup$ = ou" et que " $\cap$ = et".

Proposition 1.1.5. Soit Ω un ensemble et soient A, B, C des sous-ensembles de Ω .

- ▷ **Commutativité** : $A \cup B = B \cup A$, $A \cap B = B \cap A$
- ▷ **Associativité** : $A \cup (B \cup C) = (A \cup B) \cup C$, $A \cap (B \cap C) = (A \cap B) \cap C$
- ▷ **Distributivité** : $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- ▷ **Lois de Morgan** : $(A \cup B)^c = A^c \cap B^c$ et $(A \cap B)^c = A^c \cup B^c$.

On peut également considérer ces opérations ensemblistes sur un nombre fini ou infini dénombrable d'événements. En particulier, on a les définitions suivantes.

Définition 1.1.6. Soit $I \subseteq \mathbb{N}$ (fini ou non) et soit $(A_n)_{n \in I}$ une suite de sous-ensembles de Ω . On a

- ▷ $\bigcup_{n \in I} A_n = \{\omega \in \Omega : \text{il existe } n \in I \text{ tel que } \omega \in A_n\}$,
- ▷ $\bigcap_{n \in I} A_n = \{\omega \in \Omega : \omega \in A_n \text{ pour tout } n \in I\}$.

Exemple 1.1.7. Considérons l'expérience qui consiste à lancer 5 fois un dé. On a donc

$$\Omega = \{1, 2, 3, 4, 5, 6\}^5.$$

Considérons pour tout $n \in \{1, \dots, 5\}$, l'événement

$$A_n = \text{le } n^{\text{ème}} \text{ résultat est pair}.$$

Alors

- ▷ $\bigcap_{n=1}^5 A_n$ est l'événement "tous les résultats sont pairs"
- ▷ $\bigcup_{n=1}^5 A_n$ est l'événement "au moins un résultat est pair"

Exemple 1.1.8. On lance une pièce infiniment souvent. On a $\Omega = \{\text{PILE}, \text{FACE}\}^{\mathbb{N}}$. On considère les événements

$$A_n = \{\omega = (\omega_1, \omega_2, \dots) : \omega_n = \text{PILE}\}$$

i.e. A_n est l'événement "le $n^{\text{ème}}$ lancer est PILE". Alors, $\bigcup_{n=1}^N A_n$ correspond à l'événement "il y a au moins un PILE dans les N premiers lancers" et $\bigcap_{n=1}^N A_n$ correspond à l'événement "chacun des N premiers lancers est PILE". On peut également imaginer des événements

limites en considérant par exemple

$$\bigcup_{n=1}^{\infty} A_n \quad \text{qui correspond à "il y a un PILE à un moment",}$$

$$\bigcup_{n=N}^{\infty} A_n \quad \text{qui correspond à "il y a un PILE après le } N^{\text{ème}} \text{ lancer"}$$

ou

$$\bigcap_{n=N}^{\infty} A_n \quad \text{qui correspond à "tous les lancers après le } N^{\text{ème}} \text{ sont des PILE".}$$

Enfin, rappelons les deux définitions suivantes.

Définition 1.1.9. Soit Ω un ensemble. Soient A, B deux sous-ensembles de Ω . On dit que A et B sont *disjoints* si $A \cap B = \emptyset$. On parle également d'événements *incompatibles*.

Définition 1.1.10. L'ensemble des parties de Ω , noté $\mathcal{P}(\Omega)$, est l'ensemble des sous-ensembles de Ω .

1.2 Approche fréquentiste des probabilités

Notre objectif est d'assigner à chaque événement A un nombre entre 0 et 1, que l'on notera $\mathbb{P}(A)$ et qui mesure la confiance que l'on a en le fait que l'événement A va se produire.

Exemple 1.2.1.

- ▷ Si on lance deux dés, "*Le deuxième lancer est pile*" a intuitivement une probabilité $1/2$.
- ▷ Si on lance deux dés, afin de donner la probabilité de l'événement "*La somme des deux dés est égale à 8*", il suffit également de passer en revue tous les cas.
- ▷ Dans un tournoi de tennis pro, on sent que la probabilité de l'événement "*Le gagnant appartient au Big Four*" est proche de 1, mais il faudrait des données pour estimer cela plus précisément.
- ▷ Si on s'intéresse au passage du bus, il faut des données pour pouvoir estimer la probabilité de l'événement "*Le bus arrive dans 10 minutes maximum*".

Une manière de définir la probabilité d'un événement est d'utiliser une méthode statistique. On suppose que l'expérience d'intérêt est répétée un très grand nombre de fois sous les mêmes conditions. Si on s'intéresse à l'événement A , on note N_n le nombre de fois que l'événement s'est

réalisé lors des n premières expériences. On en déduit la fréquence de réalisation de A lors des n premières expériences, donnée par

$$f_n = \frac{N_n}{n}.$$

On définit la probabilité de A comme la limite de la fréquence f_n lorsque le nombre d'expériences n tend vers l'infini.

Exemple 1.2.2. On lance un dé et on considère l'événement $A = \text{“Le résultat est 1”}$. Après 15 lancers, on a

Résultat du dé	1	2	3	4	5	6
Effectif	2	4	3	3	1	2
Fréquence	0,133	0,266	0,2	0,2	0,066	0,133

et donc $f_{15} = 0,133$, et après 400 lancers

Résultat du dé	1	2	3	4	5	6
Effectif	67	56	68	62	80	67
Fréquence	0,1675	0,14	0,17	0,155	0,2	0,1675

et donc $f_{400} = 0,1675$ qui se rapproche de la valeur intuitive $1/6 \approx 0,166$, mais ce n'est pas le cas pour tous les résultats !

Même si cette définition est très intuitive (vous devez la garder en tête!), certaines questions apparaissent :

- ▷ Comment être sûr que la fréquence va converger ?
- ▷ Est-ce que la fréquence va converger rapidement ?
- ▷ Comment être sûr que cette limite restera la même si on recommence l'expérience un grand nombre de fois ?

Remarquons de plus que dans le cas particulier du dé, on a un modèle probabiliste en tête, ce qui n'est pas toujours le cas...

1.3 Mesures de probabilité

L'objectif de la théorie des probabilités est de donner un cadre formel qui permet d'écrire rigoureusement toutes les questions liées au hasard. Le point de départ est une construction abstraite de la notion de *probabilité* comme une *mesure mathématique* qui attribue à chaque événement A une probabilité $\mathbb{P}(A)$ qui représente une estimation raisonnable de la vraisemblance de la réalisation de l'événement A .

Pour qu'une telle mesure soit pertinente, elle doit naturellement vérifier les propriétés suivantes :

- ▷ la probabilité d'un événement est un nombre entre 0 et 1,
- ▷ la probabilité d'un événement certain doit être 1,
- ▷ la probabilité qu'un événement n'ait pas lieu doit être 1 moins la probabilité qu'il ait lieu,
- ▷ si deux événements ne peuvent pas avoir lieu en même temps, la probabilité que l'un ou l'autre se réalise doit être la somme de leurs probabilités.

On peut démontrer que, de manière contre-intuitive, il existe des sous-ensembles pour lesquels il est impossible de définir une telle mesure (il y a même des sous-ensembles de la droite réelle dont on ne peut pas définir la longueur). Il est donc crucial, dans un premier temps, de bien définir la classe d'ensembles que l'on souhaite mesurer, autrement dit la classe des événements. Puisque toute opération ensembliste peut créer de nouveaux événements, on va s'assurer qu'un "bon" système d'événements est fermé sous ces opérations ensemblistes.

Définition 1.3.1. Soit Ω un ensemble. Une σ -algèbre $\mathcal{F}$ sur Ω est une partie de $\mathcal{P}(\Omega)$ qui satisfait les conditions suivantes :

1. $\Omega \in \mathcal{F}$,
2. Stabilité par passage au complémentaire : si $A \in \mathcal{F}$, alors $A^c \in \mathcal{F}$,
3. Stabilité par union dénombrable : si $A_n \in \mathcal{F}$ pour tout $n \in \mathbb{N}$, alors $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$.

Évidemment, les deux premières propriétés impliquent que $\emptyset \in \mathcal{F}$. Cela permet d'obtenir la stabilité de $\mathcal{F}$ par union *finie* : Si $A_1, \dots, A_n \in \mathcal{F}$, alors $A_1 \cup \dots \cup A_n \in \mathcal{F}$. De plus, puisque

$$\left(\bigcap_n A_n \right)^c = \bigcup_n A_n^c,$$

les deux dernières propriétés impliquent que $\mathcal{F}$ est stable par intersection finie et dénombrable. Enfin, si $A, B \in \mathcal{F}$, alors $A \setminus B = A \cap B^c \in \mathcal{F}$.

À retenir pour la suite. Pour les applications, l'idée essentielle est que $\mathcal{F}$ rassemble les événements auxquels on souhaite attribuer une probabilité et qu'elle est stable par les opérations ensemblistes usuelles. Les détails techniques des preuves de fermeture et de continuité des mesures peuvent être lus comme un complément mathématique ; ils servent surtout à garantir la cohérence du formalisme.

Rappelez-vous que

- ▷ Ω correspond à tous les résultats possibles d'une expérience,
- ▷ $\mathcal{F}$ est un ensemble d'événements liés à cette expérience, il représente les questions que l'on peut se poser,
- ▷ si on peut se demander si un événement se réalise, on peut se demander si l'événement ne se réalise pas (traduit 2),

▷ si on peut se demander si des événements se réalisent, on peut se demander si l'un ou l'autre se réalise (traduit 3).

Exemple 1.3.2. Si on lance le dé, on sait que

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Si $k \in \{1, \dots, 6\}$, on peut se demander si le résultat est k , mais aussi si le résultat est pair, si le résultat est impair, si le résultat appartient à $\{2, 4, 5\}$... Dans ce cas, on a $\mathcal{F} = \mathcal{P}(\Omega)$.

Exemple 1.3.3. Si on attend le bus,

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

On peut se demander si le bus arrivera dans moins de 5 minutes, dans plus de 15 minutes, entre 7 et 10 minutes... Dans ce cas, on veut que notre σ -algèbre contienne tous les intervalles

$$[a, b], [a, b[,]a, b],]a, b[$$

de $[0, +\infty[$. La plus petite σ -algèbre qui contient ces intervalles s'appelle la σ -algèbre de Borel sur $[0, +\infty[$. Elle est très riche (et très importante!) mais moins grande que $\mathcal{P}([0, +\infty[)$.

Définition 1.3.4 (Axiomes de Kolmogorov). Soit $\mathcal{F}$ une σ -algèbre sur un ensemble Ω . Une *mesure de probabilité* est une application

$$\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$$

telle que

1. l'univers a une probabilité 1 : $\mathbb{P}(\Omega) = 1$,
2. σ -additivité : pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'événements deux-à-deux disjoints on a

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

Dans ce cas, on dit que $(\Omega, \mathcal{F}, \mathbb{P})$ est un *espace probabilisé*.

Proposition 1.3.5 (Règles de calcul). Soient $A, B \in \mathcal{F}$ deux événements.

1. **Vide** : $\mathbb{P}(\emptyset) = 0$

2. **Union finie** : Si $A_0, \dots, A_n \in \mathcal{F}$ sont deux-à-deux disjoints, alors

$$\mathbb{P}(A_0 \cup \dots \cup A_n) = \mathbb{P}(A_0) + \dots + \mathbb{P}(A_n)$$

3. **Complémentaire** : $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

4. **Monotonie** : si $A \subseteq B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$ et $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$

5. **Union** : $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

6. **Inégalité de Boole** : Si $(A_n)_{n \in \mathbb{N}}$ est une suite de $\mathcal{F}$, alors

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

Démonstration. 1. Considérons la suite $(A_n)_{n \in \mathbb{N}}$ de $\mathcal{F}$ définie par $A_0 = \Omega$ et $A_n = \emptyset$ pour tout $n \geq 1$. Alors, les ensembles A_n sont deux à deux disjoints et leur union donne Ω . Par σ -additivité, on obtient

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = \mathbb{P}(\Omega) + \sum_{n=1}^{+\infty} \mathbb{P}(\emptyset) = 1 + \sum_{n=1}^{+\infty} \mathbb{P}(\emptyset)$$

ce qui implique que $\mathbb{P}(\emptyset) = 0$.

2. Il suffit de considérer la suite $(B_j)_{j \in \mathbb{N}}$ d'ensembles de $\mathcal{F}$ définie par $B_j = A_j$ si $j \leq n$ et $B_j = \emptyset$ si $j > n$. La σ -additivité et le premier point donnent directement la conclusion.

3. Puisque les événements A et A^c sont disjoints, on a

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

De plus, $A \cup A^c = \Omega$ et donc $\mathbb{P}(A \cup A^c) = 1$. On en tire que

$$1 = \mathbb{P}(A) + \mathbb{P}(A^c),$$

d'où le résultat.

4. On remarque que les événements A et $B \setminus A$ sont disjoints et que $A \cup (B \setminus A) = B$. Par conséquent, on a

$$\mathbb{P}(B) = \mathbb{P}(A \cup (B \setminus A)) = \mathbb{P}(A) + \underbrace{\mathbb{P}(B \setminus A)}_{\geq 0} \geq \mathbb{P}(A).$$

La relation précédente donne également

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

5. On peut écrire $A \cup B = A \cup (B \setminus (A \cap B))$, et les événements A et $B \setminus (A \cap B)$ sont disjoints.

On obtient donc

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

par le point 4.

6. Considérons la suite $(B_n)_{n \in \mathbb{N}}$ de $\mathcal{F}$ définie par $B_0 = A_0$ et

$$B_n = A_n \setminus \bigcup_{j=0}^{n-1} A_j$$

pour tout $n \geq 1$. Par construction, on a $B_n \subset A_n$ pour tout $n \in \mathbb{N}$ et les ensembles B_n sont deux à deux disjoints. De plus,

$$\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} B_n.$$

Par conséquent, la σ -additivité permet d'écrire

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(B_n) \leq \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$$

par monotonie. ■

Proposition 1.3.6 (Continuité).

1. Si $(A_n)_{n \in \mathbb{N}}$ est une suite croissante de $\mathcal{F}$, c'est-à-dire si $A_n \subseteq A_{n+1}$ pour tout $n \in \mathbb{N}$, alors

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n).$$

2. Si $(A_n)_{n \in \mathbb{N}}$ est une suite décroissante de $\mathcal{F}$, c'est-à-dire si $A_{n+1} \subseteq A_n$ pour tout $n \in \mathbb{N}$, alors

$$\mathbb{P}\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n).$$

Démonstration. 1. Soit $A = \bigcup_{n \in \mathbb{N}} A_n$. Posons également $B_0 = A_0$ et pour tout $n \geq 1$, $B_n = A_n \setminus A_{n-1}$. Les ensembles construits appartiennent à $\mathcal{F}$ et sont deux à deux disjoints. De plus, pour tout $n \in \mathbb{N}$, on a $A_n = B_0 \cup \dots \cup B_n$ et donc $A = \bigcup_{j \in \mathbb{N}} B_j$. On en tire que

$$\mathbb{P}(A_n) = \sum_{j=0}^n \mathbb{P}(B_j) \quad \text{et} \quad \mathbb{P}(A) = \sum_{j=0}^{+\infty} \mathbb{P}(B_j),$$

d'où

$$\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \sum_{j=0}^{+\infty} \mathbb{P}(B_j) = \mathbb{P}(A).$$

2. Pour tout $n \in \mathbb{N}$, posons $B_n = A_0 \setminus A_n$. Les éléments ainsi construits appartiennent à $\mathcal{F}$ et la

suite $(B_n)_{n \in \mathbb{N}}$ est croissante. Vu le premier point, on a

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(B_n) = \mathbb{P}(A_0) - \lim_{n \rightarrow +\infty} \mathbb{P}(A_n).$$

D'autre part, on a

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \mathbb{P}\left(A_0 \setminus \bigcap_{n \in \mathbb{N}} A_n\right) = \mathbb{P}(A_0) - \mathbb{P}\left(\bigcap_{n \in \mathbb{N}} A_n\right)$$

et la conclusion s'ensuit. ■

Les propriétés que nous venons d'établir sont très importantes. Elles permettent notamment de montrer qu'il peut exister des événements ayant une probabilité nulle sans pour autant être impossibles.

Définition 1.3.7. Soit $A \in \mathcal{F}$ un événement. On dit que A est *stochastiquement certain* si $\mathbb{P}(A) = 1$. On dit que A est *de probabilité nulle* si $\mathbb{P}(A) = 0$.

Exemple 1.3.8. Reprenons l'exemple 1.1.8 d'un lancer infini de pièce. On considère l'événement A_n défini par “les n premiers lancers sont PILE”. Alors,

$$A = \bigcap_{n=1}^{+\infty} A_n$$

correspond à l'événement “tous les résultats sont PILE”. Puisque la suite $(A_n)_{n \geq 1}$ est décroissante, on a

$$\mathbb{P}(A) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \lim_{n \rightarrow +\infty} \left(\frac{1}{2}\right)^n = 0.$$

Remarquons néanmoins que $A \neq \emptyset$.

1.4 Mesures de probabilités dans un cas équiprobable

Le cas du lancer du dé représente un cas simple car toutes les issues sont équiprobables : on a 1 chance sur 6 d'obtenir chaque résultat. On souhaite définir une mesure de probabilité qui traduit cette situation.

On considère une expérience dont les résultats forment un ensemble fini (on peut donc les numéroter)

$$\Omega = \{\omega_1, \dots, \omega_N\} \quad (\text{et } \mathcal{F} = \mathcal{P}(\Omega)).$$

On souhaite construire $\mathbb{P}$ de manière à ce que chaque résultat ait la même probabilité de se produire, c'est-à-dire

$$\mathbb{P}(\{\omega_1\}) = \mathbb{P}(\{\omega_2\}) = \dots = \mathbb{P}(\{\omega_N\}).$$

Notons p cette probabilité commune. Pour avoir une mesure de probabilité, on doit imposer

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(\{\omega_1, \dots, \omega_N\}) = \underbrace{\mathbb{P}(\{\omega_1\})}_{=p} + \underbrace{\mathbb{P}(\{\omega_2\})}_{=p} + \dots + \underbrace{\mathbb{P}(\{\omega_N\})}_{=p} = Np$$

et on obtient

$$p = \frac{1}{N}.$$

Si on considère un événement A , on peut l'identifier à l'ensemble des résultats pour lesquels il se réalise. S'il y a k résultats possibles pour A , on peut écrire

$$A = \{\omega_{i_1}, \dots, \omega_{i_k}\}$$

pour certains $i_1, \dots, i_k \in \{1, \dots, N\}$. On doit donc poser

$$\mathbb{P}(A) = \mathbb{P}(\{\omega_{i_1}\} \cup \dots \cup \{\omega_{i_k}\}) = \underbrace{\mathbb{P}(\{\omega_{i_1}\})}_{=\frac{1}{N}} + \dots + \underbrace{\mathbb{P}(\{\omega_{i_k}\})}_{=\frac{1}{N}} = \frac{k}{N},$$

ce qu'on écrit souvent sous la forme

$$\mathbb{P}(A) = \frac{\text{Nombre de cas favorables pour } A}{\text{Nombre de cas possibles}} = \frac{\#A}{\#\Omega}.$$

On vérifie facilement qu'il s'agit d'une mesure de probabilité puisque si $A_1, \dots, A_n$ sont disjoints, alors $\#(A_1 \cup \dots \cup A_n) = \#A_1 + \dots + \#A_n$.

Exemple 1.4.1. Si on lance deux dés distinguables, on a les résultats possibles suivants :

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

L'événement A donné par “La somme des deux dés est égale à 8” correspond au sous-ensemble de résultats

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

Il y a donc 5 résultats possibles pour lesquels la somme des deux dés est égale à 6, parmi les 36 résultats possibles. Donc

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

Il peut donc s'avérer utile de savoir estimer le nombre d'éléments d'un ensemble. Cette branche des mathématiques s'appelle *l'analyse combinatoire* (ou le *dénombrement*). Le résultat élémentaire

suisant sert de base aux différentes méthodes de dénombrement.

Proposition 1.4.2 (Règle de multiplication). *Soient E_1, E_2 deux expériences. On regarde la succession des deux expériences. Supposons qu'il y a n_1 résultats pour E_1 et pour chaque résultat de E_1 , n_2 résultats pour E_2 . Alors le nombre de résultats de la succession est donné par $n_1 n_2$.*

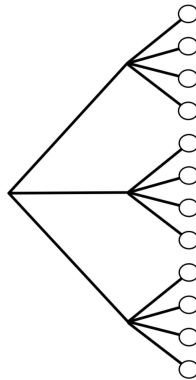


FIGURE 1.1 – Il y a 3 issues pour la première expérience et pour chacune de ces issues, il y a 4 issues possibles. Au total, on a 12 résultats possibles.

On peut bien sûr généraliser à un nombre k d'expériences successives $E_1, \dots, E_k$. Supposons qu'il y a n_1 résultats pour E_1 et pour chaque résultat de E_j où $j \in \{1, \dots, k-1\}$, n_{j+1} résultats pour E_{j+1} . Alors le nombre de résultats de la succession est donné par $n_1 n_2 \dots n_k$.

Exemple 1.4.3 (Sous-ensembles). Soit B un ensemble fini de n éléments. Alors $\#\mathcal{P}(B) = 2^n$ puisque, pour chaque élément de B , on peut décider de le choisir ou non.

Exemple 1.4.4 (Tirage avec remise – ordre important). On suppose avoir un ensemble de n objets distincts. On choisit k fois un élément de cet ensemble, en le remettant à chaque fois. Comme on a n possibilités à chaque tirage, il y a n^k résultats possibles.

Génération de mots de passe : On suppose travailler sur un alphabet de 26 lettres. Le nombre de mots de passe de 10 lettres que l'on peut construire est 26^{10} .

Exemple 1.4.5 (Tirage sans remise – ordre important). On suppose avoir un ensemble de n objets distincts. On choisit k fois un élément de cet ensemble, sans le remettre ensuite dans l'ensemble. Ainsi,

- ▷ lors du premier tirage, on a n possibilités,
- ▷ lors du deuxième tirage, on a $n - 1$ possibilités,
- ▷ ...

▷ lors du $k^{\text{ème}}$ tirage, on a $n - k + 1$ possibilités.

Au total, on a donc

$$n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!} =: A_n^k$$

possibilités.

Plan de salle : Une salle de spectacle contient 150 places. De plus, 135 tickets ont été vendus pour la représentation de ce soir. On peut asseoir les spectateurs de A_{150}^{135} manières différentes.

Exemple 1.4.6 (Permutations). Une permutation des éléments d'un ensemble de n éléments est un arrangement des éléments de cet ensemble dans un certain ordre. Il s'agit donc d'un tirage de n éléments parmi n sans remise (dont l'ordre est important). Il y a donc $n!$ permutations des n éléments.

Mots : On peut construire exactement $4!$ mots de 4 lettres sur l'alphabet $\{a, b, c, d\}$ en utilisant une et une seule fois chaque lettre.

Exemple 1.4.7 (Tirage sans remise – ordre sans importance). On suppose qu'on tire à nouveau k éléments parmi n sans remise, mais l'ordre n'a plus d'importance. Lorsque l'on effectue un tirage sans remise où l'ordre est important, on sait qu'apparaissent $k!$ permutations des mêmes objets, qui donnent ici lieu à une seule solution. Ainsi, si on se donne un ensemble de cardinalité n , il possède

$$\frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!} =: C_n^k$$

sous-ensembles de k éléments. On retrouve le *coefficient binomial* C_n^k .

Plan de salle bis : Une salle de spectacle contient 150 places et 135 tickets ont été vendus. Il y a C_{150}^{135} manières de choisir les sièges qui seront occupés.

Exemple 1.4.8 (Problème de Bose-Einstein). Combien de possibilités y a-t-il de mettre k particules indistinguables dans n boîtes? Chaque configuration correspond à une suite de $\bullet$ et de $|$. Par exemple, pour $k = 3$ et $n = 2$, on a les 4 configurations suivantes :

$$\begin{array}{cc} \bullet \bullet \bullet | & \bullet \bullet | \bullet \\ \bullet | \bullet \bullet & | \bullet \bullet \bullet \end{array}$$

Il suffit donc de choisir où mettre les k $\bullet$ parmi les $k + n - 1$ positions possibles (on doit

placer $n - 1$ parois). Il y a donc

$$C_{k+n-1}^k = \frac{(k+n-1)!}{k!(n-1)!}$$

possibilités.

Exemple 1.4.9 (Anniversaires). Terminons par un exemple classique mais surprenant de calcul de probabilités basé sur l'analyse combinatoire. On s'intéresse au problème suivant :

“Il y a k personnes dans une pièce. Quelle est la probabilité que deux personnes au moins fêtent leur anniversaire le même jour ?”

On suppose qu'il y a 365 dates d'anniversaire possibles et qu'elles sont *équiprobables*. Bien sûr, on suppose que $k \leq 365$. On suppose également que les dates d'anniversaire sont indépendantes les unes des autres (on verra cette notion de manière rigoureuse dans le chapitre 2) : par exemple, il n'y a pas de jumeaux.

Pour calculer cette probabilité, il est plus simple de passer à l'événement complémentaire :

$$\begin{aligned} & \mathbb{P}(\text{deux personnes au moins fêtent leur anniversaire le même jour}) \\ &= 1 - \mathbb{P}(\text{toutes les dates d'anniversaires sont différentes}). \end{aligned}$$

On va calculer la probabilité que toutes les dates d'anniversaires soient différentes via la formule

$$\frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}}.$$

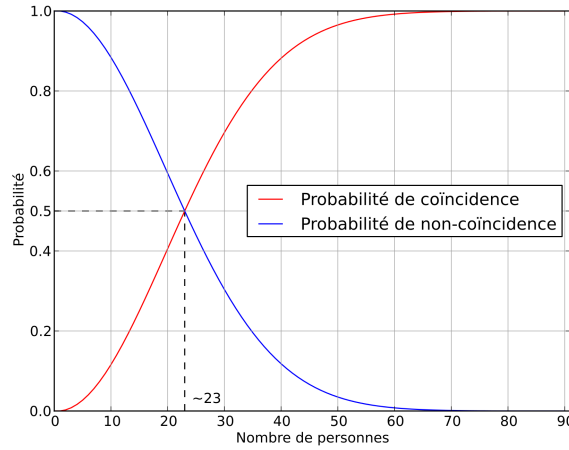
Dans notre situation, on a

- ▷ Nombre de cas possibles : il y a 365^k possibilités (tirage avec remise - ordre important)
- ▷ Nombre de cas favorables : il y a $A_{365}^k = \frac{365!}{(365-k)!}$ possibilités (tirage sans remise - ordre important).

Ainsi

$$\mathbb{P}(\text{2 personnes au moins fêtent leur anniversaire le même jour}) = 1 - \frac{365!}{(365-k)!365^k}.$$

Pour $k = 23$, on peut calculer que cette probabilité est déjà proche de $1/2$. Pour $k = 57$, la probabilité est supérieure à $99/100$.



1.5 Mesures de probabilité sur un ensemble fini ou dénombrable

Supposons que $\Omega = \{\omega_j : j \in J\}$ avec $J \subset \mathbb{N}$ (fini ou infini). Si $(p_j)_{j \in J}$ est une suite de nombres de $[0, 1]$ tels que

$$\sum_{j \in J} p_j = 1,$$

alors on peut définir une mesure de probabilité sur $(\Omega, \mathcal{P}(\Omega))$ telle que

$$\mathbb{P}(\{\omega_j\}) = p_j \quad \text{pour tout } j \in J.$$

Il suffit de poser

$$\mathbb{P}(A) = \sum_{j \in J: \omega_j \in A} p_j$$

Exemple 1.5.1.

- ▷ Si $\#J = N$, alors $p_j = \frac{1}{N}$ correspond au cas équiprobable.
- ▷ Si $\Omega = \{0, \dots, n\}$ et $p \in]0, 1[$, la mesure binomiale correspond à la suite $p_j = C_n^j p^j (1-p)^{n-j}$.
- ▷ Si $\Omega = \mathbb{N}$ et $p \in]0, 1[$, la mesure géométrique correspond à $p_j = (1-p)^{j-1} p$.
- ▷ Si $\Omega = \mathbb{N}_0$, la mesure de Poisson de paramètre $\lambda > 0$ correspond à $p_j = e^{-\lambda} \frac{\lambda^j}{j!}$.

1.6 Mesures de probabilité sur $\mathbb{R}$

Soit $\Omega = \mathbb{R}$ sur lequel on met la σ -algèbre de Borel $\mathcal{B}$, qui est la plus petite σ -algèbre qui contient tous les intervalles. Si f est une fonction positive intégrable telle que

$$\int_{\mathbb{R}} f(x) dx = 1,$$

alors on peut définir une mesure de probabilité sur $(\mathbb{R}, \mathcal{B})$ en posant

$$\mathbb{P}(A) = \int_A f(x) dx$$

Exemple 1.6.1.

- ▷ $f(x) = \mathbb{1}_{[0,1]}$ correspond à un tirage équiprobable entre 0 et 1.
- ▷ $f(x) = e^{-x} \mathbb{1}_{[0,+\infty[}(x)$ est une densité exponentielle.
- ▷ $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ est une densité gaussienne.

2 Probabilités conditionnelles et indépendance

Les probabilités conditionnelles interviennent lorsque la connaissance de la réalisation d'un événement permet de modifier la confiance que l'on a en le fait qu'un autre événement se produise. Dans ce chapitre, suivant cette idée, nous introduisons naturellement les probabilités conditionnelles et montrons qu'elles donnent des nouvelles mesures de probabilité, selon les axiomes de Kolmogorov. Nous nous intéressons ensuite à l'indépendance d'événements : ce sont des événements pour lesquels la connaissance de la réalisation d'un événement n'impacte pas la probabilité de réalisation du second.

2.1 Probabilités conditionnelles

Soient Ω un ensemble, $\mathcal{F}$ la σ -algèbre des événements sur Ω , et $\mathbb{P}$ une mesure de probabilité. Supposons que l'on sait qu'un événement B s'est produit. Si A est un autre événement, on cherche à calculer la probabilité de A avec l'information supplémentaire dont on dispose. On note

$$\mathbb{P}(A|B) = \text{la probabilité de } A \text{ sachant } B.$$

Exemple 2.1.1. Lors du lancer d'un dé, on sait que le résultat est un nombre pair, autrement dit $B = \{2, 4, 6\}$. Cette nouvelle information augmente notre confiance en le fait que l'événement $A = \{2\}$ se produise, et on estime intuitivement à présent sa probabilité à $1/3$ à la place de $1/6$. Donc

$$\mathbb{P}(A) = \frac{1}{6} \quad \text{et} \quad \mathbb{P}(A|B) = \frac{1}{3}.$$

Avant de définir proprement $\mathbb{P}(A|B)$, regardons quelques cas.

- ▷ Si $A \cap B = \emptyset$ (événements *incompatibles*), alors il est clair que A ne peut pas se réaliser en même temps que B . Donc $\mathbb{P}(A|B) = 0$.

Exemple du dé : $B = \{2, 4, 6\}$ et $A = \{1, 3\}$.

- ▷ Si $A \cap B = B$, c'est-à-dire si $B \subseteq A$, alors on est *certain* que A va se réaliser. Donc $\mathbb{P}(A|B) = 1$.

Exemple du dé : $B = \{2, 4, 6\}$ et $A = \{2, 4, 5, 6\}$.

- ▷ Si $A \cap B \neq \emptyset$, alors l'événement A est réalisable, mais seulement les résultats de A qui sont dans B .

Exemple du dé : $B = \{2, 4, 6\}$ et $A = \{2, 3\}$.

Tout se passe comme si on avait *restreint* l'univers Ω à B . Etant donné un événement A , on ne s'intéresse plus qu'à $A \cap B$. On va donc regarder $\mathbb{P}(A \cap B)$. De plus, puisqu'on sait que B est réalisé, on doit avoir

$$\mathbb{P}(B|B) = 1.$$

On va donc normaliser $\mathbb{P}(A \cap B)$ par $\mathbb{P}(B)$.

Définition 2.1.2. Supposons que $\mathbb{P}(B) \neq 0$. La *probabilité conditionnelle de A sachant B* est définie par

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

Remarque 2.1.3. Remarquons en particulier que

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

pour tout A, B avec $\mathbb{P}(B) \neq 0$.

Remarque 2.1.4. Dans le cas équiprobable, cette formule devient

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

Exemple 2.1.5. Supposons que parmi 200 étudiants, 60 sont des filles en math, 50 sont des filles en biologie, 60 sont des garçons en math et 30 sont des garçons en biologie.

	Math	Bio	
F	60	50	110
G	60	30	90
	120	80	200

On choisit un étudiant au hasard. Dans ce tableau, on lit que

$$\mathbb{P}(F \cap \text{Math}) = \frac{60}{200} = 0.3 \quad \text{et} \quad \mathbb{P}(\text{Math}) = \frac{120}{200} = 0.6$$

La probabilité de choisir une fille sachant qu'on a choisi un mathématicien est donné par la proportion de filles parmi les mathématiciens :

$$\mathbb{P}(F|\text{Math}) = \frac{60}{120} = 0.5$$

Les probabilités conditionnelles nous donnent plus d'informations que les probabilités simples. En effet, dans l'exemple précédent, on remarque que

$$\underbrace{\mathbb{P}(F|\text{Math})}_{=0.5} < \underbrace{\mathbb{P}(F)}_{=0.55} < \underbrace{\mathbb{P}(F|\text{Bio})}_{=0.625}.$$

La proportion de filles dans la population des biologistes est donc plus importante que la proportion de filles dans la population, et la proportion de filles dans la population des mathématiciens est moins importante.

On verra que les probabilités conditionnelles permettent de retrouver les probabilités simples, via la formule des probabilités totales.

Proposition 2.1.6. *Soit B un événement tel que $\mathbb{P}(B) \neq 0$. L'application*

$$\mathbb{P}(\cdot | B) : \mathcal{F} \rightarrow [0, 1]$$

est une mesure de probabilité.

Démonstration. Si $A \in \mathcal{F}$, on a

$$0 \leq \mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \leq \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$$

par monotonie de $\mathbb{P}$ puisque $A \cap B \subseteq B$. Par conséquent, $\mathbb{P}(\cdot | B)$ est à valeurs dans $[0, 1]$.

De plus, on a $\mathbb{P}(\Omega \cap B) = \mathbb{P}(B)$, d'où $\mathbb{P}(\Omega | B) = 1$.

Enfin, si les événements A_n , $n \in \mathbb{N}$, sont deux à deux disjoints, on a

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n | B\right) &= \frac{\mathbb{P}((\bigcup_{n \in \mathbb{N}} A_n) \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\bigcup_{n \in \mathbb{N}} (A_n \cap B))}{\mathbb{P}(B)} \\ &= \frac{\sum_{n \in \mathbb{N}} \mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} \\ &= \sum_{n \in \mathbb{N}} \mathbb{P}(A_n | B) \end{aligned}$$

puisque les événements $A_n \cap B$ sont deux à deux disjoints. ■

Par conséquent, on peut appliquer à $\mathbb{P}(\cdot | B)$ toutes les propriétés des probabilités. En particulier,

- ▷ $\mathbb{P}(A^c | B) = 1 - \mathbb{P}(A | B)$,
- ▷ $\mathbb{P}(\emptyset | B) = 0$,
- ▷ si $A \subseteq C$, alors $\mathbb{P}(A | B) \leq \mathbb{P}(C | B)$ et $\mathbb{P}(C \setminus A | B) = \mathbb{P}(C | B) - \mathbb{P}(A | B)$,
- ▷ $\mathbb{P}(A \cup C | B) = \mathbb{P}(A | B) + \mathbb{P}(C | B) - \mathbb{P}(A \cap C | B)$

pour tous $A, B, C \in \mathcal{F}$ avec $\mathbb{P}(B) \neq 0$.

2.2 Probabilités composées

On peut bien entendu réécrire la définition de la probabilité conditionnelle sous la forme

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

En utilisant la définition de la probabilité conditionnelle, on obtient directement la règle de multiplication suivante. C'est cette règle de multiplication qui permet de justifier l'utilisation d'arbres pondérés pour calculer des probabilités.

Proposition 2.2.1 (Règle de multiplication). *Pour tous $A_1, \dots, A_n \in \mathcal{F}$ tels que $\mathbb{P}(A_1 \cap \dots \cap A_n) > 0$, on a*

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

Démonstration. On remarque que le membre de droite est égal à

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

qui se simplifie pour laisser $\mathbb{P}(A_n \cap A_{n-1} \cap \dots \cap A_1)$. ■

Exemple 2.2.2. On tire 3 cartes sans remise d'un jeu de 52. Quelle est la probabilité d'avoir 3 as ? Avec des notations évidentes, on a

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

2.3 Loi des probabilités totales

La loi des probabilités totales lie la probabilité d'un événement avec ses probabilités conditionnelles. On l'utilise lorsqu'il est plus facile de calculer les probabilités d'un événement en ayant des informations supplémentaires. Soient A et B deux événements. Alors, les événements $A \cap B$ et $A \cap B^c$ sont disjoints et on a $A = (A \cap B) \cup (A \cap B^c)$. Par conséquent, si $\mathbb{P}(B) \in]0, 1[$, on a

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

qui est une moyenne pondérée de la probabilité de A sachant que B s'est produit et celle sachant que B ne s'est pas produit, les poids respectifs étant donnés par la probabilité que B se produise ou non.

Exemple 2.3.1. On tire 2 cartes sans remise dans un jeu de 52 cartes. Quelle est la probabilité que la deuxième carte soit un coeur ? On se rend compte intuitivement qu'il y a moins de chance que la deuxième carte soit un coeur si la première était déjà un coeur (puisqu'il en reste alors moins). On considère donc les événements

- ▷ A = la deuxième carte est un coeur
- ▷ B = la première carte est un coeur.

On calcule

$$\mathbb{P}(B) = \frac{13}{52}, \quad \mathbb{P}(A|B) = \frac{12}{51}, \quad \mathbb{P}(A|B^c) = \frac{13}{51}.$$

Au total,

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c) = \frac{12}{51} \cdot \frac{13}{52} + \frac{13}{51} \cdot \left(1 - \frac{13}{52}\right) = \frac{13}{52} = \frac{1}{4}.$$

Intuitivement c'était clair car on doit avoir la même probabilité que la deuxième carte tirée soit un coeur, un carreau, un trèfle ou un pique.

Cette formule se généralise au cas de plusieurs événements deux à deux disjoints dont l'union donne l'univers tout entier.

Définition 2.3.2. On dit qu'une suite (finie ou non) d'événements $(B_j)_{j \in J}$ forme un *système complet d'événements* ou une *partition* de Ω si

- ▷ $B_j \cap B_k = \emptyset$ si $j \neq k$
- ▷ $\bigcup_{j \in J} B_j = \Omega$.

Proposition 2.3.3 (Loi des probabilités totales). Soient $(B_j)_{j \in J}$ un système complet d'événements de Ω tels que $\mathbb{P}(B_j) > 0$ pour tout $j \in J$. Alors, pour tout $A \in \mathcal{F}$, on a

$$\mathbb{P}(A) = \sum_{j \in J} \mathbb{P}(A|B_j)\mathbb{P}(B_j).$$

Démonstration. Tout événement A peut s'écrire sous la forme de l'union disjointe des événements $A \cap B_j$, $j \in J$. Par définition d'une mesure de probabilité, on a donc

$$\mathbb{P}(A) = \sum_{j \in J} \mathbb{P}(A \cap B_j)$$

et on conclut en remarquant que $\mathbb{P}(A \cap B_j) = \mathbb{P}(A|B_j)\mathbb{P}(B_j)$ pour tout $j \in J$. ■

Remarque 2.3.4. Si certains B_j sont de probabilité nulle, on les omet simplement de la somme : la formule s'applique aux indices tels que $\mathbb{P}(B_j) > 0$. Plus généralement, il suffit que les B_j soient deux à deux disjoints et que $\mathbb{P}(\cup_{j \in J} B_j) = 1$.

2.4 Formule de Bayes

La formule de Bayes permet de calculer $\mathbb{P}(A|B)$ à partir de $\mathbb{P}(B|A)$.

Proposition 2.4.1 (Formule de Bayes). *Soient A et B deux événements de probabilités non-nulles. On a*

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

Plus généralement, si $(B_j)_{j \in J}$ est un système complet d'événements de Ω tels que $\mathbb{P}(B_j) > 0$ pour tout $j \in J$ et si A est un événement tel que $\mathbb{P}(A) > 0$, alors

$$\mathbb{P}(B_k | A) = \frac{\mathbb{P}(B_k)}{\sum_{j \in J} \mathbb{P}(B_j) \mathbb{P}(A | B_j)} \mathbb{P}(A | B_k). \quad (2.1)$$

Démonstration. On a $\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A)$. Ainsi, on obtient

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

d'où la formule annoncée. La deuxième formule de Bayes combine la première avec la loi des probabilités totales. ■

Dans la formule de Bayes, les $\mathbb{P}(B_j)$ sont les *probabilités a priori*, tandis que les $\mathbb{P}(A|B_j)$ sont les *vraisemblances*. La formule permet de calculer les *probabilités a posteriori* $\mathbb{P}(B_j|A)$.

Exemple 2.4.2. Supposons que la probabilité d'attraper une maladie D soit de 1%. Supposons également qu'il existe un test de dépistage pour la maladie avec les caractéristiques suivantes : le test rend un résultat positif 97% du temps si le patient est réellement atteint de la maladie D , et rend un résultat négatif 95% du temps si le patient n'est pas atteint de la maladie D . Si nous faisons le test sur une personne choisie au hasard :

- ▷ Quelle est la probabilité qu'elle soit indiquée comme malade ?
- ▷ Si le test rend un résultat positif, quelle est la probabilité que la personne soit réellement malade ?

On calcule

$$\mathbb{P}(T_+) = \mathbb{P}(D)\mathbb{P}(T_+ | D) + \mathbb{P}(D^c)\mathbb{P}(T_+ | D^c) \approx 0.06.$$

$$\mathbb{P}(D | T_+) = \frac{\mathbb{P}(D)}{\mathbb{P}(T_+)} \mathbb{P}(T_+ | D) \approx 0.16.$$

Remarquez notamment comment la fiabilité du test dépend de la prévalence de la maladie dans la population ! Si la maladie était plus rare il faudrait un test beaucoup plus efficace ! Avec une maladie rare (1 pour 10000) on a sous exactement les mêmes conditions : $\mathbb{P}(T_+) = 0,05$ et $\mathbb{P}(D | T_+) \approx 0,002$!

Lecture physique : fiabilité d'un détecteur. Supposons qu'une particule soit effectivement présente avec probabilité 0,10. Un détecteur donne un signal dans 95% des cas lorsqu'une particule est présente, mais produit aussi un faux signal dans 2% des cas lorsqu'elle est absente. Si S désigne l'événement "le détecteur signale quelque chose" et P l'événement "une particule est présente", alors

$$\mathbb{P}(P | S) = \frac{0,95 \times 0,10}{0,95 \times 0,10 + 0,02 \times 0,90} \approx 0,84.$$

La probabilité a posteriori dépend donc à la fois des performances du détecteur et de la fréquence a priori du phénomène recherché.

Exemple 2.4.3. Lors d'études médicales dans un laboratoire sur des souris, les données montrent que

- ▷ si une souris développe l'anticorps A, alors elle a également développé l'anticorps B,
- ▷ si une souris ne porte pas l'anticorps A, alors 4 fois sur 5, elle ne porte pas l'anticorps B,
- ▷ le tiers de la population des souris a développé l'anticorps A.

Si on prend une souris qui porte l'anticorps B, quelle est la probabilité qu'elle ait développé l'anticorps A ? On cherche donc à calculer $\mathbb{P}(A|B)$. Or, on connaît $\mathbb{P}(B|A) = 1$ et $\mathbb{P}(B|A^c) = \frac{1}{5}$. On sait aussi que

$$\mathbb{P}(A) = \frac{1}{3} \quad \text{et} \quad \mathbb{P}(A^c) = \frac{2}{3}.$$

Par la formule de Bayes, on a

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)} = \frac{1 \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + \frac{1}{5} \cdot \frac{2}{3}} = \frac{5}{7}.$$

Exemple 2.4.4. Adam, Benjamin et David ont retrouvé une chaussette perdue noire (N) et ne savent pas à qui elle appartient. On sait que

- ▷ Adam a 3 chaussettes noires et 5 colorées,
- ▷ Benjamin a 2 chaussettes noires et 7 colorées,
- ▷ David a 4 chaussettes noires et 3 colorées.

Nos trois copains sont tous les trois distraits mais Adam a deux fois plus de chance de perdre sa chaussette que les deux autres. Selon vous, qui va repartir avec la chaussette ? On calcule tout d'abord toutes les probabilités dont on aura besoin :

$$\mathbb{P}(N|A) = \frac{3}{8}, \quad \mathbb{P}(N|B) = \frac{2}{9}, \quad \mathbb{P}(N|D) = \frac{4}{7}.$$

De plus, comme Adam a deux fois plus de chance de perdre sa chaussette que Benjamin et David, on a

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

Ainsi par la formule de Bayes II, on a pour Adam

$$\mathbb{P}(A|N) = \frac{\mathbb{P}(N|A)\mathbb{P}(A)}{\mathbb{P}(N|A)\mathbb{P}(A) + \mathbb{P}(N|B)\mathbb{P}(B) + \mathbb{P}(N|D)\mathbb{P}(D)} = \frac{\frac{3}{8} \cdot \frac{1}{2}}{\frac{3}{8} \cdot \frac{1}{2} + \frac{2}{9} \cdot \frac{1}{4} + \frac{4}{7} \cdot \frac{1}{4}} \approx 0.49.$$

De même, pour Benjamin on calcule

$$\mathbb{P}(B|N) = \frac{\mathbb{P}(N|B)\mathbb{P}(B)}{\mathbb{P}(N|A)\mathbb{P}(A) + \mathbb{P}(N|B)\mathbb{P}(B) + \mathbb{P}(N|D)\mathbb{P}(D)} = \frac{\frac{2}{9} \cdot \frac{1}{4}}{\frac{3}{8} \cdot \frac{1}{2} + \frac{2}{9} \cdot \frac{1}{4} + \frac{4}{7} \cdot \frac{1}{4}} \approx 0.14$$

et pour David

$$\mathbb{P}(D|N) = \frac{\mathbb{P}(N|D)\mathbb{P}(D)}{\mathbb{P}(N|A)\mathbb{P}(A) + \mathbb{P}(N|B)\mathbb{P}(B) + \mathbb{P}(N|D)\mathbb{P}(D)} = \frac{\frac{4}{7} \cdot \frac{1}{4}}{\frac{3}{8} \cdot \frac{1}{2} + \frac{2}{9} \cdot \frac{1}{4} + \frac{4}{7} \cdot \frac{1}{4}} \approx 0.37.$$

Ainsi, c'est Adam qui va repartir avec la chaussette. Remarquez que le dénominateur est égal à $\mathbb{P}(N)$: on ne le calcule qu'une fois !

2.5 Indépendance

Remarquons que l'on peut avoir les cas suivants.

- ▷ $\mathbb{P}(A|B) > \mathbb{P}(A)$: A devient plus probable si on sait que B s'est produit. C'est le cas en particulier si $B \subseteq A$ (cas extrême).
- ▷ $\mathbb{P}(A|B) < \mathbb{P}(A)$: A devient moins probable si on sait que B s'est produit. C'est le cas en particulier si $A \cap B = \emptyset$ (cas extrême).
- ▷ $\mathbb{P}(A|B) = \mathbb{P}(A)$: le fait que B se produise n'a *aucune influence* sur la confiance que l'on

a en le fait que A se produise ou non. Dans ce cas, on a donc

$$\mathbb{P}(A|B) \underbrace{=}_{\text{def}} \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \underbrace{=}_{\text{hyp}} \mathbb{P}(A)$$

donc $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Dans ce cas, on a aussi $\mathbb{P}(B|A) = \mathbb{P}(B)$ puisque

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

Définition 2.5.1. Soient A et B deux événements. On dit que A et B sont *indépendants* si

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

Dans ce cas, on note $A \perp\!\!\!\perp B$.

Remarque 2.5.2. On pourrait également définir l'indépendance de A et B en demandant que $\mathbb{P}(A|B) = \mathbb{P}(A)$. Les avantages de notre définition sont qu'elle est valide même si $\mathbb{P}(B) = 0$ et qu'elle est symétrique en A et B .

Exemple 2.5.3. On tire une carte dans un jeu classique de 52 cartes. On considère les événements

- ▷ A = la carte tirée est un as,
- ▷ B = la carte tirée est un pique.

On remarque que $\mathbb{P}(A) = \frac{4}{52}$, $\mathbb{P}(B) = \frac{1}{4}$ et $\mathbb{P}(A \cap B) = \frac{1}{52} = \mathbb{P}(A)\mathbb{P}(B)$. Les événements A et B sont donc indépendants.

Proposition 2.5.4. Soient A et B deux événements.

1. Si $\mathbb{P}(B) = 0$, alors $A \perp\!\!\!\perp B$,
2. $A \perp\!\!\!\perp B$ si et seulement si $A \perp\!\!\!\perp B^c$,
3. $A \perp\!\!\!\perp \emptyset$, $A \perp\!\!\!\perp \Omega$.

Démonstration. 1. Puisque $A \cap B \subseteq B$, on a par monotonie de $\mathbb{P}$ que $\mathbb{P}(A \cap B) = 0$. Ainsi,

$$\mathbb{P}(A \cap B) = 0 = 0 \cdot \mathbb{P}(A) = \mathbb{P}(B)\mathbb{P}(A),$$

ce qui montre que les événements A et B sont indépendants.

2. Supposons que $A \perp\!\!\!\perp B$. On sait que $A = (A \cap B) \cup (A \cap B^c)$. Comme cette union est disjointe, il vient

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

puisque les événements A et B sont indépendants. On en tire que

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

ce qui suffit. Puisque $(B^c)^c = B$, la réciproque est également vérifiée.

3. Puisque $\mathbb{P}(\emptyset) = 0$, la première partie découle du point 1. Pour la deuxième partie, on remarque que $\emptyset^c = \Omega$ et on utilise le point 2. ■

Remarque 2.5.5. Attention, le fait que $A \perp\!\!\!\perp B$ et $B \perp\!\!\!\perp C$ n'implique pas que $A \perp\!\!\!\perp C$! Par exemple, reprenons l'exemple du jeu de carte avec

- ▷ A = la carte tirée est un as,
- ▷ B = la carte tirée est un pique,
- ▷ C = la carte tirée est un roi.

On sait que A et B sont indépendants et par le même raisonnement, B et C sont indépendants. Mais bien sûr A et C ne le sont pas !

Remarque 2.5.6. Si A et B sont disjoints, alors les événements A et B sont indépendants si et seulement si l'un des deux est de probabilité nulle. En effet, si $A \perp\!\!\!\perp B$, on a

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

ce qui implique que $\mathbb{P}(A) = 0$ ou $\mathbb{P}(B) = 0$. La réciproque est donnée par la proposition précédente.

On peut étendre la définition d'indépendance au cas de n événements.

Définition 2.5.7. On dit que des événements $A_1, \dots, A_n$ sont *indépendants* si pour toute partie non vide $J \subseteq \{1, \dots, n\}$, on a

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

Cette définition est plus forte que l'indépendance *deux à deux* pour laquelle on prend $\#J = 2$, c'est-à-dire on demande que

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j) \quad \forall i \neq j.$$

Exemple 2.5.8. Deux dés distinguables sont lancés. On considère les événements

- ▷ A = la somme des deux dés vaut 7,
- ▷ B = le résultat du premier dé est 4,
- ▷ C = le résultat du deuxième dé est 3.

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

On a

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{6}{36} = \frac{1}{6}.$$

De plus,

$$\mathbb{P}(A \cap B) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(A)\mathbb{P}(B), \quad \mathbb{P}(A \cap C) = \frac{1}{36} = \mathbb{P}(A)\mathbb{P}(C), \quad \mathbb{P}(B \cap C) = \frac{1}{36} = \mathbb{P}(B)\mathbb{P}(C)$$

mais

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{(4, 3)\}) = \frac{1}{36} \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

3 Variables aléatoires

Les variables aléatoires permettent de modéliser de manière rigoureuse et systématique des phénomènes aléatoires en associant à chaque issue possible d'une expérience un nombre réel. Elles offrent un cadre mathématique unifié pour l'étude d'expériences variées, facilitant le calcul des probabilités associées à divers événements d'intérêt.

3.1 Variables aléatoires

Bien souvent, les résultats fournis par une expérience aléatoire peuvent se traduire par une "grandeur", très souvent un nombre réel. La notion de *variable aléatoire* fournit un cadre unique pour l'étude de ce genre de situation.

Exemple 3.1.1. On considère les deux expériences suivantes.

- ▷ On lance une pièce. On peut traduire par 0 le fait d'obtenir *pile* et par 1 le fait d'obtenir *face*.
- ▷ Une urne possède 3 boules noires et 3 boules blanches. On peut traduire par 0 le fait de tirer une boule noire et par 1 le fait de tirer une boule blanche.

Ces deux expériences différentes (donc définies sur des espaces Ω différents) peuvent être *modélisées* de manière unique puisque dans les deux cas, on a

$$\text{Probabilité d'obtenir 0} = \text{Probabilité d'obtenir 1} = \frac{1}{2}.$$

De plus, on n'est pas toujours intéressés par l'issue même de l'expérience, mais plutôt à un *résultat* lié à cette issue.

Exemple 3.1.2.

- ▷ On lance deux dés et on s'intéresse à la somme des résultats.
- ▷ On lance 10 fois d'affilée une pièce de monnaie et on s'intéresse au nombre de fois que l'on a obtenu *pile*.
- ▷ On lance plusieurs fois d'affilée une pièce de monnaie et on s'intéresse au nombre de lancers nécessaires pour obtenir *pile*.

On considère donc une *fonction* qui à tout résultat de l'expérience associe un nombre réel. Comme le nombre réel dépend de l'expérience aléatoire, on peut déterminer les chances d'obtenir un nombre donné.

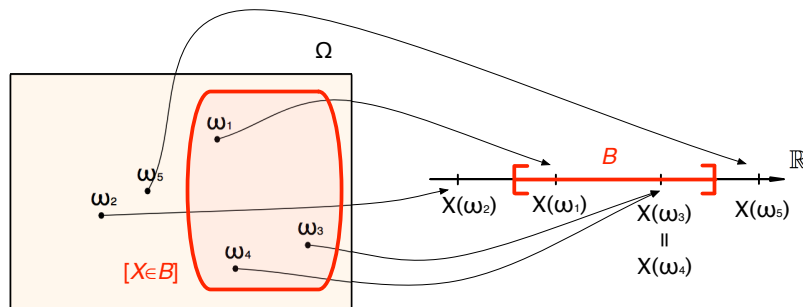
Définition 3.1.3. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Une *variable aléatoire* (abrégé en v.a.) sur $(\Omega, \mathcal{F}, \mathbb{P})$ est une application

$$X : \Omega \rightarrow \mathbb{R}$$

telle que pour tout sous-ensemble B de la σ -algèbre de Borel $\mathcal{B}$ de $\mathbb{R}$, l'ensemble

$$\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

(autrement dit, c'est un événement : il aura donc une probabilité).



C'est donc une application qui transforme les issues de l'expérience en nombres réels. L'application X n'est pas aléatoire, c'est la valeur à laquelle elle est appliquée qui est aléatoire !

Exemple 3.1.4. Considérons l'expérience consistant à lancer deux dés. L'espace probabilisé associé est $(\Omega, \mathcal{F})$, où $\Omega = \{(1, 1), (1, 2), \dots, (6, 6)\}$ est l'ensemble des issues possibles, et $\mathcal{F}$ est l'ensemble des parties de Ω . Soit X la variable aléatoire définie par la somme des résultats obtenus sur les deux dés, c'est-à-dire

$$X : \Omega \rightarrow \mathbb{R} : (j, k) \mapsto X(j, k) = j + k.$$

Exemple 3.1.5. Considérons l'expérience consistant à lancer une pièce de monnaie jusqu'à obtenir un PILE. L'espace probabilisé associé est $(\Omega, \mathcal{F})$, où

$$\Omega = \{\text{PILE}, (\text{FACE}, \text{PILE}), (\text{FACE}, \text{FACE}, \text{PILE}), \dots\}$$

est l'ensemble des séquences de lancers de pièce, et $\mathcal{F}$ est l'ensemble des parties de Ω . Soit X la variable aléatoire représentant le nombre de lancers nécessaires avant d'obtenir un PILE, c'est-à-dire

$$X : \Omega \rightarrow \mathbb{R} : \omega \mapsto X(\omega) = \#\{\text{composantes de } \omega\}.$$

Remarque 3.1.6. En général, on dira “soit X une variable aléatoire” sans se préoccuper de l’espace sur lequel elle est définie^a. On se permet donc d’écrire par exemple

- ▷ soit X le temps d’attente avant un appel,
- ▷ soit X le résultat du lancer d’un dé,
- ▷ soit X la somme des résultats de deux lancers de dé,
- ▷ ...

^a. C’est un des intérêt des variables aléatoires, on travaille dans $\mathbb{R}$ quelle que soit l’expérience !

La valeur prise par X est aléatoire : on ne peut pas la prédire. L’idée n’est donc pas de déterminer X mais de se donner des outils pour calculer ou estimer

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

pour tout sous-ensemble B de la σ -algèbre de Borel de $\mathbb{R}$.

Notations. Par convention, si X est une variable aléatoire, on note

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(X \in [a, b]) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in [a, b]\})$$

$$\mathbb{P}(X < a) = \mathbb{P}(X \in]-\infty, a[) = \mathbb{P}(\{\omega \in \Omega : X(\omega) < a\})$$

$$\mathbb{P}(X = a) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = a\})$$

$$\mathbb{P}(X \in B) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$$

...

et on adopte des notations similaires pour les intervalles (semi-)ouverts.

Définition 3.1.7. La loi de X (ou *distribution de X*) est définie par

$$\mathbb{P}_X : \mathcal{B} \rightarrow [0, 1] : B \mapsto \mathbb{P}(X \in B).$$

La valeur prise par X est aléatoire : on ne peut pas la prédire. Le seul objet qui nous intéresse, c’est sa loi car si on la connaît, on peut mesurer les risques avec précision. L’idée n’est donc pas de déterminer X (c’est impossible) mais de se donner des outils pour calculer ou estimer $\mathbb{P}(X \in B)$.

Proposition 3.1.8.

- ▷ La loi $\mathbb{P}_X$ de X est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B})$.
- ▷ Réciproquement, toute mesure de probabilité sur $(\mathbb{R}, \mathcal{B})$ est la loi d’une variable

aléatoire

Démonstration. 1. Clairement, $\mathbb{P}_X$ est à valeurs dans $[0, 1]$. De plus, on a

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

Enfin, si $(B_n)_{n \in \mathbb{N}}$ est une suite d'éléments de $\mathcal{B}$ deux à deux disjoints, on a

$$\begin{aligned} \mathbb{P}_X \left(\bigcup_{n \in \mathbb{N}} B_n \right) &= \mathbb{P} \left(X \in \bigcup_{n \in \mathbb{N}} B_n \right) = \mathbb{P}(\{\omega \in \Omega : \exists n \in \mathbb{N} \text{ tel que } X(\omega) \in B_n\}) \\ &= \mathbb{P} \left(\bigcup_{n \in \mathbb{N}} \{\omega \in \Omega : X(\omega) \in B_n\} \right) \\ &= \mathbb{P} \left(\bigcup_{n \in \mathbb{N}} \{X \in B_n\} \right) \\ &= \sum_{n \in \mathbb{N}} \mathbb{P}(X \in B_n) = \sum_{n \in \mathbb{N}} \mathbb{P}_X(B_n) \end{aligned}$$

puisque les événements $\{X \in B_n\}$ sont deux à deux disjoints.

2. Supposons que $\tilde{\mathbb{P}}$ est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B})$. Considérons la variable aléatoire

$$X : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto x.$$

Pour tout $B \in \mathcal{B}$, on a

$$\mathbb{P}_X(B) = \tilde{\mathbb{P}}(X \in B) = \tilde{\mathbb{P}}(\{x \in \mathbb{R} : x \in B\}) = \tilde{\mathbb{P}}(B).$$

Cela signifie que $\mathbb{P}_X = \tilde{\mathbb{P}}$. ■

3.2 Variables aléatoires discrètes et continues

Il y a deux familles très importantes de variables aléatoires.

Définition 3.2.1. Une variable aléatoire X est *discrète* si l'ensemble $X(\Omega)$ des valeurs prises par X est dénombrable. Dans ce cas, la loi de X est caractérisée par sa *fonction de masse*

$$p_X : X(\Omega) \rightarrow [0, 1] : x \mapsto p_X(x) = \mathbb{P}(X = x).$$

On écrit $X \sim p_X$. Remarquons que

$$\mathbb{P}_X(B) = \sum_{x \in X(\Omega) \cap B} p_X(x).$$

Exemple 3.2.2. Les variables aléatoires de l'Exemple 3.1.4 et de l'Exemple 3.1.5 sont discrètes.

Soit X une variable aléatoire discrète. Notons $X(\Omega) = \{x_j : j \in J\}$ avec $J \subset \mathbb{N}$. Comme $\mathbb{P}(X \in \{x_j : j \in J\}) = 1$, on a

$$\sum_{j \in J} p_X(x_j) = 1$$

Réciproquement, toute fonction $p : \{x_j : j \in J\} \rightarrow [0, 1]$ avec $J \subset \mathbb{N}$ qui satisfait

$$\sum_{j \in J} p(x_j) = 1$$

est la fonction de masse d'une variable aléatoire X telle que

$$p_j := p(x_j) = \mathbb{P}(X = x_j).$$

Notations.

- ▷ On dit que la loi de X est donnée par (x_j, p_j) , $j \in J$.
- ▷ Si $X(\Omega) = \{x_1, \dots, x_N\}$, on résume souvent la fonction de masse de X (et donc sa loi) via le tableau

valeurs	x_1	x_2	x_3	$\dots$	x_N
probabilités	p_1	p_2	p_3	$\dots$	p_N

On peut également représenter la fonction de masse grâce à un diagramme en bâtons.

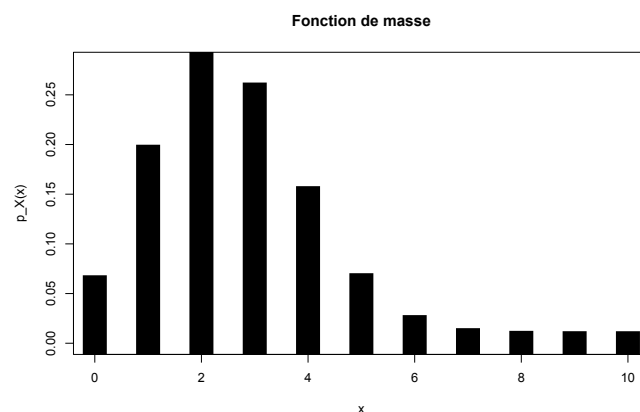


FIGURE 3.1 – Diagramme en bâtons d'une variable aléatoire X à valeurs dans $\{0, \dots, 10\}$.

Exemple 3.2.3. Considérons la variable aléatoire X qui donne la somme des résultats obtenus lors du lancer de deux dés. L'ensemble des valeurs prises par X est

$$X(\Omega) = \{2, 3, 4, \dots, 12\}$$

où on rappelle que $\Omega = \{(1, 1), (1, 2), \dots, (6, 6)\}$. On peut évidemment supposer que l'on travaille dans un contexte équiprobable et on peut facilement calculer toutes les probabilités en passant en revue les cas. Par exemple, on a

$$\mathbb{P}_X(\{2\}) = \mathbb{P}(X = 2) = \mathbb{P}(\{(1, 1)\}) = \frac{1}{36}$$

et

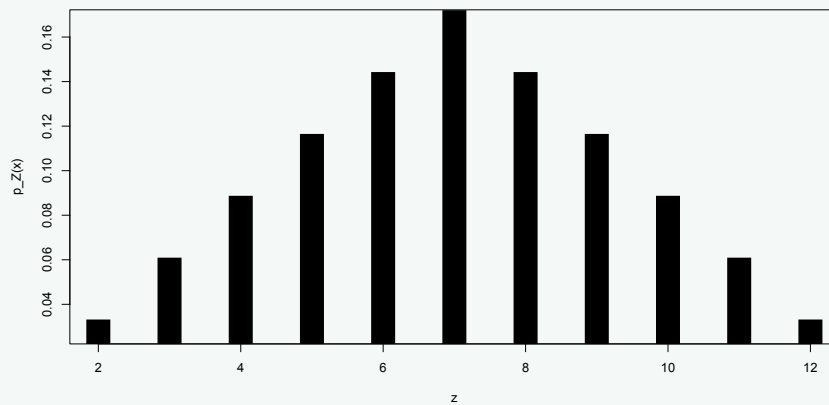
$$\mathbb{P}_X(\{3\}) = \mathbb{P}(X = 3) = \mathbb{P}(\{(1, 2), (2, 1)\}) = \frac{2}{36} = \frac{1}{18}.$$

De même, on calcule

$$\mathbb{P}(X = 4) = \frac{3}{36} = \frac{1}{12}, \mathbb{P}(X = 5) = \frac{4}{36} = \frac{1}{9}, \mathbb{P}(X = 6) = \frac{5}{36}, \mathbb{P}(X = 7) = \frac{6}{36} = \frac{1}{6},$$

$$\mathbb{P}(X = 8) = \frac{5}{36}, \mathbb{P}(X = 9) = \frac{4}{36} = \frac{1}{9}, \mathbb{P}(X = 10) = \frac{3}{36} = \frac{1}{12}, \mathbb{P}(X = 11) = \frac{2}{36} = \frac{1}{18}, \mathbb{P}(X = 12) = \frac{1}{36}.$$

Le diagramme en bâtons est donné par



et on peut également résumer l'information dans le tableau

valeurs	2	3	4	5	6	7	8	9	10	11	12
probabilités	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Définition 3.2.4. Une variable aléatoire X est *continue* s'il existe une fonction intégrable $f : \mathbb{R} \rightarrow [0, +\infty[$ telle que

$$\mathbb{P}_X(B) = \int_B f(x) dx$$

pour tout $B \in \mathcal{B}$. Dans ce cas, on écrit $X \sim f$ et on dit que f est la *fonction de densité* de X .

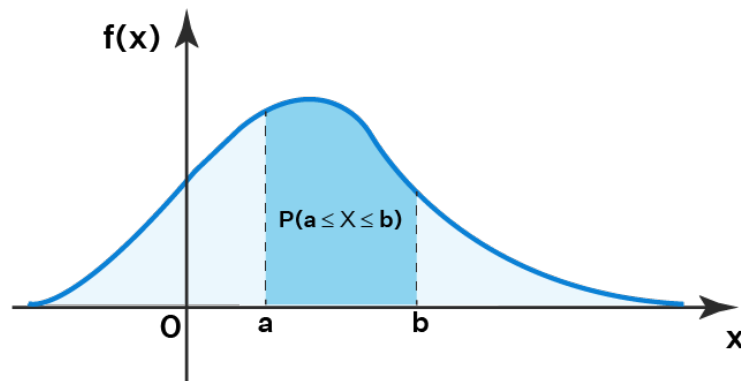


FIGURE 3.2 – La probabilité $\mathbb{P}(a \leq X \leq b)$ est donnée par l'aire sous la courbe de la fonction de densité f sur l'intervalle $[a, b]$.

Soit X une variable aléatoire continue de fonction de densité f . Comme $\mathbb{P}(X \in \mathbb{R}) = 1$, on a

$$\int_{\mathbb{R}} f(x) dx = 1.$$

Réciproquement, toute fonction positive et intégrable f d'intégrale égale à 1 est la fonction de densité d'une variable aléatoire continue.

Exemple 3.2.5.

- ▷ Le choix aléatoire d'un nombre entre 0 et 1 peut être modélisé par la densité $\mathbb{1}_{[0,1]}$. On parle de loi uniforme.
- ▷ La durée de vie d'un composant électronique peut être modélisée par la densité $f(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0,+\infty[}(x)$. On parle de loi exponentielle.

Remarque 3.2.6. Une fonction de densité peut avoir des points de discontinuité et n'est pas nécessairement bornée.

L'aire sous la courbe de f nous permet de calculer toutes les probabilités souhaitées. En particulier, on a

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x) dx = \mathbb{P}(a < X < b) \quad \text{et} \quad \mathbb{P}(X = a) = 0.$$

Exemple 3.2.7. Supposons que la durée de vie d'une ampoule en années, notée X , est modélisée par la densité

$$f(x) = e^{-x} \mathbb{1}_{[0,+\infty[}(x).$$

Calculons la probabilité que l'ampoule fonctionne entre 5 et 10 ans. On a

$$\mathbb{P}(5 \leq X \leq 10) = \int_5^{10} e^{-x} dx = [-e^{-x}]_5^{10} = -e^{-10} + e^{-5} \approx 6,70 \times 10^{-3}.$$

Calculons maintenant la probabilité que la lampe fonctionne au moins 15 ans. On a

$$\mathbb{P}(X \geq 15) = \int_{15}^{+\infty} e^{-x} dx = [-e^{-x}]_{15}^{+\infty} = e^{-15} \approx 3,06 \times 10^{-7}.$$

Remarque 3.2.8. Il existe des variables aléatoires qui ne sont ni discrètes ni continues. Par exemple, considérons une cible de fléchettes avec un rayon R . Nous définissons une variable aléatoire X qui représente la distance au centre de la cible, sous la condition que la fléchette a atterri sur la cible. Si la fléchette ne touche pas la cible, nous supposons que la variable prend une valeur fixe C . Autrement dit, X est définie par

$$X = \begin{cases} \text{distance au centre de la cible} & \text{si la fléchette touche la cible,} \\ C & \text{si la fléchette ne touche pas la cible.} \end{cases}$$

Dans ce cas, la variable X n'est ni discrète ni continue. Elle est mixte car elle combine une partie continue (la distance au centre lorsque la fléchette touche la cible) et une partie discrète (la valeur fixe C lorsque la fléchette ne touche pas la cible).

3.3 Fonction de répartition

Nous avons vu que la loi d'une variable aléatoire discrète est entièrement déterminée par sa fonction de masse, tandis que la loi d'une variable aléatoire continue est entièrement déterminée par sa fonction de densité. En général, les variables aléatoires ne sont ni discrètes, ni continues. On ne peut donc pas spécifier leur loi via les fonctions de masse ou de densité. Dans cette section nous associons à toute variable aléatoire, quelle que soit sa nature, une nouvelle fonction qui caractérise sa loi.

Définition 3.3.1. Soit X une variable aléatoire. La *fonction de répartition* de X est la fonction F_X définie pour tout $x \in \mathbb{R}$ par

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

Les propriétés des fonctions de répartition sont rassemblées dans la proposition suivante.

Proposition 3.3.2. Soit X une variable aléatoire. Alors

▷ F_X est croissante,

- ▷ F_X est continue à droite,
- ▷ $\lim_{x \rightarrow -\infty} F_X(x) = 0$ et $\lim_{x \rightarrow +\infty} F_X(x) = 1$.

Démonstration. 1. Si $x \leq y$, alors $\{X \leq x\} \subseteq \{X \leq y\}$. La monotonie de la mesure de probabilité implique que

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

2. Soit $x \in \mathbb{R}$. Il faut montrer que

$$\lim_{y \rightarrow x^+} F_X(y) = F_X(x).$$

Soit $(x_n)_{n \in \mathbb{N}}$ une suite qui décroît vers x . Alors, on a

$$\{X \leq x\} = \bigcap_{n \in \mathbb{N}} \{X \leq x_n\}$$

et puisqu'il s'agit d'une intersection d'événements décroissante, on obtient que

$$F_X(x) = \mathbb{P}(X \leq x) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq x_n) = \lim_{n \rightarrow +\infty} F_X(x_n).$$

Puisque cette relation est vérifiée pour toute suite qui décroît vers x , on en tire la conclusion.

3. On procède de la même manière en prenant une suite qui décroît vers $-\infty$ et une suite qui croît vers $+\infty$. ■

La fonction de répartition permet de calculer les probabilités liées à X .

Proposition 3.3.3. *De plus, pour tous réels $a < b$, on a*

- ▷ $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$
- ▷ $\mathbb{P}(X > a) = 1 - F_X(a)$
- ▷ $\mathbb{P}(X < a) = F_X(a^-)$
- ▷ $\mathbb{P}(X = a) = F_X(a) - F_X(a^-)$

où $F_X(a^-) = \lim_{x \rightarrow a^-} F_X(x)$.

En particulier, F_X est continue en un point $a \in \mathbb{R}$ si et seulement si $\mathbb{P}(X = a) = 0$.

Démonstration. 1. Puisque

$$\{X \in]a, b]\} = \{X \leq b\} \setminus \{X \leq a\},$$

on obtient

$$\mathbb{P}(a < X \leq b) = \mathbb{P}(X \in]a, b]) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F_X(b) - F_X(a).$$

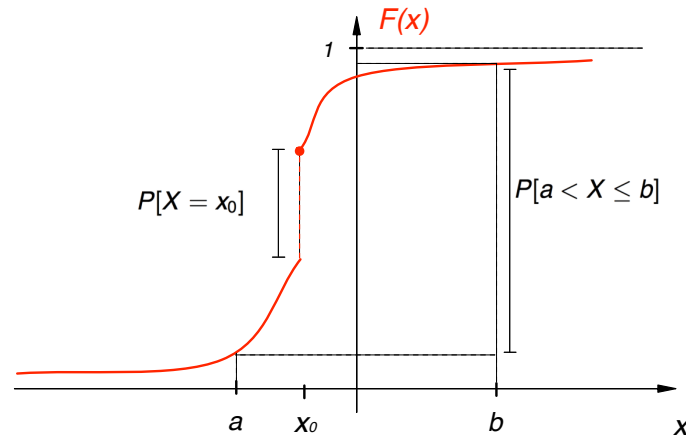


FIGURE 3.3 – La hauteur du saut de F_X en x_0 est égal à $\mathbb{P}(X = a)$.

2. On a $\{X > a\} = \{X \leq a\}^c$, d'où

$$\mathbb{P}(X > a) = 1 - \mathbb{P}(X \leq a) = 1 - F_X(a).$$

3. Par définition,

$$F_X(a^-) = \lim_{x \rightarrow a^-} F_X(x).$$

Soit $(x_n)_{n \in \mathbb{N}}$ une suite qui croît strictement vers a . Alors

$$\{X < a\} = \bigcup_{n \in \mathbb{N}} \{X \leq x_n\}$$

et comme la suite d'événements est croissante, on en tire que

$$\mathbb{P}(X < a) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq x_n) = \lim_{n \rightarrow +\infty} F_X(x_n) = F_X(a^-).$$

4. Il suffit d'appliquer le point précédent puisque

$$\{X = a\} = \{X \leq a\} \setminus \{X < a\}.$$

■

Théorème 3.3.4. *La fonction de répartition de X caractérise la loi de X . Par conséquent, on écrit $X \sim F$.*

Ce résultat est admis, mais on peut s'en convaincre intuitivement grâce à l'égalité

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$$

qui montre qu'on peut calculer les probabilités qui nous intéressent via la fonction de répartition. On verra que, dans le cas d'une variable aléatoire discrète ou continue, ce résultat est évident.

Définition 3.3.5. Soient X et Y deux variables aléatoires.

▷ On dit que $X = Y$ *presque sûrement*, et on note $X \stackrel{ps}{=} Y$, si

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

En particulier, il faut que les variables aléatoires X et Y soient définies sur le même espace.

▷ On dit que $X = Y$ *en loi* ou que X et Y sont *identiquement distribuées*, et on note $X \stackrel{\mathcal{L}}{=} Y$, si $\mathbb{P}_X = \tilde{\mathbb{P}}_Y$ c'est-à-dire si

$$\mathbb{P}(X \in B) = \tilde{\mathbb{P}}(Y \in B)$$

pour tout sous-ensemble B de la σ -algèbre de Borel $\mathcal{B}$ de $\mathbb{R}$, où X est défini sur $(\Omega, \mathcal{F}, \mathbb{P})$ et Y est défini sur $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$.

Exemple 3.3.6.

▷ Soit X la variable aléatoire qui rend 1 si on répond correctement à un QCM avec 3 distracteurs et 0 sinon. Alors, les valeurs prises par X sont données par l'ensemble $\{0, 1\}$ et on a $\mathbb{P}(X = 0) = \frac{3}{4} = 1 - \mathbb{P}(X = 1)$.

▷ Soit Y la variable aléatoire qui rend 1 si on tire un coeur d'un jeu de cartes classique et 0 sinon. Alors, les valeurs prises par Y sont données par $\{0, 1\}$ et on a $\tilde{\mathbb{P}}(Y = 0) = \frac{3}{4} = 1 - \tilde{\mathbb{P}}(Y = 1)$.

On en tire donc que $X \stackrel{\mathcal{L}}{=} Y$.

Remarque 3.3.7.

▷ On a donc obtenu que $X \stackrel{\mathcal{L}}{=} Y$ ssi $F_X = F_Y$.

▷ L'égalité presque sûre implique l'égalité en loi. L'inverse n'est pas vrai, même si les variables sont définies sur le même espace. Par exemple, lors du lancer d'une pièce de monnaie, on peut considérer la variable aléatoire X définie par

$$X(\omega) = \begin{cases} 1 & \text{si } \omega = \text{pile} \\ 0 & \text{si } \omega = \text{face.} \end{cases}$$

et la variable aléatoire $Y = 1 - X$. Alors $X \stackrel{\mathcal{L}}{=} Y$ mais $\mathbb{P}(X = Y) = 0$.

Fonction de répartition dans le cas discret : Si X est une variable aléatoire discrète, sa fonction de répartition F_X est

- ▷ une *fonction en escalier*
- ▷ dont les sauts se situent exactement aux points $x \in \mathbb{R}$ tels que $\mathbb{P}(X = x) > 0$
- ▷ et la taille du saut y est donnée par $\mathbb{P}(X = x)$.

On peut donc retrouver la fonction de masse à partir de la fonction de répartition, ce qui montre que celle-ci caractérise bien la loi.

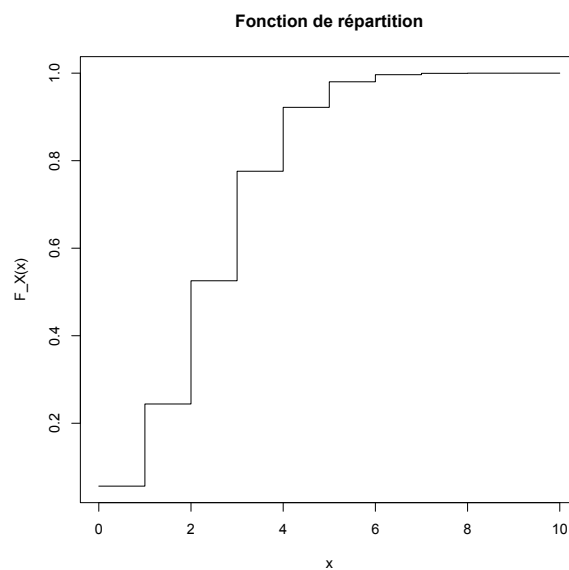


FIGURE 3.4 – Fonction de répartition d’une variable aléatoire discrète.

Fonction de répartition dans le cas continu : Si X est une variable aléatoire continue de fonction de densité f , alors

- ▷ sa fonction de répartition F_X est une fonction *continue* sur $\mathbb{R}$,
- ▷ si f est continue en x ¹, alors F_X est dérivable en x et

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

On peut donc retrouver la fonction de densité à partir de la fonction de répartition, ce qui montre que celle-ci caractérise bien la loi.

Terminons cette section en introduisant la fonction quantile.

Définition 3.3.8. Soit X une variable aléatoire de fonction de répartition F .

1. On peut montrer que ça arrive partout sauf en un nombre au plus dénombrable de points

▷ On définit sa *fonction quantile*

$$F^{-1} :]0, 1[\rightarrow \mathbb{R} : \alpha \mapsto F^{-1}(\alpha) := \inf \{x \in \mathbb{R} : F(x) \geq \alpha\}.$$

C'est l'inverse généralisé de F .

▷ Le *quantile d'ordre* α de X est le nombre

$$x_\alpha := F^{-1}(\alpha).$$

▷ La *médiane* de X est le quantile d'ordre $\frac{1}{2}$.

▷ Les *quartiles* sont les quantiles d'ordre $\frac{1}{4}$ et $\frac{3}{4}$.

3.4 Espérance et moments d'une variable aléatoire

Il est souvent utile d'avoir un résumé des caractéristiques principales de la loi de X , grâce à des indications numériques faciles à calculer et à interpréter. Les deux caractéristiques les plus importantes sont *l'espérance et la variance*.

Nous verrons que l'espérance est un indicateur de position ; la variance est un indicateur de dispersion (représente la variabilité présente dans les valeurs prises par la variable aléatoire autour de sa valeur centrale).

Commençons par établir un lien avec l'approche fréquentiste : Supposons que $\{\omega_1, \dots, \omega_n\}$ sont les résultats d'une expérience répétée n fois. On note $\{x_1, \dots, x_J\}$ l'ensemble des images de $\omega_1, \dots, \omega_n$ par X et on forme le tableau des effectifs et des fréquences

	x_1	$\dots$	x_J
effectifs	n_1	$\dots$	n_J
fréquences	f_1	$\dots$	f_J

où $f_j = \frac{n_j}{n}$. La moyenne et la variance de l'échantillon sont données respectivement par

$$\bar{x} = \frac{1}{n} \sum_{j=1}^J n_j x_j = \sum_{j=1}^J f_j x_j$$

et

$$s^2 = \frac{1}{n} \sum_{j=1}^J n_j (x_j - \bar{x})^2 = \sum_{j=1}^J f_j (x_j - \bar{x})^2$$

Dans un premier temps, nous définissons par analogie avec l'approche ci-dessus un indicateur de position.

Définition 3.4.1. Soit X une variable aléatoire.

▷ Si X est discrète de distribution (x_j, p_j) , $j \in J$, l'espérance de X est définie par

$$\mathbb{E}[X] = \sum_{j \in J} x_j p_j$$

à condition que cette série soit absolument convergente.

▷ Si X est continue de fonction de densité f , l'espérance de X est définie par

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f(x) dx$$

à condition que la fonction $x \mapsto x f(x)$ soit intégrable sur $\mathbb{R}$.

Dans les deux cas, si X possède une espérance, on note

$$X \in L^1.$$

Notons qu'il existe un cadre général qui permet de définir l'espérance d'une variable aléatoire de manière unique (sans se limiter aux cas discret et continu, et sans distinction entre ces cas). Cette définition repose sur l'intégrale de X par rapport à la mesure $\mathbb{P}$. Cela nécessite néanmoins de devoir généraliser la notion d'intégrale au cas de mesures de probabilité.

Exemple 3.4.2.

▷ Soit X le résultat du lancer d'un dé. On a $\mathbb{P}(X = n) = \frac{1}{6}$ pour tout $n \in \{1, \dots, 6\}$ et donc

$$\mathbb{E}[X] = \sum_{n=1}^6 n \mathbb{P}(X = n) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3,5$$

▷ On suppose que Y est le résultat d'un deuxième dé et on s'intéresse à $Z = X + Y$. La variable aléatoire Z prend ses valeurs dans $\{2, \dots, 12\}$ et on a déjà vu que la distribution de Z est donnée par

valeurs	2	3	4	5	6	7	8	9	10	11	12
probabilités	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

On a

$$\mathbb{E}[Z] = \sum_{n=2}^{12} n \mathbb{P}(Z = n) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{1}{18} + 4 \cdot \frac{1}{12} + \dots + 12 \cdot \frac{1}{36} = 7.$$

En particulier, on voit que $\mathbb{E}[Z] = \mathbb{E}[X] + \mathbb{E}[Y]$ où $Z = X + Y$. Ce résultat est vrai en toute généralité : on verra que l'espérance est un opérateur linéaire.

Exemple 3.4.3.

- ▷ Soit X le résultat d'un tirage aléatoire d'un nombre entre 0 et 1 modélisé par la fonction de densité $f(x) = \mathbf{1}_{[0,1]}$. On a

$$\mathbb{E}[X] = \int_0^1 x \, dx = \frac{1}{2}.$$

- ▷ Soit X la durée de vie de l'ampoule modélisé par la densité $f(x) = e^{-x}\mathbf{1}_{[0,+\infty[}(x)$. On a

$$\mathbb{E}[X] = \int_0^{+\infty} x e^{-x} dx = [-x e^{-x}]_0^{+\infty} + \int_0^{+\infty} e^{-x} dx = 1.$$

Proposition 3.4.4 (Propriétés de l'espérance). *Soient $X, Y \in L^1$ et soit $\alpha \in \mathbb{R}$.*

1. **Linéarité** : $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ et $\mathbb{E}[\alpha X] = \alpha \mathbb{E}[X]$.
2. **Monotonie** : Si $X \leq Y$, alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$.
3. **Constante** : $\mathbb{E}[\alpha] = \alpha$.
4. **Inégalité de Markov** : Si $X \geq 0$, alors

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

pour tout $a > 0$.

5. Si $X \geq 0$, alors $\mathbb{E}[X] = 0$ si et seulement si $X \stackrel{ps}{=} 0$.

Démonstration. 1. Le cas de la multiplication par α dans le cas discret est immédiat. En effet, si la loi de X est donnée par (x_j, p_j) , $j \in J$, alors la loi de αX est donnée par $(\alpha x_j, p_j)$, $j \in J$. Dans le cas continu, on peut montrer que si $\alpha > 0$, la fonction de densité de αX est donnée par $\frac{1}{\alpha} f(\frac{x}{\alpha})$. En effet, pour tout $x \in \mathbb{R}$, on a

$$\mathbb{P}(\alpha X \leq x) = \mathbb{P}(X \leq \frac{x}{\alpha}) = \int_{-\infty}^{\frac{x}{\alpha}} f(t) dt = \int_{-\infty}^x \frac{1}{\alpha} f(\frac{u}{\alpha}) du$$

par la formule du changement de variables². Il vient donc

$$\mathbb{E}[\alpha X] = \int_{\mathbb{R}} \frac{x}{\alpha} f(\frac{x}{\alpha}) dx = \alpha \int_{\mathbb{R}} t f(t) dt = \alpha \mathbb{E}[X].$$

On procède de manière semblable si $\alpha < 0$. Nous reviendrons sur la somme lorsque nous étudierons les lois jointes.

2. Si $X \geq 0$, alors $\mathbb{E}[X]$ est une série à termes positifs ou une intégrale d'une fonction positive. Il est donc clair que $\mathbb{E}[X] \geq 0$. Si $X \leq Y$, alors $Y - X \geq 0$ et donc $\mathbb{E}[Y - X] \geq 0$. La linéarité de l'espérance permet de conclure.

3. C'est évident.

2. Nous y reviendrons dans la Section 3.7.

4. Considérons la variable aléatoire Y définie par

$$Y = \begin{cases} 1 & \text{si } X \geq a \\ 0 & \text{sinon.} \end{cases}$$

Remarquons que

$$Y \leq \frac{X}{a}.$$

La monotonie et la linéarité de l'espérance donnent alors

$$\mathbb{E}[Y] \leq \mathbb{E}\left[\frac{X}{a}\right] = \frac{\mathbb{E}[X]}{a}.$$

Pour conclure, il suffit de remarquer que

$$\mathbb{E}[Y] = 1 \mathbb{P}(X \geq a) + 0 \mathbb{P}(X < a) = \mathbb{P}(X \geq a).$$

5. Pour tout $a > 0$, l'inégalité de Markov donne

$$\mathbb{P}(X \geq a) = 0.$$

Ainsi, puisque X est positive,

$$\mathbb{P}(X = 0) = 1 - \mathbb{P}(X > 0) = 1 - \mathbb{P}\left(\bigcup_{n \geq 1} \{X \geq \frac{1}{n}\}\right)$$

et

$$\mathbb{P}\left(\bigcup_{n \geq 1} \{X \geq \frac{1}{n}\}\right) \leq \sum_{n \geq 1} \mathbb{P}(X \geq \frac{1}{n}) = 0.$$

■

Souvent, on peut s'intéresser à une autre variable aléatoire Y qui est une fonction de X , c'est-à-dire $Y = h(X)$. Pour calculer $\mathbb{E}[Y]$, il faut dans un premier temps déterminer la distribution de Y . Le résultat ci-dessous, souvent pris comme définition, montre qu'on peut calculer l'espérance de Y directement à partir de la distribution de X , sans avoir à connaître la loi exacte de Y .

Proposition 3.4.5. *Soit $h : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction telle que $h(X) \in L^1$.*

▷ *Si X est discrète de distribution (x_j, p_j) , $j \in J$, l'espérance de $h(X)$ est donnée par*

$$\mathbb{E}[h(X)] = \sum_{j \in J} h(x_j) p_j$$

à condition que la série converge absolument.

▷ Si X est continue de fonction de densité f , l'espérance de $h(X)$ est donnée par

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x)f(x)dx$$

à condition que la fonction hf soit intégrable sur $\mathbb{R}$.

Démonstration. Nous le démontrons dans un cas particulier. Supposons que X est une variable aléatoire qui prend un nombre fini de valeurs différentes. Alors $h(X(\Omega))$ est évidemment fini. Si on écrit $X(\Omega) = \{x_1, \dots, x_n\}$ et $h(X(\Omega)) = \{y_1, y_2, \dots, y_m\}$ (avec $m \leq n$), on a $X(\Omega) = \cup_{i=1}^m h^{-1}(\{y_i\})$ et par σ -additivité, il vient

$$\mathbb{P}(h(X) = y_i) = \mathbb{P}\left(\bigcup_{j:h(x_j)=y_i} \{X = x_j\}\right) = \sum_{j:h(x_j)=y_i} \mathbb{P}(X = x_j)$$

et donc

$$\begin{aligned} \mathbb{E}[h(X)] &= \sum_{i=1}^m y_i \mathbb{P}(h(X) = y_i) = \sum_{i=1}^m y_i \left(\sum_{j:h(x_j)=y_i} \mathbb{P}(X = x_j) \right) \\ &= \sum_{i=1}^m \sum_{j:h(x_j)=y_i} h(x_j) \mathbb{P}(X = x_j) = \sum_{j=1}^n h(x_j) \mathbb{P}(X = x_j), \end{aligned}$$

ce qui démontre le résultat. ■

En prenant $h(x) = x^k$ ou $h(x) = (x - \mathbb{E}[X])^k$ pour $k \in \mathbb{N}_0$, on obtient les définitions suivantes.

Définition 3.4.6. Si X est une variable aléatoire telle que $X^k \in L^1$, on note

$$X \in L^k$$

et

▷ le *moment d'ordre k* de X est défini par $\mathbb{E}[X^k]$

▷ le *moment centré d'ordre k* de X est défini par $\mathbb{E}[(X - \mathbb{E}[X])^k]$.

Si $X \in L^2$, le moment centré d'ordre 2 est appelé la *variance* de X et on note

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

La racine carrée de la variance s'appelle l'*écart-type*

$$\sigma_X = \sqrt{\text{Var}[X]}.$$

Exemple 3.4.7. Soit X le résultat du lancer d'un dé. Puisqu'on le modélise par un contexte équiprobable, on a $\mathbb{P}(X = n) = \frac{1}{6}$ pour tout $n \in \{1, \dots, 6\}$. On calcule

$$\mathbb{E}[X^2] = \sum_{n=1}^6 n^2 \mathbb{P}(X = n) = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) = \frac{91}{6}$$

et

$$\mathbb{E}[X^3] = \sum_{n=1}^6 n^3 \mathbb{P}(X = n) = \frac{1}{6}(1^3 + 2^3 + 3^3 + 4^3 + 5^3 + 6^3) = \frac{147}{2}.$$

On peut aussi calculer

$$\text{Var}[X] = \mathbb{E}[(X - 3,5)^2] = \sum_{n=1}^6 (n - 3,5)^2 \frac{1}{6} = \frac{35}{12}.$$

Remarque 3.4.8. On peut montrer que $L^k \subset L^q$ si $k \geq q$. En particulier, si $X \in L^2$, on a $X \in L^1$.

Proposition 3.4.9. Si $X \in L^2$ et $\alpha \in \mathbb{R}$, alors

1. $\text{Var}[X] \geq 0$,
2. $\text{Var}[X] = 0$ si et seulement si X est presque sûrement constant et égal à $\mathbb{E}[X]$,
3. $\text{Var}[\alpha X] = \alpha^2 \text{Var}[X]$ et $\text{Var}[X + \alpha] = \text{Var}[X]$,
4. $\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Démonstration. 1. Il est clair que $\text{Var}[X] \geq 0$ puisque $(X - \mathbb{E}[X])^2 \geq 0$.

2. Supposons que $\text{Var}[X] = 0$. Alors, $(X - \mathbb{E}[X])^2$ est presque sûrement égal à zéro. Ainsi, X prend une unique valeur donnée par $\mathbb{E}[X]$. La réciproque de ce résultat est évidente.

3. Il suffit de remarquer que

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

en utilisant deux fois la linéarité de l'espérance. De même, on a

$$\text{Var}[X + \alpha] = \mathbb{E}[(X + \alpha - \mathbb{E}[X + \alpha])^2] = \mathbb{E}[(X + \alpha - (\mathbb{E}[X] + \alpha))^2] = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

en utilisant à nouveau la linéarité de l'espérance.

4. On a successivement

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

où on a utilisé la linéarité de l'espérance et le fait que $\mathbb{E}[X]$ est une constante. ■

Dans la pratique, on calcule souvent la variance avec la formule donnée au point 4.

Exemple 3.4.10. Pour le lancer du dé, on a

$$\text{Var}[X] = \sum_{n=1}^6 n^2 \frac{1}{6} - (3,5)^2 = \dots = \frac{35}{12}.$$

Définition 3.4.11. Soit $X \in L^2$ une variable aléatoire d'espérance μ et d'écart-type σ . Centrer et réduire X revient à considérer la variable aléatoire Z définie par

$$Z = \frac{X - \mu}{\sigma}.$$

L'espérance $\mathbb{E}[X]$ d'une variable aléatoire X nous renseigne sur la valeur “moyenne” ou “attendue” de X . C'est une indication de *position*.

Proposition 3.4.12 (Décomposition de König-Huygens). Si $X \in L^2$, alors

$$\mathbb{E}[(X - a)^2] = \text{Var}[X] + (\mathbb{E}[X] - a)^2 \text{ pour tout } a \in \mathbb{R}.$$

En particulier,

$$\mathbb{E}[X] = \operatorname{argmin}_{a \in \mathbb{R}} \mathbb{E}[(X - a)^2].$$

Démonstration. On utilise les propriétés de l'espérance pour écrire

$$\begin{aligned}
 \mathbb{E}[(X - a)^2] &= \mathbb{E}[(X - \mathbb{E}[X]) + (\mathbb{E}[X] - a)]^2 \\
 &= \mathbb{E}[(X - \mathbb{E}[X])^2] + \mathbb{E}[2(X - \mathbb{E}[X])(\mathbb{E}[X] - a)] + \mathbb{E}[(\mathbb{E}[X] - a)^2] \\
 &= \mathbb{E}[(X - \mathbb{E}[X])^2] + 0 + (\mathbb{E}[X] - a)^2.
 \end{aligned}$$

Cette valeur est donc minimale lorsque $a = \mathbb{E}[X]$. ■

Ainsi,

- ▷ $\mathbb{E}[X]$ est la meilleure approximation de X par une constante en norme L^2
- ▷ $\text{Var}[X]$ est l'écart minimal qu'on peut obtenir.

Le résultat suivant montre que la variance est une mesure de la *dispersion* de X autour de son espérance.

Proposition 3.4.13 (Inégalité de Tchebychev). *Soit $X \in L^2$ une variable aléatoire d'espérance μ et d'écart-type σ . Alors pour tout $r > 0$,*

$$\mathbb{P}(|X - \mu| \geq r\sigma) \leq \frac{1}{r^2}.$$

Démonstration. Cela découle directement de l'inégalité de Markov appliquée à la variable aléatoire positive $(X - \mathbb{E}[X])^2$. En effet, on a

$$\mathbb{P}(|X - \mu| \geq r\sigma) = \mathbb{P}((X - \mu)^2 \geq r^2\sigma^2) \leq \frac{\mathbb{E}[(X - \mu)^2]}{r^2\sigma^2} = \frac{1}{r^2}.$$

■

Ainsi, seule une grande variance peut permettre à une variable X de différer significativement de son espérance puisque dans ce cas, le risque de voir une valeur extrême de X augmente. On peut interpréter cette inégalité de la manière suivante : la variable aléatoire X prend ses valeurs dans l'intervalle $[\mu - r\sigma, \mu + r\sigma]$ avec une probabilité plus grande ou égale à $(1 - \frac{1}{r^2})$.

Pour $r = 2$, on trouve une probabilité d'au moins $\frac{3}{4}$ d'être dans l'intervalle $[\mu - 2\sigma, \mu + 2\sigma]$. Pour $r = 3$, la probabilité d'être dans l'intervalle $[\mu - 3\sigma, \mu + 3\sigma]$ est supérieure ou égale à $8/9 \approx 0,888$. Pour $r = 4$, on trouve une probabilité supérieure ou égale à $15/16 \approx 0,9375$ d'être dans l'intervalle $[\mu - 4\sigma, \mu + 4\sigma]$.

Terminons cette section en introduisant deux fonctions associées à une variable aléatoire donnée qui permettent de retrouver les moments de cette variable aléatoire et qui, de plus, caractérisent entièrement sa loi.

Définition 3.4.14. Soit X une variable aléatoire.

- ▷ La *fonction caractéristique* de X est la fonction φ_X définie pour tout $t \in \mathbb{R}$ par

$$\varphi_X(t) = \mathbb{E}[e^{itX}].$$

- ▷ La *fonction génératrice des moments* de X est la fonction M_X définie par

$$M_X(t) = \mathbb{E}[e^{tX}]$$

pour tout t tel que l'espérance est définie.

Ces fonctions caractérisent la loi de X :

$$X \stackrel{\mathcal{L}}{=} Y \text{ si et seulement si } \varphi_X(t) = \varphi_Y(t) \text{ pour tout } t \in \mathbb{R}$$

et si X et Y admettent une fonction génératrice des moments pour tout $|t| < \varepsilon$,

$$X \stackrel{\mathcal{L}}{=} Y \text{ si et seulement si } M_X(t) = M_Y(t) \text{ pour tout } |t| < \varepsilon.$$

De plus, si ils existent, elles permettent de retrouver les moments de X :

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}[X^k] \quad \text{et} \quad M_X^{(k)}(0) = \mathbb{E}[X^k]$$

En effet, considérons le cas de la fonction génératrice des moments. On a

$$\begin{aligned} \mathbb{E}[e^{tX}] &= \mathbb{E}\left[1 + tX + \frac{1}{2}t^2X^2 + \frac{1}{3!}t^3X^3 + \dots\right] \\ &= 1 + t\mathbb{E}[X] + \frac{1}{2}t^2\mathbb{E}[X^2] + \frac{1}{3!}t^3\mathbb{E}[X^3] + \dots \end{aligned}$$

Cette réécriture (formelle) justifie la terminologie utilisée.

Remarque 3.4.15. Dans le cas où X est une variable aléatoire continue de fonction de densité f , on a

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx} f(x) dx$$

et la fonction caractéristique de X est donc la transformée de Fourier de la fonction de densité f . De même,

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{\mathbb{R}} e^{tx} f(x) dx$$

et la fonction génératrice des moments de X est la transformée de Laplace de f .

3.5 Lois discrètes usuelles

N'importe quelle suite de nombres positifs dont la somme vaut 1 définit une loi de probabilité discrète. Il existe certaines suites qui apparaissent un peu partout dans la vraie vie et qui permettent de modéliser un grand nombre de phénomènes. Nous allons en présenter quelques-unes.

3.5.1 Loi discrète uniforme

Définition 3.5.1. Une variable aléatoire discrète X qui prend ses valeurs dans $\{1, \dots, n\}$ avec la même probabilité est appelée *uniforme* de paramètre n . On note $X \sim \text{Unif}(\{1, \dots, n\})$.

La loi uniforme de paramètre n est donnée par

$$\mathbb{P}(X = k) = \frac{1}{n}$$

pour tout $k \in \{1, \dots, n\}$.

Exemple 3.5.2.

▷ Résultat d'un dé : $X \sim \text{Unif}(\{1, \dots, 6\})$.

▷ Nombre choisi au hasard entre 1 et 100 : $X \sim \text{Unif}(\{1, \dots, 100\})$.

Proposition 3.5.3. Si $X \sim \text{Unif}(\{1, \dots, n\})$, alors $\mathbb{E}[X] = \frac{n+1}{2}$ et $\text{Var}[X] = \frac{n^2-1}{12}$.

Démonstration. On a

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

De même, on calcule

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

et donc

$$\begin{aligned} \text{Var}[X] &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{(n+1)(4n+2-3n-3)}{12} \\ &= \frac{(n+1)(n-1)}{12} \\ &= \frac{n^2-1}{12}. \end{aligned}$$

■

3.5.2 Loi de Bernoulli et loi binomiale

Définition 3.5.4. Une variable aléatoire n'ayant que deux issues possibles, 1 et 0 avec les probabilités p et $1-p$, est appelée une variable aléatoire de *Bernoulli* de taux de succès p . On note $X \sim \text{Bern}(p)$. La loi de Bernoulli est donnée par

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

Exemple 3.5.5.

- ▷ Lancer d'une pièce : $X = 1$ si "pile" et $X = 0$ si "face". Alors $X \sim \text{Bern}(1/2)$.
- ▷ On veut faire 6 lors du lancer d'un dé : $X = 1$ si le résultat est 6, $X = 0$ sinon. Alors $X \sim \text{Bern}(1/6)$.

Proposition 3.5.6. Si $X \sim \text{Bern}(p)$, alors $\mathbb{E}[X] = p$ et $\text{Var}[X] = p(1 - p)$.

Démonstration. On calcule directement que

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

De même,

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

donc

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

■

Un *schéma de Bernoulli* de paramètre p est une expérience aléatoire n'ayant que deux issues possibles (succès/échec) de probabilités p et $1 - p$.

Définition 3.5.7. Une variable aléatoire qui compte le nombre de succès lors de n répétitions indépendantes d'un même schéma de Bernoulli de paramètre p est appelée *binomiale* de paramètres (n, p) . On note $X \sim \text{Bin}(n, p)$. La *loi binomiale* est donnée par

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

pour tout $k \in \{0, \dots, n\}$.

Exemple 3.5.8.

- ▷ Nombre de "pile" obtenus lors du lancer de 10 pièces : $X \sim \text{Bin}(10, 1/2)$.
- ▷ Un détecteur enregistre chaque particule incidente indépendamment avec une efficacité p . Si n particules traversent le dispositif et si X désigne le nombre de particules détectées, alors $X \sim \text{Bin}(n, p)$.

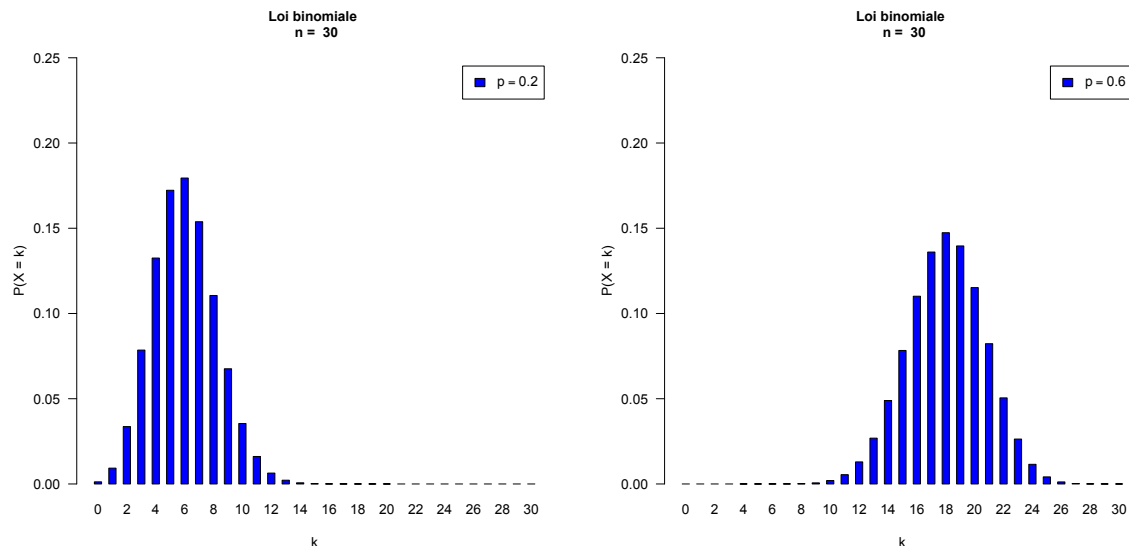


FIGURE 3.5 – Diagrammes en bâtons de lois binomiales de paramètres $(30, 0.2)$ et $(30, 0.6)$.

Interprétation physique. La loi de Bernoulli décrit une seule détection (signal / pas de signal) ; la loi binomiale décrit le nombre de détections parmi un nombre fixé n de particules ou de photons. Le paramètre p s'interprète alors comme une efficacité de détection.

Remarque 3.5.9. La loi binomiale de paramètres $(1, p)$ est simplement la loi de Bernoulli de paramètre p .

Proposition 3.5.10. Si $X \sim \text{Bin}(n, p)$, alors $\mathbb{E}[X] = np$ et $\text{Var}[X] = np(1 - p)$.

Démonstration. La preuve est laissée à titre d'exercice. Ce résultat se démontrera plus facilement plus tard en utilisant le fait qu'une variable aléatoire suivant une loi binomiale peut être vue comme la somme de variables aléatoires *indépendantes* qui suivent une loi de Bernoulli. ■

3.5.3 Loi géométrique

Définition 3.5.11. Une variable aléatoire qui compte le nombre de répétitions indépendantes d'un même schéma de Bernoulli de paramètre p nécessaires pour obtenir un succès est appelée *géométrique* de paramètre p . On note $X \sim \text{Geom}(p)$. La *loi géométrique* est donnée par

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

pour tout $k \in \mathbb{N}$.

Exemple 3.5.12.

- ▷ Nombre de lancers nécessaires pour obtenir “pile” : $X \sim \text{Geom}(1/2)$.
- ▷ Nombre de particules successives qu’il faut envoyer vers un détecteur d’efficacité p avant d’obtenir le premier signal : $X \sim \text{Geom}(p)$.

Proposition 3.5.13. Si $X \sim \text{Geom}(p)$, alors $\mathbb{E}[X] = \frac{1}{p}$ et $\text{Var}[X] = \frac{1-p}{p^2}$.

Démonstration. Pour tout $t \in [0, 1[$, on pose

$$g(t) = \sum_{k=0}^{+\infty} t^k = \frac{1}{1-t}$$

puisqu’il s’agit d’une série géométrique de raison t . Ainsi, d’une part on a

$$g'(t) = \sum_{k=1}^{+\infty} k t^{k-1}$$

et d’autre part

$$g'(t) = \frac{1}{(1-t)^2}.$$

Par conséquent, on obtient

$$\mathbb{E}[X] = \sum_{k \in \mathbb{N}} k(1-p)^{k-1}p = pg'(1-p) = p \frac{1}{(1-(1-p))^2} = \frac{1}{p}.$$

On procède de manière semblable pour calculer $\mathbb{E}[X^2]$ en dérivant g une seconde fois, et cela permet d’obtenir facilement la variance. ■

Remarque 3.5.14. Il existe deux paramétrisations classiques de la loi géométrique. Une autre convention compte le nombre d’échecs avant le premier succès : elle prend ses valeurs dans $\mathbb{N}_0$ et sa fonction de masse est $\mathbb{P}(X = k) = (1-p)^k p$.

Proposition 3.5.15. Soit $X \sim \text{Geom}(p)$.

1. $\mathbb{P}(X > k) = (1-p)^k$
2. La distribution géométrique est “sans mémoire” : pour tous $j, k \geq 1$,

$$\mathbb{P}(X \geq k+j \mid X > k) = \mathbb{P}(X \geq j).$$

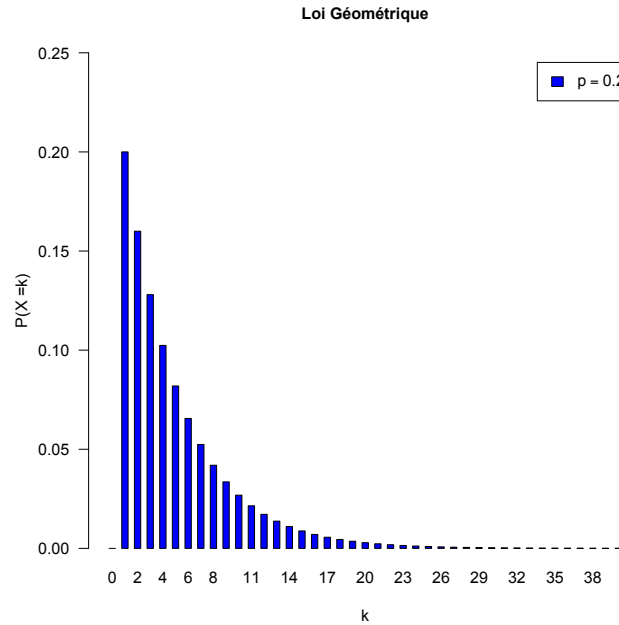


FIGURE 3.6 – Diagramme en bâtons d'une loi géométrique de paramètre 0.2.

Démonstration. 1. On a

$$\begin{aligned}
 \mathbb{P}(X > k) &= 1 - \mathbb{P}(X \leq k) = 1 - \sum_{j=1}^k \mathbb{P}(X = j) = 1 - \sum_{j=1}^k (1-p)^{j-1} p \\
 &= 1 - p \sum_{m=0}^{k-1} (1-p)^m \\
 &= 1 - p \frac{(1-p)^k - 1}{(1-p) - 1} \\
 &= 1 + (1-p)^k - 1 \\
 &= (1-p)^k
 \end{aligned}$$

où on a utilisé la formule donnant la somme des $k-1$ premiers termes d'une suite géométrique.

2. Par définition des probabilités conditionnelles, on a

$$\mathbb{P}(X \geq k+j \mid X > k) = \frac{\mathbb{P}(X \geq k+j \text{ et } X > k)}{\mathbb{P}(X > k)} = \frac{\mathbb{P}(X \geq k+j)}{\mathbb{P}(X > k)}$$

puisque $\{X \geq j+k\} \subseteq \{X > k\}$. Par le premier point et puisque X est à valeurs entières, on trouve que

$$\frac{\mathbb{P}(X \geq k+j)}{\mathbb{P}(X > k)} = \frac{\mathbb{P}(X > k+j-1)}{\mathbb{P}(X > k)} = \frac{(1-p)^{k+j-1}}{(1-p)^k} = (1-p)^{j-1}.$$

En appliquant à nouveau le premier point, on obtient

$$(1-p)^{j-1} = \mathbb{P}(X > j-1) = \mathbb{P}(X \geq j),$$

ce qui donne le résultat attendu. ■

3.5.4 Loi de Poisson

Définition 3.5.16. Une variable aléatoire discrète X suit une *loi de Poisson* de paramètre $\lambda > 0$ si sa fonction de masse est donnée par

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

pour tout $k \in \mathbb{N}_0$. On note $X \sim \text{Pois}(\lambda)$.

La loi de Poisson permet de modéliser le nombre de succès parmi un très grand nombre de répétitions indépendantes de schémas de Bernoulli avec une faible probabilité de succès (événements rares) :

$$\text{Bin}(n, \lambda/n) \approx \text{Pois}(\lambda).$$

Exemple 3.5.17.

- ▷ Nombre de photons détectés pendant une fenêtre temporelle fixée lorsque le flux moyen est stable.
- ▷ Nombre de désintégrations radioactives observées pendant une durée fixée.
- ▷ Plus généralement, nombre d'événements rares indépendants observés dans un intervalle de temps ou d'espace fixé.

Comptage en physique. Une signature importante du modèle de Poisson est l'égalité

$$\mathbb{E}[X] = \text{Var}[X] = \lambda.$$

Dans un dispositif de comptage idéal, la fluctuation typique est donc de l'ordre de $\sqrt{\lambda}$, et l'incertitude relative de comptage est de l'ordre de $1/\sqrt{\lambda}$.

Proposition 3.5.18. Si $X \sim \text{Pois}(\lambda)$, alors $\mathbb{E}[X] = \lambda$ et $\text{Var}[X] = \lambda$.

Démonstration. On a

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

puisque

$$\sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} = \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^\lambda.$$

De même, on calcule

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{k \in \mathbb{N}} e^{-\lambda} \frac{\lambda^k}{k!} k^2 = \lambda \sum_{k=1}^{+\infty} e^{-\lambda} k \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda \sum_{k=0}^{+\infty} e^{-\lambda} (k+1) \frac{\lambda^k}{k!} \\ &= \lambda \left(\sum_{k=0}^{+\infty} e^{-\lambda} k \frac{\lambda^k}{k!} + \sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} \right) \\ &= \lambda(\mathbb{E}[X] + 1) = \lambda(\lambda + 1) \end{aligned}$$

d'où

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

■

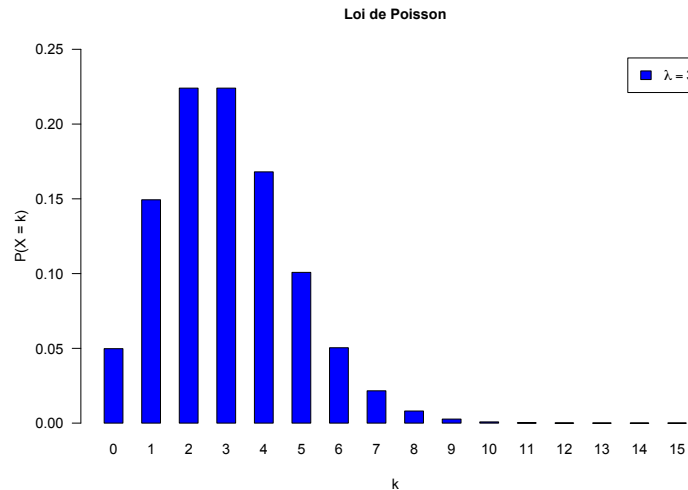


FIGURE 3.7 – Diagramme en bâtons d'une loi de Poisson de paramètre 3.

3.6 Lois continues usuelles

N'importe quelle fonction positive dont l'intégrale sur $\mathbb{R}$ vaut 1 définit une loi de probabilité. Il existe certaines fonctions qui apparaissent un peu partout dans la vraie vie et qui permettent de modéliser un grand nombre de phénomènes. Nous allons en présenter quelques-unes.

3.6.1 Loi continue uniforme

Définition 3.6.1. Soient $a < b$ des réels. Une variable aléatoire continue X suit une *loi uniforme* sur $[a, b]$ si sa densité est constante sur $[a, b]$ et nulle en dehors, c'est-à-dire si

$$f(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x).$$

On note $X \sim \mathcal{U}([a, b])$.

Exemple 3.6.2.

- ▷ Nombre réel choisi au hasard entre 1 et 100 : $X \sim \mathcal{U}([1, 100])$.
- ▷ Endroit de panne sur une route rectiligne.

Proposition 3.6.3. Si $X \sim \mathcal{U}([a, b])$, alors $\mathbb{E}[X] = \frac{a+b}{2}$ et $\text{Var}[X] = \frac{(b-a)^2}{12}$.

Démonstration. On calcule

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{a+b}{2}.$$

De plus,

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2 \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = \frac{1}{b-a} \frac{b^3 - a^3}{3} = \frac{a^2 + ab + b^2}{3}$$

et donc

$$\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2} \right)^2 = \frac{(b-a)^2}{12}.$$

■

3.6.2 Loi exponentielle

Définition 3.6.4. Soit $\lambda > 0$. Une variable aléatoire continue X suit une *loi exponentielle* de paramètre λ si sa densité est donnée par

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0, +\infty[}(x).$$

On note $X \sim \text{Exp}(\lambda)$.

Exemple 3.6.5.

- ▷ Temps d'attente avant la désintégration d'un noyau radioactif dans un modèle à

taux constant.

- ▷ Durée de vie d'un composant soumis à un taux de panne constant.
- ▷ Temps d'attente entre deux événements d'un processus de Poisson homogène.

Demi-vie. Si $T \sim \text{Exp}(\lambda)$ modélise un temps de désintégration, alors

$$\mathbb{P}(T > t) = e^{-\lambda t}.$$

La demi-vie $t_{1/2}$ est définie par $\mathbb{P}(T > t_{1/2}) = 1/2$, d'où

$$t_{1/2} = \frac{\log 2}{\lambda}.$$

Proposition 3.6.6. Si $X \sim \text{Exp}(\lambda)$, alors $\mathbb{E}[X] = \frac{1}{\lambda}$ et $\text{Var}[X] = \frac{1}{\lambda^2}$.

Démonstration. Par définition, on a

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \lambda e^{-\lambda x} \mathbb{1}_{[0, +\infty[}(x) dx = \int_0^{+\infty} x \lambda e^{-\lambda x} dx.$$

En intégrant par parties, il vient

$$\int_0^{+\infty} x \lambda e^{-\lambda x} dx = [-x e^{-\lambda x}]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx = 0 + \left[\frac{-1}{\lambda} e^{-\lambda x} \right]_0^{+\infty} = \frac{1}{\lambda}.$$

On procède de même en intégrant deux fois par parties pour calculer

$$\mathbb{E}[X^2] = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx$$

et on en déduit directement la valeur de la variance. ■

La loi exponentielle jouit de propriétés semblables à celles de la loi géométrique. C'est justifié par le point 3 de la proposition suivante, où la notation $\lceil \cdot \rceil$ fait référence à la partie entière par excès (arrondie vers le haut).

Proposition 3.6.7. Soit $X \sim \text{Exp}(\lambda)$. Pour tous $t, s \geq 0$, on a

- ▷ $\mathbb{P}(X \geq t) = \mathbb{P}(X > t) = e^{-\lambda t}$,
- ▷ la distribution exponentielle est “sans mémoire” :

$$\mathbb{P}(X \geq t + s \mid X > t) = \mathbb{P}(X \geq s),$$

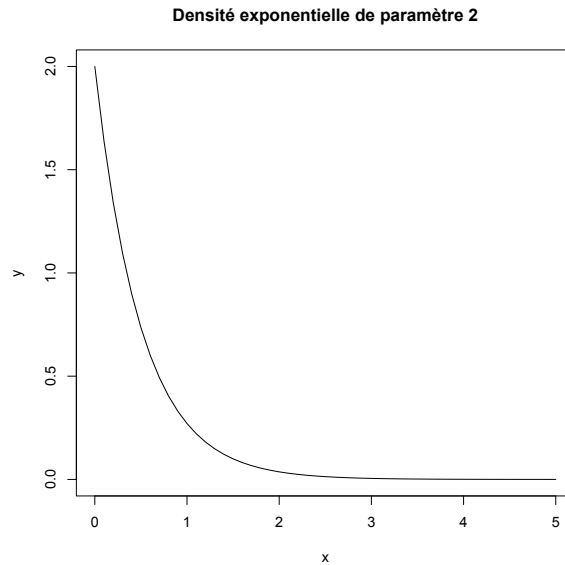


FIGURE 3.8 – Fonction de densité d’une loi exponentielle de paramètre 2.

$$\triangleright \lceil X \rceil \sim \text{Geom}(1 - e^{-\lambda}).$$

Démonstration. 1. On calcule directement que

$$\mathbb{P}(X \geq t) = \mathbb{P}(X > t) = \int_t^{+\infty} \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_t^{+\infty} = e^{-\lambda t}.$$

2. Par définition de la probabilité conditionnelle, on a

$$\begin{aligned} \mathbb{P}(X \geq t + s \mid X > t) &= \frac{\mathbb{P}(X \geq t + s \text{ et } X > t)}{\mathbb{P}(X > t)} = \frac{\mathbb{P}(X \geq t + s)}{\mathbb{P}(X > t)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\ &= e^{-\lambda s} \\ &= \mathbb{P}(X \geq s). \end{aligned}$$

3. Remarquons que $\lceil X \rceil$ prend ses valeurs dans $\mathbb{N} \setminus \{0\}$. De plus, pour tout $k \geq 1$, on a

$$\mathbb{P}(\lceil X \rceil > k) = \mathbb{P}(X > k) = e^{-\lambda k} = (1 - (1 - e^{-\lambda}))^k,$$

ce qui suffit par la Proposition 3.5.15. ■

Remarque 3.6.8. Dans un modèle exponentiel de désintégration radioactive, la propriété “sans mémoire” signifie qu’un noyau qui a déjà survécu pendant un temps t a, à cet instant, la même loi de durée de vie résiduelle qu’au temps initial. Autrement dit, dans ce modèle à taux constant, le noyau ne “vieillit” pas : la probabilité de survivre encore pendant une

durée s ne dépend pas du temps déjà écoulé.

3.6.3 Loi normale

La loi normale (ou gaussienne) est probablement la plus importante de toutes les lois de probabilité. Elle apparaît absolument partout ! Elle est définie via sa fonction de densité comme suit.

Définition 3.6.9. Soient $\mu \in \mathbb{R}$ et $\sigma > 0$. Une variable aléatoire continue X suit une *loi normale* (ou *gaussienne*) de paramètre (μ, σ^2) si sa densité est donnée par

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

On note $X \sim \mathcal{N}(\mu, \sigma^2)$.

Exemple 3.6.10.

- ▷ Erreurs de mesure résultant de l'accumulation de nombreux petits effets indépendants.
- ▷ Fluctuations autour d'une valeur de référence lorsqu'aucun mécanisme particulier ne domine.

Modèle de mesure. Une écriture très fréquente en physique expérimentale est

$$X = \theta + \varepsilon,$$

où θ est la grandeur physique à mesurer et ε l'erreur de mesure. Lorsque $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, on a $X \sim \mathcal{N}(\theta, \sigma^2)$: σ quantifie alors la résolution ou le niveau de bruit de la mesure.

Proposition 3.6.11. Si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors $\mathbb{E}[X] = \mu$ et $\text{Var}[X] = \sigma^2$.

Démonstration. Pour pouvoir prouver ce résultat, il faut se rappeler que

$$\int_{\mathbb{R}} e^{-x^2/2} dx = \sqrt{2\pi}$$

et effectuer le changement de variable $t = \frac{x-\mu}{\sigma}$. Ce calcul est laissé en exercice. ■

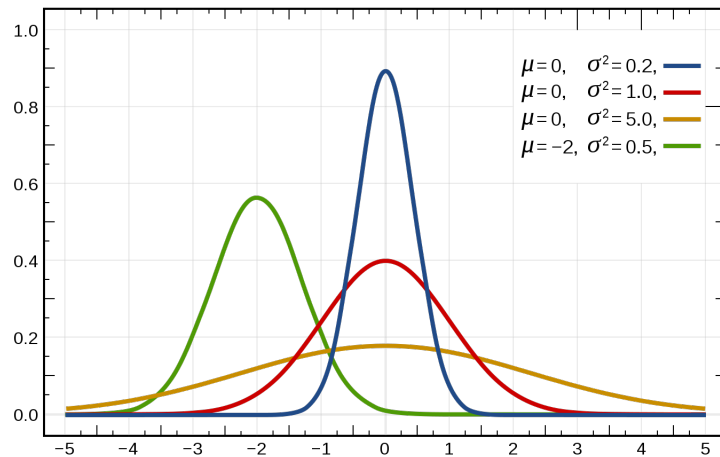


FIGURE 3.9 – Quelques représentations de densités de lois normales pour des paramètres différents. On a une courbe en “cloche”, symétrique par rapport à l’axe $x = \mu$. Plus σ est grand, plus la courbe est “aplatie”.

Définition 3.6.12. La loi normale $\mathcal{N}(0, 1)$ est appelée la loi normale *centrée et réduite* (ou loi normale *standard*).

On ne peut pas calculer les probabilités associées à une loi normale “à la main”.

En effet, même si on prend la loi normale centrée et réduite, sa fonction de répartition n’admet pas de forme fermée ; on la note

$$\Phi(x) = \int_{-\infty}^x \varphi(u) du.$$

Cette fonction n’est calculable que par intégration numérique. On utilisera les *tables de la loi normale* (cfr. Figure 3.10) ou un logiciel de calcul.

On utilise énormément les propriétés de symétrie des lois normales, qui découlent toutes de l’identité $\varphi(x) = \varphi(-x)$ pour tout $x \in \mathbb{R}$.

Proposition 3.6.13. Si $X \sim \mathcal{N}(0, 1)$ alors pour tout $x \geq 0$, on a

1. $\mathbb{P}(X < -x) = \mathbb{P}(X > x)$ et en particulier $\mathbb{P}(X > 0) = \mathbb{P}(X < 0) = \frac{1}{2}$,
2. $\Phi(-x) = 1 - \Phi(x)$,
3. $\mathbb{P}(|X| > x) = 2\mathbb{P}(X > x)$.

Démonstration. Exercice. ■

Remarquons que les propriétés de l’espérance et de la variance donnent

$$\mathbb{E} \left[\frac{X - \mu}{\sigma} \right] = \frac{1}{\sigma} \mathbb{E}[X - \mu] = \frac{1}{\sigma} (\mathbb{E}[X] - \mu) = 0$$

et

$$\text{Var} \left[\frac{X - \mu}{\sigma} \right] = \frac{1}{\sigma^2} \text{Var}[X - \mu] = \frac{1}{\sigma^2} \text{Var}[X] = 1.$$

Proposition 3.6.14. Si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors

$$\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

Démonstration. On calcule la fonction de répartition de $\frac{X - \mu}{\sigma}$. Pour tout $x \in \mathbb{R}$, on a

$$\begin{aligned} \mathbb{P}\left(\frac{X - \mu}{\sigma} \leq x\right) &= \mathbb{P}(X \leq \sigma x + \mu) \\ &= \int_{-\infty}^{\sigma x + \mu} \frac{1}{\sqrt{2\pi}\sigma} e^{-(y - \mu)^2 / 2\sigma^2} dy \\ &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2 / 2} dt \end{aligned}$$

où on a effectué le changement de variables $t = \frac{y - \mu}{\sigma}$. ■

Ainsi, lors de calcul de probabilités liées à une loi normale, on se ramènera toujours à la loi standard, ce qui revient à centrer et réduire la variable aléatoire.

Exemple 3.6.15. Soit $X \sim \mathcal{N}(300, 20^2)$. On a

$$\begin{aligned} \mathbb{P}(295 \leq X \leq 320) &= \mathbb{P}\left(\frac{295 - 300}{20} \leq \frac{X - 300}{20} \leq \frac{320 - 300}{20}\right) \\ &= \mathbb{P}\left(-\frac{1}{4} \leq Z \leq 1\right) \\ &= \mathbb{P}(Z \leq 1) - \mathbb{P}\left(Z \leq -\frac{1}{4}\right) \end{aligned}$$

où $Z \sim \mathcal{N}(0, 1)$. Pour ce calcul, on utilise un logiciel ou des tables de probabilités.

Terminons par quelques valeurs remarquables de la loi normale. Si $X \sim \mathcal{N}(\mu, \sigma^2)$, on calcule

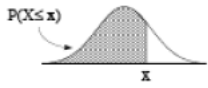
$$\begin{aligned} \mathbb{P}(|X - \mu| \leq r\sigma) &= \mathbb{P}(-r\sigma \leq X - \mu \leq r\sigma) \\ &= \mathbb{P}\left(-r \leq \underbrace{\frac{X - \mu}{\sigma}}_{Z \sim \mathcal{N}(0, 1)} \leq r\right) \\ &= \mathbb{P}(|Z| \leq r) \end{aligned}$$

et on remarque que ces probabilités sont les mêmes pour toutes les lois normales. En particulier,

$$\begin{aligned} \mathbb{P}(|X - \mu| \leq \sigma) &\approx 0.68 \\ \mathbb{P}(|X - \mu| \leq 2\sigma) &\approx 0.95 \\ \mathbb{P}(|X - \mu| \leq 3\sigma) &\approx 0.997 \end{aligned}$$

Ce résultat nous donne les probabilités que X prenne ses valeurs dans l'intervalle $[\mu - r\sigma, \mu + r\sigma]$

Fournit la probabilité $P(X \leq x)$
pour $X \sim N(0,1)$



x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0,0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0,1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0,2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0,3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0,4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0,5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0,6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0,7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0,8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0,9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1,0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1,1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1,2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1,3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1,4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1,5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1,6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1,7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1,8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1,9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2,0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2,1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2,2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2,3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2,4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2,5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2,6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2,7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2,8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2,9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3,0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3,1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3,2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3,3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3,4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

FIGURE 3.10 – Table de probabilités d’une loi normale centrée et réduite.

de manière plus précise que les estimations obtenues par l’inégalité de Tchebychev.

3.6.4 Deux lois dérivées de la Gaussienne

Les deux lois présentées ci-dessous font intervenir la notion d’indépendance de variables aléatoires. Nous y reviendrons plus tard.

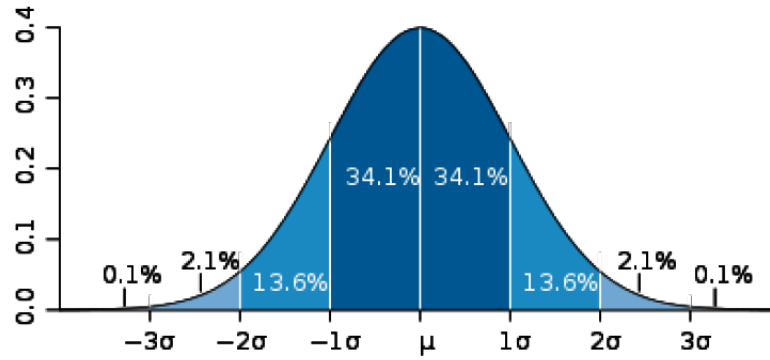
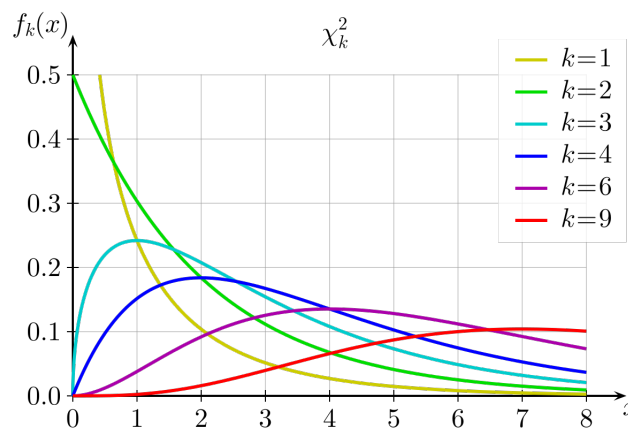
▷ **Loi χ^2 à ν degrés de liberté :** $X \sim \chi_\nu^2$ si

$$X = \sum_{i=1}^{\nu} Z_i^2$$

où les Z_i sont des $\mathcal{N}(0,1)$ indépendantes. Dans ce cas, on peut montrer que

$$\triangleright f_X(x) = \frac{1}{2^{\nu/2} \Gamma(\nu/2)} e^{-\frac{x}{2}} x^{(\nu-2)/2} \mathbb{1}_{[0,+\infty[}(x)$$

$$\triangleright \mathbb{E}[X] = \nu \text{ et } \text{Var}[X] = 2\nu.$$

FIGURE 3.11 – Quelques valeurs remarquables de la loi normale $\mathcal{N}(\mu, \sigma^2)$.FIGURE 3.12 – Fonctions de densités de variables aléatoires de loi χ^2 pour différents degrés de liberté. On remarque que plus le degré de liberté augmente, plus la dispersion augmente et l'asymétrie diminue.

▷ **Loi de Student à ν degrés de liberté** : $X \sim t_\nu$ si

$$X = \frac{Z}{\sqrt{Y/\nu}}$$

avec $Z \sim \mathcal{N}(0, 1)$ indépendant de $Y \sim \chi_\nu^2$. Dans ce cas, on peut montrer que

$$\triangleright f_X(x) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}$$

▷ $\mathbb{E}[X] = 0$ si $\nu > 1$ et $\text{Var}[X] = 1 + \frac{2}{\nu-2}$ si $\nu > 2$. Dans les autres cas, ces quantités n'existent pas.

3.7 Transformation de variables aléatoires

Comme déjà signalé auparavant, on est parfois amené à considérer une nouvelle variable aléatoire qui est une fonction $h(X)$ d'une variable aléatoire connue. C'est le cas par exemple lorsque l'on change les unités considérées (ce qui correspond à une variable aléatoire du type αX) ou qu'on

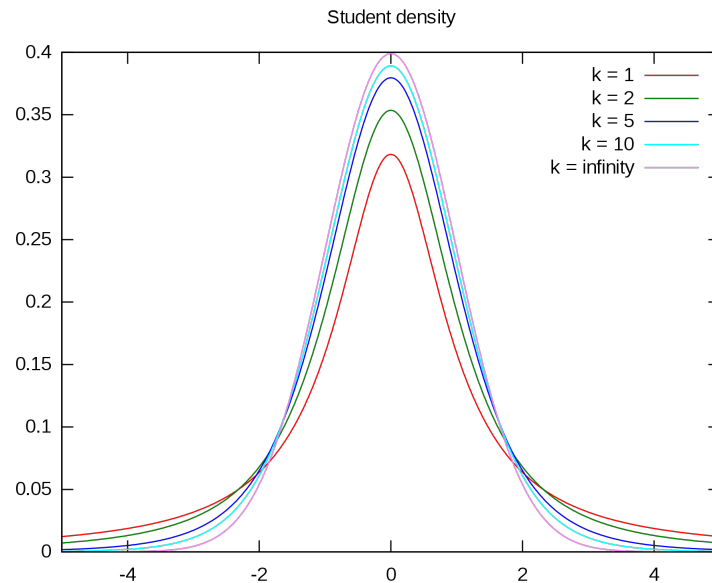


FIGURE 3.13 – Fonctions de densités de variables aléatoires de loi de student pour différents degrés de liberté. On remarque que ces densités ont plus de masses dans les queues de distribution que la loi normale standard, et peuvent donc prendre des valeurs plus extrêmes avec des probabilités plus grandes.

change d'échelle de position (variable aléatoire du type $\alpha + X$).

Un second exemple est donné par l'aire d'un cercle dont le rayon est choisi aléatoirement : Soit R une variable aléatoire positive représentant le rayon d'un cercle. L'aire A du cercle est alors donnée par la variable aléatoire $A = \pi R^2$.

On pourrait aussi imaginer qu'un observateur se trouve à une distance D aléatoire d'une source lumineuse. L'intensité lumineuse I perçue par l'observateur est inversement proportionnelle au carré de la distance. Plus précisément, I est la variable aléatoire donnée par $I = \frac{P}{4\pi D^2}$ où P est la puissance émise par la source lumineuse (supposée constante).

Nous avons déjà vu dans la Proposition 3.4.5 comment calculer l'espérance de Y à partir de la distribution de X , sans avoir à connaître la distribution de Y . Néanmoins, on peut parfois avoir besoin de comprendre plus précisément la loi de Y .

Cas discret : Dans le cas discret, on peut obtenir la fonction de masse de $h(X)$ en recherchant toutes les valeurs x prises par X telles que $h(x) = y$. Cela donne la formule suivante :

$$\mathbb{P}(h(X) = y) = \sum_{x: h(x)=y} \mathbb{P}(X = x).$$

Pour une fonction h bijective, la situation est particulièrement simple, car il n'existe qu'une seule valeur de x telle que $h(x) = y$, à savoir $h^{-1}(y)$. On peut alors utiliser :

$$\mathbb{P}(h(X) = y) = \mathbb{P}(X = h^{-1}(y))$$

pour obtenir la fonction de masse de $h(X)$ à partir de celle de X .

Exemple 3.7.1. Imaginons que deux équipes de basketball (A et B) jouent sept matchs. Soit X le nombre de victoires de l'équipe A (donc $X \sim \text{Bin}(7, \frac{1}{2})$ si les équipes sont équilibrées et les jeux sont indépendants). Considérons les fonctions

$$g(x) = 7 - x$$

et

$$h(x) = \begin{cases} 1 & \text{si } x \geq 4 \\ 0 & \text{si } x < 4. \end{cases}$$

Alors $g(X) = 7 - X$ est le nombre de victoires de l'équipe B, et $h(X)$ vaut 1 si et seulement si l'équipe A a gagné la majorité des matchs. Pour $g(X)$, on remarque que $g(X) = k$ si et seulement si $X = 7 - k$. Ainsi, $g(X)$ prend les valeurs $\{0, \dots, 7\}$ et

$$\mathbb{P}(g(X) = k) = \mathbb{P}(X = 7 - k) = C_7^{7-k} \left(\frac{1}{2}\right)^{7-k} \left(\frac{1}{2}\right)^k = C_7^k \left(\frac{1}{2}\right)^{7-k} \left(\frac{1}{2}\right)^k$$

vu les propriétés du coefficient binomial. Ainsi, $g(X) \sim \text{Bin}(7, \frac{1}{2})$ (remarquons que les variables aléatoires X et $g(X)$ ont la même loi mais ne sont égales nulle part). Pour $h(X)$, on remarque que

$$\{h(X) = 1\} = \{X = 4\} \cup \{X = 5\} \cup \{X = 6\} \cup \{X = 7\}$$

et donc

$$\mathbb{P}(h(X) = 1) = \mathbb{P}(X = 4) + \mathbb{P}(X = 5) + \mathbb{P}(X = 6) + \mathbb{P}(X = 7) = \mathbb{P}(X \geq 4).$$

Puisque $X \sim \text{Bin}(7, \frac{1}{2})$, nous pouvons calculer cette probabilité et on trouve $\mathbb{P}(h(X) = 1) = \mathbb{P}(X \geq 4) = \frac{1}{2}$. On en tire que $\mathbb{P}(h(X) = 0) = 1 - \mathbb{P}(h(X) = 1) = \frac{1}{2}$ et que $h(X) \sim \text{Bern}(\frac{1}{2})$ (on aurait pu le deviner).

Cas continu : Dans le cas continu, on ne peut plus utiliser le raisonnement précédent puisque $\mathbb{P}(X = x) = 0$ pour tout $x \in \mathbb{R}$. On va donc partir de la fonction de répartition X pour trouver celle de $h(X)$, ce qui donnera la loi de $h(X)$. Il faut donc traduire l'événement $h(X) \leq y$ en un événement équivalent impliquant X . Pour une fonction h générale, il n'existe pas de formule simple que l'on puisse utiliser. Mais lorsque h est strictement croissante et continue, la traduction est facile : $h(X) \leq y$ est équivalent à $X \leq h^{-1}(y)$, donc :

$$F_{h(X)}(y) = \mathbb{P}(h(X) \leq y) = \mathbb{P}(X \leq h^{-1}(y)) = F_X(h^{-1}(y)).$$

En dérivant $F_{h(X)}$ par rapport à y , on obtient alors la fonction de densité de $h(X)$.

Exemple 3.7.2. Soient $X \sim \mathcal{N}(0, 1)$ et $Y = e^X$. On dit que Y suit une loi *log-normale*.

On considère ici la fonction $h(x) = e^x$. On a $F_Y(y) = 0$ pour $y \leq 0$. Pour $y > 0$,

$$F_Y(y) = \mathbb{P}(e^X \leq y) = \mathbb{P}(X \leq \log(y)) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\log(y)} e^{-\frac{x^2}{2}} dx$$

et en effectuant le changement de variables $t = e^x$, il vient

$$F_Y(y) = \frac{1}{\sqrt{2\pi}} \int_0^y \frac{1}{t} e^{-\frac{\log^2(t)}{2}} dt.$$

On en tire que la fonction de densité de Y est donnée par

$$f_Y(y) = \frac{1}{y\sqrt{2\pi}} e^{-\frac{\log^2(y)}{2}} \mathbb{1}_{]0, +\infty[}(y).$$

Dans l'exemple précédent, nous avons utilisé la formule du changement de variables. En fait, cette formule donne un résultat général.

Proposition 3.7.3 (Changement de variables). *Soient X une variable aléatoire continue de fonction de densité f_X et h un changement de variables^a. Si $Y = h(X)$, alors Y est une variable aléatoire continue de fonction de densité*

$$f_Y(y) = \frac{f_X(h^{-1}(y))}{|h'(h^{-1}(y))|}.$$

^a. On demande en plus de la monotonie stricte que la fonction soit de classe C^1 et que sa dérivée ne s'annule en aucun point.

Démonstration. Soit B un ensemble borélien de $\mathbb{R}$. En utilisant la formule de changement de variables dans $\mathbb{R}$ ³, on a

$$\int_B \frac{f_X(h^{-1}(y))}{|h'(h^{-1}(y))|} dy = \int_{h^{-1}(B)} f_X(x) dx = \mathbb{P}(X \in h^{-1}(B)) = \mathbb{P}(h(X) \in B) = \mathbb{P}(Y \in B)$$

en utilisant la formule

$$(h^{-1})'(y) = \frac{1}{h'(h^{-1}(y))}$$

qui donne la dérivée de h^{-1} . Cela implique la conclusion. ■

Lorsque la fonction g n'est pas monotone et ne définit pas un changement de variables, on travaille au cas par cas.

Exemple 3.7.4. Soit $X \sim \mathcal{N}(0, 1)$.

3. La valeur absolue provient juste du fait qu'on ne change pas l'ordre des bornes d'intégration.

▷ On calcule la loi de $|X|$ en remarquant que $\mathbb{P}(|X| \leq x) = 0$ si $x < 0$ et, pour $x \geq 0$,

$$\begin{aligned}\mathbb{P}(|X| \leq x) &= \mathbb{P}(-x \leq X \leq x) = 2\mathbb{P}(X \leq x) - 1 \\ &= 2 \int_{-\infty}^x f_X(t) dt - 1 \\ &= 2 \int_{-\infty}^0 f_X(t) dt + 2 \int_0^x f_X(t) dt - 1 \\ &= 2 \int_0^x f_X(t) dt\end{aligned}$$

puisque la fonction f_X est paire et d'intégrale égale à 1. Donc

$$f_{|X|}(x) = 2f_X(x)\mathbb{1}_{[0,+\infty[}(x) = \frac{2}{\sqrt{2\pi}}e^{-x^2/2}\mathbb{1}_{[0,+\infty[}(x).$$

▷ La loi de X^2 est obtenue en remarquant que $\mathbb{P}(X^2 \leq x) = 0$ si $x < 0$ et, pour $x \geq 0$, on a

$$\mathbb{P}(X^2 \leq x) = \mathbb{P}(-\sqrt{x} \leq X \leq \sqrt{x}) = 2\mathbb{P}(X \leq \sqrt{x}) - 1.$$

Donc

$$f_{X^2}(x) = \frac{f_X(\sqrt{x})}{\sqrt{x}}\mathbb{1}_{[0,+\infty[}(x) = \frac{1}{\sqrt{2\pi x}}e^{-x/2}\mathbb{1}_{[0,+\infty[}(x).$$

Cette loi est la loi χ^2 à 1 degré de liberté.

4 Lois jointes

Des variables aléatoires $X_1, \dots, X_d$ définies sur un même espace probabilisé (liées à une même expérience) peuvent avoir des liens entre elles. On parle de *vecteur aléatoire* $(X_1, \dots, X_d)$. Les outils présentés précédemment s'appliquent à chacune de ces variables aléatoires mais ne permet pas de capturer ces liens éventuels. Afin de capturer cette information, on étudie la distribution jointe des variables aléatoires. Cela nous mènera en particulier à la notion d'indépendance de variables aléatoires.

Pour simplifier les choses, nous nous restreindrons à la dimension $d = 2$. On parle alors de *couple aléatoire* ou de vecteur aléatoire *bivarié*. On suppose donc avoir deux variables aléatoires X et Y définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ et on s'intéresse aux valeurs prises par le couple (X, Y) . On va distinguer trois cas : le cas discret, le cas continu et le cas mixte. On va s'intéresser à

- ▷ la distribution *jointe* de (X, Y) qui donne une description des probabilités que (X, Y) appartienne à un sous-ensemble de $\mathbb{R}^2$
- ▷ la distribution *marginale* de X qui est la distribution de X (en ignorant Y)
- ▷ la distribution *conditionnelle* de X étant donné $Y = y$, qui est la distribution de X sous l'information supplémentaire que $Y = y$.

4.1 Loi jointe et fonction de répartition

On suppose donc avoir deux variables aléatoires X et Y définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ et on s'intéresse aux valeurs prises par le couple (X, Y) . L'ensemble des valeurs prises par le couple (X, Y) est *inclus* dans l'ensemble

$$\{(X(\omega), Y(\omega)) : \omega \in \Omega\}$$

et on s'intéresse aux probabilités de chacune de ces valeurs. La distribution de (X, Y) est donnée par sa loi.

Définition 4.1.1. La *loi de* (X, Y) (ou *distribution jointe de* (X, Y)) est l'application définie par

$$\mathbb{P}_{(X,Y)} : \mathcal{B}^2 \rightarrow [0, 1] : B \mapsto \mathbb{P}((X, Y) \in B)$$

où $\mathcal{B}^2$ est la σ -algèbre de Borel sur $\mathbb{R}^2$ (la plus petite σ -algèbre qui contient les intervalles $[a, b] \times [c, d]$).

Remarquons que si $I = [a, b] \times [c, d]$, alors

$$\mathbb{P}((X, Y) \in I) = \mathbb{P}(a \leq X \leq b, c \leq Y \leq d).$$

Comme dans le cas univarié, on va s'intéresser aux intervalles de la forme

$$]-\infty, x] \times]-\infty, y]$$

avec $x, y \in \mathbb{R}$.

Définition 4.1.2. La *fonction de répartition* de (X, Y) est la fonction $F_{(X,Y)}$ définie pour tout $(x, y) \in \mathbb{R}^2$ par

$$F_{(X,Y)}(x, y) = \mathbb{P}(X \leq x, Y \leq y).$$

Evidemment, la loi détermine entièrement la fonction de répartition. Réciproquement, si on connaît la fonction de répartition, on peut retrouver la loi. En particulier, on remarque que

$$\begin{aligned} \mathbb{P}((X, Y) \in]a, b] \times]c, d]) &= \mathbb{P}(a < X \leq b, c < Y \leq d) \\ &= F_{X,Y}(b, d) - F_{X,Y}(a, d) - F_{X,Y}(b, c) + F_{X,Y}(a, c). \end{aligned}$$

On spécifiera donc la distribution d'un couple aléatoire (X, Y) en spécifiant sa loi ou sa fonction de répartition.

4.2 Lois bivariées discrètes

Dans cette section, on suppose avoir deux variables aléatoires discrètes X et Y définies sur le même espace probabilisé. Notons

▷ $x_i, i \in I$, les valeurs possibles de X

▷ $y_j, j \in J$, les valeurs possibles de Y

avec $I, J \subset \mathbb{N}$. L'ensemble des valeurs possibles du couple aléatoire (X, Y) est alors *inclus* dans

$$\{(x_i, y_j) : i \in I, j \in J\}.$$

Définition 4.2.1. La *fonction de masse* du couple aléatoire discret (X, Y) est la fonction $p_{(X,Y)}$ définie par

$$p_{(X,Y)}(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j)$$

pour tous $i \in I, j \in J$.

Pour simplifier les notations lorsque le contexte est clair, on écrit

$$p_{ij} := p_{(X,Y)}(x_i, y_j).$$

Lorsque les ensembles des valeurs prises par X et Y respectivement sont finis, la loi de probabilité

jointe est transcrite dans un tableau de contingence

	x_1	x_2	$\dots$	x_{N_1}
y_1	p_{11}	p_{21}	$\dots$	$p_{N_1 1}$
y_2	p_{12}	p_{22}	$\dots$	$p_{N_1 2}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
y_{N_2}	p_{1N_2}	p_{2N_2}	$\dots$	$p_{N_1 N_2}$

La fonction de répartition, et donc la loi, de (X, Y) est complètement caractérisée par la fonction de masse puisque

$$\mathbb{P}(X \leq x, Y \leq y) = \sum_{i \in I: x_i \leq x} \sum_{j \in J: y_j \leq y} p_{ij}.$$

En particulier, puisque $\mathbb{P}((X, Y) \in \mathbb{R}^2) = 1$, on trouve

$$\sum_{i \in I} \sum_{j \in J} p_{ij} = 1.$$

La fonction de répartition et la fonction de masse du vecteur bivarié discret donnent donc des descriptions équivalentes de sa loi ; on spécifiera le couple aléatoire en spécifiant l'une des deux. La fonction de répartition n'est toutefois en général pas simple à exprimer et est généralement ignorée pour les couples discrets.

Exemple 4.2.2. Considérons l'expérience consistant à lancer deux dés (distinguables). Notons X le résultat du premier dé et Y le résultat du deuxième dé. L'ensemble des valeurs prises par le couple (X, Y) est donné par

$$\{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (6, 6)\}.$$

La loi du vecteur (X, Y) est donnée par

X	1	2	3	4	5	6	
Y							
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Exemple 4.2.3. On peut facilement construire un autre vecteur aléatoire bivarié possédant les mêmes entrées, par exemple en lançant un dé une seule fois, en prenant X le résultat et $Y = 7 - X$. Alors le couple (X, Y) a la loi jointe donnée dans le tableau

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

Exemple 4.2.4. On suppose qu'on lance deux dés (distinguables). Notons X la somme des résultats des deux dés et Y la valeur absolue de la différence des résultats de chaque dé.

La loi du vecteur (X, Y) est donnée dans le tableau

X	2	3	4	5	6	7	8	9	10	11	12	
Y												
0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	$\frac{6}{36}$
1	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{10}{36}$
2	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	$\frac{8}{36}$
3	0	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	0	$\frac{6}{36}$
4	0	0	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	0	0	$\frac{4}{36}$
5	0	0	0	0	0	$\frac{2}{36}$	0	0	0	0	0	$\frac{2}{36}$
	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

Dans les exemples précédents, nous avons rapporté dans les marges du tableau les distributions de X et Y . On les appelle les distributions *marginales*. Elles s'obtiennent par projection comme le montre le résultat suivant.

Proposition 4.2.5. *Soit (X, Y) un couple aléatoire discret. Les fonctions de masses marginales de X et Y sont données par*

$$p_X(x_i) = \mathbb{P}(X = x_i) = \sum_{j \in J} p_{ij} =: p_{i\bullet} \quad \forall i \in I$$

et

$$p_Y(y_j) = \mathbb{P}(Y = y_j) = \sum_{i \in I} p_{ij} =: p_{\bullet j} \quad \forall j \in J.$$

Démonstration. On démontre le résultat pour X , celui pour Y s'obtient de manière similaire. Fixons $i \in I$. Puisque Y prend ses valeurs dans $\{y_j : j \in J\}$, on a

$$\begin{aligned}
 p_X(x_i) = \mathbb{P}(X = x_i) &= \mathbb{P}(X = x_i \text{ et il existe } j \in J \text{ tel que } Y = y_j) \\
 &= \mathbb{P}\left(\bigcup_{j \in J} \{X = x_i, Y = y_j\}\right) \\
 &= \sum_{j \in J} \mathbb{P}(X = x_i, Y = y_j) \\
 &= \sum_{j \in J} p_{ij}
 \end{aligned}$$

puisque les événements $\{X = x_i, Y = y_j\}$ avec $j \in J$ sont deux à deux disjoints. ■

On résume en général la loi du vecteur aléatoire bivarié (X, Y) et de ses marginales via le tableau

de contingence complet

	x_1	x_2	$\dots$	x_{N_1}	
y_1	p_{11}	p_{21}	$\dots$	$p_{N_1 1}$	$p_{\bullet 1}$
y_2	p_{12}	p_{22}	$\dots$	$p_{N_1 2}$	$p_{\bullet 2}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
y_{N_2}	p_{1N_2}	p_{2N_2}	$\dots$	$p_{N_1 N_2}$	$p_{\bullet N_2}$
	$p_{1\bullet}$	$p_{2\bullet}$	$\dots$	$p_{N_1\bullet}$	1

Dans le chapitre 1, nous avons vu que connaître la réalisation d'un événement permettait en général d'affiner la probabilité de réalisation d'un autre événement. Ici, on considère une situation où on connaît la valeur qu'a prise l'une des variables aléatoires et on est désireux de savoir si cette information permet d'affiner la probabilité que l'autre variable prenne sa valeur dans une région donnée (et par suite la distribution de l'autre variable), ou l'espérance ou la variance de l'autre variable. Au contraire des distributions marginales, les distributions conditionnelles permettent d'appréhender le lien entre X et Y .

Supposons à présent que nous avons observé la valeur x_i de X . Cette information peut nous permettre d'affiner notre connaissance sur les valeurs prises par Y ainsi que leurs probabilités. On s'intéresse donc à des probabilités conditionnelles.

Définition 4.2.6. Soit (X, Y) un couple aléatoire discret. La *fonction de masse conditionnelle* de Y sachant que $X = x_i$ est la fonction $p_{Y|X=x_i}$ donnée par

$$p_{Y|X=x_i}(y_j) = \mathbb{P}(Y = y_j | X = x_i) = \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(X = x_i)} = \frac{p_{ij}}{p_{i\bullet}}$$

pour tout $j \in J$.

On a donc défini une nouvelle variable aléatoire $Y|X = x_i$ dont la fonction de masse est décrite dans le tableau ci-dessous.

valeurs possibles	y_1	y_2	$\dots$	y_{N_2}
probabilités	$\frac{p_{i1}}{p_{i\bullet}}$	$\frac{p_{i2}}{p_{i\bullet}}$	$\dots$	$\frac{p_{iN_2}}{p_{i\bullet}}$

Remarquons qu'on a bien

$$\sum_{j \in J} \mathbb{P}(Y = y_j | X = x_i) = \sum_{j \in J} \frac{p_{ij}}{p_{i\bullet}} = \frac{1}{p_{i\bullet}} \sum_{j \in J} p_{ij} = 1$$

par la Proposition 4.2.5.

Comme pour toute distribution discrète univariée, on peut calculer l'espérance et la variance de Y sachant que $X = x_i$. Elles seront dites *conditionnelles*. On a naturellement

$$\mathbb{E}[Y | X = x_i] = \sum_{j=1}^{N_2} y_j \frac{p_{ij}}{p_{i\bullet}}$$

et

$$\text{Var}[Y|X = x_i] = \sum_{j=1}^{N_2} (y_j - \mathbb{E}[Y|X = x_i])^2 \frac{p_{ij}}{p_{i\bullet}} = \sum_{j=1}^{N_2} y_j^2 \frac{p_{ij}}{p_{i\bullet}} - (\mathbb{E}[Y|X = x_i])^2.$$

Exemple 4.2.7. Reprenons l'exemple 4.2.2 où X est le résultat du premier dé et Y le résultat du deuxième dé. On peut par exemple supposer que $X = 1$. Intuitivement, il est clair que le résultat du premier dé n'a pas d'influence sur le résultat du deuxième dé et, par conséquent, la variable Y doit prendre chacune des valeurs entre 1 et 6 avec une probabilité $1/6$. On a en effet

$$\mathbb{P}(Y = j|X = 1) = \frac{p_{1j}}{p_{1\bullet}} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$$

pour tout $j \in \{1, \dots, 6\}$.

Exemple 4.2.8. Reprenons à présent l'exemple 4.2.3 où X est le résultat du premier dé et $Y = 7 - X$. Contrairement à l'exemple précédent, la connaissance de X détermine entièrement la valeur prise par Y . Par exemple, si on suppose que $X = 1$ alors Y doit prendre la valeur 6 avec une probabilité 1. On a en effet

$$\mathbb{P}(Y = 6|X = 1) = \frac{p_{16}}{p_{1\bullet}} = \frac{\frac{1}{6}}{\frac{1}{6}} = 1$$

et

$$\mathbb{P}(Y = j|X = 1) = \frac{p_{1j}}{p_{1\bullet}} = \frac{0}{\frac{1}{6}} = 0$$

pour tout $j \in \{1, \dots, 5\}$.

Les deux exemples précédents sont extrêmes : dans le premier cas, la valeur prise par X n'a aucun impact sur Y , alors que dans le deuxième cas, cette valeur détermine entièrement Y . L'exemple suivant montre un cas intermédiaire.

Exemple 4.2.9. Dans l'exemple 4.2.4, X est le résultat de la somme des résultats de deux dés et Y est la valeur absolue de leur différence. La $i^{\text{ème}}$ colonne du tableau suivant donne la distribution conditionnelle de Y sachant que $X = i$:

	2	3	4	5	6	7	8	9	10	11	12
0	1		$\frac{1}{3}$		$\frac{1}{5}$		$\frac{1}{5}$		$\frac{1}{3}$		1
1		1		$\frac{1}{2}$		$\frac{1}{3}$		$\frac{1}{2}$		1	
2			$\frac{2}{3}$		$\frac{2}{5}$		$\frac{2}{5}$		$\frac{2}{3}$		
3				$\frac{1}{2}$		$\frac{1}{3}$		$\frac{1}{2}$			
4					$\frac{2}{5}$		$\frac{2}{5}$				
5						$\frac{1}{3}$					
	1	1	1	1	1	1	1	1	1	1	1

Remarquons que sachant que $\{X = 3\} = \{(1, 2), (2, 1)\}$, le résultat de l'expérience est encore aléatoire, mais plus Y ! Si on calcule en plus les espérances et les variances conditionnelles, on obtient :

	2	3	4	5	6	7	8	9	10	11	12
0	1		$\frac{1}{3}$		$\frac{1}{5}$		$\frac{1}{5}$		$\frac{1}{3}$		1
1		1		$\frac{1}{2}$		$\frac{1}{3}$		$\frac{1}{2}$		1	
2			$\frac{2}{3}$		$\frac{2}{5}$		$\frac{2}{5}$		$\frac{2}{3}$		
3				$\frac{1}{2}$		$\frac{1}{3}$		$\frac{1}{2}$			
4					$\frac{2}{5}$		$\frac{2}{5}$				
5						$\frac{1}{3}$					
	1	1	1	1	1	1	1	1	1	1	1
$\mathbb{E}$	0	1	1.33	2	2.4	3	2.4	2	1.33	1	0
Var	0	0	0.89	1	2.24	2.67	2.24	1	0.89	0	0

Les fonctions $x_i \mapsto \mathbb{E}[Y|X = x_i]$ et $x_i \mapsto \text{Var}[Y|X = x_i]$ portent souvent une information importante sur le lien entre X et Y .

Jusqu'ici, on a supposé que X prenait une certaine valeur fixée x_i . Si on rend à X son caractère aléatoire, on obtient deux nouvelles variables aléatoires. En effet, pour chaque valeur x_i prise par X , on trouve une espérance conditionnelle $\mathbb{E}[Y|X = x_i]$. Cette valeur peut être vue comme un résultat aléatoire qui dépend du résultat de X . Autrement dit, on a une nouvelle variable aléatoire qui prend la valeur $\mathbb{E}[Y|X = x_i]$ avec une probabilité $\mathbb{P}(X = x_i)$.

Définition 4.2.10. L'espérance conditionnelle $\mathbb{E}[Y|X]$ de Y sachant X est la variable aléatoire dont la fonction de masse est donnée par

valeurs possibles	$\mathbb{E}[Y X = x_1]$	$\mathbb{E}[Y X = x_2]$	$\dots$	$\mathbb{E}[Y X = x_{N_1}]$
probabilités	$p_{1\bullet} = \mathbb{P}(X = x_1)$	$p_{2\bullet} = \mathbb{P}(X = x_2)$	$\dots$	$p_{N_1\bullet} = \mathbb{P}(X = x_{N_1})$

La *variance conditionnelle* $\text{Var}[Y|X]$ de Y sachant X est la variable aléatoire dont la fonction de masse est donnée par

valeurs possibles	$\text{Var}[Y X = x_1]$	$\text{Var}[Y X = x_2]$	$\dots$	$\text{Var}[Y X = x_{N_1}]$
probabilités	$p_{1\bullet} = \mathbb{P}(X = x_1)$	$p_{2\bullet} = \mathbb{P}(X = x_2)$	$\dots$	$p_{N_1\bullet} = \mathbb{P}(X = x_{N_1})$

Le résultat suivant montre que la moyenne des valeurs prises par la variable aléatoire *espérance conditionnelle* est exactement l'espérance de Y .

Théorème 4.2.11 (Théorème de l'espérance totale). *Si X et Y sont des variables aléatoires discrètes qui admettent une espérance, on a*

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

Autrement dit,

$$\mathbb{E}[Y] = \sum_{i \in I} \mathbb{E}[Y|X = x_i] \mathbb{P}(X = x_i).$$

Démonstration. On a

$$\begin{aligned}
 \mathbb{E}[\mathbb{E}[Y|X]] &= \sum_{i \in I} \mathbb{E}[Y|X = x_i] \mathbb{P}(X = x_i) = \sum_{i \in I} \sum_{j \in J} y_j \mathbb{P}(Y = y_j | X = x_i) \mathbb{P}(X = x_i) \\
 &= \sum_{j \in J} y_j \left(\sum_{i \in I} \mathbb{P}(Y = y_j | X = x_i) \mathbb{P}(X = x_i) \right) \\
 &= \sum_{j \in J} y_j \left(\sum_{i \in I} \mathbb{P}(X = x_i, Y = y_j) \right) \\
 &= \sum_{j \in J} y_j \left(\sum_{i \in I} p_{ij} \right) \\
 &= \sum_{j \in J} y_j \mathbb{P}(Y = y_j) \\
 &= \mathbb{E}[Y]
 \end{aligned}$$

où on a utilisé la Proposition 4.2.5. ■

Exemple 4.2.12. Reprenons l'exemple 4.2.4. On vérifie directement que $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$.

Exemple 4.2.13. Supposons que deux machines fabriquent les pièces d'un appareil. La première machine fabrique des pièces qui fonctionnent en moyenne 1000 jours, alors que les pièces de la deuxième machine fonctionnent en moyenne 1500 jours. Sachant que $\frac{3}{4}$ des pièces proviennent de la deuxième machine, quelle est la durée moyenne de fonctionnement d'une pièce choisie au hasard ? On note T le temps de fonctionnement d'une pièce choisie au hasard et X la variable aléatoire qui donne 1 si la machine choisie est la première et 2 si c'est la deuxième. On calcule

$$\mathbb{E}[T] = \mathbb{E}[T|X = 1]\mathbb{P}(X = 1) + \mathbb{E}[T|X = 2]\mathbb{P}(X = 2) = 1000 \cdot \frac{1}{4} + 1500 \cdot \frac{3}{4} = 1375.$$

On vient de montrer que $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$. Par contre, en général, $\mathbb{E}[\text{Var}[Y|X]] \neq \text{Var}[Y]$. Néanmoins, on a le résultat suivant.

Théorème 4.2.14. $\mathbb{E}[\text{Var}[Y|X]] = \text{Var}[Y] - \text{Var}[\mathbb{E}[Y|X]]$.

Démonstration. Puisque $\text{Var}[Y|X] = \mathbb{E}[Y^2|X] - (\mathbb{E}[Y|X])^2$, on a

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

En utilisant deux fois le théorème de l'espérance totale, il vient alors

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

■

Bien sûr, tout ce que l'on vient de faire pour les distributions conditionnelles de Y sachant X peut se faire pour les distributions conditionnelles de X sachant Y .

4.3 Lois bivariées continues

Dans cette section, on suppose avoir deux variables aléatoires continues X et Y définies sur le même espace probabilisé.

Définition 4.3.1. Un couple aléatoire (X, Y) est *continu* s'il existe une fonction de deux

variables $f : \mathbb{R}^2 \rightarrow [0, +\infty[$ telle que pour que tout sous-ensemble $B \in \mathcal{B}^2$, on a

$$\mathbb{P}((X, Y) \in B) = \iint_B f(x, y) dy dx.$$

Dans ce cas, on dit que f est la *fonction de densité jointe* de (X, Y) et on la note $f_{(X,Y)}$.

Autrement dit, la probabilité que le couple aléatoire (X, Y) prenne ses valeurs dans $[a, b] \times [c, d]$ est donnée par le *volume sous la surface* dans le rectangle $[a, b] \times [c, d]$ de f . La fonction f doit être positive et vérifier

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{(X,Y)}(x, y) dy dx = 1.$$

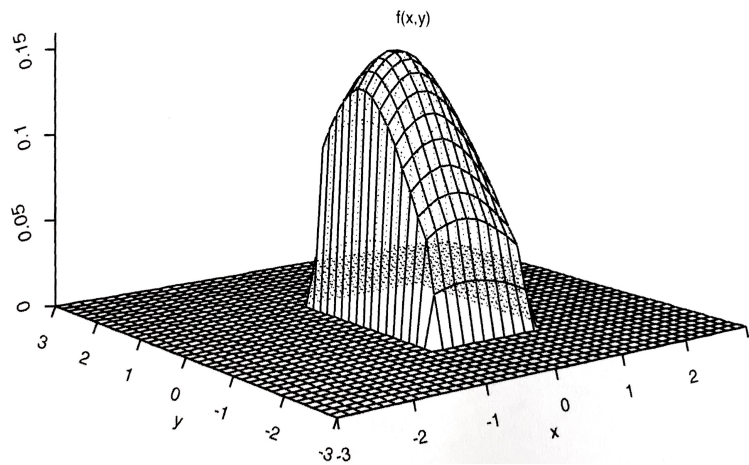


FIGURE 4.1 – Volume sous une fonction de densité jointe $f_{(X,Y)}$ sur le rectangle $[-0.5, 1] \times [-1.5, 1]$.

Comme dans le cas univarié et dans le cas bivarié discret, on peut retrouver la fonction de répartition jointe à partir de la fonction de densité jointe, et inversement. Plus précisément, la fonction de répartition s'exprime grâce à la densité

$$F_{(X,Y)}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{(X,Y)}(u, v) dv du$$

et réciproquement

$$f_{(X,Y)}(x, y) = \frac{\partial}{\partial x} \frac{\partial}{\partial y} F_{(X,Y)}(x, y)$$

en tout point de dérivabilité de la fonction de répartition.

Exemple 4.3.2. Considérons deux voies de détection d'un même dispositif expérimental. Soit X la réponse normalisée du premier capteur et Y celle du second, toutes deux comprises entre 0 et 1. Supposons que (X, Y) admette la fonction de densité

$$f_{(X,Y)}(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & \text{si } (x, y) \in [0, 1] \times [0, 1] \\ 0 & \text{sinon.} \end{cases}$$

Notez bien qu'il s'agit d'un choix de modélisation. On s'intéresse par exemple à la probabilité

d'obtenir simultanément une réponse d'au moins $1/2$ sur le premier capteur et d'au moins $1/4$ sur le second, c'est-à-dire

$$\mathbb{P}(1/2 \leq X \leq 1, 1/4 \leq Y \leq 1).$$

On calcule cette probabilité de la manière suivante

$$\begin{aligned} \int_{1/2}^1 \int_{1/4}^1 f_{(X,Y)}(x,y) dy dx &= \int_{1/2}^1 \left(\int_{1/4}^1 \frac{6}{5}(x+y^2) dy \right) dx \\ &= \int_{1/2}^1 \left[\frac{6}{5} \left(xy + \frac{y^3}{3} \right) \right]_{y=1/4}^{y=1} dx \\ &= \int_{1/2}^1 \left(\frac{9x}{10} + \frac{63}{160} \right) dx = \left[\frac{9x^2}{20} + \frac{63x}{160} \right]_{x=1/2}^{x=1} = \frac{171}{320}, \end{aligned}$$

En procédant de la même façon, on obtient

$$F_{(X,Y)}(x,y) = \begin{cases} 0 & \text{si } x < 0 \text{ ou } y < 0 \\ \frac{1}{5}xy(3x+2y^2) & \text{si } (x,y) \in [0,1] \times [0,1] \\ \frac{1}{5}x(3x+2) & \text{si } x \in [0,1] \text{ et } y > 1 \\ \frac{1}{5}y(3+2y^2) & \text{si } x > 1 \text{ et } y \in [0,1] \\ 1 & \text{sinon.} \end{cases}$$

Notez qu'on peut aussi s'intéresser à des ensembles non-rectangulaires. Par exemple, on peut s'intéresser à la probabilité que la réponse du premier capteur soit inférieure à celle du second. Ceci correspond à calculer la probabilité

$$\mathbb{P}(X \leq Y) = \mathbb{P}[(X,Y) \in B_{\text{tr}}],$$

où $B_{\text{tr}} = \{(x,y) \in [0,1] \times [0,1] : x \leq y\}$.

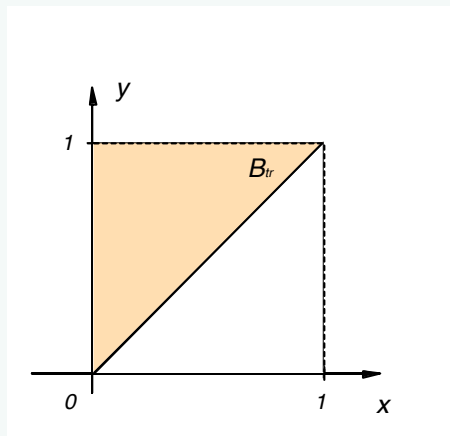


FIGURE 4.2 – Ensemble B_{tr} sur lequel on doit calculer l'intégrale de f .

On calcule cette intégrale de la manière suivante. On remarque que si on fixe x entre 0 et 1, alors y doit varier entre x et 1. Ainsi,

$$\begin{aligned}\mathbb{P}(X \leq Y) &= \mathbb{P}((X, Y) \in B_{\text{tr}}) \\ &= \int \int_{B_{\text{tr}}} f(x, y) dy dx \\ &= \int_0^1 \left(\int_x^1 f(x, y) dy \right) dx = \dots = \frac{1}{2}.\end{aligned}$$

On peut déduire de la distribution jointe les distributions marginales. Exactement comme dans le cas discret, elles s'obtiennent par projection.

Proposition 4.3.3. *Soit (X, Y) un couple aléatoire continu de densité jointe $f_{(X,Y)}$. Alors, les densités marginales de X et Y sont données respectivement par*

$$f_X(x) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x, y) dy$$

et

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x, y) dx.$$

Démonstration. On remarque que

$$\mathbb{P}(X \in B) = \mathbb{P}((X, Y) \in B \times \mathbb{R}) = \int_B \left(\int_{\mathbb{R}} f_{(X,Y)}(x, y) dy \right) dx$$

par le Théorème de Tonelli-Fubini. Cela implique directement que

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy.$$

On procède de manière similaire pour Y . ■

Exemple 4.3.4. Reprenons l'exemple 4.3.2 des deux capteurs. On a

$$f_X(x) = \int_0^1 \frac{6}{5}(x + y^2) dy = \left[\frac{6xy}{5} + \frac{2y^3}{5} \right]_{y=0}^{y=1} = \frac{6x}{5} + \frac{2}{5}$$

si $x \in [0, 1]$, et $f_X(x) = 0$ ailleurs. De même, on a

$$f_Y(y) = \int_0^1 \frac{6}{5}(x + y^2) dx = \left[\frac{3x^2}{5} + \frac{6xy^2}{5} \right]_{x=0}^{x=1} = \frac{3}{5} + \frac{6y^2}{5}$$

si $y \in [0, 1]$, et $f_Y(y) = 0$ ailleurs.

Comme dans le cas discret, si nous avons observé la valeur x de X , nous pouvons affiner notre

connaissance de Y . Par analogie avec le cas discret, où on a

$$\mathbb{P}(Y = y_j | X = x_i) = \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(X = x_i)},$$

on adopte la définition suivante.

Définition 4.3.5. Soit (X, Y) un couple aléatoire continu et soit $x \in \mathbb{R}$ tel que $f_X(x) > 0$. La *fonction de densité conditionnelle* de Y sachant que $X = x$ est la fonction $f_{Y|X=x}$ définie pour tout $y \in \mathbb{R}$ par

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)}.$$

Il s'agit bien d'une densité puisque

$$\int_{-\infty}^{+\infty} f_{Y|X=x}(y) dy = 1$$

et elle définit donc une variable aléatoire. De plus, on a par exemple

$$\begin{aligned} \triangleright f(x, y) &= f_{X|Y=y}(x) f_Y(y), \\ \triangleright \mathbb{P}(a \leq Y \leq b | X = x) &= \int_a^b f_{Y|X=x}(y) dy. \end{aligned}$$

Enfin, on peut calculer l'espérance et la variance *conditionnelles* de Y sachant que $X = x$. On a

$$\mathbb{E}[Y | X = x] = \int_{-\infty}^{+\infty} y f_{Y|X=x}(y) dy$$

et

$$\begin{aligned} \text{Var}[Y | X = x] &= \int_{-\infty}^{+\infty} (y - \mathbb{E}[Y | X = x])^2 f_{Y|X=x}(y) dy \\ &= \int_{-\infty}^{+\infty} y^2 f_{Y|X=x}(y) dy - (\mathbb{E}[Y | X = x])^2. \end{aligned}$$

Exemple 4.3.6. Dans le cas des deux capteurs, ceci donne en particulier (pour chaque $x \in [0, 1]$)

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)} = \begin{cases} \frac{3(x + y^2)}{3x + 1} & \text{si } y \in [0, 1] \\ 0 & \text{sinon.} \end{cases}$$

De là, on calcule

$$\mathbb{E}[Y | X = x] = \int_{-\infty}^{+\infty} y f_{Y|X=x}(y) dy = \int_0^1 y \frac{3(x + y^2)}{3x + 1} dy = \dots = \frac{6x + 3}{12x + 4}$$

et

$$\begin{aligned}\text{Var}[Y|X = x] &= \int_{-\infty}^{+\infty} y^2 f_{Y|X=x}(y) dy - (\mathbb{E}[Y|X = x])^2 \\ &= \int_0^1 y^2 \frac{3(x+y^2)}{3x+1} dy - \left(\frac{6x+3}{12x+4} \right)^2 = \dots = \frac{60x^2 + 44x + 3}{80(3x+1)^2}.\end{aligned}$$

Définition 4.3.7.

▷ La variable aléatoire *espérance conditionnelle* $\mathbb{E}[Y|X]$ est la variable aléatoire

$$\mathbb{E}[Y|X] : \Omega \rightarrow \mathbb{R} : \omega \mapsto \mathbb{E}[Y|X = x] \quad \text{si} \quad X(\omega) = x.$$

▷ La variable aléatoire *variance conditionnelle* $\text{Var}[Y|X]$ est la variable aléatoire

$$\text{Var}[Y|X] : \Omega \rightarrow \mathbb{R} : \omega \mapsto \text{Var}[Y|X = x] \quad \text{si} \quad X(\omega) = x.$$

Alors, comme dans le cas discret, on a le théorème de l'espérance totale suivant.

Théorème 4.3.8 (Théorème de l'espérance totale). *On a*

▷ $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$, *c'est-à-dire*

$$\mathbb{E}[Y] = \int_{-\infty}^{+\infty} \mathbb{E}[Y|X = x] f_X(x) dx.$$

▷ $\mathbb{E}[\text{Var}[Y|X]] = \text{Var}[Y] - \text{Var}[\mathbb{E}[Y|X]]$.

si les intégrales sont bien définies.

Démonstration. On a

$$\begin{aligned}\mathbb{E}[\mathbb{E}[Y|X]] &= \int_{\mathbb{R}} \mathbb{E}[Y|X = x] f_X(x) dx = \int_{\mathbb{R}} \int_{\mathbb{R}} y f_{Y|X=x}(y) dy f_X(x) dx \\ &= \int_B \int_{\mathbb{R}} y \frac{f_{(X,Y)}(x,y)}{f_X(x)} dy f_X(x) dx \\ &= \int_{\mathbb{R}} y \int_B f_{(X,Y)}(x,y) dx dy \\ &= \int_{\mathbb{R}} y f_Y(y) dy \\ &= \mathbb{E}[Y]\end{aligned}$$

avec $B = \{x \in \mathbb{R} : f_X(x) \neq 0\}$ et puisque $f_{(X,Y)}(x,y) = 0$ si $x \notin B$.

La preuve du second point est identique à celle faite dans le cas discret puisqu'elle n'utilise que le théorème de l'espérance totale. ■

Exemple 4.3.9. Dans le cas des deux capteurs, on a

$$\begin{aligned}\mathbb{E}[\mathbb{E}[Y|X]] &= \int_{-\infty}^{+\infty} \mathbb{E}[Y|X=x] f_X(x) dx = \int_0^1 \left(\frac{6x+3}{12x+4} \right) \left(\frac{2}{5} (3x+1) \right) dx \\ &= \int_0^1 \frac{1}{10} (6x+3) dx \\ &= \left[\frac{1}{10} (3x^2 + 3x) \right]_{x=0}^{x=1} = \frac{3}{5} = \mathbb{E}[Y].\end{aligned}$$

4.4 Lois bivariées mixtes

Rien n'impose que toutes les composantes d'un vecteur aléatoire soient de même nature : on peut imaginer un vecteur ayant certaines composantes discrètes, certaines continues, certaines « mixtes », etc. Nous nous contenterons de présenter un exemple.

Exemple 4.4.1. Soit (X, Y) un vecteur aléatoire avec $X \sim \text{Exp}(1)$ un temps d'attente (en heures) et Y qui marque si l'on a dû attendre plus d'une heure ou non (donc $Y = 1$ si $X \geq 1$ et 0 sinon). Alors les lois marginales sont données par

$$X \sim \text{Exp}(1) \quad \text{et} \quad Y \sim \text{Bern}(p)$$

avec $p = \mathbb{P}(X \geq 1) = \int_1^\infty e^{-x} dx = 1/e$. La loi jointe s'obtient grâce à

$$\mathbb{P}(X \leq x, Y = 0) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-x} & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

et

$$\mathbb{P}(X \leq x, Y = 1) = \begin{cases} 0 & \text{si } x < 1 \\ e^{-1} - e^{-x} & \text{si } x \geq 1 \end{cases}.$$

4.5 Indépendance de variables aléatoires

Soit (X, Y) un couple aléatoire. Lorsque l'on s'intéresse aux lois conditionnelles, il se peut que la connaissance de X ne nous donne aucune information sur Y et réciproquement. C'est le cas de l'exemple 4.2.2 du lancer de deux dés.

La définition de l'indépendance de variables aléatoires se base sur ce qui a été vu dans le chapitre 2. On va dire que X et Y sont indépendantes si les événements $\{X \leq x\}$ et $\{Y \leq y\}$ sont indépendants quels que soient les réels x et y . Autrement dit, on adopte la définition suivante.

Définition 4.5.1. Soient X et Y deux variables aléatoires définies sur un même espace probabilisé. On dit que X et Y sont *indépendantes* si

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

pour tous $x, y \in \mathbb{R}$, c'est-à-dire si la fonction de répartition jointe se factorise en le produit des fonctions de répartition marginales. Dans ce cas, on note $X \perp\!\!\!\perp Y$.

Cette définition se particularise au cas où le couple (X, Y) est discret de la manière suivante.

Proposition 4.5.2. Soit (X, Y) un couple aléatoire discret. Les assertions suivantes sont équivalentes :

▷ X et Y sont indépendants ;

▷ on a

$$p_{ij} = p_{i\bullet}p_{\bullet j}$$

pour tous $i \in I, j \in J$;

▷ on a

$$\mathbb{P}(Y = y_j | X = x_i) = p_{\bullet j}$$

pour tous $i \in I, j \in J$ avec $p_{i\bullet} \neq 0$.

Démonstration. $1 \Leftrightarrow 2$. Supposons que $X \perp\!\!\!\perp Y$. Alors, les événements $\{X = x_i\}$ et $\{Y = y_j\}$ sont indépendants et donc

$$p_{ij} = \mathbb{P}(X = x_i \text{ et } Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j) = p_{i\bullet}p_{\bullet j}.$$

La réciproque est immédiate puisque la fonction de répartition jointe se réécrit à partir de la fonction de masse jointe.

$2 \Rightarrow 3$. On a

$$\mathbb{P}(Y = y_j | X = x_i) = \frac{\mathbb{P}(X = x_i \text{ et } Y = y_j)}{\mathbb{P}(X = x_i)} = \frac{p_{ij}}{p_{i\bullet}} = p_{\bullet j}.$$

$3 \Rightarrow 2$. On a

$$p_{ij} = \mathbb{P}(X = x_i \text{ et } Y = y_j) = \mathbb{P}(Y = y_j | X = x_i)\mathbb{P}(X = x_i) = p_{\bullet j}p_{i\bullet}.$$

■

Ainsi, tout ce qui est conditionnel devient marginal et les fonctions de masse jointes sont données par le produit des fonctions de masse.

Exemple 4.5.3. On lance deux dés distinguables. On considère la variable aléatoire X qui donne le résultat du premier dé et la variable aléatoire Y qui donne le résultat du deuxième dé.

	1	2	3	4	5	6	
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Pour tous i, j , on a $p_{ij} = p_{i\bullet}p_{\bullet j}$ et donc $X \perp\!\!\!\perp Y$. Cette indépendance se traduit aussi par le fait que, pour chaque x_i , la distribution conditionnelle de Y sachant que $X = x_i$ coïncide avec la distribution marginale de Y .

	1	2	3	4	5	6
1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
3	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
4	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
5	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
6	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
	1	1	1	1	1	1

y_j	$p_{\bullet j}$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$
	1

De même, dans le cas continu, on a la caractérisation suivante.

Proposition 4.5.4. Soit (X, Y) un couple aléatoire continu. Les assertions suivantes sont équivalentes.

▷ X et Y sont indépendants ;

▷ on a

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

pour tous $x, y \in \mathbb{R}$;

▷ on a

$$f_{Y|X=x}(y) = f_Y(y)$$

pour tous $x, y \in \mathbb{R}$ tels que $f_X(x) \neq 0$.

Démonstration. $1 \Rightarrow 2$. Si $X \perp\!\!\!\perp Y$, alors

$$F_{(X,Y)}(x, y) = F_X(x)f_Y(y) = \int_{-\infty}^x f_X(x)dx \int_{-\infty}^y f_Y(y)dy = \int_{-\infty}^x \int_{-\infty}^y f_X(x)f_Y(y)dydx$$

et donc $f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$.

$2 \Rightarrow 1$. On a

$$\begin{aligned} F_{(X,Y)}(x, y) &= \int_{-\infty}^x \int_{-\infty}^y f_{(X,Y)}(x, y)dydx = \int_{-\infty}^x \int_{-\infty}^y f_X(x)f_Y(y)dydx \\ &= \int_{-\infty}^x f_X(x)dx \int_{-\infty}^y f_Y(y)dy \\ &= F_X(x)F_Y(y). \end{aligned}$$

$2 \Leftrightarrow 3$. Il suffit d'utiliser la définition de la fonction de densité conditionnelle. ■

Exemple 4.5.5. A l'arrêt des grands amphis, on observe

- ▷ le temps X (en min) jusqu'à la prochaine arrivée d'un bus 58 (direction ville),
- ▷ le temps Y (en min) jusqu'à la prochaine arrivée d'un bus 48 (direction ville).

Supposons qu'un 58 passe exactement toutes les 10 minutes, qu'un 48 passe exactement toutes les 8 minutes et que (X, Y) admette la densité

$$f(x, y) = \begin{cases} \frac{1}{80} & \text{si } (x, y) \in [0, 10] \times [0, 8] \\ 0 & \text{sinon.} \end{cases}$$

On vérifie alors que

$$f_X(x) = \begin{cases} \frac{1}{10} & \text{si } x \in [0, 10] \\ 0 & \text{sinon} \end{cases} \quad \text{et} \quad f_Y(y) = \begin{cases} \frac{1}{8} & \text{si } y \in [0, 8] \\ 0 & \text{sinon,} \end{cases}$$

ce qui implique que $f(x, y) = f_X(x)f_Y(y) \forall x, y$. On a donc $X \perp\!\!\!\perp Y$.

En particulier, dans les deux cas, il est clair que si X et Y sont indépendantes, alors

$$\mathbb{E}[X|Y = y] = \mathbb{E}[X] \quad \text{et} \quad \text{Var}[X|Y = y] = \text{Var}[X].$$

Comme dans le cas de l'indépendance de plus de deux événements, on peut adopter l'extension suivante de la notion d'indépendance de variables aléatoires.

Définition 4.5.6. Des variables aléatoires $X_1, X_2, \dots, X_d$ sont *indépendantes* si les événements $\{X_1 \leq x_1\}, \dots, \{X_d \leq x_d\}$ sont indépendants pour tous $x_1, \dots, x_d \in \mathbb{R}$.

A nouveau, cette définition est plus forte que l'indépendance des variables aléatoires deux à deux ! On demande que la fonction de répartition de chaque vecteur aléatoire que l'on peut former à partir de $X_1, X_2, \dots, X_d$ se factorise en le produit des fonctions de répartition correspondantes. Soit (X, Y) un vecteur aléatoire bivarié. Pour finir cette section, on considère à présent une fonction $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ et on va s'intéresser à des espérances du type $\mathbb{E}[h(X, Y)]$. Par exemple, $\mathbb{E}[X + Y]$ ou $\mathbb{E}[XY]$. Comme dans le cas unidimensionnel, on souhaite calculer $\mathbb{E}[h(X, Y)]$ sans avoir à déterminer la loi de $h(X, Y)$

Proposition 4.5.7. Soit $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ telle que $h(X, Y) \in L^1$.

▷ Cas discret :

$$\mathbb{E}[h(X, Y)] = \sum_{i \in I} \sum_{j \in J} h(x_i, y_j) p_{ij}$$

▷ Cas continu :

$$\mathbb{E}[h(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x, y) f_{(X, Y)}(x, y) dy dx$$

Démonstration. Donnons l'idée de la preuve dans le cas discret. Notons $\{t_k : k \in K\}$ l'ensemble des valeurs $\{h(x_i, y_j) : i \in I, j \in J\}$. On a

$$\mathbb{P}(h(X, Y) = t_k) = \mathbb{P}\left(\bigcup_{i \in I, j \in J : h(x_i, y_j) = t_k} \{X = x_i, Y = y_j\}\right) = \sum_{i \in I, j \in J : h(x_i, y_j) = t_k} \mathbb{P}(X = x_i, Y = y_j)$$

et donc

$$\begin{aligned} \mathbb{E}[h(X, Y)] &= \sum_{k \in K} t_k \sum_{i \in I, j \in J : h(x_i, y_j) = t_k} \mathbb{P}(X = x_i, Y = y_j) \\ &= \sum_{k \in K} \sum_{i \in I, j \in J : h(x_i, y_j) = t_k} h(x_i, y_j) \mathbb{P}(X = x_i, Y = y_j) \\ &= \sum_{i \in I} \sum_{j \in J} h(x_i, y_j) \mathbb{P}(X = x_i, Y = y_j). \end{aligned}$$

■

Exemple 4.5.8. Supposons que X et Y sont deux variables aléatoires indépendantes de loi uniforme sur $[0, 1]$. On souhaite calculer

$$\mathbb{E}[|X - Y|].$$

Comme les variables aléatoires sont indépendantes, on a $f_{(X, Y)}(x, y) = \mathbf{1}_{[0, 1]}(x) \mathbf{1}_{[0, 1]}(y)$. On

applique le résultat précédent pour écrire

$$\begin{aligned}\mathbb{E}[|X - Y|] &= \int_0^1 \int_0^1 |x - y| dx dy \\ &= \int_0^1 \int_y^1 (x - y) dx dy + \int_0^1 \int_0^y (y - x) dx dy = \dots = \frac{1}{3}\end{aligned}$$

Comme dans le cas unidimensionnel, on peut s'intéresser également à la loi de $h(X, Y)$ et pas seulement à son espérance. On pourrait même supposer que $h(X, Y)$ est également un vecteur aléatoire (par exemple, considérer $(X + Y, X - Y)$). Dans le cas discret, on passe en revue les différentes valeurs prises par X et Y qui donnent une valeur de $h(X, Y)$ fixée, comme on l'a fait dans la preuve précédente. Dans le cas continu, on a l'équivalent de la formule de changement de variables.

Proposition 4.5.9 (Changement de variables). *Si $X = (X_1, X_2)$ est un vecteur aléatoire continu de densité f_X et si $g : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ est un difféomorphisme de classe C^1 , alors $Y = g(X)$ est continu de densité*

$$f_Y(y) = f_X(g^{-1}(y)) \left| \det Dg^{-1}(y) \right|.$$

Équivalemment,

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{\left| \det Dg(g^{-1}(y)) \right|}.$$

La Proposition 4.5.7 nous permet de retrouver la linéarité de l'espérance. On peut également montrer que si deux variables aléatoires sont indépendantes, alors l'espérance de leur produit est égal au produit des espérances.

Théorème 4.5.10. *Soit (X, Y) un vecteur aléatoire bivarié. Si X et Y admettent des espérances, on a*

1. $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$,
2. si $X \perp\!\!\!\perp Y$, alors $\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y]$.

Démonstration. 1. On fait la preuve dans le cas discret (le cas continu est laissé comme exercice).

On a

$$\begin{aligned}
 \mathbb{E}[X + Y] &= \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} (x_i + y_j) p_{ij} \\
 &= \sum_{i=1}^{N_1} x_i \sum_{j=1}^{N_2} p_{ij} + \sum_{j=1}^{N_2} y_j \sum_{i=1}^{N_1} p_{ij} \\
 &= \sum_{i=1}^{N_1} x_i p_{i\bullet} + \sum_{j=1}^{N_2} y_j p_{\bullet j} \\
 &= \mathbb{E}[X] + \mathbb{E}[Y]
 \end{aligned}$$

où on a utilisé la Proposition 4.2.5.

2. On fait la preuve dans le cas continu (le cas discret est laissé comme exercice). On a

$$\begin{aligned}
 \mathbb{E}[XY] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_{(X,Y)}(x, y) dy dx \\
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_X(x) f_Y(y) dy dx \\
 &= \left(\int_{-\infty}^{+\infty} x f_X(x) dx \right) \left(\int_{-\infty}^{+\infty} y f_Y(y) dy \right) = \mathbb{E}[X] \mathbb{E}[Y].
 \end{aligned}$$

■

Remarque 4.5.11. Pour le 2e point, on ne peut pas retirer l'hypothèse d'indépendance ! En effet, prenons $X \sim \text{Bern}(p)$ et $Y = 1 - X$. Alors $XY = 0$ est la variable constante 0, d'où $\mathbb{E}[XY] = 0$, alors que $\mathbb{E}[X]\mathbb{E}[Y] = p(1 - p)$.

Le résultat suivant est intuitif et sera admis.

Proposition 4.5.12. Si $X \perp\!\!\!\perp Y$, alors quelles que soient les fonctions h, g pour lesquelles $h(X)$ et $g(Y)$ sont des v.a., on a $h(X) \perp\!\!\!\perp g(Y)$. En particulier, si $h(X), g(Y) \in L^1$, on a

$$\mathbb{E}[h(X)g(Y)] = \mathbb{E}[h(X)]\mathbb{E}[g(Y)].$$

Démonstration. L'idée de la preuve est que si X et Y sont indépendants, alors les événements $\{h(X) \leq x\}$ et $\{g(Y) \leq y\}$ sont indépendants. Ainsi

$$F_{(h(X), g(Y))}(x, y) = \mathbb{P}(h(X) \leq x, g(Y) \leq y) = \mathbb{P}(h(X) \leq x) \mathbb{P}(g(Y) \leq y) = F_{h(X)}(x) F_{g(Y)}(y).$$

■

Le résultat suivant est une application directe de cette propriété.

Proposition 4.5.13. *Si $X \perp\!\!\!\perp Y$, alors*

$$M_{X+Y}(t) = M_X(t)M_Y(t) \quad \text{et} \quad \varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$$

là où les fonctions sont définies.

Démonstration. On utilise le résultat précédent pour écrire

$$M_{X+Y}(t) = \mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX}e^{tY}] = \mathbb{E}[e^{tX}]\mathbb{E}[e^{tY}] = M_X(t)M_Y(t).$$

■

Ceci permet de prouver certaines *propriétés d'additivité*. On utilise pour cela le fait que la fonction génératrice des moments

$$M_X(t) = \mathbb{E}[e^{tX}]$$

caractérise la loi de la variable aléatoire X . On peut faire de même avec la fonction caractéristique.

Théorème 4.5.14. *Supposons que $X \perp\!\!\!\perp Y$.*

1. *Si $X \sim \text{Bin}(n_1, p)$ et $Y \sim \text{Bin}(n_2, p)$, alors $X + Y \sim \text{Bin}(n_1 + n_2, p)$.*
2. *Si $X \sim \text{Pois}(\lambda_1)$ et $Y \sim \text{Pois}(\lambda_2)$, alors $X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$.*
3. *Si $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, alors $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.*

Démonstration. 1. On calcule $M_X(t) = (1 - p + pe^t)^n$ si $X \sim \text{Bin}(n, p)$. Par conséquent, l'indépendance entre X et Y fournit

$$M_{X+Y}(t) = M_X(t)M_Y(t) = (1 - p + pe^t)^{n_1}(1 - p + pe^t)^{n_2} = (1 - p + pe^t)^{n_1+n_2} = M_Z(t),$$

où $Z \sim \text{Bin}(n_1 + n_2, p)$. Puisque deux variables aléatoires partageant la même fonction génératrice des moments ont la même loi, on en déduit que $X + Y \sim \text{Bin}(n_1 + n_2, p)$.

2. On fait de même en utilisant le fait que $M_X(t) = e^{\lambda(e^t-1)}$ si $X \sim \text{Poi}(\lambda)$.

3. On fait de même en utilisant le fait que $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$ si $X \sim \mathcal{N}(\mu, \sigma^2)$. ■

4.6 Covariance et corrélation

En utilisant les propriétés de l'espérance, si $Z = X + Y$, on a

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

La variance d'une somme n'est donc pas égale à la somme des variances. La variance de la somme peut être plus grande ou plus petite suivant que la *covariance entre X et Y* est positive ou négative.

Définition 4.6.1. La *covariance entre X et Y* est la quantité définie par

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

Bruit commun entre deux capteurs. Supposons que deux capteurs mesurent la même grandeur θ et que

$$X = \theta + U + \varepsilon_1, \quad Y = \theta + U + \varepsilon_2,$$

où U représente une fluctuation commune au dispositif et $\varepsilon_1, \varepsilon_2$ des bruits propres aux capteurs, centrés et indépendants entre eux et de U . Alors

$$\text{Cov}(X, Y) = \text{Var}(U) > 0.$$

Une corrélation positive peut donc provenir d'une source de bruit commune, même si les bruits propres des deux capteurs sont indépendants.

L'interprétation de la covariance est illustrée dans les Figures 4.3 et 4.4.

L'espérance est linéaire : la moyenne d'une somme est la somme des moyennes, et ce quelle que soit la structure de dépendance entre les différentes variables sommées. Nous savons que l'espérance du produit de variables aléatoires indépendantes est le produit des espérances individuelles. On a vu que ceci n'est plus vrai si les variables ne sont pas indépendantes. Il y a donc une différence entre l'espérance du produit et le produit des espérances. La covariance capture exactement cette différence.

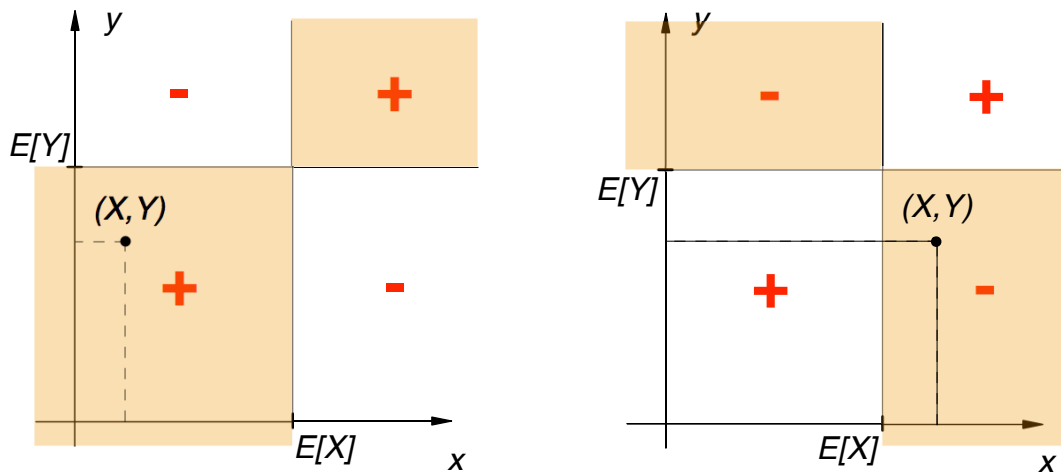


FIGURE 4.3 – A gauche, (X, Y) se réalise dans la zone où $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) > 0$. A droite, (X, Y) se réalise dans la zone où $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) < 0$. Si (X, Y) se réalise plus souvent dans les zones "+" (ou de façon plus extrême), $\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$ sera positif, et inversement.

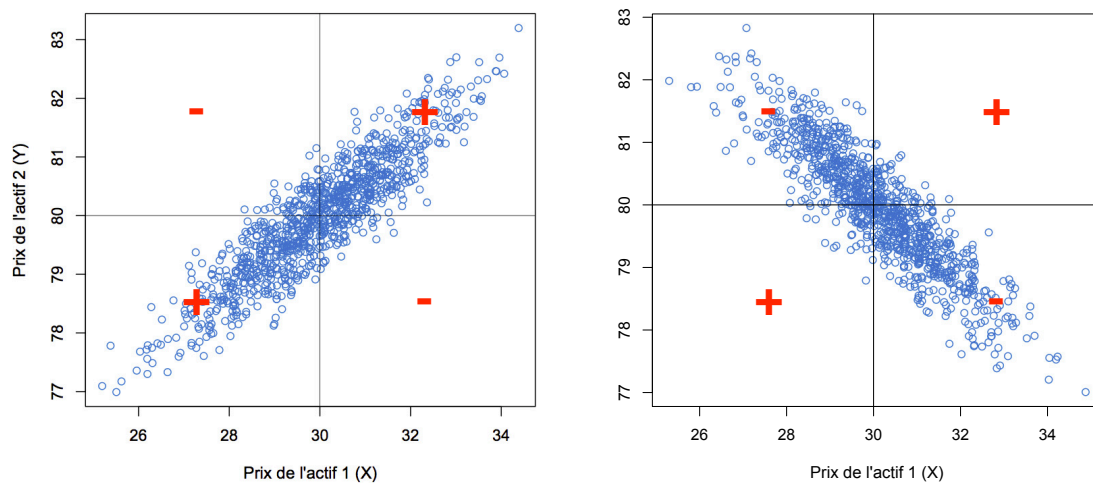


FIGURE 4.4 – A gauche, $\text{Cov}[X, Y] > 0$; on parlera de dépendance positive (notons que dans ce cas, $\text{Var}[X + Y] > \text{Var}[X] + \text{Var}[Y]$). A droite, $\text{Cov}[X, Y] < 0$; on parlera de dépendance négative (notons que dans ce cas, $\text{Var}[X + Y] < \text{Var}[X] + \text{Var}[Y]$).

Proposition 4.6.2. Soient X et Y deux variables aléatoires qui possèdent une variance.

1. $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2 \text{Cov}[X, Y]$.
2. $\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y]$.
3. Si $X \perp\!\!\!\perp Y$, alors $\text{Cov}[X, Y] = 0$.
4. Si $X \perp\!\!\!\perp Y$, alors $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$.

Démonstration. 1. Le premier point a déjà été montré en introduction de la définition de la corrélation.

2. On a

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

3. Si $X \perp\!\!\!\perp Y$, alors on sait que $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ et le point 2 donne la conclusion.

4. C'est une conséquence directe des points 1 et 3. ■

Développons l'expression de la covariance dans les cas discrets et continus. Si on utilise la définition de la covariance, il vient

$$\text{Cov}[X, Y] = \begin{cases} \sum_{i \in I} \sum_{j \in J} (x_i - \mathbb{E}[X])(y_j - \mathbb{E}[Y]) p_{ij} & (\text{cas discret}) \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) f(x, y) dy dx & (\text{cas continu}). \end{cases}$$

Si on utilise l'expression $\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ donnée par le point 2 de la proposition précédente, on trouve que

$$\text{Cov}[X, Y] = \begin{cases} \sum_{i \in I} \sum_{j \in J} x_i y_j p_{ij} - \mathbb{E}[X]\mathbb{E}[Y] & (\text{cas discret}) \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f(x, y) dy dx - \mathbb{E}[X]\mathbb{E}[Y] & (\text{cas continu}). \end{cases}$$

Dans la pratique, c'est souvent cette dernière expression que l'on utilise.

Définition 4.6.3. Si $\text{Cov}[X, Y] = 0$, on dit que les variables X et Y sont *non-corrélées*.

En particulier, le point 3 de la proposition précédente montre que des variables aléatoires indépendantes sont non-corrélées. Attention, la non-corrélation n'implique pas l'indépendance ! La covariance n'est donc pas une mesure de dépendance parfaite.

Exemple 4.6.4. On lance deux dés distinguables. Comme dans l'exemple 4.2.4, soit X le résultat de la somme des résultats des deux dés et soit Y la valeur absolue de leur différence.

	2	3	4	5	6	7	8	9	10	11	12
0	$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$
1		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$	
2			$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$		
3				$\frac{2}{36}$		$\frac{2}{36}$		$\frac{2}{36}$			
4					$\frac{2}{36}$		$\frac{2}{36}$				
5						$\frac{2}{36}$					

On vérifie que $\text{Cov}[X, Y] = 0$ mais les variables X et Y ne sont pas indépendantes.

Exemple 4.6.5. Si $X \sim \mathcal{N}(0, 1)$ et $Y = X^2$ alors $\mathbb{E}[XY] = \mathbb{E}[X^3] = 0$, du coup $\text{Cov}[X, Y] = 0$ mais X n'est clairement pas indépendant de Y .

Donnons quelques propriétés supplémentaires de la covariance (exercice).

Proposition 4.6.6. Soient X, Y, Z des variables aléatoires qui admettent une variance et $c \in \mathbb{R}$. Alors

1. $\text{Cov}[X, X] = \text{Var}[X]$,
2. $\text{Cov}[X, Y] = \text{Cov}[Y, X]$,
3. $\text{Cov}[X, c] = 0$,
4. $\text{Cov}[cX, Y] = c \text{Cov}[X, Y]$,
5. $\text{Cov}[X + Z, Y] = \text{Cov}[X, Y] + \text{Cov}[Z, Y]$.

Démonstration. Les résultats sont directs et sont laissés à titre d'exercice. ■

Seul compte le *signe* de la covariance, étant donné que la valeur numérique de cette quantité dépend des valeurs et des unités des variables individuelles et n'est pas standardisée. Le résultat suivant, dont la preuve est admise, va nous permettre de prendre en considération la valeur de la covariance par rapport aux variances des variables aléatoires.

Proposition 4.6.7 (Inégalité de Cauchy-Schwarz). Si $X, Y \in L^2$, on a

$$|\text{Cov}[X, Y]| \leq \sqrt{\text{Var}[X]\text{Var}[Y]}$$

avec égalité si et seulement si $Y \stackrel{ps}{=} \alpha X + \beta$ ou $X \stackrel{ps}{=} \alpha Y + \beta$ pour certains $\alpha, \beta \in \mathbb{R}$.

Démonstration. Si X, Y sont centrées, alors

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

pour tout $\lambda \in \mathbb{R}$. On a donc un polynôme de degré 2 en λ qui est toujours positif ou nul et donc

$$4\mathbb{E}[XY]^2 - 4\mathbb{E}[X^2]\mathbb{E}[Y^2] \leq 0$$

c'est-à-dire

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

On a l'égalité si et seulement si le polynôme admet une racine, c'est-à-dire il existe $\lambda \in \mathbb{R}$ tel que $\mathbb{E}[(X + \lambda Y)^2] = 0$. On en tire que $X + \lambda Y = 0$ presque sûrement. On obtient la conclusion générale appliquant le résultat aux versions centrées de X et Y . ■

Ce résultat mène naturellement à l'introduction de la notion suivante.

Définition 4.6.8. Soient X et Y deux variables aléatoires telles que $\text{Var}[X] > 0$ et $\text{Var}[Y] > 0$. La *corrélation entre X et Y* est la quantité

$$\text{Corr}[X, Y] = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]\text{Var}[Y]}}.$$

La corrélation est donc une version normalisée de la covariance dont on peut interpréter le signe mais également la valeur : en effet, les signes de $\text{Corr}[X, Y]$ et de $\text{Cov}[X, Y]$ coïncident, on a $-1 \leq \text{Corr}[X, Y] \leq 1$ et $\text{Corr}[X, Y] = 1$ ou -1 si et seulement si $Y \stackrel{ps}{=} \alpha X + \beta$ ou $X \stackrel{ps}{=} \alpha Y + \beta$ (dépendance *linéaire*).

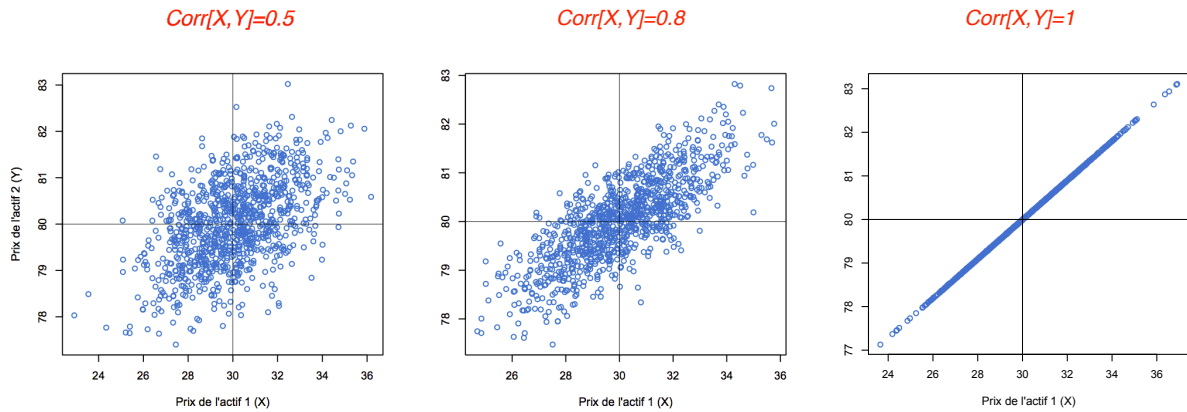


FIGURE 4.5 – Plus $|\text{Corr}[X, Y]|$ est proche de 1, plus la relation linéaire entre X et Y est forte.

Signalons que l'opérateur espérance $\mathbb{E}[\cdot]$ peut être aussi appliqué à un vecteur ou à une matrice, auquel cas on prend l'espérance composante par composante.

Définition 4.6.9. Soit $Z = (X_1, \dots, X_n)$ un vecteur aléatoire. On définit

▷ son vecteur moyen

$$\mu_Z = \mathbb{E}[Z] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])$$

▷ sa matrice de variance-covariance

$$\begin{aligned} \Sigma_Z = \text{Var}[Z] &= \mathbb{E}[(Z - \mathbb{E}[Z])(Z - \mathbb{E}[Z])^T] \\ &= \begin{pmatrix} \text{Var}[X_1] & \text{Cov}[X_1, X_2] & \dots & \text{Cov}[X_1, X_n] \\ \text{Cov}[X_2, X_1] & \text{Var}[X_2] & \dots & \text{Cov}[X_2, X_n] \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}[X_n, X_1] & \text{Cov}[X_n, X_2] & \dots & \text{Var}[X_n] \end{pmatrix} \end{aligned}$$

c'est-à-dire

$$(\Sigma_Z)_{ij} = \text{Cov}[X_i, X_j]$$

Les propriétés de l'espérance et du produit matriciel nous amènent les résultats utiles suivants.

Proposition 4.6.10. Soit $X \in L^2$ un vecteur aléatoire de $\mathbb{R}^d$. Soient A une matrice et b un vecteur de dimensions compatibles. Alors

$$\mathbb{E}[AX + b] = A\mathbb{E}[X] + b \text{ et } \text{Var}[AX + b] = A\text{Var}[X]A^T.$$

En particulier, pour tout $b \in \mathbb{R}^d$, on a

$$b^T \text{Var}[X] b = \text{Var}[b^T X] \geq 0,$$

donc une matrice de variance-covariance est toujours semi-définie positive.

Remarque 4.6.11. Soit X un vecteur aléatoire de dimension d . Si elle existe, la matrice $\text{Var}[X]$ est symétrique et semi-définie positive. On peut donc la représenter sous la forme $\text{Var}[X] = AA^T$ avec A une matrice carrée (on dit que A est une racine carrée de $\text{Var}[X]$)^a.

- ▷ *Centrer* la loi de X revient à considérer le vecteur $Y = X - \mathbb{E}[X]$ qui, par construction, est d'espérance $\mathbb{E}[Y] = 0$. Cette opération est permise dès que X est intégrable.
- ▷ *Réduire* la loi de X revient à considérer le vecteur $Z = A^{-1}X$ qui, par construction, est de variance $\text{Var}[Z] = \text{Id}$. Cette opération est permise dès que X est de carré intégrable et que $\text{Var}[X]$ est définie positive.

^a. Comme $\text{Var}[X]$ est symétrique, on peut la diagonaliser par une matrice orthogonale P . Ainsi, on peut écrire $\text{Var}[X] = P^T \text{diag}(\lambda_1, \dots, \lambda_d) P$. De plus, comme la matrice $\text{Var}[X]$ est semi-définie positive, ses valeurs propres $\lambda_1, \dots, \lambda_d$ sont positives ou nulles. Il suffit alors de poser $A = P^T \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d})$.

Pourquoi la covariance est-elle utile en pratique ?

Deux mesures réalisées dans une même expérience peuvent partager une source commune d'incertitude : une dérive de l'appareil, une erreur de calibration, une variation de température, etc. Dans ce cas, leurs erreurs peuvent ne plus être indépendantes. La covariance intervient alors naturellement lorsqu'une grandeur physique est calculée à partir de plusieurs quantités mesurées.

Supposons par exemple qu'une vitesse soit obtenue à partir de la mesure d'une distance D et d'un temps T :

$$V = \frac{D}{T}.$$

En raison des erreurs de mesure, D et T sont modélisées par des variables aléatoires. L'incertitude sur V dépend alors des incertitudes sur D et T , mais aussi de la manière dont celles-ci sont liées.

Lorsque les fluctuations de D et T autour de leurs valeurs moyennes

$$d = \mathbb{E}[D], \quad t = \mathbb{E}[T]$$

sont suffisamment petites, on peut approcher la fonction

$$g(x, y) = \frac{x}{y}$$

par son développement de Taylor au premier ordre autour du point (d, t) . Comme

$$\frac{\partial g}{\partial x}(d, t) = \frac{1}{t}, \quad \frac{\partial g}{\partial y}(d, t) = -\frac{d}{t^2},$$

on obtient que

$$V = \frac{D}{T} \approx \frac{d}{t} + \frac{1}{t}(D - d) - \frac{d}{t^2}(T - t).$$

Le terme d/t étant constant, il ne contribue pas à la variance. On utilise alors la formule

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

Ainsi,

$$\begin{aligned} \text{Var}(V) &\approx \text{Var}\left(\frac{1}{t}(D - d) - \frac{d}{t^2}(T - t)\right) \\ &= \frac{1}{t^2} \text{Var}(D) + \frac{d^2}{t^4} \text{Var}(T) - 2\frac{d}{t^3} \text{Cov}(D, T). \end{aligned}$$

On obtient donc

$$\boxed{\text{Var}(V) \approx \frac{\text{Var}(D)}{t^2} + \frac{d^2}{t^4} \text{Var}(T) - 2\frac{d}{t^3} \text{Cov}(D, T).}$$

Cette formule fait apparaître naturellement la covariance. Si D et T sont indépendantes,

alors

$$\text{Cov}(D, T) = 0,$$

et leurs contributions à la variance de V s'additionnent. En revanche, si D et T sont corrélées, il faut également tenir compte du terme de covariance.

5 Théorèmes limites

Dans ce chapitre, nous présentons deux théorèmes d'une importance primordiale en théorie des probabilités et dans ses applications en statistiques. En particulier, nous montrons que l'approche axiomatique des probabilités que nous avons adoptée permet de justifier l'approche fréquentiste des probabilités.

5.1 Loi des grands nombres

On souhaite par exemple évaluer l'autonomie des batteries d'un nouvel iPhone qui, selon la publicité, est de 28 heures. Comment vérifier, à partir de mesures expérimentales, si l'autonomie annoncée est compatible avec la réalité ?

Si X désigne la variable aléatoire donnant l'autonomie d'un appareil choisi au hasard, la quantité qui nous intéresse est son autonomie moyenne

$$\mu = \mathbb{E}[X].$$

On pourrait par exemple chercher à déterminer, à partir d'un échantillon de batteries, si

$$\mathbb{E}[X] < 28$$

ou non. Il s'agit d'un problème de décision statistique. Avant de pouvoir prendre une telle décision, il faut donc être capable d'obtenir de l'information sur l'espérance de X , alors même que sa loi n'est pas nécessairement connue.

De manière générale, considérons une variable aléatoire X d'espérance

$$\mu = \mathbb{E}[X]$$

et de variance

$$\sigma^2 = \text{Var}(X).$$

Soient

$$X_1, \dots, X_n$$

des variables aléatoires indépendantes et identiquement distribuées (*i.i.d.*) selon la même loi que X . Elles représentent n répétitions indépendantes d'une même expérience. Une quantité naturelle construite à partir de ces observations est leur moyenne empirique

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Mesures répétées en physique expérimentale.

Lorsqu'une grandeur physique est mesurée, on répète souvent la même mesure afin de réduire l'effet du bruit aléatoire. Un modèle simple consiste à écrire

$$X_i = \theta + \varepsilon_i,$$

où θ est la grandeur physique inconnue que l'on souhaite mesurer et où les erreurs ε_i sont supposées centrées, indépendantes et de variance σ^2 .

Dans ce cas,

$$\mathbb{E}[X_i] = \theta,$$

et la moyenne des n mesures s'écrit

$$\bar{X}_n = \theta + \frac{1}{n} \sum_{i=1}^n \varepsilon_i.$$

De plus, par indépendance,

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(\varepsilon_i) = \frac{\sigma^2}{n}.$$

Ainsi,

$$\sigma_{\bar{X}_n} = \frac{\sigma}{\sqrt{n}}.$$

La dispersion de la moyenne diminue donc lorsque le nombre de mesures augmente : répéter une mesure permet de réduire l'effet du bruit aléatoire. Nous allons également voir que $\bar{X}_n$ se rapproche de θ lorsque n devient grand.

Dans les deux situations, la même question apparaît : que peut-on dire de $\bar{X}_n$ lorsque le nombre n d'observations devient grand ?

La *loi des grands nombres* répond précisément à cette question : sous des hypothèses appropriées, la moyenne empirique $\bar{X}_n$ se rapproche de l'espérance μ lorsque n augmente. Elle fournit ainsi la justification mathématique fondamentale de l'utilisation d'une moyenne d'observations pour estimer une espérance inconnue.

Le *théorème central limite*, étudié dans la section suivante, ira plus loin en décrivant les fluctuations de $\bar{X}_n$ autour de μ . Il permettra ensuite, dans la partie statistique du cours, de quantifier l'incertitude associée à une estimation et de construire des procédures de décision.

Conséquence pratique. Multiplier le nombre de mesures par 4 divise l'écart-type de la moyenne par 2. Pour gagner un facteur 10 sur l'incertitude statistique, il faut donc environ 100 fois plus de mesures.

Pour une suite donnée de mesures $X_1, X_2, \dots$, on peut représenter l'évolution de la moyenne

$\bar{X}_n$ lorsque le nombre n de mesures utilisées augmente. Une nouvelle répétition de l'expérience produit une autre trajectoire de la moyenne empirique.

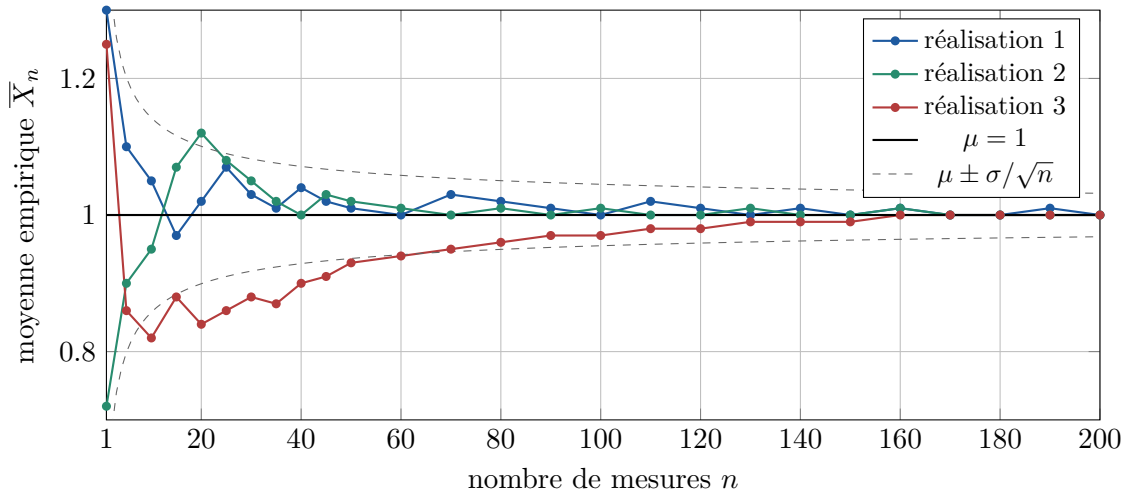


FIGURE 5.1 – Trois réalisations de la moyenne empirique $\bar{X}_n$ obtenues à partir de suites indépendantes de mesures de même espérance $\mu = 1$ et d'écart-type $\sigma = 0.45$. Les trajectoires fluctuent différemment au début, mais se stabilisent progressivement autour de μ lorsque le nombre de mesures augmente. Les courbes en pointillés représentent l'ordre de grandeur $\mu \pm \sigma/\sqrt{n}$ des fluctuations typiques de la moyenne empirique ; elles ne constituent pas des bornes.

De manière générale, on considère une variable aléatoire X d'espérance $\mu = \mathbb{E}[X]$ et de variance $\sigma^2 = \text{Var}[X]$. Soient $X_1, \dots, X_n$ des variables aléatoires *indépendantes et identiquement distribuées* (*i.i.d*) selon la même loi que X . Elles correspondent à n répétitions indépendantes de l'expérience liée à X .

Définition 5.1.1. Soient $X_1, \dots, X_n$ des variables aléatoires i.i.d selon la même loi que X . La variable aléatoire

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$$

est la *moyenne empirique* des variables $X_1, \dots, X_n$.

Cette nouvelle variable aléatoire vérifie les deux propriétés suivantes.

Proposition 5.1.2. Soit X une variable aléatoire d'espérance $\mu = \mathbb{E}[X]$ et de variance $\sigma^2 = \text{Var}[X]$. Soient $X_1, \dots, X_n$ des variables aléatoires i.i.d selon la même loi que X . La moyenne empirique vérifie

$$\mathbb{E}[\bar{X}_n] = \mu \quad \text{et} \quad \text{Var}[\bar{X}_n] = \frac{\sigma^2}{n}.$$

Démonstration. On a

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n} \mathbb{E}\left[\sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n}(n\mu) = \mu$$

et

$$\text{Var}[\bar{X}_n] = \frac{1}{n^2} \text{Var}\left[\sum_{i=1}^n X_i\right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}.$$

■

Donc la distribution de $\bar{X}_n$ reste d'espérance μ quel que soit n , et devient de plus en plus concentrée quand n grandit.

Ainsi, pour n grand, la probabilité est très grande que $\bar{X}_n$ prenne une valeur proche de $\mu = \mathbb{E}[X]$. Le résultat suivant montre que cette probabilité peut s'approcher arbitrairement de 1.

Théorème 5.1.3 (Loi faible des grands nombres). *Soient $X_1, X_2, \dots$ des variables aléatoires i.i.d d'espérance μ et de variance σ^2 . Alors pour tout $\varepsilon > 0$,*

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) = 0.$$

Démonstration. Fixons $\varepsilon > 0$. Puisque $\mathbb{E}[\bar{X}_n] = \mu$ et $\text{Var}[\bar{X}_n] = \frac{\sigma^2}{n}$, l'inégalité de Tchebychev permet d'écrire

$$\mathbb{P}\left(|\bar{X}_n - \mu| \geq a \frac{\sigma}{\sqrt{n}}\right) = \mathbb{P}\left(|\bar{X}_n - \mathbb{E}[\bar{X}_n]| \geq a \sqrt{\text{Var}[\bar{X}_n]}\right) \leq \frac{1}{a^2}$$

pour tout $a > 0$. Pour $a = \sqrt{n}\varepsilon/\sigma$, on obtient

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

qui tend vers 0 lorsque n tend vers l'infini. ■

Il existe une version plus *forte* de la loi des grands nombres. Nous en admettrons la preuve (ainsi que le fait que cette version forte implique la version faible).

Théorème 5.1.4 (Loi forte des grands nombres). *Soient $X_1, X_2, \dots$ des variables aléatoires i.i.d d'espérance μ et de variance σ^2 . Alors*

$$\mathbb{P}\left(\lim_{n \rightarrow +\infty} \bar{X}_n = \mu\right) = 1.$$

Cela signifie qu'il est *certain* que la suite $\bar{X}_n$ va tendre vers la valeur μ (on dit que $\bar{X}_n$ tend *presque sûrement* vers μ).

Montrons que la loi des grands nombres permet de **justifier l'approche fréquentiste des probabilités** présentée en début de cours. Pour cela, considérons une expérience aléatoire et $(\Omega, \mathcal{F}, \mathbb{P})$ l'espace probabilisé associé. Soit $A \in \mathcal{F}$ un événement fixé dont on souhaite estimer la probabilité $p = \mathbb{P}(A)$. Notons X la variable aléatoire qui vaut 1 si A se produit et 0 sinon. Soient $X_1, \dots, X_n$ des variables aléatoires i.i.d selon X (cela correspond donc à n répétitions indépendantes de l'expérience aléatoire). Alors $X_1 + \dots + X_n$ compte le nombre de fois que l'événement A s'est réalisé lors des n expériences.

On a $\mu = \mathbb{E}[X] = 1 \times p + 0 \times (1 - p) = p$ et la loi (forte) des grands nombres permet d'affirmer que si n tend vers l'infini

$$\frac{\text{nombre de réalisations de } A \text{ parmi les } n}{n} = \bar{X}_n \longrightarrow p$$

avec une probabilité 1.

5.2 Théorème central limite

Soit X une variable aléatoire d'espérance μ et de variance σ^2 . On vient de montrer que si $X_1, X_2, \dots$ sont i.i.d selon la même loi que X , alors

- ▷ $\mathbb{E}[\bar{X}_n] = \mu$,
- ▷ $\text{Var}[\bar{X}_n] = \frac{\sigma^2}{n}$,
- ▷ $\bar{X}_n \rightarrow \mu$ lorsque n tend vers l'infini.

Cela nous apprend que pour estimer μ , une méthode consiste à calculer $\bar{X}_n$. Cette méthode est pertinente car si $n \rightarrow +\infty$, l'estimation se fait finalement sans erreur. Par conséquent, si on répète une infinité de fois l'expérience, on pourra calculer sans se tromper $\mathbb{E}[X]$. Évidemment, on ne pourra jamais prendre $n \rightarrow +\infty$ et en pratique, on travaillera avec un nombre n d'expériences fixé. Une erreur sera commise dans l'estimation de μ par $\bar{X}_n$. On s'intéresse donc à la distribution de l'erreur $\bar{X}_n - \mu$. La difficulté vient du fait que la loi de $\bar{X}_n - \mu$ dépend a priori de la distribution de X , qui en pratique est très souvent inconnue !

Exemple 5.2.1.

- ▷ Si $X_1, \dots, X_n$ sont i.i.d. $\mathcal{N}(\mu, \sigma^2)$, alors $X_1 + X_2 + \dots + X_n \sim \mathcal{N}(n\mu, n\sigma^2)$, et donc $\bar{X}_n \sim \mathcal{N}(\frac{n\mu}{n}, \frac{n\sigma^2}{n^2}) = \mathcal{N}(\mu, \frac{\sigma^2}{n})$.
- ▷ Si $X_1, \dots, X_n$ sont i.i.d. $\text{Bern}(p) = \text{Bin}(1, p)$, alors $\bar{X}_n \sim \frac{1}{n}\text{Bin}(n, p)$.
- ▷ Si $X_1, \dots, X_n$ sont i.i.d. $\text{Pois}(\lambda)$, alors $\bar{X}_n \sim \frac{1}{n}\text{Pois}(n\lambda)$.

Mais on est sauvé par le théorème le plus important de la théorie des probabilités et des statistiques ! Si on regarde la densité de $\bar{X}_n$ dans la Figure 5.2, on se rend compte que plus n augmente, plus la courbe ressemble à la courbe de la densité de la loi normale. Ceci est vrai en toute généralité comme nous le verrons dans le Théorème Central Limite. Commençons pour

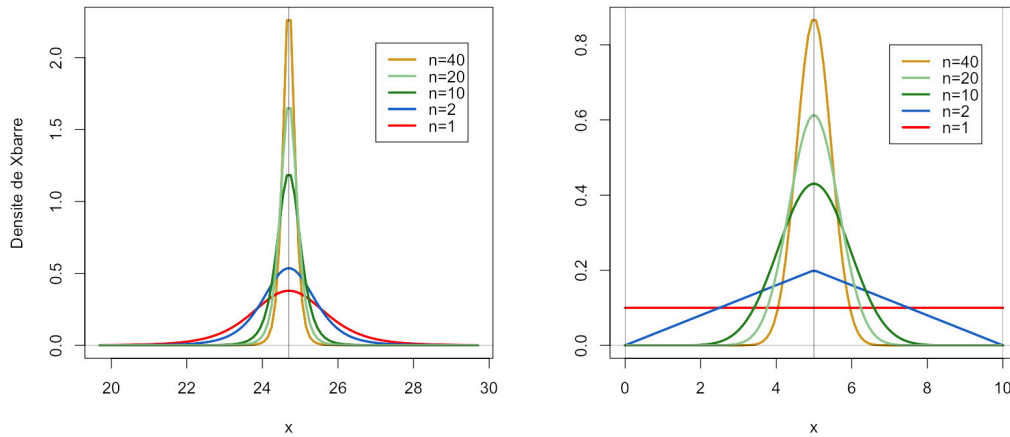


FIGURE 5.2 – Densités de $\bar{X}_n$ pour $n = 1, 2, 10, 20, 40$ dans le cas où les observations sont i.i.d. pour deux lois de X différentes. Dans les deux cas, les densités gardent la même espérance mais la variance diminue lorsque n augmente.

cela par introduire une notion de convergence.

Définition 5.2.2. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variable aléatoires et soit X une variable aléatoire. Pour tout $n \in \mathbb{N}$, on note

- ▷ F_n la fonction de répartition de X_n
- ▷ F la fonction de répartition de X .

On dit que la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si

$$\lim_{n \rightarrow +\infty} F_n(x) = F(x)$$

en tout point $x \in \mathbb{R}$ de continuité de F . Dans ce cas, on note

$$X_n \xrightarrow{\mathcal{L}} X.$$

Exemple 5.2.3.

- ▷ Si $X_n \sim \text{Bin}(n, \lambda/n)$, alors $X_n \xrightarrow{\mathcal{L}} \text{Pois}(\lambda)$.
- ▷ Si $M_n = n(1 - \max\{X_1, \dots, X_n\})$ avec X_j i.i.d. de loi uniforme sur $[0, 1]$, alors $M_n \xrightarrow{\mathcal{L}} \text{Exp}(1)$.
- ▷ Si $X_n \sim t_n$, alors $X_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$.

Les vérifications sont laissées à titre d'exercice. Dans les trois cas, la notation est abusive mais acceptée.

Une caractérisation équivalente de la convergence en loi est donnée par la convergence ponctuelle

des fonctions caractéristiques :

$$X_n \xrightarrow{\mathcal{L}} X \iff \lim_{n \rightarrow +\infty} \varphi_{X_n}(t) = \varphi_X(t) \quad \forall t \in \mathbb{R}.$$

Théorème 5.2.4 (Théorème Central Limite). *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires i.i.d de L^2 d'espérance μ et de variance σ^2 . Alors*

$$Z_n = \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sum_{j=1}^n X_j - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Démonstration. Quitte à remplacer X par $X - \mathbb{E}[X]$, on peut supposer que les variables aléatoires sont centrées. Notons φ la fonction caractéristique de X . Comme les variables aléatoires X_n sont i.i.d. de même loi que X , on a

$$\varphi_{\sqrt{n}\bar{X}_n}(t) = \mathbb{E} \left[\exp \left(it \frac{X_1 + \dots + X_n}{\sqrt{n}} \right) \right] = \left(\varphi \left(\frac{t}{\sqrt{n}} \right) \right)^n.$$

Puisque $X \in L^2$, on peut montrer que la fonction φ est de classe C^2 , et de plus on a $\varphi'(0) = i\mu = 0$ et $\varphi''(0) = -\mathbb{E}[X^2] = -\sigma^2$. En décomposant φ à l'ordre 2, on a

$$\varphi(x) = 1 + \varphi'(0)x + \varphi''(0)\frac{x^2}{2} + o(x^2) = 1 - \sigma^2\frac{x^2}{2} + o(x^2)$$

lorsque $x \rightarrow 0$. En particulier, pour tout t fixé, on a

$$\left(\varphi \left(\frac{t}{\sqrt{n}} \right) \right)^n = \left(1 - \sigma^2 \frac{t^2}{2n} + \xi_n \right)^n = \exp \left(n \log \left(1 - \sigma^2 \frac{t^2}{2n} + \xi_n \right) \right)$$

où $\xi_n = o(\frac{1}{n})$ ¹ Par conséquent, on obtient

$$\lim_{n \rightarrow +\infty} \varphi_{\sqrt{n}\bar{X}_n}(t) = \lim_{n \rightarrow +\infty} \left(1 - \sigma^2 \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \right)^n = \exp \left(-\sigma^2 \frac{t^2}{2} \right)$$

qui est la fonction caractéristique d'une loi $\mathcal{N}(0, \sigma^2)$. On a donc

$$\sqrt{n} \bar{X}_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

1. On notera que ξ_n peut être complexe. Néanmoins, on peut étendre le logarithme défini sur $]0, +\infty]$ en un logarithme complexe défini sur $\mathbb{C} \setminus]-\infty, 0]$. En particulier, pour tout complexe z tel que $|z| < 1$, on peut considérer la série de puissances suivante

$$\log(1+z) = \sum_{k=1}^{+\infty} \frac{(-1)^{k+1}}{k} z^k.$$

Cela permet d'obtenir que $\log(1+z) \sim z$ lorsque $z \rightarrow 0$.

En divisant par σ , on obtient

$$\frac{\sqrt{n} \bar{X}_n}{\sigma} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

ce qui est exactement l'énoncé dans le cas centré. Le cas général s'obtient en remplaçant X_j par $X_j - \mu$. ■

Par conséquent, si n est grand, la loi de la variable aléatoire Z_n définie par

$$Z_n = \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

est bien approximée par la loi $\mathcal{N}(0, 1)$. On pourra donc estimer des probabilités pour Z_n (et donc pour $\bar{X}_n$) en faisant comme s'il s'agissait d'une loi normale centrée et réduite. Ce qui rend le TCL si important est qu'il est vérifié quelle que soit la distribution de la variable X (qui admet une espérance et une variance).

Présentons une conséquence importante du TCL qui permet d'approcher les lois binomiales par des lois normales.

Corollaire 5.2.5 (Théorème de Moivre-Laplace). *Supposons que $(X_n)_{n \geq 1}$ est une suite de variables aléatoires telle que $X_n \sim \text{Bin}(n, p)$. Alors, on a*

$$\frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Démonstration. Soit $(Y_n)_{n \geq 1}$ une suite de variables aléatoires i.i.d. de loi de Bernoulli de paramètre p . Alors, on sait que $Y_1 + \dots + Y_n \sim \text{Bin}(n, p)$ et donc $\bar{Y}_n$ a la même loi que $\frac{X_n}{n}$. Le Théorème Central Limite permet d'écrire

$$\frac{\bar{Y}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

et donc

$$\frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

■

On a donc

$$\text{Bin}(n, p) \approx \mathcal{N}(np, np(1-p))$$

si n est grand.

Exemple 5.2.6. Supposons qu'un détecteur reçoive 100 particules et détecte chacune d'elles indépendamment avec une efficacité $p = 1/2$. On souhaite calculer la probabilité d'en détecter entre 40 et 60. Soit donc $Y \sim \text{Bin}(100, 1/2)$ le nombre de particules détectées.

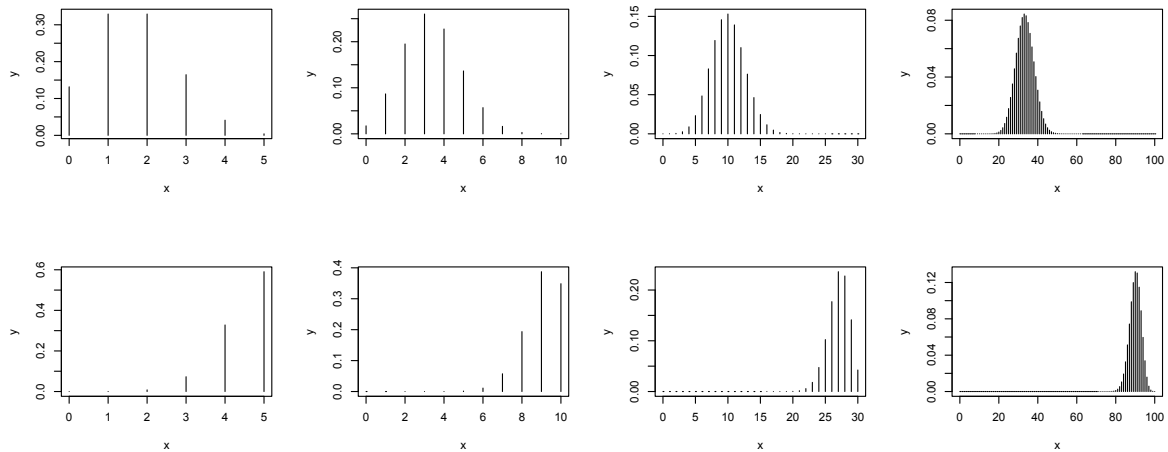


FIGURE 5.3 – Convergence vers la loi normale : diagrammes en bâtons des lois binomiales de paramètres (n, p) pour $n = 5, 10, 30, 100$ (de gauche à droite) et $p = \frac{1}{3}, \frac{9}{10}$ (haut et bas).

On a $Y \approx \mathcal{N}(50, 25)$ et on calcule

$$\begin{aligned}
 \mathbb{P}(40 < Y \leq 60) &= \mathbb{P}\left(\frac{40 - 50}{5} < \frac{Y - 50}{5} \leq \frac{60 - 50}{5}\right) \\
 &= \mathbb{P}\left(-2 < \frac{Y - 50}{5} \leq 2\right) \\
 &\approx \mathbb{P}(-2 < Z \leq 2) \\
 &\approx 0.9545
 \end{aligned}$$

où $Z \sim \mathcal{N}(0, 1)$. Il y a donc à peu près 95,45% de chance de détecter entre 40 et 60 particules sur 100 dans ce modèle. On aurait pu calculer cette valeur de manière exacte :

$$\begin{aligned}
 \mathbb{P}(40 < Y \leq 60) &= \mathbb{P}(Y = 41) + \mathbb{P}(Y = 42) + \dots + \mathbb{P}(Y = 60) \\
 &= \sum_{k=41}^{60} C_{100}^k \left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{100-k} \approx 0.9540.
 \end{aligned}$$

Cette valeur exacte est très proche de l'approximation fournie par le TCL.

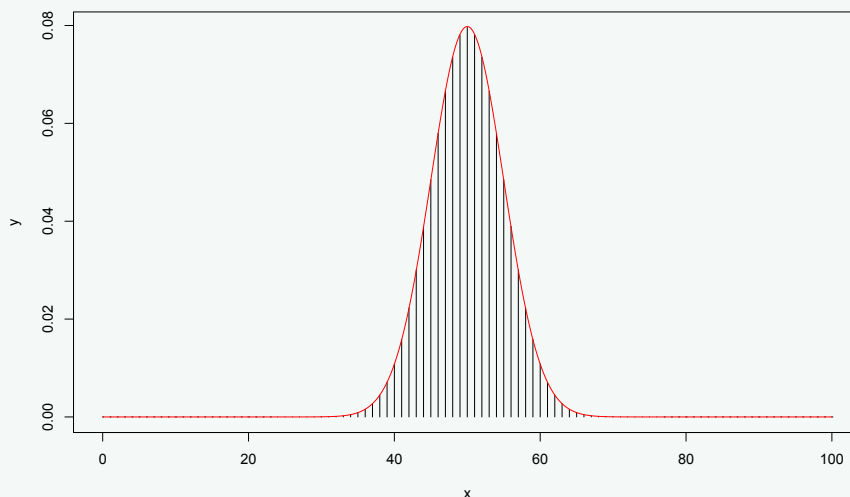


FIGURE 5.4 – Diagramme en bâtons de la loi $\text{Bin}(100, 1/2)$ et densité de la loi normale $\mathcal{N}(50, 25)$.

5.3 Une application : méthode de Monte-Carlo

La loi des grands nombres transforme une espérance en une procédure numérique. Si $U \sim \mathcal{U}([0, 1])$ et si f est intégrable, alors

$$I = \int_0^1 f(x) dx = \mathbb{E}[f(U)].$$

Si $U_1, \dots, U_n$ sont i.i.d. uniformes sur $[0, 1]$, on définit l'estimateur de Monte-Carlo

$$I_n = \frac{1}{n} \sum_{j=1}^n f(U_j).$$

La loi des grands nombres donne $I_n \rightarrow I$ presque sûrement. Si $f(U)$ est de carré intégrable, alors

$$\text{Var}(I_n) = \frac{\text{Var}(f(U))}{n},$$

et le TCL montre que l'erreur typique est de l'ordre de $1/\sqrt{n}$.

Pourquoi est-ce utile en physique ? La même idée permet d'estimer des intégrales difficiles, des moyennes sur des configurations aléatoires et des observables issues de simulations. L'avantage majeur est que la vitesse statistique $1/\sqrt{n}$ ne dépend pas directement de la dimension de l'espace d'intégration, ce qui rend les méthodes de Monte-Carlo particulièrement utiles en grande dimension.

Exemple 5.3.1. Pour estimer $I = \int_0^1 e^{-x^2} dx$, on simule $U_1, \dots, U_n \sim \mathcal{U}([0, 1])$ et on utilise

$$I_n = \frac{1}{n} \sum_{j=1}^n e^{-U_j^2}.$$

Une estimation de l'incertitude statistique est

$$\frac{s_n}{\sqrt{n}},$$

où s_n^2 est la variance empirique des valeurs $e^{-U_j^2}$.

5.4 Extension au cas multivarié

Les résultats présentés concernant la convergence en loi s'adaptent facilement dans le cas de vecteurs aléatoires.

Définition 5.4.1. Soit $(\mathbf{X}_n)_n$ une suite de vecteurs aléatoires de dimension d et soit $\mathbf{X}$ un vecteur aléatoire de dimension d . Pour tout $n \in \mathbb{N}$, on note

- ▷ F_n la fonction de répartition de $\mathbf{X}_n$
- ▷ F la fonction de répartition de $\mathbf{X}$.

On dit que la suite $(\mathbf{X}_n)_{n \in \mathbb{N}}$ converge en loi vers $\mathbf{X}$ si

$$\lim_{n \rightarrow +\infty} F_n(x) = F(x)$$

en tout point $x \in \mathbb{R}^d$ de continuité de F . Dans ce cas, on note

$$\mathbf{X}_n \xrightarrow{\mathcal{L}} \mathbf{X}.$$

A nouveau, on a

$$\mathbf{X}_n \xrightarrow{\mathcal{L}} \mathbf{X} \iff \lim_{n \rightarrow +\infty} \varphi_{\mathbf{X}_n}(t) = \varphi_{\mathbf{X}}(t) \quad \forall t \in \mathbb{R}^d.$$

Si $\mathbf{X}_n \xrightarrow{\mathcal{L}} \mathbf{X}$, alors on a la convergence en loi des marginales, mais l'inverse n'est pas vrai². On peut néanmoins avoir de meilleures propriétés si une des deux suites converge vers une constante.

Théorème 5.4.2 (Slutsky). Soient $(\mathbf{X}_n)_n$ et $(\mathbf{Y}_n)_n$ deux suites de vecteurs aléatoires de dimension d et d' respectivement. Si $\mathbf{X}_n \xrightarrow{\mathcal{L}} \mathbf{X}$ et si $\mathbf{Y}_n \xrightarrow{\mathcal{L}} \mathbf{c}$ où $\mathbf{c} \in \mathbb{R}^{d'}$, alors

$$(\mathbf{X}_n, \mathbf{Y}_n) \xrightarrow{\mathcal{L}} (\mathbf{X}, \mathbf{c}).$$

2. Il faut en fait avoir la convergence en loi de toutes les combinaisons linéaires des composantes de $\mathbf{X}_n$.

Corollaire 5.4.3. Soient $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ deux suites de variables aléatoires. Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} c$ où $c \in \mathbb{R}$, alors

- ▷ $X_n + Y_n \xrightarrow{\mathcal{L}} X + c$
- ▷ $X_n Y_n \xrightarrow{\mathcal{L}} cX$
- ▷ si $c \neq 0$, alors $\frac{X_n}{Y_n} \xrightarrow{\mathcal{L}} \frac{X}{c}$.

Afin d'énoncer un TCL multivarié, il faut étendre la notion de loi gaussienne au cas d -dimensionnel.

Définition 5.4.4. Soient $Z_1, \dots, Z_d$ des variables aléatoires i.i.d. selon une loi $\mathcal{N}(0, 1)$. Alors, la loi du vecteur aléatoire $\mathbf{Z} = (Z_1, \dots, Z_d)$ est donnée par la densité jointe

$$f_{\mathbf{Z}}(x_1, \dots, x_d) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{x_1^2 + \dots + x_d^2}{2}} = \frac{1}{(2\pi)^{d/2}} e^{-\frac{\|\mathbf{x}\|^2}{2}}.$$

- ▷ son vecteur moyen est le vecteur nul,
- ▷ sa matrice de variance-covariance est la matrice Id,
- ▷ on note

$$\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \text{Id}).$$

Si $\mathbf{X}$ est un vecteur aléatoire tel que

$$\mathbf{X} = D\mathbf{Z} + \mu$$

où D est une matrice carrée de dimension d telle que $DD^T = \Sigma$ et μ un vecteur de dimension d , alors on dit que $\mathbf{X}$ suit une loi gaussienne multivariée et on note

$$\mathbf{X} \sim \mathcal{N}(\mu, \Sigma).$$

Son vecteur moyen est μ et sa matrice de variance-covariance Σ . Ce vecteur admet une densité si et seulement si la matrice Σ est inversible.

Le TCL multivarié prend la forme suivante.

Théorème 5.4.5 (TCL multivarié). Soit $(\mathbf{X}_n)_n$ une suite de vecteurs aléatoires de dimension d i.i.d de vecteur moyen μ et de matrice de variance-covariance Σ . Alors

$$\sqrt{n}(\bar{\mathbf{X}}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \Sigma).$$

Bibliographie

- [1] M. Baron (2014), *Probability and Statistics for Computer Scientists*, 2nd ed. CRC Press.
- [2] J. K. Blitzstein, J. Hwang (2019), *Introduction to Probability*, 2nd ed. CRC Press.
- [3] F. M. Dekking, C. Kraaikamp, H. P. Lopuhaä, L. E. Meester (2007), *A Modern Introduction to Probability and Statistics : Understand Why and How*. Springer Texts in Statistics.
- [4] C. Esser and Y. Swan (2019), *Un second cours de probabilité*. Université de Liège.
- [5] G. Haesbroeck (2007), *Probabilité et Statistique I*. Université de Liège.
- [6] S. Ross (2010), *A First Course in Probability*, 8th ed. Pearson Prentice Hall.
- [7] Y. Swan (2018), *Un premier cours de probabilité*. Université de Liège.
- [8] D. Paindaveine (2018), *Probabilités*. Université libre de Bruxelles – Solvay.