

Éléments du calcul des probabilités

Céline Esser

Année académique 2022–2023

Table des matières

1	Événements et probabilités	3
1.1	Théorie des ensembles	3
1.2	Approche fréquentiste de la notion de probabilité	4
1.3	Mesures de probabilité	6
1.4	Cas équiprobable	9
2	Probabilités conditionnelles et indépendance	13
2.1	Probabilités conditionnelles	13
2.2	Indépendance	15
2.3	Loi des probabilités totales	17
2.4	Formule de Bayes	19
3	Variables aléatoires	22
3.1	Variables aléatoires	22
3.2	Variables aléatoires discrètes	24
3.3	Moments d'une variable aléatoire discrète	26
3.4	Lois discrètes usuelles	30
3.4.1	Loi discrète uniforme	30
3.4.2	Loi de Bernoulli et loi binomiale	31
3.4.3	Loi géométrique	32
3.4.4	Loi de Poisson	34
3.4.5	Logiciel R	35
3.5	Variables aléatoires continues	36
3.6	Moments d'une variable aléatoire continue	37
3.7	Lois continues usuelles	39
3.7.1	Loi continue uniforme	39
3.7.2	Loi exponentielle	40
3.7.3	Loi normale	42

4 Lois jointes	46
4.1 Loi jointe et fonction de répartition	46
4.2 Lois bivariées discrètes	47
4.3 Lois bivariées continues	55
4.4 Lois bivariées mixtes	62
4.5 Indépendance de variables aléatoires	63
4.6 Covariance et corrélation	67
5 Théorèmes limites	72
5.1 Loi des grands nombres	72
5.2 Théorème central limite	75
References	78

Chapitre 1

Événements et probabilités

1.1 Théorie des ensembles

Définition 1.1.1 Un *ensemble* Ω est une collection d'objets. Un objet ω appartenant à Ω est un *élément* de Ω . On note $\omega \in \Omega$.

Les éléments d'un ensemble peuvent être de n'importe quelle nature comme illustré dans les exemples suivants.

Exemple 1.1.2

- Lundi est un élément de l'ensemble des jours de la semaine, i.e.

$$\text{Lundi} \in \{\text{Lundi, Mardi, Mercredi, Jeudi, Vendredi, Samedi, Dimanche}\}.$$

- Un alphabet est un ensemble de lettres.
- L'ULiège est un élément de l'ensemble des universités belges.
- La fonction f définie par $f(x) = \sin(x)$, $x \in \mathbb{R}$, est un élément de l'ensemble des fonctions continues sur \mathbb{R} .

Définition 1.1.3 Un *sous-ensemble* A d'un ensemble Ω est une collection d'éléments de Ω . On note $A \subseteq \Omega$.

Exemple 1.1.4

- La collection $\{\text{Lundi, Mardi, Jeudi}\}$ est un sous-ensemble de l'ensemble des jours de la semaine.
- Les voyelles forment un sous-ensemble de l'ensemble de l'alphabet.
- La collection $\{\text{ULiège}\}$ est un sous-ensemble de l'ensemble des universités belges.
- Les fonctions dérivables forment un sous-ensemble de l'ensemble des fonctions continues sur \mathbb{R} .

Rappelons les relations qui peuvent exister entre des ensembles ainsi que les constructions classiques d'ensembles à partir d'ensembles donnés.

Définition 1.1.5 Soit Ω un ensemble et soient A et B des sous-ensembles de Ω .

- Inclusion : on note $A \subseteq B$ si tout élément de A est un élément de B , i.e.

$$x \in A \Rightarrow x \in B.$$

- Égalité : on note $A = B$ si $A \subseteq B$ et $B \subseteq A$.
- Intersection : on pose $A \cap B = \{\omega \in \Omega : \omega \in A \text{ et } \omega \in B\}$.
- Union : on pose $A \cup B = \{\omega \in \Omega : \omega \in A \text{ ou } \omega \in B\}$.
- Complémentaire : on pose $A^c = \Omega \setminus A = \{\omega \in \Omega : \omega \notin A\}$.
- Différence : on pose $A \setminus B = \{\omega \in \Omega : \omega \in A \text{ et } \omega \notin B\} = A \cap B^c$.

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 1.1.6 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 $(A \cap B)^c = A^c \cup B^c$.

Définition 1.1.7 Soit Ω un ensemble. Soient A, B deux sous-ensembles de Ω . On dit que A et B sont *disjoints* si $A \cap B = \emptyset$.

Enfin, nous introduisons une notation pour l'ensemble de toutes les parties d'un ensemble donné.

Définition 1.1.8 L'ensemble des parties de Ω , noté $\mathcal{P}(\Omega)$, est l'ensemble des sous-ensembles de Ω .

Par exemple,

$$\mathcal{P}(\{1, 2, 3\}) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

1.2 Approche fréquentiste de la notion de probabilité

Considérons une expérience aléatoire, c'est-à-dire une expérience dont l'issue n'est pas connue avec certitude. 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.

Exemple 1.2.1

- 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 pièces de monnaie indistinguables,

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

- 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 .

Exemple 1.2.2

- “Le deuxième lancer est pile” correspond au sous-ensemble $\{(Pile,Pile), (Face,Pile)\}$.
- “La somme des deux dés est égale à 8” correspond au sous-ensemble

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

- “Le gagnant appartient au Big Four” correspond au sous-ensemble

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

- “Le bus arrive dans 10 minutes maximum” correspond au sous-ensemble $[0, 10]$.

Définition 1.2.3 (Approche intuitive) La *probabilité* d'un événement A , que l'on notera $\mathbb{P}[A]$, est un nombre entre 0 et 1 qui mesure la confiance que l'on a en le fait que l'événement A va se produire.

Exemple 1.2.4

- “Le deuxième lancer est pile” a intuitivement une probabilité $1/2$.
- Pour 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.
- 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.
- 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 1.2.5 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	0,1674

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.

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 $A_1, A_2, \dots \in \mathcal{F}$, alors $A_1 \cup A_2 \cup \dots \in \mathcal{F}$.

Évidemment, les deux premières propriétés impliquent que $\emptyset \in \mathcal{F}$. De plus, puisque

$$(A_1 \cap A_2 \cap \dots)^c = A_1^c \cup A_2^c \cup \dots,$$

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

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

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

(i.e. une application qui à chaque événement A associe une valeur $\mathbb{P}[A]$ comprise entre 0 et 1) telle que

1. l'univers a une probabilité 1 : $\mathbb{P}[\Omega] = 1$,
2. σ -additivité : si A_1, A_2, A_3, \dots sont deux à deux disjoints (suite finie ou infinie), alors

$$\mathbb{P}[A_1 \cup A_2 \cup A_3 \cup \dots] = \mathbb{P}[A_1] + \mathbb{P}[A_2] + \mathbb{P}[A_3] + \dots$$

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

Remarquons que si la suite A_1, A_2, A_3, \dots est infinie, la σ -additivité s'écrit

$$\mathbb{P}\left[\bigcup_{j=1}^{+\infty} A_j\right] = \sum_{j=1}^{+\infty} \mathbb{P}[A_j]$$

et qu'il s'agit donc d'une série convergente (puisque une probabilité appartient à $[0, 1]$).

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

1. Complémentaire : $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$
2. Vide : $\mathbb{P}[\emptyset] = 0$
3. Monotonie : si $A \subseteq B$, alors $\mathbb{P}[A] \leq \mathbb{P}[B]$ et $\mathbb{P}[B \setminus A] = \mathbb{P}[B] - \mathbb{P}[A]$
4. Union : $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$

Démonstration : 1. 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 premier point.

2. Comme $\emptyset = \Omega^c$, on a

$$\mathbb{P}[\emptyset] = 1 - \mathbb{P}[\Omega] = 0$$

par le premier point.

3. 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

$$\mathbb{P}[B \setminus A] = \mathbb{P}[B] - \mathbb{P}[A].$$

4. 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 3. ■

Exemple 1.3.6 On considère le lancer d'un dé. On sait que $\Omega = \{1, 2, 3, 4, 5, 6\}$ et $\mathcal{F} = \mathcal{P}(\Omega)$.

- Complémentaire : La probabilité d'obtenir 1 ou 2 ($A = \{1, 2\}$) est égale à 1 moins la probabilité d'obtenir 3, 4, 5 ou 6 ($A^c = \{3, 4, 5, 6\}$).
- Vide : La probabilité d'obtenir aucun résultat possible (un résultat n'appartenant pas à $\{1, 2, 3, 4, 5, 6\}$) est nulle.
- Monotonie : La probabilité d'obtenir 2 ou 6 ($A = \{2, 6\}$) est plus petite ou égale à la probabilité d'obtenir un nombre pair ($B = \{2, 4, 6\}$).
- Union : La probabilité d'obtenir 1 ou 2 ($A = \{1, 2\}$) ou un nombre pair ($B = \{2, 4, 6\}$) est égale à la probabilité d'obtenir un résultat appartenant à $\{1, 2, 4, 6\}$. Ainsi, c'est bien la probabilité d'obtenir 1 ou 2, plus la probabilité d'obtenir un nombre pair, moins la probabilité d'obtenir 2 car on l'a compté deux fois!

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 *stochastiquement incertain* si $\mathbb{P}[A] = 0$.

Exemple 1.3.8 On répète une infinité de fois un lancer de pièce de monnaie et on s'intéresse à l'événement $A = \text{“Le résultat est toujours PILE”}$. On a

$$\mathbb{P}[A] = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \dots = 0$$

mais remarquons que $A \neq \emptyset$.

1.4 Cas équiprobable

Le cas du lancé 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 suivant 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 $n_1 n_2$.

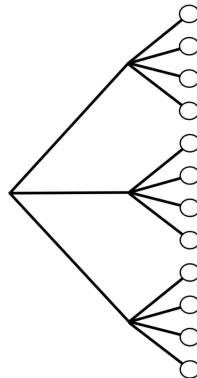


FIGURE 1.1 – Il y a 3 issues pour la première expérience et pour chacune de ces issues, il y a 4 issues possibles. Au total, on a 12 résultats possibles.

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 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 (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

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

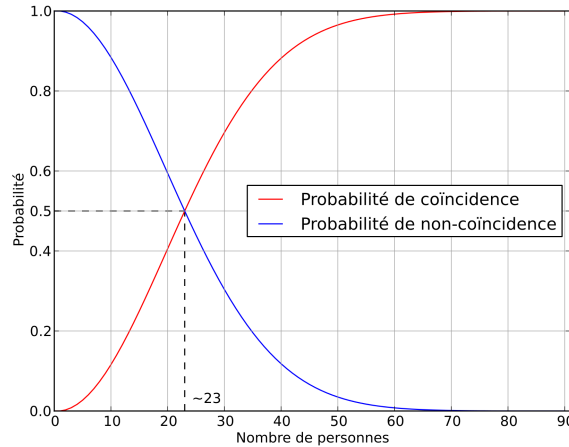
Dans notre situation, on a

- 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

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

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



Chapitre 2

Probabilités conditionnelles et indépendance

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

$$\boxed{\mathbb{P}[A|B] = \text{la probabilité de } A \text{ sachant } B.}$$

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 choisit 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 ensembles A_1, A_2, \dots sont deux à deux disjoints, on a

$$\mathbb{P}\left[\bigcup_n A_n | B\right] = \frac{\mathbb{P}[(\bigcup_n A_n) \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[\bigcup_n (A_n \cap B)]}{\mathbb{P}[B]} = \frac{\sum_n \mathbb{P}[A_n \cap B]}{\mathbb{P}[B]} = \sum_n \mathbb{P}[A_n | B]$$

puisque les événements $A_1 \cap B, A_2 \cap B, \dots$ 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 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 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 2.2.1 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 2.2.2 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 2.2.3 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 $\mathbb{P}[A] = 4/52$, $\mathbb{P}[B] = 1/4$ et $\mathbb{P}[A \cap B] = 1/52 = \mathbb{P}[A]\mathbb{P}[B]$. Les événements A et B sont donc indépendants.

Proposition 2.2.4 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 2.2.5 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!

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

Définition 2.2.6 On dit que des événements A_1, \dots, A_n sont *indépendants* si pour toute

partie finie J de $\{1, \dots, n\}$, 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 2.2.7 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

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

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, la définition d'une mesure de probabilité implique que

$$\mathbb{P}[A] = \mathbb{P}[A \cap B] + \mathbb{P}[A \cap B^c] = \mathbb{P}[A|B]\mathbb{P}[B] + \mathbb{P}[A|B^c]\mathbb{P}[B^c]$$

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.

Proposition 2.3.2 (Loi des probabilités totales) Soient B_1, \dots, B_n des événements de probabilités non-nulles tels que

$$\Omega = B_1 \cup \dots \cup B_n \quad \text{et} \quad B_i \cap B_j = \emptyset \quad \text{pour tous } i \neq j.$$

Alors pour tout événement A , on a

$$\mathbb{P}[A] = \sum_{j=1}^n \mathbb{P}[A|B_j]\mathbb{P}[B_j].$$

Démonstration : Tout événement A peut s'écrire sous la forme de l'union disjointe des événements $A \cap B_1, \dots, A \cap B_n$. Par définition d'une mesure de probabilité, on a donc

$$\mathbb{P}[A] = \mathbb{P}[(A \cap B_1) \cup \dots \cup (A \cap B_n)] = \mathbb{P}[A \cap B_1] + \dots + \mathbb{P}[A \cap B_n].$$

On conclut en remarquant que $\mathbb{P}[A \cap B_j] = \mathbb{P}[A|B_j]\mathbb{P}[B_j]$ pour tout $j \in \{1, \dots, n\}$. ■

Remarque 2.3.3 On peut considérer que cette formule reste valide si $\mathbb{P}[B_j] = 0$ car même si $\mathbb{P}[A|B_j]$ n'est pas bien défini, on le multiplie ensuite par 0.

Remarque 2.3.4 En utilisant les mêmes arguments que dans la preuve, cette loi reste valide si on considère une famille dénombrable d'événements $(B_n)_{n \in \mathbb{N}}$ deux-à-deux disjoints tels que $\Omega = \bigcup_{n \in \mathbb{N}} B_n$. Ainsi, on a

$$\mathbb{P}[A] = \sum_{n \in \mathbb{N}} \mathbb{P}[A|B_n]\mathbb{P}[B_n]$$

pour tout événement A .

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 I) Soient A et B deux événements de probabilités non-nulles. On a

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B]}.$$

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. ■

Exemple 2.4.2 Les tests médicaux ne sont pas fiables à 100% : on peut avoir des faux positifs et des faux négatifs. On suppose avoir les informations suivantes à propos d'un virus et d'un test :

- la proportion de personnes infectées dans la population est de 10%
- le test donne une réponse positive dans 95% des cas où le patient est infecté
- 14% des tests effectués sont positifs.

On souhaite tester l'efficacité du test en calculant la probabilité d'être infecté alors que le test est négatif. On définit les événements

- A = le patient est infecté,
- B = le test est négatif.

On sait que

$$\mathbb{P}[A] = \frac{10}{100}, \quad \mathbb{P}[B] = \frac{86}{100} \quad \text{et} \quad \mathbb{P}[B|A] = 1 - \mathbb{P}[B^c|A] = \frac{5}{100}.$$

Par la formule de Bayes, on a

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B]} = \frac{\frac{5}{100} \cdot \frac{10}{100}}{\frac{86}{100}} = \frac{5}{860}.$$

La deuxième formule de Bayes combine la première avec la loi des probabilités totales.

Proposition 2.4.3 (Formule de Bayes II) Soient A et B deux événements de probabilités non-nulles. On a

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B|A]\mathbb{P}[A] + \mathbb{P}[B|A^c]\mathbb{P}[A^c]}.$$

Démonstration : C'est immédiat en utilisant la formule de Bayes I et la loi des probabilités totales qui donne $\mathbb{P}[B] = \mathbb{P}[B|A]\mathbb{P}[A] + \mathbb{P}[B|A^c]\mathbb{P}[A^c]$. ■

Exemple 2.4.4 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

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

Comme pour la loi des probabilités totales, on peut étendre la formule de Bayes au cas de plusieurs événements deux-à-deux disjoints.

Proposition 2.4.5 (Formule de Bayes III) Soient A_1, \dots, A_n des événements de probabilités non-nulles tels que

$$\Omega = A_1 \cup \dots \cup A_n \quad \text{et} \quad A_i \cap A_j = \emptyset \quad \text{pour tous } i \neq j.$$

Si B est un événement de probabilité non-nulle, on a

$$\mathbb{P}[A_i|B] = \frac{\mathbb{P}[B|A_i]\mathbb{P}[A_i]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B|A_i]\mathbb{P}[A_i]}{\sum_{j=1}^n \mathbb{P}[B|A_j]\mathbb{P}[A_j]}.$$

Démonstration : On utilise à nouveau la formule de Bayes I et la loi des probabilités totales qui permet d'écrire le dénominateur sous la forme

$$\mathbb{P}[B] = \sum_{j=1}^n \mathbb{P}[B|A_j]\mathbb{P}[A_j].$$

■

Exemple 2.4.6 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 ils sont tous les trois distraits, on a

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

Ainsi par la formule de Bayes III, 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 !

Chapitre 3

Variables aléatoires

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 Une *variable aléatoire* (abrégé en v.a.) est une application

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

telle que pour tout intervalle $[a, b]$ de \mathbb{R} , l'ensemble

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

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! Par convention, si X est une variable aléatoire, on note

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

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

Exemple 3.1.4 Considérons un lancer de dé. On a $\Omega = \{1, 2, 3, 4, 5, 6\}$. On joue à un jeu de hasard dont les règles sont les suivantes.

- Si le résultat du lancer de dé est un nombre pair, on gagne 2 €.
- Si le résultat est 1, on perd 3€.
- Si le résultat est 3 ou 5, on ne gagne rien et on ne perd rien.

On s'intéresse donc à l'application

$$X : \Omega \rightarrow \mathbb{R} : \begin{cases} 1 \mapsto -3 \\ 2 \mapsto 2 \\ 3 \mapsto 0 \\ 4 \mapsto 2 \\ 5 \mapsto 0 \\ 6 \mapsto 2 \end{cases} .$$

Ce qui va nous intéresser, c'est l'argent gagné et non le résultat du dé. Pour connaître nos chances, on calcule donc

$$\mathbb{P}[X = -3] = \frac{1}{6}, \quad \mathbb{P}[X = 0] = \frac{1}{3} \quad \text{et} \quad \mathbb{P}[X = 2] = \frac{1}{2}.$$

Dans l'exemple précédent, les trois probabilités calculées $\mathbb{P}[X = -3]$, $\mathbb{P}[X = 0]$ et $\mathbb{P}[X = 2]$ nous permettent de calculer toutes les probabilités de la forme $\mathbb{P}[X \in I]$. Ces trois valeurs donnent la loi de X .

Définition 3.1.5 La *loi de X* est une description complète des probabilités $\mathbb{P}[X \in I]$ pour tout intervalle $I \subseteq \mathbb{R}$.

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 $\mathbb{P}[X \in I]$.

Etant donné que \mathbb{P} est une mesure de probabilité, on peut déjà dire par mal de choses sur les propriétés de la loi d'une variable aléatoire.

Proposition 3.1.6 Soit X une variable aléatoire. Alors

1. $\mathbb{P}[X \in \mathbb{R}] = 1$ et $\mathbb{P}[X \in I] \in [0, 1]$ pour tout intervalle $I \subseteq \mathbb{R}$,
2. si $I, J \subseteq \mathbb{R}$ sont deux intervalles, alors

$$\mathbb{P}[X \in I \cup J] = \mathbb{P}[X \in I] + \mathbb{P}[X \in J] - \mathbb{P}[X \in I \cap J],$$

3. pour tout $x \in \mathbb{R}$,

$$\mathbb{P}[X \leq x] = 1 - \mathbb{P}[X > x],$$

4. si $x \leq y$,

$$\mathbb{P}[X \leq x] \leq \mathbb{P}[X \leq y].$$

Démonstration : Tout est immédiat au vu des règles de calcul données dans la Proposition 1.3.5. ■

Définition 3.1.7 Soit X une variable aléatoire. La *fonction de répartition* de X est la fonction F_X définie pour tout $x \in \mathbb{R}$ par

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

En utilisant la Proposition 3.1.6, on obtient directement les propriétés suivantes de la fonction de répartition.

Proposition 3.1.8

1. F_X est une fonction croissante.
2. Si $a \leq b$,

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

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

En particulier, la fonction de répartition de X caractérise la loi de X . On va donc adopter la définition suivante.

Définition 3.1.9 Deux variables aléatoires X et Y ont la *même loi* si $F_X = F_Y$. Dans ce cas, on écrit $X \stackrel{\mathcal{L}}{=} Y$.

3.2 Variables aléatoires discrètes

Définition 3.2.1 Une variable aléatoire X est *discrète* si elle ne peut prendre qu'un nombre dénombrable de valeurs différentes.

Exemple 3.2.2

- Les résultats mentionnés précédemment sont tous discrets.
- Le nombre de fichiers endommagés par un virus.
- Le nombre de visites sur Wikipedia.

Exemple 3.2.3 Les variables aléatoires suivantes ne sont pas discrètes.

- La durée de vie d'un composant électronique.
- La température.
- Le temps d'attente dans une file.

Définition 3.2.4 Soit X une variable aléatoire discrète prenant les valeurs $\{x_1, \dots, x_N\}$ avec $N \in \mathbb{N}_0$ (cas fini) ou $\{x_n : n \geq 1\}$ (cas infini). La *fonction de masse* de X est la fonction p_X qui à une valeur x_n associe la probabilité que X soit égal à x_n , i.e.

$$p_X(x_n) = \mathbb{P}[X = x_n]$$

avec $n \in \{1, \dots, N\}$ ou $n \in \mathbb{N}_0$. On note

$$X \sim p_X.$$

Pour simplifier les notations, on écrira que X prend les valeurs $\{x_1, \dots, x_N\}$ avec $N \in \mathbb{N}_0$ ou $N = +\infty$. Comme X doit prendre une des valeurs x_n , on a $\Omega = \bigcup_{n=1}^N \{X = x_n\}$ et donc

$$\sum_{n=1}^N p_X(x_n) = 1.$$

Réciproquement, toute fonction $p : \{x_1, \dots, x_N\} \rightarrow [0, 1]$ avec $N \in \mathbb{N}_0$ ou $N = +\infty$ qui satisfait

$$\sum_{n=1}^N p(x_n) = 1$$

est la fonction de masse d'une variable aléatoire X .

La fonction de masse de X permet de calculer la loi de X , c'est-à-dire toutes les probabilités $\mathbb{P}[X \in I]$. En effet, on a

$$\mathbb{P}[X \in I] = \sum_{n \text{ tel que } x_n \in I} p_X(x_n).$$

Par conséquent, on l'appelle également la loi de X .

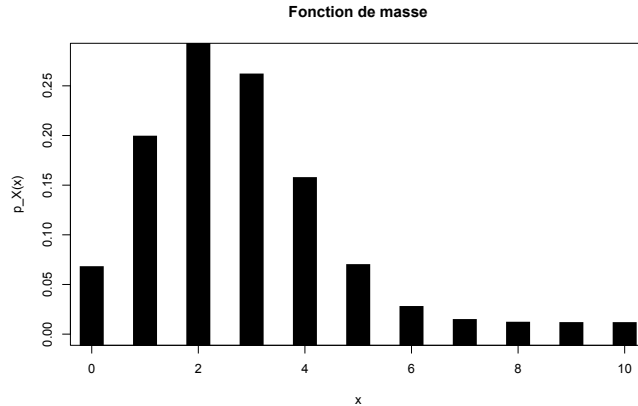
Exemple 3.2.5 On lance deux dés et on s'intéresse à la v.a. X qui donne la somme des résultats. On sait que X prend ses valeurs dans $\{2, 3, \dots, 12\}$. On a

$$\mathbb{P}[X \in [3, 6]] = \mathbb{P}[X \in \{3, 4, 5, 6\}] = \mathbb{P}[X = 3] + \mathbb{P}[X = 4] + \mathbb{P}[X = 5] + \mathbb{P}[X = 6].$$

De manière générale,

$$\mathbb{P}[X \in I] = \sum_{n \in \{2, \dots, 12\} \cap I} \mathbb{P}[X = n].$$

On peut représenter la fonction de masse grâce à un diagramme en bâtons. Par exemple, le diagramme ci-dessous donne les différentes probabilités $\mathbb{P}[X = x]$ pour une variable aléatoire X à valeurs dans $\{0, \dots, 10\}$.



3.3 Moments d’une variable aléatoire discrète

La loi d’une variable aléatoire discrète X (qui peut être donnée par sa fonction de masse ou sa fonction de répartition) donne la collection de toutes les probabilités liées à X , donc l’information complète sur son comportement. 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*. L’espérance est un indicateur de position ; la variance est un indicateur de dispersion. Avant de définir ces notions, faisons un lien avec l’approche fréquentiste des probabilités.

Soit $\{x_1, \dots, x_n\}$ un échantillon. On note y_1, \dots, y_J les modalités de cet échantillon. On forme le tableau des effectifs et des fréquences

	y_1	\dots	y_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

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

et

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

Définition 3.3.1 Soit X une variable aléatoire discrète.

— Si X prend les valeurs $\{x_1, \dots, x_N\}$ avec $N \in \mathbb{N}$ (cas fini), *l’espérance* de X est la quantité

$$\mathbb{E}[X] = \sum_{n=1}^N p_X(x_n) x_n.$$

— Si X prend les valeurs $\{x_n : n \geq 1\}$ (cas infini) et si $\sum_{n=1}^{+\infty} p_X(x_n) |x_n| < +\infty$,

l'espérance de X est la quantité

$$\mathbb{E}[X] = \sum_{n=1}^{+\infty} p_X(x_n)x_n.$$

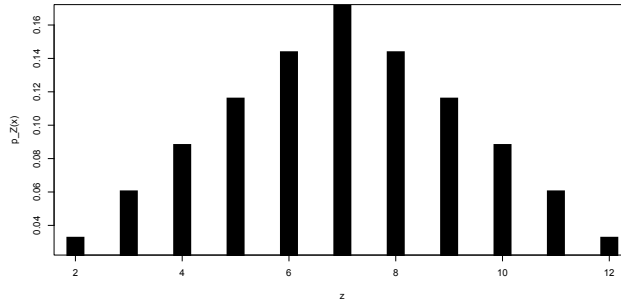
Exemple 3.3.2

— 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 calcule facilement $\mathbb{P}[Z = n]$:

$$\begin{aligned} \mathbb{P}[Z = 2] &= \frac{1}{36}, \mathbb{P}[Z = 3] = \frac{2}{36} = \frac{1}{18}, \mathbb{P}[Z = 4] = \frac{3}{36} = \frac{1}{12}, \\ \mathbb{P}[Z = 5] &= \frac{4}{36} = \frac{1}{9}, \mathbb{P}[Z = 6] = \frac{5}{36}, \mathbb{P}[Z = 7] = \frac{6}{36} = \frac{1}{6}, \mathbb{P}[Z = 8] = \frac{5}{36}, \\ \mathbb{P}[Z = 9] &= \frac{4}{36} = \frac{1}{9}, \mathbb{P}[Z = 10] = \frac{3}{36} = \frac{1}{12}, \mathbb{P}[Z = 11] = \frac{2}{36} = \frac{1}{18}, \mathbb{P}[Z = 12] = \frac{1}{36}. \end{aligned}$$



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 dans l'exemple précédent que $\mathbb{E}[Z] = \mathbb{E}[X] + \mathbb{E}[Y]$ où $Z = X + Y$. C'est vrai en toute généralité.

Proposition 3.3.3 Soient X et Y des variables aléatoires discrètes qui possèdent une espérance 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* : $\mathbb{E}[\alpha] = \alpha$.

Démonstration : Nous reviendrons sur la linéarité de l'espérance lorsque nous étudierons les lois jointes. Le deuxième point est admis et le troisième est immédiat. ■

Souvent, on peut s'intéresser à une autre variable aléatoire discrète Y qui est une *fonction de X* , c'est-à-dire $Y = g(X)$. Dans ce cas, on définit

$$\mathbb{E}[g(X)] = \sum_n g(x_n)p_X(x_n)$$

pour autant que cette expression ait du sens.

Exemple 3.3.4 Soit X le résultat du lancer d'un dé. On est dans un contexte équiprobable et on a donc $\mathbb{P}[X = n] = \frac{1}{6}$ 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

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

Définition 3.3.5 Soit X une variable aléatoire discrète. Si elle existe, la *variance* de X est la quantité

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

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

Exemple 3.3.6 Reprenons le lancer du dé. On a

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

Proposition 3.3.7

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

Démonstration : 1. Il est clair que $\text{Var}[X] \geq 0$ puisque la variance est obtenue en sommant des termes positifs ou nuls.

2. Supposons que $\text{Var}[X] = 0$. Alors les termes apparaissant dans la somme de la définition de la variance sont tous nuls, c'est-à-dire $x_n = \mathbb{E}[X]$. 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. ■

Dans la pratique, on calcule souvent la variance avec la formule suivante.

Proposition 3.3.8 Soit X une variable aléatoire qui admet une espérance et une variance. Alors

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

Démonstration : 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. ■

Exemple 3.3.9 Pour le lancer du dé, on a

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

L'inégalité célèbre suivante montre que l'espérance d'une variable aléatoire est bien un indicateur de *position* tandis que sa variance donne une indication sur sa *dispersion*.

Proposition 3.3.10 (Inégalité de Tchebychev) Soit X une variable aléatoire qui admet une espérance μ et une variance σ^2 . Alors pour tout $r > 0$,

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

Démonstration : On définit une deuxième variable aléatoire Y en posant

$$Y = \begin{cases} 1 & \text{si } |X - \mu|^2 \geq r^2\sigma^2, \\ 0 & \text{sinon.} \end{cases}$$

Par définition, on a $Y \leq \frac{|X - \mu|^2}{r^2\sigma^2}$ et par monotonie de l'espérance, on a donc

$$\mathbb{E}[Y] \leq \mathbb{E}\left[\frac{|X - \mu|^2}{r^2\sigma^2}\right].$$

Or, d'une part, on a

$$\mathbb{E}[Y] = 1 \cdot \mathbb{P}[Y = 1] + 0 \cdot \mathbb{P}[Y = 0] = \mathbb{P}[|X - \mu|^2 \geq r^2\sigma^2] = \mathbb{P}[|X - \mu| \geq r\sigma].$$

D'autre part, par linéarité de l'espérance, on a

$$\mathbb{E}\left[\frac{|X - \mu|^2}{r^2\sigma^2}\right] = \frac{1}{r^2\sigma^2} \mathbb{E}[|X - \mu|^2] = \frac{1}{r^2\sigma^2} \sigma^2 = \frac{1}{r^2}$$

puisque $\mu = \mathbb{E}[X]$ et $\sigma^2 = \text{Var}[X]$. Ainsi, l'inégalité devient

$$\mathbb{P}[|X - \mu| \geq r\sigma] \leq \frac{1}{r^2},$$

ce qui permet de conclure. ■

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 $8/9 \approx 0,888$.

3.4 Lois discrètes usuelles

On peut montrer que 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.

3.4.1 Loi discrète uniforme

Définition 3.4.1 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 3.4.2

- 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 3.4.3 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 par la Proposition 3.3.8,

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

■

3.4.2 Loi de Bernoulli et loi binomiale

Définition 3.4.4 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 3.4.5

- 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 3.4.6 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 par la Proposition 3.3.8,

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

■

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 3.4.7 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 $k \in \{0, \dots, n\}$.

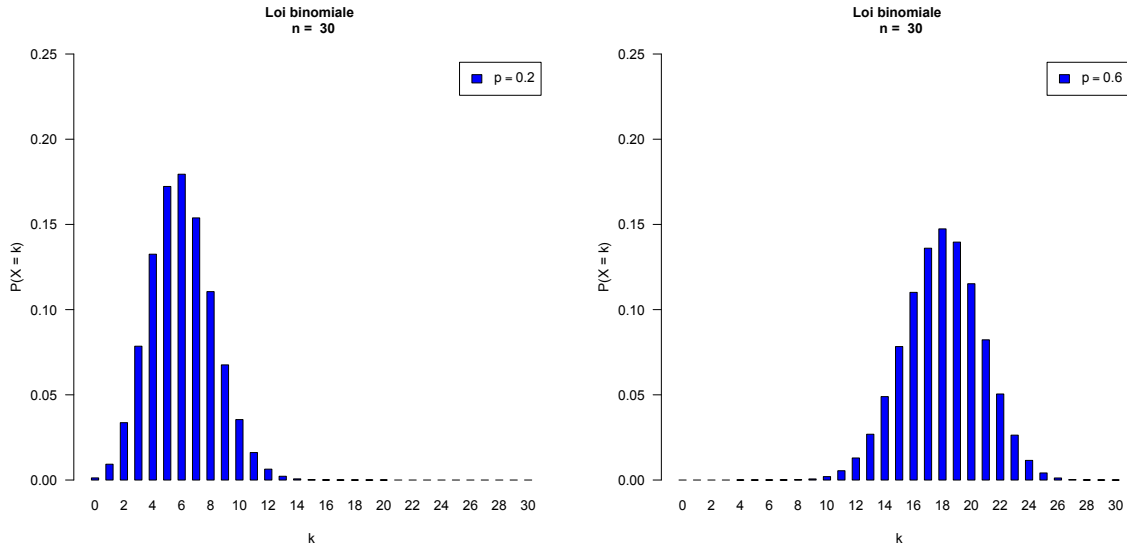


FIGURE 3.1 – Diagrammes en bâtons de lois binomiales de paramètres $(30, 0.2)$ et $(30, 0.6)$.

Exemple 3.4.8

- Nombre de “pile” obtenus lors du lancer de 10 pièces : $X \sim \text{Bin}(10, 1/2)$.
- Nombre d’étudiants malades dans un amphî de 100 étudiants avec la probabilité d’être malade égale à 0.6 : $X \sim \text{Bin}(100, 0.6)$.

Remarque 3.4.9 La loi binomiale de paramètres $(1, p)$ est simplement la loi de Bernoulli de paramètre p .

Proposition 3.4.10 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 grâce à la Proposition 4.6.2, 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. ■

3.4.3 Loi géométrique

Définition 3.4.11 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 $k \in \mathbb{N}_0$.

Exemple 3.4.12 Nombre de lancers nécessaires pour obtenir “pile” : $X \sim \text{Geom}(1/2)$.

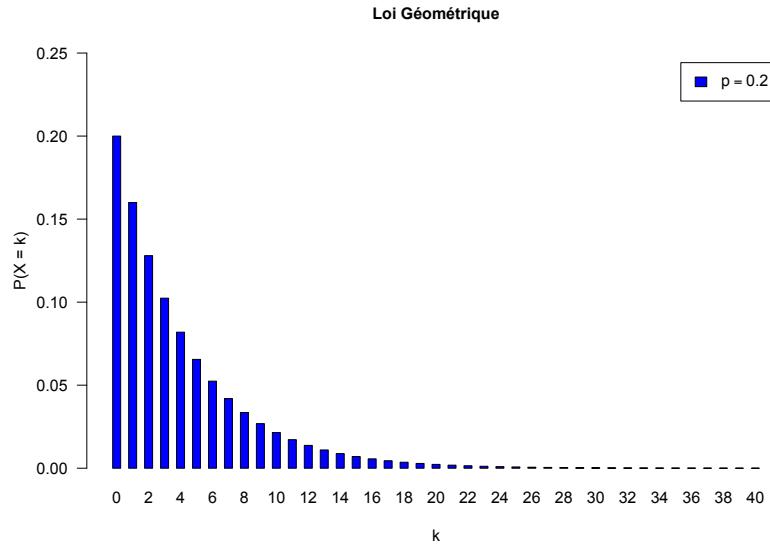


FIGURE 3.2 – Diagramme en bâtons d’une loi géométrique de paramètre 0.2.

Proposition 3.4.13 Si $X \sim \text{Geom}(p)$, alors $\mathbb{E}[X] = \frac{1}{p}$ et $\text{Var}[X] = \frac{1-p}{p^2}$.

Démonstration : La preuve est laissée à titre d’exercice lorsque vous aurez étudié les séries géométriques. ■

Remarque 3.4.14 Il existe deux paramétrisations classiques de la loi géométrique. La seconde prend ses valeurs dans \mathbb{N} et sa fonction de masse est $\mathbb{P}[X = k] = (1 - p)^k p$. Elle modélise le temps du dernier échec.

Proposition 3.4.15 Soit $X \sim \text{Geom}(p)$.

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

$$\mathbb{P}[X \geq k + j \mid X > k] = \mathbb{P}[X \geq j].$$

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

$$\mathbb{P}[X \geq k + j | X > k] = \frac{\mathbb{P}[X \geq k + j \text{ et } X > k]}{\mathbb{P}[X > k]} = \frac{\mathbb{P}[X \geq k + j]}{\mathbb{P}[X > k]}$$

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. ■

3.4.4 Loi de Poisson

Définition 3.4.16 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 $k \in \mathbb{N}$. 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 3.4.17

- Nombre d'accidents sur la route par jour.
- Nombre de clients par heure dans un supermarché.
- Nombre de visites sur Wikipedia par unité de temps.

Proposition 3.4.18 Si $X \sim \text{Pois}(\lambda)$, alors $\mathbb{E}[X] = \lambda$ et $\text{Var}[X] = \lambda$.

Démonstration : La preuve est laissée en exercice à faire lorsque vous aurez étudié la série définissant l'exponentielle. ■

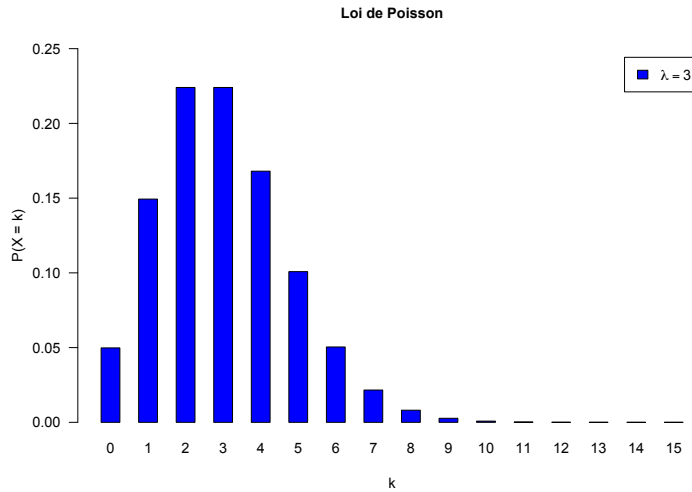


FIGURE 3.3 – Diagramme en bâtons d'une loi de Poisson de paramètre 3.

3.4.5 Logiciel R

Les lois discrètes que nous venons de voir sont très importantes et largement utilisées. Des fonctions existent dans R (et de nombreux logiciels) pour

- obtenir leur fonction de masse,
- obtenir leur fonction de répartition,
- simuler des nombres aléatoires suivant cette loi.

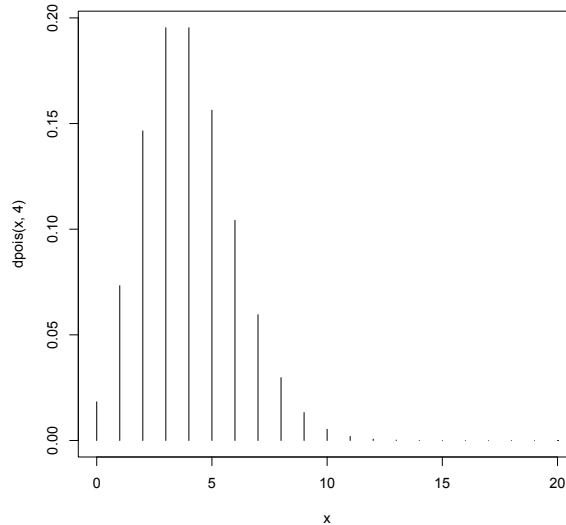
Exemple 3.4.19 Dans le cas de la loi de Poisson de paramètre λ , on a

- `dpois(x, λ)` : fonction de masse en x ,
- `ppois(x, λ)` : fonction de répartition en x ,
- `rpois(100, λ)` : échantillon de taille 100.

On a de même `dbinom(x, n, p)`, `dexp(x, λ)`, `dgeom(x, p)`... On peut utiliser ces commandes pour obtenir les diagrammes en bâtons : par exemple, si on regarde les 20 premières valeurs d'une loi de Poisson de paramètre 4, on effectue les commandes

```
> x <- 0:20
> plot(x, dpois(x, 4), type="h")
```

et on obtient le graphique suivant.



On peut aussi calculer par exemple la probabilité que cette variable aléatoire prenne ses valeurs entre 4 et 12

```
> sum(dpois(4:12, 4))
[1] 0.5662562
```

ou la probabilité qu'elle prenne une valeur strictement supérieure à 5

```
> 1 - ppois(5,4)
[1] 0.2148696
```

3.5 Variables aléatoires continues

Définition 3.5.1 On dit qu'une variable aléatoire X est *continue* s'il existe une fonction intégrable $f : \mathbb{R} \rightarrow [0, +\infty[$ telle que

$$\mathbb{P}[X \in [a, b]] = \mathbb{P}[a \leq X \leq b] = \int_a^b f(x)dx.$$

La fonction f s'appelle la *densité* de probabilité de X . Elle caractérise entièrement la loi de X . On note

$$X \sim f.$$

Exemple 3.5.2

- Le choix aléatoire d'un nombre entre 0 et 1 peut être modélisé par la densité $f(x) = 1$ si $x \in [0, 1]$ et $f(x) = 0$ sinon.
- La durée de vie d'un composant électronique peut être modélisée par la densité de probabilité $f(x) = \lambda e^{-\lambda x}$ si $x \geq 0$ et $f(x) = 0$ sinon, pour un paramètre $\lambda > 0$.

L'aire sous la courbe de f nous permet de calculer toutes les probabilités souhaitées. En particulier, l'aire totale sous la courbe doit être égale à 1.

Proposition 3.5.3 *Supposons que $X \sim f$.*

1. On a $\int_{\mathbb{R}} f(x)dx = 1$.
2. Pour tout $x \in \mathbb{R}$, $F_X(x) = \int_{-\infty}^x f(y)dy$.
3. Pour tout $x \in \mathbb{R}$, on a $\mathbb{P}[X = x] = 0$.

Démonstration : 1. Sans être rigoureux (il faudrait passer à la limite pour considérer l'intervalle $] - \infty, +\infty[$), on remarque que $1 = \mathbb{P}[X \in] - \infty, +\infty[] = \int_{\mathbb{R}} f(x)dx$.

2. De la même manière, $F_X(x) = \mathbb{P}[X \in] - \infty, x] = \int_{-\infty}^x f(y)dy$.

3. On a $\mathbb{P}[X = x] = \int_x^x f(y)dy = 0$. ■

En particulier, on peut travailler indifféremment avec des intervalles ouverts ou fermés lorsque l'on calcule les probabilités liées à X :

$$\mathbb{P}[a \leq X \leq b] = \mathbb{P}[a < X < b]$$

pour tous $a, b \in \mathbb{R}$.

Signalons également que tout fonction f définie sur \mathbb{R} , intégrable et à valeurs positives telle que

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

est la fonction de masse d'une variable aléatoire continue.

Exemple 3.5.4 On suppose que la durée de vie X d'une ampoule par unité de temps est modélisée par la densité de probabilité

$$f(x) = \begin{cases} e^{-x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0. \end{cases}$$

On a

$$\int_{\mathbb{R}} f(x)dx = \int_0^{+\infty} e^{-x}dx = [-e^{-x}]_0^{+\infty} = 1.$$

De plus, si on veut connaître la probabilité que l'ampoule dure plus de 10 unités de temps, on calcule

$$\mathbb{P}[X > 10] = \mathbb{P}[X \geq 10] = \int_{10}^{+\infty} e^{-x}dx = [-e^{-x}]_{10}^{+\infty} = e^{-10}.$$

3.6 Moments d'une variable aléatoire continue

Comme dans le cas discret, on peut considérer l'espérance et la variance d'une variable aléatoire continue.

Définition 3.6.1 Soit X une variable aléatoire continue de densité de probabilité f . L'espérance de X est la quantité

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

si la fonction $x \mapsto xf(x)$ est intégrable sur \mathbb{R} , c'est-à-dire si

$$\int_{\mathbb{R}} |x|f(x)dx < +\infty.$$

Exemple 3.6.2

— Soit X le résultat d'un tirage aléatoire d'un nombre entre 0 et 1. On a

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

— Soit X la durée de vie de l'ampoule de l'exemple précédent. On a

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

L'espérance dans le cas continu a les mêmes propriétés que dans le cas discret (ce n'est pas un hasard mais les justifications sortent du cadre de ce cours).

Proposition 3.6.3 Soient X et Y des variables aléatoires continues qui possèdent une espérance 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* : $\mathbb{E}[\alpha] = \alpha$.

Démonstration : La preuve de ce résultat est omise. ■

Si Y est une variable aléatoire qui est une *fonction de* X , c'est-à-dire $Y = g(X)$, on définit

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

pour autant que cette expression ait du sens. Par exemple,

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2 f(x)dx \quad \text{et} \quad \mathbb{E}[X^3] = \int_{\mathbb{R}} x^3 f(x)dx.$$

Définition 3.6.4 Soit $X \sim f$ une variable aléatoire continue. Si elle existe, la *variance* de X est la quantité

$$\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{\mathbb{R}} (x - \mathbb{E}[X])^2 f(x)dx.$$

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

Les propriétés de la variance étudiées dans le cas discret ne dépendaient que des propriétés de l'espérance donnée dans la Proposition 3.3.3. En utilisant l'équivalent continu de la Proposition 3.6.3, on retrouve donc exactement les mêmes résultats dans le cas continu et les mêmes interprétations.

Proposition 3.6.5

1. $\text{Var}[X] \geq 0$.
2. $\text{Var}[X] = 0$ si et seulement si X est 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. Soit X une variable aléatoire qui admet une espérance et une variance. Alors

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

Proposition 3.6.6 (Inégalité de Tchebychev) Soit X une variable aléatoire qui admet une espérance μ et une variance σ^2 . Alors pour tout $r > 0$,

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

3.7 Lois continues usuelles

On peut montrer que 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.

3.7.1 Loi continue uniforme

Définition 3.7.1 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

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

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

Exemple 3.7.2

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

Proposition 3.7.3 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

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

De plus,

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

et en utilisant le point 4 de la Proposition 3.6.5, on obtient

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

■

3.7.2 Loi exponentielle

Définition 3.7.4 Soit $\lambda > 0$. Une variable aléatoire continue X suit une *loi exponentielle* de paramètre λ si sa densité est donnée par

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

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

Exemple 3.7.5

- Durée de vie d'un composant électronique.
- Temps d'attente dans une file.

Proposition 3.7.6 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

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

En intégrant par parties, il vient

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

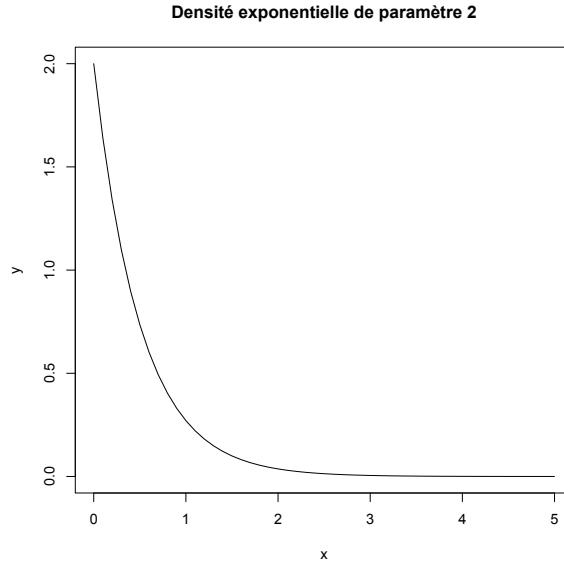
On procède de même en intégrant deux fois par parties pour calculer

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

et on utilise le point 4 de la Proposition 3.6.5 pour calculer la variance (exercice). ■

La commande suivante dans R permet de visualiser la fonction de densité d'une loi exponentielle de paramètre 2 :

```
> x<-seq(0,5,by=0.1)
> y<-dexp(x,2)
> plot(x,y,type="l",xlab="x",main="Densité exponentielle de paramètre 2")
```



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 3.7.7 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” :

$$\mathbb{P}[X \geq t + s \mid X > t] = \mathbb{P}[X \geq s],$$

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

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 3.4.15. ■

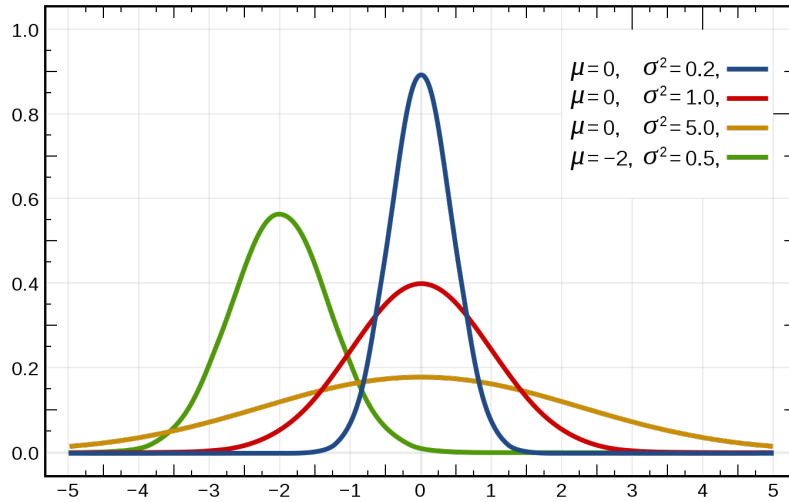


FIGURE 3.4 – Quelques représentations de densités de lois normales pour des paramètres différents. On a une courbe en “cloche”, symétrique par rapport à l’axe $x = \mu$. Plus σ est grand, plus la courbe est “aplatie”.

3.7.3 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 3.7.8 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

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

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

Exemple 3.7.9

- Taille d’un individu adulte (avec une précision infinie).
- Erreurs de mesure dans des expériences.

Proposition 3.7.10 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. ■

Définition 3.7.11 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, on a

$$\mathbb{P}[X \leq x] = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

qu’on ne peut pas exprimer sous forme de fonction calculable. On fait appel à des logiciels, ou à des tables de probabilités en se ramenant à la loi normale centrée et réduite via des dilatations et des translations. 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 3.7.12 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 $t = \frac{y - \mu}{\sigma}$. ■

Exemple 3.7.13 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.

```
> pnorm(1) - pnorm(-1/4)
[1] 0.4400511
```

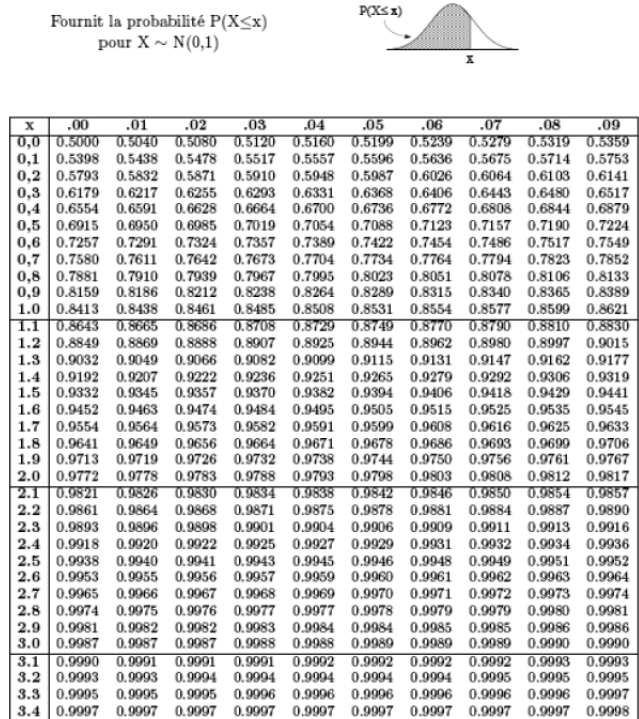


FIGURE 3.5 – Table de probabilités d’une loi normale centrée et réduite.

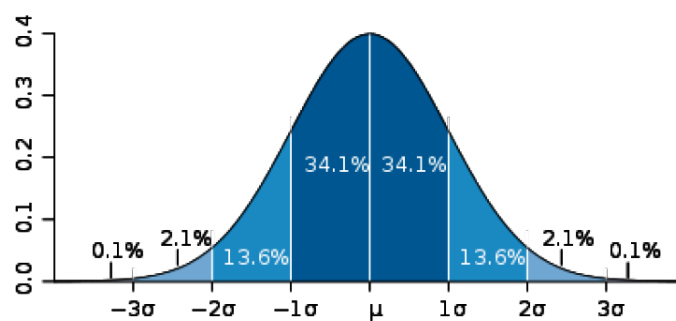
Remarque 3.7.14 Dans R, on peut utiliser directement la commande `pnorm(x, mean=300, sd=20)` qui donne la fonction de répartition de $X \sim \mathcal{N}(300, 20^2)$ en x sans se ramener à la loi centrée et réduite.

```
> pnorm(320, mean=300, sd=20) - pnorm(295, mean=300, sd=20)
[1] 0.4400511
```

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.

FIGURE 3.6 – Quelques valeurs remarquables de la loi normale $\mathcal{N}(\mu, \sigma^2)$.

Chapitre 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. Afin de capturer cette information, on étudie la distribution jointe des variables aléatoires. On parle alors de vecteur aléatoire (X_1, \dots, X_d) . 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.

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) est une description complète des probabilités

$$\mathbb{P}[(X, Y) \in I]$$

pour tout rectangle I de \mathbb{R}^2 .

Remarquons que si $I = [a, b] \times [c, d]$, alors

$$\mathbb{P}[(X, Y) \in I] = \mathbb{P}[a \leq X \leq b, c \leq Y \leq d]$$

et on a des relations semblables pour les intervalles (semi)-ouverts. 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 du couple aléatoire discret (X, Y) est la fonction $F_{X,Y}$ définie pour tout $(x, y) \in \mathbb{R}^2$ par

$$F_{X,Y}(x, y) = \mathbb{P}[X \leq x \text{ et } Y \leq y].$$

Evidemment, la loi détermine entièrement la fonction de répartition. Réciproquement, si on connaît la fonction de répartition, on peut retrouver la loi via la relation

$$\begin{aligned} \mathbb{P}[(X, Y) \in]a, b] \times]c, d] &= \mathbb{P}[a < X \leq b, c < Y \leq d] \\ &= F_{X,Y}(b, d) - F_{X,Y}(a, d) - F_{X,Y}(b, c) + F_{X,Y}(a, c). \end{aligned}$$

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

- $x_i, i \in \{1, \dots, N_1\}$, les valeurs possibles de X
- $y_j, j \in \{1, \dots, N_2\}$, les valeurs possibles de Y

avec $N_1, N_2 \in \mathbb{N} \cup \{+\infty\}$. L'ensemble des valeurs possibles du couple aléatoire (X, Y) est alors inclus dans

$$\{(x_i, y_j) : i \in \{1, \dots, N_1\}, j \in \{1, \dots, N_2\}\}.$$

Définition 4.2.1 La fonction de masse du couple aléatoire discret (X, Y) est la fonction $p_{X,Y}$ qui à tout élément (x_i, y_j) associe la probabilité que (X, Y) soit égal à (x_i, y_j) , c'est-à-dire

$$p_{X,Y}(x_i, y_j) = \mathbb{P}[X = x_i \text{ et } Y = y_j]$$

pour tout $i \in \{1, \dots, N_1\}$ et tout $j \in \{1, \dots, N_2\}$.

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, comme illustré dans la Table 4.1.

	x_1	x_2	\dots	x_{N_1}
y_1	p_{11}	p_{12}	\dots	p_{1N_1}
y_2	p_{21}	p_{22}	\dots	p_{2N_1}
\vdots	\vdots	\vdots		\vdots
y_{N_2}	p_{N_21}	p_{N_22}	\dots	$p_{N_2N_1}$

TABLE 4.1 – Distribution jointe d'un couple aléatoire discret.

La distribution jointe de (X, Y) est complètement caractérisée par la fonction de masse. En effet, si $I = [a, b] \times [c, d]$ est un intervalle de \mathbb{R}^2 , on a

$$\mathbb{P}[(X, Y) \in I] = \mathbb{P}[a \leq X \leq b, c \leq Y \leq d] = \sum_{i:a \leq x_i \leq b} \sum_{j:c \leq y_j \leq d} p_{ij}.$$

On a des relations semblables lorsque l'on considère des intervalles (semi-)ouverts. En particulier, en considérant $I = \mathbb{R}^2$, on trouve

$$\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} p_{ij} = 1.$$

De plus, la fonction de répartition du couple aléatoire discret (X, Y) prend la forme

$$F_{X,Y}(x, y) = \sum_{i:x_i \leq x} \sum_{j:y_j \leq y} p_{ij}.$$

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'un 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 (distinguable). 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 donné dans le tableau ci-dessous.

	1	2	3	4	5	6			
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$	$\mathbb{E}[Y]$ $= 3.5$	$\text{Var}[Y]$ $= \frac{35}{12}$ ≈ 2.9
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		
							$\mathbb{E}[X] = 3.5$		
							$\text{Var}[X] = \frac{35}{12} \approx 2.9$		

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 ci-dessous.

	1	2	3	4	5	6			
1	0	0	0	0	0	$\frac{1}{6}$	$\frac{1}{6}$	$\mathbb{E}[Y]$ $= 3.5$	$\text{Var}[Y]$ $= \frac{35}{12}$ ≈ 2.9
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		
								$\mathbb{E}[X] = 3.5$	
								$\text{Var}[X] = \frac{35}{12} \approx 2.9$	

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 ci-dessous.

	2	3	4	5	6	7	8	9	10	11	12			
0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	$\frac{6}{36}$	$\mathbb{E}[Y]$ $= \frac{35}{18}$ ≈ 1.94	$\text{Var}[Y]$ $= \frac{665}{324}$ ≈ 2.05
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		
												$\mathbb{E}[X] = 7$		
												$\text{Var}[X] = \frac{35}{6} \approx 5.83$		

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 masse marginales de X et Y sont données par

$$p_X(x_i) = \sum_{j=1}^{N_2} p_{ij} =: p_{i\bullet}$$

pour tout $i \in \{1, \dots, N_1\}$, et

$$p_Y(y_j) = \sum_{i=1}^{N_1} p_{ij} =: p_{\bullet j}$$

pour tout $j \in \{1, \dots, N_2\}$.

Démonstration : On démontre le résultat pour X , celui pour Y s'obtient de manière similaire. Fixons $i \in \{1, \dots, N_1\}$. Puisque Y prend ses valeurs dans $\{y_1, \dots, y_{N_2}\}$, on a

$$\begin{aligned} p_X(x_i) = \mathbb{P}[X = x_i] &= \mathbb{P}[X = x_i \text{ et il existe } j \in \{1, \dots, N_2\} \text{ tel que } Y = y_j] \\ &= \mathbb{P}\left[\bigcup_{j \in \{1, \dots, N_2\}} \{X = x_i, Y = y_j\}\right] \\ &= \sum_{j=1}^{N_2} \mathbb{P}[X = x_i, Y = y_j] \\ &= \sum_{j=1}^{N_2} p_{ij} \end{aligned}$$

puisque les événements $\{X = x_i, Y = y_j\}$ avec $j \in \{1, \dots, N_2\}$ 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 présenté dans la Table 4.2.

	x_1	x_2	\dots	x_{N_1}			
y_1	p_{11}	p_{21}	\dots	$p_{N_1 1}$	$p_{\bullet 1}$	$\mathbb{E}[X_2]$	$\text{Var}[X_2]$
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		
	$\mathbb{E}[X_1]$						
	$\text{Var}[X_1]$						

TABLE 4.2 – Distribution jointe d'un vecteur aléatoire bivarié discret, avec les distributions marginales. Les parenthèses indiquent l'éventualité d'un nombre infini dénombrable de valeurs différentes.

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 condition-*

nelle 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 $j = 1, \dots, N_2$.

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

$$\sum_{j=1}^{N_2} \mathbb{P}[Y = y_j|X = x_i] = \sum_{j=1}^{N_2} \frac{p_{ij}}{p_{i\bullet}} = \frac{1}{p_{i\bullet}} \sum_{j=1}^{N_2} p_{ij} = 1$$

par la Proposition 4.2.5.

Comme pour toute distribution discrète univarié, 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	$\mathbb{E}[Y X = x_1]$	$\mathbb{E}[Y X = x_2]$...	$\mathbb{E}[Y X = x_{N_1}]$
probabilités	$p_{1\bullet} = \mathbb{P}[X = x_1]$	$p_{2\bullet} = \mathbb{P}[X = x_2]$...	$p_{N_1\bullet} = \mathbb{P}[X = x_{N_1}]$

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	$\text{Var}[Y X = x_1]$	$\text{Var}[Y X = x_2]$...	$\text{Var}[Y X = x_{N_1}]$
probabilités	$p_{1\bullet} = \mathbb{P}[X = x_1]$	$p_{2\bullet} = \mathbb{P}[X = x_2]$...	$p_{N_1\bullet} = \mathbb{P}[X = x_{N_1}]$

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,

$$\mathbb{E}[Y] = \sum_{i=1}^{N_1} \mathbb{E}[Y | X = x_i] \mathbb{P}[X = x_i].$$

Démonstration : On a

$$\begin{aligned} \mathbb{E}[\mathbb{E}[Y|X]] &= \sum_{i=1}^{N_1} \mathbb{E}[Y|X = x_i] \mathbb{P}[X = x_i] = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} y_j \mathbb{P}[Y = y_j | X = x_i] \mathbb{P}[X = x_i] \\ &= \sum_{j=1}^{N_2} y_j \left(\sum_{i=1}^{N_1} \mathbb{P}[Y = y_j | X = x_i] \mathbb{P}[X = x_i] \right) \\ &= \sum_{j=1}^{N_2} y_j \left(\sum_{i=1}^{N_1} \mathbb{P}[X = x_i, Y = y_j] \right) \\ &= \sum_{j=1}^{N_2} y_j \left(\sum_{i=1}^{N_1} p_{ij} \right) \\ &= \sum_{j=1}^{N_2} y_j \mathbb{P}[Y = y_j] \\ &= \mathbb{E}[Y] \end{aligned}$$

où on a utilisé la Proposition 4.2.5. ■

Exemple 4.2.12 Reprenons l'exemple 4.2.4. On vérifie directement que $\mathbb{E}[\mathbb{E}[Y | X]] = \mathbb{E}[Y]$.

	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
$\mathbb{E}[Y X = x_i]$	0	1	1.33	2	2.4	3	2.4	2	1.33	1	0
$\text{Var}[Y X = x_i]$	0	0	0.89	1	2.24	2.67	2.24	1	0.89	0	0
$p_{i\bullet}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
	$\mathbb{E}[\mathbb{E}[Y X]] = 0 \times \frac{1}{36} + 1 \times \frac{2}{36} + \dots = \frac{35}{18} = \mathbb{E}[Y]$										

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 S'il existe une fonction de deux variables $f : \mathbb{R}^2 \rightarrow [0, +\infty[$ telle que pour que tout intervalle $I = [a, b] \times [c, d]$ de \mathbb{R}^2 , on a

$$\mathbb{P}[(X, Y) \in I] = \mathbb{P}[a \leq X \leq b, c \leq Y \leq d] = \int_a^b \int_c^d f(x, y) dy dx,$$

alors on dit que f est la *fonction de densité jointe* de (X, Y) et que le couple aléatoire (X, Y) est *continu*.

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

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

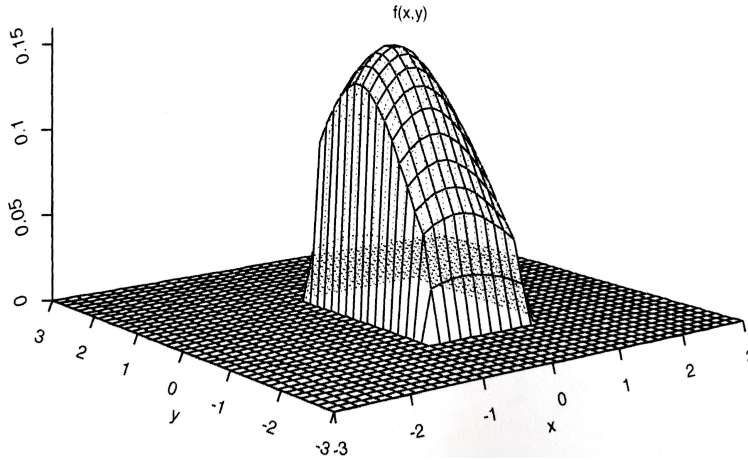


FIGURE 4.1 – Volume sous une fonction de densité jointe f sur le rectangle $[-0.5, 1] \times [-1.5, 1]$.

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

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du$$

et réciproquement

$$f(x, y) = \frac{\partial}{\partial x} \frac{\partial}{\partial y} F(x, y)$$

en tout point de dérivabilité de la fonction de répartition.

En pratique, l'évaluation de cette "intégrale double" se fait de façon séquentielle (en traitant d'abord une variable comme constante) :

$$\begin{aligned} \mathbb{P}[a < X \leq b, c < Y \leq d] &= \int_a^b \int_c^d f(x, y) dy dx \\ &= \int_a^b \underbrace{\left(\int_c^d f(x, y) dy \right)}_{\text{une fonction de } x} dx \end{aligned}$$

ou, de manière équivalente (le résultat est le même!),

$$\begin{aligned} \mathbb{P}[a < X \leq b, c < Y \leq d] &= \int_c^d \int_a^b f(x, y) dx dy \\ &= \int_c^d \underbrace{\left(\int_a^b f(x, y) dx \right)}_{\text{une fonction de } y} dy \end{aligned}$$

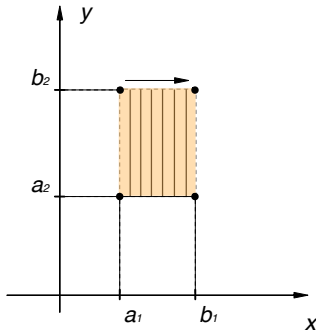


FIGURE 4.2 – $\int_{a_1}^{b_1} \left(\int_{a_2}^{b_2} \dots dy \right) dx$

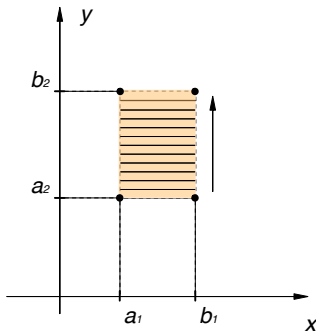


FIGURE 4.3 – $\int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} \dots dx \right) dy$

Si la fonction qu'on intègre est de la forme $f(x, y) = f_1(x)f_2(y)$ pour tous x, y , alors

$$\begin{aligned} \int_c^d \int_a^b f(x, y) dx dy &= \int_c^d \int_a^b f_1(x)f_2(y) dx dy \\ &= \left(\int_a^b f_1(x) dx \right) \left(\int_c^d f_2(y) dy \right). \end{aligned}$$

Exemple 4.3.2 Une chaîne de restauration rapide vend des hamburgers selon deux modalités distinctes : un comptoir traditionnel et un drive-in. Soit X la proportion du temps où le comptoir traditionnel est occupé le jeudi et soit Y la proportion du temps où le drive-in est occupé le jeudi. Supposons que (X, Y) admette la fonction de densité

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & \text{si } (x, y) \in [0, 1] \times [0, 1] \\ 0 & \text{sinon.} \end{cases}$$

Notez bien qu'il s'agit là d'un choix arbitraire de modélisation, contrairement à l'exemple du lancer de dés où la loi sous-jacente était dictée par l'expérience! Le manager est satisfait si le comptoir traditionnel est occupé au moins la moitié du temps et le drive-in au moins un quart du temps, ce qui arrive avec la probabilité

$$\mathbb{P}[1/2 \leq X \leq 1, 1/4 \leq Y \leq 1].$$

On calcule cette probabilité de la manière suivante

$$\begin{aligned} \int_{1/2}^1 \int_{1/4}^1 f(x, y) dy dx &= \int_{1/2}^1 \left(\int_{1/4}^1 \frac{6}{5}(x + y^2) dy \right) dx \\ &= \int_{1/2}^1 \left[\frac{6}{5} \left(xy + \frac{y^3}{3} \right) \right]_{y=1/4}^{y=1} dx \\ &= \int_{1/2}^1 \left(\frac{9x}{10} + \frac{63}{160} \right) dx = \left[\frac{9x^2}{20} + \frac{63x}{160} \right]_{x=1/2}^{x=1} = \frac{171}{320}, \end{aligned}$$

ou encore

$$\begin{aligned} \int_{1/4}^1 \int_{1/2}^1 f(x, y) dx dy &= \int_{1/4}^1 \left(\int_{1/2}^1 \frac{6}{5}(x + y^2) dx \right) dy \\ &= \int_{1/4}^1 \left[\frac{6}{5} \left(\frac{x^2}{2} + y^2 x \right) \right]_{x=1/2}^{x=1} dy \\ &= \int_{1/4}^1 \left(\frac{9}{20} + \frac{3y^2}{5} \right) dy = \left[\frac{9y}{20} + \frac{y^3}{5} \right]_{y=1/4}^{y=1} = \frac{171}{320}. \end{aligned}$$

En procédant de la même façon, on obtient

$$F(x, y) = \begin{cases} 0 & \text{si } x < 0 \text{ ou } y < 0 \\ \frac{1}{5}xy(3x + 2y^2) & \text{si } (x, y) \in [0, 1] \times [0, 1] \\ \frac{1}{5}x(3x + 2) & \text{si } x \in [0, 1] \text{ et } y > 1 \\ \frac{1}{5}y(3 + 2y^2) & \text{si } x > 1 \text{ et } y \in [0, 1] \\ 1 & \text{sinon.} \end{cases}$$

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 . Alors, les densités marginales de X et Y sont données respectivement par

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$$

et

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$$

Remarque 4.3.4 Il faut noter l'analogie avec les formules discrètes

$$\begin{cases} \mathbb{P}[X = x_i] = \sum_{j=1}^{N_2} \mathbb{P}[(X, Y) = (x_i, y_j)] \\ \mathbb{P}[Y = y_j] = \sum_{i=1}^{N_1} \mathbb{P}[(X, Y) = (x_i, y_j)]. \end{cases}$$

Exemple 4.3.5 Reprenons l'exemple 4.3.2 du fast food. 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.

Remarque 4.3.6 On a montré comment calculer des probabilités via la formule

$$\mathbb{P}[(X, Y) \in I] = \int \int_I f(x, y) dy dx$$

pour des zones *rectangulaires*. Mais les applications conduisent souvent à des zones non rectangulaires. Par exemple, le manager du fast food pourrait s'intéresser à la probabilité que le comptoir traditionnel soit moins utilisé que le drive-in. Ceci correspond à calculer la probabilité

$$\mathbb{P}[X \leq Y] = \mathbb{P}[(X, Y) \in B_{\text{tr}}],$$

où $B_{\text{tr}} = \{(x, y) \in [0, 1] \times [0, 1] : x \leq y\}$. On calcule cette intégrale de la manière suivante. On

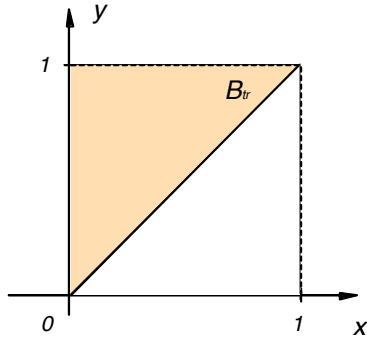


FIGURE 4.4 – Ensemble B_{tr} sur lequel on doit calculer l'intégrale de f .

remarque que si on fixe x entre 0 et 1, alors y doit varier entre x et 1. Ainsi,

$$\begin{aligned} \mathbb{P}[X \leq Y] &= \mathbb{P}[(X, Y) \in B_{\text{tr}}] \\ &= \int \int_{B_{\text{tr}}} f(x, y) dy dx \\ &= \int_0^1 \left(\int_x^1 f(x, y) dy \right) dx = \dots = \frac{1}{2}. \end{aligned}$$

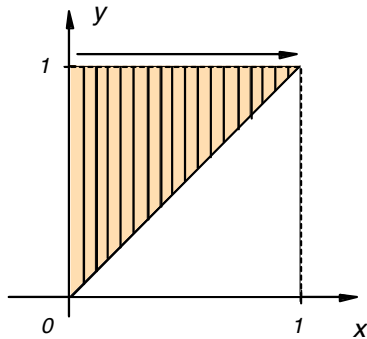


FIGURE 4.5 – Si x est fixé entre 0 et 1, alors y varie entre x et 1.

De manière semblable, on peut d'abord considérer un y entre 0 et 1. Pour cet y fixé, x doit varier entre 0 et y . On calcule donc

$$\begin{aligned} \mathbb{P}[X \leq Y] &= \mathbb{P}[(X, Y) \in B_{\text{tr}}] \\ &= \int \int_{B_{\text{tr}}} f(x, y) dy dx \\ &= \int_0^1 \left(\int_0^y f(x, y) dx \right) dy = \dots = \frac{1}{2}. \end{aligned}$$

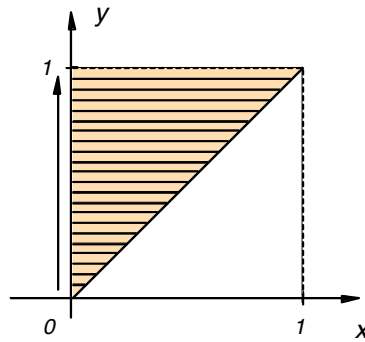


FIGURE 4.6 – Si y est fixé entre 0 et 1, alors x varie entre 0 et y .

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 4.3.7 Soit (X, Y) un couple aléatoire continu. 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 d'une densité d'une variable aléatoire. On trouve directement que

- $\int_{-\infty}^{+\infty} f_{X|Y=y}(x) dx = 1$,
- $f(x, y) = f_{X|Y=y}(x) f_Y(y)$,
- $\mathbb{P}[a \leq Y \leq b | X = x] = \int_a^b f_{Y|X=x}(y) dy$.

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 4.3.8 Dans le cas du fast food, 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 4.3.9 L'espérance conditionnelle $\mathbb{E}[Y|X]$ de Y sachant X est la variable aléatoire prenant la valeur $\mathbb{E}[Y|X = x]$ avec densité $f_X(x)$.

La variance conditionnelle $\text{Var}[Y|X]$ de Y sachant X est la variable aléatoire prenant la valeur $\text{Var}[Y|X = x]$ avec densité $f_X(x)$.

Alors, comme dans le cas discret, on a le théorème de l'espérance totale suivant.

Théorème 4.3.10 (Théorème de l'espérance totale) Si X et Y sont des variables aléatoires continues qui admettent une espérance, on a

$$\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y],$$

c'est-à-dire

$$\mathbb{E}[Y] = \int_{-\infty}^{+\infty} \mathbb{E}[Y|X = x] f_X(x) dx.$$

Enfin, on retrouve que

$$\mathbb{E}[\text{Var}[Y|X]] = \text{Var}[Y] - \text{Var}[\mathbb{E}[Y|X]].$$

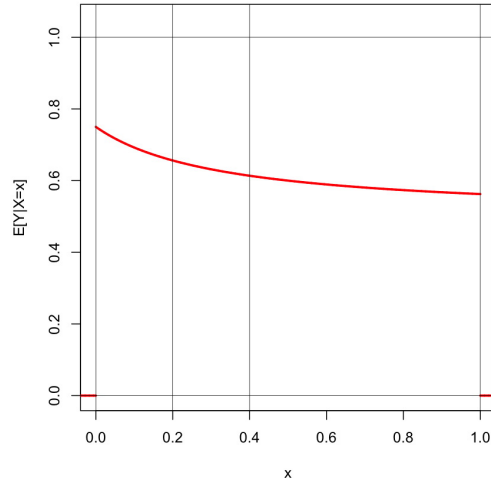


FIGURE 4.7 – Graphe de $x \mapsto \mathbb{E}[Y|X = x]$: au plus le comptoir traditionnel est occupé, au moins le drive-in l’est (en moyenne).

Exemple 4.3.11 Dans le cas du fast food, 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 lancé 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 tout $i \in \{1, \dots, N_1\}$ et tout $j \in \{1, \dots, N_2\}$;

- on a

$$\mathbb{P}[Y = y_j | X = x_i] = p_{\bullet j}$$

pour tout $i \in \{1, \dots, N_1\}$ et tout $j \in \{1, \dots, N_2\}$.

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 $f_X(x) \neq 0$.

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 une fonction $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ et on va s'intéresser à des espérances du type $\mathbb{E}[g(X, Y)]$. Par exemple, $\mathbb{E}[X + Y]$ ou $\mathbb{E}[XY]$.

Définition 4.5.7 L'espérance de $g(X, Y)$ est définie par

- Cas discret :

$$\mathbb{E}[g(X, Y)] = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} g(x_i, y_j) p_{ij}.$$

- Cas continu :

$$\mathbb{E}[g(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dy dx.$$

Cette définition 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 4.5.8 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

$$\begin{aligned} \mathbb{E}[XY] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f(x, y) dy dx \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_X(x) f_Y(y) dy dx \\ &= \left(\int_{-\infty}^{+\infty} x f_X(x) dx \right) \left(\int_{-\infty}^{+\infty} y f_Y(y) dy \right) = \mathbb{E}[X] \mathbb{E}[Y]. \end{aligned}$$

■

Remarque 4.5.9 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 4.5.10 Si $X \perp\!\!\!\perp Y$, alors on a $h(X) \perp\!\!\!\perp g(Y)$ quelles que soient les fonctions g, h . En particulier,

$$\mathbb{E}[h(X)g(Y)] = \mathbb{E}[h(X)]\mathbb{E}[g(Y)].$$

Ceci permet de prouver certaines propriétés d'additivité. On utilise pour cela la fonction génératrice des moments

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

qui caractérise la loi de la variable aléatoire X .

Théorème 4.5.11 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 $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

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

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

L'interprétation de la covariance est illustrée dans les Figures 4.8 et 4.6.

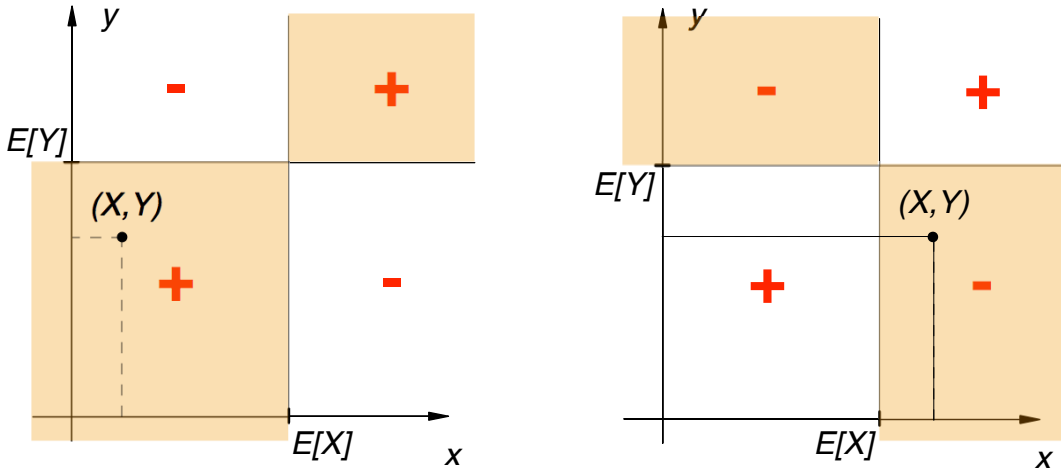


FIGURE 4.8 – A gauche, (X, Y) se réalise dans la zone où $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) > 0$. A droite, (X, Y) se réalise dans la zone où $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) < 0$. Si (X, Y) se réalise plus souvent dans les zones "+" (ou de façon plus extrême), $\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$ sera positif, et inversement.

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

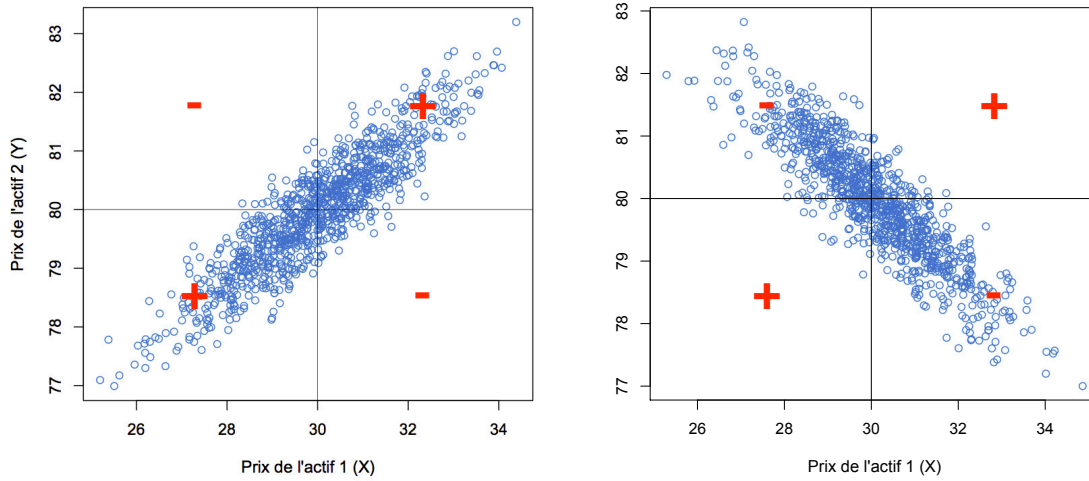


FIGURE 4.9 – A gauche, $\text{Cov}[X, Y] > 0$; on parlera de dépendance positive (notons que dans ce cas, $\text{Var}[X + Y] > \text{Var}[X] + \text{Var}[Y]$). A droite, $\text{Cov}[X, Y] < 0$; on parlera de dépendance négative (notons que dans ce cas, $\text{Var}[X + Y] < \text{Var}[X] + \text{Var}[Y]$)

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.

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 défi-

inition de la covariance, il vient

$$\text{Cov}[X, Y] = \begin{cases} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} (x_i - \mathbb{E}[X])(y_j - \mathbb{E}[Y])p_{ij} & \text{(cas discret)} \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) f(x, y) dy dx & \text{(cas continu)}. \end{cases}$$

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

$$\text{Cov}[X, Y] = \begin{cases} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} x_i y_j p_{ij} - \mathbb{E}[X]\mathbb{E}[Y] & \text{(cas discret)} \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f(x, y) dy dx - \mathbb{E}[X]\mathbb{E}[Y] & \text{(cas continu)}. \end{cases}$$

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. $\text{Cov}[X, c] = 0$,

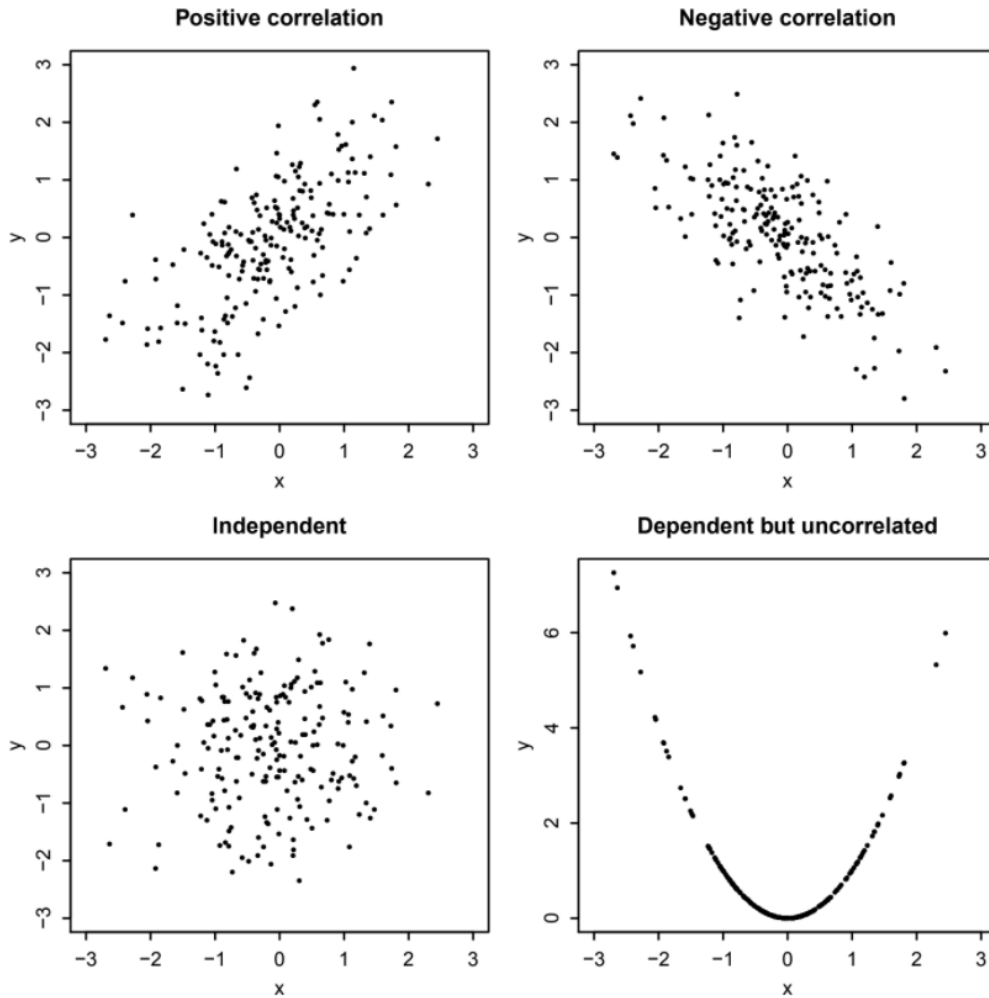


FIGURE 4.10 – Illustration de la corrélation et de la dépendance.

- 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) *On a*

$$|\text{Cov}[X, Y]| \leq \sqrt{\text{Var}[X]\text{Var}[Y]}$$

avec égalité si et seulement si $Y = \alpha X + \beta$ pour certains $\alpha, \beta \in \mathbb{R}$.

Ce résultat mène naturellement à l'introduction de la notion suivante.

Définition 4.6.8 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 $Y = \alpha X + \beta$ (dépendance *linéaire*).

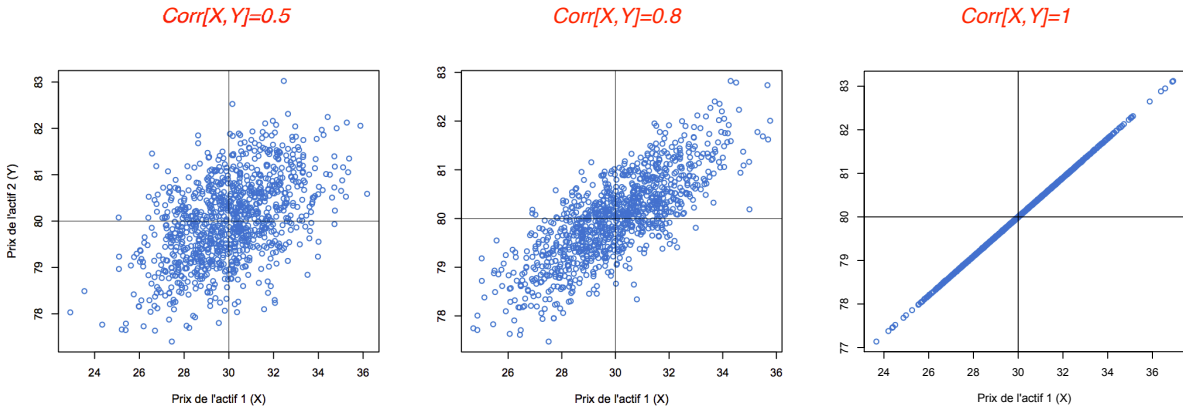


FIGURE 4.11 – Plus $|\text{Corr}[X, Y]|$ est proche de 1, plus la relation linéaire entre X et Y est forte.

Chapitre 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 estimer l'autonomie des batteries du nouvel iPhone qui, selon la publicité, est de 28h. Comment vérifier que la durée annoncée est effectivement la bonne ? Il s'agit d'un problème de décision statistique. Si on suppose que X est la variable aléatoire qui donne l'autonomie des batteries des nouveaux iPhone, on souhaite vérifier si

$$\mathbb{E}[X] < 28$$

ou non. On souhaite donc avoir une estimation de l'espérance de X même si on n'en connaît pas la loi de l'autonomie des batteries !

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

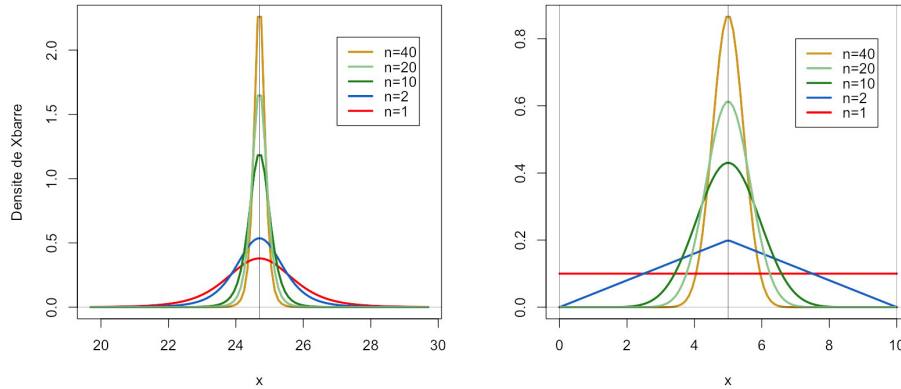


FIGURE 5.1 – Densités de \bar{X}_n pour $n = 1, 2, 10, 20, 40$ dans le cas où les observations sont i.i.d. pour deux lois de X différentes. Dans les deux cas, les densités gardent la même espérance mais la variance diminue lorsque n augmente.

moyenne empirique vérifiée

$$\mathbb{E}[\bar{X}_n] = \mu \quad \text{et} \quad \text{Var}[\bar{X}_n] = \frac{\sigma^2}{n}.$$

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 $\mathbb{E}[\bar{X}] = \mu$ et $\text{Var}[\bar{X}_n] = \frac{\sigma^2}{n}$, l'inégalité de Tchebychev donnée dans la Proposition 3.3.10 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 X_n tend *presque sûrement* vers μ).

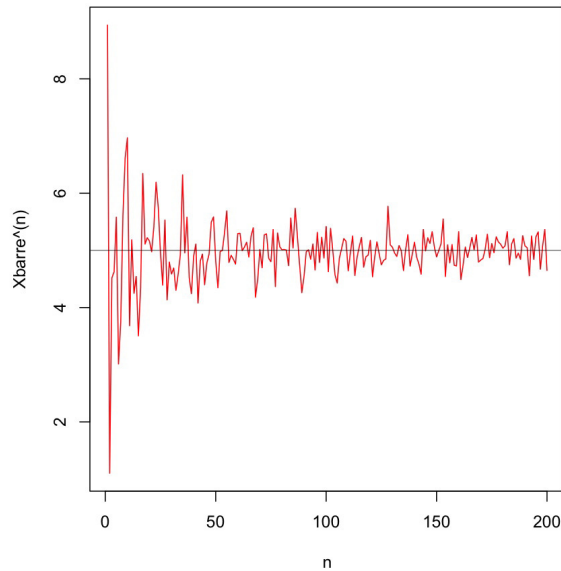


FIGURE 5.2 – Pour chaque n , on a engendré X_1, \dots, X_n i.i.d. de loi $\mathcal{U}([0, 10])$ et on a calculé \bar{X}_n . Ce graphe de \bar{X}_n en fonction de n montre que \bar{X}_n tend vers $\mu = \mathbb{E}[X_i] = 5$.

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 consister 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 5.1, 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 le montre le théorème suivant, dont la preuve sera admise.

Théorème 5.2.2 (Théorème Central Limite) Soient X_1, X_2, \dots des variables aléatoires i.i.d d'espérance μ et de variance σ^2 . Pour tout $x \in \mathbb{R}$,

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left[\frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq x \right] = \Phi(x),$$

où Φ est la fonction de répartition de la loi $\mathcal{N}(0, 1)$.

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

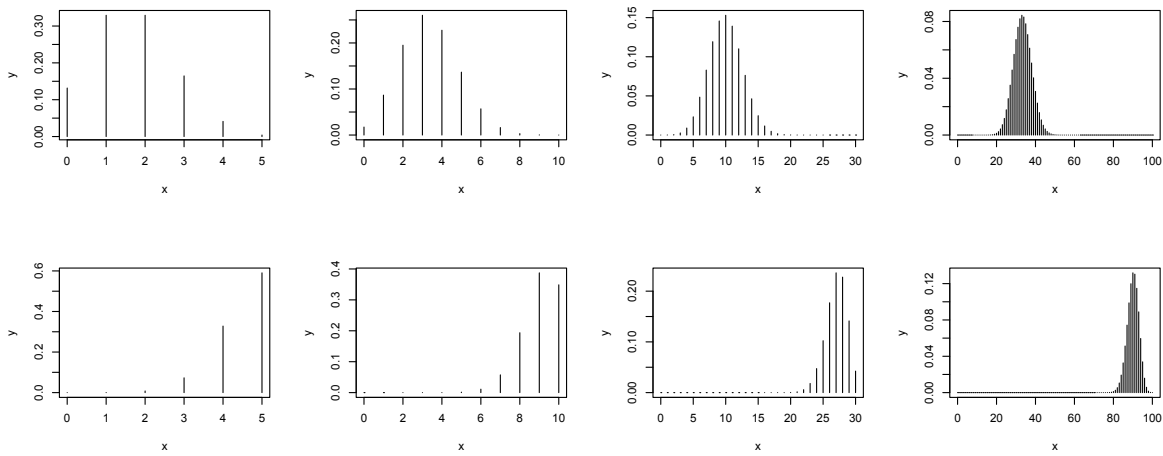


FIGURE 5.3 – Convergence vers la loi normale : diagrammes en bâtons des lois binomiales de paramètres (n, p) pour $n = 5, 10, 30, 100$ (de gauche à droite) et $p = \frac{1}{3}, \frac{9}{10}$ (haut et bas).

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. Considérons une expérience aléatoire et $(\Omega, \mathcal{F}, \mathbb{P})$ l’espace probabilisé associé. Soit A un événement fixé de probabilité $p = \mathbb{P}[A]$. Notons X la variable aléatoire qui vaut 1 si A se produit et 0 sinon : $X \sim \text{Bern}(p)$. Soient X_1, \dots, X_n des variables aléatoires i.i.d de loi $\text{Bern}(p)$. On sait que $\mu = \mathbb{E}[X] = p$ et $\sigma^2 = \text{Var}[X] = p(1 - p)$. Donc le TCL affirme que, pour tout $x \in \mathbb{R}$,

$$\mathbb{P} \left[\frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \leq x \right] \rightarrow \Phi(x).$$

Si X_1, \dots, X_n sont i.i.d de loi $\text{Bern}(p)$, alors

$$Y = X_1 + \dots + X_n \sim \text{Bin}(n, p).$$

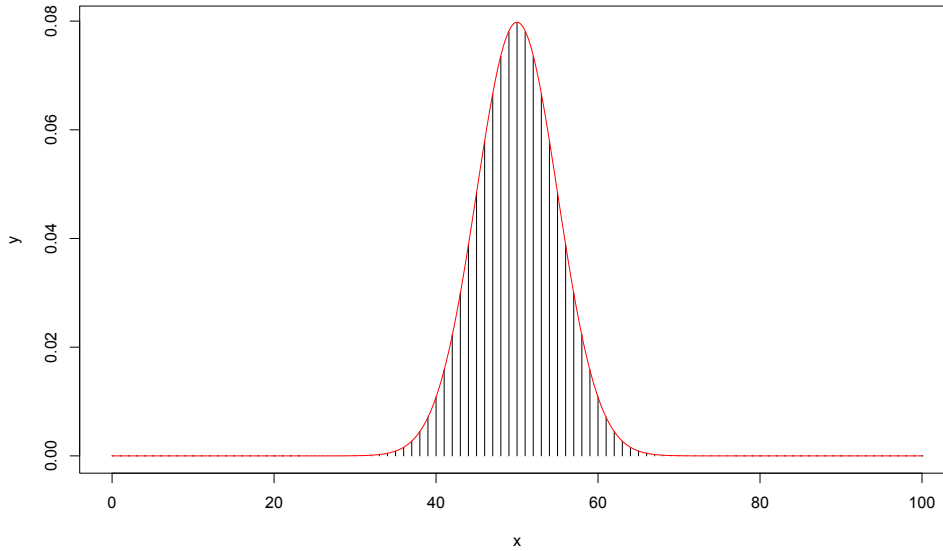
On a donc

$$\frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{Y - np}{\sqrt{np(1-p)}} \approx \mathcal{N}(0, 1) \iff Y \approx \mathcal{N}(np, np(1-p)).$$

Le résultat précédent permet donc d’approximer des probabilités de la loi binomiale à partir de la loi normale lorsque n est grand !

Proposition 5.2.3 (Théorème de De Moivre - Laplace) *Si n est grand, alors*

$$\text{Bin}(n, p) \approx \mathcal{N}(np, np(1-p)).$$

FIGURE 5.4 – Diagramme en bâtons de la loi $\text{Bin}(100, 1/2)$ et densité de la loi normale $\mathcal{N}(50, 25)$.

Exemple 5.2.4 Si on considère le lancer d’une pièce de monnaie et l’événement “A= Face”, alors $p = \frac{1}{2}$. On souhaite calculer la probabilité d’obtenir entre 40 et 60 fois “Face” en 100 lancers d’une pièce de monnaie. Soit donc $Y \sim \text{Bin}(100, 1/2)$ la variable aléatoire qui compte le nombre de fois qu’on a obtenu “ Face”. On a $Y \approx \mathcal{N}(50, 25)$ et on calcule

$$\begin{aligned}
 \mathbb{P}[40 < Y \leq 60] &= \mathbb{P}\left[\frac{40 - 50}{5} < \frac{Y - 50}{5} \leq \frac{60 - 50}{5}\right] \\
 &= \mathbb{P}\left[-2 < \frac{Y - 50}{5} \leq 2\right] \\
 &\approx \mathbb{P}\left[-2 < Z \leq 2\right] \\
 &\approx 0.9545
 \end{aligned}$$

où $Z \sim \mathcal{N}(0, 1)$. Il y a donc à peu près 95.45% de chance qu’on ait entre 40 et 60 fois “Face” en 100 lancers d’une pièce de monnaie. 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.

Bibliographie

- [1] M. Baron (2014), *Probability and Statistics for computer scientists, 2d Edition*. CRC Press.
- [2] J. K. Blitzstein, J. Hwang (2019), *Introduction to probability, 2d Edition* 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] S. Ross (2010), *A first course in Probability*. 8th Edition, Pearson Prentice Hall.
- [6] D. Paindaveine (2018), *Probabilités*. Université Libre de Bruxelles - Solvay.