Introduction aux probabilités

Nicolas MEYER

Université de Montpellier
Licence 2

2022 - 2023

Table des matières

Chapitre 1

Espace probabilisé

Le but de ce chapitre est de modéliser la notion de hasard. On définit un cadre mathématique pour les notions d'expérience aléatoire, d'événements, de probabilité, d'indépendance, etc.

1 Tribu et événement

1.1 Définitions et premiers exemples

Définition 1.1 (Expérience aléatoire, univers, événement élémentaire)**.**

1. *On appelle* expérience aléatoire *toute expérience dont on ne peut pas prévoir le résultat à l'avance (résultat soumis au hasard).*
2. *On appelle* univers*, et on note* Ω*, l'ensemble des résultats possibles de cette expérience.*
3. *On appelle* événement élémentaire *tout élément* ω *de* Ω*. L'univers* Ω *correspond donc à la réunion de tous les événements élémentaires.*

Exemples 1.2.

1. On s'intéresse au chiffre qui apparaît lors du lancer d'un dé à six faces. L'univers est alors $\Omega = \{1, 2, 3, 4, 5, 6\}$. Cet univers est fini.
2. On considère une infinité de lancers de pièce dont l'une des faces est notée 0 et l'autre 1. Un événement élémentaire est donc une suite (u_n) à valeurs dans $\{0, 1\}$. Cet univers est infini.
3. On tire une carte dans un jeu de 52 cartes. L'univers correspond alors aux 52 cartes du jeu et un événement élémentaire correspond alors à une carte.
4. On arrive à une station de tramway et on s'intéresse au temps d'attente jusqu'au prochain tram. L'univers est ici $\Omega = [0, \infty[$.

1.2 Tribu

Dans la suite, on souhaitera calculer la probabilité d'apparition de certaines parties de l'univers Ω. On note $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω :

$$A \in \mathcal{P}(\Omega) \iff A \subset \Omega\,.$$

Dans le cas où Ω est fini, toutes les parties pourront être considérées. Dans le cas où Ω est infini, $\mathcal{P}(\Omega)$ est trop gros : on doit se restreindre à un sous-ensemble $\mathcal{F}$ de $\mathcal{P}(\Omega)$ qui doit respecter un certain nombre de propriétés.

Définition 1.3 (Tribu)**.** *Soit Ω un univers et $\mathcal{F}$ un sous-ensemble de $\mathcal{P}(\Omega)$. On dit que $\mathcal{F}$ est une* tribu *(ou σ-*algèbre*) sur Ω si les trois conditions suivantes sont vérifiées.*

1. *$\Omega \in \mathcal{F}$.*
2. *Si A appartient à $\mathcal{F}$, alors son complémentaire $\bar{A}$ appartient à $\mathcal{F}$.*
3. *Si $(A_n)_{n\in\mathbb{N}}$ est une suite de $\mathcal{F}$, alors la réunion $\bigcup_{n=0}^{+\infty} A_n$ appartient à $\mathcal{F}$.*

Les éléments de la tribu $\mathcal{F}$ sont appelés les événements.

Intuitivement, une tribu est l'ensemble des événements de l'univers Ω que l'on peut observer.

On rappelle au passage que l'ensemble $\cup_{n=0}^{+\infty} A_n$ correspond aux ω qui appartiennent à au moins un A_n :

$$\omega \in \bigcup_{n=0}^{+\infty} A_n \iff \exists n \in \mathbb{N} : \omega \in A_n\,.$$

Exemples 1.4. Voici trois exemples classiques de tribus.

1. La tribu grossière ou triviale : $\mathcal{F} = \{\emptyset, \Omega\}$.
2. La tribu pleine : $\mathcal{F} = \mathcal{P}(\Omega)$.
3. La tribu engendrée par une partie A de Ω : $\mathcal{F} = \{A, \bar{A}, \emptyset, \Omega\}$, où $A \subset \Omega$ est un sous-ensemble fixé.

On déduit immédiatement de la définition qu'une tribu contient l'ensemble vide, est stable par union finie, et par intersection dénombrable ou finie. Bref, une tribu est stable par toutes les opérations ensemblistes souhaitées, du moment que ces opérations concernent un nombre *dénombrable* d'événements.

Certains événements d'une tribu sont remarquables : l'ensemble vide $\emptyset$ qu'on appelle *événement impossible* et l'univers Ω qu'on appelle *événement certain*. Enfin, si A est un événement, alors $\bar{A}$ est appelé *événement contraire de A*.

Lorsque Ω est un ensemble fini, on prendra toujours la tribu pleine $\mathcal{F} = \mathcal{P}(\Omega)$ qui est alors également finie. C'est le cas dans l'exemple du lancer de dé où $\Omega = \{1, 2, 3, 4, 5, 6\}$. Si le cardinal Ω est infini, cela se corse un peu. En pratique dans ce cours, le choix de la tribu sera un problème qu'on mettra discrètement sous le tapis.

Donnons à présent quelques exemples d'événements pour des expériences aléatoires classiques.

Exemples 1.5.

1. On considère le lancer d'un dé dont l'univers est $\Omega = \{1, 2, 3, 4, 5, 6\}$. Un exemple d'événement est donné par les résultats pairs : $A = \{2, 4, 6\}$.
2. On considère une infinité de lancers de dé dont l'univers est donné par $\Omega = \{1, 2, 3, 4, 5, 6\}^{\mathbb{N}^*}$. Cela signifie qu'un événement élémentaire est à une suite à valeurs dans $\{1, 2, 3, 4, 5, 6\}$. Pour tout $i \in \mathbb{N}^*$, on considère l'événement F_i : "on obtient un 6 au i-ème lancer". Si i est fixé, l'événement F_i est constitué de toutes les suites à valeurs dans $\{1, 2, 3, 4, 5, 6\}$ ayant un 6 en position i. On peut alors construire d'autres événements, comme "obtenir au moins un 6 parmi tous les lancers" qui s'écrit $\cup_{i\in\mathbb{N}^*} F_i$, ou encore "obtenir uniquement des 6 entre le 3-ème et le 7-ième lancer" qui s'écrit $\cap_{i=3}^{7} F_i$.

Exercice 1.6. À partir des événements F_i de l'exemple précédent, exprimer en français les événements

$$A = \bigcup_{j=1}^{+\infty} \bigcap_{i=j}^{+\infty} F_i \quad \text{et} \quad B = \bigcap_{j=1}^{+\infty} \bigcup_{i=j}^{+\infty} F_i \, .$$

2 Probabilité

2.1 Définition et propriétes

On introduit désormais la notion de probabilité. Pour cela, on a besoin de la notion d'événements incompatibles.

Définition 1.7 (Événements incompatibles)**.** *Deux événements A et B sont dits* incompatibles *si $A \cap B = \emptyset$. Plus généralement, on dit que $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements* 2 à 2 incompatibles *si $A_i \cap A_j = \emptyset$ pour tout $i \neq j$.*

Définition 1.8 (Probabilité)**.** *Soit Ω un univers et $\mathcal{F}$ une tribu sur Ω. On appelle* probabilité sur la tribu $\mathcal{F}$ *toute application $\mathbb{P} : \mathcal{F} \to [0, 1]$ qui vérifie :*

1. $\mathbb{P}(\Omega) = 1$.
2. *($\sigma-$additivité). Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements 2 à 2 incompatibles de $\mathcal{F}$, alors*

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

On dit alors que $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé.

La propriété de $\sigma-$additivité assure que calculer la probabilité d'une union d'événements incompatibles revient à calculer chaque probabilité séparément et les sommer (voir Figure 1.1).

Donnons un exemple de probabilité que l'on utilisera constamment.

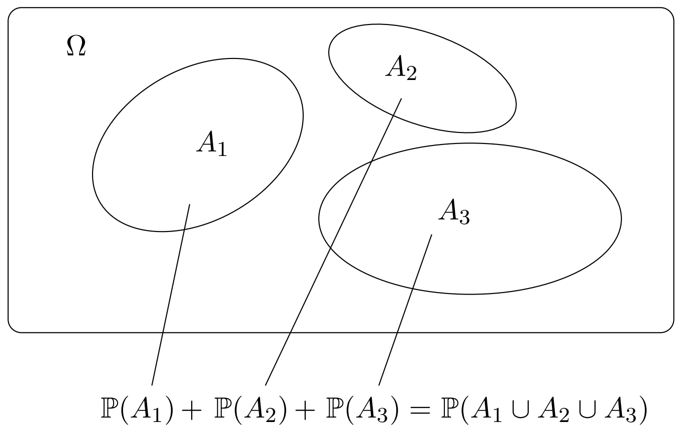


FIGURE 1.1 – Illustration de la σ-additivité. La probabilité d'une union disjointe est la somme des probabilités.

Exemple 1.9. On considère le lancer d'un dé. L'univers est $\Omega = \{1,2,3,4,5,6\}$ que l'on munit de la tribu pleine $\mathcal{F} = \mathcal{P}(\Omega)$. L'application $\mathbb{P} : \mathcal{F} \to [0,1], A \mapsto \#A/6$ est une probabilité, où $\#A$ désigne le cardinal de l'ensemble A. On obtient alors par exemple

$$\mathbb{P}(\{\text{obtenir un chiffre pair}\}) = \frac{\#\{\text{obtenir un chiffre pair}\}}{6} = \frac{3}{6} = \frac{1}{2}\,.$$

Plus généralement, si l'univers Ω est fini, alors l'application $\mathbb{P}$ définie par

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega}\,, \quad A \in \mathcal{F}\,,$$

est une probabilité sur $\mathcal{P}(\Omega)$ appelée *équiprobabilité* ou *probabilité uniforme*. On retrouve le ratio bien connu $\frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}$.

Proposition 1.10. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.*

1. *(Ensemble vide). $\mathbb{P}(\emptyset) = 0$. En particulier, la σ-additivité est vraie pour toute suite finie.*
2. *(Monotonie). Si $A, B \in \mathcal{F}$ sont tels que $A \subset B$, alors*

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)\,,$$

où $B \setminus A = \{\omega \in B : \omega \notin A\}$ désigne la soustraction ensembliste. On en déduit en particulier que $\mathbb{P}(A) \leq \mathbb{P}(B)$ et, avec $B = \Omega$, que $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$.

3. *(Additivité forte). Si $A, B \in \mathcal{F}$, alors*

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

4. *(Sous-σ-additivité). Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements, alors*
$$\mathbb{P}\left(\bigcup_{n=0}^{+\infty} A_n\right) \leq \sum_{n=0}^{+\infty} \mathbb{P}(A_n)\,.$$

5. *(Croissance monotone). Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements croissante pour l'inclusion ($A_n \subset A_{n+1}$), alors*
$$\mathbb{P}\left(\bigcup_{n=0}^{+\infty} A_n\right) = \lim_{n\to+\infty} \mathbb{P}(A_n)\,.$$
Une illustration de cette propriété est donnée Figure 1.3.

6. *(Décroissance monotone). Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements décroissante pour l'inclusion ($A_{n+1} \subset A_n$), alors*
$$\mathbb{P}\left(\bigcap_{n=0}^{+\infty} A_n\right) = \lim_{n\to+\infty} \mathbb{P}(A_n)\,.$$
Une illustration de cette propriété est donnée Figure 1.4.

Démonstration.

1. (Ensemble vide). On choisit $A_0 = \Omega$ et $A_n = \emptyset$ pour tout $n \geq 1$. La σ-additivité donne alors
$$\mathbb{P}(\Omega) = \mathbb{P}(\Omega) + \sum_{n=1}^{+\infty} \mathbb{P}(\emptyset)\,,$$
d'où $\mathbb{P}(\emptyset) = 0$.

2. (Monotonie). On applique la σ-additivité avec les événements incompatibles $A_0 = A$ et $A_1 = B \setminus A$ (et $A_n = \emptyset$ pour $n \geq 2$) dont l'union est B, ce qui donne
$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)\,, \tag{1.1}$$
cf Figure 1.2. Comme $\mathbb{P}(B \setminus A) \geq 0$, on en déduit que $\mathbb{P}(A) \leq \mathbb{P}(B)$. Par ailleurs, avec $B = \Omega$ dans (1.1), on obtient $1 = \mathbb{P}(A) + \mathbb{P}(\bar{A})$, ce qui conduit au résultat souhaité.

3. (Additivité forte). Les événements A et $B \setminus A$ étant incompatibles, la σ-additivité implique que
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)\,.$$
De même, les événements $A \cap B$ et $B \setminus A$ étant incompatibles d'union B, la σ-additivité implique que
$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A)\,.$$
En regroupant ces deux égalités, on obtient le résultat souhaité :
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)\,.$$

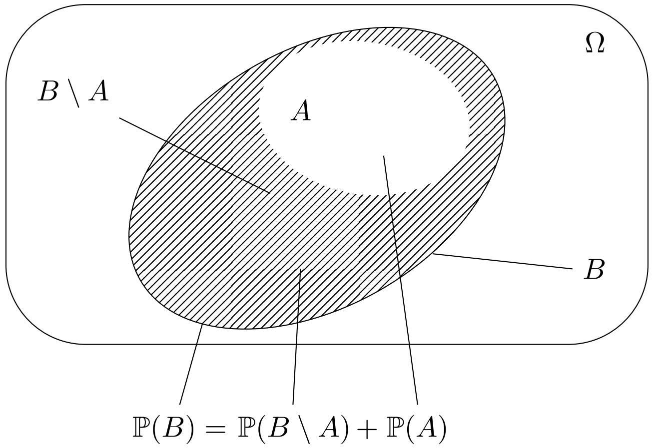


FIGURE 1.2 – Illustration de la monotonie : $\mathbb{P}(A) \leq \mathbb{P}(B)$.

4. (Sous-σ-additivité). On considère la suite d'événements $(B_n)_{n\in\mathbb{N}}$ définie par $B_0 = A_0$ et, pour $n \geq 1$,
$$B_n = A_n \setminus \left(\bigcup_{j=0}^{n-1} A_j \right).$$
Les événements B_n sont alors 2 à 2 incompatibles et on a :
$$\bigcup_{n=0}^{+\infty} A_n = \bigcup_{n=0}^{+\infty} B_n\,.$$
La σ-additivité permet alors d'écrire
$$\mathbb{P}\left(\bigcup_{n=0}^{+\infty} A_n \right) = \mathbb{P}\left(\bigcup_{n=0}^{+\infty} B_n \right) = \sum_{n=0}^{+\infty} \mathbb{P}(B_n) \leq \sum_{n=0}^{+\infty} \mathbb{P}(A_n)\,,$$
où la dernière inégalité résulte de l'inclusion $B_n \subset A_n$ et de la propriété de monotonie.
5. (Croissance monotone). On reprend la suite (B_n) précédente. La suite $(A_n)_{n\in\mathbb{N}}$ étant croissante, on a la relation
$$B_0 \cup B_1 \cup \cdots \cup B_n = A_n\,.$$
Ceci donne alors
$$\mathbb{P}\left(\bigcup_{n=0}^{+\infty} A_n \right) = \mathbb{P}\left(\bigcup_{n=0}^{+\infty} B_n \right) = \sum_{n=0}^{+\infty} \mathbb{P}(B_n) = \lim_{n\to+\infty} \sum_{k=0}^{n} \mathbb{P}(B_k) = \lim_{n\to+\infty} \mathbb{P}(A_n)\,.$$

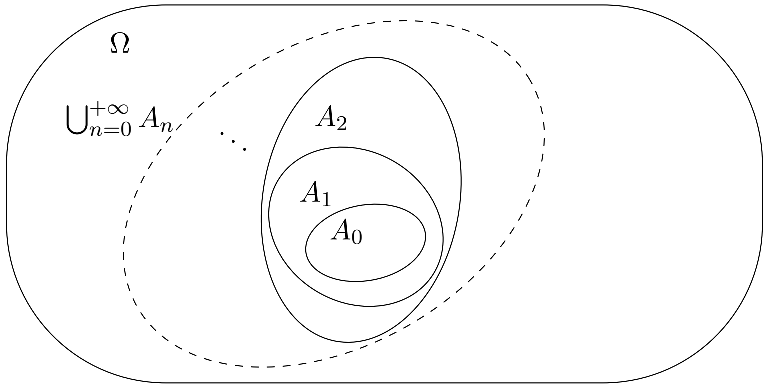


FIGURE 1.3 – Une suite d'ensembles croissante pour l'inclusion.

6. (Décroissance monotone). On considère la suite d'événements $(\bar{A}_n)_{n\in\mathbb{N}}$. D'une part, cette suite étant croissante, le point précédent et la propriété de monotonie assurent que

$$\mathbb{P}\left(\bigcup_{n=0}^{+\infty} \bar{A}_n\right) = \lim_{n\to+\infty} \mathbb{P}(\bar{A}_n) = 1 - \lim_{n\to+\infty} \mathbb{P}(A_n)\,.$$

D'autre part, l'égalité ensembliste $\bigcup_{n=0}^{+\infty} \bar{A}_n = \overline{\bigcap_{n=0}^{+\infty} A_n}$ implique que

$$\mathbb{P}\left(\bigcup_{n=0}^{+\infty} \bar{A}_n\right) = \mathbb{P}\left(\overline{\bigcap_{n=0}^{+\infty} A_n}\right) = 1 - \mathbb{P}\left(\bigcap_{n=0}^{+\infty} A_n\right).$$

En combinant les deux égalités on arrive au résultat souhaité.

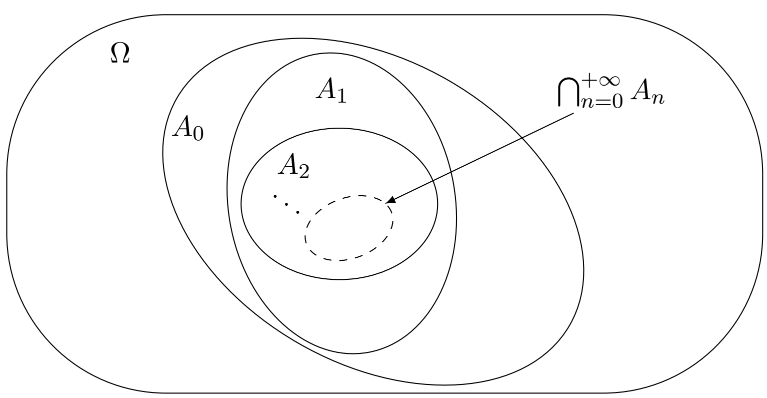


FIGURE 1.4 – Une suite d'ensembles décroissante pour l'inclusion.

□

Exemple 1.11. On lance une infinité de fois un dé. L'univers est alors $\Omega = \{1,2,3,4,5,6\}^{\mathbb{N}^*}$. On note A_n l'événement "le premier 6 apparaît au n-ème lancer". L'événement A_n est réalisé si on a obtenu un chiffre différent de 6 lors des $n-1$ premiers lancers et un 6 au n-ème lancer, d'où

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

Notons qu'en toute rigueur il faudrait invoquer ici la notion d'indépendance des lancers, notion qui sera définie dans la section suivante.

On définit à présent les événements B_n par

$$B_n = A_1 \cup \cdots \cup A_n\,, \quad n \geq 1\,.$$

En d'autres termes, B_n est réalisé si un 6 apparaît au cours des n premiers lancers. Les A_k étant des événements 2 à 2 incompatibles, on en déduit la probabilité de B_n :

$$\mathbb{P}(B_n) = \sum_{k=1}^{n} \mathbb{P}(A_k) = \sum_{k=1}^{n} \left(\frac{5}{6}\right)^{k-1} \times \frac{1}{6} = 1 - \left(\frac{5}{6}\right)^{n}.$$

Par ailleurs, la suite $(B_n)_{n\in\mathbb{N}}$ est croissante et vérifie l'égalité

$$\bigcup_{n=1}^{+\infty} B_n = \bigcup_{n=1}^{+\infty} (A_1 \cup \cdots \cup A_n) = A_1 \cup \cdots \cup A_n \cdots = \bigcup_{n=1}^{+\infty} A_n\,.$$

Une application de la propriété de croissance monotone donne alors

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{+\infty} B_n\right) = \lim_{n\to+\infty} \mathbb{P}(B_n) = \lim_{n\to+\infty} 1 - \left(\frac{5}{6}\right)^{n} = 1\,.$$

Ainsi, la probabilité d'obtenir un 6 en un nombre fini de lancers vaut 1. On vient de mettre en évidence un événement de probabilité 1 différent de l'univers Ω. Par passage au complémentaire, on en déduit que l'événement "obtenir aucun 6", bien que différent de l'ensemble vide, est de probabilité nulle.

Formule de Poincaré On peut étendre la propriété d'additivité forte à plus de deux événements. Dans le cas de trois événements A, B, et C, on obtient par exemple

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

Plus généralement, la probabilité de l'union de n événements $A_1, \ldots, A_n$ est donnée par la formule de Poincaré (ou encore formule du crible ou formule d'inclusion-exclusion) :

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \left(\sum_{1\leq i_1 < \ldots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) \right).$$

Il faut retenir que la probabilité d'une union peut se retrouver à l'aide des probabilités de toutes les intersections possibles des événements considérés : on ajoutera les probabilités quand elles concernent un nombre impair d'intersections, on les retranchera quand elles concernent un nombre pair.

2.2 Construction d'une probabilité sur un univers fini ou dénombrable

2.2.1 Le cas fini

On a vu que dans le cas d'équiprobabilité, la probabilité d'un événement A est définie par $\mathbb{P}(A) = \#A/\#\Omega$. En notant $\Omega = \{w_1, \ldots, w_n\}$, cela revient à définir l'équiprobabilité via les probabilités élémentaires

$$p_i = \mathbb{P}(\{\omega_i\}) = \frac{1}{n}\,, \quad i = 1, \ldots, n\,.$$

On notera plus simplement $\mathbb{P}(\{\omega_i\}) = \mathbb{P}(\omega_i)$.

De manière générale, si $\Omega = \{w_1, \ldots, w_n\}$, une probabilité sur Ω est entièrement déterminée par la donnée des probabilités élémentaires $p_1, \ldots, p_n$, quantités positives qui somment à 1. La probabilité d'un événement A est alors donnée par

$$\mathbb{P}(A) = \sum_{i:\omega_i \in A} p_i\,.$$

Exemple 1.12. On considère le lancer de deux dés et on s'intéresse à la somme du résultat obtenu. Ici $\Omega = \{2, 3, \ldots, 12\}$ mais il n'y a pas équiprobabilité. Les probabilités élémentaires sont données par

$$(p_2, p_3, \ldots, p_{12}) = \left(\frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \frac{4}{36}, \frac{5}{36}, \frac{6}{36}, \frac{5}{36}, \frac{4}{36}, \frac{3}{36}, \frac{2}{36}, \frac{1}{36}\right).$$

La probabilité d'obtenir un nombre premier est alors donné par la somme

$$\mathbb{P}(\text{nombre premier}) = p_2 + p_3 + p_5 + p_7 + p_{11} = \frac{1}{36} + \frac{2}{36} + \frac{4}{36} + \frac{6}{36} + \frac{2}{36} = \frac{15}{36}\,.$$

2.2.2 Le cas infini dénombrable

Si $\Omega = (\omega_i)_{i \in \mathbb{N}}$ est un univers infini dénombrable, définir une probabilité $\mathbb{P}$ revient alors à se donner une suite de nombres positifs $(p_i)_{i \in \mathbb{N}}$ telle que la série $\sum p_i$ soit convergente de somme égale à 1. La probabilité $\mathbb{P}$ est alors donnée par

$$\mathbb{P}(A) = \sum_{i:\omega_i \in A} p_i\,,$$

pour tout événement A.

Exemple 1.13. On considère le lancer d'une pièce équilibrée et on compte le nombre de lancers nécessaires pour obtenir Pile. L'univers est donc $\Omega = \mathbb{N}^*$ et on trouve facilement les probabilités élémentaires $p_1 = 1/2$, $p_2 = 1/4$, etc., $p_i = 1/2^i$. On vérifie alors que $\sum_{i=1}^{\infty} p_i = 1$ (somme d'une série géométrique).

Exercice 1.14. Quelle est la probabilité que le premier Pile apparaisse au bout d'un nombre pair de lancers ? d'un nombre impair ?

Remarque 1.15. Si Ω est infini dénombrable, disons $\Omega = \mathbb{N}$, alors on ne peut pas le munir d'une probabilité uniforme. En effet, si c'était possible il existerait $p > 0$ tel que $\mathbb{P}(\{n\}) = p$ pour tout $n \in \mathbb{N}$, et donc

$$\mathbb{P}(\mathbb{N}) = \mathbb{P}\left(\bigcup_{n=0}^{+\infty} \{n\}\right) = \sum_{n=0}^{+\infty} \mathbb{P}(\{n\}) = \sum_{n=0}^{+\infty} p = +\infty\,,$$

ce qui contredit $\mathbb{P}(\mathbb{N}) = 1$.

3 Conditionnement

La notion de conditionnement permet de tenir compte de l'information que l'on dispose pour évaluer la probabilité d'un événement. Dans tout ce qui suit, $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé.

3.1 Probabilité conditionnelle

Définition 1.16 (Probabilité conditionnelle)**.** *Soit A et B deux événements tels que $\mathbb{P}(A) > 0$. La* probabilité conditionnelle de B sachant A*, notée $\mathbb{P}(B \mid A)$ (ou parfois $\mathbb{P}_A(B)$), est définie par*

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

On prendra garde au fait qu'il n'y a pas de relation d'ordre entre $\mathbb{P}(B \mid A)$ et $\mathbb{P}(B)$: il se peut que $\mathbb{P}(B \mid A)$ soit faible et $\mathbb{P}(B)$ grande (et inversement).

Exemple 1.17. Une classe de Licence est composée de 80 garçons et 20 filles. Parmi ces élèves, 15 filles et 5 garçons suivent un cours d'espagnol. On choisit un élève au hasard dans cette classe. La probabilité que ce soit une fille vaut bien évidemment $\mathbb{P}(F) = 20/100 = 1/5$. Si on sait que l'élève choisi suit le cours d'espagnol, alors la probabilité que ce soit une fille vaut

$$\mathbb{P}(F \mid Esp) = \frac{\mathbb{P}(F \cap Esp)}{\mathbb{P}(Esp)} = \frac{15/100}{20/100} = \frac{15}{20} = \frac{3}{4}\,.$$

Si on fixe un événement A de probabilité positive, alors l'application $B \mapsto \mathbb{P}(B \mid A)$ définie sur la tribu $\mathcal{F}$ est une probabilité. Ce résultat est énoncé dans la proposition suivante.

Proposition 1.18. *Soit $A \in \mathcal{F}$ tel que $\mathbb{P}(A) > 0$. Alors l'application de $\mathcal{F}$ dans $[0,1]$ qui à un événement $B \in \mathcal{F}$ associe $\mathbb{P}(B \mid A)$ définit une nouvelle probabilité sur Ω, appelée* probabilité conditionnelle sachant A *:*

1. *$\mathbb{P}(\Omega \mid A) = 1$.*
2. *Si $(B_n)_{n\in\mathbb{N}}$ est une suite d'événements 2 à 2 disjoints, alors :*

$$\mathbb{P}\left(\bigcup_{n=0}^{+\infty} B_n \,\middle|\, A\right) = \sum_{n=0}^{+\infty} \mathbb{P}(B_n \mid A)\,.$$

En particulier, l'application $B \mapsto \mathbb{P}(B \mid A)$ vérifie toutes les propriétés énoncées dans la proposition 1.10.

Démonstration. Notons tout d'abord que $\mathbb{P}(B \mid A)$ est une quantité positive car quotient de deux probabilités, et inférieure à 1 car $\mathbb{P}(B\cap A) \leq \mathbb{P}(A)$ (propriété de monotonie). L'application $\mathbb{P}(\cdot \mid A)$ est donc bien à valeurs dans $[0,1]$.

Le premier point résulte des égalités suivantes :

$$\mathbb{P}(\Omega \mid A) = \frac{\mathbb{P}(\Omega \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A)}{\mathbb{P}(A)} = 1\,.$$

Concernant le second point, il résulte de la σ-additivité de la probabilité $\mathbb{P}$ et du fait que les événements $B_n \cap A$ sont 2 à 2 incompatibles :

$$\mathbb{P}\left(\bigcup_{n=0}^{+\infty} B_n \,\middle|\, A\right) = \frac{\mathbb{P}\Big(\big(\bigcup_{n=0}^{+\infty} B_n\big)\cap A\Big)}{\mathbb{P}(A)} = \frac{\mathbb{P}\Big(\bigcup_{n=0}^{+\infty}(B_n \cap A)\Big)}{\mathbb{P}(A)} = \sum_{n=0}^{+\infty} \frac{\mathbb{P}(B_n\cap A)}{\mathbb{P}(A)} = \sum_{n=0}^{+\infty} \mathbb{P}(B_n|A)\,.$$

□

Exemple 1.19. Vous rencontrez un ancien camarade de classe qui vous annonce qu'il est l'heureux papa de deux enfants, l'un d'entre eux étant un garçon. On souhaite calculer la probabilité qu'il ait deux garçons. On considère donc l'univers $\Omega = \{GG, FF, FG, GF\}$, où par exemple GF désigne avoir un garçon puis un fille. On le munit de la tribu pleine $\mathcal{F} = \mathcal{P}(\Omega)$ et de l'équiprobabilité. La probabilité de l'événement $A = \{GG, FG, GF\}$ vaut $\mathbb{P}(A) = \#A/\#\Omega = 3/4$, tandis que celle de l'événement $B = \{GG\}$ vaut $\mathbb{P}(B) = \#B/\#\Omega = 1/4$. Comme $A\cap B = B$, la probabilité cherchée est donnée par

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

Exercice 1.20. Que devient cette probabilité si votre camarade vous annonce "l'aîné s'appelle Pierre" ?

Proposition 1.21. *Si $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$, alors*

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

Cette formule est utile pour inverser le conditionnement. En effet, si on connaît la probabilité conditionnelle de B sachant A, on peut en déduire la probabilité conditionnelle de A sachant B :

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

3.2 Quelques formules utiles

On donne ici plusieurs résultats d'usage constant faisant intervenir des probabilités conditionnelles. L'idée générale est qu'une combinaison astucieuse des conditionnements permet de calculer de nombreuses probabilités.

Proposition 1.22 (Formule des probabilités composées)**.** *Soit* $A_1, \dots, A_n$ *des événements tels que* $\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) > 0$*. Alors,*

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

Cette formule est utile quand on enchaîne des expériences aléatoires : l'événement A_1 correspond alors à un événement d'intérêt de la première expérience, l'événement A_2 de la deuxième expérience, etc. On peut ainsi déterminer une probabilité concernant la dernière expérience à partir des précédentes.

Démonstration. Notons tout d'abord que toutes les probabilités conditionnelles sont bien définies puisque

$$0 < \mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) \leq \mathbb{P}(A_1 \cap \cdots \cap A_{n-2}) \leq \cdots \leq \mathbb{P}(A_1)\,.$$

Maintenant, pour $k = 2, \dots, n$, on écrit

$$\mathbb{P}(A_k \mid A_{k-1} \cap \cdots \cap A_1) = \frac{\mathbb{P}(A_k \cap A_{k-1} \cap \cdots \cap A_1)}{\mathbb{P}(A_{k-1} \cap \cdots \cap A_1)}\,.$$

En multipliant terme à terme ces $n-1$ égalités on obtient un produit téléscopique qui se simplifie en

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

ce qui conclut la preuve. □

Le prochain résultat repose sur la notion de partition, aussi appelée système complet d'événements.

Définition 1.23 (Partition)**.** *Soit* $I \subset \mathbb{N}$ *et* $(A_i)_{i \in I}$ *une suite (finie ou non) d'événements. On dit que cette suite forme une* partition *de* Ω *si*

1. *pour tout* $i \neq j$, $A_i \cap A_j = \emptyset$ *(ensembles 2 à 2 disjoints),*
2. $\Omega = \bigcup_{i \in I} A_i$.

Il faut penser une partition comme les pièces d'un puzzle : elles ne doivent pas se chevaucher et doivent recouvrir tout l'espace (cf Figure 1.5 pour une illustration).

Exemple 1.24. Si $A \in \mathcal{F}$, alors le couple d'événements $(A, \bar{A})$ forment une partition de Ω.

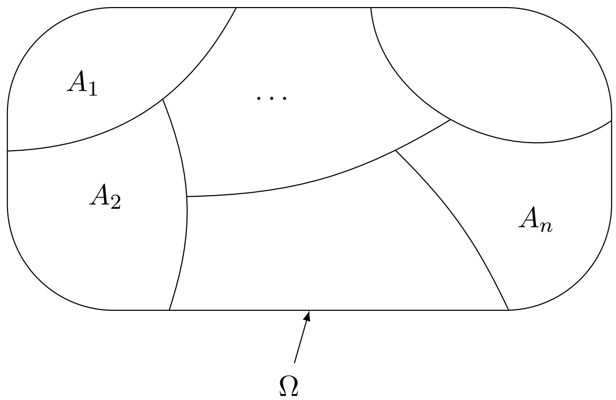


FIGURE 1.5 – Une partition $(A_1, \dots, A_n)$ de Ω.

Étant donnée une partition, le résultat suivant permet de calculer la probabilité $\mathbb{P}(B)$ à partir de chaque probabilité conditionnelle $\mathbb{P}(B \mid A_i)$. En résumé, si on connaît la probabilité de B sur sachant les A_i on la connaît partout.

Proposition 1.25 (Formule des probabilités totales)**.** *Soit* $(A_i)_{i\in I}$ *une partition de* Ω *telle que* $\mathbb{P}(A_i) > 0$ *pour tout* $i \in I$*. Pour tout événement* B*, on a*

$$\mathbb{P}(B) = \sum_{i\in I} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)\,.$$

Démonstration. Il suffit de remarquer que les événements $B \cap A_i$ sont 2 à 2 incompatibles et que leur réunion vaut B. Par suite, on obtient le résultat voulu :

$$\mathbb{P}(B) = \mathbb{P}\Big(\bigcup_{i\in I}(B \cap A_i)\Big) = \sum_{i\in I} \mathbb{P}(B \cap A_i) = \sum_{i\in I} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)\,.$$

La relation $B = \bigcup_{i\in I}(B \cap A_i)$ est illustrée Figure 1.6.

□

En pratique, il vaut mieux réécrire cette ligne de calculs avec les probabilités ad hoc.

Un cas particulier intéressant de cette formule concerne le cas d'une partition formée par deux événements A et $\bar{A}$. La formule devient alors

$$\mathbb{P}(B) = \mathbb{P}(B \mid A)\mathbb{P}(A) + \mathbb{P}(B \mid \bar{A})\mathbb{P}(\bar{A})\,.$$

Exemple 1.26. On considère deux urnes : la première contient 2 boules noires et 4 boules rouges, tandis que la seconde contient 6 boules noires, 2 boules rouges, et 4 boules bleues. Une personne choisit au hasard une urne et y pioche une boule. On souhaite calculer la probabilité que la boule

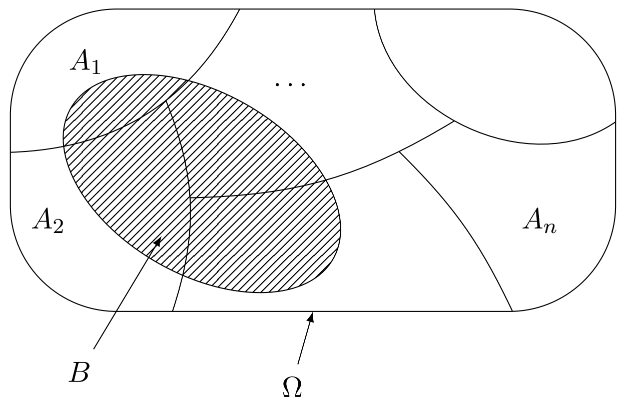


FIGURE 1.6 – Illustration de $B = \bigcup_{i \in I}(B \cap A_i)$.

tirée soit rouge. On note U_i l'événement "la boule tirée provient de l'urne i", avec $i = 1, 2$, et R l'événement "la boule tirée est rouge". Si on connaît l'urne dans laquelle la boule a été tirée, la probabilité que cette boule soit rouge est alors facile à déterminer. Comme les événements U_1 et U_2 forment une partition de l'univers, on peut appliquer la formule des probabilités totales :

$$\mathbb{P}(R) = \mathbb{P}(R \mid U_1)\mathbb{P}(U_1) + \mathbb{P}(R \mid U_2)\mathbb{P}(U_2) = \frac{4}{6} \times \frac{1}{2} + \frac{2}{12} \times \frac{1}{2} = \frac{5}{12}\,.$$

La probabilité de tirer une boule rouge vaut donc 5/12.

On arrive à présent à la formule de Bayes qui permet d'inverser le conditionnement : on calcule une probabilité "sachant B" grâce à des probabilités "sachant A_i".

Proposition 1.27 (Formule de Bayes)**.** *Soit $(A_i)_{i \in I}$ une partition de Ω telle que $\mathbb{P}(A_i) > 0$ pour tout $i \in I$. Alors pour tout événement B tel que $\mathbb{P}(B) > 0$ et pour tout indice $j \in I$, on a*

$$\mathbb{P}(A_j \mid B) = \frac{\mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}{\sum_{i \in I} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}\,.$$

Démonstration. Il suffit d'écrire

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

et de développer le dénominateur à l'aide de la formule des probabilités totales. □

Comme pour la formule des probabilités totales, la formule de Bayes est d'usage courant avec la partition $(A, \bar{A})$ où A est un événement donné. On obtient alors

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B \mid A)\mathbb{P}(A) + \mathbb{P}(B \mid \bar{A})\mathbb{P}(\bar{A})}\,.$$

Exemple 1.28. Une maladie touche une personne sur 10 000. Un test de dépistage est proposé. Si une personne est malade, le test est positif à 99%. Si une personne n'est pas malade, le test est positif à 0.1%. On souhaite savoir si ce test est efficace, c'est-à-dire si une personne testée positive est effectivement malade.

Pour cela, on note M l'événement "la personne est malade" et P l'événement "le test est positif". Le quantité qui nous intéresse est $\mathbb{P}(M \mid P)$. Les données se traduisent alors par

$$\mathbb{P}(M) = 10^{-4}\,, \quad \mathbb{P}(\bar{M}) = 0,9999\,, \quad \mathbb{P}(T \mid M) = 0.99\,, \quad \mathbb{P}(T \mid \bar{M}) = 10^{-3}\,.$$

La formule de Bayes donne alors

$$\mathbb{P}(M \mid T) = \frac{\mathbb{P}(T \mid M)\mathbb{P}(M)}{\mathbb{P}(T \mid M)\mathbb{P}(M) + \mathbb{P}(T \mid \bar{M})\mathbb{P}(\bar{M})} = \frac{10^{-4} \times 0.99}{0.99 \times 10^{-4} + 10^{-3} \times 0.9999} \approx 0.09\,.$$

Il n'y a que 9% de chances qu'une personne positive au test soit effectivement malade. C'est tout le problème des tests de dépistage pour les maladies rares : ils doivent être excessivement performants sans quoi ils donneront beaucoup trop de faux positifs.

4 Indépendance

4.1 Indépendance de deux événements

Intuitivement, deux événements A et B sont indépendants si le fait de savoir que A est réalisé n'a pas d'influence sur la probabilité de réalisation de B, et réciproquement. Cela revient à écrire $\mathbb{P}(B \mid A) = \mathbb{P}(B)$, sous réserve que $\mathbb{P}(A) > 0$. Pour éviter de devoir vérifier la positivité de $\mathbb{P}(A)$, et dans un souci de symétrie, on adopte la définition suivante.

Définition 1.29 (Événements indépendants)**.** *Deux événements A et B sont dits* indépendants *si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.*

Exemple 1.30. On tire une carte dans un jeu de 52 cartes. L'univers est alors l'ensemble des cartes du jeu que l'on munit de la tribu pleine et de la probabilité uniforme. On note A l'événement "la carte tirée est un roi" et B l'événement "la carté tirée est un trêfle". On a alors :

$$\mathbb{P}(A) = \frac{4}{52}\,, \quad \mathbb{P}(B) = \frac{13}{52}\,, \quad \mathbb{P}(A \cap B) = \frac{1}{52}\,.$$

Ainsi, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Les événements A et B sont donc indépendants.

Exercice 1.31. Montrer que si l'as de pique a été égaré et que le paquet ne contient ainsi que 51 cartes, alors les événements A et B ne sont plus indépendants.

Attention. Il ne faut pas confondre incompatibles et indépendants. La notion d'événements incompatibles est une notion ensembliste et ne dépend pas de la probabilité choisie. L'indépendance est une notion liée au choix de la probabilité $\mathbb{P}$. Si l'on change de probabilité, on ne conserve pas nécessairement l'indépendance des événements.

Le cas de deux événements indépendants non incompatibles est donné dans l'exemple précédent (leur intersection correspond au roi de trèfle).

Réciproquement, si on considère un événement A de probabilité dans $]0,1[$, alors le produit $\mathbb{P}(A)\mathbb{P}(\bar{A})$ est non nul alors que $\mathbb{P}(A\cap\bar{A})=0$. Les deux événements A et $\bar{A}$ sont donc incompatibles mais dépendants.

Proposition 1.32. *Si deux événements A et B sont indépendants, alors il en est de même de A et $\bar{B}$, $\bar{A}$ et B, $\bar{A}$ et $\bar{B}$.*

Démonstration. On montre uniquement le premier point. On part de l'égalité $A\cap\bar{B}=A\setminus(A\cap B)$ qui donne
$$\mathbb{P}(A\cap\bar{B})=\mathbb{P}(A)-\mathbb{P}(A\cap B)\,.$$
L'indépendance de A et B permet alors de conclure :
$$\mathbb{P}(A\cap\bar{B})=\mathbb{P}(A)-\mathbb{P}(A)\mathbb{P}(B)=\mathbb{P}(A)(1-\mathbb{P}(B))=\mathbb{P}(A)\mathbb{P}(\bar{B})\,.$$
□

Ainsi, pour vérifier l'indépendance de deux événements on peut étudier leur complémentaire.

Exemple 1.33. Si on reprend l'exemple précédent, les événements "ne pas tirer un roi" et "tirer un trèfle" sont donc indépendants.

Attention. On prendra garde au fait, qu'en général, $\mathbb{P}(A\cap B)\neq\mathbb{P}(A)\mathbb{P}(B)$. Cette relation n'est vraie que sous l'hypothèse d'indépendance de A et B. Pour calculer une quantité du type $\mathbb{P}(A\cap B)$ lorsqu'on n'a pas fait d'hypothèses particulières, on repassera toujours par $\mathbb{P}(A\mid B)\mathbb{P}(B)$ ou bien $\mathbb{P}(B\mid A)\mathbb{P}(A)$.

4.2 Indépendance de plusieurs événements

On souhaite généraliser la notion d'indépendance à une famille d'événements. Deux notions d'indépendance apparaissent alors.

Définition 1.34 (Indépendance 2 à 2 et indépendance mutuelle)**.** *Soit $I\subset\mathbb{N}$ et $(A_i)_{i\in I}$ une suite (finie ou non) d'événements. Ces événements sont dits :*

1. 2 à 2 indépendants *si pour tout $i\neq j$ les événements A_i et A_j sont indépendants :*
$$\mathbb{P}(A_i\cap A_j)=\mathbb{P}(A_i)\mathbb{P}(A_j)\,,$$
2. mutuellement indépendants *si pour tout sous-ensemble d'indices $(i_1,\dots,i_k)\subset I$, on a :*
$$\mathbb{P}(A_{i_1}\cap\dots\cap A_{i_k})=\mathbb{P}(A_{i_1})\cdots\mathbb{P}(A_{i_k})\,.$$

Conséquence. Trois événements A, B et C sont mutuellement indépendants si et seulement si les quatre égalités suivantes sont vérifiées : $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, $\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$, $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$, et $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$. Les trois premières égalités donnent seulement l'indépendance 2 à 2 des trois événements.

Il est clair que l'indépendance mutuelle implique l'indépendance 2 à 2 : il suffit de prendre $i_1 = i$ et $i_2 = j$. La réciproque est fausse.

Exemple 1.35. On lance deux pièces. On munit l'univers $\Omega = \{PP, PF, FP, FF\}$ de la probabilité uniforme et on considère les événements A : "la première pièce tombe sur Pile ", B : "la deuxième pièce tombe sur Pile ", et C : "les deux pièces tombent sur le même côté". On obtient alors

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

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

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

On vérifie alors que ces trois événements sont 2 à 2 indépendants mais ne sont pas mutuellement indépendants puisque

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

Exercice 1.36. On lance deux dés et on considère les événements A : "le premier dé donne un résultat impair" ; B : "le deuxième dé donne un résultat pair" ; C : "la somme des deux dés est un nombre pair". Montrer que ces trois événements sont deux à deux indépendants mais ne sont pas mutuellement indépendants.

En pratique, c'est souvent l'indépendance mutuelle qui nous intéresse. Dans la suite, une famille d'événements indépendants désignera ainsi des événements mutuellement indépendants.

Remarque 1.37. On se rend compte que pour vérifier proprement l'indépendance de plusieurs événements, cela peut s'avérer (très) long ! En pratique, c'est le contexte de l'expérience qui assurera l'indépendance des événements. C'est le cas lorsqu'on répète plusieurs fois une expérience (par exemple un lancer de dé), le résultat de chacune de ces expériences n'ayant aucune influence sur les autres.

Chapitre 2

Variables aléatoires discrètes

Lorsqu'on réalise une expérience aléatoire, on s'intéresse souvent à une fonction du résultat plutôt qu'au résultat brut lui-même. Par exemple, si on joue à la roulette au casino, ce qui nous intéresse est notre gain potentiel plutôt que le chiffre (ou la couleur) qui sort. La notion de variable aléatoire permet de modéliser ceci.

1 Loi d'une variable aléatoire discrète

On rappelle qu'un ensemble $\mathcal{X} = (x_k)_{k\in K}$ est dit au plus dénombrable s'il est fini ou dénombrable, c'est-à-dire si on peut énumérer ses éléments ($K \subset \mathbb{N}$). En pratique, on peut voir un tel ensemble comme une partie de $\mathbb{N}$, ou $\mathbb{N}$ tout entier.

1.1 Définitions et premiers exemples

Définition 2.1 (Variable aléatoire discrète)**.** *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace probabilisé. Soit* $\mathcal{X} = (x_k)_{k\in K}$ *un ensemble au plus dénombrable de* $\mathbb{R}$*. On appelle* variable aléatoire discrète *toute application*

$$\begin{array}{rcccc} X & : & \Omega & \to & \mathcal{X} \\ & & \omega & \mapsto & X(\omega)\,, \end{array}$$

vérifiant

$$\forall k \in K, \quad \{X = x_k\} = X^{-1}(\{x_k\}) = \{\omega \in \Omega : X(\omega) = x_k\} \in \mathcal{F}\,.$$

Cette définition se justifie de la manière suivante : on souhaite déterminer des probabilités du type $\mathbb{P}(\{X = x_k\})$ (ou faisant intervenir ce type d'événement) ; ceci nécessite que l'ensemble $\{X = x_k\}$ soit bien un événement, c'est-à-dire appartienne à la tribu $\mathcal{F}$.

Remarques 2.2.

1. La définition d'une variable aléatoire discrète permet de considérer les ensembles $\{X = x_k\}$ comme des événements. Par suite, tout autre sous-ensemble du type $\{X \in A\} = \{\omega \in \Omega :$

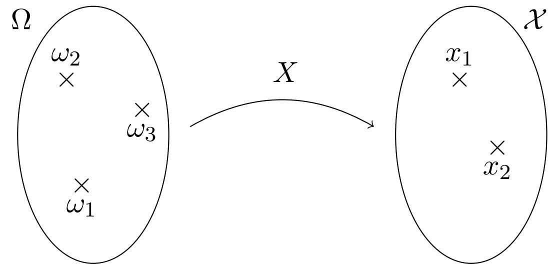


FIGURE 2.1 – Une variable aléatoire discrète est une fonction d'un espace probabilisé Ω vers un espace discret $\mathcal{X} \subset \mathbb{R}$.

$X(\omega) \in A\}$ est également un événement puisqu'il peut s'écrire comme union des événements élémentaires $\{X = x_k\}$ pour $x_k \in A$.

2. On simplifie les notations en écrivant $\mathbb{P}(X = x_k)$ au lieu de $\mathbb{P}(\{X = x_k\})$.
3. Il faut prendre garde au fait qu'une variable aléatoire discrète X n'est pas nécessairement à valeurs dans $\mathbb{N}$. Elle peut prendre ses valeurs dans $\mathbb{R}$: $\pi, \sqrt{2}, \ldots$ L'ensemble d'arrivée $\mathcal{X}$ est appelé le *support* de la variable aléatoire X.
4. La terminologie "variable aléatoire" est trompeuse. Une variable aléatoire n'est pas une variable, c'est une fonction, et elle n'a rien d'aléatoire : une fois qu'un événement élémentaire ω est choisi dans Ω, la valeur $X(\omega)$ est déterministe. Il faut voir le processus en deux étapes : la première est aléatoire et consiste à choisir un élément ω selon la probabilité $\mathbb{P}$; la seconde consiste à associer à cet élément la valeur d'intérêt $X(\omega)$.
5. La notion de variable aléatoire discrète est stable par toutes les opérations classiques : combinaison linéaire, produit, maximum, minimum, etc.

Exemples 2.3.

1. On lance deux pièces de monnaie et on compte le nombre de pile obtenu. On part alors de l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ avec $\Omega = \{PP, PF, FP, FF\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, et $\mathbb{P}$ la probabilité uniforme. On considère la variable aléatoire $X : \Omega \to \mathcal{X} = \{0, 1, 2\}$ définie par
$$X(PP) = 2\,, \quad X(PF) = 1\,, \quad X(FP) = 1\,, \quad X(FF) = 0\,,$$
qui donne bien le nombre de pile obtenu.
2. On lance deux dés et on s'intéresse à la somme des résultats. Cette expérience peut se modéliser via la variable aléatoire $Z : \Omega \to \mathcal{X} = \{2, 3, \ldots, 12\}$ qui part de l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ donné par $\Omega = \{1, \ldots, 6\}^2$, $\mathcal{F} = \mathcal{P}(\Omega)$, et $\mathbb{P}$ la probabilité uniforme $\mathbb{P}$, et qui est définie par
$$Z(\omega_1, \omega_2) = \omega_1 + \omega_2\,.$$

Puisque notre intérêt réside davantage dans l'ensemble $\mathcal{X}$ que dans l'univers Ω, l'idée est de transporter la probabilité $\mathbb{P}$ sur Ω vers l'ensemble $\mathcal{X}$. Ceci motive la notion de loi d'une variable aléatoire.

Définition 2.4 (Loi d'une variable aléatoire discrète)**.** *Soit X une variable aléatoire discrète à valeurs dans $\mathcal{X} = (x_k)_{k\in K}$. On appelle* loi de la variable aléatoire X *la suite des $(p_k)_{k\in K}$, avec*

$$\forall k \in K\,, \quad p_k = \mathbb{P}(X = x_k)\,.$$

Donner la loi d'une variable aléatoire discrète X revient donc à donner le support $(x_k)_{k\in K}$ de X et les probabilités $(p_k)_{k\in K}$ associées. On peut la représenter dans un tableau avec les x_k sur une ligne et les p_k associés en-dessous.

Exemple 2.5. On reprend l'exemple du lancer de deux pièces. La variable aléatoire X qui donne le nombre de pile vérifie

$$\begin{aligned}
\mathbb{P}(X=0) &= \mathbb{P}(\{\omega\in\Omega : X(\omega)=0\}) = \mathbb{P}(\{FF\}) = \frac{1}{4}\,,\\
\mathbb{P}(X=1) &= \mathbb{P}(\{\omega\in\Omega : X(\omega)=1\}) = \mathbb{P}(\{FP, PF\}) = \frac{1}{2}\,,\\
\mathbb{P}(X=2) &= \mathbb{P}(\{\omega\in\Omega : X(\omega)=2\}) = \mathbb{P}(\{PP\}) = \frac{1}{4}\,.
\end{aligned}$$

D'où la loi de X résumée dans le tableau suivant.

x_k	0	1	2
p_k	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Exemple 2.6. On considère une succession de lancers de dé et on note Y la variable aléatoire qui donne le nombre de lancers de dé nécessaires pour obtenir le chiffre 6 :

$$\begin{array}{rrcl}
Y \;:\; & \Omega = \{(u_n)_{n\geq 1} : u_n \in \{1,\dots,6\}\} & \to & \mathcal{X}\\
& (u_n)_{n\geq 1} & \mapsto & \min\{n\geq 1 : u_n = 6\}\,,
\end{array}$$

Notons qu'a priori l'espace d'arrivée est $\mathbb{N}^* \cup \{\infty\}$ puisque dans Ω il y a des suites qui ne comportent aucun 6. Cependant, on a vu dans l'exemple 1.11 du Chapitre 1 que la probabilité d'obtenir aucun 6 est nulle. On peut donc restreindre le support de Y à $\mathcal{X} = \mathbb{N}^*$. La loi de Y est alors donnée par

$$p_k = \mathbb{P}(X=k) = \mathbb{P}(\min\{n\geq 1 : u_n = 6\} = k) = \mathbb{P}(u_1 \neq 6, \dots, u_{k-1}\neq 6, u_k = 6) = \left(\frac{5}{6}\right)^{k-1}\frac{1}{6}\,,$$

où un argument d'indépendance des lancers est utilisé à la dernière égalité.

On peut résumer la loi de Y dans le tableau suivant.

x_k	1	2	$\cdots$	k	$\cdots$
p_k	$\frac{1}{6}$	$\frac{5}{6}\times\frac{1}{6}$	$\cdots$	$\left(\frac{5}{6}\right)^{k-1}\frac{1}{6}$	$\cdots$

Un point important. Il faut s'habituer à la notation $\mathbb{P}(X = x)$ qui peut surprendre. Elle correspond à la probabilité d'avoir choisi un élément ω de Ω dont l'image par X vaut x :

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

Plus généralement, si A est une partie de $\mathcal{X}$ (ou de $\mathbb{R}$) on écrira alors

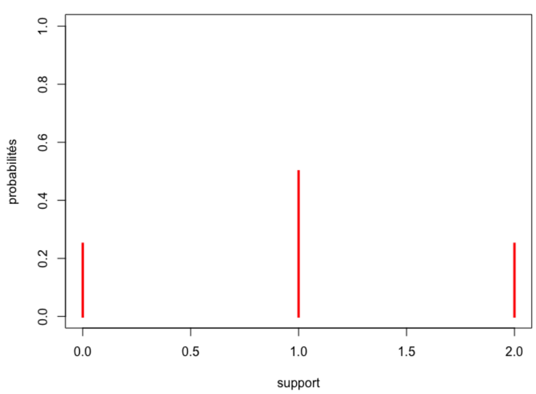
$$\mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) .$$

Par exemple, avec $A =]-\infty, x]$, pour $x \in \mathbb{R}$ fixé, on écrit

$$\mathbb{P}(X \leq x) = \mathbb{P}(X \in]-\infty, x]) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\}) .$$

L'idée de cette notation est d'oublier complètement comment a été généré le hasard en mettant Ω sous le tapis.

On représente parfois la loi d'une variable aléatoire discrète sous forme de diagramme en bâtons. Celles des exemples précédents sont données dans en Figure 2.2.



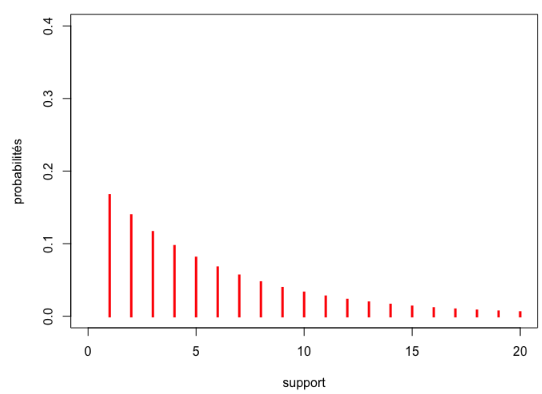


FIGURE 2.2 – Diagrammes en bâtons de la loi modélisant le lancer de deux pièces (gauche) et de la loi modélisant l'apparition du premier 6 dans une succession de lancers de dé (droite).

Proposition 2.7. *Soit X une variable aléatoire discrète à valeurs dans $\mathcal{X} = (x_k)_{k \in K}$ et de loi $p_k = \mathbb{P}(X = x_k)$. Alors on a*

1. *pour tout $k \in K$, $0 \leqslant p_k \leqslant 1$,*
2. *$\sum_{k \in K} p_k = 1$.*

Dans tout ce qui suit, on illustrera les différents résultats sur les exemples 2.5 et 2.6 que l'on désignera respectivement par "Lancer de deux pièces" et "Premier 6".

1.2 Fonction de répartition

Définition 2.8 (Fonction de répartition)**.** *Soit X une variable aléatoire discrète. On appelle* fonction de répartition *de X la fonction F définie par*

$$\begin{array}{rcccl} F & : & \mathbb{R} & \to & \mathbb{R} \\ & & x & \mapsto & F(x) = \mathbb{P}(X \leq x). \end{array}$$

On notera parfois $F = F_X$ pour préciser qu'on parle bien de la variable aléatoire X.

Exemples 2.9.

1. *Lancer de deux pièces.* La fonction de répartition de cette variable aléatoire vaut 0 sur $]-\infty, 0[$, $1/4$ sur $[0, 1[$, $3/4$ sur $[1, 2[$, et 1 sur $[2, +\infty[$. Son graphe est représenté en Figure 2.3.
2. *Premier 6.* Notons que pour tout entier $k \geq 1$, on a

$$\mathbb{P}(X \leq k) = \sum_{j=1}^{k} \mathbb{P}(X = j) = \sum_{j=1}^{k} \left(\frac{5}{6}\right)^{k-1} \frac{1}{6} = \frac{1}{6} \times \frac{1 - (5/6)^k}{1 - 5/6} = 1 - \left(\frac{5}{6}\right)^k.$$

Par suite, la fonction de répartition F de X vaut 0 sur $]-\infty, 1[$, et

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \leq \lfloor x \rfloor) = 1 - \left(\frac{5}{6}\right)^{\lfloor x \rfloor}, \quad x \geq 1,$$

où $\lfloor x \rfloor$ désigne la partie entière de x. Son graphe est représentée en Figure 2.3.

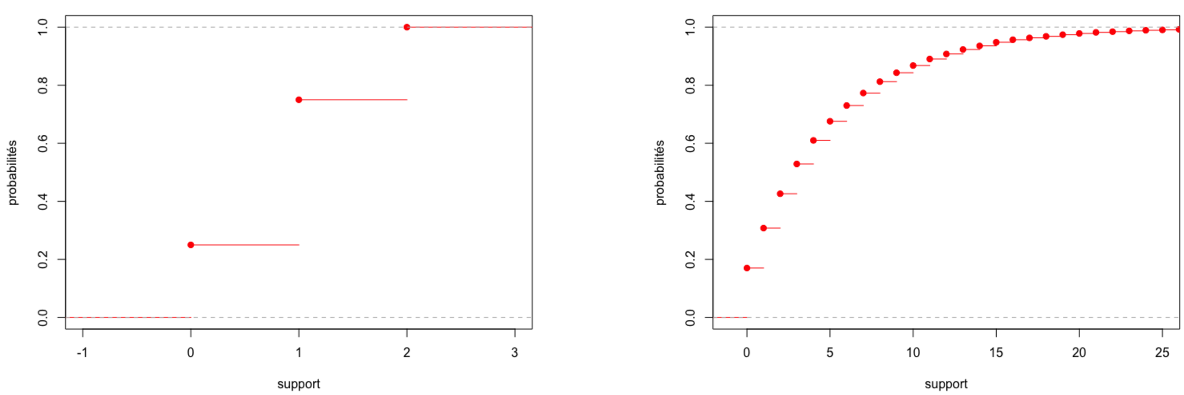


FIGURE 2.3 – Fonction de répartition de la loi modélisant le lancer de deux pièces (gauche) et de la loi modélisant l'apparition du premier 6 dans une succession de lancers de dé (droite).

En pratique. Si X est une variable aléatoire discrète de support $(x_k)_{k\in K}$ et de loi $(p_k)_{k\in K}$, alors la valeur de sa fonction de répartition F en un point x se calcule en sommant les probabilités p_k qui correspondent aux valeurs x_k inférieures à x :

$$F(x) = \mathbb{P}(X \le x) = \sum_{k : x_k \le x} p_k \,.$$

La fonction de répartition d'une variable aléatoire discrète est donc une fonction en escalier. On recense ci-dessous d'autres propriétés vérifiées par une fonction de répartition.

Proposition 2.10. *Soit X une variable aléatoire discrète de $(\Omega, \mathcal{F}, \mathbb{P})$. Sa fonction de répartition F vérifient les propriétés suivantes :*

1. *pour tout $x \in \mathbb{R}$, $F(x) \in [0,1]$,*
2. *F est croissante,*
3. $\lim_{x\to-\infty} F(x) = 0$ *et* $\lim_{x\to+\infty} F(x) = 1$,
4. *F est continue à droite en tout point,*
5. *pour tout $x \in \mathbb{R}$, on a $\mathbb{P}(X = x) = F(x) - F(x-)$, où $F(x-) = \lim_{\epsilon\to 0+} F(x-\epsilon)$.*

Démonstration.

1. Ce point est clair car $F(x) = \mathbb{P}(X \le x)$ est une probabilité.
2. Si $x \le y$, alors $\{X \le x\} \subset \{X \le y\}$ et la monotonie de $\mathbb{P}$ permet de conclure.
3. Traitons par exemple le cas $+\infty$. La fonction F étant croissante et majorée par 1, elle admet une limite finie en $+\infty$ qui vérifie

   $$\lim_{x\to+\infty} F(x) = \lim_{n\to+\infty} F(n) \,.$$

   On définit alors la suite croissante d'événements $A_n = \{X \le n\}$ et on applique la propriété de croissance monotone, ce qui donne

   $$\lim_{n\to+\infty} F(n) = \lim_{n\to+\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n=0}^{+\infty} A_n \right) = \mathbb{P}(\Omega) = 1 \,.$$

4. Soit $x \in \mathbb{R}$. Montrons que F est continue à droite en x. Cela revient à montrer que pour toute suite $(x_n)_{n\in\mathbb{N}}$ décroissante de limite x, on a

   $$\lim_{n\to+\infty} F(x_n) = F(x) \,.$$

   On considère alors la suite décroissante d'événements $B_n = \{X \le x_n\}$ et on applique la propriété de décroissance monotone :

   $$\lim_{n\to+\infty} F(x_n) = \lim_{n\to+\infty} \mathbb{P}(X \le x_n) = \mathbb{P}\left(\bigcap_{n=0}^{+\infty} B_n \right) = \mathbb{P}(X \le x) = F(x) \,.$$

5. Pour $x \in \mathbb{R}$, on considère les ensembles $C_n = \{x - y_n < X \leq x\}$, où (y_n) est une suite décroissante convergeant vers 0. Les ensembles C_n sont décroissants d'intersection

$$\bigcap_{n=0}^{+\infty} C_n = \bigcap_{n=0}^{+\infty} \{x - y_n < X \leq x\} = \{X = x\}\,.$$

La propriété de de décroissance monotone et l'égalité $\mathbb{P}(C_n) = F(x) - F(x - y_n)$ assurent alors que

$$F(x) - F(x-) = \lim_{n\to+\infty} (F(x) - F(x - y_n)) = \lim_{n\to+\infty} \mathbb{P}(C_n) = \mathbb{P}\left(\bigcap_{n=0}^{+\infty} C_n\right) = \mathbb{P}(X = x)\,.$$

$\square$

La dernière propriété montre que les endroits où F présente des sauts correspondent aux points $x_k \in \mathcal{X}$ où X a des chances de tomber. Par ailleurs, la hauteur associée de ce saut correspond à la probabilité p_k de tomber sur x_k.

Ce dernier point donc que si on connaît la fonction de répartition F de X, alors on connaît les probabilités $\mathbb{P}(X = x_k)$, où x_k désigne les valeurs que prend X. Ainsi, on connaît complètement la loi de X. On obtient ainsi le résultat suivant.

Proposition 2.11. *La fonction de répartition d'une variable aléatoire discrète caractérise sa loi : deux variables aléatoires discrètes ont même loi si et seulement si elles ont même fonction de répartition.*

2 Moments d'une variable aléatoire discrète

On considère une variable aléatoire discrète X définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $\mathcal{X} = (x_k)_{k\in K}$ de loi $(p_k)_{k\in K}$.

2.1 Espérance

L'espérance correspond à la notion de "moyenne", mais celle-ci est définie a priori, c'est-à-dire que l'on peut calculer l'espérance d'une variable aléatoire sans avoir réalisé aucune expérience.

Avant de donner la définition de l'espérance, rappelons qu'une série $\sum u_n$ converge absolument si la série $\sum |u_n|$ converge. Ceci assure alors la convergence de la série $\sum u_n$.

Définition 2.12 (Espérance d'une variable aléatoire discrète)**.** *On dit que la variable aléatoire X admet une* espérance *si la série $\sum_{k\in K} x_k p_k$ est absolument convergente. On définit alors son* espérance *par*

$$\mathbb{E}[X] = \sum_{k\in K} x_k p_k = \sum_{k\in K} x_k \mathbb{P}(X = x_k)\,.$$

Si l'ensemble $\mathcal{X}$ est fini, alors l'espérance de X est toujours bien définie. De plus, si $\mathcal{X}$ est inclus dans $\mathbb{R}_+$ ou $\mathbb{R}_-$, alors la preuve de l'existence de l'espérance se fait simultanément avec son calcul. Les cas problématiques correspondent au cas où $\mathcal{X}$ est infini avec une infinité de valeurs positives et une infinité de valeurs négatives.

Exemples 2.13.

1. Si X est une variable aléatoire qui ne prend qu'une seule valeur $a \in \mathbb{R}$ avec probabilité 1, autrement dit $X = a$, alors $\mathbb{E}[X] = a \times 1 = a$.
2. *Lancer de deux pièces.* On considère la variable aléatoire X qui compte le nombre de Pile lors du lancer de deux pièces. Son espérance vaut
$$\mathbb{E}[X] = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1\,.$$
C'est assez intuitif : si on lance deux pièces, on obtient en moyenne un Pile.
3. *Premier 6.* On considère la variable aléatoire discrète X qui donne le premier 6 lors d'une succession de lancers de dé. L'espérance de X est donnée par la formule
$$\mathbb{E}[X] = \frac{1}{6} \sum_{k=1}^{+\infty} k \left(\frac{5}{6}\right)^{k-1}.$$
On reconnaît la dérivée de la fonction $x \mapsto \sum_{k=0}^{+\infty} x^k = \frac{1}{1-x}$ évaluée en $\frac{5}{6}$. Cette dérivée vaut
$$\sum_{k=1}^{+\infty} k x^{k-1} = \frac{1}{(1-x)^2}\,.$$
On en déduit alors l'espérance de X :
$$\mathbb{E}[X] = \frac{1}{6} \frac{1}{\left(1 - \frac{5}{6}\right)^2} = 6\,.$$

Attention. On prendra garde à bien vérifier l'hypothèse d'absolue convergence dans le cas où X n'est pas une variable aléatoire finie. On donne ci-dessous deux exemples pathologiques.

Exemples 2.14.

1. On considère la variable aléatoire X à valeurs dans $\mathcal{X} = \mathbb{N}^*$ dont la loi est donnée par
$$\forall k \geq 1, \quad p_k = \mathbb{P}(X = k) = \frac{1}{k(k+1)}\,.$$
Les p_k définissent bien une loi puisque
$$\sum_{n=1}^{+\infty} p_n = \sum_{n=1}^{+\infty} \frac{1}{n(n+1)} = \sum_{n=1}^{+\infty} \left(\frac{1}{n} - \frac{1}{n+1}\right) = 1\,.$$

Par ailleurs, on remarque que

$$\sum_{k=1}^{n} |k| p_k = \sum_{k=1}^{n} k p_k = \sum_{k=1}^{n} \frac{1}{k+1} .$$

On reconnaît la série harmonique qui diverge. La variable aléatoire X n'admet donc pas d'espérance.

2. Voici un autre exemple encore plus pathologique. On considère une variable aléatoire X prenant les valeurs $x_n = (-1)^n \, n$ pour tout $n \geq 1$. En d'autres termes, $\mathcal{X} = \{-1, 2, -3, 4, \ldots\}$. On suppose que la variable aléatoire X suit la loi donnée par $p_n = \frac{1}{n(n+1)}$, pour tout $n \geq 1$. La série de terme général $x_n p_n$ est alors égale à

$$x_n p_n = \frac{(-1)^n}{n+1} .$$

On reconnaît le terme général d'une série alternée. La série $\sum x_n p_n$ est donc convergente mais pas absolument convergente : on dit qu'elle est semi-convergente. En particulier, la variable aléatoire X n'admet pas d'espérance.

Dans de nombreux cas, on est amené à devoir calculer non pas l'espérance de X mais d'une fonction g de X. Le théorème suivant permet d'effectuer ce calcul sans connaître la loi de $g(X)$.

Théorème 2.15 (Théorème de transfert)**.** *Soit g une fonction de $\mathbb{R}$ dans $\mathbb{R}$ et X une variable aléatoire discrète sur $\mathcal{X} = (x_k)_{k \in K}$ de loi $(p_k)_{k \in K}$. La variable aléatoire discrète $Y = g(X)$ admet une espérance si la série $\sum_{k \in K} |g(x_k)| p_k$ converge et alors :*

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \sum_{k \in K} g(x_k) p_k .$$

On retiendra que pour calculer l'espérance de la variable aléatoire $Y = g(X)$ il est inutile de déterminer sa loi. Il suffit de remplacer les x_k par $g(x_k)$.

Démonstration. On considère l'ensemble $\mathcal{Y} = \{y_j, j \in J\} = \{g(x_k), k \in K\}$ des valeurs prises par Y. L'ensemble K peut alors être partitionné en fonction de l'image par g de chaque élément x_k via

$$E_j = \{k \in K : g(x_k) = y_j\} .$$

On définit également $q_j = \sum_{k \in E_j} p_k$ de sorte que la loi de Y soit caractérisée par les q_j. On obtient alors

$$\mathbb{E}[Y] = \sum_{j \in J} y_j q_j = \sum_{j \in J} y_j \left(\sum_{k \in E_j} p_k \right) = \sum_{j \in J} \left(\sum_{k \in E_j} g(x_k) p_k \right) .$$

Cette double somme correspond en fait à une somme sur K puisque les ensembles E_j forment une partition de cet ensemble. On en déduit alors

$$\mathbb{E}[Y] = \sum_{k \in K} g(x_k) p_k ,$$

et tous les calculs faisant intervenir des sommes sont justifiés par l'absolue convergence de la série $\sum_{k\in K} g(x_k)p_k$ $\square$

Exemple 2.16. On reprend l'exemple du lancer de deux pièces pour lequel on souhaite calculer l'espérance de X^2. Une application directe du théorème de transfert donne

$$\mathbb{E}[X^2] = 0^2 \times \frac{1}{4} + 1^2 \times \frac{1}{2} + 2^2 \times \frac{1}{4} = \frac{3}{2}\,.$$

On peut également obtenir ce résultat en déterminant la loi de X^2. Cette variable aléatoire prend ses valeurs dans l'ensemble $\{0,1,4\}$ et sa loi est donnée par les probabilités $\mathbb{P}(X^2=0) = \mathbb{P}(X=0) = \frac{1}{4}$, $\mathbb{P}(X^2=1) = \mathbb{P}(X=1) = \frac{1}{2}$, $\mathbb{P}(X^2=4) = \mathbb{P}(X=2) = \frac{1}{4}$. On en déduit alors l'espérance de X^2 via le calcul

$$\mathbb{E}[X^2] = 0 \times \frac{1}{4} + 1^2 \times \frac{1}{2} + 4 \times \frac{1}{4} = \frac{3}{2}\,.$$

Exercice 2.17. Calculer la valeur de $\mathbb{E}[\mathrm{e}^X]$ où X est la variable aléatoire donnée dans l'exemple précédent.

Propriétés 2.18. *Soit X et Y deux variables aléatoires admettant une espérance, et a et b deux réels.*

1. *(Linéarité). La variable aléatoire $aX+bY$ admet une espérance donnée par*

$$\mathbb{E}[aX+bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]\,.$$

2. *(Positivité). Si $X \geq 0$, alors $\mathbb{E}[X] \geq 0$*
3. *(Croissance). Si $X \leq Y$, alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$.*

Si X a une espérance nulle, on dit alors qu'elle est *centrée*. La propriété de linéarité assure que si X a pour espérance μ, alors $X-\mu$ est une variable aléatoire centrée.

Démonstration. 1. On traite tout d'abord de la multiplication par un scalaire. Si X prend les valeurs $\mathcal{X} = (x_i)_{i\in I}$ avec les probabilités $(p_i)_{i\in I}$, alors la formule de transfert avec $g(x) = ax$ donne

$$\mathbb{E}[aX] = \sum_{i\in I} a x_i p_i = a \sum_{i\in I} x_i p_i = a\mathbb{E}[X]\,.$$

Il reste à prouver que $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$. On suppose que X prend les valeurs $\mathcal{X} = (x_i)_{i\in I}$ avec les probabilités $(p_i)_{i\in I}$, Y prend les valeurs $\mathcal{Y} = (y_j)_{j\in J}$ avec les probabilités $(q_j)_{j\in J}$, et $Z = X+Y$ prend les valeurs $\mathcal{Z} = (z_k)_{k\in K}$ avec les probabilités $(r_k)_{k\in K}$. L'espérance de Z vaut alors

$$\begin{aligned}\mathbb{E}[Z] = \sum_{k\in K} z_k r_k &= \sum_{k\in K} z_k \Bigg(\sum_{(i,j):x_i+y_j=z_k} \mathbb{P}(X=x_i, Y=y_j) \Bigg) \\ &= \sum_{k\in K} \sum_{(i,j):x_i+y_j=z_k} z_k \mathbb{P}(X=x_i, Y=y_j)\,.\end{aligned}$$

Puisque tout z_k est de la forme $x_i + y_j$, on obtient

$$\begin{aligned}\mathbb{E}[Z] &= \sum_{(i,j)\in I\times J} (x_i + y_j)\mathbb{P}(X = x_i, Y = y_j) \\ &= \sum_{i\in I} x_i\Big(\sum_{j\in J} \mathbb{P}(X = x_i, Y = y_j)\Big) + \sum_{j\in J} y_j\Big(\sum_{i\in I} \mathbb{P}(X = x_i, Y = y_j)\Big)\end{aligned}$$

La première somme entre parenthèses vaut $\mathbb{P}(X = x_i)$ tandis que la deuxième vaut $\mathbb{P}(Y = y_j)$ (on applique la formule des probabilités totales). On conclut alors

$$\mathbb{E}[Z] = \sum_{i\in I} x_i\mathbb{P}(X = x_i) + \sum_{j\in J} y_j\mathbb{P}(Y = y_j) = \mathbb{E}[X] + \mathbb{E}[Y]\,.$$

2. L'inégalité $X \geq 0$ signifie que X ne prend que des valeurs positives : $\mathcal{X} \subset \mathbb{R}_+$. La définition de l'espérance assure alors que $\mathbb{E}[X] \geq 0$.
3. On se ramène au cas précédent en considérant la variable aléatoire $Z = Y - X$ qui est positive. Ainsi, $\mathbb{E}[Z] = \mathbb{E}[Y - X] \geq 0$, et on utilise la linéarité de l'espérance pour conclure : $\mathbb{E}[Y] - \mathbb{E}[X] \;\geq 0$.

□

Une récurrence immédiate donne la relation

$$\mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]\,,$$

pour toute suite $X_1, \ldots, X_n$ de variables aléatoires discrètes admettant une espérance.

Exemple 2.19. On lance un dé n fois et on désigne par X_i la variable aléatoire donnant le chiffre qui apparaît lors du i-ème lancer. Pour i fixé, la loi de X_i est donnée par

$$p_k = \mathbb{P}(X_i = k) = \frac{1}{6}\,, \quad k = 1, \ldots, 6\,,$$

de sorte que l'espérance de X_i est donnée par

$$\mathbb{E}[X_i] = \sum_{k=1}^{6} k \times \frac{1}{6} = \frac{7}{2}\,.$$

La linéarité de l'espérance donne alors l'espérance de la somme des n lancers de dés :

$$\mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = \frac{7n}{2}\,.$$

2.2 Variance

On souhaite désormais mesurer la dispersion d'une variable aléatoire autour de son espérance.

Définition 2.20 (Variance)**.** *Soit X une variable aléatoire admettant une espérance $\mathbb{E}[X]$. On dit que X admet une* variance *si la série*

$$\sum_{k \in K} (x_k - \mathbb{E}[X])^2 p_k$$

converge, et on définit dans ce cas la variance de X par la quantité

$$\mathrm{Var}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] = \sum_{k \in K} (x_k - \mathbb{E}[X])^2 p_k \,.$$

On définit l'écart-type *de X par*

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

On remarque que la variance est une quantité positive. Par ailleurs, comme pour l'espérance, si X prend un nombre fini de valeurs alors sa variance est toujours bien définie.

Exemple 2.21. *Lancer de deux pièces.* L'espérance de X vaut 1. On obtient alors

$$\mathrm{Var}(X) = (0-1)^2 \times \frac{1}{4} + (1-1)^2 \times \frac{1}{2} + (2-1)^2 \times \frac{1}{4} = \frac{1}{2}\,.$$

La formule suivante est très pratique pour calculer la variance d'une variable aléatoire. Elle est parfois appelée formule de Koenig-Huygens.

Proposition 2.22 (Formule de Koenig-Huygens)**.** *Soit X une variable aléatoire admettant une variance. Alors*

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

Démonstration. Notons μ l'espérance de X. On développe le carré qui apparaît dans la définition de la variance et on applique la linéarité de l'espérance :

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

□

Cela signifie que si X est une variable aléatoire discrète admettant une variance, alors

$$\mathrm{Var}(X) = \sum_{k \in K} x_k^2 p_k - \Big(\sum_{k \in K} x_k p_k\Big)^2\,.$$

En pratique, si on doit étudier les propriétés d'une variable aléatoire X, on commencera quasiment toujours par calculer son espérance. Pour déterminer la variance de X, il suffira alors de calculer $\mathbb{E}[X^2]$ et d'utiliser la formule précédente. Le calcul de $\mathbb{E}[X^2]$ s'effectue via le théorème de transfert (inutile de déterminer la loi de X^2).

Exemple 2.23. *Premier 6.* On rappelle que l'espérance de X vaut 6. Pour appliquer la formule de Koenig-Huygens, il s'agit de calculer l'espérance de $\mathbb{E}[X^2]$ dont l'expression est donnée par la formule de transfert :

$$\mathbb{E}[X^2] = \sum_{k=1}^{+\infty} k^2 \left(\frac{5}{6}\right)^{k-1} \frac{1}{6}\,.$$

On utilise alors l'astuce qui consiste à écrire $k^2 = k(k-1) + k$:

$$\mathbb{E}[X^2] = \sum_{k=1}^{+\infty} k(k-1) \left(\frac{5}{6}\right)^{k-1} \frac{1}{6} + \sum_{k=1}^{+\infty} k \left(\frac{5}{6}\right)^{k-1} \frac{1}{6}\,.$$

La deuxième somme correspond à l'espérance de X et vaut donc 6. Pour la première, on reconnaît la dérivée seconde de la fonction

$$\phi : x \mapsto \sum_{k=0}^{+\infty} x^k = \frac{1}{1-x}$$

évaluée en 5/6. On trouve $\phi''(x) = \frac{2}{(1-x)^3}$, puis $\phi''(5/6) = 432$, ce qui donne

$$\mathbb{E}[X^2] = \frac{5}{6} \times \frac{1}{6} \times \phi''(5/6) + 6 = 66\,.$$

On obtient alors

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 66 - 36 = 30\,.$$

L'écart-type vaut alors $\sigma(X) = \sqrt{30} \approx 5.48$.

Propriétés 2.24. *Soit X une variable aléatoire admettant une variance, et a et b deux réels.*

1. $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$,
2. $\mathrm{Var}(X) = 0$ *si et seulement si X est constante.*

Démonstration. 1. On part de la formule de Koenig-Huygens

$$\mathrm{Var}(aX + b) = \mathbb{E}[(aX + b)^2] - \mathbb{E}[aX + b]^2\,.$$

La linéarité de l'espérance entraîne que

$$\mathbb{E}[(aX + b)^2] = \mathbb{E}[a^2X^2 + 2abX + b^2] = a^2\mathbb{E}[X^2] + 2ab\mathbb{E}[X] + b^2\,,$$

et

$$\mathbb{E}[aX + b]^2 = (a\mathbb{E}[X] + b)^2 = a^2\mathbb{E}[X]^2 + 2ab\mathbb{E}[X] + b^2\,.$$

La différence de ces deux quantités donne alors

$$\mathrm{Var}(aX + b) = a^2(\mathbb{E}[X^2] - \mathbb{E}[X]^2) = a^2\,\mathrm{Var}(X)\,,$$

où l'on a une nouvelle fois utilisé la formule de Koenig-Huygens pour X.

2. Si $\mathrm{Var}(X) = \sum_{k\in K}(x_k - \mathbb{E}[X])^2 p_k = 0$, cela implique que pour tout $k \in K$, on a $x_k = \mathbb{E}[X]$ (les probabilités p_k sont supposées strictement positives). Ainsi, la seule valeur que peut prendre X est $\mathbb{E}[X]$, ce qui signifie que cette variable aléatoire est constante. Réciproquement, si X est constante égale à a, alors son espérance vaut a et sa variance est nulle :
$$\mathrm{Var}(X) = (a-a)^2 \times 1 = 0\,.$$
□

On dit qu'une variable aléatoire X est *réduite* si $\mathrm{Var}(X) = 1$. La propriété précédente implique que si X admet pour variance σ^2, alors X/σ est une variable aléatoire réduite. Plus généralement, si X est une variable aléatoire telle que $\mathbb{E}[X] = \mu$ et $\mathrm{Var}(X) = \sigma^2$, alors $(X-\mu)/\sigma$ est une *variable aléatoire centrée réduite.*

Attention. Contrairement à l'espérance, la variance n'est en général pas linéaire, c'est-à-dire que si X et Y sont deux variables aléatoires, $\mathrm{Var}(X+Y) \neq \mathrm{Var}(X)+\mathrm{Var}(Y)$. On verra plus loin que cette propriété est toutefois vraie dans le cas de deux variables aléatoires indépendantes.

La variance permet de mesurer l'écart d'une variable aléatoire à son espérance comme l'illustre l'inégalité suivante, appelée inégalité de Bienaymé-Tchebychev.

Proposition 2.25 (Inégalité de Bienaymé-Tchebychev)**.** *Soit X une variable aléatoire discrète admettant une variance. Alors,*
$$\forall t > 0\,, \quad \mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}\,.$$

La probabilité que X s'écarte de sa moyenne d'une distance t est donc contrôlée par sa variance et une décroissance en t^2.

Démonstration. On note x_k les valeurs prises par X et p_k les probabilités associées. On note également μ l'espérance de X. Pour $t > 0$, on définit l'ensemble A_t par
$$A_t = \{x \in \mathbb{R} : |x_k - \mu| \geq t\}\,.$$
On obtient alors l'égalité suivante :
$$\mathbb{P}(|X-\mu| \geq t) = \mathbb{P}(X \in A_t) = \sum_{k:x_k\in A_t} p_k\,.$$
Il suffit alors d'écrire
$$\mathrm{Var}(X) = \sum_{k\in K}(x_k-\mu)^2 p_k \geq \sum_{k:x_k\in A_t}(x_k-\mu)^2 p_k \geq \sum_{k:x_k\in A_t} t^2 p_k = t^2\mathbb{P}(|X-\mu| \geq t)\,,$$
ce qui conclut la preuve. □

Cette inégalité permet de contrôler la probabilité qu'un phénomène aléatoire s'écarte de sa moyenne, c'est-à-dire d'un comportement standard. On dit que c'est une inégalité de *concentration.* Bien que l'inégalité de Bienaymé-Tchebychev soit très générale, elle donne des bornes assez mauvaises en pratique.

2.3 Moments

On étend à présent les notions d'espérance et de variance.

Définition 2.26 (Moment)**.** *Soit X une variable aléatoire discrète et $m \in \mathbb{N}^*$. On appelle*

1. moment d'ordre m *de X l'espérance de X^m :*

$$\mathbb{E}[X^m] = \sum_{k \in K} x^m p_k\,,$$

2. moment centré d'ordre m *de X l'espérance de $(X - \mathbb{E}[X])^m$:*

$$\mathbb{E}[(X - \mathbb{E}[X])^m] = \sum_{k \in K} (x_k - \mathbb{E}[X])^m p_k\,,$$

sous réserve que ces quantités sont bien définies (absolue convergence des séries considérées).

Remarques 2.27.

1. Si X prend un nombre fini de valeurs, alors X admet des moments (et des moments centrés) de tout d'ordre.
2. Le moment d'ordre 1 n'est rien d'autre que l'espérance. Le moment centré d'ordre 2 correspond à la variance.

Exemple 2.28. *Lancer de deux pièces.* Pour $m \in \mathbb{N}^*$, le moment d'ordre m vaut

$$\mathbb{E}[X^m] = 0^m \times \frac{1}{4} + 1^m \times \frac{1}{2} + 2^m \times \frac{1}{4} = \frac{1}{2} + 2^{m-2}\,.$$

Exercice 2.29. Calculer les moments centrés dans l'exemple précédent.

Proposition 2.30. *Soit X une variable aléatoire discrète. Si X admet un moment d'ordre $m \in \mathbb{N}^*$, alors X admet des moments d'ordre q pour tout $q \in \{1, \dots, m\}$.*

Démonstration. Il s'agit de montrer que la série $\sum_{k \in K} |x|^q p_k$ converge. On souhaite bien évidemment utiliser l'inégalité $|u|^q \leq |u|^m$, mais celle-ci n'est valable que pour $|u| \geq 1$. On pose alors

$$A = \{x \in \mathbb{R} : |x_k| \geq 1\}\,,$$

de sorte que pour tout $k \in A$, on a $|x_k|^q \leq |x_k|^m$. Il s'ensuit :

$$\sum_{k \in K} |x|^q p_k = \sum_{k : x_k \in A} |x|^q p_k + \sum_{k : x_k \notin A} |x|^q p_k \leq \sum_{k : x_k \in A} |x|^m p_k + \sum_{k : x_k \notin A} p_k \leq \sum_{k \in K} |x|^m p_k + 1\,.$$

Cette dernière quantité correspond à $\mathbb{E}[|X|^m] + 1$ qui est finie par hypothèse. Ceci prouve la convergence de $\sum_{k \in K} |x|^q p_k$, donc l'existence du moment d'ordre q. □

On généralise à présent l'inégalité de Bienaymé-Tchebychev.

Proposition 2.31 (Inégalité de Markov)**.** *Soit X une variable aléatoire discrète. Si X admet un moment d'ordre m, alors*

$$\forall t > 0\,, \quad \mathbb{P}(|X| \geq t) \leq \frac{\mathbb{E}[|X|^m]}{t^m}\,.$$

Cette inégalité permet de contrôler les queues de distribution d'une variable aléatoire X, c'est-à-dire les quantités $\mathbb{P}(X \geq t)$ et $\mathbb{P}(X \leq -t)$, pour $t > 0$. Sa preuve est similaire à celle de l'inégalité de Bienaymé-Tchebychev. On remarque qu'on retrouve bien cette dernière en considérant la variable aléatoire $X - \mathbb{E}[X]$ et $m = 2$.

3 Indépendance et corrélation

Cette section est consacrée aux relations qui peuvent exister entre des variables aléatoires. La première notion développée est celle d'indépendance. On présente ensuite une notion plus souple, la corrélation, qui mesure le lien linéaire qui peut exister entre deux variables aléatoires.

3.1 Indépendance

Définition 2.32 (Variables aléatoires indépendantes)**.** *Deux variables aléatoires discrètes X et Y à valeurs dans $\mathcal{X} = (x_i)_{i\in I}$ et $\mathcal{Y} = (y_j)_{j\in J}$ sont dites* indépendantes *si pour tout couple $(i, j) \in I \times J$ elles vérifient*

$$\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j)\,.$$

Cela signifie que les événements $\{X = x_i\}$ et $\{Y = y_j\}$ sont indépendants au sens du chapitre précédent : la réalisation de l'événement $\{X = x_i\}$ n'influence pas celle de l'événement $\{Y = y_j\}$. Comme pour l'indépendance d'événements, l'indépendance de variables aléatoires est fastidieuse à vérifier. Dans la majorité des cas, c'est le contexte qui permettra de considérer les variables aléatoires comme indépendantes ou non. Par exemple si on lance deux dés, le résultat X donné par l'un est indépendant du résultat Y donné par l'autre.

On peut généraliser la définition précédente. Soit n variables aléatoires $X_1, \dots, X_n$ prenant leurs valeurs dans $\mathcal{X}_1, \dots, \mathcal{X}_n$. Elles sont indépendantes si pour tout $(x_1, \dots, x_n) \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$,

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n)\,.$$

L'indépendance est une hypothèse forte, que l'on souhaite assouplir. L'idée de la covariance est de tenir uniquement compte des relations linéaires entre deux variables aléatoires.

3.2 Covariance

Définition 2.33 (Covariance)**.** *Soient X et Y deux variables aléatoires discrètes admettant des moments d'ordre 2. On définit la* covariance de X et de Y *par*

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

La deuxième égalité dans la définition de la covariance résulte de la linéarité de l'espérance. En effet, on développe

$$(X-\mathbb{E}[X])(Y-\mathbb{E}[Y]) = XY - \mathbb{E}[Y]X - \mathbb{E}[X]Y + \mathbb{E}[X]\mathbb{E}[Y]$$

et on applique l'espérance de chaque côté de l'égalité ce qui donne bien le résultat attendu.

Remarque 2.34. L'inégalité classique

$$|XY| \leq \frac{1}{2}(X^2+Y^2) \,.$$

et l'existence des moments d'ordre 2 de X et Y impliquent que la quantité $\mathbb{E}[XY]$, et donc la covariance, est bien définie.

Exemple 2.35. On considère une variable aléatoire discrète X à valeurs dans $\mathcal{X} = \{-1,0,1\}$ de loi donnée par $\mathbb{P}(X=-1) = \mathbb{P}(X=1) = 1/4$ et $\mathbb{P}(X=0) = 1/2$. On pose $Y = X^2$. Alors,

$$\mathbb{E}[X] = -1 \times \frac{1}{4} + 0 \times \frac{1}{2} + 1 \times \frac{1}{4} = 0 \,.$$

et

$$\mathbb{E}[XY] = \mathbb{E}[X^3] = (-1)^3 \times \frac{1}{4} + 0^3 \times \frac{1}{2} + 1^3 \times \frac{1}{4} = 0 \,.$$

La covariance des variables aléatoires X et Y vaut donc $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$.

Exercice 2.36. On dispose d'une pièce équilibrée dont le côté Pile est numéroté 1 et le côté Face 0. On lance deux fois la pièce et on désigne par X (resp. Y) la variable aléatoire donnant la valeur maximale (resp. minimale) obtenue au cours des deux lancers. Déterminer la loi de X et de Y. Calculer $E[X]$, $E[Y]$, puis $\mathrm{Cov}(X,Y)$.

Propriétés 2.37. *Soit X et Y deux variables aléatoires discrètes admettant un moment d'ordre 2.*

1. $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$.
2. *(Symétrie)* $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$.
3. *(Bilinéarité) Soit Z une variable aléatoire discrète admettant un moment d'ordre deux et a, b deux réels. Alors*

$$\mathrm{Cov}(aX+bZ,Y) = a\,\mathrm{Cov}(X,Y) + b\,\mathrm{Cov}(Z,Y) \,,$$

et

$$\mathrm{Cov}(X,aY+bZ) = a\,\mathrm{Cov}(X,Y) + b\,\mathrm{Cov}(X,Z) \,.$$

4. $\operatorname{Var}(X+Y) = \operatorname{Var}(X) + 2\operatorname{Cov}(X,Y) + \operatorname{Var}(Y)$.

Démonstration. Les deux premiers points découlent de la définition de la covariance. Concernant le troisième, la linéarité de l'espérance permet d'écrire

$$\mathbb{E}[(aX+bZ)Y] = a\mathbb{E}[XY] + b\mathbb{E}[ZY]\,,$$

et

$$\mathbb{E}[aX+bZ]\mathbb{E}[Y] = a\mathbb{E}[X]\mathbb{E}[Y] + b\mathbb{E}[Z]\mathbb{E}[Y]\,.$$

La différence terme à terme de ces deux égalités donne alors

$$\mathbb{E}[(aX+bZ)Y] - \mathbb{E}[aX+bZ]\mathbb{E}[Y] = a(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) + b(\mathbb{E}[ZY] - \mathbb{E}[Z]\mathbb{E}[Y])\,,$$

ou encore par la deuxième expression de la covariance :

$$\operatorname{Cov}(aX+bZ,Y) = a\operatorname{Cov}(X,Y) + b\operatorname{Cov}(Z,Y)\,.$$

La linéarité par rapport à la deuxième composante se montre de la même manière.

Enfin, la dernière propriété découle des égalités suivantes :

$$\operatorname{Var}(X+Y) = \mathbb{E}[(X+Y)^2] - (\mathbb{E}[X+Y])^2 = \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - 2\mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]^2\,.$$

En regroupant les termes en X, en Y, et en XY respectivement, on obtient bien

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + 2\operatorname{Cov}(X,Y) + \operatorname{Var}(Y)\,.$$

□

La troisième propriété est appelée "bilinéarité" car elle assure que la covariance est linéaire en chacune de ses composantes.

La dernière propriété assure que si $\operatorname{Cov}(X,Y) = 0$, alors $\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$. Cependant, on se gardera bien d'écrire cette inégalité sans hypothèse sur la covariance de X et Y.

3.3 Coefficient de corrélation

Définition 2.38 (Coefficient de corrélation, variables aléatoires décorrélées)**.** *Soit X et Y deux variables aléatoires discrètes admettant des variances non nulles. Le* coefficient de corrélation *entre X et Y est défini par :*

$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sigma(X)\sigma(Y)} = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}}\,.$$

On dit que deux variables aléatoires sont décorrélées *si $\rho(X,Y) = \operatorname{Cov}(X,Y) = 0$. Dans ce cas,*

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)\,.$$

Exemple 2.39. Les variables aléatoires X et Y de l'exemple 2.35 sont décorrélées.

Le coefficient de corrélation permet de mesurer la "degré de linéarité" entre deux variables aléatoires aléatoires X et Y. C'est ce qu'énonce le résultat suivant.

Proposition 2.40 (Inégalité de Cauchy-Schwarz)**.** *Soit X et Y deux variables aléatoires discrètes admettant des moments d'ordre 2. Alors*

$$-1 \leq \rho(X, Y) \leq 1 .$$

Plus précisément :

1. $\rho(X, Y) = 1$ *si et seulement si il existe* $a > 0$ *et* $b \in \mathbb{R}$ *tels que* $Y = aX + b$.
2. $\rho(X, Y) = -1$ *si et seulement si il existe* $a > 0$ *et* $b \in \mathbb{R}$ *tels que* $Y = -aX + b$ *;*

Démonstration. L'astuce consiste à considérer la variable aléatoire $tX+Y$ pour $t \in \mathbb{R}$. Sa variance est donnée par

$$\operatorname{Var}(tX + Y) = t^2 \operatorname{Var}(X) + 2t \operatorname{Cov}(X, Y) + \operatorname{Var}(Y) .$$

Cette quantité est positive pour tout $t \in \mathbb{R}$ (c'est une variance), donc le trinôme en t a un discriminant négatif, c'est à dire

$$4 \operatorname{Cov}(X, Y)^2 - 4 \operatorname{Var}(X) \operatorname{Var}(Y) \leq 0 ,$$

ce qui revient à $\rho(X, Y) \in [-1, 1]$.

Si $\rho(X, Y) = 1$, alors le discriminant précédent est nul donc le trinôme admet une racine : il existe $t_0 \in \mathbb{R}$ tel que $\operatorname{Var}(t_0 X + Y) = 0$. Ceci implique qu'il existe $b \in \mathbb{R}$ tel que $aX + Y = b$, c'est-à-dire $Y = -t_0 X + b$. En particulier l'égalité $\rho(X, Y) = 1$ donne

$$1 = \rho(X, Y) = \frac{\operatorname{Cov}(X, -t_0 X + b)}{\sigma(X)\sigma(-t_0 X + b)} = \frac{-t_0 \operatorname{Var}(X) + \operatorname{Cov}(X, b)}{\sigma(X)|t_0|\sigma(X)} = \frac{-t_0}{|t_0|} ,$$

ce qui prouve que $-t_0 > 0$. Si $Y = aX + b$ avec $a > 0$, un calcul direct montre que $\rho(X, Y) = 1$. Enfin, l'équivalence dans le cas $\rho(X, Y) = -1$ se prouve de manière analogue. □

On remarque que les propriétés de la covariance sont semblables à celles du produit scalaire. La propriété précédente correspond par exemple à l'inégalité de Cauchy-Schwarz vue dans un espace euclidien. Le coefficient de corrélation entre deux variables aléatoires est quant à lui équivalent au cosinus de l'angle entre deux vecteurs.

Remarque 2.41. Plus le coefficient de corrélation est proche de 1 en valeur absolue, plus les variables X et Y sont linéairement liées. Un coefficient de corrélation nul signifie donc que les deux variables ne sont pas linéairement liées. Mais elles peuvent être liées par un autre type de relation. C'est par exemple le cas de l'exemple 2.35 où $Y = X^2$. Dans cet exemple, si X est connu, alors il n'y a plus aucune incertitude sur Y.

Comme évoqué précédemment, la notion de variables aléatoires décorrélées est plus faible que celle de variables aléatoires indépendantes.

Proposition 2.42. *Soit X et Y des variables aléatoires discrètes admettant des moments d'ordre 2. Si X et Y sont indépendantes, alors elles sont décorrélées.*

Démonstration. On suppose que X prend ses valeurs dans $\mathcal{X} = (x_i)_{i\in I}$ avec probabilité $p_i = \mathbb{P}(X = x_i)$ et Y dans $\mathcal{Y} = (y_j)_{j\in J}$ avec probabilité $q_j = \mathbb{P}(Y = y_j)$. La variable aléatoire XY prend alors ses valeurs dans $\{x_i y_j, (x_i, y_j) \in \mathcal{X} \times \mathcal{Y}\}$ avec probabilité

$$r_{i,j} = \mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j)\,,$$

où la dernière égalité résulte de l'hypothèse d'indépendance.

On en déduit que

$$\mathbb{E}[XY] = \sum_{(i,j)\in I\times J} x_i y_j \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j) = \sum_{i\in I} x_i \mathbb{P}(X = x_i) \sum_{j\in J} y_j \mathbb{P}(Y = y_j) = \mathbb{E}[X]\mathbb{E}[Y]\,.$$

On obtient alors $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$: les variables aléatoires X et Y sont bien décorrélées. □

En particulier, deux variables aléatoires X et Y indépendantes admettant des moments d'ordre 2 vérifient

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)\,.$$

On rappelle que ce résultat est faux dans le cas général ! Cette formule peut s'étendre à plusieurs variables aléatoires indépendantes : si $X_1, \dots, X_n$ sont des variables aléatoires indépendantes admettant des moments d'ordre 2, alors

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)\,.$$

Attention. La réciproque de la proposition est fausse : on peut trouver deux variables aléatoires qui ne sont pas indépendantes mais dont le coefficient de corrélation est nul. C'est cohérent avec ce qui a été raconté jusqu'ici : un coefficient de corrélation nul signifie l'absence de relation linéaire tandis que l'indépendance signifie l'absence de toute relation.

Exemple 2.43. On reprend les variables aléatoires décorrélées de l'exemple 2.35. Au vu de la relation $Y = X^2$, il semble évident que ces deux variables aléatoires ne sont pas indépendantes. Cela se justifie en prenant par exemple l'événement $\{X = 0, Y = 0\}$:

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

4 Lois discrètes usuelles

On considère toujours un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. On donne dans cette section quelques lois discrètes d'usage fréquent.

4.1 Lois à support fini

4.1.1 Loi uniforme sur un ensemble fini

La loi uniforme correspond au cas d'équiprobabilité.

Définition 2.44 (Loi uniforme sur $\{x_1, \ldots, x_n\}$)**.** *On dit qu'une variable aléatoire X suit une* loi uniforme *sur l'ensemble $\mathcal{X} = \{x_1, \ldots, x_n\}$, et on note $X \sim \mathcal{U}(\{x_1, \ldots, x_n\})$, si X prend ses valeurs dans $\mathcal{X}$ avec*

$$\forall k = 1, \ldots, n\,, \quad \mathbb{P}(X = x_k) = \frac{1}{n}\,.$$

Un cas particulier fréquent est celui où $x_k = k$ pour tout $k = 1, \ldots, n$.

Exemple 2.45. On lance un dé équilibré et on note X la variable aléatoire donnant le numéro obtenu. Alors X suit une loi uniforme sur $\{1, \ldots, 6\}$: $X \sim \mathcal{U}(\{1, \ldots, 6\})$.

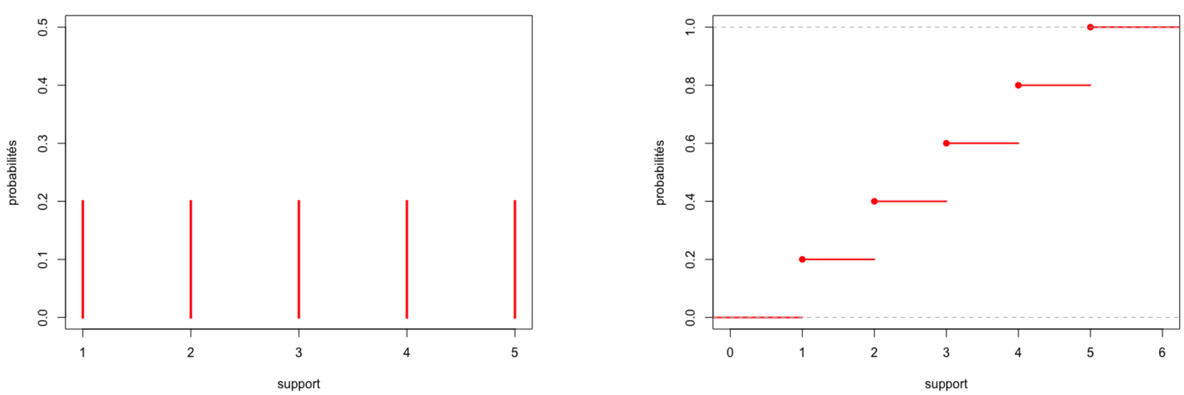


FIGURE 2.4 – Loi uniforme $\mathcal{U}(1, \ldots, 5)$: diagramme en bâtons (gauche) et fonction de répartition (droite).

Proposition 2.46. *Si X suit une loi uniforme sur $\{x_1, \ldots, x_n\}$, alors*

$$\mathbb{E}[X] = \frac{x_1 + \ldots + x_n}{n} \quad \textit{et} \quad \mathrm{Var}(X) = \frac{x_1^2 + \ldots + x_n^2}{n} - \left(\frac{x_1 + \ldots + x_n}{n}\right)^2\,.$$

En particulier, si $X \sim \mathcal{U}(\{1, \ldots, n\})$, alors

$$\mathbb{E}[X] = \frac{n+1}{2} \quad \textit{et} \quad \mathrm{Var}(X) = \frac{n^2 - 1}{12}\,.$$

Démonstration. Le cas général résulte de la définition de l'espérance et de la formule de Koenig-Huygens. Dans le cas où $X \sim \mathcal{U}(\{1,\dots,n\})$, on utilise la formule bien connue

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}\,,$$

et on obtient $\mathbb{E}[X] = \frac{n+1}{2}$. Pour le calcul de la variance, on rappelle la formule

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}\,.$$

On en déduit alors la valeur de $\mathrm{Var}(X)$:

$$\mathrm{Var}(X) = \frac{1}{n}\sum_{k=1}^{n} k^2 - \left(\frac{n+1}{2}\right)^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}\,.$$

□

Exemple 2.47. Dans le cas du lancer d'un dé équilibré, on obtient alors $\mathbb{E}[X] = \frac{7}{2} = 3.5$ (soit le milieu de l'intervalle $[1,6]$) et $\mathrm{Var}(X) = \frac{35}{12}$, puis $\sigma(X) = \sqrt{\frac{35}{12}} \approx 1.7$.

4.1.2 Loi de Bernoulli

La loi de Bernoulli correspond au cas d'une variable aléatoire binaire.

Définition 2.48 (Loi de Bernoulli)**.** *On dit qu'une variable aléatoire X suit une* loi de Bernoulli *de paramètre $p \in]0,1[$, et on note $X \sim \mathcal{B}(p)$, si X prend les valeurs 0 et 1 avec*

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

Une expérience aléatoire n'ayant que deux issues possibles est souvent appelée expérience de Bernoulli. On nomme alors l'une de ces issues "succès" et l'autre "échec", la probabilité du succès étant p, celle de l'échec $q = 1-p$.

Exemple 2.49. On lance une pièce équilibrée et on considère la variable aléatoire X qui vaut 1 si on obtient Pile et 0 si on obtient Face. Alors X suit une loi de Bernoulli de paramètre $1/2$.

Plus généralement, si la probabilité d'apparition de Pile vaut p, alors $X \sim \mathcal{B}(p)$.

Remarques 2.50.

1. La loi de Bernoulli de paramètre $p = 1/2$ correspond à une loi uniforme sur l'ensemble $\{0,1\}$.
2. Une variante de la loi de Bernoulli est la loi de Rademacher. On dit qu'une variable aléatoire X suit une loi de Rademacher si $\mathcal{X} = \{-1,1\}$ et $\mathbb{P}(X=1) = \mathbb{P}(X=-1) = 1/2$. Cela correspond a une loi uniforme sur $\{-1,1\}$.

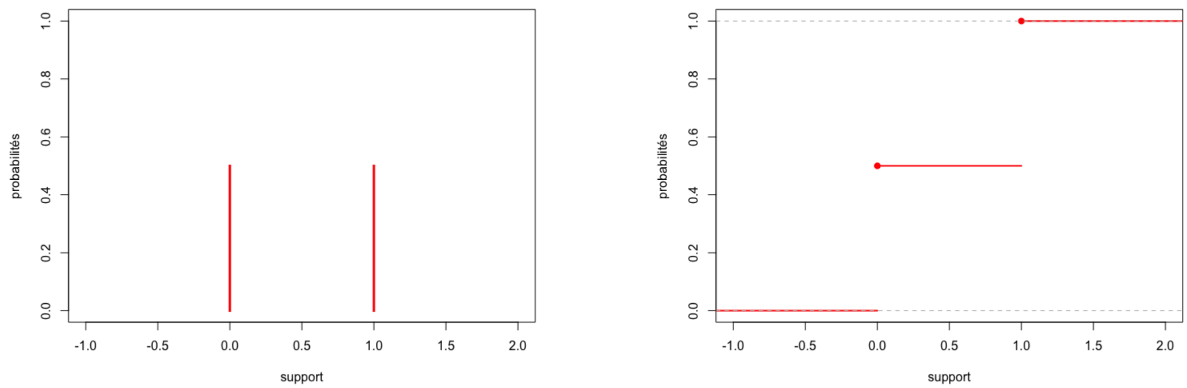


FIGURE 2.5 – Loi de Bernoulli $\mathcal{B}(1/2)$: diagramme en bâtons (gauche) et fonction de répartition (droite).

Proposition 2.51. *Si $X \sim \mathcal{B}(p)$, alors*

$$\mathbb{E}[X] = p \quad \textit{et} \quad \mathrm{Var}(X) = p(1-p)\,.$$

Démonstration. L'espérance et la variance s'obtiennent directement à partir de leur définition :

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

et

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = (0-p)^2 \times (1-p) + (1-p)^2 \times p = p(1-p)\,.$$

□

Par exemple, dans le cas du lancer d'une pièce équilibrée, on obtient $\mathbb{E}[X] = 1/2$ et $\mathrm{Var}(X) = 1/4$.

4.1.3 Loi binomiale

Définition 2.52 (Loi binomiale)**.** *On dit qu'une variable aléatoire X suit une* loi binomiale *de paramètres $n \in \mathbb{N}^*$ et $p \in]0,1[$, et on note $X \sim \mathcal{B}(n,p)$, si X prend ses valeurs dans $\{0,\dots,n\}$ avec*

$$\forall k = 0,\dots,n\,, \quad \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}\,.$$

On vérifie qu'il s'agit bien d'une loi de probabilité grâce à la formule du binôme de Newton :

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1\,.$$

Comme dans le cas d'une loi de Bernoulli, on notera souvent $q = 1 - p$.

Exemple 2.53. On lance 6 fois de suite une pièce déséquilibrée pour laquelle la probabilité d'obtenir Pile vaut 1/4. La variable aléatoire X qui compte le nombre de Pile obtenu suit une loi binomiale de paramètre $n = 6$ et $p = 1/4$: $X \sim \mathcal{B}(6, 1/4)$.

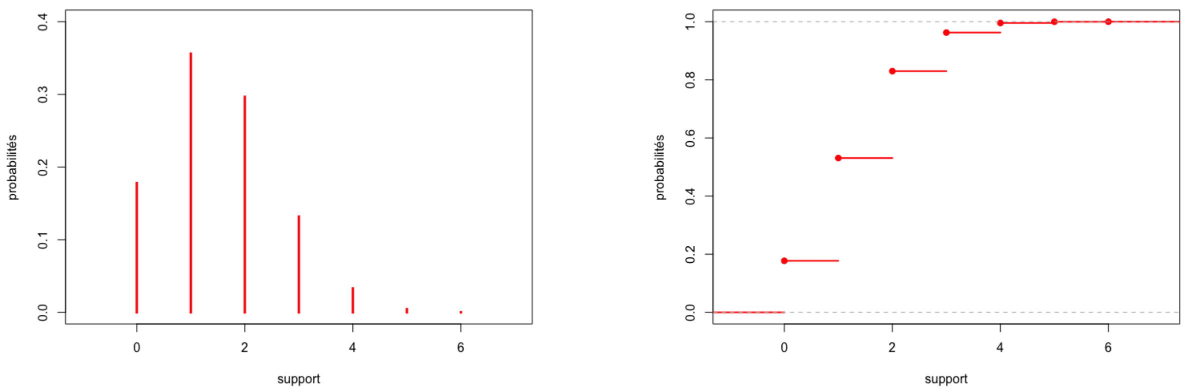


FIGURE 2.6 – Loi binomiale $\mathcal{B}(6, 1/4)$: diagramme en bâtons (gauche) et fonction de répartition (droite).

Proposition 2.54. *La somme de n variables aléatoires indépendantes de loi de Bernoulli de paramètre p est une variable aléatoire de loi binomiale de paramètres n et p. Cela signifie que si $X_1, \dots, X_n$ sont indépendantes de loi $\mathcal{B}(p)$, alors $Y = X_1 + \cdots + X_n \sim \mathcal{B}(n, p)$.*

Démonstration. Les X_i pouvant prendre des valeurs dans $\{0, 1\}$, leur somme Y prend ses valeurs dans $\{0, \dots, n\}$. Il reste à calculer les probabilités $p_k = \mathbb{P}(Y = k)$. Or l'événement $\{Y = k\}$ est réalisé si et seulement si k des n variables X prennent la valeur 1 et $n - k$ la valeur 0. Pour $I \subset \{0, \dots, n\}$, on définit les ensembles E_I par

$$E_I = \{\forall i \in I,\ X_i = 1,\ \text{ et } \forall i \notin I,\ X_i = 0\}\,.$$

L'indépendance des X_i implique que

$$\mathbb{P}(E_I) = \prod_{i \in I} \mathbb{P}(X_i = 1) \prod_{i \notin I} \mathbb{P}(X_i = 0) = p^k q^{n-k}\,.$$

Or pour tout $I \subset \{0, \dots, n\}$ de cardinal k fixé, les E_I forment des événements incompatibles et il y en a $\binom{n}{k}$ différents, d'où

$$\mathbb{P}(Y = k) = \mathbb{P}\Big(\bigcup_{\substack{I \subset \{1,\dots,n\} \\ \#I = k}} E_I \Big) = \sum_{\substack{I \subset \{1,\dots,n\} \\ \#I = k}} \mathbb{P}(E_I) = \binom{n}{k} p^k q^{n-k}\,.$$

□

Cette proposition est très utile pour calculer l'espérance et la variance de X : il suffit de sommer les espérances des X_i et, comme les X_i sont indépendantes, de sommer leur variance. Par ailleurs, elle conduit au résultat suivant : si $X \sim \mathcal{B}(n,p)$ et $Y \sim \mathcal{B}(m,p)$ sont deux variables aléatoires indépendantes, alors $X+Y \sim \mathcal{B}(n+m,p)$.

Proposition 2.55. *Si $X \sim \mathcal{B}(n,p)$, alors*

$$\mathbb{E}[X] = np \quad et \quad \mathrm{Var}(X) = np(1-p)\,.$$

Démonstration. D'après la proposition précédente, X peut se décomposer sous la forme $X = X_1 + \cdots + X_n$, où les variables aléatoires X_j sont indépendantes de loi de Bernoulli de paramètre p. Par suite, la linéarité de l'espérance assure que

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = np\,.$$

De plus, la linéarité de la variance dans le cas des variables aléatoires indépendantes donne

$$\mathrm{Var}[X] = \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n] = np(1-p)\,.$$

□

Exemple 2.56. Si l'on reprend l'exemple précédent du lancer de pièce avec $n = 6$ et $p = 1/4$, on obtient $\mathbb{E}[X] = \frac{3}{2}$ et $\mathrm{Var}(X) = \frac{9}{8}$.

Exercice 2.57. Un joueur de tennis réussit son premier service 80% du temps. En moyenne, combien de fois aura-t-il réussi son premier service lors d'un match où il aura servi 50 fois ?

4.2 Lois usuelles infinies

4.2.1 Loi géométrique

La loi géométrique est utilisée pour modéliser des temps d'attente.

Définition 2.58 (Loi géométrique)**.** *On dit qu'une variable aléatoire X suit une* loi géométrique *de paramètre $p \in]0,1[$, et on note $X \sim \mathcal{G}(p)$, si X prend des valeurs dans $\mathbb{N}^*$ avec*

$$\forall k \in \mathbb{N}^*, \quad \mathbb{P}(X = k) = p(1-p)^{k-1}\,.$$

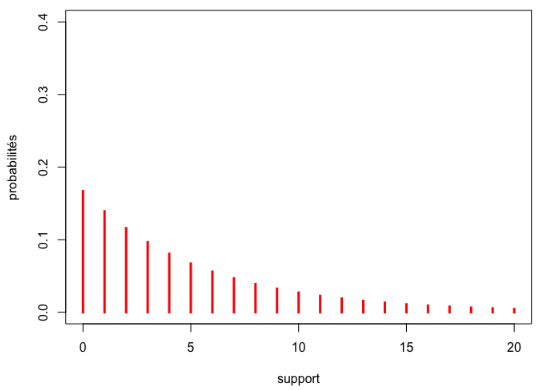
Son nom vient du fait que la suite $(\mathbb{P}(X=k))_{k\geqslant 1}$ est géométrique de raison $1-p$ (et non p, attention).

Cette loi correspond au nombre d'essais avant le premier succès dans une succession d'expériences de Bernoulli, d'où l'idée de temps d'attente. Par exemple, la probabilité d'attendre un temps strictement supérieur à k vaut $\mathbb{P}(X > k) = 1 - \mathbb{P}(X \leq k)$ qui se réécrit

$$1 - \sum_{j=1}^{k} \mathbb{P}(X = j) = 1 - \sum_{j=1}^{k} p(1-p)^{j-1} = 1 - p\sum_{j=0}^{k-1}(1-p)^j = 1 - p\frac{1-(1-p)^k}{p} = (1-p)^k\,.$$

Cela coïncide bien évidemment avec la probabilité d'échouer lors des k premiers essais, chaque échec se produisant avec probabilité $1-p$.

Exemple 2.59. On lance un dé équilibré jusqu'à obtenir un 6. La variable aléatoire X donnant le nombre de lancers nécessaire suit une loi géométrique de paramètre $p = 1/6$: $X \sim \mathcal{G}(1/6)$.



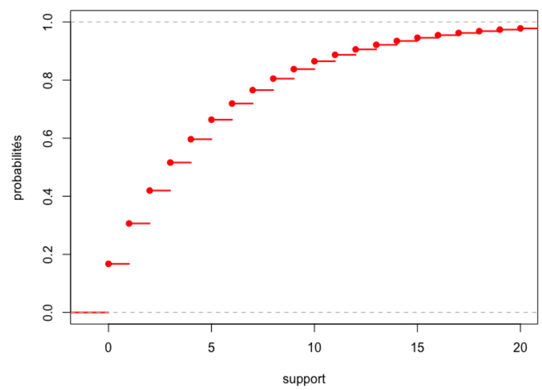


FIGURE 2.7 – Loi géométrique $\mathcal{B}(6, 1/4)$: diagramme en bâtons (gauche) et fonction de répartition (droite).

Proposition 2.60. *Si $X \sim \mathcal{G}(p)$, alors*

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

Le calcul a déjà été effectué dans le cas $p = 1/6$, voir les exemples 2.13 et 2.23. Il se généralise aisément, l'outil central étant l'utilisation de la fonction $x \mapsto \sum_{k=0}^{+\infty} x^k = (1-x)^{-1}$ et de ses dérivées.

Exemple 2.61. Dans l'expérience du lancer de dé, le nombre moyen de lancers pour voir apparaître un 6 vaut donc $\mathbb{E}[X] = 6$. La variance de cette variable aléatoire vaut $\mathrm{Var}(X) = 30$ et son écart-type vaut $\sigma(X) = \sqrt{30} \approx 5.5$.

Une propriété caractéristique de la loi géométrique est son "absence de mémoire". C'est ce qu'énonce la proposition suivante.

Proposition 2.62. *Soit X une variable aléatoire de loi géométrique de paramètre p : $X \sim \mathcal{G}(p)$. Alors*

$$\forall k, l \in \mathbb{N}^*, \quad \mathbb{P}(X > k + l \mid X > l) = \mathbb{P}(X > k).$$

Si X modélise un temps d'attente, cette propriété s'interprète de la manière suivante : si on a déjà attendu un temps l, alors la probabilité d'attendre encore un temps k correspond à la probabilité d'attendre un temps k dès le début.

Démonstration. Il suffit de remarquer que $\{X > k + l\} \cap \{X > l\} = \{X > k + l\}$, de sorte que

$$\mathbb{P}(X > k + l \mid X > l) = \frac{\mathbb{P}(X > k + l)}{\mathbb{P}(X > l)} = \frac{(1-p)^{k+l}}{(1-p)^k} = (1-p)^l = \mathbb{P}(X > l)\,,$$

où l'on a utilisé à plusieurs reprises la relation $\mathbb{P}(X > k) = (1-p)^k$ établie plus haut. □

Exemple 2.63. Vous avez déjà lancé le dé 10 fois sans voir de 6 apparaître. La probabilité de ne pas obtenir de 6 lors des 4 prochains lancers est alors identique à celle de ne pas obtenir de 6 lors des 4 lancers du début :

$$\mathbb{P}(X > 14 \mid X > 10) = \mathbb{P}(X > 10)\,.$$

La loi géométrique vérifie une autre propriété importante : elle est stable par minimum.

Proposition 2.64. *Soit X_1, X_2 deux variables aléatoires indépendantes suivant une loi géométrique de paramètre respectif $p_1 \in]0,1[$ et $p_2 \in]0,1[$. La variable aléatoire $Z = \min(X_1, X_2)$ suit alors une loi loi géométrique de paramètre $1 - (1-p_1)(1-p_2)$:*

$$Z \sim \mathcal{G}(1 - (1-p_1)(1-p_2))\,.$$

Démonstration. On rappelle encore une fois que si $X \sim \mathcal{G}(p)$, alors $\mathbb{P}(X > k) = (1-p)^k$ pour tout $k \in \mathbb{N}$. Ainsi, la fonction de répartition de X est donnée par $\mathbb{P}(X \leq k) = 1 - (1-p)^k$.

Montrons que Z suit une loi géométrique. L'indépendance de X_1 et X_2 permet d'écrire

$$\mathbb{P}(Z > k) = \mathbb{P}(X_1 > k, X_2 > k) = \mathbb{P}(X_1 > k)\mathbb{P}(X_2 > k) = (1-p_1)^k(1-p_2)^k\,.$$

La fonction de répartition F de Z est alors donnée par la relation

$$F(x) = \mathbb{P}(Z \leq x) = 1 - [(1-p_1)(1-p_2)]^k\,,$$

On reconnaît la fonction de répartition d'une loi géométrique. Comme la fonction de répartition d'une variable aléatoire caractérise sa loi, on en déduit que Z suit une loi géométrique de paramètre $p = 1 - (1-p_1)(1-p_2)$. □

Le paramètre de Z se retrouve facilement si l'on interprète la loi géométrique en terme de temps d'attente avant un succès. En effet, le minimum $Z = \min(X_1, X_2)$ correspond au nombre d'essais nécessaires avant le premier succès quand on mène simultanément une expérience de Bernoulli de paramètre p_1 et une expérience de Bernoulli de paramètre p_2. Le paramètre de la loi géométrique Z correspond alors à la probabilité p qu'au moins une expérience soit un succès, qui est égale à

$$p = \mathbb{P}(\{X_1 = 1\} \cup \{X_2 = 1\}) = 1 - \mathbb{P}(\{X_1 = 0\} \cap \{X_2 = 0\}) = 1 - (1-p_1)(1-p_2)\,,$$

où la dernière égalité résulte de l'indépendance de X_1 et X_2.

Notons qu'une récurrence directe permet d'étendre la proposition précédente au cas de n variables aléatoires $X_1, \ldots, X_n$ indépendantes de loi géométrique de paramètre respectif $p_1, \ldots, p_n$. Dans ce cas, la variable aléatoire $Z = \min(X_1, \ldots, X_n)$ suit une loi géométrique de paramètre $p = 1 - (1-p_1)\cdots(1-p_n)$:

$$Z \sim \mathcal{G}(1 - (1-p_1)\cdots(1-p_n))\,.$$

4.2.2 Loi de Poisson

La loi de Poisson permet de modéliser les événements rares. Il n'existe cependant pas d'expérience simple permettant de mettre en évidence cette loi.

Définition 2.65 (Loi de Poisson)**.** *On dit qu'une variable aléatoire X suit une* loi de Poisson *de paramètre $\lambda > 0$, et on note $X \sim \mathcal{P}(\lambda)$, si X prend ses valeurs dans $\mathbb{N}$ avec*

$$\forall k \in \mathbb{N}\,, \quad \mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}\,.$$

On rappelle que si $x \in \mathbb{R}$, alors

$$\mathrm{e}^x = \sum_{n=0}^{+\infty} \frac{x^n}{n!}\,,$$

ce qui justifie que la définition précédente correspond bien à une loi de probabilité.

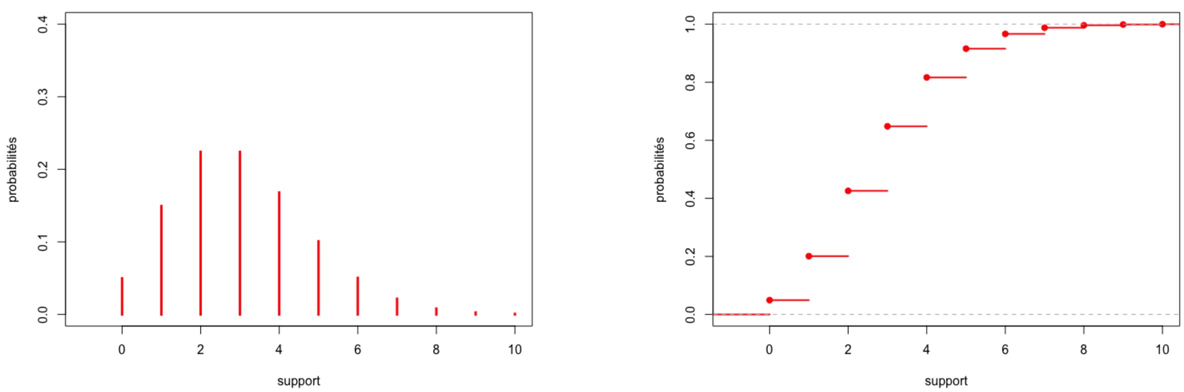


FIGURE 2.8 – Loi de Poisson $\mathcal{P}(3)$: diagramme en bâtons (gauche) et fonction de répartition (droite).

Proposition 2.66. *Si X suit une loi de Poisson de paramètre $\lambda > 0$, alors*

$$\mathbb{E}[X] = \lambda \quad \textit{et} \quad \mathrm{Var}(X) = \lambda\,.$$

Démonstration. On définit $\phi(x) = \mathrm{e}^x = \sum_{k=0}^{+\infty} \frac{x^k}{k!}$. On remarque alors que

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

Concernant la variance, on calcule d'abord $\mathbb{E}[X^2]$ avec l'astuce $k^2 = k(k-1)+k$ déjà utilisée pour la loi géométrique :

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

Ainsi, $\mathbb{E}[X^2] = \lambda^2 + \lambda$. On obtient alors le résultat souhaité :

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

□

La loi de Poisson simule l'occurence d'un événement rare au sein d'une très grande population. Le paramètre λ correspond donc au nombre moyen d'apparition de l'événement rare. Plus λ est proche de 0, plus la probabilité d'observer un tel événement est faible.

Proposition 2.67. *Soit $(p_n)_{n\in\mathbb{N}}$ une suite de $]0,1[$ telle que $\lim_{n\to+\infty} np_n = \lambda > 0$. Soit X_n une variable aléatoire de loi binomiale $\mathcal{B}(n,p_n)$. Alors*

$$\forall k \in \mathbb{N}\,, \quad \mathbb{P}(X_n = k) \underset{n\to+\infty}{\longrightarrow} e^{-\lambda}\frac{\lambda^k}{n!}\,.$$

On rappelle que si on répète n fois une expérience de Bernoulli dont la probabilité de succès est p, alors la variable aléatoire X donnant le nombre de succès parmi les n expériences suit une loi binomiale de paramètres n et p. La proposition dit que si le nombre d'expériences n est grand et que la probabilité de succès p est petite (de l'ordre de λ/n, avec $\lambda > 0$ fixé), alors X suit approximativement une loi de Poisson de paramètre $\lambda = np$.

Exemple 2.68. On suppose que la probabilité de gagner au loto est d'environ $p = 10^{-6}$ (en réalité, c'est beaucoup moins, de l'ordre de 10^{-10}). Ainsi, si $n = 10^6$ personnes jouent une grille, le nombre de vainqueurs peut être modélisé par une loi de Poisson de paramètre $\lambda = np = 1$. En moyenne, il y aura un vainqueur : c'est l'espérance d'une loi de Poisson de paramètre $\lambda = 1$.

Bilan loi discrètes

Loi	Uniforme	Bernoulli	Binomiale	Géométrique	Poisson
Support	$\{1,\ldots,n\}$	$\{0,1\}$	$\{0,\ldots,n\}$	$\mathbb{N}^*$	$\mathbb{N}$
p_k	$\frac{1}{n}$	$p_0 = 1-p$ et $p_1 = p$	$\binom{n}{k}p^k(1-p)^{n-k}$	$(1-p)^{k-1}p$	$\frac{e^{-\lambda}\lambda^k}{k!}$
Espérance	$\frac{n+1}{2}$	p	np	$\frac{1}{p}$	λ
Variance	$\frac{n^2-1}{12}$	$p(1-p)$	$np(1-p)$	$\frac{1-p}{p^2}$	λ

Chapitre 3

Variables aléatoires à densité

On considère désormais des variables aléatoires dont le support n'est plus un ensemble discret. Parmi celles-ci, on se limitera à l'étude de celles ayant une représentation sous forme intégrale, appelées variables aléatoires à densité.

On considère dans tout ce chapitre un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.

1 Loi d'une variable aléatoire à densité

1.1 Densité

On généralise tout d'abord la notion de variable aléatoire, qui étend celle déjà vue dans le cadre discret.

Définition 3.1 (Variable aléatoire)**.** *Une* variable aléatoire *est une application*

$$\begin{array}{cccc} X & : & (\Omega, \mathcal{F}, \mathbb{P}) & \to & \mathbb{R} \\ & & \omega & \mapsto & X(\omega)\,, \end{array}$$

telle que pour tout intervalle I de $\mathbb{R}$, on a

$$\{X \in I\} = \{\omega \in \Omega : X(\omega) \in I\} \in \mathcal{F}\,.$$

Remarquons que cette définition prolonge bien la notion de variables aléatoires discrètes. En effet, il suffit de choisir $I = \{x_k\}$ pour les différentes valeurs x_k prises par X et alors

$$\{X \in I\} = \{X = x_k\} \in \mathcal{F}\,.$$

Remarques 3.2.

1. Comme dans le cas discret, l'idée de la définition est de s'assurer que l'événement $\{X \in I\}$ est dans $\mathcal{F}$: on peut donc calculer sa probabilité. Cependant, il s'avère nécessaire de considérer des intervalles quelconques I et non plus des points. On verra en effet que les événements du type $\{X = x\}$ pour un réel $x \in \mathbb{R}$ fixé seront généralement de probabilité nulle, donc sans intérêt.

2. Comme dans le cas discret, la notion de variable aléatoire est stable par toutes les opérations classiques : somme, produit, différence, maximum, minimum, etc.

On va étudier un type particulier de variables aléatoires appelées variables aléatoires à densité. L'idée est la suivante. Pour une variable aléatoire discrète sa loi peut être représentée via un diagramme en bâtons. On souhaite étendre cette représentation dans la cas d'une variable aléatoire X qui prend un nombre non dénombrable de valeurs dans $\mathbb{R}$. On ne peut alors plus définir la loi de X via la probabilité que X prenne une valeur fixée (cette dernière est nulle, comme on le verra plus tard). Cependant, on peut toujours considérer la probabilité que X tombe dans un intervalle, disons $[a, b]$, avec $a < b$. Dans le cas d'une variable aléatoire discrète, il s'agit alors de sommer la hauteur des bâtons compris entre a et b. Dans le cas non discret, on remplace la somme par une intégrale : on est ainsi amené à calculer une aire sous une courbe.

Définition 3.3 (Densité)**.** *Une* densité *est une fonction* $f : \mathbb{R} \to \mathbb{R}$ *telle que*

1. f *est positive :*
$$\forall x \in \mathbb{R}, \quad f(x) \geq 0,$$
2. f *est d'intégrale 1 :*
$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1.$$

Il faudrait ici être plus précis sur la notion d'intégrale. Un cadre rigoureux est celui de l'intégrale de Lebesgue qui dépasse le cadre de ce cours. On se placera donc dans le cas où les fonctions f à intégrer sont "sympathiques", par exemple continues par morceaux.

Exemples 3.4.

1. *Uniforme.* On considère la fonction f constante égale à 1 sur $[0, 1]$ et nulle ailleurs. On vérifie aisément que cette fonction est une densité, elle est représentée en Figure 3.1 (à gauche).
2. *Exponentielle.* La fonction définie par $x \mapsto \mathrm{e}^{-x}$ sur $[0, +\infty[$ et nulle ailleurs est une densité. On vérifie en effet que
$$\int_0^{+\infty} \mathrm{e}^{-x}\,\mathrm{d}x = \big[-\mathrm{e}^{-x}\big]_0^{+\infty} = 1.$$
Son graphe est donné en Figure 3.1 (à droite).

Définition 3.5 (Variable aléatoire à densité)**.** *Une variable aléatoire X est dite* à densité *(ou* absolument continue*) s'il existe une densité f telle que, pour tout intervalle $I \subset \mathbb{R}$,*
$$\mathbb{P}(X \in I) = \int_I f(x)\,\mathrm{d}x.$$
On dit alors que X est une variable aléatoire de densité f.

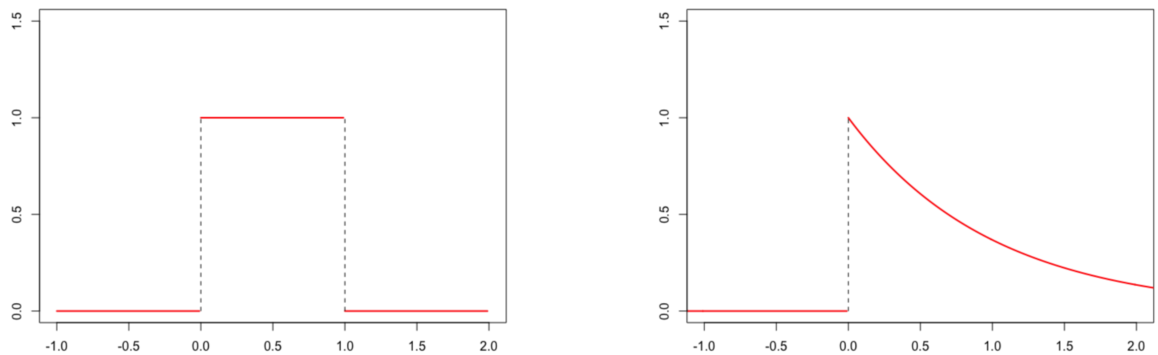


FIGURE 3.1 – La densite *uniforme* (gauche) et la densité *exponentielle* (droite).

Interprétation. Si X est une variable aléatoire absolument continue de densité f et $I \subset \mathbb{R}$, alors la probabilité que X prenne une valeur dans I correspond à l'aire sur I de la surface entre f (qui est positive) et l'axe des abscisses. Cette probabilité est donc d'autant plus grande que f prend des grandes valeurs sur I.

Remarquons d'ores et déjà que si a est un réel fixé, alors la probabilité qu'une variable aléatoire X de densité f prenne la valeur a vaut 0 puisque

$$\mathbb{P}(X = a) = \mathbb{P}(X \in [a, a]) = \int_a^a f(x)\,\mathrm{d}x = 0\,.$$

Ceci se comprend assez bien : vu que X peut prendre une infinité non dénombrable de valeurs dans $\mathbb{R}$, la probabilité de tomber exactement sur une valeur précise est nulle.

Remarque 3.6. On a donc défini deux types de variables aléatoires : les variables aléatoires discrètes et les variables aléatoires à densité. Il faut noter qu'il existe des variables aléatoires mixtes : ni discrète, ni à densité. Néanmoins, les variables aléatoires à densité constituent une grande sous-classe de la famille des variables aléatoires car leur représentation intégrale permet d'effectuer de nombreux calculs en pratique.

Exemples 3.7.

1. *Uniforme.* On a vu que la fonction f définie sur $[0, 1]$ par $f(x) = 1$ est une densité. Soit X une variable aléatoire de densité f. On dit alors que X suit une loi uniforme sur $[0, 1]$. Voici quelques calculs que l'on peut faire :

$$\mathbb{P}(X \leqslant 1/4) = 1/4\,, \quad \mathbb{P}(1/3 \leqslant X < 7/8) = 7/8 - 1/3 = 13/24\,, \quad \mathbb{P}(X = 1/3) = 0\,.$$

2. *Exponentielle.* On a vu que la fonction f définie par $f(x) = \mathrm{e}^{-x}$ sur $[0, +\infty[$ et 0 ailleurs est une densité. Soit X une variable aléatoire de densité f. On dit alors que X suit une loi exponentielle. On peut alors par exemple calculer les quantités suivantes :

$$\mathbb{P}(X \leq 1) = \int_{-\infty}^{1} f(x)\,\mathrm{d}x = \int_{0}^{1} \mathrm{e}^{-x}\,\mathrm{d}x = [-\mathrm{e}^{-x}]_0^1 = 1 - \mathrm{e}^{-1}\,,$$

$$\mathbb{P}(3 \leq X \leq 5) = \int_{3}^{5} \mathrm{e}^{-x}\,\mathrm{d}x = [-\mathrm{e}^{-x}]_3^5 = \mathrm{e}^{-3} - \mathrm{e}^{-5}\,.$$

Attention. Si la notion de variable aléatoire discrète, et plus généralement de variable aléatoire, est stable par les opérations classiques, ce n'est pas le cas pour celle de variable aléatoire à densité. Prenons par exemple une variable aléatoire X de loi uniforme sur $[0, 1]$ et définissons $Y = 1 - X$ qui suit également une loi uniforme sur $[0, 1]$ (exercice). Alors, la variable aléatoire $X + Y$ est une variable aléatoire discrète constante égale à 1. Elle n'admet donc pas de densité.

1.2 Fonction répartition

On étend la définition d'une fonction de répartition au cas des variables aléatoires à densité.

Définition 3.8 (Fonction de répartition)**.** *Soit X une variable aléatoire de densité f. La fonction de répartition de X est la fonction F définie par*

$$\forall x \in \mathbb{R}\,, \quad F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t\,.$$

Comme pour le cas d'une variable aléatoire discrète, la fonction de répartition d'une variable aléatoire à densité permet de calculer la probabilité de tomber dans n'importe quel intervalle :

$$\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \leq b) - \mathbb{P}(X < a) = \int_{-\infty}^{b} f(t)\,\mathrm{d}t - \int_{-\infty}^{a} f(t)\,\mathrm{d}t = \int_{a}^{b} f(t)\,\mathrm{d}t\,.$$

Notons que rien ne change si les bornes de l'intervalle sont ouvertes ou fermées puisque la probabilité de tomber sur a ou sur b vaut 0.

Exemples 3.9. On reprend les deux exemples précédents.

1. *Uniforme.* La densité est donnée par $f(x) = 1$ sur $[0, 1]$ et $f(x) = 0$ sinon. Ainsi, F est nulle avant 0, égale à 1 après 1, et vaut

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_{0}^{x} 1\,\mathrm{d}t = x\,,$$

sur $[0, 1]$. Sa représentation est donnée en Figure 3.2 à gauche.

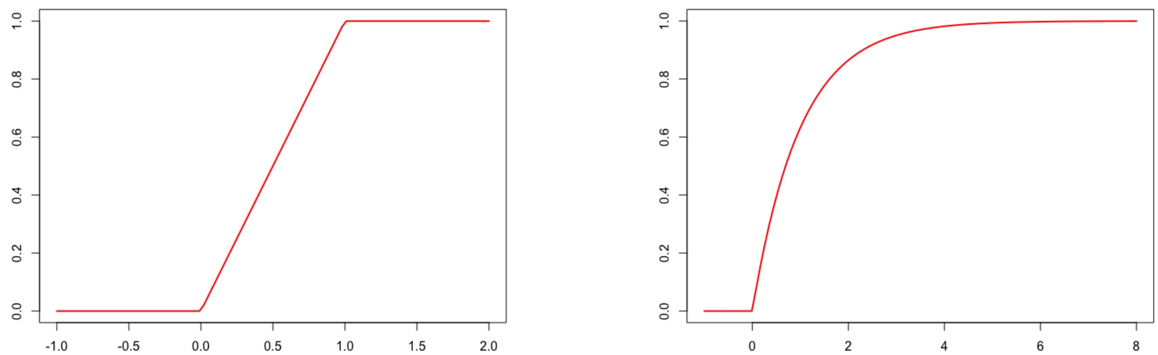


FIGURE 3.2 – La fonction de répartition d'une loi uniforme (gauche) et d'une loi exponentielle fonction de répartition (droite).

2. *Exponentielle.* La densité est donnée par $f(x) = \mathrm{e}^{-x}$ sur $[0, +\infty[$. Ainsi, F est nulle sur $]-\infty, 0[$ et vaut
$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_{0}^{x} \mathrm{e}^{-t}\,\mathrm{d}t = \big[-\mathrm{e}^{-t}\big]_0^x = 1 - \mathrm{e}^{-x}\,,$$
pour $x \in [0, +\infty[$. La représentation graphique de F est donnée en Figure 3.2 à droite.

La fonction de répartition d'une variable aléatoire à densité possède les mêmes propriétés que dans le cas discret.

Proposition 3.10. *Soit X une variable aléatoire de densité f et de fonction de répartition F.*

1. *F est une fonction croissante.*
2. $\lim_{x\to-\infty} F(x) = 0$ *et* $\lim_{x\to+\infty} F(x) = 1$.
3. *F est une fonction continue sur $\mathbb{R}$.*
4. *En tout point où f est continue, F est dérivable et on a $F'(x) = f(x)$.*

Démonstration. Les deux premiers points se prouvent comme dans le cas discret.

La continuité à droite se montre également comme dans le cas discret. Montrons la continuité à gauche. Soit a un réel fixé. Remarquons tout d'abord que comme F est une fonction croissante, elle admet une limite à gauche en a, que l'on note $F(a-)$ et qui peut s'écrire
$$F(a-) = \lim_{n\to+\infty} F(a - 1/n)\,.$$
Il s'agit donc de montrer que cette limite coïncide avec $F(a)$. Comme la densité f est supposée continue par morceaux, elle est bornée au voisinage de a, disons par une réel $M > 0$. On en

déduit alors que

$$|F(a) - F(a - 1/n)| = \left| \int_{-\infty}^{a} f(t)\,\mathrm{d}t - \int_{-\infty}^{a-1/n} f(t)\,\mathrm{d}t \right| \leq \int_{a-1/n}^{a} |f(t)|\,\mathrm{d}t \leq \frac{M}{n} \to 0\,, \quad n \to +\infty$$

ce qui prouve que $\lim_{n\to+\infty} F(a - 1/n) = F(a)$. La fonction de répartition F est donc bien continue sur $\mathbb{R}$.

Soit maintenant a un point de continuité de f. Montrons que

$$\lim_{h\to 0} \frac{F(a+h) - F(a)}{h} = f(a)\,.$$

Soit $\epsilon > 0$ fixé. Comme f est continue en a, il existe $\eta > 0$ tel que pour tout $x \in [a - \eta, a + \eta]$, l'inégalité $|f(x) - f(a)| \leq \epsilon$ est vérifiée. Par suite, on obtient

$$\left|\frac{F(a+h) - F(a)}{h} - f(a)\right| = \left|\frac{1}{h}\int_{a}^{a+h} (f(t) - f(a))\,\mathrm{d}t\right| \leq \frac{1}{h}\int_{a}^{a+h} |f(t) - f(a)|\,\mathrm{d}t\,.$$

En prenant $|h| \leq \eta$ on conclut que

$$\left|\frac{F(a+h) - F(a)}{h} - f(a)\right| \leq \epsilon\,,$$

ce qui conclut la preuve. □

La dernière propriété se vérifie facilement sur la loi uniforme (resp. exponentielle), où hormis en 0 et en 1 (resp. hormis en 0), la fonction de répartition d'une loi uniforme sur $[0, 1]$ (resp. exponentielle) est dérivable de dérivée f.

En pratique Dans le cas d'une variable aléatoire à densité, le calcul de la fonction de répartition se ramène ainsi à un calcul de primitive. Plus précisément, soit X une variable aléatoire continue de densité f ayant un support de la forme $[a, b]$ (c'est-à-dire que f est nulle en dehors de l'intervalle $[a, b]$), avec a éventuellement égal à $-\infty$ et b éventuellement égal à $+\infty$ (et dans ce cas les bornes de l'intervalle sont ouvertes). Alors :

- à gauche du support de f, la fonction de répartition est nulle : $F(x) = 0$,
- à droite du support de f la fonction de répartition vaut 1 : $F(x) = 1$,
- à l'intérieur du support de f, la fonction de répartition F est donnée par la primitive de f qui s'annule en a.

Remarque 3.11. Réciproquement, si une variable aléatoire X admet une fonction de répartition F qui est continue partout, dérivable sauf éventuellement en un nombre fini de points et de dérivée notée f aux points de dérivabilité, alors on peut montrer que X est une variable aléatoire de densité f.

On résume ces considérations dans la proposition suivante qui est l'analogue du résultat déjà énoncé dans le cas discret.

Proposition 3.12. *La fonction de répartition d'une variable aléatoire à densité caractérise sa loi : deux variables aléatoires à densité ont même loi si et seulement si elles ont même fonction de répartition.*

On peut alors se demander comment utiliser la fonction de répartition pour déterminer une loi. Donnons un exemple.

Exemple 3.13. On considère une variable aléatoire U suivant une loi uniforme sur $[0,1]$. On rappelle que cela signifie que U a pour densité la fonction égale à 1 sur $[0,1]$ et nulle ailleurs. On définit alors la variable aléatoire $Y = -\log(U)$. Quelle est la loi de Y ? On va passer par la fonction de répartition pour répondre à cette question. Remarquons déjà que puisque $U \in [0,1]$, la variable aléatoire Y prend ses valeurs dans $[0,+\infty[$. La fonction de répartition F_Y de Y vérifie alors $F_Y(x) = 0$ si $x < 0$ et

$$F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}(U \geq \mathrm{e}^{-x}) = 1 - \mathbb{P}(U \leq \mathrm{e}^{-x}) = 1 - \mathrm{e}^{-x} \ .$$

On reconnaît alors la fonction de répartition d'une loi exponentielle de paramètre 1. La proposition précédente assure donc que Y suit une loi exponentielle de paramètre 1. En particulier, on peut retrouver la densité f_Y de la loi exponentielle en dérivant F_Y : $f_Y(x) = F_Y'(x) = \mathrm{e}^{-x}$ si $x \geq 0$ et $f_Y(x) = 0$ sinon.

2 Moments d'une variable aléatoire à densité

On considère dans tout ce qui suit une variable aléatoire absolument continue X de densité f.

2.1 Espérance

Définition 3.14 (Espérance d'une variable aléatoire à densité)**.** *L'espérance d'une variable aléatoire X de densité f est définie par la quantité*

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

sous réserve de convergence de cette intégrale.

Remarque 3.15. Noter que, contrairement au cas discret, il est inutile ici de vérifier a priori l'absolue convergence de l'intégrale puisque la quantité $xf(x)$ est de signe constant au voisinage des bornes.

L'expression de l'espérance est analogue au cas discret : les valeurs prises par X sont les x de l'intégrande (x_k dans le cas discret) et les probabilités deviennent infinitésimales sous la forme $f(x)\,\mathrm{d}x$ (p_k dans le cas discret). Enfin la somme est remplacée par une intégrale.

Tout comme dans le cas discret, l'espérance s'interprète comme la moyenne des valeurs prises par la variable aléatoire X pondérées par les "probabilités" $f(x)\,\mathrm{d}x$.

Exemples 3.16.

1. *Uniforme.* Si X suit une loi uniforme sur $[0,1]$, sa densité est donnée par $f(x)=1$ sur $[0,1]$ et 0 ailleurs. L'espérance de X vaut donc
$$\mathbb{E}[X]=\int_0^1 x\,\mathrm{d}x=\frac{1}{2}\,.$$
Autrement dit, si on tire "au hasard" un nombre entre 0 et 1 on obtient en moyenne $1/2$.
2. *Exponentielle.* Si X suit une loi exponentielle, sa densité est donnée par $f(x)=\mathrm{e}^{-x}$ sur $[0,+\infty[$ et 0 ailleurs. Ainsi, l'espérance de X vaut
$$\mathbb{E}[X]=\int_0^{+\infty} x\,\mathrm{e}^{-x}\,\mathrm{d}x\,,$$
et une intégration par parties donne
$$\mathbb{E}[X]=\big[-x\,\mathrm{e}^{-x}\big]_0^{+\infty}+\int_0^{+\infty}\mathrm{e}^{-x}\,\mathrm{d}x=0+\big[-\mathrm{e}^{-x}\big]_0^{+\infty}=1\,.$$

Contre-exemple 3.17. Un exemple classique de non-existence de l'espérance est la loi de Cauchy. Cette loi est donnée par la densité $f(x)=\frac{1}{\pi(1+x^2)}$. On vérifie que cette fonction est bien une densité puisqu'elle est positive et d'intégrale
$$\int_{\mathbb{R}}\frac{1}{\pi(1+x^2)}\,\mathrm{d}x=\frac{1}{\pi}\big[\arctan(x)\big]_{-\infty}^{+\infty}=\frac{1}{\pi}\Big[\frac{\pi}{2}-\Big(-\frac{\pi}{2}\Big)\Big]=1\,.$$
Concernant l'espérance, on remarque que l'intégrale généralisée $\int xf(x)\,\mathrm{d}x$ n'est pas convergente car $xf(x)$ est équivalent à $1/x$ en $\pm\infty$.

Comme dans le cas discret, il arrive souvent qu'on veuille calculer non pas l'espérance de X mais l'espérance d'une fonction de X. Pour cela, on s'appuie sur le théorème de transfert suivant, analogue du cas discret. On admet sa preuve.

Théorème 3.18 (Théorème de transfert)**.** *Soit X une variable aléatoire de densité f et $\phi:\mathbb{R}\to\mathbb{R}$ une fonction. Alors l'espérance de $\phi(X)$ vaut*
$$\mathbb{E}[\phi(X)]=\int_{\mathbb{R}}\phi(x)f(x)\,\mathrm{d}x\,,$$
sous réserve d'absolue convergence de l'intégrale, c'est-à-dire $\int_{\mathbb{R}}|\phi(x)|f(x)\,\mathrm{d}x<+\infty$.

Remarques 3.19.

1. Notons qu'il faut ici supposer la condition $\int_{\mathbb{R}} |\phi(x)| f(x)\,\mathrm{d}x < +\infty$ car on ne sait rien du signe de $\phi(x)$.
2. On retiendra donc que pour calculer $\mathbb{E}[\phi(X)]$, il suffit de remplacer x par $\phi(x)$ dans l'intégrale. Le théorème de transfert permet donc de calculer l'espérance d'une variable aléatoire $Y = \phi(X)$ sans connaître la loi de Y.

Exemples 3.20.

1. *Uniforme.* Si X suit une loi uniforme sur $[0,1]$, alors l'espérance de X^2 est donnée par
$$\mathbb{E}[X^2] = \int_0^1 x^2\,\mathrm{d}x = \Big[\frac{x^3}{3}\Big]_0^1 = \frac{1}{3}\,.$$
2. *Exponentielle.* De même, si X suit une loi exponentielle, alors l'espérance de X^2 est donnée par
$$\mathbb{E}[X^2] = \int_0^{+\infty} x^2\,\mathrm{e}^{-x}\,\mathrm{d}x\,.$$
Une intégration par parties donne alors
$$\mathbb{E}[X^2] = \Big[-x^2\,\mathrm{e}^{-x}\Big]_0^{+\infty} + \int_0^{+\infty} 2x\,\mathrm{e}^{-x}\,\mathrm{d}x = 0 + 2\int_0^{+\infty} x\,\mathrm{e}^{-x}\,\mathrm{d}x\,.$$
On reconnaît dans le deuxième terme l'expression de l'espérance de X, ce qui donne
$$\mathbb{E}[X^2] = 2\mathbb{E}[X] = 2\,.$$

Les propriétes de l'espérance dans le cas d'une variable aléatoire à densité sont les mêmes que dans le cas discret. On les regroupe ci-dessous. Les preuves sont les mêmes que dans le cas discret.

Propriétés 3.21. *Soit X et Y deux variables aléatoires à densité admettant une espérance et a et b deux réels.*

1. *(Linéarité). La variable aléatoire $aX + bY$ admet une espérance donnée par*
$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]\,.$$
2. *(Positivité). Si $X \geq 0$ (X ne prend que des valeurs positives), alors $\mathbb{E}[X] \geq 0$.*
3. *(Croissance). Si $X \leq Y$, alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$.*

Ces résultats sont en fait valables pour toute variable aléatoire au sens de la définition 3.1 (à condition bien sûr de définir l'espérance dans un cadre général).

2.2 Variance

L'espérance est une mesure de la tendance centrale d'une variable aléatoire. La variance permet de mesurer l'écart autour de cette valeur centrale.

Définition 3.22 (Variance)**.** *Soit X une variable aléatoire de densité f admettant une espérance $\mathbb{E}[X]$. On définit la* variance *de X par la quantité*

$$\mathrm{Var}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] = \int_{\mathbb{R}} (x - \mathbb{E}[X])^2 f(x)\,\mathrm{d}x\,,$$

sous réserve de convergence de cette intégrale.

*On définit l'*écart-type *comme la racine carrée de la variance : $\sigma(X) = \sqrt{\mathrm{Var}(X)}$.*

La variance permet de mesurer la moyenne des écarts à la moyenne : c'est une mesure de dispersion.

Exemples 3.23.

1. *Uniforme.* Si X suit une loi uniforme, son espérance vaut $\mathbb{E}[X] = \frac{1}{2}$. Ainsi, sa variance vaut

$$\mathrm{Var}(X) = \int_0^1 (x - \frac{1}{2})^2\,\mathrm{d}x = \left[\frac{(x-\frac{1}{2})^3}{3}\right]_0^1 = \frac{1}{12}\,.$$

2. *Exponentielle.* Si X suit une loi exponentielle, son espérance vaut $\mathbb{E}[X] = 1$. Sa variance est alors donnée par la formule :

$$\mathrm{Var}(X) = \int_0^{+\infty} (x-1)^2\,\mathrm{e}^{-x}\,\mathrm{d}x\,.$$

Une intégration par parties donne alors

$$\mathrm{Var}(X) = \big[-(x-1)^2\,\mathrm{e}^{-x}\big]_0^{+\infty} + 2\int_0^{+\infty} (x-1)\,\mathrm{e}^{-x}\,\mathrm{d}x\,.$$

Le premier terme vaut $\big[-(x-1)^2\,\mathrm{e}^{-x}\big]_0^{+\infty} = 1$, tandis que le deuxième terme vaut

$$2\Big(\int_0^{+\infty} x\,\mathrm{e}^{-x}\,\mathrm{d}x - \int_0^{+\infty} \mathrm{e}^{-x}\,\mathrm{d}x\Big) = 0\,.$$

Ainsi, $\mathrm{Var}(X) = 1$.

Comme dans le cas discret, la variance vérifie plusieurs propriétés, notamment la formule de Koenig-Huygens.

Propriétés 3.24. *Soit X une variable aléatoire à densité admettant une variance.*

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

2. *Pour tous réels a, b,* $\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X)$.

La formule de Koenig-Huygens se réécrit :

$$\mathrm{Var}(X) = \int_{\mathbb{R}} x^2 f(x)\,\mathrm{d}x - \left(\int_{\mathbb{R}} x f(x)\,\mathrm{d}x\right)^2 .$$

Démonstration. 1. On note $\mu = \mathbb{E}[X]$ et on développe le carré dans la définition de la variance de X :

$$\mathrm{Var}(X) = \int_{\mathbb{R}} (x-\mu)^2 f(x)\,\mathrm{d}x = \int_{\mathbb{R}} x^2 f(x)\,\mathrm{d}x - 2\int_{\mathbb{R}} x\mu f(x)\,\mathrm{d}x + \int_{\mathbb{R}} \mu^2 f(x)\,\mathrm{d}x .$$

Le premier terme correspond à $\mathbb{E}[X^2]$, le deuxième à $-2\mu\mathbb{E}[X] = -2\mu^2$, et le troisième à μ^2. On obtient alors le résultat voulu.

2. On utilise la formule de Koenig-Huygens. D'une part, l'espérance de $(aX+b)^2$ vaut

$$\mathbb{E}[(aX+b)^2] = \mathbb{E}[a^2X^2 + 2abX + b^2] = a^2\mathbb{E}[X^2] + 2ab\mathbb{E}[X] + b^2 .$$

D'autre part, le carré de $\mathbb{E}[aX+b]$ vaut

$$\mathbb{E}[aX+b]^2 = (a\mathbb{E}[X]+b)^2 = a^2\mathbb{E}[X]^2 + 2ab\mathbb{E}[X] + b^2 .$$

La différence des deux quantités donne alors

$$\mathrm{Var}(aX+b) = \mathbb{E}[(aX+b)^2] - \mathbb{E}[aX+b]^2 = a^2(\mathbb{E}[X^2] - \mathbb{E}[X]^2) = a^2\,\mathrm{Var}(X) .$$

□

Exemples 3.25.

1. *Uniforme.* Dans le cas d'une loi uniforme, on a vu que $\mathbb{E}[X] = \frac{1}{2}$ et que $\mathbb{E}[X^2] = \frac{1}{3}$. La formule de Koenig-Huygens donne bien la valeur de la variance calculée précédemment :

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12} .$$

2. *Exponentielle.* Dans le cas d'une loi exponentielle, on a vu que $\mathbb{E}[X] = 1$ et que $\mathbb{E}[X^2] = 2$. Ainsi, $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 1$, qui coïncide encore une fois avec la valeur calculée précédemment.

L'inégalité de Bienaymé-Tchebychev permet de quantifier la phrase "la variance mesure l'écart d'une variable aléatoire à sa moyenne". La preuve est similaire au cas discret.

Proposition 3.26 (Inégalité de Bienaymé-Tchebychev)**.** *Soit X une variable aléatoire à densité admettant une variance. Alors*

$$\forall t > 0, \quad \mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2} .$$

Ainsi, la probabilité que X s'écarte de sa moyenne d'une distance t est contrôlée par sa variance et une décroissance en t^{-2}.

Exemple 3.27. Soit X une variable aléatoire de loi exponentielle. On rappelle que X prend alors ses valeurs dans $[0, +\infty[$ et que $\mathrm{Var}(X) = 1$. L'inégalité de Bienaymé-Tchebychev donne alors que pour tout $t > 0$,

$$\mathbb{P}(|X-1| \geq t) \leq \frac{1}{t^2}\,.$$

Par exemple, en prenant $t = 2$, on obtient

$$\mathbb{P}(|X-1| \geq 2) = \mathbb{P}(X > 3) \leq \frac{1}{4}\,.$$

Il n'y a donc qu'une chance sur quatre que X soit supérieur à 3. Notons toutefois que cette borne est très médiocre. En effet, la probabilité précédente peut se calculer directement et vaut

$$\mathbb{P}(X \geq 3) = \int_3^{+\infty} \mathrm{e}^{-x}\;\mathrm{d}x = [\mathrm{e}^{-x}]_3^{+\infty} = \mathrm{e}^{-3} \approx 0.04\,.$$

2.3 Moments

On généralise à présent les notions d'espérance et de variance.

Définition 3.28 (Moment d'une variable aléatoire à densité)**.** *Soit X une variable aléatoire de densité f et $m \in \mathbb{N}^*$. On appelle*

1. moment d'ordre *m de X l'espérance de X^m :*

$$\mathbb{E}[X^m] = \int_{\mathbb{R}} x^m f(x)\,\mathrm{d}x\,,$$

2. moment centré d'ordre *m de X l'espérance de $(X - \mathbb{E}[X])^m$:*

$$\mathbb{E}[(X - \mathbb{E}[X])^m] = \int_{\mathbb{R}} (x - \mathbb{E}[X])^m f(x)\,\mathrm{d}x\,,$$

sous réserve de convergence des intégrales considérées.

L'espérance correspond donc au moment d'ordre 1 et la variance au moment centré d'ordre 2.

On rappelle le vocabulaire déjà utilisé dans le cas discret : une variable aléatoire est dite

— *centrée* si $\mathbb{E}[X] = 0$,

— *réduite* si $\mathrm{Var}(X) = 1$.

— *centrée réduite* si les deux conditions précédentes sont vérifiées.

Si X est une variable aléatoire admettant une variance, alors la variable aléatoire $\frac{X - \mathbb{E}[X]}{\sigma(X)}$ est centrée réduite.

Proposition 3.29. *Soit X une variable aléatoire à densité admettant un moment d'ordre $m \in \mathbb{N}^*$. Alors X admet un moment d'ordre p pour tout $p \in \{1, \ldots, m\}$.*

Démonstration. La preuve est similaire au cas discret. □

On a vu que la variance permettait de mesurer l'écart à la moyenne avec une décroissance en t^{-2}. Plus généralement, les moments permettent de mesurer la décroissance de la queue de distribution d'une variable aléatoire. C'est l'objet de l'inégalité de Markov.

Proposition 3.30 (Inégalité de Markov)**.** *Soit X une variable aléatoire de densité f. Si X admet un moment d'ordre m, alors pour tout $t > 0$,*

$$\mathbb{P}(|X| \geq t) \leq \frac{\mathbb{E}[|X|^m]}{t^m}\,.$$

Démonstration. Pour $t > 0$, on considère l'ensemble $E_t = \{x \in \mathbb{R} : |x| \geq t\}$, qui se décompose en $E_t^+ \cup E_t^-$ avec

$$E_t^+ = \{x \in \mathbb{R}_+ : x \geq t\} \quad \text{et} \quad E_t^- = \{x \in \mathbb{R}_- : x \leq -t\}$$

On décompose alors la probabilité $\mathbb{P}(|X| \geq t)$ via

$$\mathbb{P}(|X| \geq t) = \int_{E_t} f(x)\,\mathrm{d}x = \int_{E_t^+} f(x)\,\mathrm{d}x + \int_{E_t^-} f(x)\,\mathrm{d}x \leq \int_{E_t^+} \frac{x^m}{t^m} f(x)\,\mathrm{d}x + \int_{E_t^-} \frac{(-x)^m}{t^m} f(x)\,\mathrm{d}x\,.$$

En rassemblant les deux intégrales, on obtient la borne

$$\int_{E_t^+} \frac{x^m}{t^m} f(x)\,\mathrm{d}x + \int_{E_t^-} \frac{(-x)^m}{t^m} f(x)\,\mathrm{d}x = \int_{E_t} \frac{|x|^m}{t^m} f(x)\,\mathrm{d}x = \frac{\mathbb{E}[|X|^m]}{t^m}\,,$$

ce qui conclut la preuve. □

Dans le cas d'une loi exponentielle, qui admet des moments d'ordre m pour tout $m \in \mathbb{N}^*$, on retrouve le fait que la queue de distribution de X décroît plus vite que toute fonction puissance. C'est normal vu que cette queue est de la forme e^{-x}.

3 Lois à densité usuelles

On détaille ici quelques lois à densité fréquemment utilisées en probabilité.

3.1 Loi uniforme sur $[a, b]$

La loi uniforme est le prolongement à un intervalle $[a, b]$ de la loi uniforme discrète.

Définition 3.31 (Loi uniforme)**.** *Une variable aléatoire X suit une loi uniforme sur $[a,b]$, $a<b$, si X admet pour densité la fonction f définie par*

$$f(x) = \begin{cases} \frac{1}{b-a} & \textit{si } a \leq x \leq b\,, \\ 0 & \textit{sinon.} \end{cases}$$

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

La densité étant constante sur $[a,b]$, la probabilité que X tombe dans un intervalle $I \subset [a,b]$ dépend uniquement de la longueur de cet intervalle. On a déjà étudié plus haut le cas $X \sim \mathcal{U}([0,1])$.

La densité et la fonction de répartition d'une loi uniforme sur $[-4,-1]$ est donnée en Figure 3.3.

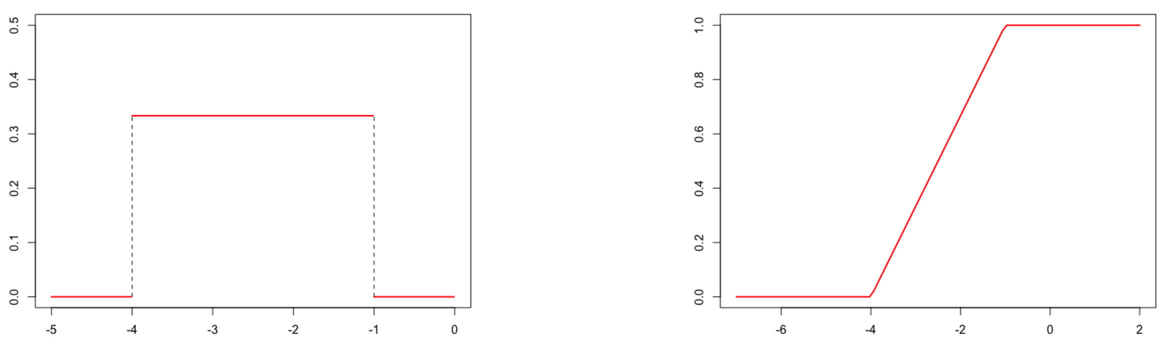


FIGURE 3.3 – Loi uniforme sur $[-4,-1]$: densité (gauche) et fonction de répartition (droite).

L'espérance et la variance d'une loi uniforme sur $[0,1]$ ont déjà été calculé précédemment. L'extension au cas d'un intervalle $[a,b]$ ne pose pas de problème.

Proposition 3.32. *Si $X \sim \mathcal{U}([a,b])$, alors*

$$\mathbb{E}[X] = \frac{a+b}{2}\,, \quad \textit{et} \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12}\,.$$

En particulier, on trouve que l'espérance correspond au milieu du segment $[a,b]$, ce qui semble assez naturel.

Proposition 3.33. *Si $X \sim \mathcal{U}([a,b])$, alors sa fonction de répartition est donnée par*

$$F(x) = \begin{cases} 0 & \textit{si } x < a\,, \\ \frac{x-a}{b-a} & \textit{si } a \leq x \leq b\,, \\ 1 & \textit{si } x > b\,. \end{cases}$$

Démonstration. Puisque X prend ses valeurs dans $[a, b]$, se fonction de répartition est nulle avant a et égale à 1 après b. Par ailleurs, si $x \in [a, b]$, la valeur de $F(x)$ est donnée par

$$F(x) = \mathbb{P}(X \leq x) = \int_a^x f(t)\,\mathrm{d}t = \int_a^x \frac{1}{b-a}\,\mathrm{d}t = \frac{x-a}{b-a}\,.$$

□

3.2 Loi exponentielle de paramètre λ

Définition 3.34 (Loi exponentielle)**.** *Une variable aléatoire X suit une loi exponentielle de paramètre $\lambda > 0$ si X admet pour densité la fonction f définie par*

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \textit{si } x \geq 0\,, \\ 0 & \textit{sinon.} \end{cases}$$

On note alors $X \sim \mathcal{E}(\lambda)$.

Notons que f est bien une densité puisqu'elle est positive et vérifie

$$\int_{\mathbb{R}} f(t)\,\mathrm{d}t = \int_0^\infty \lambda\,\mathrm{e}^{-\lambda x}\,dx = \Big[-\mathrm{e}^{-\lambda x} \Big]_0^\infty = 1\,.$$

La loi exponentielle est utilisée pour modéliser des durées de vie ou des temps d'attente avant l'arrivée d'un événement spécifique. C'est la version continue de la loi géométrique. On remarque que plus λ est grand, plus X a des chances de prendre des valeurs proches de 0.

La densité et la fonction de répartition d'une loi exponentielle de paramètre $\lambda = 3$ est donnée en Figure 3.4.

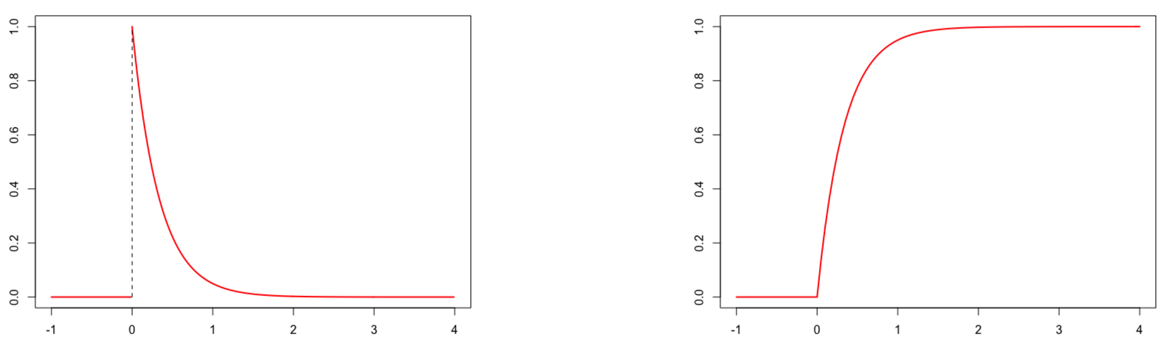


FIGURE 3.4 – Loi exponentielle de paramètre $\lambda = 3$: densité (gauche) et fonction de répartition (droite).

Le calcul de l'espérance et de la variance d'une loi exponentielle de paramètre 1 a déjà été effectué précédemment. Généralisons ceci au cas $\lambda > 0$ quelconque.

Proposition 3.35. *Si $X \sim \mathcal{E}(\lambda)$, alors*

$$\mathbb{E}[X] = \frac{1}{\lambda}\,, \quad \textit{et} \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}\,.$$

Démonstration. L'espérance se calcule à partir d'une intégration par parties :

$$\mathbb{E}[X] = \int_0^{+\infty} x\lambda\, \mathrm{e}^{-\lambda x}\, \mathrm{d}x = \Big[-x\, \mathrm{e}^{-\lambda x} \Big]_0^{+\infty} + \int_0^{+\infty} \mathrm{e}^{-\lambda x}\, \mathrm{d}x\,.$$

Le premier terme est nul tandis que le deuxième terme vaut

$$\int_0^{+\infty} \mathrm{e}^{-\lambda x}\, \mathrm{d}x = \Big[-\frac{\mathrm{e}^{-\lambda x}}{\lambda} \Big]_0^{+\infty} = \frac{1}{\lambda}\,,$$

ce qui prouve le résultat. Concernant la variance, il suffit de calculer $\mathbb{E}[X^2]$. Une intégration par parties donne alors

$$\mathbb{E}[X^2] = \int_0^{+\infty} x^2\lambda\, \mathrm{e}^{-\lambda x}\, \mathrm{d}x = \big[-x^2\, \mathrm{e}^{-\lambda x} \big]_0^{+\infty} + \int_0^{+\infty} 2x\, \mathrm{e}^{-\lambda x}\, \mathrm{d}x = 2\int_0^{+\infty} x\, \mathrm{e}^{-\lambda x}\, \mathrm{d}x\,.$$

Cette dernière intégrale correspond, à un facteur $2/\lambda$ près, à l'espérance de X. Ainsi,

$$\mathbb{E}[X^2] = \frac{2}{\lambda}\mathbb{E}[X] = \frac{2}{\lambda^2}\,.$$

On conclut alors avec la formule de Koenig-Huygens :

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

□

On peut généraliser les résultats précédents. Des intégrations par parties successives donnent en effet les moments d'ordre $m \in \mathbb{N}^*$ de la loi exponentielle :

$$\forall m \in \mathbb{N}^*\,, \quad \mathbb{E}[X^m] = \frac{m!}{\lambda^m}\,.$$

Cette formule peut également se démontrer par récurrence sur m.

Proposition 3.36. *Si $X \sim \mathcal{E}(\lambda)$, alors sa fonction de répartition est donnée par*

$$F(x) = \begin{cases} 0 & \textit{si } x < 0\,, \\ 1 - \mathrm{e}^{-\lambda x} & \textit{si } x \geq 0\,. \end{cases}$$

Démonstration. La densité de X étant nulle sur $]-\infty, 0[$, on a bien $F(x) = 0$ sur cet intervalle. Par ailleurs, si $x \geq 0$, on obtient

$$F(x) = \int_0^x \lambda\, \mathrm{e}^{-\lambda t}\, \mathrm{d}t = \Big[-\mathrm{e}^{-\lambda x} \Big]_0^x = 1 - \mathrm{e}^{-\lambda x}\,.$$

□

La loi exponentielle généralise la loi géométrique. Elle satisfait donc les mêmes propriétés, à savoir absence de mémoire et stabilité par minimum. Ces résultats reposent essentiellement sur la relation
$$\mathbb{P}(X > x) = 1 - F(x) = \mathrm{e}^{-\lambda x}\,, \quad x \geq 0\,,$$
avec $X \sim \mathcal{E}(\lambda)$.

Proposition 3.37. *Soit $X \sim \mathcal{E}(\lambda)$. Pour tout $x, y \geq 0$, on a l'égalité*
$$\mathbb{P}(X > x + y \mid X > y) = \mathbb{P}(X > x)\,.$$

Démonstration. L'égalité ensembliste $\{X > x + y\} \cap \{X > y\} = \{X > x + y\}$ implique que
$$\mathbb{P}(X > x + y \mid X > y) = \frac{\mathbb{P}(X > x + y, X > y)}{\mathbb{P}(X > y)} = \frac{\mathbb{P}(X > x + y)}{\mathbb{P}(X > y)}\,.$$
On obtient alors le résultat souhaité :
$$\mathbb{P}(X > x + y \mid X > y) = \frac{\mathrm{e}^{-\lambda(x+y)}}{\mathrm{e}^{-\lambda y}} = \mathrm{e}^{-\lambda x}\,.$$
□

Proposition 3.38. *Soit X et Y deux variables aléatoires indépendantes suivant des lois exponentielles de paramètre respectif $\lambda > 0$ et $\mu > 0$. Alors la variable aléatoire* $\min(X, Y)$ *suit une loi exponentielle de paramètre $\lambda + \mu$:*
$$\min(X, Y) \sim \mathcal{E}(\lambda + \mu)\,.$$

Ce résultat se généralise par récurrence au cas de n variables aléatoires indépendantes de loi exponentielle.

Démonstration. Montrons que la fonction de répartition de $Z = \min(X, Y)$ est celle d'une loi exponentielle paramètre $\lambda + \mu$. Pour tout $z \geq 0$, on a
$$\mathbb{P}(Z > z) = \mathbb{P}(\min(X, Y) > z) = \mathbb{P}(X > z, Y > z) = \mathbb{P}(X > z)\mathbb{P}(Y > z) = \mathrm{e}^{-\lambda z}\,\mathrm{e}^{-\mu z}\,,$$
où l'avant-dernière égalité est une conséquence de l'indépendance de X et Y. On en déduit alors que pour tout $z > 0$,
$$F_Z(z) = 1 - \mathbb{P}(Z > z) = \mathrm{e}^{-(\lambda+\mu)z}\,.$$
On a également $F_Z(z) = 0$ pour $z < 0$. La fonction F_Z correspond donc à la fonction de répartition d'une loi exponentielle paramètre $\lambda + \mu$. Comme la fonction de répartition d'une variable aléatoire caractérise la loi, on conclut que $\min(X, Y)$ suit une loi exponentielle de paramètre $\lambda + \mu$. □

Les propriétes similaires entre loi exponentielle et loi géométrique proviennent de la remarque suivante : si $X \sim \mathcal{E}(\lambda)$, alors sa partie entière supérieure $\lceil X \rceil$ suit une loi géométrique de paramètre $1 - \mathrm{e}^{-\lambda}$ (exercice).

3.3 Loi normale

La loi normale est une des lois les plus importantes en probabilités et en statistique.

Définition 3.39 (Loi normale)**.** *Une variable aléatoire X suit une loi* normale *ou* gaussienne *de paramètres $\mu \in \mathbb{R}$ et $\sigma > 0$ si X admet pour densité la fonction f définie par*

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(- \frac{(x-\mu)^2}{2\sigma^2} \right).$$

On note alors $X \sim \mathcal{N}(\mu, \sigma^2)$. Si $\mu = 0$ et $\sigma = 1$, on parle de loi normale centrée réduite*.*

Cela signifie que pour tout $a < b$,

$$\mathbb{P}(a \leq X \leq b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\left(- \frac{(x-\mu)^2}{2\sigma^2} \right) \mathrm{d}x\,.$$

L'appellation "loi gaussienne" vient du scientifique allemand Johann Carl Friedrich Gauss qui a beaucoup étudié cette loi. La densité de la loi normale n'admet pas de primitive explicite. Calculer des probabilités avec la loi normale se fait donc via des simulations sur ordinateur. Les résultats numériques sont alors reportés dans des tables. Notons toutefois qu'il est possible de montrer que

$$\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{x^2}{2}}\, \mathrm{d}x = \sqrt{2\pi}\,, \tag{3.1}$$

ce qui prouve que la densité d'une loi normale centrée réduite est bien une densité. Ouf !

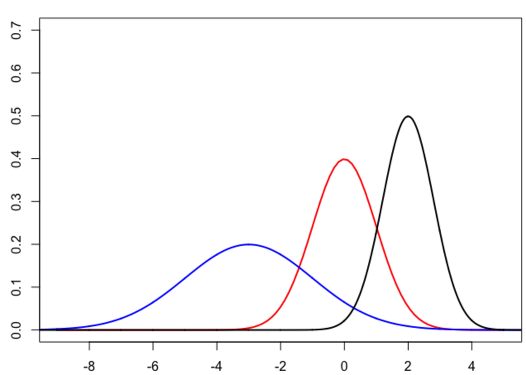


FIGURE 3.5 – Densités de la loi normale : $\mathcal{N}(0,1)$ en rouge, $\mathcal{N}(-3,4)$ en bleu, $\mathcal{N}(2,0.64)$ en noir.

Les densités pour différentes valeurs de μ et σ sont données en Figure 3.5.

La loi normale vérifie une propriété de stabilité par transformation affine. C'est ce qu'énonce la proposition suivante.

Proposition 3.40 (Transformation affine d'une loi normale)**.** *Soit $X \sim \mathcal{N}(\mu, \sigma^2)$ une variable aléatoire de loi normale et $(a, b) \in \mathbb{R}^* \times \mathbb{R}$. La variable aléatoire $Y = aX + b$ suit alors une loi normale $\mathcal{N}(a\mu + b, a^2\sigma^2)$.*

On dit que la loi normale est stable par transformation affine.

Démonstration. Il s'agit de montrer que pour tout couple (c, d) de réels vérifiant $c \leq d$, on a

$$\mathbb{P}(c \leq Y \leq d) = \int_c^d \frac{1}{\sigma\sqrt{a^2 2\pi}} \exp\left(- \frac{(x - (a\mu + b))^2}{2a^2\sigma^2} \right) \mathrm{d}x \, .$$

Traitons le cas $a > 0$. On part de l'égalité

$$\mathbb{P}(c \leq Y \leq d) = \mathbb{P}\left(\frac{c-d}{a} \leq X \leq \frac{d-b}{a} \right) = \int_{\frac{c-b}{a}}^{\frac{d-b}{a}} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(- \frac{(x-\mu)^2}{2\sigma^2} \right) \mathrm{d}x \, ,$$

et on applique le changement de variable $y = ax + b$. On obtient alors

$$\begin{aligned} \mathbb{P}(c \leq Y \leq d) &= \int_c^d \frac{1}{\sigma\sqrt{2\pi}} \exp\left(- \frac{((y-b)/a - \mu)^2}{2\sigma^2} \right) \frac{\mathrm{d}y}{a} \\ &= \int_c^d \frac{1}{a\sigma\sqrt{2\pi}} \exp\left(- \frac{[y - (a\mu + b)]^2}{2\sigma^2 a^2} \right) \mathrm{d}y \, , \end{aligned}$$

ce qui prouve bien que $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$. Le cas $a < 0$ se traite de manière analogue. □

En particulier on a les relations fondamentales suivantes :

1. Si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors $\frac{X-\mu}{\sigma}$ suit une loi normale centrée réduite $\mathcal{N}(0, 1)$.
2. Si $X \sim \mathcal{N}(0, 1)$, alors $\sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2)$.

Il faut retenir ces relations qui seront utiles lors de la lecture de la table de la loi normale.

Proposition 3.41. *Si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors*

$$\mathbb{E}[X] = \mu \quad \textit{et} \quad \mathrm{Var}(X) = \sigma^2 \, .$$

En particulier, si $X \sim \mathcal{N}(0, 1)$, alors $\mathbb{E}[X] = 0$ et $\mathrm{Var}(X) = 1$. Notons que cela justifie a posteriori les adjectifs "centrée" et "réduite" pour la loi $\mathcal{N}(0, 1)$.

Démonstration. Montrons d'abord le résultat dans le cas où X suit une loi normale centrée réduite ($\mu = 0$ et $\sigma = 1$). L'espérance est alors donnée par

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} \frac{x}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, \mathrm{d}x \, .$$

La domination $x \, \mathrm{e}^{-x^2} = o(1/x^2)$ en $+\infty$ et $-\infty$ assure que l'intégrale converge bien. La fonction intégrée étant impaire et l'intervalle d'intégration centré en 0, on en déduit que $\mathbb{E}[X] = 0$.

Passons à la variance qui vaut $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2]$. Cette dernière quantité correspond à

$$\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} \frac{x^2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \,\mathrm{d}x\,.$$

où la convergence de l'intégrale résulte du même argument de domination vu précédemment. Une intégration par parties donne alors :

$$\mathbb{E}[X^2] = \left[-\frac{x}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \,\mathrm{d}x = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \,\mathrm{d}x\,,$$

et l'intégrale en (3.1) permet de conclure que $\mathrm{Var}(X) = \mathbb{E}[X^2] = 1$.

Dans le cas général, si $Y \sim \mathcal{N}(\mu, \sigma^2)$, alors la variable aléatoire X définie par $X = \frac{Y-\mu}{\sigma}$ suit une loi normale centrée réduite, ce qui permet de conclure :

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

et

$$\mathrm{Var}(Y) = \mathrm{Var}(\sigma X + \mu) = \sigma^2 \,\mathrm{Var}(X) = \sigma^2\,.$$

□

Plus généralement, les moments d'ordre m d'une variable aléatoire X suivant une loi normale centrée réduite sont nuls pour m impair et valent

$$\mathbb{E}[X^m] = \frac{(2m)!}{2^m\, m!}$$

pour m pair. Ceci peut se montrer par récurrence sur m à l'aide d'une intégration par parties.

Notons qu'en l'absence de primitive explicite pour la densité f de la loi normale, il n'existe pas de forme close pour la fonction de répartition. Pour $X \sim \mathcal{N}(\mu, \sigma^2)$, on peut uniquement écrire

$$\mathbb{P}(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2} \right) \mathrm{d}t\,.$$

La proposition 3.40 permet toutefois de ramener l'étude de X à celle d'une loi normale centrée réduite via

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma} \right) = \Phi\left(\frac{x-\mu}{\sigma} \right),$$

où Φ est la notation usuelle pour la fonction de répartition de la loi normale centrée réduite :

$$\forall x \in \mathbb{R}\,, \quad \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}} \,\mathrm{d}t\,,$$

voir Figure 3.6 pour une représentation graphique de cette fonction. Ce sont les valeurs de cette fonction qui sont données dans les tables.

Une autre propriété importante vérifiée par les variables aléatoires gaussiennes est leur stabilité par addition. C'est l'objet de la proposition suivante dont la preuve est admise.

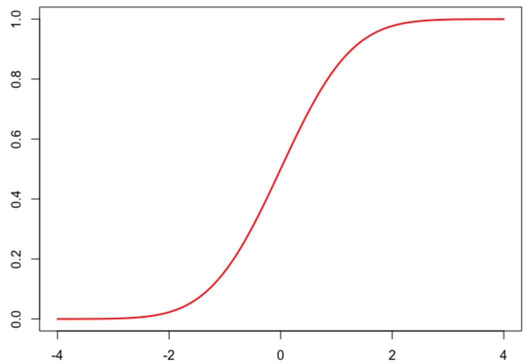


FIGURE 3.6 – Fonction de répartition Φ de la loi normale $\mathcal{N}(0,1)$.

Proposition 3.42 (Somme de variables aléatoires gaussiennes indépendantes)**.** *Soit $X_1,\dots,X_n$ des variables aléatoires indépendantes suivant des lois normales de paramètres respectifs $\mu_1,\dots,\mu_n$ et $\sigma_1,\dots,\sigma_n$. Alors la somme $X_1+\dots+X_n$ suit une loi gaussienne de paramètres $\mu_1+\dots+\mu_n$ et de variance $\sqrt{\sigma_1^2+\dots+\sigma_n^2}$:*

$$X_1+\dots+X_n \sim \mathcal{N}\Big(\mu_1+\dots+\mu_n, \sigma_1^2+\dots+\sigma_n^2\Big)\,.$$

En particulier si $i=1,\dots,n$, on a $\mu_i=\mu\in\mathbb{R}$ et $\sigma_i=\sigma>0$, alors $X_i\mathcal{N}(\mu,\sigma^2)$ et la somme a pour loi

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)\,. \tag{3.2}$$

Il faut retenir qu'une somme de variables aléatoires X_i indépendantes de loi gaussienne suit encore une loi gaussienne. Pour retrouver les paramètres de la somme, c'est-à-dire son espérance et sa variance, il suffit de sommer les espérances des X_i et, puisque les X_i sont indépendantes, de sommer les variances des X_i. On insiste : il faut sommer les variances, c'est-à-dire les σ_i^2 et pas les σ_i.

Bilan loi continues

Loi	Uniforme	Exponentielle	Normale	Normale centrée réduite
Support	$[a,b]$	$\mathbb{R}_+$	$\mathbb{R}$	$\mathbb{R}$
Densité	$f(x)=\frac{1}{b-a}$	$f(x)=\lambda e^{-\lambda x}$	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$
Espérance	$\frac{a+b}{2}$	$\frac{1}{\lambda}$	μ	0
Variance	$\frac{(b-a)^2}{12}$	$\frac{1}{\lambda^2}$	σ^2	1

3.4 Théorème central limite

La loi normale est fondamentale car elle apparaît dans un des théorèmes les plus importants en probabilités : le théorème central limite, souvent abrégé en TCL.

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes et de même loi (on écrit souvent "iid" pour indépendantes et identiquement distribuées) admettant une variance. Il faut voir cette suite comme la répétition d'une expérience aléatoire dans les mêmes conditions. On note S_n la somme partielle des n premières variables aléatoires :

$$S_n = X_1 + \cdots + X_n\,.$$

L'espérance de S_n vaut

$$\mathbb{E}[S_n] = \mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = n\mathbb{E}[X_1]\,,$$

tandis que la variance de S_n vaut

$$\mathrm{Var}[S_n] = \mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = n\,\mathrm{Var}(X_1)\,,$$

où la linéarité résulte ici de l'indépendance des X_i. SI on centre et on réduit S_n on obtient la variable aléatoire

$$\frac{S_n - n\mathbb{E}[X_1]}{\sqrt{n\,\mathrm{Var}(X_1)}}\,.$$

Le théorème central limite assure alors qu'asymptotiquement cette variable aléatoire centrée réduite suit une loi normale $\mathcal{N}(0,1)$. Ceci s'écrit formellement

$$\frac{S_n - n\mathbb{E}[X_1]}{\sqrt{n\,\mathrm{Var}(X_1)}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1)\,,$$

où $\xrightarrow[n\to+\infty]{\mathcal{L}}$ désigne la "convergence en loi", une notion qui dépasse largement le cadre de ce cours. Avec nos outils, on peut écrire cette convergence sous la forme

$$\mathbb{P}\left(a \leq \frac{S_n - n\mathbb{E}[X_1]}{\sqrt{n\,\mathrm{Var}(X_1)}} \leq b\right) \to \int_a^b \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x\,,$$

pour tout $a \leq b$. Cela signifie que la probabilité que la variable aléatoire centrée réduite issue de S_n appartienne à l'intervalle $[a,b]$ converge vers la probabilité qu'une variable aléatoire de loi $\mathcal{N}(0,1)$ appartienne à cet intervalle.

En pratique Il faut retenir que lorsque n est grand (souvent $n \geq 30$ fera l'affaire même si cela dépend du contexte), une somme de variables aléatoires iid se comporte approximativement comme une loi normale. Le centrage se détermine alors en calculant l'espérance $\mathbb{E}[S_n] = n\mathbb{E}[X_1]$ et le dénominateur correspond à la racine de la variance $\mathrm{Var}(S_n) = n\,\mathrm{Var}(X_1)$. La variable aléatoire centrée réduite se comporte alors approximativement comme une loi normale $\mathcal{N}(0,1)$ quand n est grand, ce que l'on écrit

$$\frac{S_n - n\mathbb{E}[X_1]}{\sqrt{n\,\mathrm{Var}(X_1)}} \approx \mathcal{N}(0,1)\,.$$

Une autre manière d'écrire ce résultat consiste à écrire

$$S_n \approx \mathcal{N}(n\mathbb{E}[X_1], n\,\mathrm{Var}(X_1))\,.$$

Remarque 3.43. Il faut se rendre compte des hypothèses très faibles d'un tel théorème : seule l'existence de la variance est requise. Si cette hypothèse est vérifiée, le résultat est alors valable pour n'importe quel type de loi : discrète, à densité, etc.

Exemple 3.44. On lance 100 fois un dé et on note X_i la variable aléatoire donnant 1 si le i-ème lancer donne un 6 et 0 sinon. La variable aléatoire $S_n = X_1 + \cdots + X_n$ correspond alors au nombre de 6 obtenus. Cette variable aléatoire suit une loi binomiale de paramètre $n = 100$ et $p = \frac{1}{6}$. Son espérance vaut $\mathbb{E}[S_n] = \frac{100}{6}$ et sa variance $\mathrm{Var}(S_n) = \frac{500}{36}$. Le théorème central limite donne alors l'approximation

$$S_n = X_1 + \cdots + X_n \approx \mathcal{N}\left(\frac{100}{6}, \frac{500}{6}\right).$$

Cette approximation est représentée en Figure 3.7.

Bilan Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes d'espérance μ et de variance σ^2. On pose $S_n = X_1 + \cdots + X_n$.

— Si les X_i suivent une loi normale ($X_i \sim \mathcal{N}(\mu, \sigma^2)$), alors la proposition 3.42 assure que la somme S_n suit une loi normale :

$$S_n \sim \mathcal{N}(n\mu, n\sigma^2)\,.$$

— Si les X_i sont iid de loi quelconque de variance finie $\sigma^2 > 0$ et d'espérance $\mu \in \mathbb{R}$, alors la somme S_n suit approximativement une loi normale quand n est grand :

$$S_n \approx \mathcal{N}(n\mu, n\sigma^2)\,.$$

On prendre bien garde à la différence entre le symbole $\sim$ qui signifie "suit la loi" et le symbole $\approx$ qui signifie "peut être approché par la loi".

Exemple 3.45. On note $X_1, \ldots, X_{100}$ les poids de 100 étudiants tirés au hasard que l'on modélise par des lois normales d'espérance 75kg et d'écart-type 2kg. On suppose que les poids des étudiants sont indépendants. On obtient alors la loi de la somme $S_n = X_1 + \cdots + X_n$:

$$S_n \sim \mathcal{N}(7500, 400)\,.$$

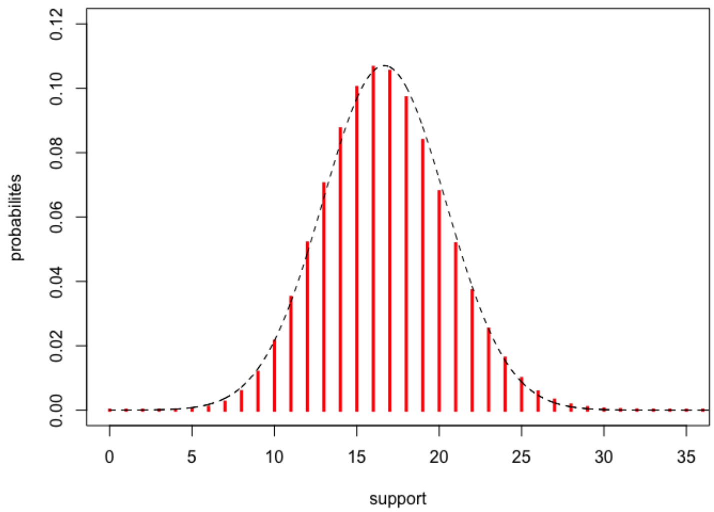


FIGURE 3.7 – Approximation de la loi $S_n \sim \mathcal{B}(100, 1/6)$ (bâtons rouges) par une loi normale $\mathcal{N}(100/6, 500/6)$ (courbe en pointillé).

Supposons désormais que l'on ne connaisse pas la loi des X_i et que la seule information disponible est le poids moyen des étudiants qui est de $\mu = 75$kg et l'écart-type de ces poids qui est de $\sigma = 2$kg. Dans ce cas, on peut déterminer approximativement la loi de la somme S_n. En effet, si on considère toujours les poids comme indépendants, comme $n = 100$ est grand, le théorème central limite donne l'approximation

$$S_n \approx \mathcal{N}(7500, 400)\,.$$

C'est le même résultat qu'avant, mais c'est un résultat approché.