

Chapitre 2

Notions de base en probabilités

JEAN-JIL DUCHAMPS*

ISIFC 2ème année, Statistiques pour l'ingénieur, 2020-2021

N.B. : dans ce cours, les définitions ne sont rigoureuses que pour les espaces de probabilités discrets. On abordera les variables aléatoires continues dans la Section 3, mais sans parler d'ensembles et de fonctions mesurables, ce qui demanderait l'introduction de beaucoup plus de théorie au préalable.

1 Éléments de probabilités

1.1 Probabilités sur un espace fini, événements

1.1.a) Un peu de vocabulaire

Expérience aléatoire, épreuve tirer une carte d'un jeu de 32 cartes ; durée de vie d'un appareil, etc.

Univers (Ω), espace fondamental ensembles des résultats possibles. Exemples :

- $\Omega = \{As\spadesuit, Valet\spadesuit, Dame\spadesuit, \dots\} \rightsquigarrow \text{Card}(\Omega) = 32.$
- $\Omega = \mathbb{R} \rightsquigarrow \text{Card}(\Omega) = \infty.$

Éventualité un élément de Ω (une carte, un réel, etc.).

Événement un sous-ensemble $E \subset \Omega$. Exemple : l'événement *tirer un trèfle* correspond à l'ensemble $\{7\clubsuit, 8\clubsuit, \dots, Roi\clubsuit, As\clubsuit\}$. Si E contient l'issue de l'épreuve, on dit que E est **réalisé**.

Événement élémentaire un ensemble qui contient une seule éventualité (ex : *tirer l'as de trèfle*).

Événement impossible aucune éventualité, l'ensemble vide \emptyset .

Événement certain contient toutes les éventualités : Ω .

Événements incompatibles/mutuellement exclusifs ensembles A et B disjoints, c'est-à-dire que $A \cap B = \emptyset$ (ex : *tirer un trèfle* et *tirer un carreau*).

Événement contraire l'événement contraire de A est son complémentaire dans Ω , soit $\bar{A} = \Omega \setminus A$.

*Laboratoire de mathématiques de Besançon, jean-jil.duchamps@univ-fcomte.fr.

1.1.b) Loi de probabilité : définition et propriétés

Définition 1 (Mesure de probabilité). Soit (Ω, \mathcal{A}) un univers muni d'événements – en pratique on considère $\mathcal{A} = \mathcal{P}(\Omega)$. On appelle **mesure de probabilité** une application $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ avec

(i) $\mathbb{P}(\Omega) = 1$

(ii) Si A_1, A_2, \dots est une suite d'événements mutuellement exclusifs – c'est-à-dire si $A_i \cap A_j = \emptyset$ pour tous $i \neq j$ – on a

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

Le couple (Ω, \mathbb{P}) définit un **espace probabilisé**. Ceci *modélise* une expérience aléatoire. Sous cette définition, un événement A se réalise avec probabilité $\mathbb{P}(A) \in [0, 1]$.

- Quel sens donner à cette phrase? Intuitivement, l'idée est que si l'on répète l'expérience un grand nombre de fois, la *proportion* des fois où l'événement A se sera réalisé sera environ $\mathbb{P}(A)$.
- Si $\mathbb{P}(A) = 1$, l'événement se réalise à chaque répétition de l'expérience.
- Si $\mathbb{P}(A) = 0$, l'événement ne se réalise jamais.

Premières propriétés (importantes)

Proposition 2.

(i) $\mathbb{P}(\emptyset) = 0$ (l'événement \emptyset est impossible).

(ii) Si $A \subset B$, on a $\mathbb{P}(A) \leq \mathbb{P}(B)$.

(iii) Si A est fini, sa probabilité est égale à la somme des probabilités de tous ses événements élémentaires :

$$A = \{a_1, a_2, \dots, a_k\} \implies \mathbb{P}(A) = \mathbb{P}(\{a_1\}) + \mathbb{P}(\{a_2\}) + \dots + \mathbb{P}(\{a_k\}).$$

Abus de notation : pour une éventualité $a \in \Omega$, on écrira souvent $\mathbb{P}(a)$, voire $p(a)$ dans certains contextes, à la place de $\mathbb{P}(\{a\})$. Alors la dernière formule de la proposition se note plus simplement $\mathbb{P}(A) = \mathbb{P}(a_1) + \mathbb{P}(a_2) + \dots + \mathbb{P}(a_k)$.

Conséquence du dernier point de la proposition : si Ω est un univers fini, il suffit de s'intéresser aux événements élémentaires pour connaître la probabilité de tout événement. Cela reste vrai pour un univers discret (i.e. dénombrable, par exemple $\Omega = \mathbb{N}$), puisqu'on peut alors énumérer toutes les éventualités, mais attention, c'est *faux* pour un univers continu (quand $\Omega = \mathbb{R}$).

Proposition 3. Soit (Ω, \mathbb{P}) un espace probabilisé, et $A, B \subset \Omega$ deux événements. Alors on a

(i) $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$.

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

1.1.c) Équiprobabilité

Définition 4. Soit (Ω, \mathbb{P}) un espace probabilisé, avec Ω fini. On dit que (Ω, \mathbb{P}) est **équiprobable** lorsque tous les événements élémentaires ont la même probabilité. Autrement dit, si $\Omega = \{e_1, \dots, e_n\}$, on a

$$\text{pour tout } i \in \{1, \dots, n\}, \quad \mathbb{P}(e_i) = \frac{1}{n}.$$

Proposition 5. Soit (Ω, \mathbb{P}) un espace probabilisé équiprobable. Alors, pour tout $A \subset \Omega$, on a

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

1.2 Probabilités conditionnelles, indépendance

1.2.a) Probabilités conditionnelles

Exemple introductif On jette un dé jusqu'à obtenir un numéro pair. Quelle est la probabilité d'obtenir un 2 ?

\leadsto Modélisation de l'expérience par $\Omega = \{1, 2, 3, 4, 5, 6\}$ équiprobable. Intuitivement, on se ramène à considérer le sous-ensemble des pairs $\Omega_p := \{2, 4, 6\}$. Les événements étant équiprobables, la probabilité que le premier numéro pair obtenu soit un 2 est $1/\text{Card}(\Omega_p) = 1/3$. On notera

$$\mathbb{P}(\text{on obtient } 2 \mid \text{on obtient un numéro pair}) = \frac{1}{3}.$$

Définition 6 (Probabilité conditionnelle). Soit (Ω, \mathbb{P}) un espace probabilisé et soient A, B deux événements, avec $\mathbb{P}(B) \neq 0$ (donc $\mathbb{P}(B) \in]0, 1]$). La **probabilité conditionnelle** de A sachant B est définie par

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

Remarque 7.

- Si A est disjoint de B , on a $\mathbb{P}(A \mid B) = 0$.
- Intuitivement, $\mathbb{P}(A \mid B)$ représente la proportion des éventualités où A est réalisé parmi celles où B est réalisé.
- Dans un espace équiprobable, on a alors simplement

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

Si l'on revient à l'exemple précédent, on calcule donc bien

$$\begin{aligned} \mathbb{P}(\{2\} \mid \{2, 4, 6\}) &= \frac{\text{Card}(\{2\} \cap \{2, 4, 6\})}{\text{Card}(\{2, 4, 6\})} \\ &= \frac{\text{Card}(\{2\})}{\text{Card}(\{2, 4, 6\})} \\ &= \frac{1}{3}, \end{aligned}$$

ce qui justifie formellement le résultat donné plus haut.

La proposition suivante est fondamentale.

Proposition 8 (Formule des probabilités totales). Soit (Ω, \mathbb{P}) un espace probabilisé, et B_1, B_2, \dots une suite d'événements mutuellement exclusifs (c'est-à-dire tels que $B_i \cap B_j = \emptyset$ pour tous $i \neq j$), tels que :

- pour tout $i \geq 1$, $\mathbb{P}(B_i) \neq 0$.

- $\bigcup_{i \geq 1} B_i = \Omega$.

Alors pour tout événement $A \subset \Omega$, on a

$$\mathbb{P}(A) = \sum_{i \geq 1} \mathbb{P}(A | B_i) \mathbb{P}(B_i).$$

On note que sous les hypothèses précédentes, les B_i définissent une partition de l'univers Ω . C'est-à-dire que chaque éventualité $\omega \in \Omega$ peut être classée dans exactement un seul des B_i (il existe un unique $i \geq 1$ tel que $\omega \in B_i$). Cette formule se comprend bien quand les probabilités $\mathbb{P}(A | B_i)$ sont facilement exprimables.

Exemple 9. On modélise une expérience aléatoire définie de la manière suivante :

- Le jour suivant, il y a 50% de chance de faire beau, 30% de chance de pluie et 20% de chance de tempête.
- Je me fie à Météo France et prend un parapluie quand il est prévu de la pluie.
- Or on connaît bien les erreurs de Météo France :
 - S'il fait beau le lendemain, on prévoit de la pluie dans 15% des cas.
 - S'il pleut le lendemain, on prévoit de la pluie dans 90% des cas.
 - Si c'est la tempête le lendemain, on prévoit de la pluie dans 95% des cas.

Quelle est la probabilité que je sois mouillé le lendemain ?

Solution : en utilisant la formule des probabilités totales, on a

$$\begin{aligned} \mathbb{P}(\text{mouillé}) &= 0 \times \mathbb{P}(\text{beau temps}) + \frac{10}{100} \times \mathbb{P}(\text{pluie}) + \frac{5}{100} \times \mathbb{P}(\text{tempête}) \\ &= 0 \times 0.5 + 0.1 \times 0.3 + 0.05 \times 0.2 \\ &= 0.04, \end{aligned}$$

j'ai donc 4% de chance d'être mouillé de lendemain.

On voit donc que la notion de probabilités conditionnelles est assez intuitive ; on peut en effet leur donner du sens et les utiliser dans des situations réelles. La proposition suivante permet de calculer de nouvelles probabilités conditionnelles.

Proposition 10 (Règle de Bayes). Soit (Ω, \mathbb{P}) un espace probablisé, et B_1, B_2, \dots une suite d'événements mutuellement exclusifs (c'est-à-dire tels que $B_i \cap B_j = \emptyset$ pour tous $i \neq j$), tels que :

- pour tout $i \geq 1$, $\mathbb{P}(B_i) \neq 0$.
- $\bigcup_{i \geq 1} B_i = \Omega$.

Alors pour tout événement $A \subset \Omega$ tel que $\mathbb{P}(A) \neq 0$, on a

$$\forall i \geq 1, \quad \mathbb{P}(B_i | A) = \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A | B_i) \mathbb{P}(B_i)}{\sum_{j \geq 1} \mathbb{P}(A | B_j) \mathbb{P}(B_j)}.$$

On voit que l'on a simplement utilisé la définition de l'espérance conditionnelle et la formule des probabilités totales.

Revenons à l'exemple précédent : quelle la probabilité qu'il fasse beau sachant que Météo France prévoit du beau temps ?

Solution : avec la règle de Bayes, on a (exercice : compléter les trous)

$$\begin{aligned} \mathbb{P}(\text{beau temps} \mid \text{beau temps prévu}) &= \dots = \frac{0.85 \times 0.5}{0.85 \times 0.5 + 0.1 \times 0.3 + 0.05 \times 0.2} \\ &= \frac{0.425}{0.465} \simeq 0.91, \end{aligned}$$

donc s'il est prévu du beau temps, ce sera vérifié dans 91% des cas.

1.2.b) Indépendance

En général, le fait de savoir qu'un événement B est réalisé influe sur la probabilité qu'un autre événement A se réalise (par exemple, si Météo France prévoit du beau temps, on a beaucoup plus de chance d'avoir du beau temps que si l'on ne sait rien). Ceci se traduit pas le fait que $\mathbb{P}(A \mid B) \neq \mathbb{P}(A)$ dans beaucoup de situations.

Mais il existe des cas où $\mathbb{P}(A \mid B) = \mathbb{P}(A)$. C'est le cas si deux événements "ne peuvent pas influencer l'un sur l'autre", par exemple :

A *l'équipe de France gagne la prochaine coupe du monde.*

B *les manchots empereurs souffriront d'un épidémie virale sévère dans les 10 prochaines années.*

\leadsto on a envie de dire que ces événements sont indépendants.

Remarque 11. Si $\mathbb{P}(A \mid B) = \mathbb{P}(A)$, alors on a

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

On définit donc l'indépendance de la manière suivante.

Définition 12. Soit (Ω, \mathbb{P}) un espace probabilisé, et A, B deux événements de Ω . On dit que A et B sont **indépendants** si

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

Exemple 13. On jette deux pièces et on suppose équiprobable les quatre résultats possibles. Soit A l'événement *la première pièce montre pile* et B l'événement *la deuxième pièce montre face*. Intuitivement, ces deux événements doivent être indépendants. On le montre en calculant :

- $\mathbb{P}(A \cap B) = \frac{1}{4}$ (une seule éventualité possible)
- $\mathbb{P}(A)\mathbb{P}(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

Il y a égalité, donc A et B sont bien indépendants.

Notons que l'on a $\mathbb{P}(A \cap \bar{B}) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$, donc si A et B sont indépendants, on a alors

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

On a donc le résultat suivant.

Proposition 14. Soit (Ω, \mathbb{P}) un espace probabilisé, et A, B deux événements de Ω . Si A et B sont indépendants, alors A et \bar{B} le sont aussi.

Par symétrie, ce résultat montre que sous la même hypothèse, on a également que \bar{A} et B sont indépendants, et que \bar{A} et \bar{B} sont indépendants.

2 Variables aléatoires et lois de probabilités discrètes

2.1 Motivation

La notion de variable aléatoire est fondamentale en probabilités et en statistiques. C'est une notion simple mathématiquement mais avec des subtilités conceptuelles.

Exemple introductif Épreuve : lancer de dé équilibré à 6 faces. Quel univers choisir ?

- Choix 1 : $\Omega_1 := \{1, 2, 3, 4, 5, 6\}$ chaque éventualité représente le chiffre apparaissant sur la face supérieure du dé à la fin du lancer.
- Choix 2 : $\Omega_2 := \{\text{ensemble des trajectoires possibles du dé depuis la main jusqu'à la table}\}$.
 $\sim \omega \in \Omega_2 =$ fonction à variable t le temps et à valeur vectorielles... compliqué ! Mais de chaque trajectoire ω on peut extraire certaines quantités simples et intéressantes :
 - $X(\omega)$ le chiffre de la face supérieure du dé.
 - $T(\omega)$ le temps écoulé jusqu'à l'arrêt du dé.
 - $D(\omega)$ la distance à l'origine de la trajectoire, etc.

On s'intéresse à la répartition (*moyenne, fluctuations, etc.*) des réalisations de chaque quantité lorsque l'épreuve est répétée un grand nombre de fois \sim peut-on prédire le comportement du système ?

En pratique, on se place donc dans un univers de réalisations Ω compliqué et qu'on ne peut peut-être pas décrire complètement, mais que l'on décrit à travers des *quantités observées* ($X, T, D, \text{etc.}$) qui prennent différentes valeurs selon la réalisation ω . Ce sont ces quantités que l'on appelle *variables aléatoires* et que l'on va maintenant toujours manipuler en statistiques.

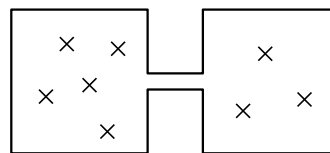
2.2 La notion de variable aléatoire

2.2.a) Un exemple classique de physique statistique

On considère n particules de gaz réparties dans une boîte à deux compartiments symétriques qui communiquent.

On modélise la position de la i -ème particule ($i = 1, \dots, n$) en définissant $x_i \in \{1, 2\}$ suivant si elle est dans le premier compartiment ou le deuxième.

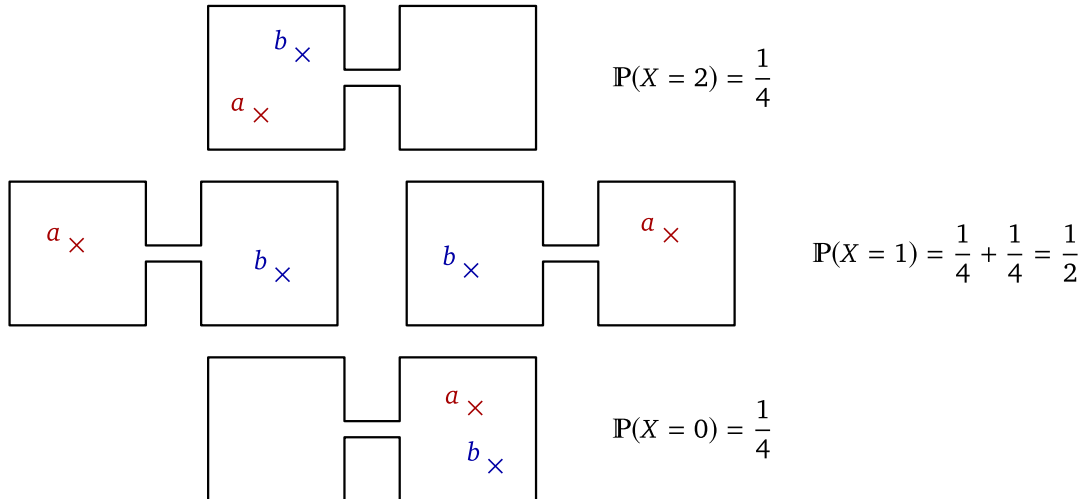
Nombre de configurations possibles : 2^n . Or n peut être très grand. Par exemple, pour deux grammes de dihydrogène H_2 , on a déjà environ une mole, soit $n = N_A = 6.022 \times 10^{23}$ particules. Les 2^n configurations ont chacune une probabilité 2^{-n} de se produire. On ne peut pas toutes les considérer séparément (c'est impossible de même les énumérer).



Répartition correspondant à la configuration
 $(x_i)_{i=1}^n = (1, 1, 2, 2, 1, 1, 1, 2)$.

→ *Modèle probabiliste*. On définit $X \in \{0, \dots, n\}$ comme le nombre de particules dans le compartiment premier compartiment (à gauche). X est une variable aléatoire qui dépend de la répartition (au hasard) des particules.

Exemple pour $n = 2$.



Il y a quatre configurations possibles, et elles sont équiprobables. On trouve la “répartition” de X (*mesure de probabilité* sur $\{0, 1, 2\}$) en additionnant les probabilités des configurations correspondantes à chaque valeur de X .

2.2.b) Définition

Définition 15 (Variable aléatoire). Soit (Ω, \mathbb{P}) un espace probabilitisé. Une variable aléatoire (réelle), abrégé v.a., est une application

$$X : \begin{cases} \Omega & \longrightarrow \mathbb{R} \\ \omega & \longmapsto X(\omega). \end{cases}$$

Vocabulaire

- Si $X(\Omega) = \{X(\omega), \omega \in \Omega\}$ (ensemble image de Ω par X) est fini ou dénombrable, alors la v.a. X est dite **discrète**. Exemple : résultat d’un lancer de dé, $X(\Omega) = \{1, 2, 3, 4, 5, 6\}$.
- Si $X(\Omega)$ n’est pas dénombrable, alors la v.a. X est dite **continue**. Exemple : durée totale du lancer de dé, $T(\Omega) = \mathbb{R}_+$.

Dans ce chapitre, on ne s’intéressera qu’aux variables aléatoire discrètes.

Remarque 16. Contrairement à ce que l’on a fait dans nos premiers exemples, Ω n’a plus besoin d’être décrit très précisément. Ce sont sur les v.a. X, Y , etc. que l’on va travailler directement.

Opérations sur les variables aléatoires Il est important d’évoquer une manière de *créer* de nouvelles variables aléatoires. Supposons que l’on dispose de X_1, X_2, \dots, X_k des variables aléatoires, et d’une fonction $\varphi : \mathbb{R}^k \rightarrow \mathbb{R}$. Alors l’application

$$\begin{cases} \Omega & \longrightarrow \mathbb{R} \\ \omega & \longmapsto \varphi(X_1(\omega), \dots, X_k(\omega)) \end{cases}$$

définit bien une variable aléatoire, que l'on note alors $\varphi(X_1, \dots, X_k)$.

En particulier, en appliquant ceci à $\varphi_1(x, y) = x + y$ ou $\varphi_2(x, y) = xy$, on voit que si X et Y sont des variables aléatoires, $Z_1 := X + Y$ ou $Z_2 := XY$ en sont aussi.

Enfin, notons que l'application $\omega \mapsto c$, pour $c \in \mathbb{R}$ fixé, définit une variable aléatoire constante (étonnant ?). Ainsi, on peut retenir que *les constantes sont des variables aléatoires*.

2.3 Loi de probabilité et fonction de répartition

Dans toute la section, (Ω, \mathbb{P}) désigne un espace probabilisé et X une variable aléatoire. $X(\Omega)$, l'ensemble image de Ω par l'application X est un sous-ensemble de \mathbb{R} que l'on peut considérer comme un *nouvel univers*.

Définition 17 (Loi de probabilité). À chaque événement $B \subset X(\Omega)$ on associe une valeur notée $\mathbb{P}_X(B)$ et définie comme

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

L'application \mathbb{P}_X définit une mesure de probabilité sur $X(\Omega)$ appelée la **loi (ou distribution) de X** .

Notation importante : on notera souvent $\mathbb{P}_X(B) = \mathbb{P}(X \in B)$. On s'autorise aussi différentes notations qui vont de paire avec nos intuitions sur les probabilités (*quelle est la probabilité que X soit égal à k ? compris entre a et b ?*), par exemple :

- si B est un singleton $\{k\}$, on pourra noter $\mathbb{P}_X(B) = \mathbb{P}(X = k)$.
- si B est un intervalle $[a, b]$, on pourra noter $\mathbb{P}_X(B) = \mathbb{P}(a \leq X \leq b)$.

Exemple 18. Épreuve : on lance deux dés, et on définit X comme la v.a. qui compte la somme des points des deux faces supérieures des dés. Quelle est la probabilité que cette somme fasse 5 ?

Ici on a $B = \{5\}$, et l'univers Ω peut être défini comme l'ensemble des paires (n, m) de résultats de chaque dé, soit $\Omega := \{1, 2, 3, 4, 5, 6\}^2$, avec l'équiprobabilité de chaque éventualité. Si $\omega = (n, m)$, on a $X(\omega) = n + m$, on peut donc calculer :

$$\begin{aligned} \mathbb{P}(X = 5) &= \mathbb{P}_\Omega((1, 4)) + \mathbb{P}_\Omega((2, 3)) + \mathbb{P}_\Omega((3, 2)) + \mathbb{P}_\Omega((4, 1)) \\ &= 4 \times 1/36 = 1/9. \end{aligned}$$

Remarque 19.

- $\{X = 5\}$ est un événement élémentaire (éventualité) de $X(\Omega)$ MAIS $\{\omega \in \Omega \mid X(\omega) = 5\} = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$ ne l'est pas.
- Dans cet exemple, (Ω, \mathbb{P}) est équiprobable, MAIS $(X(\Omega), \mathbb{P}_X)$ ne l'est pas !
- On peut s'intéresser à plusieurs variables aléatoires. Si X et Y sont deux v.a., on notera par exemple

$$\mathbb{P}(X \in A, Y \in B) := \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A \text{ et } Y(\omega) \in B\}).$$

À noter qu'il existe une notion de loi de probabilité pour un couple (ou plus généralement un n -uplet ou une suite) de variables aléatoires, mais on ne rentrera pas dans ces détails.

Pour continuer avec notre exemple, on aimerait calculer la probabilité que la somme des deux faces soit inférieure ou égale à 5. Pour cela, on introduit la notion de fonction de répartition.

Définition 20 (Fonction de répartition). On appelle **fonction de répartition** de X et on note F_X la fonction

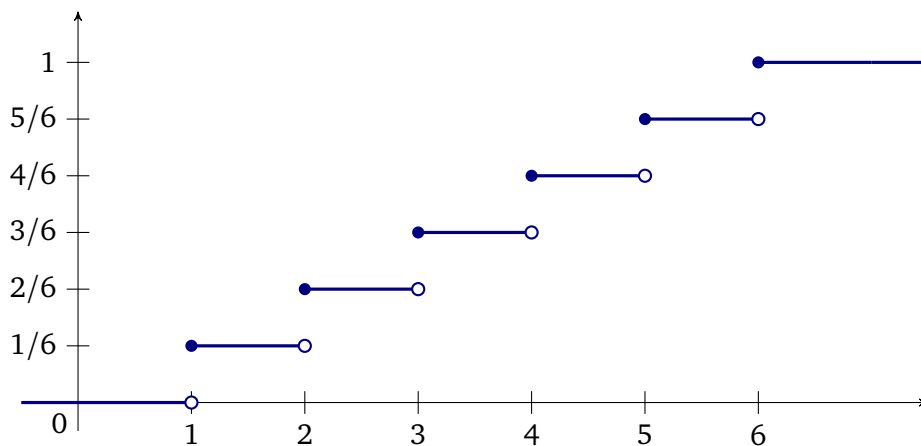
$$F_X : \begin{cases} \mathbb{R} & \longrightarrow [0, 1] \\ t & \longmapsto \mathbb{P}(X \leq t), \end{cases}$$

où l'on utilise la convention de notation $\mathbb{P}(X \leq t) = \mathbb{P}_X(]-\infty, t])$.

Exemple 21. *Épreuve* : lancer d'un dé à six faces. $X : \Omega \rightarrow \{1, 2, 3, 4, 5, 6\}$ résultat du jet.
Loi de probabilité de X : équiprobable, $\mathbb{P}(X = i) = 1/6$ pour tout i .

Construisons F_X :

- pour $t < 1$, $F_X(t) = \mathbb{P}(X \leq t) \leq \mathbb{P}(X < 1) = 0$.
- pour $t \in [1, 2[$, $F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}(X = 1) = 1/6$.
- pour $t \in [i, i + 1[$ avec $1 \leq i \leq 5$, $F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}(X = i) = i/6$.
- pour $t \geq 6$, $F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}(X \leq 6) = 1$.



Graphe de F_X .

Proposition 22 (Propriétés de la fonction de répartition).

(i) F_X est continue à droite, croissante, et satisfait

$$\lim_{t \rightarrow -\infty} F_X(t) = 0, \quad \lim_{t \rightarrow \infty} F_X(t) = 1.$$

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

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

Attention à la place des comparateurs $<$ et \leq dans le dernier point ci-dessus.

Remarque 23. La loi d'une v.a. **discrète** est complètement déterminée par :

- sa fonction de masse ($\mathbb{P}(X = s)$, $s \in X(\Omega)$).
- sa fonction de répartition.

2.4 Espérance et variance

On se donne une v.a. discrète $X : \Omega \rightarrow X(\Omega) \subset \mathbb{R}$, avec $X(\Omega)$ dénombrable, que l'on peut alors énumérer :

$$X(\Omega) = \{s_1, s_2, \dots, s_k, \dots\} = \{s_k, k \in \mathbb{K}\},$$

où $\mathbb{K} \subset \mathbb{N}$ (avec éventuellement $\mathbb{K} = \mathbb{N}$) et $s_k \in \mathbb{R}$ pour tout $k \in \mathbb{K}$.

Définition 24 (Espérance mathématique). On dit qu'une variable aléatoire X admet une espérance si la somme suivante converge (c'est toujours le cas si $X(\Omega)$ est fini) :

$$\sum_{k \in \mathbb{K}} |s_k| p_k, \quad \text{avec } p_k = \mathbb{P}(X = s_k).$$

Dans ce cas, l'espérance de X est le réel, noté $\mathbb{E}[X]$ défini par :

$$\mathbb{E}[X] = \sum_{k \in \mathbb{K}} s_k p_k.$$

Plus généralement, pour toute application $\varphi : \mathbb{R} \rightarrow \mathbb{R}$, on définit (à condition que la somme ait un sens) :

$$\mathbb{E}[\varphi(X)] = \sum_{k \in \mathbb{K}} \varphi(s_k) p_k.$$

L'espérance formalise l'idée de moyenne : intuitivement, si l'on répète une expérience aléatoire un grand nombre de fois et que l'on mesure à chaque fois la réalisation de la même variable X_1, X_2, \dots , on va observer que la *moyenne* des X_i s'approche de l'espérance de X :

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[X].$$

On va voir ce résultat (la loi des grands nombres) formellement un peu plus loin dans le cours.

Exemples 25.

1. Un premier exemple important : **l'espérance d'une v.a. constante**. Si pour $c \in \mathbb{R}$ fixé on a $X(\omega) = c$ pour tout $\omega \in \Omega$, on a alors $X(\Omega) = \{c\}$ et $\mathbb{P}(X = c) = 1$, donc

$$\mathbb{E}[X] = c \times 1 = c.$$

2. Alice joue aux dés avec Bob. Elle lance un dé et gagne dix euros si elle tire un 5 ou un 6 ; elle perd cinq euros dans tous les autres cas. Qui d'Alice ou de Bob est avantagé par ces règles ?

Réponse : ça dépend comment on le voit, mais du point de vue de l'espérance des gains d'Alice, on a

$$\mathbb{E}[\text{Gains}_A] = 10 \times \frac{2}{6} - 5 \times \frac{4}{6} = 0.$$

Le jeu est donc équitable (mais pas très intéressant peut-être).

Remarque 26. Pour un univers Ω fini ou dénombrable, on a par définition

$$p_k = \mathbb{P}(X = s_k) = \sum_{\substack{\omega \in \Omega: \\ X(\omega) = s_k}} \mathbb{P}(\omega),$$

ainsi on a pour tout $k \in \mathbb{K}$,

$$s_k p_k = s_k \mathbb{P}(X = s_k) = \sum_{\substack{\omega \in \Omega: \\ X(\omega) = s_k}} s_k \mathbb{P}(\omega) = \sum_{\substack{\omega \in \Omega: \\ X(\omega) = s_k}} X(\omega) \mathbb{P}(\omega).$$

Notons que chaque $\omega \in \Omega$ est dans un seul de ces ensembles $A_k = \{\omega \in \Omega \mid X(\omega) = s_k\}$ (on dit que les A_k forment une partition de Ω). En sommant l'expression précédente sur $k \in \mathbb{K}$, on se ramène donc à sommer l'expression $X(\omega) \mathbb{P}(\omega)$ sur chacun des $\omega \in \Omega$:

$$\mathbb{E}[X] = \sum_{k \in \mathbb{K}} s_k p_k = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega).$$

Cette remarque nous permet de déduire facilement la propriété fondamentale de linéarité de l'espérance (cela découle de la linéarité de la somme elle-même).

Proposition 27 (Linéarité de l'espérance). *Soient X, Y deux variables aléatoires, et a, b deux réels. Alors*

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

Si l'espérance formalise l'idée de moyenne, la variance caractérise les *fluctuations* d'une variable aléatoire *autour de sa moyenne*.

Définition 28 (Variance). La **variance** d'une variable aléatoire (discrète) X est donnée (si la formule suivante est bien définie) par

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

L'écart-type de X est donné par

$$\sigma(X) := \sqrt{\text{Var}(X)}.$$

L'équivalence entre les deux formules données pour la variance se montre en utilisant la linéarité de l'espérance. En effet, si l'on note $m = \mathbb{E}[X]$ pour se rappeler que c'est un réel constant, on a donc

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

En pratique, c'est la deuxième formule, *beaucoup plus commode* à manipuler en général, que l'on utilisera. En utilisant la définition de l'espérance de variables aléatoires, on peut donc écrire pour calculer la variance :

$$\text{Var}(X) = \sum_{k \in \mathbb{K}} s_k^2 p_k - \left(\sum_{k \in \mathbb{K}} s_k p_k \right)^2.$$

Exemple 29. Pour le jeu d’Alice et Bob, on a $\text{Var}(\text{Gains}_A) = 10^2 \times 1/3 + (-5)^2 \times 2/3 = 50$. Cette valeur n’est pas évidente à interpréter. Par contre, l’écart-type a plus facilement du sens (même dimension – au sens du système de mesure physique – que la v.a.) : ici, on a $\sigma(\text{Gains}_A) = \sqrt{50} \simeq 7.07$, ce qui nous indique que le résultat du jeu se trouve *en général* à plus ou moins 7 euros de sa valeur moyenne.

De la même manière que l’espérance apparaît dans la loi des grands nombres, la variance intervient dans le théorème central limite, qui donne *l’ordre de grandeur des fluctuations* de la moyenne d’une suite d’observations autour de l’espérance. On reviendra sur ce point à la fin du chapitre.

Quand on calcule la variance d’une somme de deux variables aléatoires X et Y , un simple calcul nous donne

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

Ceci nous amène à définir la **covariance** de X et Y par

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

Proposition 30 (Propriétés de la (co)variance). *Si X, Y, Z sont des variables aléatoires et a et b sont des réels, on a :*

(i) $\text{Var}(aX) = a^2 \text{Var}(X)$.

(ii) $\text{Var}(X) = \text{Cov}(X, X)$.

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

(iv) $\text{Cov}(X, Y) = \frac{1}{2}(\text{Var}(X + Y) - \text{Var}(X) - \text{Var}(Y)) = \frac{1}{4}(\text{Var}(X + Y) - \text{Var}(X - Y))$.

(v) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.

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

De manière formelle (hors programme), ces propriétés nous disent que la variance est une forme quadratique, et que la covariance est la forme bilinéaire associée (dans ce contexte, le point (iv) de la proposition est appelé *identité de polarisation*).

2.5 Indépendance de variables aléatoires

Rappel : Deux événements A et B sont indépendants si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. On a donc indépendance si le “et” logique de l’intersection se traduit en produit lors de la mesure de probabilité. La notion d’indépendance de variables aléatoires suit la même logique.

Définition 31 (Indépendance de variables aléatoires). Soit (Ω, \mathbb{P}) un espace probabilisé et X, Y deux variables aléatoires. On dit que X est **indépendant de** Y , et on note $X \perp Y$, si pour tout $A \subset X(\Omega)$, pour tout $B \subset Y(\Omega)$, on a

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

En fait, cette propriété est équivalente à une propriété en apparence plus forte.

Proposition 32. *Deux v.a. X et Y sont **indépendantes** si et seulement si pour toutes fonctions $\varphi, \psi : \mathbb{R} \rightarrow \mathbb{R}$, on a*

$$\mathbb{E}[\varphi(X) \psi(Y)] = \mathbb{E}[\varphi(X)] \mathbb{E}[\psi(Y)].$$

Ce n'est pas immédiat et l'on ne le montre pas ici, mais c'est souvent ce qu'on va utiliser en pratique. En particulier, l'espérance du produit de deux variables indépendantes s'exprime simplement $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Ce qui est valable pour deux v.a. est aussi valable pour k v.a. X_1, \dots, X_k . De la même façon, on dira qu'elles sont indépendantes si pour tous $A_1, \dots, A_k \subset \mathbb{R}$ on a

$$\mathbb{P}(X_1 \in A_1, \dots, X_k \in A_k) = \prod_{i=1}^k \mathbb{P}(X_i \in A_i),$$

ou de manière équivalente, si pour toutes fonctions $\varphi_1, \dots, \varphi_k : \mathbb{R} \rightarrow \mathbb{R}$, on a

$$\mathbb{E}\left[\prod_{i=1}^k \varphi_i(X_i)\right] = \prod_{i=1}^k \mathbb{E}[\varphi_i(X_i)].$$

Proposition 33. Soient X et Y deux variables aléatoires indépendantes. Alors $\text{Cov}(X, Y) = 0$.

La preuve tient en deux lignes : $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$, or par l'indépendance de X et Y , on a $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, donc la covariance est nulle.

Attention, la réciproque de ce résultat n'est pas vraie : deux variables de covariance nulle ne sont pas toujours indépendantes (on pourra évoquer des contre-exemples en TD).

Les propriétés suivantes sont fondamentales pour des variables aléatoires indépendantes.

Proposition 34. Soient X_1, \dots, X_k des variables aléatoires indépendantes. Alors on a

- (i) $\mathbb{E}[X_1 X_2 \dots X_k] = \mathbb{E}[X_1] \mathbb{E}[X_2] \dots \mathbb{E}[X_k]$.
- (ii) $\text{Var}(X_1 + X_2 + \dots + X_k) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_k)$.

Le premier point est un cas particulier de la caractérisation de l'indépendance, et le second, qui se révélera **très important** dans la suite, se démontre par récurrence, en utilisant le fait que la covariance de variables indépendantes est nulle (à faire en exercice pour deux v.a. X et Y).

2.6 Quelques lois usuelles

On va maintenant passer en revue quelques lois de probabilité classiques, à connaître absolument pour faire des statistiques.

2.6.a) Loi de Bernoulli

Une épreuve a deux résultats possibles, le *succès* ou l'*échec*, alternative caractérisée par la variable aléatoire X telle que

$$X(\omega) = \begin{cases} 1 & \text{si } \omega \in \{\text{succès}\} \\ 0 & \text{si } \omega \in \{\text{échec}\}. \end{cases}$$

Alors on a $X(\Omega) = \{0, 1\}$, et :

- si l'on note $p = \mathbb{P}(X = 1) \in [0, 1]$,
- on a alors $\mathbb{P}(X = 0) = 1 - p$.

La loi de X est caractérisée par ce nombre p . On dit que X suit une **loi de Bernoulli de paramètre p** , et on note $X \sim \text{Ber}(p)$.

Fonction de répartition :

$$F_X(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 - p & \text{si } 0 \leq t < 1 \\ 1 & \text{si } t \geq 1 \end{cases}$$

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

Variance : on calcule

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= 1^2 \times p + 0^2 \times (1 - p) - p^2 \\ &= p - p^2. \end{aligned}$$

Pour résumer pour la loi de Bernoulli $X \sim \text{Ber}(p)$:

$$\mathbb{P}(X = 1) = p, \quad \mathbb{E}[X] = p, \quad \text{Var}(X) = p(1 - p), \quad \sigma(X) = \sqrt{p(1 - p)}.$$

2.6.b) Loi binomiale

On s'intéresse au nombre de succès lors d'une répétition de n épreuves de Bernoulli identiques (par exemple, on lance une pièce 10 fois et on s'intéresse au nombre de "pile"). Formellement, on se donne n v.a. de Bernoulli X_1, X_2, \dots, X_n , indépendantes et de même paramètre p . La variable aléatoire

$$X := \sum_{i=1}^n X_i$$

compte le nombre de succès parmi les n épreuves.

On veut maintenant calculer la loi de X , qui prend ses valeurs dans $\{0, 1, \dots, n\}$. Soit $k \in \{0, \dots, n\}$. L'événement $\{X = k\}$ est réalisé si le vecteur aléatoire (X_1, \dots, X_n) a k coordonnées à 1 et $n - k$ coordonnées à 0. Soit $\mathbf{x} = (x_1, \dots, x_n) \in \{0, 1\}^n$ un tel vecteur, fixé. On a, par l'indépendance des (X_i) ,

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n \mathbb{P}(X_i = x_i) = p^k (1 - p)^{n-k},$$

puisque il y a dans ce produit k termes qui valent $\mathbb{P}(X_i = 1) = p$ et $n - k$ termes qui valent $\mathbb{P}(X_i = 0) = 1 - p$. De plus, le nombre de vecteurs \mathbf{x} avec k coordonnées à 1 est connu : c'est le nombre de façon de choisir k places parmi n , soit $\binom{n}{k} = n! / (k!(n - k)!)$. Finalement, on a simplement

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

Cette formule caractérise donc la loi de X ; on dit que X suit une **loi binomiale de paramètres (n, p)** et on note $X \sim \text{Bin}(n, p)$.

Fonction de répartition : (dans ce cas, pas très utile de retenir)

$$F_X(t) = \begin{cases} 0 & \text{si } t < 0 \\ \sum_{k=1}^{\lfloor t \rfloor} \binom{n}{k} p^k (1-p)^{n-k} & \text{si } 0 \leq t < n \\ 1 & \text{si } t \geq n \end{cases}$$

Espérance : en utilisant la linéarité de l'espérance,

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}\left[\sum_{i=1}^n X_i\right] \\ &= \sum_{i=1}^n \mathbb{E}[X_i] \\ &= \sum_{i=1}^n p \\ &= np. \end{aligned}$$

Variance : en utilisant le fait que la variance d'une somme de v.a. *indépendantes* est la somme des variances,

$$\begin{aligned} \text{Var}(X) &= \sum_{i=1}^n \text{Var}(X_i) \\ &= np(1-p). \end{aligned}$$

Ainsi pour résumer pour la loi binomiale $X \sim \text{Bin}(n, p)$:

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \mathbb{E}[X] = np, \quad \text{Var}(X) = np(1-p), \quad \sigma(X) = \sqrt{np(1-p)}.$$

2.6.c) Loi géométrique

On s'intéresse cette fois au nombre d'épreuves de Bernoulli nécessaires pour obtenir le premier succès. C'est-à-dire que l'on répète, de manière indépendante, une épreuve de Bernoulli de paramètre p , jusqu'au premier succès. La variable aléatoire X qui donne le "rang" du premier succès est dite suivre une loi géométrique de paramètre p .

Les valeurs de X possibles sont les entiers naturels non nuls $\mathbb{N}^* = \{1, 2, 3, \dots\}$. On a

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

(c'est la probabilité que l'on observe $k - 1$ échecs puis 1 succès). Quand X suit une **loi géométrique de paramètre p** , on note $X \sim \text{Geo}(p)$.

Fonction de répartition :

$$F_X(t) = \begin{cases} 0 & \text{si } t < 1 \\ 1 - (1-p)^k & \text{si } t \in [k, k+1[. \end{cases}$$

Pour cette loi, il est **important (et facile) de retenir que** $\mathbb{P}(X > k) = 1 - F_X(k) = (1 - p)^k$. On retrouve ça facilement : le premier succès arrive strictement après le k -ième essai ssi les k premiers essais sont des échecs.

$$\text{Espérance : } \mathbb{E}[X] = \frac{1}{p}.$$

$$\text{Variance : } \text{Var}(X) = \frac{1-p}{p^2}, \quad \sigma(X) = \frac{\sqrt{1-p}}{p}$$

Un fait important sur la loi géométrique : le **manque de mémoire**, c'est-à-dire qu'on a

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

On peut le retrouver simplement avec un calcul (ou intuitivement, en pensant à ce que ceci veut dire en termes d'échecs d'une suite d'épreuve). Les lois géométriques sont les seules lois à valeurs dans \mathbb{N}^* qui vérifient cette propriété de manque de mémoire.

2.6.d) Loi de Poisson

La loi de Poisson (du nom de Siméon Denis Poisson, 1838) apparaît quand l'on cherche à compter le nombre d'événements *rare*s se produisant dans un *intervalle de temps fixé*. Exemple classique (mais peu réaliste en fait) : "si les bus sont censés passer tous les M minutes, combien de bus vont s'arrêter ici durant le prochain quart d'heure?" Le nombre de bus suivra alors une loi de Poisson.

Définition : soit un réel strictement positif $\lambda \in \mathbb{R}_+^*$ et une v.a. X telle que pour tout $k \in \mathbb{N}$,

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

On dit alors que X suit une **loi de Poisson de paramètre** λ et on note $X \sim \text{Poi}(\lambda)$.

Fonction de répartition : (dans ce cas, pas très utile de retenir)

$$F_X(t) = \begin{cases} 0 & \text{si } t < 0 \\ e^{-\lambda} \sum_{k=0}^{\lfloor t \rfloor} \frac{\lambda^k}{k!} & \text{si } t \geq 0. \end{cases}$$

Les calculs (laissés en exercice) donnent une espérance et une variance de λ . Pour résumer, on a pour la loi de Poisson $X \sim \text{Poi}(\lambda)$:

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad \mathbb{E}[X] = \lambda, \quad \text{Var}(X) = \lambda, \quad \sigma(X) = \sqrt{\lambda}.$$

Un fait important sur la loi de Poisson : la **stabilité par somme**, c'est-à-dire que si $X \sim \text{Poi}(\lambda)$ et $Y \sim \text{Poi}(\mu)$, alors on a $X + Y \sim \text{Poi}(\lambda + \mu)$. Ainsi la loi de Poisson est très particulière et on la retrouvera dans de nombreux problèmes de statistiques.

3 Variables aléatoires et lois de probabilité continues

3.1 Variables aléatoires continues, fonctions de densité et de répartition

Rappel Une variable aléatoire continue est une application $X : \Omega \rightarrow X(\Omega) \subset \mathbb{R}$, avec $X(\Omega)$ ensemble image non dénombrable. En pratique, toute variable qui peut prendre des *valeurs non entières* sera une variable aléatoire continue. Par exemple, la mesure précise d'une longueur, d'une durée, dans une expérience aléatoire, pourront être modélisées par des variables aléatoires continues.

On caractérise la **loi de probabilité** d'une v.a. continue X par les valeurs

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

pour tous $a \leq b$ réels. On note que contrairement aux v.a. discrètes du chapitre précédent, les valeurs ponctuelles $\mathbb{P}(X = x)$ pour un $x \in \mathbb{R}$ fixé *ne caractérisent pas* la loi de X . En général (c'est le cas pour tous les exemples de v.a. continues qui nous intéressent), on a $\mathbb{P}(X = x) = 0$ pour tout $x \in \mathbb{R}$.

Exemple 35. Que veut dire "tirer un nombre au hasard entre 0 et 1"? Avec R, on verra que la fonction `runif` permet de faire ceci; c'est une fonctionnalité présente dans tous les langages de programmation modernes, qui sert de base pour simuler toute variable aléatoire continue. Le principe est le suivant :

- on ne tirera jamais deux fois la même valeur.
- la valeur est *choisie uniformément* dans le segment $[0, 1]$.

Choisir uniformément un réel dans $[0, 1]$ n'est pas forcément très clair pour l'instant. Intuitivement, aucun endroit du segment n'est favorisé plus qu'un autre, et donc si l'on découpe le segment en n petits intervalles successifs $[k/n, (k+1)/n]$, la v.a. devrait être répartie *de manière équiprobable* dans l'un de ces segments. Ainsi on aurait $\mathbb{P}(X \in [k/n, (k+1)/n]) = 1/n$ pour tout $k \in \{0, \dots, n-1\}$.

Formellement, une variable aléatoire X uniforme dans $[0, 1]$ est donc choisie selon la loi de probabilité telle que

$$\forall 0 \leq a \leq b \leq 1, \quad \mathbb{P}(a \leq X \leq b) = b - a.$$

La probabilité de se retrouver dans un segment est égale à la longueur du segment, ce qui correspond bien à l'intuition qu'on a en découpant le segment $[0, 1]$. On notera $X \sim \mathcal{U}([0, 1])$ si X suit une **loi uniforme sur** $[0, 1]$.

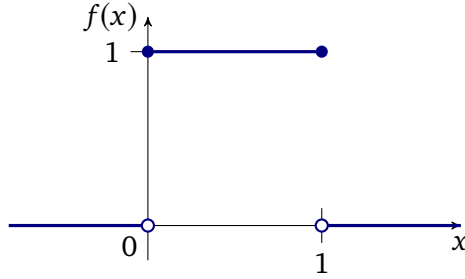
Définition 36 (Fonction densité). On dit qu'une variable aléatoire continue X admet une (**fonction**) **densité** f si $f : \mathbb{R} \rightarrow \mathbb{R}_+$ est telle que pour tous $a \leq b$ réels,

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

Exemple 37. La loi uniforme sur $[0, 1]$ admet la fonction de densité

$$f : x \mapsto \mathbb{1}_{[0,1]}(x) = \begin{cases} 1 & \text{si } x \in [0, 1] \\ 0 & \text{sinon.} \end{cases}$$

On remarque que la densité n'est pas nécessairement une fonction continue. Dans nos exemple, elle sera toujours continue par morceau (ici elle est constante par morceaux).



Graphique de la densité de la loi $\mathcal{U}([0, 1])$.

Proposition 38. Soit f la fonction de densité d'une v.a. continue X . Alors on a :

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

(ii) pour tout $a \in \mathbb{R}$ fixé, $\mathbb{P}(X = a) = 0$.

Remarque 39. On rappelle que \mathbb{P}_X est une mesure de probabilité. Comme $\{a\}$ et $]a, b]$ sont deux ensembles disjoints dont l'union vaut $[a, b]$, on a donc

$$\begin{aligned} \mathbb{P}(X \in [a, b]) &= \underbrace{\mathbb{P}(X = a)}_{=0} + \mathbb{P}(X \in]a, b]) \\ &= \mathbb{P}(X \in]a, b]). \end{aligned}$$

On peut faire la même opération sur la borne de droite de l'intervalle, et résumer de la manière suivante : toute variable aléatoire *continue* (à densité) X est telle que pour tous $a < b$ réels, on a

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

Attention, ce n'est bien sûr pas valable pour les variables aléatoires discrètes.

Rappel : Fonction de répartition On a défini la fonction de répartition d'une v.a. X comme

$$F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}_X(]-\infty, t]).$$

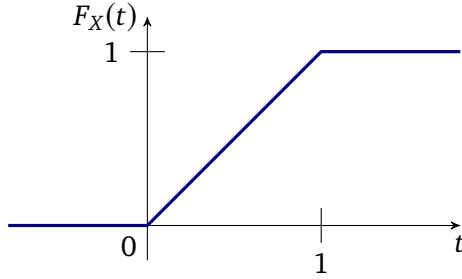
Ainsi, pour une variable de densité f , on a

$$F_X(t) = \int_{-\infty}^t f(s) ds.$$

La fonction de répartition est donc une *primitive* de la fonction de densité : on a donc $F'_X = f$. Ceci est à prendre dans un sens un peu "faible" puisqu'on autorise la densité à n'être pas continue : la fonction de répartition n'est pas dérivable aux points de discontinuité de la densité.

Exemple 40. On revient à la loi uniforme sur $[0, 1]$. Il est facile de calculer :

$$F_X(t) = \mathbb{P}(X \leq t) = \begin{cases} 0 & \text{si } t < 0 \\ t & \text{si } 0 \leq t \leq 1 \\ 1 & \text{si } t > 1. \end{cases}$$



Graphique de la fonction de répartition de la loi $\mathcal{U}([0, 1])$.

On remarque (cf. schéma) qu'elle n'est pas dérivable en 0 ni en 1, mais qu'ailleurs sa dérivée coïncide bien avec la densité $f = \mathbb{1}_{[0,1]}$.

Proposition 41 (Propriétés de la fonction de répartition).

(i) F_X est continue pour une v.a. continue, croissante, à valeurs dans $[0, 1]$, et satisfait encore

$$\lim_{t \rightarrow -\infty} F_X(t) = 0, \quad \lim_{t \rightarrow \infty} F_X(t) = 1.$$

(ii) $\mathbb{P}(X > t) = 1 - F_X(t)$.

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

3.2 Espérance et variance d'une variable aléatoire continue

Définition 42 (Espérance). Pour X une v.a. continue de densité f , on définit l'**espérance** de X comme

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} tf(t) dt,$$

si cette valeur est bien définie.

Plus généralement, pour $\varphi : \mathbb{R} \rightarrow \mathbb{R}$, on définit

$$\mathbb{E}[\varphi(X)] := \int_{-\infty}^{\infty} \varphi(t)f(t) dt,$$

Comme dans le cas discret, cette définition formalise l'idée de moyenne. Aussi, la *linéarité de l'espérance* tient toujours.

Exemple 43. Soit $X \sim \mathcal{U}([0, 1])$. Intuitivement, si une suite de nombres est uniformément répartie sur le segment $[0, 1]$, elle doit avoir pour moyenne le milieu du segment, c'est-à-dire $1/2$. On vérifie que la définition de l'espérance colle bien avec cette intuition :

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} t\mathbb{1}_{[0,1]}(t) dt = \int_0^1 t dt = \left[\frac{t^2}{2} \right]_0^1 = \frac{1}{2}.$$

Remarque : il est inutile d'intégrer sur un intervalle où la fonction de densité est nulle. En général, on écrira donc directement $\mathbb{E}[\varphi(X)] = \int_0^1 \varphi(t) dt$ pour une v.a. uniforme sur $[0, 1]$, c'est-à-dire que l'on se passe de la première étape du calcul ci-dessus.

Définition 44 (Variance). Pour X une v.a. continue de densité f , on définit la **variance** de X comme pour une v.a. discrète, c'est-à-dire

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

à condition que cette valeur soit bien définie.

L'écart-type est encore défini comme $\sigma(X) = \sqrt{\text{Var}(X)}$.

Exemple 45. Revenons à notre variable X uniforme dans $[0, 1]$. Pour calculer sa variance, on calcule d'abord

$$\mathbb{E}[X^2] = \int_0^1 t^2 dt = \left[\frac{t^3}{3} \right]_0^1 = \frac{1}{3}.$$

Comme on sait que $\mathbb{E}[X] = 1/2$, on a donc $\text{Var}(X) = 1/3 - (1/2)^2 = 1/12$, et $\sigma(X) = 1/2\sqrt{3} \approx 0.29$.

3.3 Quelques lois continues

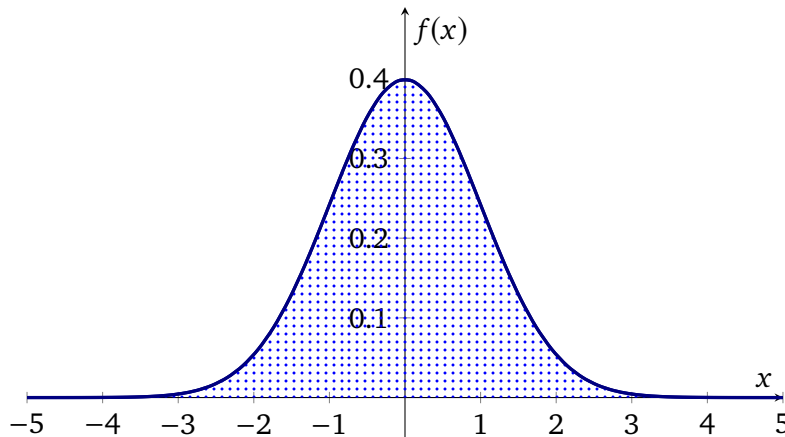
3.3.a) Loi normale

Définition 46 (Loi normale). X suit la **loi normale** $\mathcal{N}(0, 1)$ **centrée réduite** si sa densité de probabilité est donnée par

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

On note alors $X \sim \mathcal{N}(0, 1)$.

Densité de la loi normale centrée réduite



Rappel : on a $\int_{-\infty}^{\infty} f(x) dx = 1$, c'est-à-dire que l'aire sous la courbe vaut 1.

Ainsi pour $X \sim \mathcal{N}(0, 1)$ on a

$$\mathbb{P}(a \leq X \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

→ Pas d'expression analytique connue pour l'intégrale (idem pour la fonction de répartition). On utilise un tableau de valeurs ou de l'intégration numérique.

Remarque 47.

- La fonction de répartition de la loi normale n'a pas d'expression analytique, on peut néanmoins déduire de la symétrie de sa densité le fait que $F_X(0) = \mathbb{P}(X \leq 0) = \mathbb{P}(X \geq 0) = 1/2$.
- C'est une loi associée à Carl Friedrich Gauss (début du XIX^e siècle), même si De Moivre ou Laplace l'ont étudiée dès le XVIII^e. Par conséquent, on la trouve aussi sous le nom de loi gaussienne.
- C'est une *loi fondamentale* en statistique (cf. théorème central limite).

Proposition 48. Si $X \sim \mathcal{N}(0, 1)$, on a $\mathbb{E}[X] = 0$ et $\text{Var}(X) = 1$.

On remarque que $\mathbb{E}[X] = 0$ et $\text{Var}(X) = 1$ sont les deux paramètres que l'on a écrit dans $\mathcal{N}(0, 1)$. En fait, on peut définir des lois normales de moyennes et variances différentes.

Définition 49 (Loi normale $\mathcal{N}(\mu, \sigma)$). Une variable aléatoire X suit la loi normale d'espérance $\mu \in \mathbb{R}$ et de variance $\sigma^2 \in \mathbb{R}_+^*$ si elle admet pour fonction de densité :

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

On note alors $X \sim \mathcal{N}(\mu, \sigma)$, et on a $\mathbb{E}[X] = \mu$, $\text{Var}(X) = \sigma^2$, et donc $\sigma(X) = \sigma$.

Remarque 50. Plus souvent en probabilités on trouve la notation $\mathcal{N}(\mu, \sigma^2)$ – attention au carré – pour la définition précédente, c'est-à-dire que l'on met en avant la variance plutôt que l'écart-type. En pratique, comme on travaillera plus souvent avec l'écart-type (par exemple en R on simulera une loi normale en stipulant l'écart-type dans la fonction `rmnorm`, et non la variance), dans le cadre de ce cours on choisit cette définition.

3.3.b) Loi uniforme

On a déjà vu plus haut la loi uniforme sur $[0, 1]$. De manière générale, soit a et b deux réels, avec $a < b$; on dit qu'une v.a. X suit la **loi uniforme sur $[a, b]$** , et on note $X \sim \mathcal{U}([a, b])$, si elle admet pour densité

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

- Fonction de répartition : $F_X(t) = \frac{t - a}{b - a}$ pour tout $t \in [a, b]$.
- $\mathbb{E}[X] = \frac{a + b}{2}$, $\text{Var}(X) = \frac{(b - a)^2}{12}$

3.3.c) Loi exponentielle

C'est la loi classique qui modélise la durée de vie d'un matériel (sans usure). On note $X \sim \text{Exp}(\lambda)$ et on dit que la variable aléatoire X suit une **loi exponentielle** de paramètre λ , avec $\lambda \in \mathbb{R}_+^*$, si elle admet comme densité sur \mathbb{R}_+ :

$$f(x) = \lambda e^{-\lambda x}.$$

- Fonction de répartition : $F_X(t) = 1 - e^{-\lambda t}$.

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

Remarque 51. La loi exponentielle est l'analogie de la loi géométrique pour les lois continues. Elle vérifie aussi la propriété de **manque de mémoire** :

$$\forall t, s \geq 0, \quad \mathbb{P}(X > t + s \mid X > t) = \mathbb{P}(X > s) = e^{-\lambda s}.$$

3.3.d) Loi gamma

Quand l'on somme deux variables exponentielles X_1 et X_2 indépendantes et de même paramètre λ , on peut calculer (exercice !) que

$$\mathbb{P}(X_1 + X_2 \leq x) = \int_0^x \lambda^2 t e^{-\lambda t} dt.$$

Plus généralement, si l'on somme n variables exponentielles X_1, \dots, X_n indépendantes et de même paramètre λ , on calcule que

$$\mathbb{P}(X_1 + \dots + X_n \leq x) = \int_0^x \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} dt.$$

Ces lois font partie des lois gamma, qui sont un peu plus générales encore : on note $X \sim \Gamma(\alpha, \lambda)$ et on dit que X suit la **loi gamma** de paramètre (α, λ) , avec $\alpha > 0, \lambda > 0$ des réels, si sa fonction de densité sur \mathbb{R}_+ est

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x},$$

où $\Gamma(\alpha) := \int_0^\infty t^{\alpha-1} e^{-t} dt$ désigne la fonction gamma d'Euler (cf. [Wikipédia](#)).

- α est appelé paramètre de forme, λ paramètre d'intensité.
- Quand α vaut 1, une variable $X \sim \Gamma(1, \lambda)$ suit la loi exponentielle d'intensité λ .
- Si $X_1 \sim \Gamma(\alpha_1, \lambda)$ et $X_2 \sim \Gamma(\alpha_2, \lambda)$, avec $X_1 \perp X_2$, alors $X_1 + X_2 \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$.
- Sa fonction de répartition n'a pas de forme analytique simple, sauf par exemple si $\alpha = 1$.
- $\mathbb{E}[X] = \frac{\alpha}{\lambda}$, $\text{Var}(X) = \frac{\alpha}{\lambda^2}$.

3.3.e) Loi bêta

Les lois bêta sont aussi formées à partir de lois exponentielles (ou plus généralement de lois gamma). On considère $X_1 \sim \Gamma(\alpha, \lambda)$ et $X_2 \sim \Gamma(\beta, \lambda)$, avec $X_1 \perp X_2$. Alors la variable

$$X := \frac{X_1}{X_1 + X_2}$$

a pour fonction de densité sur $[0, 1]$

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}.$$

On note alors $X \sim \text{Beta}(\alpha, \beta)$ et on dit que X suit la **loi bêta** de paramètre (α, β) .

- En faisant varier α et β , on obtient différentes formes de distributions (plus ou moins concentrée sur la moyenne, ou concentrée près de l'une ou l'autre des extrémités, voire des deux).
- Les lois bêta sont très souvent utilisées en statistique pour l'inférence de quantités dans $[0, 1]$.
- Sa fonction de répartition n'a pas de forme analytique simple en général.
- $\mathbb{E}[X] = \frac{\alpha}{\alpha + \beta}$, $\text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$.

4 Convergences et applications

4.1 Différents types de convergence

Deux résultats fondamentaux en probabilité et statistique sont la *loi des grands nombres* et le *théorème central limite*. Pour pouvoir les appréhender, on introduit quelques notions de *convergence d'une suite de variables aléatoires*.

Définition 52 (Convergence presque sûre). Une suite de v.a. $(X_n)_{n \geq 1}$ **converge presque sûrement** vers la v.a. X si

$$\mathbb{P}\left(\left\{\omega \in \Omega, \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

On notera $X_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X$.

C'est "la plus forte" des notions de convergence de variables aléatoires. Intuitivement, si on a $X_n \rightarrow X$ p.s., cela signifie que pour toute réalisation $\omega \in \Omega$ de notre expérience aléatoire, la suite (numérique) d'observations $(X_n(\omega))_{n \geq 1}$ tend vers l'observation $X(\omega)$.

Les autres notions de convergence de variables aléatoires ne font pas intervenir la convergence de la suite d'observations $(X_n(\omega))_{n \geq 1}$ elle-même.

Définition 53 (Convergence en probabilité). Une suite de v.a. $(X_n)_{n \geq 1}$ **converge en probabilité** vers la v.a. X si pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

On notera $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$.

Le dernier type de convergence que l'on aborde ici, très utile en statistiques, est la plus faible des notions, au sens où si on converge dans un certain sens, on converge toujours en loi.

Définition 54 (Convergence en loi). Une suite de v.a. $(X_n)_{n \geq 1}$ **converge en loi** vers la loi de probabilité de fonction de répartition F si on a

$$\lim_{n \rightarrow \infty} F_{X_n}(t) = F(t),$$

en tout point t où F est continue.

On notera $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} F$. On pourra aussi remplacer F par les conventions de notations des lois usuelles, par exemple on notera $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$ si F est la fonction de répartition de la loi normale centrée réduite.

Cela signifie que lorsque n est grand, la loi de probabilité de X_n est approximativement la loi de fonction de répartition F .

Le lien entre ces différentes notions de convergence est le suivant :

convergence presque sûre \implies convergence en probabilité \implies convergence en loi,

où la dernière implication signifie que si une suite de v.a. converge vers une variable X de fonction de répartition F_X , alors elle converge aussi en loi vers F_X .

4.2 Deux résultats importants

Théorème 55 (Loi forte des grands nombres). Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles indépendantes et de même loi, d'espérance $\mathbb{E}[X]$. Alors on a la convergence presque sûre des moyennes empiriques :

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{p.s.} \mathbb{E}[X].$$

Concrètement, cela signifie que quand on fait un très grand nombre d'expériences identiques et indépendantes, la moyenne des réalisations de la v.a. à laquelle on s'intéresse tend vers l'espérance de sa loi. Ce résultat nous permettra de justifier l'idée naturelle d'estimer une espérance par une moyenne et une probabilité par une proportion.

Théorème 56 (Théorème central limite). Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles indépendantes et de même loi, d'espérance $\mathbb{E}[X]$ et d'écart-type $\sigma(X) = \sqrt{\text{Var}(X)}$. Alors on a la convergence en loi suivante :

$$\frac{(\sum_{i=1}^n X_i) - n\mathbb{E}[X]}{\sqrt{n}} = \sqrt{n}(\bar{X}_n - \mathbb{E}[X]) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma(X)),$$

ou de manière équivalente

$$\frac{(\sum_{i=1}^n X_i) - n\mathbb{E}[X]}{\sqrt{n \text{Var}(X)}} = \sqrt{n} \frac{\bar{X}_n - \mathbb{E}[X]}{\sigma(X)} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Cela signifie que la loi de toute v.a. égale à la somme d'un nombre "suffisamment grand" de variables aléatoires indépendantes et de même loi est approximativement une loi normale. Plus précisément, pour n grand, $\sum_{i=1}^n X_i$ est approximativement de loi $\mathcal{N}(n\mathbb{E}[X], \sqrt{n}\sigma(X))$. De manière remarquable, ce résultat est vrai quelle que soit la loi des X_i (tant qu'elle admet une variance). De très nombreux phénomènes naturels sont la résultante d'un grand nombre de phénomènes identiques, indépendants et additifs, ce qui justifie l'importance (et le nom) de la loi normale.