Probabilités

Département de Mathématiques Efrei Paris

Année 2023/2024 S3

L2 - L2 BN - L2 BDX - L2 PP

Table des matières

II Les variables aléatoires

I Fondamentaux des probabilités

1. Espace probabilisé

La théorie des probabilités est une branche des mathématiques qui étudie les lois régissant les phénomènes aléatoires, c'est-à-dire les phénomènes où le hasard intervient, comme par exemple lors d'un lancer de dés. Dans ce cours, nous allons décrire le cadre des probabilités et en fixer la terminologie. Nous ne séparerons pas la présentation des définitions du rôle qu'elles jouent dans la modélisation d'une situation concrète.

Un bref historique [1]

Aujourd'hui, les probabilités sont partout : elles sont au coeur de la physique quantique, dans la prédiction pratique de processus temporels (le marché de la bourse, la météo), dans la recherche en médecine, etc. Le terme "probabilité" a d'abord été utilisé au Moyen Âge en jurisprudence. Il est issu du latin "probare" qui signifie "prouver". Le calcul des probabilités était alors un "art" dont le but était de déterminer la vraisemblance d'un évènement. Le développement des probabilités comme science mathématique est donc intimement lié aux jeux. Les premiers problèmes étaient liés aux jeux de dés et jeux de cartes : il s'agissait de *compter* le nombre de possibilités pour qu'une certaine configuration apparaisse (e.g., la somme de deux dés égale à 7). Mais le problème qui a réellement lancé le développement de la théorie des probabilités est le fameux *problème des partis* :

1. Toutes les notes historiques sont principalement tirées de l'article *Calcul des probabilités*, de Michel Loève, dans l'ouvrage Abrégé d'histoire des mathématiques 1700-1900, sous la direction de Jean Dieudonné, Tome II, Hermann, 1978.

Problème des paris
n joueurs, chacun ayant misé au départ une somme d'argent k, jouent à un jeu de hasard. Le jeu s'arrête lorsqu'un des joueurs a emporté N victoires et le gagnant emporte la totalité de l'argent misée nk. Comment distribuer le prix si le jeu est interrompu avant la fin pour une raison quelconque ?

Certains des grands mathématiciens italiens du 15ème et 16ème siècle (Pacioli, Cardano, Tartaglia), la même école qui a contribué à la résolution des équations de troisième degré et à l'introduction des nombres complexes, se sont attaqués à ce problème en fournissant des réponses fausses.

Les trois pères fondateurs de la théorie des probabilités viennent alors avec :

- ⋆ Blaise Pascal, mathématicien francais, qui en 1654, dans sa correspondance avec Pierre de Fermat, trouve la solution correcte au problème des partis. La solution est très liée au fameux triangle de Pascal et utilise, sans le dire encore, le concept de probabilité de victoire.

- ⋆ Christian Huygens, le grand scientifique néerlandais du 17ème siècle, qui, en 1657, publie *De Ratiocinius in Ludo Aleae*, premier ouvrage où le concept de probabilité est énoncé rigoureusement et qui reconnaît l'importance des probabilités au-delà des jeux [2].

- ⋆ Jakob Bernoulli, mathématicien et physicien suisse, qui, dans son oeuvre posthume *Ars Conjectandi* de 1713, s'intéresse au problème d'un jeu de hasard avec deux issues (victoire/défaite) qui se répète de façon identique n fois (la fameuse *expérience de Bernoulli*).



FIGURE 1.1 – Les pères fondateurs des probabilités :
Blaise Pascal, Christian Huygens et Jakob Bernoulli.

Il s'en suit une période de deux cents ans avec quelques avancées importantes faites surtout par l'école de mathématiciens français de l'école polytechnique (Poisson, Laplace) et l'école russe de

2. "Le lecteur observera que nous nous occupons, non seulement de jeux, mais d'une théorie nouvelle, à la fois profonde et intéressante."

Saint-Pétersbourg (Tchebychev, Markov, Liapounov) (voir pages 87 et 125 pour les détails). Cela dit, jusque dans les années 1930, la théorie des probabilités reste, pour les mathématiciens, une petite partie des mathématiques appliquées avec peu d'intérêt théorique. La situation change radicalement grâce aux travaux du mathématicien russe Andreï Kolmogorov qui, dans son ouvrage *Grundbegriffe der Wahrscheinlichkeitsrechnung* de 1933, formalise rigoureusement toute la théorie des probabilités en la rattachant à la théorie de la mesure développée par Henri-Léon Lebesgue (voir section 1.4).[3]

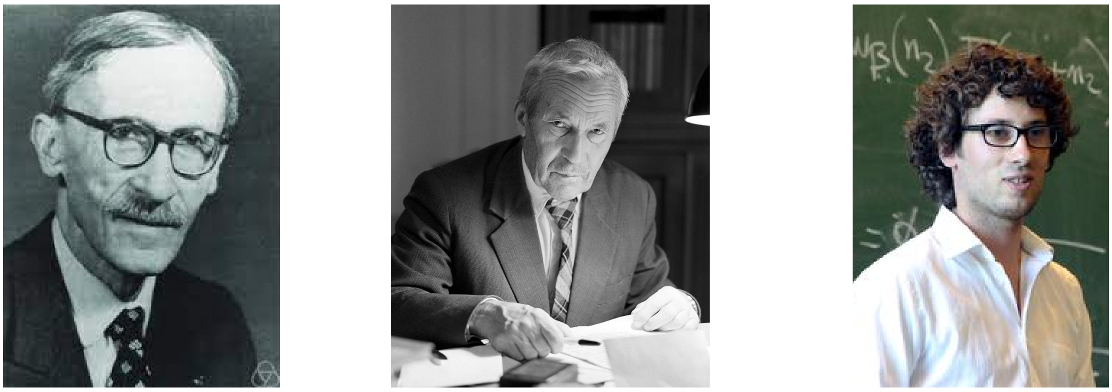


FIGURE 1.2 – Quelques mathématiciens des probabilités modernes : Paul Pierre Levy, Andreï Kolmogorov et Hugo Dominil-Copin (Médaille Fields 2022).

1.1 Concepts de base

> **Définition 1.1 (Expérience aléatoire, univers)**
> On parle d'*expérience ou épreuve aléatoire* lorsqu'on considère une expérience dont on ne peut pas prévoir le résultat de façon certaine. L'étude d'une expérience aléatoire commence par la description des résultats possibles, appelés *éventualités* ou *évènements élémentaires*. L'ensemble de tous les résultats d'une expérience aléatoire est appelé *univers* que l'on note, par habitude, Ω.

Remarque : Tout univers est par convention non vide car on suppose que l'expérience sur laquelle on travaille produit au moins un résultat possible.

3. Nous, dans ce cours, nous nous arrêterons aux mathématiques du 19ème siècle. Nous ne toucherons à aucun développement mathématique des 150 dernières années !

Exemples :

i) Dans le cas d'un lancé d'une pièce, $\Omega = \{P, F\}$ avec P désignant Pile et F désignant Face.

ii) Dans le cas d'un lancer d'un dé (à six faces), l'univers est tout naturellement $\Omega = \{1,2,3,4,5,6\}$. Le résultat d'un lancer est représenté par le nombre marqué sur la face tournée vers le ciel.

iii) Dans le cas d'un lancer de deux dés, l'univers est $\Omega = \{1,2,3,4,5,6\}^2$. Le résultat d'une telle épreuve est représenté par un couple formé par le nombre indiqué par le premier dé et celui indiqué par le second.

iv) Dans le cas de la mesure (en jours) de la durée de vie d'un téléphone portable, on peut prendre comme univers $\Omega = \mathbb{N}$.

v) Nous sommes à l'arrêt du bus et nous savons que le bus passe exactement toutes les douze minutes mais nous ignorons quand est passé le dernier bus. Alors, notre temps d'attente jusqu'à l'arrivée du prochain bus peut être vu comme une expérience aléatoire dont l'univers est $\Omega = [0, 12]$.

Définition 1.2 (Évènements)
Etant donnée une expérience aléatoire, on appelle :

- ⋆ *évènement*, un sous-ensemble de l'univers Ω,
- ⋆ *évènement certain*, l'évènement Ω,
- ⋆ *évènement impossible*, l'évènement $\varnothing$.

Remarque : Cette définition de la notion d'évènement pose problème dans le cas où l'univers Ω est indénombrable, comme dans le dernier exemple ci-dessous (pour les curieux, voir la section 1.4 afin de mieux comprendre le problème et comment le dépasser). Dans ce cours on considérera toujours des univers finis ou dénombrables, sauf mention du contraire.

Exemples : En reprenant les exemples précédents :

i) L'évènement $E =$"tomber sur Pile"$= \{P\}$.

ii) L'évènement $E =$"le résultat est paire".

iii) L'évènement $E =$"la somme des dés vaut 7".

iv) L'évènement $E =$"la durée de vie du téléphone est supérieure à trois ans".

v) L'évènement $E =$"attendre moins de cinq minutes".

1.2 Rappels sur la théorie des ensembles

1.2.1 Opérations ensemblistes

Définition 1.3 (Opérations élémentaires sur les ensembles)
Soit Ω un ensemble. Alors :

1. **Sous-ensemble**. E est un sous-ensemble de Ω, noté $E \subset \Omega$, si $\forall x \in E, x \in \Omega$.
   E est appelé un *sous-ensemble propre* si $E \subset \Omega$, $E \neq \Omega$ et $E \neq \varnothing$.
2. **Complémentaire**. Etant donné un sous-ensemble E de Ω, le complémentaire de E dans Ω est l'ensemble $\overline{E} = \{x \in \Omega \,|\, x \notin E\}$.
3. **Union**. Etant donnés deux sous-ensembles E et F de Ω, leur union est l'ensemble
$$E \cup F = \{x \in \Omega \,|\, (x \in E) \text{ ou } (x \in F)\}.$$
4. **Intersection**. Etant donnés deux sous-ensembles E et F de Ω, leur intersection est l'ensemble $E \cap F = \{x \in \Omega \,|\, (x \in E) \text{ et } (x \in F)\}$.

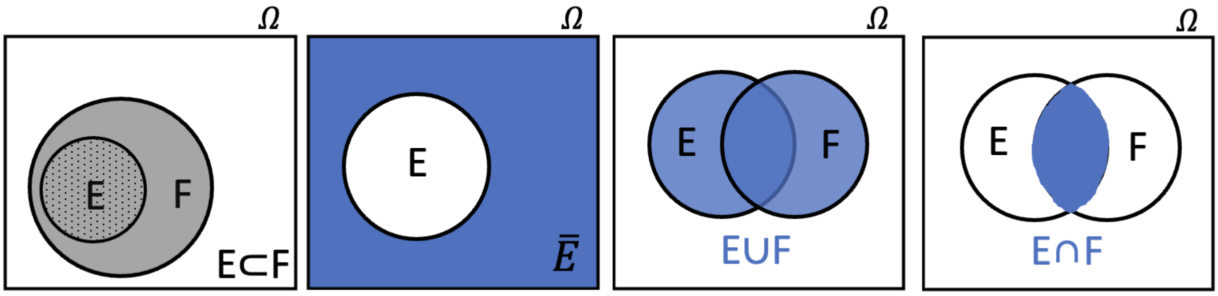


FIGURE 1.3 – Les quatre opérations sur les ensembles.

Ces quatre opérations peuvent être vues d'un point de vue logique comme correspondant à des connecteurs mais aussi comme décrivant des réalisations d'évènements d'expériences aléatoires :

Connecteurs logiques	**Opérations sur les ensembles**	**Réalisation d'évènements**
ET	Intersection $E \cap F$	les deux évènements se réalisent
OU	Union $E \cup F$	au moins un des deux évènements se réalise
NON	Complémentaire $\overline{E}$	E ne se réalise pas
$\Longrightarrow$	Sous-ensemble $E \subset F$	Chaque fois que E se réalise, F se réalise aussi

Proposition 1.1 : Structure algébrique des opérations ensemblistes

Les opérations élémentaires sur les ensembles vérifient les propriétés suivantes :

1. Commutativité de $\cup$ et $\cap$:
$$\forall (E,F) \subset \Omega^2,\ E \cup F = F \cup E \text{ et } E \cap F = F \cap E.$$
2. Associativité de $\cup$ et $\cap$:
$$\forall (E,F,G) \subset \Omega^3,\ (E \cup F) \cup G = E \cup (F \cup G) \text{ et } (E \cap F) \cap G = E \cap (F \cap G).$$
3. Idempotence de $\cup$ et $\cap$:
$$\forall E \subset \Omega,\ E \cup E = E = E \cap E.$$
4. $\varnothing$ est l'élément neutre de $\cup$ et Ω est l'élément neutre de $\cap$:
$$\forall E \subset \Omega,\ E \cup \varnothing = E = E \cap \Omega.$$
5. Ω est l'élément absorbant de $\cup$ et $\varnothing$ est l'élément absorbant de $\cap$:
$$\forall E \subset \Omega,\ E \cup \Omega = \Omega \text{ et } E \cap \varnothing = \varnothing.$$
6. Distributivité de $\cup$ sur $\cap$ et de $\cap$ sur $\cup$:
$$\forall (E,F,G) \subset \Omega^3,\ (E \cup F) \cap G = (E \cap G) \cup (F \cap G) \text{ et } (E \cap F) \cup G = (E \cup G) \cap (F \cup G).$$

Remarque : D'un point de vue purement structurel, *il n'y a donc rien qui distingue l'intersection de l'union*. Cela implique, en particulier, que toute formule générale utilisant les symboles $\cap$ et $\cup$ sera aussi vraie sous l'échange $\cap \longleftrightarrow \cup$ et $\Omega \longleftrightarrow \varnothing$. Un exemple est la formule suivante :

Proposition 1.2 : Lois de De Morgan

Soient $E, F \subset \Omega$. On a
$$\overline{E \cup F} = \overline{E} \cap \overline{F} \text{ et } \overline{E \cap F} = \overline{E} \cup \overline{F}.$$

En mots : la négation d'une conjonction est la disjonction des négations et vice-versa.

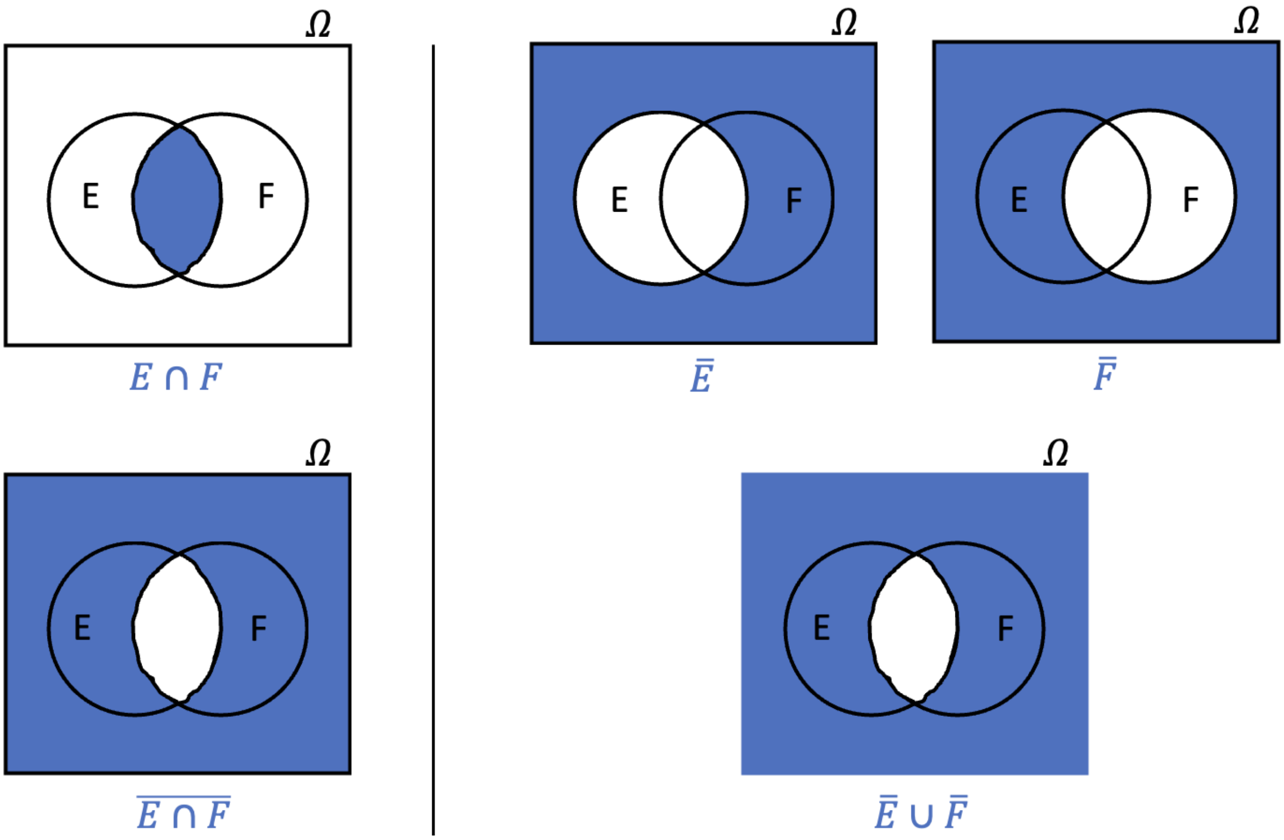


FIGURE 1.4 – Illustration de l'égalité $\overline{E \cap F} = \overline{E} \cup \overline{F}$.

Définition 1.4 (évènements incompatibles)
Soient $E, F \subset \Omega$ deux évènements. On dit que E et F sont *incompatibles* si leur réalisation conjointe est impossible : $E \cap F = \varnothing$.

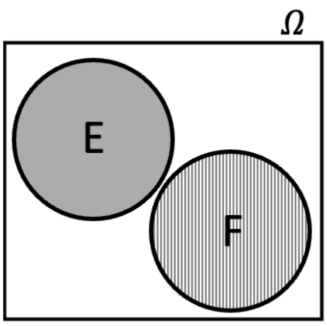


FIGURE 1.5 – Deux évènements incompatibles.

Définition 1.5 (Système complet d'évènements)
Soit $\mathscr{A} = \{A_1, \ldots, A_n\}$ une famille d'évènements possibles de Ω. On dit que $\mathscr{A}$ est un *système complet d'évènements* si les évènements satisfont les deux propriétés suivantes :

1. $\bigcup_{1 \leq i \leq n} A_i = \Omega$ (les évènements couvrent tout ce qui est possible),
2. $\forall 1 \leq i, j \leq n, i \neq j \implies A_i \cap A_j = \varnothing$ (les évènements sont deux à deux incompatibles).

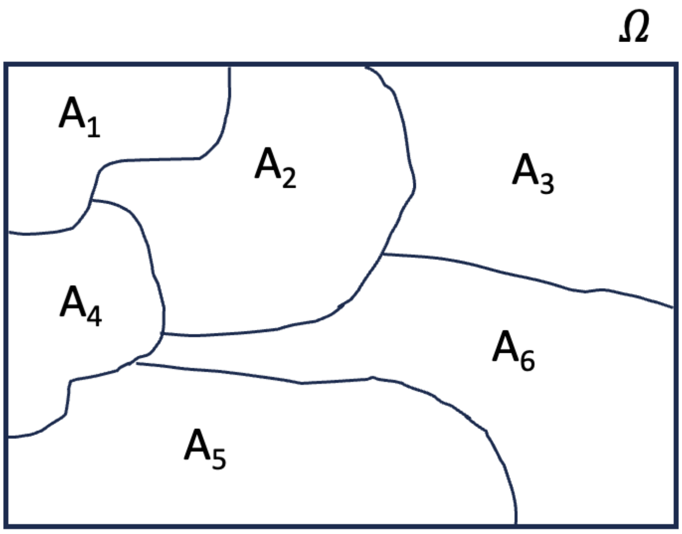


FIGURE 1.6 – Exemple d'un système complet d'évènements.

Exemples :

i) Dans l'expérience aléatoire "choisir au hasard une carte dans un paquet de 52 cartes", les évènements suivants forment un système complet d'évènements :

$$\begin{aligned} A_1 &= \text{"la carte est une figure ou un As"}, \\ A_2 &= \text{"la carte est comprise entre 7 et 10"}, \\ A_3 &= \text{"la carte est comprise entre 2 et 6"}. \end{aligned}$$

ii) Dans l'expérience aléatoire "choisir au hasard un mot du dictionnaire", les évènements suivants forment un système complet d'évènements :

$$\begin{aligned} A_1 &= \text{"le mot commence par une voyelle"}, \\ A_2 &= \text{"le mot commence par une consonne et apparaît avant le mot } \texttt{étudiant}\text{"}, \\ A_3 &= \text{"le mot commence par une consonne et apparaît après le mot } \texttt{excellent}\text{"}. \end{aligned}$$

iii) Dans l'expérience aléatoire "choisir au hasard un étudiant de l'EFREI", les évènements suivants forment un système complet d'évènements :

$$\begin{aligned}
A_1 &= \text{“l'étudiant est dans les programmes experts”},\\
A_2 &= \text{“l'étudiant est dans le cycle licence du programme grande école”},\\
A_3 &= \text{“l'étudiant est en alternance dans le cycle master du programme grande école”},\\
A_4 &= \text{“l'étudiant est dans le cycle master du programme grande école sans alternance”}.
\end{aligned}$$

1.2.2 Cardinalité

Définition 1.6 (Cardinalité)
Soit E un ensemble fini. On appelle *cardinalité*, ou *cardinal* de E, notée $|E|$ ou $card(E)$, le nombre d'éléments de E.

Remarque : La notion de cardinalité peut être étendue aux ensembles infinis. C'est d'ailleurs dans ce cadre que la notion devient vraiment intéressante. Mais la théorie des cardinaux développée par Georg Cantor dépasse le cadre de ce cours.

Fiche méthode #1 (Principe multiplicatif)
Lorsque le dénombrement d'une ensemble se décompose en une succession d'étapes indépendantes, on compte le nombre de possibilités offertes à chaque étape et les nombres de possibilités se multiplient.

♦Exemple : Combien peut-on composer d'équipes mixtes avec un garçon et une fille, sachant qu'il y a dans la classe 20 filles et 15 garçons ?

- ⋆ choix de la fille : 20 possibilités
- ⋆ choix du garçon : 15 possibilités

Au total, il y a donc $20 \times 15 = 300$ équipes mixtes possibles.

Remarque : Mathématiquement, cela se traduit par la notion de cardinal d'un produit de deux ensembles finis : soient E et F deux ensembles finis, alors $E \times F$ est un ensemble fini et $Card(E \times F) = Card(E) \times Card(F)$.

Fiche méthode #2 (Passage au complémentaire)
Pour déterminer le cardinal d'une partie A d'un ensemble Ω, il est parfois plus simple de passer au complémentaire $\overline{A}$, puis d'utiliser la formule : $Card(A) = Card(E) - Card(\overline{A})$ et qui provient directement du fait que A et $\overline{A}$ forment un système complet d'évènements de Ω.

♦Exemple : Combien y a-t-il de nombres binaires formant un octet avec au moins un "0" ?

Le passage au complémentaire donne directement un seul nombre binaire sans "0". Et donc on trouve qu'il y a $2^8 - 1$ nombres binaires formant un octet avec au moins un "0".

Fiche méthode #3 (Utilisation de l'union pour des ensembles deux à deux disjoints)
Pour déterminer le cardinal d'un ensemble E, on est parfois amené à le décomposer comme union de parties $(A_i)_{i\in[1;n]}$ formant un système complet d'évènements de E. On a alors

$$Card(E) = Card(A_1) + ... + Card(A_n).$$

Exemples :

- ♦ Dénombrer les carrés dans un quadrillage de 4 cases sur 3 (rappelons qu'un carré peut être de côté 1, 2 ou 3). Il y a 12 carrés de côté 1, 6 carrés de côté 2 et 2 carrés de côté 3. Ce qui donne $12 + 6 + 2 = 20$ carrés dans ce quadrillage.

- ⋆ A l'aide d'un dénombrement, démontrer par récurrence la formule de Pascal.
  Rappelons d'abord ce qu'est la formule de Pascal : soit E un ensemble de n éléments et notons $\mathscr{P}_p(E)$ l'ensemble de sous-ensembles de E avec $0 \leq p \leq n$ éléments. Alors, selon la formule de Pascal, on a
  $$card(\mathscr{P}_p(E)) = \binom{n}{p} = \frac{n!}{p!(n-p)!}.$$
  La formule est vraie pour des ensembles avec 0 ou 1 un élément. Soit $n \in \mathbb{N}$ et supposons que la formule est vraie pour tous les ensembles avec n éléments ou moins. Soit E un ensemble avec $n+1$ éléments et soit a un élément de E. Faisons la remarque suivante : les sous-ensembles de E à p éléments se partagent en deux sous-ensembles disjoints :
  - ceux qui contiennent a (notons A_1 l'ensemble de ces sous-ensembles).
  - ceux qui ne le contiennent pas (notons A_2 l'ensemble de ces sous-ensembles).

  D'après l'hypothèse de récurrence, on a $card(A_1) = \binom{n-1}{p-1}$ et $card(A_2) = \binom{n-1}{p}$ et donc
  $$card(\mathscr{P}_p(E)) = card(A_1) + card(A_2) = \binom{n-1}{p-1} + \binom{n-1}{p} = \binom{n}{p}.$$

1.3 Fonction de probabilité

On cherche maintenant à attribuer à chacun de ces évènements un nombre qui représente le degré de confiance que l'on a en sa réalisation. On appelle ce nombre la probabilité de l'évènement. Il s'agit d'un réel avec l'interprétation que plus la probabilité est grande et plus notre confiance dans la réalisation de l'évènement l'est aussi. Intuitivement, on identifie la probabilité d'un événement avec la fréquence asymptotique de réalisation de cet événement lorsque l'expérience est répétée infiniment souvent. Ainsi, en répétant N fois une expérience, dans les mêmes conditions, et en notant $f_N(A)$ la fréquence de réalisation de l'événement A : $P(A) = \lim_{N\longrightarrow\infty} f_N(A)$

Plusieurs problèmes se posent alors avec cette approche fréquentiste :

- ⋆ est-ce que la suite des fréquences $f_N(A)$ converge ? En cas de convergence, la limite est-elle toujours la même ?
- ⋆ comment interpréter les probabilités dans le cas d'expériences non renouvelables ?

Pour ces raisons, l'approche moderne préfère donner une définition axiomatique abstraite des probabilités et montrer ensuite que ces axiomes permettent de reconstruire notre intuition sur le sujet.

Définition 1.7 (Fonction de probabilité)
Soit Ω un univers fini ou dénombrable.
On désigne par $\mathscr{P}(\Omega)$ l'ensemble des parties de Ω. Une *fonction de probabilité* sur Ω est une fonction $P : \mathscr{P}(\Omega) \longrightarrow \mathbb{R}$ telle que :

1. $\forall E \subset \Omega, 0 \leq P(E) \leq 1$.
2. $P(\Omega) = 1$.
3. *Additivité :* pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'évènements deux à deux incompatibles, on a

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

On dit alors que le doublet (Ω, P) est un *espace probabilisé*, et pour tout évènement $E \in \mathscr{P}(\Omega)$, on appelle *probabilité de E* le nombre $P(E) \in [0,1]$.

♦Exemples :

i) Cet exemple fait le lien entre l'approche axiomatique abstraite et l'intuition que nous avons des probabilités.
Soit Ω un ensemble fini et soit $P : \mathscr{P}(\Omega) \longrightarrow \mathbb{R}$ définie par

$$\forall E \in \mathscr{P}(\Omega), P(E) = \frac{card(E)}{card(\Omega)}.$$

Autrement dit, on définit la probabilité d'un évènement donné par la formule fréquentiste “cas favorables divisé par cas possibles”.
Alors P est une fonction de probabilité sur Ω. Cette fonction de probabilité est appelée la *probabilité uniforme sur* Ω puisque les évènements élémentaires sont tous équiprobables.
Cette probabilité uniforme apparaît dans beaucoup de situations et dans ce cas, les calculs de probabilités sont réduits à des calculs de dénombrement pour compter la cardinalité de divers ensembles. D'où l'intérêt de la section 1.2.2.

ii) Soit $\Omega = \{\omega_1, \omega_2, \ldots\}$ un ensemble fini ou dénombrable. Alors, toute fonction $p : \Omega \longrightarrow [0,1]$ telle que $\sum_{\Omega} p(\omega_i) = 1$ définit une fonction de probabilité sur Ω par

$$\forall E \in \mathscr{P}(\Omega), P(E) = \sum_{x \in E} p(x).$$

Autrement dit, toute fonction de probabilité est complètement déterminée par une suite de nombres réels positifs dont la somme vaut 1.

Proposition 1.3

Soit (Ω, P) un espace probabilisé. Alors :

- $\star$ $P(\varnothing) = 0$,
- $\star$ $P(\overline{E}) = 1 - P(E)$,
- $\star$ $E \subset F \implies P(E) \leq P(F)$,
- $\star$ $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

Démonstration. D'abord, il suffit de regarder à nouveau les dessins des opérations sur les ensembles et penser la probabilité comme la surface relative d'un sous-ensemble pour avoir une compréhension intuitive de ces résultats.

Formellement (c'est un bon exercice pour se familiariser avec les propriétés des fonctions de probabilités) :

- $\star$ On a $\varnothing \cap \Omega = \varnothing$ et $\varnothing \cup \Omega = \Omega$ donc $P(\varnothing \cup \Omega) = P(\varnothing) + P(\Omega) = P(\Omega)$, ce qui donne $P(\varnothing) = 0$.

- $\star$ Puisque $E \cap \overline{E} = \varnothing$, on a $P(E \cup \overline{E}) = P(E) + P(\overline{E})$. De plus, on a $E \cup \overline{E} = \Omega$, ce qui donne $P(E) + P(\overline{E}) = P(\Omega) = 1$.

- $\star$ Si $E \subset F$, on a $F = E \cup (F \cap \overline{E})$. De plus, puisque $E \cap (F \cap \overline{E}) = \varnothing$, on a $P(F) = P(E) + P(F \cap \overline{E})$. Enfin, comme $P(F \cap \overline{E}) \geq 0$, on a $P(F) \geq P(E)$.

⋆ On a $E \cup F = E \cup (F \cap \overline{E})$. Comme $E \cap (F \cap \overline{E}) = \varnothing$, on a $P(E \cup F) = P(E) + P(F \cap \overline{E})$. De plus, $F = (E \cap F) \cup (F \cap \overline{E})$ et ces deux derniers ensembles sont aussi disjoints. Donc $P(F) = P(E \cap F) + P(F \cap \overline{E})$. D'où l'égalité $P(E \cup F) = P(E) + P(F) - P(E \cap F)$. □

Proposition 1.4 : Formule de Poincaré

Soit (Ω, P) un espace probabilisé et $\{E_1, E_2, \ldots E_n\}$ une famille quelconque d'évènements. On a :

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} E_i\right) &= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < \ldots < i_k \leq n} P(E_{i_1} \cap \ldots \cap E_{i_k}) \\ &= \sum_{i=1}^{n} P(E_i) - \sum_{1 \leq i < j \leq n}^{n} P(E_i \cap E_j) + \sum_{1 \leq i < j < k \leq n}^{n} P(E_i \cap E_j \cap E_k) + \ldots + (-1)^{n+1} P\left(\bigcap_{i=1}^{n} E_i\right). \end{aligned}$$

♦<u>Exemple :</u> Pour une famille de quatre événements E, F, G, H, la formule de Poincaré écrite explicitement est :

$$\begin{aligned} P(E \cup F \cup G \cup H) =& P(E) + P(F) + P(G) + P(H) \\ & - P(E \cap F) - P(E \cap G) - P(E \cap H) - P(F \cap G) - P(F \cap H) - P(G \cap H) \\ & + P(E \cap F \cap G) + P(E \cap F \cap H) + P(E \cap G \cap H) + P(F \cap G \cap H) \\ & - P(E \cap F \cap G \cap H) \end{aligned}$$

Démonstration. La preuve se fait par récurrence en prouvant les cas $n = 1$ et $n = 2$ d'abord, puis en utilisant le fait que

$$P\left(\bigcup_{i=1}^{n+1} E_i\right) = P\left(\bigcup_{i=1}^{n} E_i\right) + P(E_{n+1}) - P\Big(\bigcup_{i=1}^{n} (E_i \cap E_{n+1})\Big).$$ [4]

□

Fiche méthode #4 (Raisonnement en trois étapes pour déterminer la probabilité d'un évènement)

Lorsque vous devez calculer la probabilité d'un certain évènement E, suivez la méthode suivante :

1. Commencez par identifier les évènements élémentaires qui vous permettent de décrire complètement l'expérience aléatoire.
2. Décrivez l'évènement E grâce aux évènements élémentaires (en utilisant les opérations sur les ensembles $\cap, \cup, \overline{}$).
3. Enfin, utilisez les propriétés des fonctions de probabilité pour trouver $P(E)$.

4. Pour une preuve détaillée, voir le Chapitre 2 de cette leçon sur Wikiversité.

♦Exemple : Nous sommes dans un duel au Poker (Texas Hold'em). Le premier joueur a dans sa main $(8\spadesuit, 8\heartsuit)$, le deuxième a dans sa main $(2\clubsuit, J\clubsuit)$. Sur la table, le croupier a pour l'instant retourné les cartes suivantes : $(2\diamondsuit, 4\clubsuit, 9\heartsuit, A\clubsuit)$. Il reste donc une dernière carte a retourner. Pour gagner, le deuxième joueur a besoin que cette carte soit un 2 ou un J ou un trèfle. Calculer la probabilité de victoire du deuxième joueur.

1. L'univers Ω consiste ici dans toutes les possibilités pour cette dernière carte. Ω est donc l'ensemble des 44 cartes restantes qui ne sont pas déjà sur la table.
2. Notons V_2 l'évènement “le deuxième joueur gagne” et "Deux", "Valet", "Trèfle" les évènements "la dernière carte retournée est un deux" (respectivement "un valet", "un trèfle"). On a donc $V_2 = \text{Deux} \cup \text{Valet} \cup \text{Trèfle}$. De plus, on a

$$\begin{aligned}
\text{Deux} &= \{2\heartsuit, 2\spadesuit\} \\
\text{Valet} &= \{J\heartsuit, J\spadesuit, J\diamondsuit\} \\
\text{Trèfle} &= \{3\clubsuit, 4\clubsuit, 5\clubsuit, 6\clubsuit, 7\clubsuit, 8\clubsuit, 9\clubsuit, 10\clubsuit, Q\clubsuit, K\clubsuit\}.
\end{aligned}$$

3. Puisque les évènements "Deux", "Valet" et "Trèfle" sont incompatibles, on a

$$P(V_2) = P(\text{Deux}) + P(\text{Valet}) + P(\text{Trèfle}) = \frac{2}{44} + \frac{3}{44} + \frac{9}{44} = \frac{14}{44} \approx 0.32.$$

1.4 (Hors programme) Les univers indénombrables

Cette section ne fait pas partie du programme. Elle est écrite pour les étudiants curieux, qui veulent avoir un aperçu de ce que sont réellement les mathématiques. Peut-être, sur un malentendu, vous y prendrez du goût !

La définition d'une fonction de probabilité (1.7, page 16) semble assez innocente. Mais le monde des objets mathématiques est surprenant et subtil. Il se trouve que cette définition des probabilités ne fonctionne pas dès que l'on a à faire à des ensembles indénombrables. Commençons donc par rappeler la définition précise du concept d'ensemble indénombrable.

Définition 1.8 (Ensembles dénombrables, indénombrables)
Soit E un ensemble. On dit que E est *dénombrable* s'il existe une bijection de E dans un sous-ensemble de $\mathbb{N}$. Autrement, on dit que E est *indénombrable*.

Exemples :

- $\star$ Tout ensemble fini est dénombrable. En effet, si $card(E) = n$ avec $n \in \mathbb{N}^*$, alors il existe une bijection de E dans $[\![1, n]\!]$.
- $\star$ L'ensemble des nombres pairs $2\mathbb{N}$, l'ensemble des entiers relatifs $\mathbb{Z}$, l'ensemble des nombres rationnels $\mathbb{Q}$ sont tous dénombrables.

Le concept d'ensemble dénombrable/indénombrable a été introduit par le mathématicien allemand Georg Cantor qui prouva en 1891 l'existence d'ensembles indénombrables.

Théorème 1.5 : $\mathbb{R}$ est indénombrable

L'ensemble des nombres réels est un ensemble indénombrable.

Démonstration. Nous allons prouver par l'absurde que l'intervalle $]0,1[$ est indénombrable. La preuve est basée sur l'*argument diagonal de Cantor*.

Supposons qu'il existe une bijection $f : \mathbb{N} \longrightarrow]0,1[$ et écrivons chaque nombre dans $]0,1[$ dans son écriture décimale. Définissons alors le nombre α de sorte que son n-ième chiffre après la virgule soit égal au n-ième chiffre de $f(n)$ plus un modulo 10.

Par exemple, si

$$\begin{aligned}
f(0) &= 0,5486276021766767\ldots \\
f(1) &= 0,4118197681602603\ldots \\
f(2) &= 0,6718761791761260\ldots \\
f(3) &= 0,0019653165281527\ldots \\
f(4) &= 0,3333433333333333\ldots
\end{aligned}$$

on prend

$$\alpha = 0,62205\dots$$

Alors, par construction, pour tout $n \in \mathbb{N}$, $\alpha \neq f(n)$ et donc α n'admet pas d'antécédent par f ce qui contredit l'hypothèse que f est une bijection. □

Cantor a donc prouvé qu'il existe différentes tailles d'infinis. Mais il est allé plus loin : il a prouvé qu'en partant de n'importe quel ensemble (dénombrable ou pas), on peut construire une suite infinie d'ensembles de plus en plus grands.

Théorème 1.6 : Théorème de Cantor

Soit E un ensemble et notons $\mathscr{P}(E)$ l'ensemble des sous-ensembles de E. Alors $card(\mathscr{P}(E)) > card(E)$.

Démonstration. La preuve (par l'absurde) donnée par Cantor est d'une simplicité étonnante : soit f une application de E dans $\mathscr{P}(E)$ et soit B l'ensemble défini par

$$B = \{x \in E \big| x \notin f(x)\}.$$

Supposons que f soit bijective. Alors, il existe $y \in E$ tel que $f(y) = B$ et on a

$$y \in B \iff y \in f(y) \iff y \notin B$$

ce qui constitue une contradiction. Il n'y a donc pas de fonction bijective (ni surjective) de E dans $\mathscr{P}(E)$ et donc $card(\mathscr{P}(E)) > card(E)$! □

Revenons aux probabilités et prenons comme univers un carré $\Omega = [0,1] \times [0,1]$. Nous voudrions généraliser la probabilité uniforme (voir page 16) à cet univers indénombrable. Le plus naturel serait de définir la probabilité uniforme par

$$\forall E \subset \Omega, P(E) = \frac{aire(E)}{aire(\Omega)}.$$

Cela requiert donc d'être capable de donner l'aire de n'importe quel sous-ensemble de Ω.

Ainsi, le problème de définir la probabilité uniforme sur des espaces indénombrables est donc ramené à un problème de mesure de la "taille" ou de "l'aire" de ses sous-ensembles. La théorie de la mesure est une branche importante des mathématiques qui a été développée par les mathématiciens Emile Borel et Henri-Léon Lebesgue entre la fin du 19ème siècle et le début du 20ème.

L'aire doit être une application $aire : \mathscr{P}(\Omega) \longrightarrow \mathbb{R}^+$ satisfaisant

⋆ *Additivité :* pour toute suite $(A_n)_{n\in\mathbb{N}}$ de sous-ensembles deux à deux disjoints, on a

$$aire\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} aire\,(A_i)$$

⋆ *Invariance par translation :* pour tout sous-ensemble E et tout vecteur $(x,y) \in \Omega$,

$$aire(E + (x,y)) = aire(E)$$

(l'aire d'un sous-ensemble ne doit pas changer si on le déplace.)

Mais l'étrange monde des ensembles indénombrables nous réserve une surprise :

Théorème 1.7

Il n'existe pas d'application non-nulle $\mu : \mathscr{P}([0,1]\times[0,1]) \longrightarrow \mathbb{R}^+$ satisfaisant les propriétés d'additivité et d'invariance par translation.

Autrement dit, il est impossible de définir une probabilité uniforme pour l'ensemble de tous les sous-ensembles d'un carré de $\mathbb{R}^2$.

Démonstration. Une des preuves les plus simples utilise les *ensembles de Vitali*, découverts par le mathématicien italien Giuseppe Vitali en 1905. Ce furent uns des premiers exemples d'ensembles que l'"on ne peut pas mesurer". Pour simplifier, nous allons faire la preuve en dimension 1 plutôt que 2.

Commençons par définir ces ensembles de Vitali :

Soit $\mathscr{B}$ l'ensemble de tous les nombres dans $[0,1[$ dont l'écriture binaire est finie. De plus, on définit sur $[0,1]$ la relation d'équivalence

$$\begin{aligned} x \sim y &\iff \exists b \in \mathscr{B},\, x = y \oplus b \\ &\iff \begin{array}{c}\text{on peut passer de l'écriture binaire de } x \text{ à celle de} \\ y \text{ en faisant un nombre fini de modifications.}\end{array} \end{aligned}$$

Par exemple, on a

$$15/100 = 0.001001100110011001100110011\ldots \text{ et } 65/100 = 0.101001100110011001100110011\ldots,$$

donc $15/100 \sim 65/100$ puisque leurs écritures binaires ne diffèrent que par le premier chiffre. Par contre,

$$10/100 = 0.000110011001100110011001100\ldots$$

et donc $15/100 \not\sim 10/100$ puisque leurs écritures binaires diffèrent dans tous les chiffres sauf les deux premiers. Maintenant, notons $[x]$ la classe d'équivalence de $x \in [0,1]$. Puisque $\mathscr{B}$ est dénombrable, $[x]$

l'est aussi et donc il y a un nombre indénombrable de classes d'équivalence. Construisons l'ensemble V en choisissant un élément de chaque classe d'équivalence [5].

A partir de V, on peut définir d'autres ensembles similaires en en faisant des translations par un élément de $\mathscr{B}$: pour $b \in \mathscr{B}$, on définit

$$V_b = \{x \oplus b \,|\, x \in V\}.$$

Les ensembles de Vitali ainsi définis vérifient les propriétés suivantes

$$\forall (b, b') \in \mathscr{B}^2, b \neq b' \implies V_b \cap V_{b'} = \varnothing$$

$$\bigcup_{b \in \mathscr{B}} V_b = [0, 1].$$

Maintenant, utilisons ces ensembles pour prouver, par l'absurde, le théorème. Supposons qu'il existe une application $\mu : \mathscr{P}([0,1]) \longrightarrow \mathbb{R}^+$ satisfaisant les propriétés d'additivité et invariance par translation. D'après la première propriété, on a

$$\mu([0,1]) = \mu(\bigcup_{b \in \mathscr{B}} V_b) = \sum_{b \in \mathscr{B}} \mu(V_b)$$

et d'après la deuxième propriété, on a

$$\forall b \in \mathscr{B}, \mu(V_b) = \mu(V).$$

Donc, si $\mu(V) \neq 0$, on trouve $\mu([0,1]) = +\infty$ et si $\mu(V) = 0$, on trouve $\mu([0,1]) = 0$. Les deux options sont absurdes. $\square$

Face à ce constat, l'attitude des probabilités modernes est d'accepter qu'il n'est pas possible d'attribuer une probabilité à tous les sous-ensembles de l'univers et restreindre le classe d'évènements ayant une probabilité bien définie. Borel et Lebesgue ont donc introduit la définition suivante pour clarifier comment choisir le type d'évènements acceptables.

5. Pour les insatiables : cela veut dire que cette construction utilise l'*axiome du choix*, axiome ayant un statut très particulier dans les mathématiques. Voir la page Wikipedia. En réalité, le théorème lui même dépend de l'axiome du choix. Le mathématicien américain Robert Solovay a démontré en 1970 que si l'on abandonne l'axiome du choix, il est possible de construire une probabilité uniforme qui prenne en compte tous les sous-ensembles de $[0,1]$.

Définition 1.9 (Tribu ou σ-algèbre)
Soit Ω un ensemble. On appelle *tribu ou σ-algèbre sur* Ω une famille $\mathscr{A}$ de sous-ensembles de Ω telle que

1. $\mathscr{A}$ est non vide : $\varnothing \in \mathscr{A}$,
2. $\mathscr{A}$ est stable par passage au complémentaire : $A \in \mathscr{A} \implies \overline{A} \in \mathscr{A}$,
3. $\mathscr{A}$ est stable par union dénombrable : pour toute suite $(A_n)_{n\in\mathbb{N}}$ de sous-ensembles dans $\mathscr{A}$, $\bigcup_{n\in\mathbb{N}} A_n$ est à nouveau dans $\mathscr{A}$.

Exemples :

- ⋆ L'ensemble des parties $\mathscr{P}(\Omega)$ est toujours une tribu.
- ⋆ On appelle *tribu de Borel* de l'ensemble $[0,1]$ la plus petite tribu qui contient tous les intervalles du type $[a,b]$ avec $a,b \in [0,1]$. C'est la tribu la plus souvent utilisée en probabilités. En particulier, cette tribu ne contient pas les ensembles de Vitali.

On peut enfin donner la définition moderne qui est à la base de la théorie des probabilités, donnée par le mathématicien russe Andreï Kolmogorov dans les années 1930 :

Définition 1.10 (Espace probabilisé)
Un espace probabilisé est un triplet $(\Omega, \mathscr{F}, P)$ où :

- ⋆ l'univers Ω est un ensemble quelconque.
- ⋆ l'ensemble des évènements $\mathscr{F}$ est une tribu sur Ω.
- ⋆ la mesure de probabilité P est une application $P : \mathscr{F} \longrightarrow [0,1]$ telle que
  i) $P(\Omega) = 1$.
  ii) pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'évènements deux à deux incompatibles, on a

$$P\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} P(A_i).$$

1.5 Acquis d'apprentissage

À la fin de ce chapitre, vous devrez être en mesure de :

- ⋆ Décrire les évènements élémentaires d'une expérience aléatoire.
- ⋆ Décrire un évènement grâce aux évènements élémentaires et les opérations élémentaires sur les ensembles.
- ⋆ Reconnaître si deux évènements sont incompatibles.
- ⋆ Déterminer si une fonction est une fonction de probabilité.
- ⋆ Calculer la probabilité d'un évènement en utilisant les propriétés des fonctions de probabilité.
- ⋆ Utiliser la formule de Poincaré pour trois évènements.
- ⋆ Avoir une intuition géométrique des propriétés des probabilités.

1.6 Exercices

♦ = Exercice obligatoire.

[*] = Application directe du cours : sert à vérifier que vous avez bien compris les notions et techniques de base.

[**] = Demande plus de réflexion mais vous devez être capables de le résoudre et maîtriser les techniques utilisées dans cet exercice.

[***] = Exercice difficile pour aller plus loin, découvrir de nouvelles idées ou techniques. Pour les plus motivés !

Faire le point en autonomie : QCM sur les notions de base

Soient A, B et C des évènements quelconques de Ω, l'ensemble de tous les résultats possibles d'une épreuve. On appelle $\overline{A}$ le complémentaire de A dans Ω.

Parmi les propositions suivantes, trouvez la (ou les) propositions vraies :

Item	Enoncé
A	$P(\Omega) = 1$
B	$P(\overline{A}) < P(A)$
C	Si $A \subset B$ alors $P(A) \leq P(B)$
D	Si $A \subset B$ alors $A \cup B = B$
E	$A \cup B \neq B \cup A$
F	$P(A \cup B) \leq P(A) + P(B)$
G	$P(A \cup B) = P(A) + P(B)$ quand $A \cap B = \varnothing$
H	$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B \cap C)$
I	$P(\overline{A} \cup B) + P(A \cup B) = P(A)$
J	$P(\overline{A} \cap B) + P(A \cap B) - P(B) = 0$
K	$(A \cup B) \cup C \neq A \cup (B \cup C)$

♦ **Exercice** 1.1 [*]

Soient E, F et G trois évènements. Exprimez les évènements suivants en fonction de E, F, G et les opérations élémentaires sur les ensembles :

a) Seulement E se réalise.

b) E et G se réalisent mais pas F.

c) Au moins un des trois se réalise.

d) Au moins deux des trois se réalisent.

e) Les trois se réalisent.

f) Aucun se réalise.

g) Au plus un des trois se réalise.

h) Au plus deux des trois se réalisent.

i) Au plus trois des trois se réalisent.

j) Exactement deux des trois se réalisent.

♦ **Exercice** 1.2 [*] **Dénombrement**

Dans une école d'ingénieur de 1200 étudiants, 652 pratiquent les échecs, 327 jouent un instrument de musique et 453 ne font ni échecs, ni musique.

Déterminer le nombre d'étudiants musiciens et jouant aux échecs.

♦ **Exercice** 1.3 [*] **La cardinalité**

On dispose de 12 chaises qui ne diffèrent que par leur couleur : 5 bleues, 4 vertes et 3 rouges. On les empile dans un ordre quelconque.

a) Combien peut-on former de piles différentes ?

b) Dans combien de dispositions retrouve-t-on les chaises rouges au-dessus de la pile ?

Exercice 1.4 [**] **Le quadrillage**

Combien y a-t-il de carrés dans un quadrillage de n cases sur n ?

Exercice 1.5 [**] **Paris sportifs**

Un joueur remplit une grille dans laquelle il indique ses pronostics pour 7 matchs de football à venir. Pour chacun des matchs, il peut sélectionner une des 3 cases au choix :

- ⋆ 1 : pour une victoire de l'équipe qui reçoit ;
- ⋆ N : pour un match nul ;
- ⋆ 2 : pour une victoire de l'équipe qui se déplace ;

a) De combien de façons différentes un joueur peut-il remplir la grille ?

b) Combien existe-t-il de grilles dans lesquelles tous les pronostics sont faux ?

c) Combien existe-t-il de grilles avec exactement trois pronostics corrects ?

d) Pour gagner, il faut avoir trouvé au moins 6 pronostics exacts. Quel est le nombre de grilles gagnantes ?

Exercice 1.6 [***] **La différence symétrique**

Soit Ω un ensemble. La proposition 1.1 laisse penser que la structure de $(\mathscr{P}(\Omega), \cup, \cap)$ est de celle d'un anneau. En effet, $\cup$ et $\cap$ joueraient les rôles de $+$ et $\times$, alors que Ω et $\varnothing$ joueraient les rôles de 1 et 0.

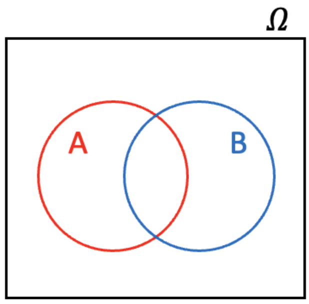
Mais il y a une différence fondamentale : dans un anneau, tout entier admet un opposé (on peut faire des soustractions), alors que les éléments de $\mathscr{P}(\Omega)$ n'admettent pas d'opposé. On peut écrire cela de la façon suivante : $\forall A \in \mathscr{P}(\Omega), A \neq \varnothing, \nexists B \in \mathscr{P}(\Omega), A \cup B = \varnothing$.

Pour avoir une structure d'anneau, on va donc définir une nouvelle opération, appelée la *différence symétrique*, notée $A \triangle B$ et définie par :

$$A \triangle B = (A \cap \overline{B}) \cup (B \cap \overline{A}).$$

$A \triangle B$ est donc l'ensemble des éléments de Ω qui appartiennent à $A \cup B$ mais pas à $A \cap B$.

a) Sur le dessin ci-dessous, hachurer l'ensemble $A \triangle B$.

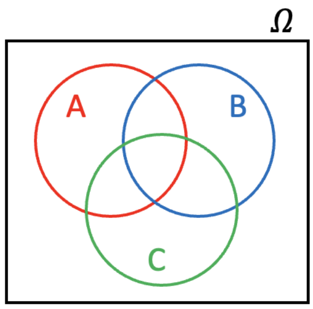


b) Soit $\Omega = \{1,2,3,4,5,6,7,8,9,10\}$, $A = \{1,2,3,4,6,7\}$ et $B = \{1,5,6,7,9\}$. Déterminer $A \triangle B$.

c) Soit $\Omega = \mathbb{R}$, $A =]-\infty, 2]$ et $B =]1; +\infty[$. Déterminer $A \triangle B$.

d) Prouver que $\triangle$ est commutative et associative.

e) Sur le dessin ci-dessous, hachurer l'ensemble $A \triangle B \triangle C$.



f) Prouver que $\varnothing$ est l'élément neutre de $\triangle$.

g) Prouver que tout élément de $\mathscr{P}(\Omega)$ admet un opposé par rapport à $\triangle$:

$$\forall A \in \mathscr{P}(\Omega), \exists B \in \mathscr{P}(\Omega), A \triangle B = \varnothing.$$

h) $\triangle$ admet-elle un élément absorbant ?

i) Prouver que $\cap$ est distributive sur $\triangle$ mais que l'inverse n'est pas vrai.

j) Prouver que $\cup$ n'est pas distributive sur $\triangle$.

Cela montre que $(\mathscr{P}(\Omega), \triangle, \cap)$ a bien une structure d'anneau, tout comme $(\mathbb{Z}, +, \times)$ ou $(\mathbb{R}, +, \times)$ par exemple. De plus, la loi $\cap$ étant commutative, l'anneau est dit commutatif.

♦ **Exercice** 1.7 [*]

Soient A et B deux évènements de l'ensemble Ω tels que $P(A) = 1/2$, $P(A \cup B) = 3/4$ et $P(\overline{B}) = 5/8$.

Calculer $P(A \cap B), P(\overline{A} \cap \overline{B}), P(\overline{A} \cup \overline{B}), P(B \cap \overline{A})$.

Exercice 1.8 [*]

60% des étudiants de l'école ne portent ni une bague ni un collier. 20% portent une bague et 30% portent un collier. Si l'on choisi un étudiant au hasard, quelle est la probabilité que l'étudiant

a) porte une bague ou un collier ?

b) porte une bague et un collier ?

♦ **Exercice** 1.9 [*]

Dans une population de 2000 nouveau-nés :

⋆ 1040 sont des garçons,

⋆ 50 présentent un ictère (maladie),

⋆ 30 sont des garçons et présentent un ictère.

On note respectivement G et I les évènements "être un garçon" et "présenter un ictère".

a) Calculer $P(G)$, $P(I \cup G)$, $P(I \cap \overline{G})$, $P(G \cap \overline{I})$.

b) Quelle est la probabilité de ne pas être un garçon et de ne pas présenter d'ictère ?

c) Les évènements I et G sont-ils compatibles ?

Exercice 1.10 [*]

Soient E et F deux évènements d'un même espace probabilisé (Ω, P).

On suppose que $P(E) = P(F) = \frac{3}{4}$. Déterminer un encadrement de $P(E \cap F)$ et $P(E \cup F)$.

Exercice 1.11 **Inégalité de Bonferroni**

Soit Ω un ensemble.

a) $[*]$ Soient E et F deux évènements de Ω. Démontrez que

$$P(E \cap F) \geq P(E) + P(F) - 1.$$

b) $[**]$ Pour $n \in \mathbb{N}^*$, soient n évènements de Ω notés $A_1, \dots, A_n$. Démontrez par récurrence que

$$P(\bigcap_{i=1}^{n} A_i) \geq \sum_{i=1}^{n} P(A_i) - (n-1).$$

Exercice 1.12 [**] **Inégalité de Boole**

Soit Ω un ensemble. Pour $n \in \mathbb{N}^*$, soient n évènements de Ω notés $A_1, \dots, A_n$.

Démontrez par récurrence que :

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i).$$

Exercice 1.13 [***] **Formule de Poincaré**

Soit $(\Omega, \mathscr{P}(\Omega), P)$ un espace probabilisé fini, $(A_i)_{1\leq i\leq n}$ une famille finie quelconque d'évènements. Montrer par récurrence que :

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n}\left((-1)^{k+1} \sum_{1\leq i_1<\ldots<i_k\leq n} P(A_{i_1}\cap\ldots\cap A_{i_k})\right)$$

Exercice 1.14 [*]

Soient E et F deux évènements. Notons G l'évènement “exactement l'un parmi E et F se réalise”. Prouvez les égalités suivantes :

a) $P(G) = P(E) + P(F) - 2P(E \cap F)$.

b) $P(E \cap \overline{F}) = P(E) - P(E \cap F)$.

c) $P(\overline{E} \cap \overline{F}) = 1 - P(E) - P(F) + P(E \cap F)$.

♦ **Exercice** 1.15 [***]

Soit Ω un ensemble infini mais dénombrable et soit P une fonction de probabilité sur Ω.

a) Prouver que P ne peut pas être telle que tous les éléments de Ω soient équiprobables.

b) Est-il possible qu'aucun élément ait une probabilité nulle de se réaliser ?

Exercice 1.16 [**]

L'école I-FREE propose à ses étudiants de deuxième année trois cours optionnels : A, B et C. Ces cours sont ouverts aux 400 étudiants de deuxième année. On dispose des informations suivantes :

- ⋆ 112 étudiants sont inscrits dans le cours A.
- ⋆ 104 étudiants sont inscrits dans le cours B.
- ⋆ 64 étudiants sont inscrits dans le cours C.
- ⋆ 48 étudiants sont inscrits dans les cours A et B.
- ⋆ 16 étudiants sont inscrits dans les cours A et C.

- $\star$ 24 étudiants sont inscrits dans les cours B et C.
- $\star$ 8 étudiants sont inscrits dans les trois cours.

On choisit un étudiant au hasard, calculer la probabilité que :

a) cet étudiant ne soit inscrit dans aucun des trois cours.

b) cet étudiant soit inscrit dans exactement un des trois cours.

Exercice 1.17 [***] **Le problème des anniversaires**

On considère un groupe de $n \geq 2$ personnes et l'évènement $E =$"il y a au moins deux personnes dans le groupe avec la même date d'anniversaire".

a) Trouver $P(E)$ (on excluera le 29 Fevrier et on supposera que le jour de naissance d'une personne peut être n'importe quel jour de l'année avec la même probabilité).

b) Trouver le taille minimale d'un groupe de personnes pour avoir $P(E) \geq 0.5$.

c) Evaluer $P(E)$ pour $n = 50$.

Exercice 1.18 [***] **Le problème des rencontres**

N personnes déposent au vestiaire leur chapeau. Suite à une panne d'électricité, chaque personne reprend du vestiaire un chapeau de façon aléatoire.

a) Quelle est la probabilité que toutes les personnes reprennent leur propre chapeau ?

b) Quelle est la probabilité qu'au moins une des personnes ne reprenne pas son propre chapeau ?

c) Quelle est la probabilité qu'aucune personne reprenne son propre chapeau ?

Aide : utilisez pour cela la formule de Poincaré et la définition des coefficients binomiaux.

d) Quelle est la limite de cette dernière probabilité lorsque N tend vers l'infini ?

Exercice 1.19 [***] **Une expérience surprenante**

Cet exercice est tiré de S. Ross, *A first course in Probability*, p. 51-53.

Supposons que l'on dispose d'une urne infinie et d'une collection infinie de balles, numérotées 1,2,3,.... Considérons l'expérience de pensée suivante :

- $\star$ une minute avant minuit, on place les balles 1 à 10 dans l'urne puis on en retire la balle 10.
- $\star$ 1/2 minute avant minuit, on place les balles 11 à 20 dans l'urne puis on en retire la balle 20.
- $\star$ 1/4 minute avant minuit, on place les balles 21 à 30 dans l'urne puis on en retire la balle 30.
- $\star$ 1/8 minute avant minuit, on place les balles 31 à 40 dans l'urne puis on en retire la balle 40.
- $\star$ etc ...

a) Les balles 1 à 9, puis 11 à 19, et ainsi de suite ont-elles une chance d'être enlevées ? Ainsi, combien de balles y aura-t-il dans l'urne à minuit ?

b) Maintenant, on modifie un peu l'expérience. Au lieu de retirer d'abord la balle numéro 10, puis 20, puis 30, puis 40, ..., on retire d'abord la balle numéro 1, puis 2, puis 3, puis 4, ...

i) Puisqu'à chaque étape on rajoute plus de balles qu'on n'en enlève, combien de balles diriez vous qu'il y a dans l'urne à minuit ?

ii) Y aura-t-il une balle qui ne sera pas enlevée lorsqu'on s'approche suffisamment près de minuit ?

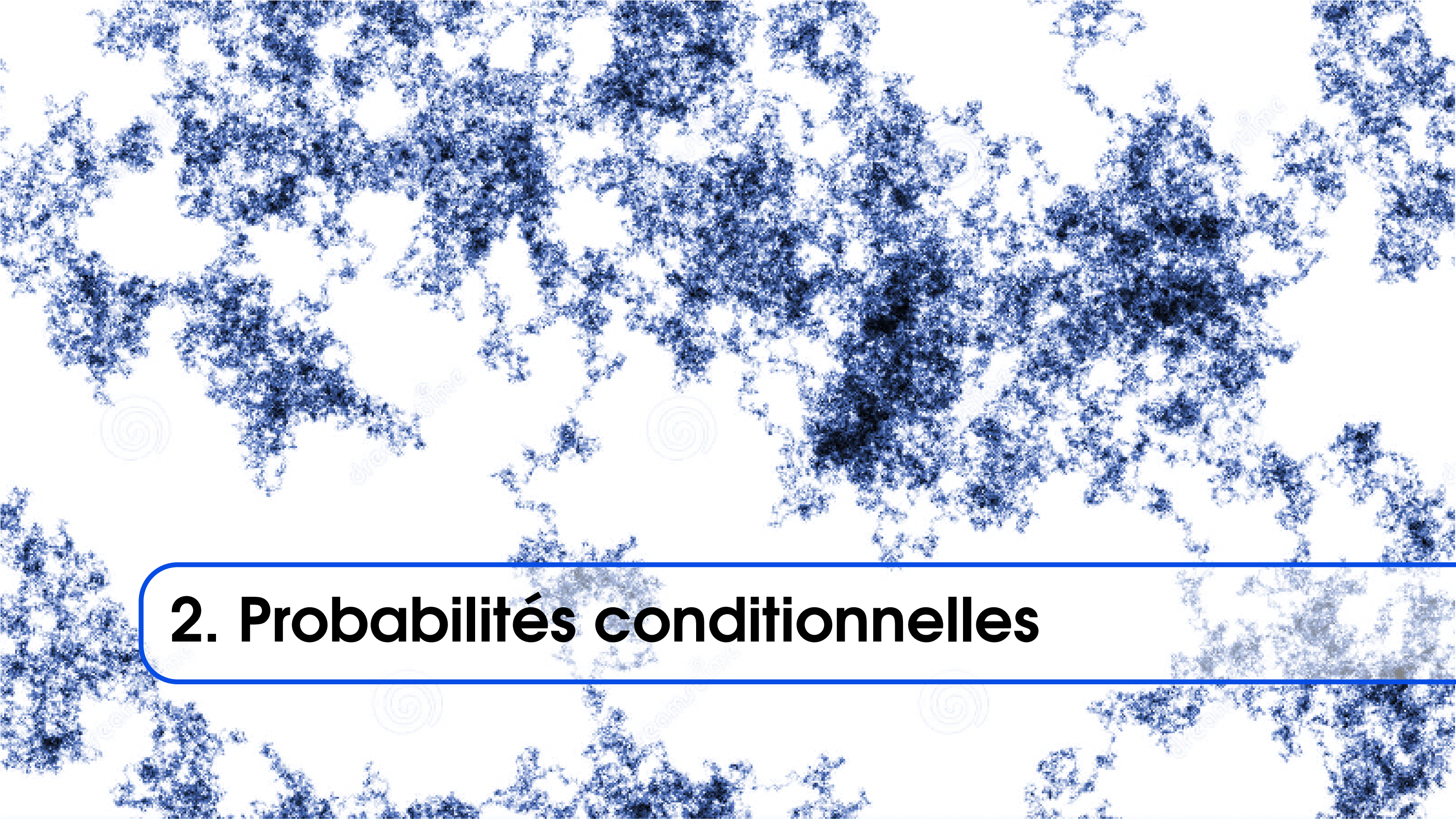
iii) Ainsi, à minuit pile, combien de balles y aura-t-il dans l'urne ?

c) Enfin, dernière modification, au lieu de choisir la balle que l'on retire dans un ordre précis, on la choisit à chaque étape de façon aléatoire.

i) Pour une balle n donnée, calculer la probabilité $P_k(n)$ qu'elle soit encore dans l'urne à l'étape k.

ii) Combien vaut la limite $\lim_{k\to+\infty} P_k(n)$?

iii) Ainsi, à minuit pile, combien de balles y-a-t-il dans l'urne ?



2. Probabilités conditionnelles

Dans les calculs pratiques sur les probabilités, nous sommes rapidement confrontés à deux types de situations qui nous poussent à introduire le concept de “probabilité conditionnelle” :

- ⋆ D’abord, il y a des situations où nous possédons des *informations partielles* sur le résultat de l’expérience aléatoire, informations qu’il faut prendre en compte car, souvent, elles modifient les probabilités du résultat. La probabilité qu’une personne ait moins de 20 ans n’est pas la même si l’on sait, en plus, que cette personne est étudiante à l’EFREI ou qu’elle est capable de prouver le théorème central limite.
- ⋆ Ensuite, même quand des informations partielles ne sont pas connues, il est souvent plus facile de calculer la probabilité que l’on veut en faisant une distinction de cas : si l’on veut calculer la probabilité pour qu’un citoyen de l’union européenne vote pour le parti écologiste aux prochaines élections européennes, le plus facile est de regarder ce qui se passe pays par pays (probabilité sachant qu’il est francais, allemand, hongrois, etc).

2.1 Définition et propriétés de base

Définition 2.1 (Probabilité conditionnelle)
Soit (Ω, P) un espace probabilisé et soient E et F deux évènements de Ω tels que F ne soit pas impossible. Alors, la *probabilité de E sachant F* (ou probabilité conditionnelle de E en F), notée $P(E|F)$ ou $P_F(E)$, est définie par :

$$P_F(E) = P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Cette définition se comprend bien : tout se passe comme si F s'était produit, donc on ne regarde plus que l'ensemble F qui “remplace” en quelque sorte l'univers Ω.

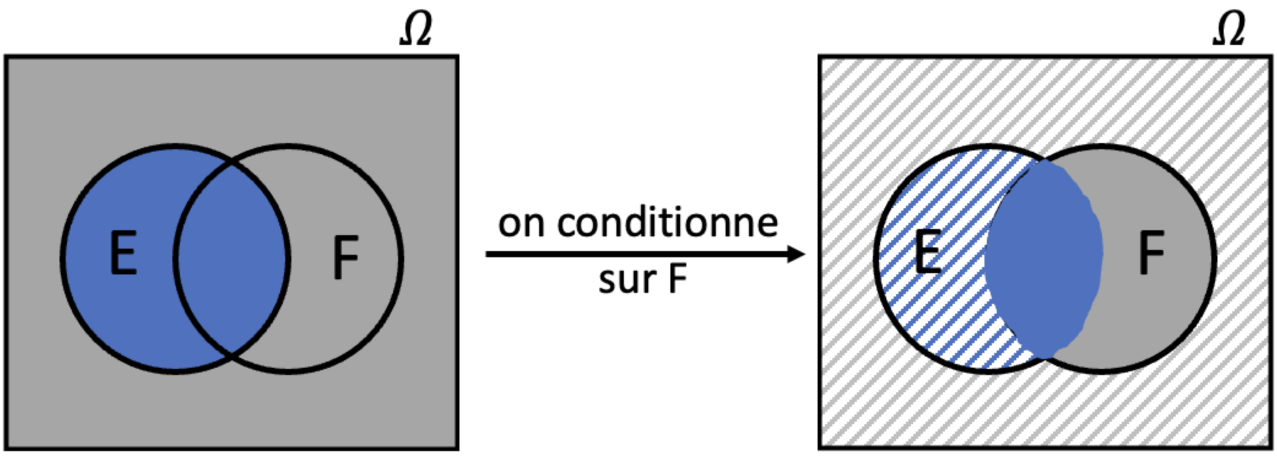
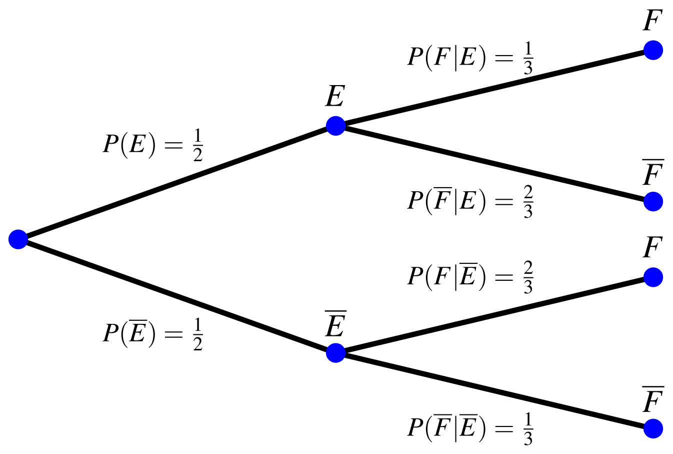


FIGURE 2.1 – La probabilité de réalisation de l'évènement E est le rapport de la surface bleue sur celle de l'univers des possibles (surface grise). Lorsque l'on conditionne sur F (figure de droite), l'univers des possibles se rétrécit et on ignore la partie hachurée.

♦<u>Exemple :</u> Soit l'expérience suivante : on tire, successivement et sans remise, deux boules dans une urne contenant initialement deux boules noires et deux boules blanches.

Soient E l'évènement "tirer une boule blanche lors du premier tirage" et F l'évènement "tirer une boule blanche lors du deuxième tirage". Il est clair que $P(E) = \frac{2}{4} = \frac{1}{2}$. Aussi, le fait que E se soit produit ou non change la probabilité que F se produise. En effet, $P(F|E) = \frac{1}{3}$ et $P(F|\overline{E}) = \frac{2}{3}$. Modélisons ces embranchements à l'aide de l'arbre pondéré suivant :



La définition de la probabilité conditionnelle nous donne la probabilité que les deux boules blanches soient tirées : $P(E \cap F) = P(F|E) \times P(E) = \frac{1}{6}$. On "lit la probabilité de l'intersection" en multipliant les probabilités conditionnelles apparaissant le long de l'arbre. On remarque qu'à chaque embranchement, la somme des probabilités fait 1. Aussi, on peut résumer ces informations dans le *tableau de contingence* suivant :

	E	$\overline{E}$
F	$P(F \cap E) = \frac{1}{6}$	$P(F \cap \overline{E}) = \frac{1}{3}$
$\overline{F}$	$P(\overline{F} \cap E) = \frac{1}{3}$	$P(\overline{F} \cap \overline{E}) = \frac{1}{6}$

Proposition 2.1 : Formule des évènements contraires

Soit (Ω, P) un espace probabilisé et soient E et F deux évènements de Ω tels que F ne soit pas impossible. Alors, on a

$$P(E|F) + P(\overline{E}|F) = 1.$$

Démonstration. Ce résultat devrait être évident intuitivement. Formellement, on a, par définition :

$$P(E|F) + P(\overline{E}|F) = \frac{P(E \cap F) + P(\overline{E} \cap F)}{P(F)}.$$

Or, puisque les évènements $E \cap F$ et $\overline{E} \cap F$ sont disjoints, on a

$$P(E \cap F) + P(\overline{E} \cap F) = P\left((E \cap F) \cup (\overline{E} \cap F)\right) = P\left((E \cup \overline{E}) \cap F\right) = P(\Omega \cap F) = P(F).$$

Ce qui montre bien le résultat. $\square$

La terminologie de probabilité conditionnelle est justifiée puisque :

Proposition 2.2

Soient (Ω, P) un espace probabilisé et F un évènement possible. Alors la fonction sur $\mathscr{P}(\Omega)$ définie par :

$$\begin{aligned} P(\cdot|F) : \mathscr{P}(\Omega) &\longrightarrow \mathbb{R} \\ E &\longmapsto P(E|F) \end{aligned}$$

est une fonction de probabilité sur Ω.

Démonstration. Voir exercice 2.2. □

Proposition 2.3 : Propriété de multiplication

Soit (Ω, P) un espace probabilisé et $E_1, \ldots, E_n$ une famille d'évènements possibles. Alors, on a

$$\begin{aligned} P(E_1 \cap E_2) &= P(E_2|E_1)P(E_1), \\ P(E_1 \cap E_2 \cap E_3) &= P(E_3|E_1 \cap E_2)\, P(E_2|E_1)\, P(E_1) \end{aligned}$$

et, plus généralement,

$$P\left(\bigcap_{1\leq i\leq n} E_i\right) = P\left(E_n| \bigcap_{1\leq i\leq n-1} E_i\right) P\left(E_{n-1}| \bigcap_{1\leq i\leq n-2} E_i\right) \cdots P(E_2|E_1)\, P(E_1).$$

Démonstration. Pour deux évènements, la formule est juste une réécriture de la définition de la probabilité conditionnelle. Pour trois évènements, on a

$$P(E_1 \cap E_2 \cap E_3) = \frac{P\big((E_1 \cap E_2) \cap E_3\big)}{P(E_1 \cap E_2)} \times \frac{P(E_1 \cap E_2)}{P(E_1)} \times P(E_1) = P(E_3|E_1 \cap E_2)\, P(E_2|E_1)\, P(E_1).$$

Même idée pour une famille de n évènements. □

♦Exemple : A Paris, il pleut en moyenne 111 jours par an. A Bogota, par contre, il pleut en moyenne 181 jours par an et à Quito il pleut 128 jours par an. Sachant que ce texte a été écrit pendant l'été et que l'auteur de cette phrase a passé 20% de l'été à Paris, 20% à Bogota et 60% à Quito, calculer la probabilité que cette phrase ait été écrite à Bogota en regardant la pluie par la fenêtre.

On pose les évènements $B =$"l'auteur de cette phrase était à Bogota au moment de l'écriture", $Pl =$"il pleuvait au moment de l'écriture", on cherche à calculer $P(B \cap Pl)$. D'après les données, on a

$$P(B \cap Pl) = P(B) \times P(Pl|B) = \frac{20}{100} \times \frac{181}{365} \approx 0,1.$$

Définition 2.2 (Couple d'évènements indépendants)
Soit (Ω, P) un espace probabilisé et E et F deux évènements. On dit que E et F sont *indépendants* si

$$P(E \cap F) = P(E)P(F).$$

Si E et F sont des évènements possibles, ceci est équivalent à demander

$$P(E|F) = P(E) \text{ ou } P(F|E) = P(F).$$

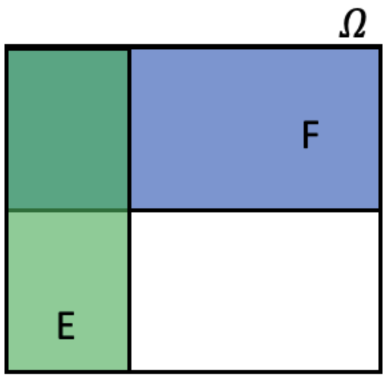


FIGURE 2.2 – E et F sont indépendants car les proportions ne changent pas lorsque l'on réduit l'univers à E ou à F : $P(E) = P(E|F) = 1/3$ et $P(F) = P(F|E) = 1/2$.

Remarque : Sur cette figure, on peut s'amuser à modifier la croix centrale, c'est à dire à déplacer la barre centrale verticale, de gauche à droite, ou la barre centrale horizontale, de haut en bas. L'indépendance est alors conservée. Par contre, si une de ces deux barres s'incline, l'indépendance n'est plus vérifiée.

Exemples :

1. Les évènements $E =$"gagner au loto" et $F =$"porter des chaussettes bleues" sont indépendants.
2. Les évènements $E =$"la monnaie est tombée du côté pile aux dix derniers lancers" et $F =$"la monnaie va tomber du côté pile au prochain lancer" sont indépendants.
3. Les évènements $E =$"mesurer plus de deux mètres de haut" et $F =$"peser plus de 100kg" ne sont pas indépendants.
4. Les évènements $E =$"il y a des nuages dans le ciel" et $F =$"il va pleuvoir dans les cinq minutes qui suivent" ne sont pas indépendants.

Proposition 2.4

Soit (Ω, P) un espace probabilisé et E et F deux évènements. On a

$$E \text{ et } F \text{ indépendants} \iff E \text{ et } \overline{F} \text{ indépendants.}$$

Démonstration. Voir exercice 2.3. □

Pour des familles de plus de deux évènements, on pourrait s'attendre à que ce soit suffisant de regarder si tous les couples d'évènements sont indépendants. Mais l'exemple suivant montre que cela n'est pas le cas.

♦Exemple : On jette deux dés non truqués. On considère les évènements suivants :

$$\begin{aligned} E &= \text{ la somme des dés est 7,} \\ F &= \text{ le résultat du premier dé est 4,} \\ G &= \text{ le résultat du deuxième dé est 3.} \end{aligned}$$

On a

$$P(E|F) = P(E|G) = P(F|G) = P(E) = P(F) = \frac{1}{6}$$

donc (E,F) (E,G) et (F,G) sont trois couples d'évènements indépendants. Or, clairement, E et $F \cap G$ ne sont pas indépendants puisque $P(E|F \cap G) = 1$.

Cet exemple motive la définition suivante :

Définition 2.3 (Triplet d'évènements indépendants)
Soient (Ω, P) un espace probabilisé et E, F et G trois évènements. On dit que E,F et G sont *indépendants* si

$$\begin{aligned} P(E \cap F) &= P(E)P(F) \\ P(E \cap G) &= P(E)P(G) \\ P(F \cap G) &= P(F)P(G) \\ P(E \cap F \cap G) &= P(E)P(F)P(G). \end{aligned}$$

2.2 Formule de Bayes

L'objectif de la formule de Bayes est de nous permettre de passer de $P(E|A)$ à $P(A|E)$. Il y a beaucoup de situations où ce problème est de grand intérêt. Prenons un exemple concret simple de statistiques en Médecine, lorsque l'on compte suffisamment d'entrées de patients avec une certaine pathologie dans un service. On reçoit des patients malades et l'on peut aisément calculer la proportion qui sont, par exemple, vaccinés. On a donc deux évènements : malade M et vacciné V. Grâce à ce décompte, on accède à la proportion de vacciné parmi les malades entrants, c'est à dire la probabilité $P(V|M)$. Mais ce qui nous intéresse est de connaître la probabilité de tomber malade sachant que l'on est vacciné, c'est à dire $P(M|V)$ [1].

Proposition 2.5 : Formule des probabilités totales

Soit $\mathscr{A} = \{A_1, \ldots, A_n\}$ un système complet d'évènements et E un évènement quelconque. Alors, on a

$$P(E) = \sum_{i=1}^{n} P(E|A_i)P(A_i).$$

Remarque : En particulier, on a, pour deux évènements possibles E et A :

$$P(E) = P(E|A)P(A) + P(E|\overline{A})P(\overline{A}).$$

Démonstration. Puisque $\mathscr{A}$ est un système complet d'évènements, on a

$$E = \bigcup_{1 \leq i \leq n} (E \cap A_i) \text{ et, pour } k \neq j, (E \cap A_k) \cap (E \cap A_j) = \varnothing.$$

Donc, on a bien

$$P(E) = \sum_{i=1}^{n} P(E \cap A_i) = \sum_{i=1}^{n} P(E|A_i)P(A_i).$$

□

♦Exemple : Une élève qui vient tout juste d'intégrer l'EFREI en première année veut évaluer ses chances de rentrer dans la majeure cybersécurité en première année de Master. Cela dépend, bien sûr, de son classement général à la fin de la troisième année de licence : elle est sûre d'intégrer cette majeure si elle est dans les 50 meilleurs étudiants, 80% de chances si elle est entre le 51-ième et le 200ième, 50% de chances si elle est entre le 201-ième et le 400-ième étudiant et 10% de chances si

1. Mais attention, pour obtenir ce passage M à V, il y a tout un protocole issu des statistiques à respecter. Il faut d'abord s'assurer du tirage aléatoire des patients en entrée, ce qui n'est à priori pas le cas. Aussi pour prouver l'influence de tel ou tel évènement sur l'évènement M, comme une prise de médicaments par exemple, il faut se doter d'un protocole plus lourd appelé test en aveugle ou en double aveugle.

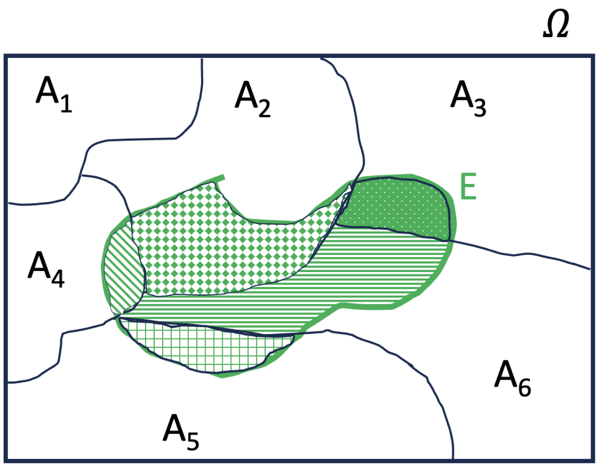


FIGURE 2.3 – Illustration de la formule des probabilités totales : la surface de E est la somme des surfaces intersectant les différentes régions A_i.

elle est plus bas dans le classement. De plus, on suppose que toutes les places du classement sont équiprobables pour un étudiant rentrant en L1 (il y a 600 étudiants en L3) et que 20% des étudiants rentrant en L1 quittent l'école avant de valider la L3. Quelle est la probabilité pour que cette nouvelle étudiante atteigne son objectif ?

Introduisons les évènements

$$\begin{aligned}
A &: \text{“être parmi les 50 meilleurs étudiants”},\\
B &: \text{“être entre le 51ème et 200ième”},\\
C &: \text{“être entre le 201ème et 400ième”},\\
D &: \text{“être en-dessous de la 400ième place”},\\
F &: \text{“quitter l'école avant la fin de la troisième année”},\\
M &: \text{“intégrer la majeure cybersécurité”}.
\end{aligned}$$

Les cinq premiers évènements forment un système complet d'évènements. Par exemple, on a

$$P(A) = P(A \cap \overline{F}) = P(A|\overline{F})P(\overline{F}) = 0.8 \times \frac{50}{600}.$$

Ainsi, la probabilité que l'on cherche est donnée par :

$$\begin{aligned}
P(M) &= P(M|A)P(A) + P(M|B)P(B) + P(M|C)P(C) + P(M|D)P(D) + P(M|F)P(F)\\
&= 0.8 \times \frac{50}{600} + 0.8 \times 0.8 \times \frac{150}{600} + 0.5 \times 0.8 \times \frac{200}{600} + 0.1 \times 0.8 \times \frac{200}{600} + 0 \times 0.2\\
&= \frac{29}{75} \approx 0.39.
\end{aligned}$$

Théorème 2.6 : Formule de Bayes

Soit $\mathscr{A} = \{A_1, \ldots, A_n\}$ un système complet d'évènements et E un évènement quelconque. Alors, on a

$$P(A_j|E) = \frac{P(E|A_j)P(A_j)}{\sum_{i=1}^{n} P(E|A_i)P(A_i)}.$$

Démonstration. Il suffit de combiner la formule des probabilités totales et la définition des probabilités conditionnelles pour obtenir la formule de Bayes. En effet :

$$P(A_j|E) = \frac{P(A_j \cap E)}{P(E)} = \frac{P(E|A_j)P(A_j)}{P(E)} = \frac{P(E|A_j)P(A_j)}{\sum_{i=1}^{n} P(E|A_i)P(A_i)}.$$

□

Remarque : Souvent, les évènements A_j sont pensés comme étant des "hypothèses" et dans ce cas la formule de Bayes nous montre comment modifier nos opinions sur ces hypothèses lorsqu'un évènement E se réalise (voir 2.3 pour plus de détails sur ce point de vue sur les probabilités comme mesure de confiance).

♦Exemple : Un médecin est appelé pour voir un enfant malade. Le médecin dispose d'informations préalables selon lesquelles 90% des enfants malades dans ce quartier ont la grippe, tandis que les 10% restants sont malades de la rougeole. Supposons pour simplifier qu'il n'y a pas d'autres maladies dans ce quartier. Un symptôme bien connu de la rougeole est une éruption cutanée. Supposons que la probabilité d'avoir une éruption cutanée si l'on a la rougeole est 0.95. Cependant, il arrive que des enfants atteints de grippe développent également des éruptions cutanées, et la probabilité d'avoir une éruption cutanée en cas de grippe est 0.1. En examinant l'enfant, le médecin trouve une éruption cutanée. Quelle est la probabilité que l'enfant ait la rougeole ?

On pose les évènements

$$G = \text{"avoir la grippe"}, \quad R = \text{"avoir la rougeole"}, \quad EC = \text{"avoir une éruption cutanée"}$$

On cherche à calculer $P(R|EC)$. Les évènements G et R forment un système complet d'évènements et, d'après la formule de Bayes, on a :

$$P(R|EC) = \frac{P(EC|R)P(R)}{P(EC|R)P(R) + P(EC|G)P(G)} = \frac{0.95 \times 0.1}{0.95 \times 0.1 + 0.1 \times 0.9} \approx 0.51.$$

L'enfant a donc presque autant de chances d'avoir la rougeole que la grippe !

Point historique #1 (Thomas Bayes et “sa” formule)
Thomas Bayes (1701-1761) est un mathématicien britannique qui était destiné à l’oubli : ses travaux sur les probabilités furent publiés en 1761 après sa mort et trouvèrent très peu de lecteurs. Mais sa formule fut redécouverte en 1774 par le jeune et brillant mathématicien français Pierre-Simon Laplace. C’est Laplace qui lui donna le nom de “théorème de Bayes” en 1781 après la visite de Richard Price, ami de Bayes, qui lui appris l’existence des travaux du mathématicien britannique.
De plus, Laplace mit la formule de Bayes au centre d’une nouvelle interprétation des probabilités, appelée aujourd’hui Bayésienne même si rien ne montre que Thomas Bayes lui-même souscrivait à cette interprétation.

FIGURE 2.4 – Thomas Bayes et Pierre-Simon Laplace.

2.3 (Hors programme) : un peu de philosophie

Le concept de probabilité est le concept le plus important des sciences modernes, d'autant plus que personne n'a la moindre idée de ce qu'il signifie.

Bertrand Russell - Prix Nobel de littérature en 1950

Après les travaux d'Andreï Kolmogorov dans les années 1930, un mathématicien n'a pas de doutes de ce qu'est une probabilité : il prendra tout simplement la définition abstraite 1.7, page 16 (ou plus exactement la définition encore plus abstraite 1.10, page 24). Mais, dès que l'on sort du cadre strict des mathématiques –est c'est justement en dehors des mathématiques que les probabilités acquièrent leur importance– la signification précise du concept de "probabilité" est bien plus évasif de ce que l'on pourrait croire naïvement.

Des affirmations telles que "la probabilité que je tombe sur six au prochain lancer du dé est 1/6" ou bien "la probabilité que cet atome de carbone 14 se désintègre dans les prochaines 5734 années est de 0.5" sont prises comme des descriptions objectives d'une certaine réalité du monde. Ici, leur interprétation semble assez claire : par exemple la première affirmation dit que, si je répétais le lancer de dés 600 fois, je tomberais sur 6 *environ* cent fois et que si je faisais une infinité de lancers, la proportion des fois où je tomberais sur 6 serait *exactement* un sixième.

Cette interprétation des probabilités est appelée l'interprétation fréquentiste :

> **Interprétation fréquentiste**
> La probabilité de l'évènement E est comprise comme la fréquence relative d'apparition de l'évènement E dans une série infinie de répétitions.

Cette interprétation était implicite chez De Moivre, Poisson et Gauss et a été définie explicitement par Mills, Ellis, Cournot et Fries. C'était le point de vue dominant dans les sciences jusqu'au début du vingtième siècle et ça continue sans doute à l'être dans la discussion courante aujourd'hui.

Mais, au delà des problèmes évidents liés à l'apparition de l'infini dans cette définition [2], l'interprétation fréquentiste devient plus problématique pour des affirmations comme "la probabilité de pluie pour demain est de 80%" ou, encore pire, "la probabilité de victoire de Carlos Alcaraz au prochain US Open de tennis est de 30%". Le problème principal est, clairement, que les évènements en question ne peuvent pas être répétés. Mais, problème encore plus profond, on peut mettre en doute le caractère *objectif* des affirmations précédentes et les voir comme des affirmations sur les connaissances et croyances *subjectives* de la personne qui les proclame.

---

2. Prenez un résultat de probabilités qui soit contre-intuitif, par exemple celui de l'exercice 1.17, et essayez de convaincre par l'expérience vos amis non scientifiques qui sont réticents à accepter le résultat. Vous allez voir que, très souvent, l'expérience vous donnera tort, ce à quoi vous rétorquerez, un peu embarrassé : "si l'on avait répété l'expérience plus de fois, vous auriez vu que le résultat est vrai".

Pour donner un exemple facile et concret : la probabilité que, sur 10000 lancers d'une monnaie, il y ait exactement 5000 piles et 5000 faces est de seulement 0.008 !

Interprétation Bayésienne ou subjective
La probabilité de l'évènement E est une mesure de la croyance subjective en la réalisation future de l'évènement E étant données les connaissances subjectives actuelles.

Ainsi, lorsque j'affirme que “la probabilité que je tombe sur six au prochain lancer du dé est 1/6”, cela montre en réalité que j'ignore tout sur la vitesse et le moment cinétique du dé au moment où il quitte ma main et sur la répartition (équitable ou pas) du poids dans le dé.

Selon l'approche Bayésienne, les probabilités sont toujours des probabilités conditionnelles (ça n'a pas de sens de parler d'une probabilité dans l'absolu) et le théorème 2.6 décrit précisément comment fonctionne l'aprentissage : la formule de Bayes décrirait comment évoluent nos croyances en certaines informations A_j au fur et à mesure que nous découvrons l'existence de certains évènements E. Cette deuxième approche des probabilités a été très féconde en statistiques et a donné lieu à une méthode appelée l'inférence Bayésienne.

Voici une liste de liens pour explorer davantage le sujet :

- Page Wikipédia de l'approche fréquentiste,
- Page Wikipédia de l'approche Bayésienne,
- Article du Stanford Encyclopedia of Philosophy sur les différentes interprétations des probabilités,
- Article du Stanford Encyclopedia of Philosophy sur l'interprétation de la formule de Bayes.

2.4 Acquis d'apprentissage

À la fin de ce chapitre, vous devrez être en mesure de :

- ⋆ Calculer la probabilité de E sachant F, pour E, F deux évènements quelconques.
- ⋆ Interpréter les probabilités conditionnelles en termes d'arbres.
- ⋆ Utiliser des tableaux de contingence pour calculer des probabilités conditionnelles.
- ⋆ Reconnaître si E et F sont deux évènements indépendants.
- ⋆ Reconnaître un système complet d'évènements.
- ⋆ Utiliser la formule de multiplication pour calculer la probabilité $P(E \cap F)$.
- ⋆ Reconnaître dans quelles situations utiliser la formule des probabilités totales.
- ⋆ Utiliser la formule de Bayes.

2.5 Exercices

♦ = Exercice obligatoire.

[*] = Application directe du cours : sert à vérifier que vous avez bien compris les notions et techniques de base.

[**] = Demande plus de réflexion mais vous devez être capables de le résoudre et maîtriser les techniques utilisées dans cet exercice.

[***] = Exercice difficile pour aller plus loin, découvrir de nouvelles idées ou techniques. Pour les plus motivés !

Faire le point en autonomie : QCM sur les notions de base

Soient A et B des évènements quelconques de Ω, l'ensemble de tous les résultats possibles d'une épreuve. Parmi les propositions suivantes, trouvez la (ou les) propositions vraies :

Item	**Enoncé**
A	B étant possible, on a $P(A\|B) = P_B(A) = \dfrac{P(A\cap B)}{P(B)}$.
B	A et B étant deux évènements possibles, on a $P(A\|B) = P(B\|A)$.
C	B étant possible, on a $P(A\|B) + P(\overline{A}\|B) = 1$.
D	$A = (A\cap B)\cup(A\cap\overline{B})$.
E	A et B étant deux évènements possibles, on a $P(B\|A) = \dfrac{P(B)P(A\|B)}{P(B)P(A\|B)+P(\overline{B})P(A\|\overline{B})}$

♦ **Exercice** 2.1 [*] **Indépendance vs. Incompatibilité**

a) Donner un exemple de deux évènements qui sont compatibles et indépendants.

b) Donner un exemple de deux évènements qui sont compatibles et ne sont pas indépendants.

c) Donner un exemple de deux évènements qui sont incompatibles et ne sont pas indépendants.

d) Prouver que deux évènements possibles et incompatibles ne peuvent pas être indépendants.

Exercice 2.2 [**] *Démonstration de la proposition 2.2*

Soient (Ω, P) un espace probabilisé et $F \subset \Omega$. On considère la fonction suivante :

$$\begin{aligned} P(\cdot|F) : \mathscr{P}(\Omega) &\longrightarrow \mathbb{R} \\ E &\longmapsto P(E|F) \end{aligned}$$

Montrez que $P(\cdot|F)$ est une fonction de probabilité sur Ω.

Exercice 2.3 [**] *Démonstration de la proposition 2.4*

Soient (Ω, P) un espace probabilisé et $E, F \subset \Omega$. Démontrez que l'on a

$$E \text{ et } F \text{ indépendants} \iff E \text{ et } \overline{F} \text{ indépendants.}$$

Exercice 2.4 [**] **Indépendance**

Soit $n \in \mathbb{N}^*$ un entier naturel non nul. On effectue n lancers indépendants d'une pièce pour laquelle la probabilité d'obtenir pile est p, avec $p \in]0;1[$.

On pose $q = 1 - p$.

a) Quelle est la probabilité d'obtenir au moins une fois pile ?

b) Quelle est la probabilité qu'au cours de ces n lancers, face ne soit jamais suivi de pile.

♦ **Exercice** 2.5 [**]

A et B sont deux personnes qui jouent aux dés en jetant alternativement une paire de dés. Le premier des joueurs qui obtient un total de sept avec les deux dés est déclaré vainqueur. A commence à jouer.

Calculer la probabilité que A soit vainqueur, puis que B soit vainqueur (on commencera par tracer un arbre donnant les déroulements possibles du jeu).

Exercice 2.6 [***] **Le jeu de craps**

Le craps est un jeu de dés solitaire où le joueur jette simultanément deux dés. Les règles sont les suivantes :

Le joueur jette un premier lancer de dé :

- ⋆ Si la somme des dés est 2,3 ou 12, alors le joueur perd.
- ⋆ Si la somme des dés est 7 ou 11, alors le joueur gagne.

Sinon, le joueur continue à jeter simultanément les deux dés jusqu'à tomber :

- ⋆ soit sur 7, alors il perd.
- ⋆ soit sur la somme obtenue au premier lancer, alors il gagne.

Calculer la probabilité de victoire du joueur.

♦ **Exercice** 2.7 [*]

Sur une série de 42 enfants hospitalisés dans un grand hôpital parisien, des anomalies à l'IRM sont constatées chez 16 des 20 malades qui vont développer des séquelles et chez 2 malades qui guériront sans séquelle.

a) Remplir le tableau de contingence grâce aux données de l'énoncé.

	S : Séquelles	$\overline{S}$: Pas de séquelle	Total
A : IRM anormale			
N : IRM normale			
Total			

b) Calculer $P(N|\overline{S})$, $P(A|S)$ et $P(S|A)$.

c) Est-ce que A et S sont des évènements indépendants ?

Exercice 2.8 [*]

Dans un groupe de 9 personnes on étudie deux évènements pour chaque personne de ce groupe : $J =$"la personne pratique le judo" et $E=$ "la personne pratique les échecs". On dresse alors le tableau des effectifs suivant :

	E	$\overline{E}$
J	1	2
$\overline{J}$	2	4

a) Remplir le tableau de probabilités suivant :

	E	$\overline{E}$
J	$P(J\cap E)$	$P(J\cap\overline{E})$
$\overline{J}$	$P(\overline{J}\cap E)$	$P(\overline{J}\cap\overline{E})$

b) Les évènements J et E sont-ils indépendants ?

Exercice 2.9 [*]

Dans la population des sujets qui développent une pathologie coronarienne, la proportion de fumeurs (plus de 10 cigarettes par jour) vaut 60%. De plus, une hypertension artérielle est présente chez 75% des fumeurs et chez 50% des non-fumeurs. Un sujet est tiré au sort dans la population.

a) Quelle est la probabilité que le sujet soit hypertendu et fumeur ?

b) Quelle est a probabilité que le sujet soit hypertendu ?

c) Quelle est la probabilité que le sujet ne soit ni hypertendu ni fumeur ?

d) Quelle est maintenant la probabilité qu'un sujet hypertendu soit fumeur ?

Exercice 2.10 [**] **Information d'un évènement sur un autre**

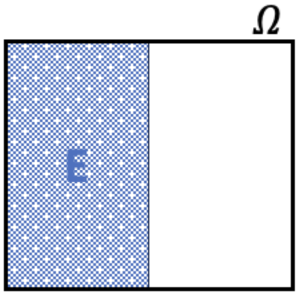
Cet exercice est tiré de S. Ross, *A first course in Probability*, p. 107.

Soient E et F deux évènements possibles dans Ω. F est dit "posséder une information négative" sur E, noté $F \searrow E$, si

$$P(E|F) \leq P(E).$$

(Autrement dit : la réalisation de F diminue la probabilité de la réalisation de E.)

a) Dans le dessin suivant, tracer un ensemble F qui satisfait à la propriété $F \searrow E$.



Pour chacune des affirmations suivantes, prouver si elles sont vraies ou fausses (vous pouvez vous appuyer sur des dessins) :

b) $F \searrow E \implies E \searrow F$.

c) $F \searrow E$ et $E \searrow G \implies F \searrow G$.

d) $F \searrow E$ et $G \searrow E \implies F \cap G \searrow E$.

Mêmes questions pour la relation $F \nearrow E$ (F "possède une information positive" sur E) définie par $P(E|F) \geq P(E)$.

Exercice 2.11 [***] **Indicatrice d'Euler**

D'après un sujet de Mines-Ponts.

Soit $n = p_1^{\alpha_1} \times ... \times p_r^{\alpha_r}$ un entier naturel non nul, décomposé en produit de facteurs premiers. On note $\varphi(n)$ (et on appelle fonction indicatrice d'Euler) le nombre d'entiers compris entre 1 et n qui sont premiers avec n.

On se propose de démontrer que

$$\varphi(n) = n\left(1 - \frac{1}{p_1}\right) \times ... \times \left(1 - \frac{1}{p_r}\right)$$

Soit $\Omega = [\![1;n]\!]$, muni de la probabilité uniforme.

a) Si d est un diviseur de n, on note D_d l'ensemble des multiples de d dans Ω. Calculer $P(D_d)$.

b) Montrez que $D_{p_1}, D_{p_2}, ..., D_{p_r}$ sont mutuellement indépendants.

c) En déduire la formule pour $\varphi(n)$.

♦ **Exercice** 2.12 [*] **Un problème de diagnostic**

Si une personne est porteur de la maladie M, un test de laboratoire a une efficacité de 95% pour le détecter. Cependant, ce même test présente 1% de "faux-positifs" : parmi les personnes non-porteuses de la maladie qui font le test, 1% auront un résultat positif. Sachant que .05% de la population est porteuse de la maladie, quelle est la probabilité que je sois porteur si mon test a été positif ?

Exercice 2.13 [**]

Au cours d'une épidémie, 20% des personnes vaccinées et 60% des personnes non vaccinées sont malades. Parmi les malades, la proportion de personnes vaccinées vaut 10%.

a) Déterminez la proportion de malades.

b) Et la proportion de vaccinés.

Exercice 2.14 [*]

Cet exercice est tiré de S. Ross, *A first course in Probability*, p. 82-83.

Dans une clinique psychiatrique, les assistants sont si occupés que, en moyenne, seulement 60% des nouveaux patients potentiels qui téléphonent peuvent parler immédiatement à un assistant social lorsqu'ils appellent. Au 40% restant, on leur demande de laisser leur numéro de téléphone : 75% des fois, les assistants les rappellent le jour même et 25% des fois ils sont rappelés le lendemain. L'expérience montre que la probabilité que l'appelant se rende effectivement à la clinique est de 0.8 s'il arrive à parler immédiatement à un assistant, de 0.6 s'il est rappelé le jour même et 0.4 s'il est rappelé le lendemain.

a) Quel pourcentage des appelants se rend effectivement à la clinique ?

b) Quel pourcentage des patients qui se sont rendus à la clinique ont du être rappelés ?

♦ **Exercice** 2.15 [*]

Une compagnie d'assurance automobile a classé ses assurés en trois classes d'âges :

⋆ moins de 25 ans,

⋆ de 25 ans à 50 ans,

⋆ plus de 50 ans.

Le tableau ci-dessous fournit deux informations :

⋆ La proportion d'assurés appartenant à chaque classe,

⋆ La probabilité qu'un assuré d'une classe donnée déclare au moins un accident au cours de l'année.

Classe	**Proportion**	**Probabilité**
moins de 25 ans	0,25	0,12
de 25 ans à 50 ans	0,53	0,06
plus de 50 ans	0,22	0,09

a) Un assuré est tiré au hasard dans le fichier de la compagnie. Quelle est la probabilité qu'il ait déclaré au moins un accident au cours de l'année ?

b) Quelle est la probabilité qu'un assuré ayant déclaré au moins un accident au cours de l'année soit agé d'au plus 25 ans.

c) Quelle est la probabilité pour qu'un assuré agé de 25 ans ou plus ait au moins un accident au cours de l'année ?

d) Quelle est la probabilité qu'un assuré n'ayant déclaré aucun accident soit agé de 25 à 50 ans.

Exercice 2.16 [*]

Dans une usine, on fabrique des composants électroniques sur trois machines. Les machines M_1, M_2 et M_3 produisent respectivement 50%, 30% et 20% des composants. Le qualiticien de l'usine estime que :

⋆ 2 % des composants fabriqués par la machine M_1 sont défectueux.

⋆ 3 % des composants fabriqués par la machine M_2 sont défectueux.

⋆ 5 % des composants fabriqués par la machine M_3 sont défectueux.

a) Quelle est la probabilité qu'un composant pris au hasard à la sortie de l'usine soit défectueux ?

b) Quelle est la probabilité d'obtenir une pièce défectueuse provenant de M_1 ?

c) Les évènements "la pièce est défectueuse" et "la pièce provient de M_1" sont-ils indépendants ?

d) Un composant est défectueux. Quelle est la probabilité pour qu'il provienne de M_1 ?

Exercice 2.17 [**]

Un jeu ordinaire de 52 cartes est divisé de façon aléatoire en quatre paquets de treize cartes. Calculez la probabilité que chaque paquet contienne exactement un as.

Exercice 2.18 [*] **Bien répondre à un QCM**

Lors d'un QCM, un étudiant a deux options : soit il connaît la bonne réponse (ce qui arrive avec une probabilité $p = 1/2$), soit il coche une réponse au hasard parmi les $m = 5$ réponses possibles (une seule bonne réponse). Sachant que l'étudiant a coché la bonne réponse, quelle est la probabilité que l'étudiant connaissait la bonne réponse ?

Exercice 2.19 [**] **Un jeu de dés**

On dispose d'une pièce de monnaie et de deux dés A et B. Le dé A a quatre faces rouges et deux faces noires et le dé B a deux faces rouges et quatre noires. On lance la pièce de monnaie au début du jeu : si on obtient face, on lance le dé A indéfiniment, sinon on utilise le dé B.

a) Calculer la probabilité d'avoir rouge au n-ième jet de dé.

b) Sachant que rouge est sorti aux deux premiers coups, quelle est la probabilité d'avoir encore rouge au troisième ?

c) Si le rouge est sorti aux n premiers coups, quelle est la probabilité d'avoir encore rouge au $(n+1)$-ième lancer ?

d) Si le rouge est sorti aux n premiers coups, quelle est la probabilité d'avoir tiré face avec la pièce de monnaie au début du jeu ?

Exercice 2.20 [*] **Les chats de Schrödinger**

Considérons l'expérience de pensée suivante : on place 20 chats dans des boites individuelles. Par ailleurs,

- ⋆ 10 boites sont vides.
- ⋆ 5 boites contiennent 3g d'arsenic.
- ⋆ 5 boites contiennent 10g d'arsenic.

Au bout d'une heure, la probabilité qu'un chat soit encore en vie dans la boite est égale à 1 si la boite est vide ; 0,6 si la boite contient 3g d'arsenic ; 0,2 si la boite contient 10g d'arsenic. On choisit une des 20 boites au hasard.

a) Quelle est la probabilité que le chat qui s'y trouve soit encore en vie au bout d'une heure ?

b) Le chat est en vie. Quelle est la probabilité que la boite ait été vide quand on l'y a placé ?

Exercice 2.21 [**] **Une machine sans mémoire**

Soit une machine qui, à l'instant $n\Delta t$, est en panne ou en marche. On l'observe aux instants $0, \Delta t, 2\Delta t, 3\Delta t, \ldots$ On suppose que, si à un instant donné la machine marche, la probabilité que la machine soit en panne à l'instant d'après est a (avec $O < a < 1$). De même, si à un instant donné la machine est en panne, la probabilité que la machine marche à l'instant d'après est b. Calculer la probabilité que la machine soit en marche à $n\Delta t$ et montrez que la machine tend à "oublier" dans quel état elle est partie.

Exercice 2.22 [***] **Suites de probabilités**

Au moment où chacun possède un tiers du marché de la téléphonie mobile, trois opérateurs A, B et C décident de mettre sur le marché un nouveau type de forfait annuel. A la fin de l'année, l'évolution des parts de marché se fait de la façon suivante :

- ⋆ Les clients de la compagnie A se répartissent indifféremment entre A, B et C l'année suivante.
- ⋆ Les clients de la compagnie B restent toujours fidèles à cette compagnie.
- ⋆ Les clients de la compagnie C seront l'année suivante clients de A avec une probabilité $\frac{1}{12}$, de B avec une probabilité $\frac{7}{12}$ et de C avec une probabilité $\frac{1}{3}$.

On note pour $n \in \mathbb{N}$, a_n, b_n et c_n les probabilités pour qu'à l'issue de la n-ième année, un consommateur décide de s'abonner chez A, B ou C l'année suivante.

a) Déterminer une relation de récurrence entre a_n, b_n et c_n et a_{n+1}, b_{n+1} et c_{n+1}

b) En déduire, l'expression de a_n, b_n et c_n en fonction de n. On fera apparaître des suites récurrentes linéaires d'ordre 2 et on utilisera le point méthode qui suit.

c) Déterminez la limite de ces suites. Un opérateur tend-il vers une position de monopôle ?

Point méthode : Soit $(u_n)_{n\in\mathbb{N}}$ une suite vérifiant : $\forall n \in \mathbb{N}$, $u_{n+2} + au_{n+1} + bu_n = 0$, avec $a, b \in \mathbb{R}$ deux constantes. Pour trouver la formule de u_n en fonction de n, on suit la méthode de l'équation caractéristique :

i. On considère l'équation caractéristique $X^2 + aX + b = 0$ dont on calcule les deux racines q_1 et q_2 (éventuellement confondues).

ii. ⋆ si $q_1 \neq q_2$, on a $\forall n \in \mathbb{N}$, $u_n = \alpha.q_1^n + \beta.q_2^n$ avec α et β deux constantes.
   ⋆ si $q_1 = q_2 = q$, on a $\forall n \in \mathbb{N}$, $u_n = (\alpha.n + \beta).q^n$ avec α et β deux constantes.

iii. On trouve les deux constantes α et β en utilisant les données du problème. Le plus souvent, il s'agit des conditions initiales u_0 et u_1 dont un système linéaire de deux équations à deux inconnues va fixer α et β.

♦ **Exercice** 2.23 [***] **Matrices de probabilités**

Une particule se déplace à chaque seconde d'un sommet à l'autre du triangle ABC selon le protocole suivant :

⋆ Lorsqu'à un instant donné, elle se situe en A, elle se fixe à l'instant suivant en B avec la probabilité 0,75 et en C avec la probabilité 0,25.

⋆ Lorsqu'à un instant donné, elle se situe en B, elle se fixe à l'instant suivant en A avec la probabilité 0,75 et en C avec la probabilité 0,25.

⋆ Lorsqu'à un instant donné, elle se situe en C, elle ira systématiquement en B à l'instant suivant.

On désigne par a_n, b_n et c_n les probabilités qu'à l'instant n, la particule se situe en A, B ou C.

a) Déterminer des relations de récurrence entre a_{n+1}, b_{n+1}, c_{n+1} et a_n, b_n, c_n.

b) En déduire l'existence d'une matrice carrée d'ordre 3, notée M telle que

$$\begin{pmatrix} a_{n+1} \\ b_{n+1} \\ c_{n+1} \end{pmatrix} = M \begin{pmatrix} a_n \\ b_n \\ c_n \end{pmatrix}.$$

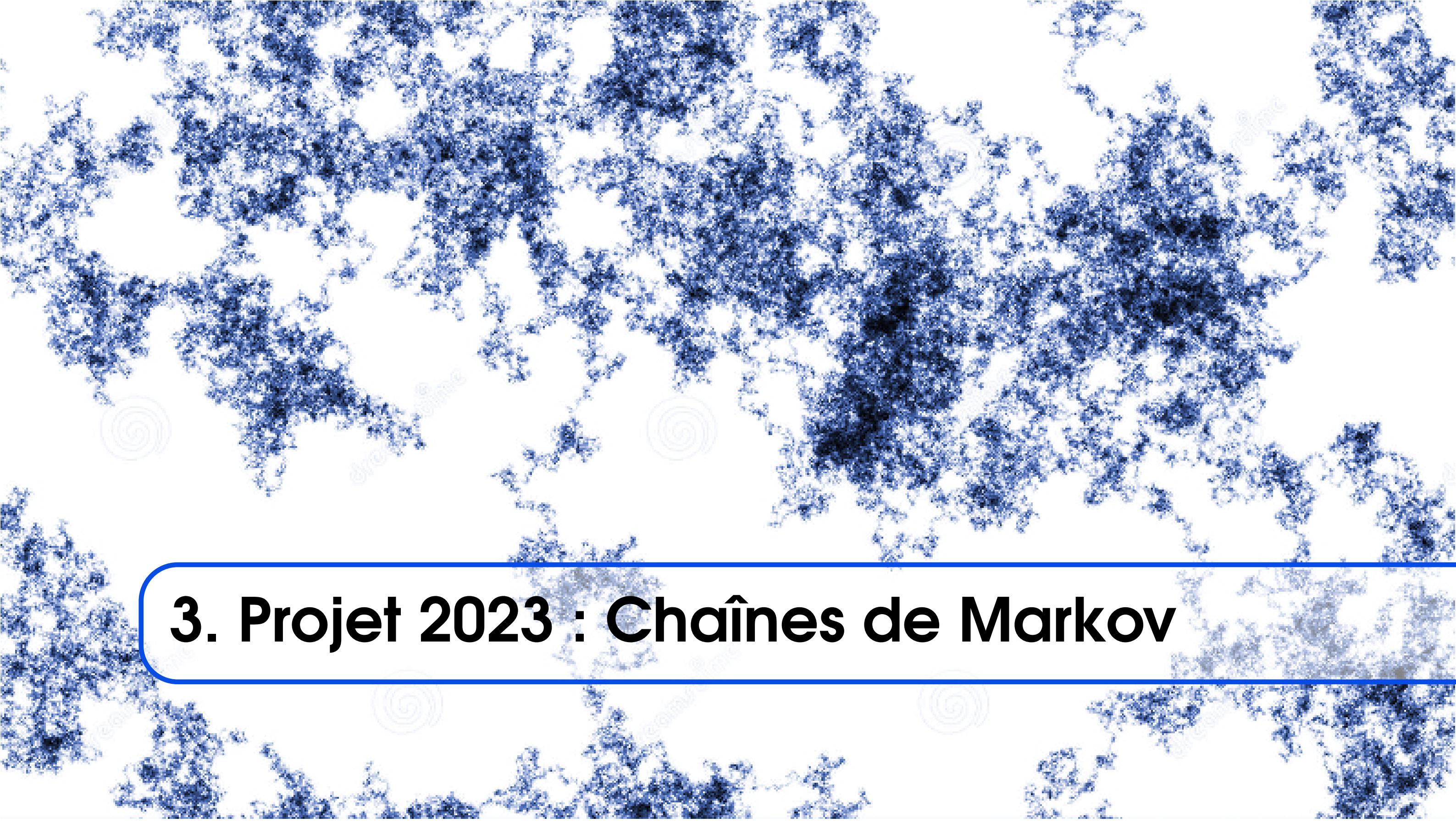
c) Montrer maintenant que l'on a

$$\begin{pmatrix} a_n \\ b_n \\ c_n \end{pmatrix} = M^n \begin{pmatrix} a_0 \\ b_0 \\ c_0 \end{pmatrix}.$$

d) Soit P la matrice $\begin{pmatrix} 12 & 3 & 1 \\ 16 & -1 & -1 \\ 7 & -2 & 0 \end{pmatrix}$. Montrer que P est inversible.

e) Calculer $P^{-1} \times M \times P$. Que pouvez-vous en dire sur M, en vous utilisant de votre cours d'Algèbre linéaire de première année ?

f) En déduire les limites quand n tend vers l'infini des probabilités a_n, b_n, c_n.



3. Projet 2023 : Chaînes de Markov

3.1 Introduction

En écologie, on appelle *écosystème* un ensemble formé par une communauté d'êtres vivants en interaction avec leur environnement. La théorie des probabilités est un outil qui nous permet de modéliser les interactions des constituants d'un écosystème dans le but de prédire leurs évolutions.

Ce projet a pour but d'introduire un type de modélisation d'écosystème. Afin de prendre en compte plusieurs constituants, on vous propose de travailler avec l'outil des graphes, puis des chaînes de Markov discrètes.

3.2 Un peu de théorie

3.2.1 Définitions

Définition 3.1 (Un graphe orienté)
Un graphe fini $G = (S, A)$ orienté est défini par :

- ⋆ un ensemble fini de n sommets que l'on note $S = \{s_1, s_2, \dots, s_n\}$.
- ⋆ un ensemble fini $A = \{a_1, a_2, \dots, a_m\}$ des paires ordonnées de sommets : chaque a_i correspond à un arc qui se traduira par une flèche sur le dessin d'un graphe.

Définition 3.2 (Une matrice des valeurs associées au graphe)
Soit $G = (S, A)$ un graphe orienté. On peut décider d'associer une valeur à chaque arc (on parle alors de graphe étiqueté) et, dans ce cas, on peut reporter l'ensemble des valeurs sur une matrice P où le coefficient $p_{i,j}$ de la ligne i et colonne j représente la valeur associée à l'arc allant du sommet j vers le sommet i.

3.2.2 Exemple

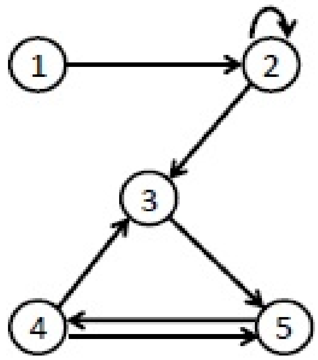


FIGURE 3.1 – $S = \{1,2,3,4,5\}$ est l'ensemble des sommets. Il y a 7 arcs comme, par exemple, celui allant du sommet 3 vers le sommet 5.

Sur ce graphe, remplissons par exemple la matrice avec les valeurs suivantes :

$$P = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0,7 & 0 & 0 & 0 \\ 0 & 0,3 & 0 & 0,7 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0,3 & 0 \end{pmatrix}$$

Sur la deuxième colonne, qui correspond au sommet 2, nous voyons deux valeurs $0,7$ et $0,3$. Ainsi, la probabilité de passer du sommet 2 (deuxième colonne) au sommet 2 (deuxième ligne) est 0,7. Et la probabilité de passer du sommet 2 (deuxième colonne) au sommet 3 (troisième ligne) est 0,3.

Revenons à la problématique d'un écosystème. Supposons qu'à l'instant initial, l'écosystème soit composé de $0,3 = 30\%$ de l'espèce 2 et $0,7 = 30\%$ de l'espèce 5 au temps initial. Alors on peut écrire ces informations dans le vecteur colonne suivant : $X_0 = \begin{pmatrix} 0 \\ 0,3 \\ 0 \\ 0 \\ 0,7 \end{pmatrix}$.

Supposons que la matrice P mime l'évolution du système de l'année n à l'année $n+1$, et ce pour tout $n \in \mathbb{N}$. On calcule la distribution des espèces dans notre écosystème l'année 1 suivante, c'est à dire X_1.

$$X_1 = P.X_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0,7 & 0 & 0 & 0 \\ 0 & 0,3 & 0 & 0,7 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0,3 & 0 \end{pmatrix} \times \begin{pmatrix} 0 \\ 0,3 \\ 0 \\ 0 \\ 0,7 \end{pmatrix} = \begin{pmatrix} 0 \\ 0,3 \times 0,7 \\ 0,3 \times 0,3 \\ 0,7 \times 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0,21 \\ 0,09 \\ 0,7 \\ 0 \end{pmatrix}$$

Lors de l'année 1, l'écosystème sera donc composé de 21% de l'espèce 2, 9% de l'espèce 3 et 70% de l'espèce 4.

On remarque que l'évolution de notre système n'est dictée que par la matrice P qui est constante. Ainsi, la prédiction de la distribution future d'un système sera entièrement contenue dans la distribution présente de ce même système. On dira alors que ce processus possédera la *propriété de Markov*.

3.3 La forêt des Landes : modèle à 3 états

Dans un premier temps, on vous demande d'étudier l'évolution de la forêt des Landes qu'on suppose composée de Résineux (Rs), de Chênes (Ch) et de Châtaigniers (Cht).

Le temps est discrétisé en intervalles de temps de durée Δt (exprimée en années). Pour tout $n \in \mathbb{N}$, le vecteur

$$X_n = \begin{pmatrix} Rs \\ Ch \\ Cht \end{pmatrix}_n$$

désignera la distribution de cet écosystème, c'est à dire les proportions de présence de chacun des arbres, à l'instant $t_n = n\Delta t$.

A l'instant initial, on suppose que

$$X_0 = \begin{pmatrix} Rs \\ Ch \\ Cht \end{pmatrix}_0 = \begin{pmatrix} 0,36 \\ 0,4 \\ 0,24 \end{pmatrix}.$$

1. Pour tout $n \in \mathbb{N}$, on suppose que les proportions d'arbres évoluent de la façon suivante : dans le cas où lors de l'année n, il n'y a pas de Ch, ni de Cht, mais qu'il y a une proportion de 100%=1 de Rs alors, l'année suivante, il y aura :
   - $\star$ une proportion $0,5$ de Rs,
   - $\star$ une proportion $0,2$ de Ch,
   - $\star$ et une proportion $0,3$ de Cht.

   Comment se traduisent ces informations sur la matrice P telle que nous l'avons définie, possédant la propriété de Markov ?
2. Supposons maintenant que la matrice P s'écrive de la manière suivante : $\begin{pmatrix} 0,5 & 0,1 & 0,1 \\ 0,2 & 0,6 & 0,2 \\ 0,3 & 0,3 & 0,7 \end{pmatrix}$.
   Dessiner un graphe qui modélise l'écosystème. On précisera notamment les valeurs sur chacun des arcs.
3. Calculez la proportion de chaque arbre pour l'année $n = 1$.
4. A l'aide d'un ordinateur, calculez la proportion de chaque arbre pour $n = 20$.
   On pourra se servir de Matlab dont une aide sur les matrices se trouve sur le lien suivant : Aide sur les matrices
5. Tracer sur un graphique l'évolution des trois espèces de l'écosystème en fonction de n.
   On pourra se servir de Matlab, l'aide pour le tracé se trouve sur les liens suivants :
   Aide sur les plots et Aide sur les graphiques en 3D
   En regardant les valeurs numériques obtenues, explicitez le comportement de l'écosystème aux temps longs.
6. En vous servant de votre cours d'Algèbre linéaire de L1, diagonaliser P. Donner alors les matrices A, A^{-1} et D (avec D diagonale) telles que $P = A.D.A^{-1}$. Que représentent les matrices A, A^{-1} et D en termes de vecteurs propres et valeurs propres ?
7. Écrire le vecteur X_n en fonction de D^n, A, A^{-1} et la distribution initiale X_0.
   Faites tendre n vers l'infini dans cette expression. Vers quelle proportion tend alors chaque terme de X_n ? Retrouver les valeurs obtenues numériquement à la question 5.

3.4 Ecosystèmes méditerranéens à 5 états

Exemple tiré du livre Modélisation et simulation d'écosystèmes, P. Coquillard et D. Hill, Masson 1997

A l'origine la forêt méditerranéenne, sur roche calcaire à faible altitude, était très certainement dominée par des chênes (chênes pubescents). Mais l'action de l'homme a éradiqué ces forêts primitives pour leur substituer parcours pastoraux, vergers, etc. Puis l'abandon de toute activité agricole a favorisé l'implantation d'une autre espèce, le pin d'Alep, après passage par un état de guarrigue. Or ces forêts de substitution, hautement inflammables, subissent de manière récurrente le passage du feu (incendies volontaires ou non) et sont donc condamnées à une perpétuelle reconstitution.

Pour étudier l'évolution à long terme de l'écosystème, on propose de modéliser cette évolution par une chaîne de Markov à 5 états : chênaie (C), Vignes,Vergers (V), pelouse (Pe), garrigue (Ga) et pinède (Pi). Le temps est de nouveau discrétisé en intervalles de temps de durée Δt (exprimée en années). Pour tout $n \in \mathbb{N}$, le vecteur

$$X_n = \begin{pmatrix} C \\ V \\ Pe \\ Ga \\ Pi \end{pmatrix}_n$$

désignera la distribution de cet écosystème à l'instant $t_n = n\Delta t$.

La matrice de transition de cette chaîne de Markov à 5 états est la matrice P de probabilité de transition relative à ces 5 états :

$$P = \begin{pmatrix} 0,8 & 0 & 0 & 0 & 0,1 \\ 0,2 & 0,7 & 0 & 0 & 0 \\ 0 & 0,3 & 0,4 & 0 & 0,25 \\ 0 & 0 & 0,6 & 0,2 & 0 \\ 0 & 0 & 0 & 0,8 & 0,65 \end{pmatrix}$$

1. Dessiner le graphe relatif à ce modèle.
2. Quelle est, selon ce modèle, la probabilité que l'on passe de l'état Pi à Pe, d'une année sur l'autre ? Et de l'état Pe à Pi ?
3. Sachant que l'on est dans l'état Pi, quelle est la probabilité d'une trajectoire du type $Pi - Pe - Ga$? Et d'une trajectoire du type $Pi - C - V$?
4. Donnez un exemple de trajectoire de probabilité nulle.

5. En vous servant de votre cours d'Algèbre linéaire de L1 expliquez, lorsque n tend vers l'infini, vers quelle proportion tend la dynamique lorsque

$$X_0 = \begin{pmatrix} C \\ V \\ Pe \\ Ga \\ Pi \end{pmatrix}_0 = 0,2 \times \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

6. La transition de Pi vers Pe mime la présence des incendies. Étudier l'influence aux temps longs des incendies, en faisant varier sa valeur dans la matrice.

3.5 Modèle plus riche d'écosystème

Imaginons maintenant un écosystème avec 20 états qui évoluent. Soit la matrice P de probabilité de transitions :

$$\begin{pmatrix}
0,1 & 0 & 0,8 & 0 & 0,2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0,2 & 0,4 & 0,05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0,1 & 0,2 & 0,05 & 0,9 & 0,1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0,5 & 0,2 & 0,05 & 0,05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0,1 & 0,2 & 0,05 & 0,05 & 0,7 & 0 & 0 & 0 & 0,2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,6 & 0,4 & 0,1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,1 & 0,2 & 0,7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,1 & 0,2 & 0,1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,1 & 0 & 0,3 & 0,3 & 0,3 & 0,4 & 0,1 & 0,3 & 0 & 0,15 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,3 & 0,05 & 0,3 & 0,1 & 0,1 & 0,1 & 0,2 & 0,05 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,1 & 0,1 & 0 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 & 0,15 & 0,05 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 & 0,05 & 0,05 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,1 & 0,1 & 0 & 0,1 & 0,2 & 0,1 & 0,2 & 0,3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,2 & 0,1 & 0,05 & 0,2 & 0,1 & 0,1 & 0,2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,2 & 0 & 0 & 0 & 0,05 & 0,1 & 0,05 & 0,1 & 0,1 & 0,2 & 0,1
\end{pmatrix}$$

1. A partir de chaque état, le ou lesquels sont à un moment ou un autre visités ?
   Donner alors les groupes d'états qui communiquent tous ensemble.
   Expliquer comment cela se voit sur la matrice.
2. Partons de la distribution :

$$X_0 = 1/16 \times \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}^T$$

   (où T désigne la matrice transposée).
   Donner le composition X_n du système pour $n = 1$, $n = 2$ et $n = 50$.
   Faire une simulation numérique de l'ensemble des états et afficher X_n en fonction de n.
   Quel temps moyen (en nombre de pas) estimez-vous pour que le système soit concentré sur les états $\{1,2,3,4,5\}$, $\{13,14,15,16,17,18,19,20\}$.
3. Partons de la distribution :

$$Y_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^T$$

   Faire une simulation numérique de l'ensemble des états et afficher Y_n en fonction de n.
   Trouver le temps moyen (en nombre de pas) que met le système pour peupler l'ensemble des cinq premiers états $\{1,2,3,4,5\}$ et le temps moyen pour peupler l'ensemble des états $\{14,15,16,17,18,19,20\}$. Lequel des deux sera le plus peuplé quand la dynamique aura convergé ?
4. Partons de la distribution :

$$Z_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^T$$

   Faire une simulation numérique de l'ensemble des états et afficher Z_n en fonction de n.
   Expliquez ce qu'il se passe.
5. Partons de la distribution :

$$A_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^T$$

   Faire une simulation numérique de l'ensemble des états et afficher A_n en fonction de n.
   En combien de temps (en nombre de pas) le système reviendra à sa distribution initiale ? Expliquez ce qu'il se passe.

3.6 Pour aller plus loin

Soit une chaîne de Markov à $k \in \mathbb{N}^*$ états tels que :

$$P = 1/2 \times \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & 0 & 1 \\ 1 & 0 & 1 & 0 & & & 0 \\ 0 & 1 & 0 & 1 & 0 & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & & & \ddots & \ddots & 0 & 1 \\ 1 & 0 & \dots & \dots & 0 & 1 & 0 \end{pmatrix}_k$$

.

1. Dessiner le graphe de cette chaîne.
2. On prend pour valeur initiale :

$$X_0 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_k .$$

Faire une simulation numérique de $n = 200$ pas pour une chaîne de taille $k = 100$.
Afficher X_n en fonction de n.
Trouver une bonne façon de représenter la distribution X_{200}. A quelle distribution vous fait-elle penser ? Approximez celle-ci par la distribution que vous pensez être correcte et donnez ses paramètres.

3.7 Précisions sur le rendu

3.7.1 Le rendu

Le projet se fait par équipes de 5. Vous avez jusqu'au vendredi 27 octobre pour inscrire vos équipes dans la partie dédiée à cet effet sur Moodle. Les étudiants non affectés à une équipe après cette date seront distribués dans des groupes de façon aléatoire par les enseignants.

Un rapport est demandé pour chacune des équipes qui fera au maximum 16 pages, i.e. environ 4 pages par partie :

1. La forêt des Landes
2. Ecosystèmes méditerranéens
3. Modèle plus riche d'écosystème
4. Pour aller plus loin

La notation tiendra compte de la qualité des courbes présentées, la concision des réponses ainsi que des interprétations.

Le dépôt de votre rapport se fera sur Moodle jusqu'au samedi 2 décembre à 23h59. Aucun délais supplémentaire ne pourra être accepté.

3.7.2 L'oral

Pour l'oral, il est attendu une présentation Powerpoint. La pédagogie et la clarté sont le but de cet oral de 15 min. Attention, si vous dépassez les 15 min de présentation, votre enseignant vous arrêtera. Les Powerpoints ne doivent évidemment pas comporter de texte écrit à la main et photographié. Toutes les formules et matrices doivent être tapées. A l'oral, il est demandé de restituer avec précision les points les plus importants de chaque partie.

A l'issu de l'oral, 15 min de questions vous seront posées : chaque étudiant sera interrogé individuellement sur le projet ou sur un point du cours.

Bon courage pour la rédaction de ce projet.

II Les variables aléatoires

4. Variables aléatoires discrètes

Souvent, dans une expérience aléatoire, ce n'est pas le résultat en lui-même qui nous intéresse mais une certaine valeur liée au résultat. Par exemple, on peut :

- ⋆ tirer au hasard une personne dans la population française et mesurer sa hauteur ou son nombre moyen d'heures de sommeil par jour,
- ⋆ tirer au hasard un étudiant de l'EFREI en troisième année et mesurer la note qu'il ou elle a eu au DE de probabilités,
- ⋆ jeter deux dés et compter la somme des dés.

Ces valeurs que l'on forme grâce au résultat d'une expérience sont appelées des "variables aléatoires" et leur étude est au coeur de la théorie des probabilités.

4.1 Définitions de base

Définition 4.1 (Variable aléatoire)
Soit (Ω, P) un espace probabilisé. On appelle *variable aléatoire* toute fonction $X : \Omega \longrightarrow \mathbb{R}$.
L'ensemble des valeurs possibles de X est appelé le *support de* X et est noté $X(\Omega)$.
L'ensemble des variables aléatoires sur Ω est noté $\Omega^{\mathbb{R}}$.

Exemples : En reprenant les exemples de l'introduction :

- ⋆ Si $\Omega = \{\text{population française}\}$ et P est la probabilité uniforme sur Ω, alors la hauteur et le nombre moyen d'heures de sommeil par jours sont bien des fonctions $H, S : \Omega \longrightarrow \mathbb{R}$ dont les supports sont respectivement $[25, 222]$ (en cm) et $[3, 18]$ (en heures).

- ⋆ Si $\Omega = \{$étudiants de l'EFREI en L3$\}$ et P est la probabilité uniforme sur Ω, alors la note de DE de probabilités est bien une fonction $N : \Omega \longrightarrow \mathbb{R}$ dont le support est l'intervalle $[0,20]$.
- ⋆ Dans le jet d'une paire de dés, $\Omega = \{1,2,3,4,5,6\}^2$ et la somme du résultat des deux dés est une fonction $S : \Omega \longrightarrow \mathbb{R}$ dont le support est $\{2,3,4,\dots,12\}$.

Définition 4.2 (Opérations sur les variables aléatoires)
Soient (Ω,P) un espace probabilisé, X,Y deux variables aléatoires et $\lambda \in \mathbb{R}$. Alors :

1. $\Omega^{\mathbb{R}}$ est un espace vectoriel réel. En d'autre mots, $X+Y$ et $\lambda.X$ sont des variables aléatoires.
2. Le produit $X.Y$ est aussi une variable aléatoire.
3. Soit f une fonction réelle définie sur $X(\Omega)$, alors la composée $f(X)$ définie par $\Omega \xrightarrow{X} \mathbb{R} \xrightarrow{f} \mathbb{R}$ est aussi une variable aléatoire.

Exemples :

- ⋆ Pour les étudiants en L3 à l'EFREI, notez X la note au DE de probabilité et Y la note au CE de probabilité. Alors $Z = 0.6X + 0.4Y$ est une variable aléatoire (c'est la note finale du module de l'année dernière).
- ⋆ Pour la population française, soit H la hauteur d'une personne, M le poids et f la fonction définie par $f(x) = \dfrac{1}{x^2}$. Alors, $IMC = P \times f(H) = \dfrac{P}{H^2}$ est une variable aléatoire (c'est l'indice de masse corporel).

Définition 4.3 (Variable aléatoire discrète)
Soient (Ω,P) un espace probabilisé et X une variable aléatoire. On dit que X est une variable aléatoire *discrète* si $X(\Omega)$ est un ensemble fini ou dénombrable.
L'ensemble des variables aléatoires discrètes sur Ω est noté $\ell(\Omega)$.

Exemples :

- ⋆ La variable aléatoire $X =$ "note au DE de probabilités" peut être discrète ou pas en fonction de notre modèle. Si l'on considère que la note peut être n'importe quel réel dans l'intervalle $[0,20]$, alors X n'est pas discrète. Si l'on est plus réaliste et l'on considère que la note est définie au demi-point près, on a $X(\Omega) = \{0, 0.5, 1, 1.5, 2, \dots, 19, 19.5, 20\}$ et donc X est discrète.
- ⋆ Même commentaire pour la variable aléatoire $H =$"hauteur d'une personne" qui peut être vue comme prenant n'importe quelle valeur dans l'intervalle $[0,222]$ (hauteur mesurée en cm), ou bien on peut penser que la hauteur est définie au millimètre près (par exemple).
- ⋆ La variable aléatoire $S =$ "somme des deux dés" est discrète puisque les valeurs possibles sont les entiers compris entre 2 et 12.

Remarque : $\ell(\Omega)$ est un sous-espace vectoriel de $\Omega^{\mathbb{R}}$.

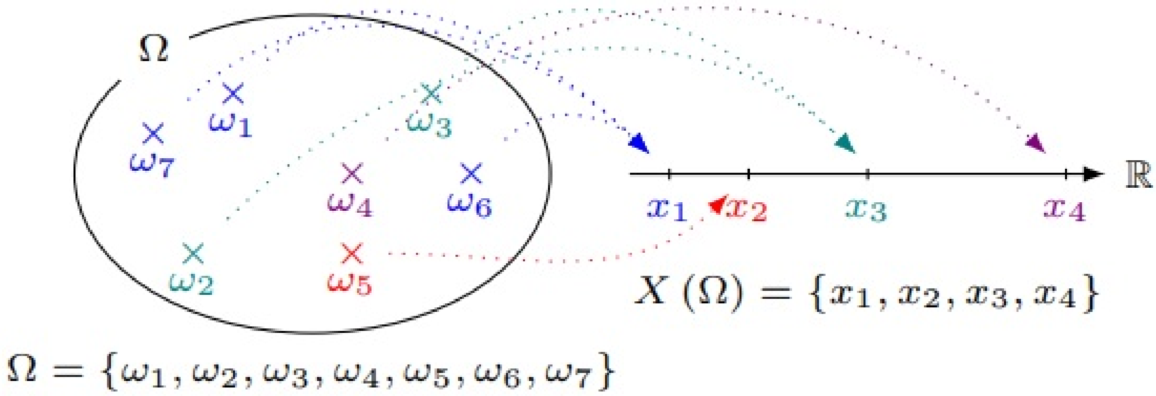


FIGURE 4.1 – Une variable aléatoire discrète dont l'univers est
$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7\}$ et le support est $X(\Omega) = \{x_1, x_2, x_3, x_4\}$.
Figure tirée du site https://chauvetmath.fr.

Théorème 4.1 : Système complet d'évènements associé à X

Soit (Ω, P) un espace probabilisé et X une variable aléatoire discrète sur Ω. On note

$$(X = x) = \{\omega \in \Omega \big| X(\omega) = x\}.$$

Alors la famille $\big((X = x)_{x \in X(\Omega)}\big)$ est un système complet d'évènements de l'univers Ω.

Exemple : Dans la figure précédente, on a

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7\} = \{\underbrace{\omega_1, \omega_6, \omega_7}_{(X=x_1)}, \underbrace{\omega_5}_{X=x_2}, \underbrace{\omega_2, \omega_3}_{X=x_3}, \underbrace{\omega_4}_{X=x_4}\} = \big((X = x)_{x \in X(\Omega)}\big)$$

La variable aléatoire discrète redéfinit ainsi complètement l'espace d'étude.

Définition 4.4 (Loi de probabilité d'une variable aléatoire discrète)
Soit (Ω, P) un espace probabilisé et X une variable aléatoire discrète sur Ω. Supposons, de plus, que l'on peut classer les valeurs prises par X par ordre croissant depuis une plus petite valeur x_1 (autrement dit, $X(\Omega) = \{x_1, x_2, \dots, x_n, \dots\}$, avec $i < j \implies x_i < x_j$).
Alors *la loi de probabilité de X* est la donnée des nombres $p_i = P(X = x_i)$.

Remarque : Par définition, la suite des nombres p_i doit satisfaire les deux propriétés suivantes :

$$\sum_i p_i = 1 \text{ et } P(x_i \leq X \leq x_j) = \sum_{k=i}^{j} p_i.$$

Exemple : On tire au hasard un étudiant de l'EFREI et on regarde son âge (en années). Ici l'univers Ω est l'ensemble des étudiants inscrits à l'EFREI en ce moment. La variable aléatoire X est l'âge et le support de X est $X(\Omega) = \{16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27\}$. Une bonne façon de représenter la loi de probabilité de X est sous forme d'un tableau :

x_i	16	17	18	19	20	21	22	23	24	25	26	27
$P(X = x_i)$	$\frac{1}{100}$	$\frac{4}{100}$	$\frac{10}{100}$	$\frac{15}{100}$	$\frac{15}{100}$	$\frac{20}{100}$	$\frac{15}{100}$	$\frac{10}{100}$	$\frac{5}{100}$	$\frac{3}{100}$	$\frac{1}{100}$	$\frac{1}{100}$

(et on vérifie bien que $\frac{1}{100} + \frac{4}{100} + \frac{10}{100} + \frac{15}{100} + \frac{15}{100} + \frac{20}{100} + \frac{15}{100} + \frac{10}{100} + \frac{5}{100} + \frac{3}{100} + \frac{1}{100} + \frac{1}{100} = 1$).

Définition 4.5 (Fonction de répartition)
On appelle *fonction de répartition* de la variable aléatoire X, que l'on note $F_X(x)$ (ou $F(x)$ si la variable aléatoire est évidente), la fonction définie pour tout réel $x \in \mathbb{R}$ par

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i).$$

Exemple : Voici le graphe représentant la fonction de répartition de la variable X de l'exemple précédent (âge des étudiants de l'EFREI) :

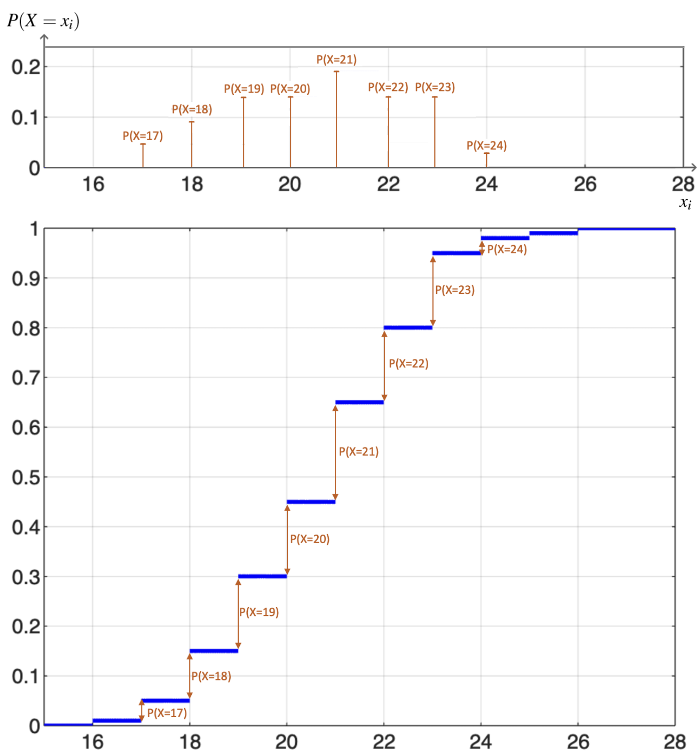


FIGURE 4.2 – Le distribution initiale est en haut. En bas, la fonction de répartition (en bleu) est une fonction constante par morceaux. On peut lire dans chaque "saut" correspondant à la valeur de la probabilité $P(X = x_i)$.

On est souvent confronté au problème de trouver la loi de probabilité d'une variable aléatoire Y qui s'exprime en fonction d'une variable aléatoire X dont la loi de probabilité est connue :

Fiche méthode #5 (Loi de probabilité de $Y = g(X)$ dans le cas discret)
Soit X une variable aléatoire discrète de support $X(\Omega) = \{x_1, x_2, \dots, x_n, \dots\}$ et dont la loi de probabilité est connue. Soit g une fonction réelle définie sur $X(\Omega)$ et $Y = g(X)$ une nouvelle variable aléatoire. Pour trouver la loi de probabilité de Y :

a) Déterminer d'abord les valeurs possibles de Y (i.e., le support $Y(\Omega)$).

b) Pour $y \in Y(\Omega)$, décrire l'évènement $(Y = y)$ en fonction du système complet d'évènements $(X = x_i)_{x_i \in X(\Omega)}$.

c) Utiliser les propriétés des fonctions de probabilités pour calculer $P(Y = y)$ grâce aux valeurs $p_i = P(X = x_i)$.

♦Exemple : On considère une variable aléatoire discrète X dont la loi de probabilité est donnée ci-dessous :

x_i	-2	-1	0	1	2	3	4
$P(X = x_i)$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{3}{16}$

Soit $Y = X^2$. Le support de Y est $Y(\Omega) = \{0, 1, 4, 9, 16\}$. De plus, on a

$$(Y = 0) = (X = 0) \qquad (Y = 1) = (X = -1) \cup (X = 1)$$
$$(Y = 4) = (X = -2) \cup (X = 2) \qquad (Y = 9) = (X = 3) \qquad (Y = 16) = (X = 4)$$

et donc la loi de probabilité de Y est :

y_i	0	1	4	9	16
$P(Y = y_i)$	$\frac{1}{8}$	$\frac{5}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{3}{16}$

4.2 Couples de variables aléatoires discrètes

Définition 4.6 (Loi de probabilité conjointe)
Soient $X, Y \in \ell(\Omega)$ deux variables aléatoires discrètes sur Ω. La *loi de probabilité conjointe de X et Y* est la donnée de des nombres $P(X = x, Y = y)$ pour toutes les couples de valeurs $(x, y) \in X(\Omega) \times Y(\Omega)$.

Exemple : Lorsque les supports des variables aléatoires sont petits, on peut tout représenter sous forme d'un tableau à deux dimensions.

On étudie l'influence de l'alcool sur les accidents de voiture mortels. Soit X l'heure de l'accident (par tranches horaires de trois heures : de minuit à 3h, de 3h à 6h, etc) et Y la variable aléatoire binaire qui comptabilise si oui ou non le conducteur avait un taux d'alcoolémie supérieur à 0,5g/l. On a les données suivantes :

	$X=[0,3[$	$X=[3,6[$	$X=[6,9[$	$X=[9,12[$	$X=[12,15[$	$X=[15,18[$	$X=[18,21[$	$X=[21,0[$
$Y=$ oui	0.085	0.041	0.014	0.009	0.014	0.031	0.051	0.068
$Y=$ non	0.047	0.042	0.082	0.088	0.115	0.126	0.108	0.068

Cela veut dire, par exemple, que si l'on prend un accident mortel au hasard, la probabilité que cet accident soit survenu entre 9h et 12h et que le conducteur ait un taux d'alcoolémie supérieur à 0,5g/l est de 0,009.

Proposition 4.2

Soient $X,Y \in \ell(\Omega)$ deux variables aléatoires discrètes sur Ω. Alors, si l'on connaît la loi de probabilité conjointe de X et Y, on connaît aussi la loi de probabilité de X et celle de Y (la réciproque est fausse !).

Démonstration. Etant donné la loi de probabilité conjointe de X et Y, la loi de probabilité de X se trouve tout simplement en sommant sur toutes les valeurs possibles de Y (et viceversa) :

$$P(X=x) = \sum_{y\in Y(\Omega)} P(X=x, Y=y).$$

(C'est une application de la formule des probabilités totales puisque la famille $\big((Y=y)_{y\in Y(\Omega)}\big)$ forme un système complet d'évènements.) $\square$

Exemple : Lorsque la loi de probabilité conjointe est présentée sous forme d'un tableau, on trouve la loi de probabilité de X et celle de Y en sommant les coefficients des lignes pour une variable et des colonnes pour l'autre :

	$X=-2$	$X=-1$	$X=0$	$X=1$	$X=2$	$\boldsymbol{P(Y=y)}$
$Y=15$	$\frac{1}{10}$	$\frac{1}{15}$	$\frac{1}{15}$	0	$\frac{1}{30}$	**4/15**
$Y=30$	$\frac{1}{10}$	0	0	$\frac{2}{15}$	$\frac{1}{10}$	**1/3**
$Y=40$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{15}$	**2/5**
$\boldsymbol{P(X=x)}$	**4/15**	**2/15**	**1/6**	**7/30**	**1/5**	**1**

Voici une loi de probabilité conjointe de X et Y différente mais qui donne lieu aux mêmes lois de probabilités pour X et pour Y :

	$X=-2$	$X=-1$	$X=0$	$X=1$	$X=2$	$\boldsymbol{P(Y=y)}$
$Y=15$	$\frac{4}{15}$	0	0	0	0	**4/15**
$Y=30$	0	$\frac{2}{15}$	$\frac{1}{6}$	$\frac{1}{30}$	0	**1/3**
$Y=40$	0	0	0	$\frac{1}{5}$	$\frac{1}{5}$	**2/5**
$\boldsymbol{P(X=x)}$	**4/15**	**2/15**	**1/6**	**7/30**	**1/5**	**1**

Définition 4.7 (Couple de variables aléatoires discrètes indépendantes)
Soient $X,Y \in \ell(\Omega)$ deux variables aléatoires discrètes sur Ω. On dit que X et Y sont *indépendantes* si

$$\forall (x,y) \in X(\Omega)\times Y(\Omega), P(X=x,Y=y) = P(X=x)P(Y=y).$$

De façon équivalente, cette condition s'écrit :

$$\forall (x,y) \in X(\Omega)\times Y(\Omega), P(X=x \big| Y=y) = P(X=x).$$

Autrement dit, les variables X et Y sont indépendantes si les évènements $(X=x)$ et $(Y=y)$ sont indépendants pour tout choix de x et y.

Exemples :

- ⋆ Si l'on reprend le dernier tableau de l'exemple précédent, les variables X et Y ne sont pas indépendantes puisque l'on a $P(X=-2\big|Y=15)=1$ or $P(X=-2)=4/15$.

- ⋆ La loi de probabilité conjointe ci-dessous représente deux variables aléatoires indépendantes :

	$X=1$	$X=2$	$X=3$	$\boldsymbol{P(Y=y)}$
$Y=0$	$\frac{1}{18}$	$\frac{1}{6}$	$\frac{1}{9}$	**1/3**
$Y=1$	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{2}{9}$	**2/3**
$\boldsymbol{P(X=x)}$	**1/6**	**1/2**	**1/3**	**1**

  Une façon directe de remarquer l'indépendance est de voir que les valeurs sur les lignes $Y=0$ et $Y=1$ sont proportionnelles. On pourrait dire la même chose des colonnes $X=1$, $X=2$ et $X=3$. Ainsi le résultat obtenu sur Y n'affectera pas la distribution de probabilité des X.

- ⋆ On tire au hasard une personne de la population française. On note X le nombre de fois que cette personne a donné "positif" à un test de la COVID-19 et Y le nombre de têtes d'ails mangés par la personne, en moyenne, par semaine. Ces deux variables aléatoires sont indépendantes.

4.3 Espérance, Variance, Ecart-type

Lorsque l'on travaille avec une variable aléatoire discrète, on a donc affaire à une (longue) liste de couples de nombres (les x_i et les $p(X = x_i)$). Pour décrire ces listes de nombres, il est utile de construire d'autres quantités qui vont résumer certaines caractéristiques importantes de la variable aléatoire. Les deux quantités les plus importantes sont l'*espérance* (moyenne théorique) et la *variance* ou l'*écart-type* (qui mesure la dispersion).

L'espérance mathématique correspond à la moyenne théorique. Par exemple, si l'on dispose d'un dé à 6 faces, dont cinq faces sont notées 1 et une face est numérotée 2, le 2 sortira en moyenne une fois sur six et le 1 apparaîtra cinq fois sur six. Ainsi, sur six tirages, on peut « espérer » avoir en moyenne cinq fois 1 et une fois 2. La moyenne de ces six tirages est alors

$$\frac{1}{6}(5\times 1+1\times 2) = 1P(X=1)+2P(X=2)$$

où X désigne le résultat d'un tirage. Cela motive la définition suivante de la moyenne théorique, appelée plus généralement espérance (mathématique).

Définition 4.8 (Espérance d'une variable aléatoire discrète)
Soit X une variable aléatoire discrète. On appelle *espérance de X*, si elle existe, la quantité

$$E(X) = \sum_{x\in X(\Omega)} xP(X=x).$$

L'ensemble des variables aléatoires discrètes sur Ω pour lesquelles l'espérance existe est noté $\ell_1(\Omega)$.

♦Exemples :

⋆ Reprenons l'exemple de l'âge des étudiants de l'EFREI (page 68) donné par la loi de probabilité :

x_i	16	17	18	19	20	21	22	23	24	25	26	27
$P(X=x_i)$	$\frac{1}{100}$	$\frac{4}{100}$	$\frac{10}{100}$	$\frac{15}{100}$	$\frac{15}{100}$	$\frac{20}{100}$	$\frac{15}{100}$	$\frac{10}{100}$	$\frac{5}{100}$	$\frac{3}{100}$	$\frac{1}{100}$	$\frac{1}{100}$

Avec ces données, l'espérance de X, qui représente l'âge moyen d'un étudiant de l'EFREI est :

$$\begin{aligned}E(X) = 16\times\frac{1}{100}+17\times\frac{4}{100}+18\times\frac{10}{100}+19\times\frac{15}{100}+20\times\frac{15}{100}+21\times\frac{20}{100}+22\times\frac{15}{100}\\ +23\times\frac{10}{100}+24\times\frac{5}{100}+25\times\frac{3}{100}+26\times\frac{1}{100}+27\times\frac{1}{100}\\ =20.77\end{aligned}$$

⋆ Dans le cas où $X(\Omega)$ est un ensemble infini dénombrable, l'espérance peut ne pas exister car rien n'empêche que la série $\sum_{x \in X(\Omega)} xP(X = x)$ soit divergente.

Prenez l'exemple d'un jeu où l'on jette une pièce jusqu'à ce que l'on tombe sur pile et le joueur gagne 2^n euros où n est le nombre de lancers effectués. Soit X la variable aléatoire représentant le gain du joueur. Alors, le gain moyen d'un joueur est donné par

$$\sum_{n=1}^{+\infty} 2^n P(X = 2^n) = \sum_{n=1}^{+\infty} 2^n \frac{1}{2^n} = \infty.$$

Proposition 4.3

L'espérance est une application linéaire sur l'espace vectoriel $\ell_1(\Omega)$.

Autrement dit, soient X et Y deux variables aléatoires discrètes sur Ω pour lesquelles l'espérance existe, et soient $\alpha, \beta \in \mathbb{R}$. Alors, on a

$$E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y).$$

Théorème 4.4 : Théorème du transfert (version discrète)

Soit $X \in \ell_1(\Omega)$ une variable aléatoire discrète, soit g une fonction réelle définie sur $X(\Omega)$. Alors, si l'espérance de $Y = g(X)$ existe, on a

$$E(g(X)) = \sum_{x \in X(\Omega)} g(x)P(X = x).$$

Démonstration. La preuve est tout simplement l'idée que l'on doit regrouper ensemble les différentes valeurs de X ayant la même valeur $y_j = g(x)$:

$$\begin{aligned}
\sum_i g(x_i)P(X = x_i) &= \sum_j \sum_{x_i | g(x_i) = y_j} g(x_i)P(X = x_i) \\
&= \sum_j y_j \sum_{x_i | g(x_i) = y_j} P(X = x_i) \\
&= \sum_j y_j P(g(X) = y_j) \\
&= E(g(X)).
\end{aligned}$$ □

♦Exemple : Reprenons l'exemple page 69. Nous avons maintenant deux façons de calculer $E(X^2)$.

- ⋆ D'abord, en utilisant la loi de probabilité de la variable aléatoire $Y = X^2$. Cela donne
$$E(X^2) = 0 \times \frac{1}{8} + 1 \times \frac{5}{16} + 4 \times \frac{1}{4} + 9 \times \frac{1}{8} + 16 \times \frac{3}{16} = \frac{87}{16}.$$
- ⋆ Ensuite, en utilisant le théorème du transfert :
$$E(X^2) = (-2)^2 \times \frac{3}{16} + (-1)^2 \times \frac{1}{4} + 0^2 \times \frac{1}{8} + 1^2 \times \frac{1}{16} + 2^2 \times \frac{1}{16} + 3^2 \times \frac{1}{8} + 4^2 \times \frac{3}{16} = \frac{87}{16}.$$

Remarque : Dans cet exemple, comme en général, on a $E(X^2) \neq \left(E(X)\right)^2$.

Définition 4.9 (Variance et écart-type d'une variable aléatoire discrète)
Soit X une variable aléatoire discrète. On appelle *variance de* X, si elle existe, la quantité
$$V(X) = E\left(\left(X - E(X)\right)^2\right) = E(X^2) - E(X)^2.$$
De plus, on appelle *écart-type de* X la quantité $\sigma(X) = \sqrt{V(X)}$.
L'ensemble des variables aléatoires discrètes sur Ω pour lesquelles la variance existe est noté $\ell_2(\Omega)$.

Si l'espérance mesure la moyenne théorique de X, l'écart-type (qui a les mêmes unités que X) permet de mesurer la dispersion de X, c'est-à-dire la distance moyenne entre le résultat d'une expérience aléatoire et $E(X)$: plus $\sigma(X)$ est "petit", plus les valeurs de X sont "concentrées" autour de $E(X)$.

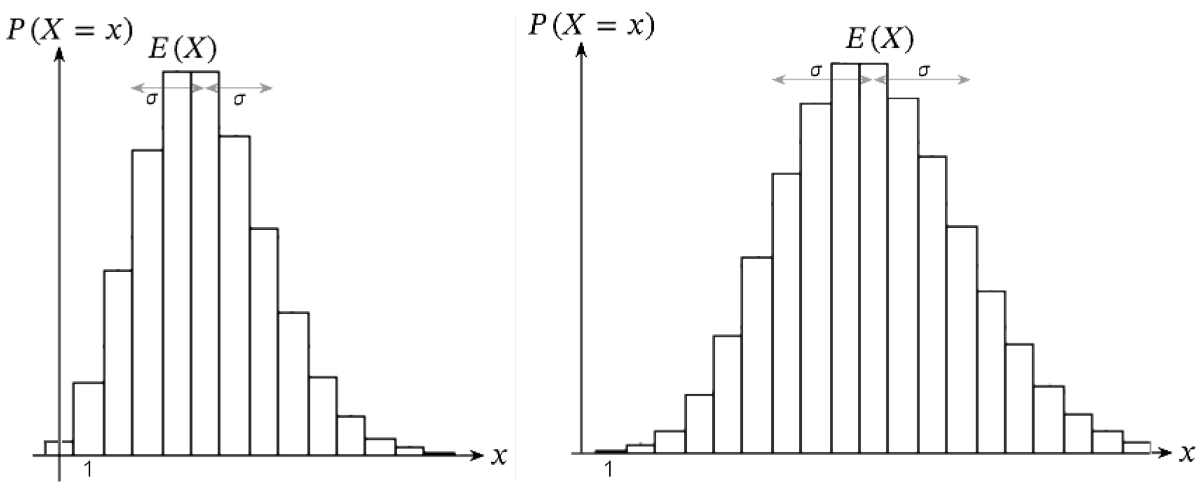


FIGURE 4.3 – Illustration de la signification de l'espérance et l'écart-type d'une variable aléatoire discrète. A gauche, $E(X) = 5$ et $\sigma = 2,2$. Et à droite, $E(X) = 10$ et $\sigma = 3,2$.

Remarque : En pratique, on utilise la formule $V(X) = E(X^2) - E(X)^2$ pour calculer la variance d'une variable aléatoire. Mais, en toute rigueur, c'est la formule $V(X) = E\left(\left(X - E(X)\right)^2\right)$ qui définit la variance et permet de comprendre sa signification. Voici la preuve que les deux formules sont égales :

$$\begin{aligned}
E\left(\left(X-E(X)\right)^2\right) &= E\left(X^2-2XE(X)+E(X)^2\right)\\
&= E(X^2)-E\left(2XE(X)\right)+E\left(E(X)^2\right)\\
&= E(X^2)-2E(X)^2+E(X)^2 \quad \text{(car } E(X) \text{ est une constante)}\\
&= E(X^2)-E(X)^2.
\end{aligned}$$

♦<u>Exemple :</u> Toujours avec l'exemple de l'âge des étudiants de l'EFREI (pages 68 et 72)). On a déjà calculé $E(X)$. Calculons maintenant $E(X^2)$ avec le théorème du transfert

$$\begin{aligned}
E(X^2) = {} & 16^2\times\frac{1}{100}+17^2\times\frac{4}{100}+18^2\times\frac{10}{100}+19^2\times\frac{15}{100}+20^2\times\frac{15}{100}+21^2\times\frac{20}{100}+22^2\times\frac{15}{100}\\
&+23^2\times\frac{10}{100}+24^2\times\frac{5}{100}+25^2\times\frac{3}{100}+26^2\times\frac{1}{100}+27^2\times\frac{1}{100}\\
&= 435.97
\end{aligned}$$

Et donc :

$$V(X)=E(X^2)-(E(X))^2=435.97-20.77^2=4.5771 \text{ et } \sigma(X)\approx 2.14.$$

Autrement dit, en moyenne, si l'on prend un étudiant au hasard, son âge aura un écart de 2.14 ans par rapport à l'âge moyen de 20.77 ans.

En général, la variance n'est pas une application linéaire sur $\ell_2(\Omega)$ (autrement dit, en général on a $V(X+Y)\neq V(X)+V(Y)$). Cependant, on a les deux résultats suivants :

Proposition 4.5

Soient $X\in\ell_2(\Omega)$ une variable aléatoire discrète et $a,b\in\mathbb{R}$. On a

$$V(aX+b)=a^2V(X).$$

Démonstration.

$$\begin{aligned}
V(aX+b) &= E\left(\left(aX+b\right)^2\right)-E(aX+b)^2\\
&= E(a^2X^2+2abX+b^2)-(aE(X)+b)^2\\
&= a^2E(X^2)+2abE(X)+b^2-\left(a^2E(X)^2+2abE(X)+b^2\right)\\
&= a^2\left(E(X^2)-E(X)^2\right)\\
&= a^2V(X).
\end{aligned}$$

□

Proposition 4.6

Soient $X, Y \in \ell_2(\Omega)$ deux variables aléatoires discrètes indépendantes. Alors

$$V(X+Y) = V(X) + V(Y).$$

Remarque : Attention ! La réciproque est fausse : il est possible d'avoir $V(X+Y) = V(X) + V(Y)$ sans que les variables aléatoires soient indépendantes.

Démonstration. Ce résultat est un corollaire du théorème 4.8 et la proposition 4.9 qui sont prouvés à l'exercice 4.5. □

Définition 4.10 (Variable aléatoire centrée réduite)
Soit X une variable aléatoire.
On dit que X est *centrée* si son espérance existe et $E(X) = 0$.
On dit que X est *réduite* si sa variance existe et $V(X) = 1$.

Proposition 4.7 : Variable centrée réduite associée à une variable quelconque

Soit $X \in \ell_2(\Omega)$ une variable aléatoire telle que sa variance existe. Alors, la variable aléatoire

$$Z = \frac{X - E(X)}{\sigma(X)}$$

est une variable aléatoire centrée réduite.

Démonstration. Par linéarité de l'espérance, on a

$$E(Z) = \frac{1}{\sigma(X)}(E(X) - E(X)) = 0.$$

De plus, d'après la proposition 4.5, on a

$$V(Z) = \frac{1}{\sigma(X)^2} V(X - E(X)) = \frac{1}{V(X)} V(X) = 1.$$ □

L'espérance et la variance sont deux quantités qui permettent de décrire grossièrement le comportement d'une variable aléatoire. Cette description peut être affinée en utilisant d'autres quantités :

Définition 4.11 (Moment d'ordre n)
Soit $X \in \ell(\Omega)$ une variable aléatoire discrète et $n \in \mathbb{N}^*$. On appelle *moment d'ordre n*, si elle existe, la quantité

$$m_n(X) = E(X^n) = \sum_{x \in X(\Omega)} x^n P(X = x).$$

On appelle *moment centré d'ordre n*, si elle existe, la quantité

$$\mu_n(X) = E\left((X - E(X))^n\right).$$

Remarque : Cette définition généralise bien l'espérance et la variance puisque $m_1 = \mu_1 = E(X)$ et $\mu_2(X) = V(X)$.

4.4 Covariance et coefficient de corrélation

Lorsque l'on étudie des couples de variables aléatoires, la question principale est de mesurer leur "'influence mutuelle" (dans quelle mesure le cancer de poumon est-il lié au tabagisme ? Le réchauffement planétaire est-il lié aux émissions de CO_2 ?). Pour quantifier ces influences, on introduit des quantités similaires à la variance d'une variable aléatoire.

Définition 4.12 (Covariance entre X et Y)
Soient $X, Y \in \ell_2(\Omega)$ deux variables aléatoires discrètes. On appelle *covariance entre X et Y* la quantité :

$$Cov(X,Y) = E(XY) - E(X)E(Y).$$

Remarque : On a $Cov(X,X) = V(X)$.

Le résultat principal qui montre que la covariance est une quantité pertinente est le théorème suivant :

Théorème 4.8 : Indépendance et covariance

Soient $X, Y \in \ell_2(\Omega)$ deux variables aléatoires discrètes. Alors, on a

$$X \text{ et } Y \text{ indépendantes} \implies Cov(X,Y) = 0.$$

Démonstration. Voir exercice 4.5. □

Remarque : On utilise souvent la contra-posée

$$Cov(X,Y) \neq 0 \implies X \text{ et } Y \text{ ne sont pas indépendantes .}$$

♦Exemples :

⋆ Commençons par donner un exemple qui montre que la réciproque est fausse. Soient X,Y deux variables aléatoires dont la loi de probabilité conjointe est donnée par

	$X=-1$	$X=0$	$X=1$	$P(Y=y)$
$Y=0$	$\frac{1}{3}$	0	$\frac{1}{3}$	**2/3**
$Y=1$	0	$\frac{1}{3}$	0	**1/3**
$P(X=x)$	**1/3**	**1/3**	**1/3**	**1**

Clairement, les deux variables ne sont pas indépendantes puisqu'elles sont construites de sorte que, si une des variables est nulle, l'autre ne l'est pas.
Cependant, on a : $E(X) = 0$ et $E(XY) = 0$ puisque les trois évènements possibles sont $(X = -1, Y = 0)$, $(X = 1, Y = 0)$ et $(X = 0, Y = 1)$ et à chaque fois le produit XY est nul. Ainsi, on trouve $Cov(X,Y) = 0$.
Remarquons que cette table a une propriété de symétrie évidente, c'est cela qui conduit la covariance à être nulle.

⋆ Reprenons maintenant l'exemple de la page 71. Soient X,Y deux variables aléatoires dont la loi de probabilité conjointe est donnée par

	$X=-2$	$X=-1$	$X=0$	$X=1$	$X=2$	$P(Y=y)$
$Y=15$	$\frac{4}{15}$	0	0	0	0	**4/15**
$Y=30$	0	$\frac{2}{15}$	$\frac{1}{6}$	$\frac{1}{30}$	0	**1/3**
$Y=40$	0	0	0	$\frac{1}{5}$	$\frac{1}{5}$	**2/5**
$P(X=x)$	**4/15**	**2/15**	**1/6**	**7/30**	**1/5**	**1**

On a

$$E(X) = -2\times\frac{4}{15} + (-1)\times\frac{2}{15} + 0\times\frac{1}{6} + 1\times\frac{7}{30} + 2\times\frac{1}{5} = -\frac{1}{30}$$
$$E(Y) = 15\times\frac{4}{15} + 30\times\frac{1}{3} + 40\times\frac{2}{5} = 30$$
$$E(XY) = (-2)\times 15\times\frac{4}{15} + (-1)\times 30\times\frac{2}{15} + 1\times 30\times\frac{1}{30} + 1\times 40\times\frac{1}{5} + 2\times 40\times\frac{1}{5} = 13$$

et donc on trouve

$$Cov(X,Y) = E(XY) - E(X)E(Y) = 12$$

ce qui permet de confirmer que X et Y ne sont pas indépendantes (on le savait déjà).

Le deuxième résultat important où la covariance rentre en jeu est le suivant :

Proposition 4.9 : Variance de $X+Y$

Soient $X, Y \in \ell_2(\Omega)$ deux variables aléatoires discrètes. Alors, on a

$$V(X+Y) = V(X) + V(Y) + 2Cov(X,Y).$$

Démonstration. Voir exercice 4.5. □

Proposition 4.10 : Propriétés de la covariance

La covariance est une application symétrique et bilinéaire.
Autrement dit : $\forall X, Y, Z \in \ell_2(\Omega), \forall \alpha, \beta \in \mathbb{R}$

$$\begin{aligned} Cov(X,Y) &= Cov(Y,X) \\ Cov(\alpha X + \beta Y, Z) &= \alpha Cov(X,Z) + \beta Cov(Y,Z) \\ Cov(X, \alpha Y + \beta Z) &= \alpha Cov(X,Y) + \beta Cov(X,Z) \end{aligned}$$

Démonstration. La symétrie se constate directement dans la définition de la covariance.
Si l'on montre la linéarité par rapport à la première variable, la symétrie force la linéarité par rapport à la deuxième variable.
Enfin, la linéarité par rapport à la première variable est une conséquence de la linéarité de l'espérance :

$$\begin{aligned} Cov(\alpha X + \beta Y, Z) &= E\left((\alpha X + \beta Y)Z\right) - E(\alpha X + \beta Y)E(Z) \\ &= \alpha E(XZ) + \beta E(YZ) - \alpha E(X)E(Z) - \beta E(Y)E(Z) \\ &= \alpha Cov(X,Z) + \beta Cov(Y,Z). \end{aligned}$$

□

Pour l'instant, nous avons vu que si la covariance est non nulle, alors les variables X et Y ne sont pas indépendantes. Mais nous n'avons pas discuté du sens de la valeur trouvée : qu'apprend-on sur le lien entre X et Y si l'on trouve que $Cov(X,Y) = 12$ ou $Cov(X,Y) = -3$?

En réalité, pour rendre cette discussion précise, nous devons introduire une nouvelle quantité :

Définition 4.13 (Coefficient de corrélation entre X et Y)
Soient $X, Y \in \ell_2(\Omega)$ deux variables aléatoires discrètes telles que leurs variances sont non nulles (i.e. les variables ne sont pas des constantes). On appelle *coefficient de corrélation entre X et Y* la quantité :

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

♦Exemple : Voici une compilation des conclusions de plusieurs études [1] sur l'efficacité des doses de vaccinations Pfizer contre la COVID. On note X le nombre de doses de vaccinations reçues par la personne et Y la variable binaire "donner positif à un test PCR dans une période de deux semaines" :

	$X=0$	$X=1$	$X=2$	$X=3$	$\boldsymbol{P(Y=y)}$
$Y=0$	0.079	0.09	0.093	0.675	**0.937**
$Y=1$	0.021	0.01	0.007	0.025	**0.063**
$\boldsymbol{P(X=x)}$	**0.10**	**0.10**	**0.10**	**0.70**	**1**

Calculons le coefficient de corrélation entre X et Y : d'abord, on a $E(X) = 2.40$ et $\sigma(X) \approx 1.02$ d'une part et $E(Y) = 0.063$ et $\sigma(Y) \approx 0.24$ d'autre part. De plus, on a $E(XY) = 0.099$. Ainsi, on trouve, pour le coefficient de corrélation

$$\rho(X,Y) = \frac{E(XY) - E(X)E(Y)}{\sigma(X)\sigma(Y)} = \frac{0.099 - 2.40 \times 0.063}{1.02 \times 0.24} \approx -0.21.$$

Le signe de $\rho(X,Y)$ montre que Y a tendance à diminuer lorsque l'on augmente le nombre de doses de vaccination. En effet :

Théorème 4.11 : Signification du coefficient de corrélation

Soient $X,Y \in \ell_2(\Omega)$ deux variables aléatoires discrètes de variance non nulle. On a

$$\begin{aligned} -1 \leq &\rho(X,Y) \leq 1 \\ \rho(X,Y) = 1 &\iff Y = aX + b \\ \rho(X,Y) = -1 &\iff Y = -aX + b \qquad \text{où } b \in \mathbb{R} \text{ et } a = \frac{\sigma(Y)}{\sigma(X)} > 0. \end{aligned}$$

Démonstration. Voir exercice 4.20. □

Autrement dit, le coefficient de corrélation mesure le degré de linéarité entre X et Y :

- Si $\rho(X,Y) = 0$, on dit que les variables aléatoires *ne sont pas linéairement corrélées* (ce qui est différent de dire qu'elles sont indépendantes).
- Si $\rho(X,Y) > 0$, Y a tendance à augmenter si X augmente.
- Si $\rho(X,Y) < 0$, Y a tendance à diminuer si X augmente.
- Si $|\rho(X,Y)| = 1$, Y est une fonction linéaire de X.

1. Il s'agit de l'expérience REACT-1 conduite par le gourvernement anglais en Juin-Juillet 2021 (voir ici) et les données du *Maccabi Healthcare Services* recueillies en 2020 et 2021 (voir ici).

Exemple : Soit X une variable aléatoire discrète donc le support est $X(\Omega) = \{-1,0,1\}$ et $P(X=-1) = P(X=0) = P(X=1)$. Considérons de plus la variable aléatoire $Y = X^2$. Les deux variables aléatoires sont clairement non-indépendantes mais le coefficient de corrélation va s'annuler car une augmentation de X peut mener, avec la même probabilité, à une augmentation ou une diminution de Y.

Explicitement, le calcul de $\rho(X,Y)$ donne :

- $E(X) = -\frac{1}{3} + \frac{1}{3} = 0$,
- $E(XY) = -P(X=-1, Y=1) + P(X=1, Y=1) = -\frac{1}{3} + \frac{1}{3} = 0$,
- Donc $E(XY) - E(X)E(Y) = 0$ et $\rho(X,Y) = 0$.

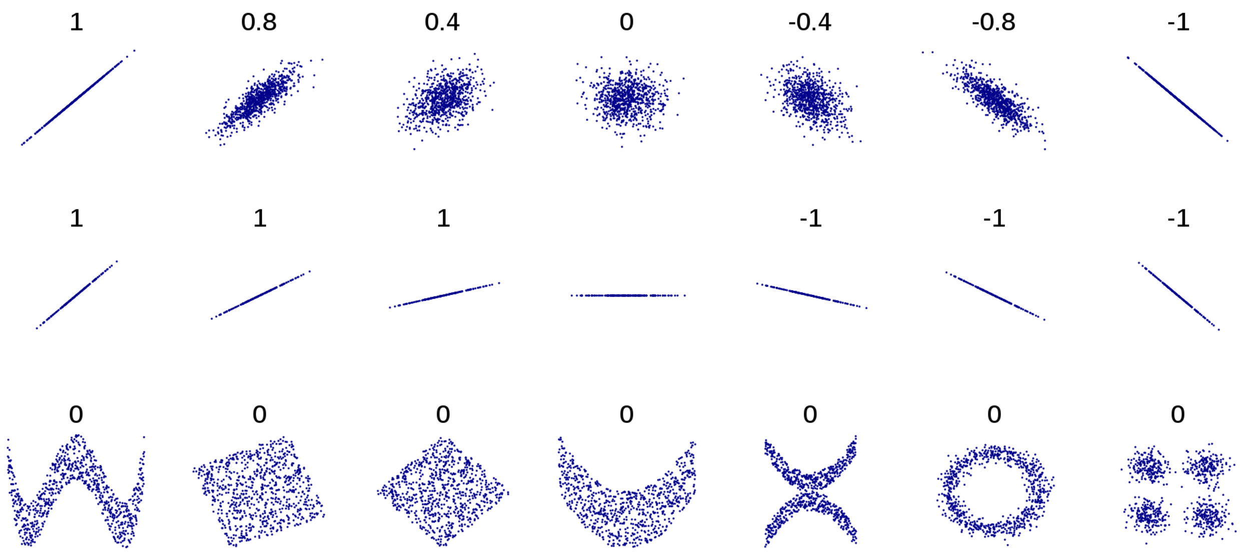


FIGURE 4.4 – Exemples de nuages de points avec des coefficients de corrélation différents dans une représentation Y en fonction de X.

Figure tirée de cette page Wikipédia.

4.5 Lois usuelles

On va maintenant regarder certaines lois de probabilités de référence que l'on rencontre souvent.

4.5.1 Loi uniforme

Définition 4.14 (Loi uniforme)
Soit X une variable aléatoire discrète. On dit que X suit une *loi uniforme* si toutes les valeurs possibles de X sont équiprobables.
En particulier, soit $n \in \mathbb{N}^*$. On note $X \sim \mathscr{U}(\llbracket 1,n \rrbracket)$ si

$$X(\Omega) = \llbracket 1,n \rrbracket \text{ et } \forall k \in \llbracket 1,n \rrbracket, P(X=k) = \frac{1}{n}.$$

Remarque : D'après la définition, si X suit une loi uniforme, $X(\Omega)$ est forcément un ensemble fini et, pour $x \in X(\Omega)$, on a $P(X=x) = \dfrac{1}{|X(\Omega)|}$.

Proposition 4.12 : Espérance et variance de la loi uniforme

Si $X \sim \mathscr{U}(\llbracket 1,n \rrbracket)$, alors $E(X) = \dfrac{n+1}{2}$ et $V(X) = \dfrac{n^2-1}{12}$.

Démonstration. Par définition, on a

$$E(X) = \sum_{k=1}^{n} kP(X=k) = \frac{1}{n}\sum_{k=1}^{n} k = \frac{n+1}{2}.$$

De même,

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

et donc

$$V(X) = E(X^2) - E(X)^2 = \frac{2(n+1)(2n+1) - 3(n+1)^2}{12} = \frac{(n+1)(n-1)}{12}.$$

□

4.5.2 Loi de Bernoulli

L'expérience type d'une loi de Bernoulli est une expérience à deux issues, l'une représentant le "succès" et l'autre l'"échec".

Définition 4.15 (Loi de Bernoulli)
Soit X une variable aléatoire discrète et $p \in]0,1[$. On dit que X suit une *loi de Bernoulli de paramètre* p, noté $X \sim \mathscr{B}(p)$, si

$$X(\Omega) = \{0,1\} \text{ et } \begin{cases} P(X=1) = p \\ P(X=0) = 1-p. \end{cases}$$

<u>Exemples :</u> Il y a une infinité d'exemples !

- ⋆ On prend une personne de la population française au hasard et on regarde si elle est végétarienne ($X = 1$) ou non ($X = 0$). Alors, $X \sim \mathscr{B}(0.02)$ (on estime qu'il y a 1.4 millions de végétariens en France.)
- ⋆ On prend un jour de l'année dernière au hasard et on regarde si ce jour il a plu à Paris ($X = 1$) ou non ($X = 0$). Alors, $X \sim \mathscr{B}(\frac{111}{365})$ (en moyenne, Paris connaît 111 jours de pluie par an).
- ⋆ On tire un étudiant de L2 de l'EFREI au hasard et on regarde si c'est un homme $X = 0$ ou une femme $X = 1$. Alors $X \sim \mathscr{B}(\frac{1}{5})$ (il y a 20% de filles en L2).

Proposition 4.13 : Espérance et variance de la loi de Bernoulli

Si $X \sim \mathscr{B}(p)$, alors $E(X) = p$ et $V(X) = p(1-p)$.

Démonstration. Il suffit de faire les calculs. □

4.5.3 Loi binomiale

Soit $n \in \mathbb{N}^*$. On répète n fois, de manière indépendante, une expérience de Bernoulli et l'on note X le nombre de succès obtenus. Par exemple, on prend n atomes de plutonium ^{240}Pu et l'on compte combien se désintègrent dans les cinq minutes qui suivent (on sait que les désintégrations se comportent de manière indépendante).

Déterminons la loi de probabilité de X. D'abord, on peut avoir de 0 à n succès, donc $X(\Omega) = [\![0,n]\!]$. De plus, quand on a $X = k$ succès parmi les n répétitions, on a aussi $n-k$ échecs. Puisque chaque expérience est indépendante des autres, la probabilité d'avoir k succès et $n-k$ échecs *dans un certain ordre donné* est $p^k(1-p)^{n-k}$ (on suit les probabilités conditionnelles le long du chemin). Mais attention, pour avoir la probabilité $P(X=k)$, il ne faut pas oublier de compter tous les chemins possibles, il y en a autant que le nombre de façons de choisir les k expériences parmi les n qui donneront un succès. Ce nombre est donné par le coefficient du binôme :

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

.

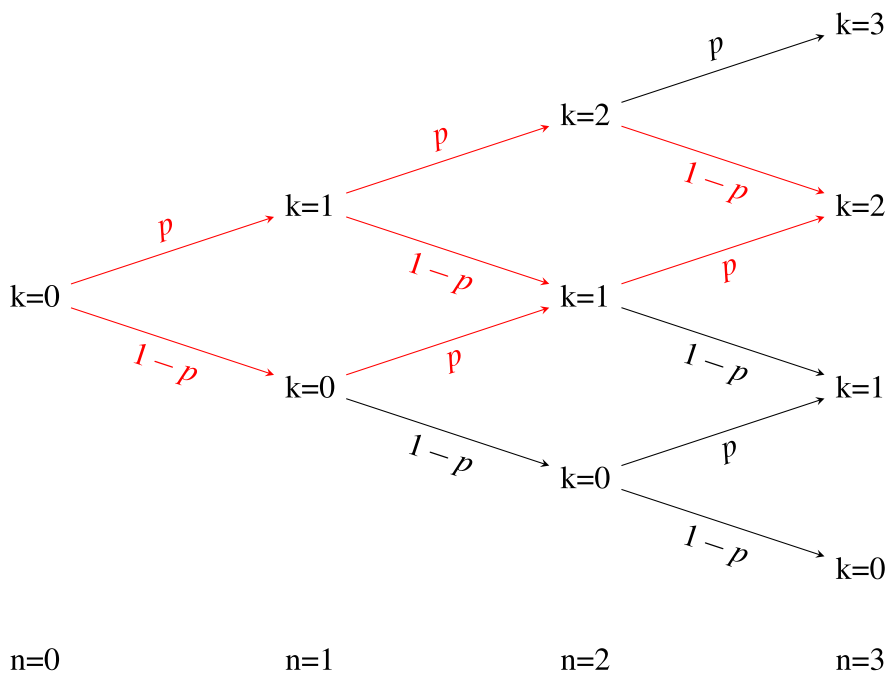


FIGURE 4.5 – Arborescence pour $n=3$. Nous avons trois façons d'avoir $k=2$ succès, ce qui correspond aux tracés en rouge. Il s'agit bien de choisir 2 succès parmi 3 tentatives, ou bien 1 échec parmi 3 tentatives, ce qui correspond bien aux $\binom{3}{2} = \binom{3}{1} = 3$ chemins.

Définition 4.16 (Loi binomiale)
Une variable aléatoire discrète X suit une *loi binomiale de paramètres n et p*, noté $X \sim \mathscr{B}(n,p)$, si X compte le nombre de succès dans n répétitions identiques et indépendantes d'une expérience de Bernoulli où la probabilité de succès est p.
Dans ce cas là, on a $X(\Omega) = [\![0,n]\!]$ et $\forall k \in [\![0,n]\!]$,

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

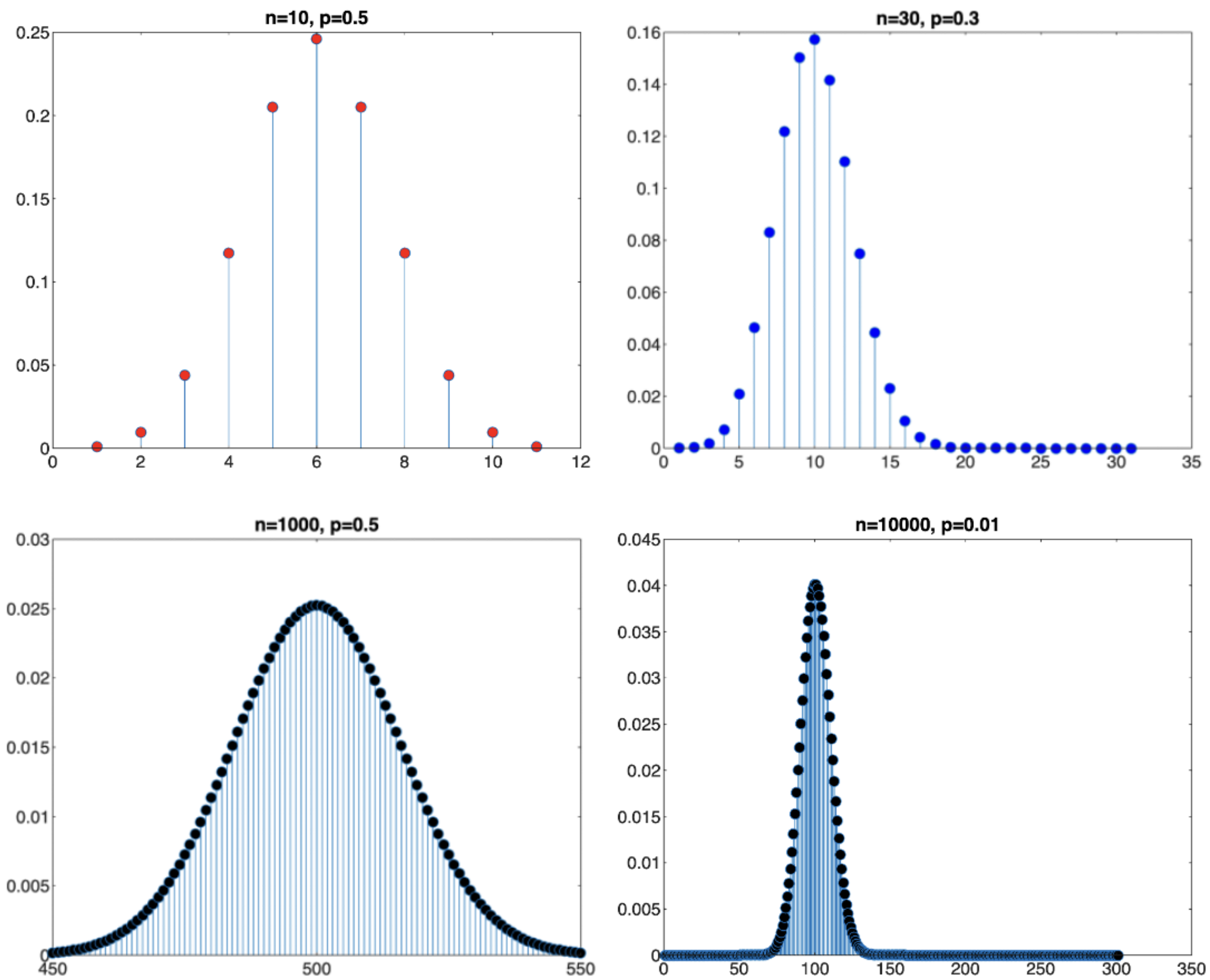


FIGURE 4.6 – La loi binomiale pour plusieurs valeurs de n et p. Dans chaque graphe, la valeur en abscisse est k et la valeur en ordonnée est $P(X = k)$. Le script Matlab qui produit (par exemple) la troisième courbe est `n=1000; p=0.5; x=450:550; y=binopdf(x,n,p); stem(x,y).`

Remarque : On peut vérifier que $\sum_{k=0}^{n} P(X = k) = 1$. En effet, si l'on se rappelle de la formule du binôme de Newton

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k},$$

on a

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

Exemples :

⋆ Lorsque je tape un symbole sur le clavier de mon ordinateur, il y a une probabilité de $p = \frac{1}{100}$ que je me trompe de touche. Si les frappes sont indépendantes et que cette page du polycopié contient environ 2500 caractères, la variable aléatoire X qui compte le nombre de coquilles (avant relecture !) est telle que $X \sim \mathscr{B}(2500, \frac{1}{100})$.

- ⋆ La surface terrestre et marine se comportent comme des corps noirs et émettent des rayonnements (i.e, des photons) infrarouges en direction de l'espace. Mais chacun de ces photons à une probabilité $p = 0.95$ d'être absorbé par l'atmosphère (c'est ce que l'on appelle l'effet de serre). Si l'on note X le nombre de ces photons absorbés par seconde, alors on a $X \sim \mathscr{B}(N, p)$ où N est le nombre moyen de photons émis par la surface terrestre en une seconde.

Remarque : La loi binomiale est associée à des tirages successifs avec remise, mais on considère souvent que, si la taille n de l'échantillon à tirer est petite par rapport à la taille de la population N, un tirage simultané de l'échantillon est assimilable avec n tirages successifs avec remise.

Par exemple : on considère la population française ($N = 67,75 \times 10^6$), et on s'intéresse à la maladie du VIH (170 000 personnes atteintes). On prend un échantillon aléatoire de $n = 1000$ personnes (tirage simultané) et on appelle X le nombre de malades dans l'échantillon. Puisque $n/N = 0.000015$ est petit devant 1, on considère que X suit une loi binomiale de paramètres n et $p = \frac{0.17}{67.75} \approx 0.002$, autrement dit on assimile le tirage simultané à un tirage successif avec remise.

Proposition 4.14 : Espérance et variance de la loi binomiale

Si $X \sim \mathscr{B}(n, p)$, alors $E(X) = np$ et $V(X) = np(1-p)$.

Démonstration. Deux façons de faire la preuve : soit en faisant les calculs directement (ce serait un bon exercice du cours d'Analyse SM201), soit (point de vue bien plus fécond) en voyant X comme une somme de variables aléatoires indépendantes $X_1, X_2, \dots, X_n$ qui suivent chacune la même loi de Bernoulli de paramètre p. Ainsi, puisque l'espérance est linéaire

$$E(X) = E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n) = nE(X_1) = np.$$

De même, puisque l'on a $V(X+Y) = V(X) + V(Y)$ si X, Y sont indépendantes,

$$V(X) = nV(X_1) = np(1-p).$$

□

Remarque : Si l'on tire de manière indépendante 200 fois sur une cible et qu'à chaque fois, on a la probabilité 30% de la toucher. Instantanément, nous affirmons qu'on touche 60 fois la cible, en moyenne. Donc intuitivement, et sans connaître la formule, on a fait le calcul np de l'espérance.

Proposition 4.15 : Somme de variables aléatoires indépendantes binomiales

Soient X et Y deux variables aléatoires discrètes <u>indépendantes</u> telles que $X \sim \mathscr{B}(n,p)$ et $Y \sim \mathscr{B}(m,p)$. Alors, on a $X+Y \sim \mathscr{B}(n+m,p)$.

Démonstration. Si l'on reprend le point de vue de la preuve précédente, ce résultat est presque évident. En effet, X est la somme de n variables suivant un loi de Bernoulli de paramètre p et Y est la somme de m variables suivant un loi de Bernoulli de paramètre p, donc il est clair que $X+Y$ est bien la somme de $n+m$ variables suivant un loi de Bernoulli de paramètre p : on enrichit simplement l'expérience de n tirages par m tirages supplémentaires. $\square$

Point historique #2 (L'expérience de Bernoulli et la loi binomiale)

Le cadre conceptuel fourni par l'expérience de Bernoulli et la loi binomiale joue un rôle central dans le développement des probabilités du 19ème siècle. Il fut introduit en 1713 par Jakob Bernoulli dans son oeuvre posthume *Ars Conjectandi*.

En particulier, il y prouva le premier théorème mathématiquement intéressant en probabilités (qu'il appela le "théorème d'or") :

Théorème d'or de Bernoulli

Soit S_n le nombre de victoires dans les n répétitions identiques du jeu dont la probabilité de victoire est p. Alors

$$\forall \varepsilon > 0, \lim_{n \to +\infty} P\left(|\frac{S_n}{n} - p| < \varepsilon\right) = 1.$$

(En mots : la probabilité de victoire correspond, sur le long terme, à la fréquence des victoires.)

Le problème mathématique principal du 19ème siècle en théorie des probabilités va être d'affiner et étendre ce théorème, grâce en particulier aux travaux de Abraham De Moivre, Pierre-Simon Laplace et Pafnouti Tchebychev.

4.5.4 Loi binomiale négative

Dans le cas d'une loi binomiale :

- ⋆ on fixe le nombre de répétitions d'une certaine expérience de Bernoulli,
- ⋆ on compte le nombre de succès obtenus.

Mais on peut s'intéresser à la situation inverse :

- ⋆ on fixe le nombre de succès que l'on veut obtenir,
- ⋆ on compte le nombre de répétitions nécessaires pour y arriver.

Définition 4.17 (Loi binomiale négative)
Une variable aléatoire discrète X suit une *loi binomiale négative de paramètres r et p*, notée $X \sim \mathscr{BN}(r,p)$, si X compte le nombre de répétitions indépendantes d'une expérience de Bernoulli (où la probabilité de succès est p) nécessaires pour atteindre $r > 0$ succès.
Dans ce cas là, on a $X(\Omega) = [\![r, +\infty[\![$ et $\forall j \in [\![r, +\infty[\![$,

$$P(X = j) = \binom{j-1}{r-1} p^r (1-p)^{j-r}.$$

♦Exemple : A partir de la L3, un étudiant de l'EFREI qui doit passer le rattrapage d'une certaine matière doit repasser chaque année ce rattrapage jusqu'à ce qu'il ou elle réussisse à le valider. En général, pour les rattrapages de mathématiques de L3, il y a 80% des copies qui valident. Si l'on note X le nombre de fois qu'un étudiant doit passer le rattrapage, on a $X \sim \mathscr{BN}(1, 0.8)$.[2]

Par exemple, on peut se demander quelle est la probabilité qu'un étudiant ne réussisse pas à valider avant la fin du Master 2 une matière qu'il ou elle n'a pas validé en L3 (avant le rattrapage). On a :

$$\begin{aligned} P(X > 3) &= 1 - P(X=1) - P(X=2) - P(X=3) \\ &= 1 - 0.8 - 0.2 \times 0.8 - 0.2^2 \times 0.8 \\ &\approx 0.008. \end{aligned}$$

Proposition 4.16 : Espérance et variance de la loi binomiale négative

Si $X \sim \mathscr{BN}(r,p)$, alors $E(X) = \dfrac{r}{p}$ et $V(X) = \dfrac{r(1-p)}{p^2}$.

Démonstration. Voir exercice 4.19. □

2. On pourrait s'attendre à ce que les répétitions ne soient pas indépendantes, mais l'expérience montre que le fait d'avoir déjà passé le rattrapage n'a pas d'influence sur le résultat du rattrapage suivant...

4.5.5 Loi de Poisson

On termine avec une loi un peu à part des autres et qui est principalement associée aux évènements rares comme les accidents, les désintégrations radioactives, les défauts de crédits, etc. On verra qu'elle est dans tous ces cas une excellente approximation de la loi binomiale.

Définition 4.18 (Loi de Poisson)
Soient X une variable aléatoire discrète et $\lambda \in \mathbb{R}^+$. On dit que X suit une *loi de Poisson de paramètre* λ, noté $X \sim \mathscr{P}(\lambda)$, si $X(\Omega) = \mathbb{N}$ et $\forall k \in \mathbb{N}$,

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

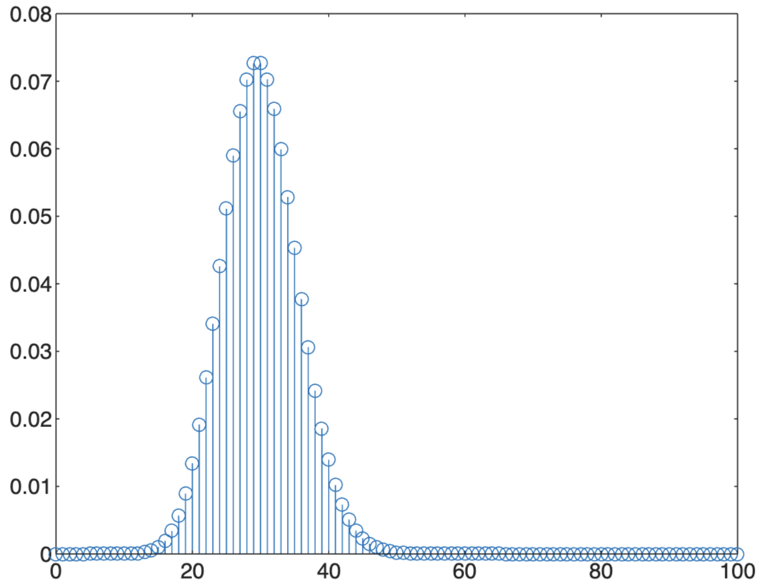


FIGURE 4.7 – Représentation de la loi de Poisson pour $\lambda = 30$.

Remarque : On vérifie bien que

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

Exemples : Voici plusieurs exemples de variables qui sont bien modélisées par des lois de Poisson.

- ⋆ Le nombre de coquilles par page de ce polycopié après relecture.
- ⋆ Le nombre de patients arrivant avec une fracture dans un hôpital un jour donné.
- ⋆ Le nombre de personnes ayant plus de 100 ans dans une communauté donnée.
- ⋆ Le nombre de vidéos sur Youtube dépassant le million de vues.
- ⋆ Le nombre de connexions à un serveur web durant un intervalle de temps T.

Proposition 4.17 : Espérance et variance de la loi de Poisson

Si $X \sim \mathscr{P}(\lambda)$, on a $E(X) = V(X) = \lambda$.

Démonstration. On a

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

De même,

$$\begin{aligned} E(X^2) = \sum_{k=0}^{+\infty} k^2 P(X=k) = e^{-\lambda} \sum_{k=1}^{+\infty} k \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} \sum_{K=0}^{+\infty} (K+1) \frac{\lambda^K}{K!} = \lambda \left(E(X) + e^{-\lambda} \sum_{K=0}^{+\infty} \frac{\lambda^K}{K!} \right) \\ = \lambda^2 + \lambda \end{aligned}$$

et donc $V(X) = E(X^2) - E(X) = \lambda$. □

La raison précise pour laquelle la loi de Poisson se révèle être pertinente dans autant de cas est donnée par le résultat suivant

Théorème 4.18 : Lien entre la loi binomiale et la loi de Poisson

Soit $\lambda \in \mathbb{R}^+$. Dans le cas où $n \to +\infty$ (grand nombre de répétitions), $p \to 0$ (l'évènement succès est rare) et $np \to \lambda$ (le nombre de succès moyen se maintient constant), on a

$$\lim_{\substack{n \to +\infty \\ p \to 0 \\ np \to \lambda}} \mathscr{B}(n,p) = \mathscr{P}(\lambda).$$

Démonstration. Mathématiquement, le théorème limite de Poisson est donc

$$\lim_{\substack{n \to +\infty \\ p \to 0 \\ np \to \lambda}} \binom{n}{k} p^k (1-p)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}.$$

La preuve de ce résultat dépasse les notions de ce cours car elle utilise l'approximation de Stirling

$$n! \underset{+\infty}{\sim} \sqrt{2\pi n} \left(\frac{n}{e} \right)^n$$

et la limite

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

□

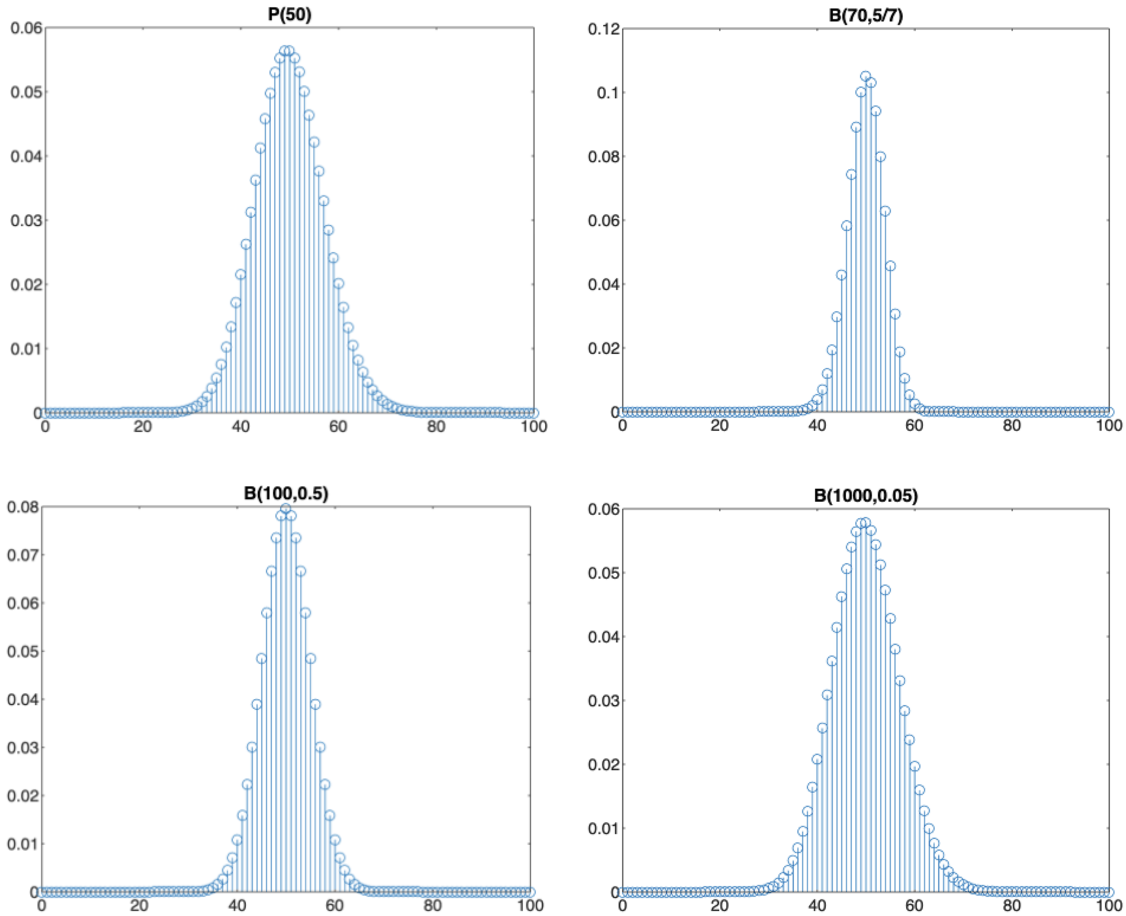


FIGURE 4.8 – En haut à gauche : représentation de la loi de Poisson pour $\lambda = 50$. Les trois autres graphes sont des représentations de lois binomiales où $np = 50$. On voit que plus n augmente (et donc p diminue), plus la loi binomiale se ressemble à la loi de Poisson.

Proposition 4.19 : Somme de variables aléatoires indépendantes de Poisson

Soient X et Y deux variables aléatoires <u>indépendantes</u> telles que $X \sim \mathscr{P}(\lambda)$ et $Y \sim \mathscr{P}(\theta)$. Alors, on a $X + Y \sim \mathscr{P}(\lambda + \theta)$.

Démonstration. Voir exercice 4.21. □

FIGURE 4.9 – Siméon Denis Poisson (1781-1840).

4.6 Acquis d'apprentissage

À la fin de ce chapitre, vous devrez être en mesure de :

⋆ Déterminer la loi de probabilité d'une variable aléatoire discrète.

⋆ Déterminer la loi de probabilité d'une nouvelle variable Y exprimée en fonction d'une variable X dont on connaît la loi de probabilité.

⋆ Calculer l'espérance, la variance et l'écart-type d'une variable aléatoire discrète.

⋆ Déterminer la loi de probabilité conjointe de deux variables aléatoires discrètes.

⋆ Déterminer si deux variables aléatoires discrètes sont indépendantes.

⋆ Calculer la covariance et le coefficient de corrélation entre deux variables aléatoires discrètes.

⋆ Reconnaître les situations où l'on doit utiliser les lois uniforme, binomiale, binomiale négative ou de Poisson.

⋆ Connaître les caractéristiques principales des lois uniforme, binomiale, binomiale négative et de Poisson.

FZ: Faire une table résumé avec l'espérance et la variance des lois usuelles.

4.7 Exercices

♦= Exercice obligatoire.

[*] = Application directe du cours : sert à vérifier que vous avez bien compris les notions et techniques de base.

[**] = Demande plus de réflexion mais vous devez être capables de le résoudre et maîtriser les techniques utilisées dans cet exercice.

[***] = Exercice difficile pour aller plus loin, découvrir de nouvelles idées ou techniques. Pour les plus motivés !

Faire le point en autonomie : QCM sur les notions de base

Soient X et Y deux variables aléatoires sur un espace probabilisé fini (Ω, P) :

Item	Enoncé
A	Si X suit une loi uniforme $\mathscr{U}([\![1,n]\!])$ alors $P(X=k)=\frac{1}{n^2}$
B	$E(X^2) \leq E(X)^2$
C	Si $X(\Omega)=[\![1,n]\!]$ alors $E(e^X)=\sum_{k=1}^n e^k P(X=k)$
D	Si X suit une loi de Bernoulli $\mathscr{B}(p)$ alors $E(X)=\frac{1}{p}$
E	Si X suit une loi binomiale $\mathscr{B}(n,p)$ alors $E(X)=np$
F	Si X suit une loi binomiale négative $\mathscr{BN}(r,p)$ alors $X(\Omega)=[\![1,r]\!]$
G	Si X suit une loi binomiale $\mathscr{B}(n,p)$ alors $P(X=0)=(1-p)^n$
H	Pour que X et Y ne soient pas indépendantes, il suffit que $Cov(X,Y) \neq 0$
I	X et Y sont indépendantes si $\forall (x,y) \in X(\Omega)\times Y(\Omega), P(X=x,Y=y)=P(X=x)P(Y=y)$.

Exercice 4.1 [*]

On considère les familles de trois enfants et l'on note X la variable aléatoire qui compte le nombre de garçons. Donner la loi de probabilité de X en supposant l'équiprobabilité des sexes.

♦ **Exercice** 4.2 [*]

On jette simultanément trois monnaies. Soient N_F et N_P le nombre de faces et de piles obtenus respectivement. On considère la variable aléatoire $X=N_F-N_P$. Donner la loi de probabilité de X.

♦ **Exercice** 4.3 [*]

Un secrétariat a établit sur plusieurs années la distribution du nombre X d'appels téléphoniques par minute les jours ouvrés (du lundi au vendredi), tabulée ci-dessous :

Nombre d'appels x	0	1	2	3	4	5	6	7
$P(X \leq x)$	$0,05$	$0,15$	$0,3$	$0,5$	$0,7$	$0,85$	$0,95$	1

a) Remplissez le tableau ci-dessous

Nombre d'appels x	0	1	2	3	4	5	6	7
$P(X=x)$								

b) Calculer $E(X)$ et $V(X)$.

c) Est-ce que, en moyenne, il y a autant de jours avec moins de deux appels par minute qu'il y en a avec plus de cinq ? Expliquer.

d) Quelle est la probabilité de recevoir au plus 2 appels par minute ?

e) Quelle est la probabilité que, sur dix minutes, il y ait exactement trois appels chaque minute ?

f) Parmi les minutes avec plus de quatre appels, quelle proportion en reçoit exactement sept ?

g) Soit Y le nombre d'appels téléphoniques par heure les jours ouvrés. Calculer $E(Y)$ et $V(Y)$.

Exercice 4.4 [*]

Cet exercice est tiré de S. Ross, *A first course in Probability*, p. 186.

Quatre autocars transportent 148 étudiants de l'EFREI au WEI. Les bus transportent, respectivement, 40,33,25 et 50 étudiants. Un étudiant parmi les 148 et un conducteur parmi les quatre sont tirés au hasard. On désigne par X le nombre d'étudiants transportés dans le bus de l'étudiant tiré au hasard et par Y le nombre d'étudiants transportés dans le bus du conducteur tiré au hasard.

a) A votre avis, parmi $E(X)$ et $E(Y)$ lequel est plus grand ?

b) Maintenant, calculez $E(X), E(Y), V(X)$ et $V(Y)$.

Exercice 4.5 [**] **Preuve du théorème 4.8 et des propositions 4.6 et 4.9.**

Soient X et Y deux variables aléatoires discrètes sur Ω.

a) Prouver que l'on a $V(X+Y) = V(X)+V(Y)+2Cov(X,Y)$.

b) Prouver que l'on a X et Y indépendantes $\Longrightarrow Cov(X,Y) = 0$.

Exercice 4.6 [*]

Supposons que, dans une certaine communauté, 15% des couples n'ont pas d'enfants, 20% ont un enfant, 35% ont deux enfants et 30% ont trois enfants. On tire un couple au hasard et on note G le nombre de garçons de cette famille et F le nombre de filles. En supposant que la naissance d'un garçon ou d'une fille est équiprobable, donner la loi de probabilité conjointe de G et F.

Exercice 4.7 [*]

Soient X et Y deux variables aléatoires discrètes dont la loi de probabilité conjointe est donnée ci-dessous :

	$Y=2$	$Y=4$	$Y=5$
$X=1$	$2c$	c	c
$X=2$	$4c$	$2c$	$3c$
$X=3$	$6c$	$3c$	$2c$

a) Trouver la valeur de la constante c.

b) Calculer $P(X \leq 2, Y \leq 4)$.

c) Calculer $P(Y=2 | X=1)$.

d) Les deux variables sont-elles indépendantes ?

♦ **Exercice** 4.8 [*] **Couples**

Soit un urne contenant une boule blanche, une verte et une boule rouge. On extrait successivement les trois boules de l'urne. On note X le rang d'apparition de la boule blanche et Y le rang d'apparition de la boule rouge. On représente la loi conjointe de X et Y sous la forme d'un tableau :

	$X=1$	$X=2$	$X=3$
$Y=1$			
$Y=2$			
$Y=3$			

Ω est l'ensemble des anagrammes du mot RBV.

a) Calculer $Card(\Omega)$.

b) Remplir le tableau de la loi conjointe.

c) Déterminer les lois de probabilité de X et Y. Sont-elles indépendantes ?

d) Calculer l'espérance et la variance de X et Y.

e) Calculer $Cov(X,Y)$ puis $\rho(X,Y)$.

♦ **Exercice** 4.9 [**] **Couples et changement de variable**

On considère le couple de variables aléatoires (X,Y) de la loi conjointe définie par la table ci-dessous :

	$Y=0$	$Y=1$	$Y=2$	$Y=3$
$X=1$	$0,1$	$0,05$	0	$0,05$
$X=2$	$0,2$	$0,1$	$0,1$	$0,2$
$X=3$	$0,1$	0	$0,05$	$0,05$

a) Vérifier la loi de (X,Y).

b) Déterminer les lois marginales de X et Y. Sont-elles indépendantes ? Sinon, comment pourriez-vous modifier trois cases pour rendre les deux lois indépendantes.

Dans la suite, on ne modifie pas la table.

c) Calculer l'espérance et la variance de X et Y.

d) Calculer $Cov(X,Y)$.

e) Soit $W = X^2$. Déterminer la loi de W.

f) Soit $Z = (X-2)^2$. Déterminer la loi de Z.

g) Soit $U = X.Y$. Déterminer la loi de U.

h) Soit $V = min(X,Y)$. Déterminer la loi de V.

i) Former la table conjointe de U et V.

♦ **Exercice** 4.10 [*]

Les infections invasives à méningocoques (IIM) sont des infections graves mais relativement rares, leur fréquence annuelle étant estimée à 1,5 cas/100 000 habitants par an.

a) Dans une ville de 50 000 habitants, quelle est la probabilité d'observer plus de 2 cas d'IIM par an ?

b) On sait que 80% des cas d'IIM surviennent chez des sujets âgés de moins de 25 ans. On sait par ailleurs que les moins de 25 ans représentent 31% de la population française. Quelle est la probabilité d'avoir une IIM sachant que l'on a moins de 25 ans ?

c) Dans un département de 1 million d'habitants, quelle est la probabilité d'observer moins de 2 cas d'IIM sur un an chez les moins de 25 ans ?

♦ **Exercice** 4.11 [**] **Une puce brownienne 1**

Soit $p \in [0;1]$. Une puce se déplace aléatoirement sur une droite d'origine O. A chaque instant, elle fait un bond d'une unité vers la droite ou d'une unité vers la gauche avec les probabilités respectives p et $q = 1 - p$. A l'instant initial, la puce est à l'origine.

Etant donné $n \in \mathbb{N}^*$, on note X_n l'abscisse de la puce à l'instant n.

Déterminer la loi de X_n, son espérance et sa variance.

Exercice 4.12 [***] **Une puce brownienne 2**

Reprenons l'exercice précédent. Mais cette fois-ci, étudions la distribution de probabilité quand $p = q = \frac{1}{2}$, c'est à dire quand il n'y a plus de dérive.

a) Montrer par récurrence que

$$\forall n \in \mathbb{N}^*, \sum_{k=1}^{n} \frac{\binom{2k}{k}}{4^k} = \frac{2n+1}{4^n} \binom{2n}{n} - 1.$$

b) Soit $k \in \mathbb{N}^*$. Calculer la probabilité que la puce soit à l'origine en l'instant $2k$.

c) Soit $n \in \mathbb{N}^*$. On note N_n la variable aléatoire égale au nombre de passages à l'origine entre l'instant 0 et n. Calculer l'espérance de N_n.

d) Donner un équivalent de N_n lorsque n tend vers l'infini en utilisant la formule de Stirling : $n! \underset{+\infty}{\sim} \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$

Exercice 4.13 [**] **Un secrétaire très occupé**

Un secrétaire effectue n appels téléphoniques vers n personnes (n est un entier supérieur à 2). Ces appels sont indépendants les uns des autres et, pour chaque appel, la probabilité d'obtenir le correspondant demandé est de $p \in]0;1[$. On note $q = 1 - p$. X désigne la variable aléatoire égale au nombre de correspondants obtenus.

a) Quelle est la loi de X ? Donner $E(X)$ et $V(X)$.

b) Après ses n recherches, le secrétaire demande une deuxième fois, et dans les mêmes conditions chacun des $n-k$ correspondants qu'il n'a pas réussi à joindre la première fois. Soit Y la variable aléatoire égale au nombre de correspondants obtenus dans cette deuxième série d'appels. On note $Z = X + Y$ le nombre de correspondants obtenus.

i) Pour $k \in [\![0;n]\!]$ et $l \in [\![0;n-k]\!]$, calculer la probabilité conditionnelle $P(Y = l | X = k)$.

ii) Déterminez la loi de Z.

♦ **Exercice** 4.14 [**] **Un examen important**

Un étudiant s'apprête à passer un examen écrit très important et est préoccupé par la possibilité de ne pas être en forme le jour J à cause du stress. Il estime que, dans sa forme normale, il a une probabilité de 0.8 de réussir un exercice donné alors que, s'il n'est pas en forme, cette probabilité chute à 0.4. De plus, il estime qu'il est deux fois plus probable qu'il ne soit pas en forme le jour de l'examen. Sachant que pour valider l'examen l'étudiant doit réussir plus de la moitié des exercices, vaut-il mieux pour l'étudiant avoir un examen avec 3 ou 5 exercices ?

Exercice 4.15 [*]

Un magasin est en rupture de stock d'un produit A, 3 semaines en moyenne, par un an. Et 2 semaines en moyenne d'un produit B. Les approvisionnements se font toutes les semaines et sont indépendants. Appelons :

- ⋆ X le nombre de ruptures de stock du produit A, sur un an,
- ⋆ Y le nombre de ruptures de stock du produit B, sur un an,
- ⋆ Z le nombre de ruptures de stock d'un produit A ou B, sur un an.

On suppose que la variable aléatoire X suit une loi de Poisson de paramètre 3 et Y suit une loi de Poisson de paramètre 2.

a) Quelle est la probabilité que le nombre de ruptures de stock du produit A soit strictement supérieur à 1 ?

b) Quelle est la probabilité que, sur un an, il n'y ait aucune rupture de stock ?

c) Quelle est la probabilité qu'il y ait au moins 2 ruptures de stock ?

♦ **Exercice** 4.16 [**] **Pannes des Vélib'.**

Les vélos en libre service Vélib' ont principalement trois types de pannes : crevaison des pneus, déraillement de la chaîne et défaillance des freins. Durant la première année d'utilisation d'un Vélib', on désigne par

$$\begin{aligned} W = &\text{ nombre de pannes dues aux pneus} \\ X = &\text{ nombre de pannes dues à la chaîne} \\ Y = &\text{ nombre de pannes dues aux freins} \\ Z = &\text{ nombre de pannes quelconques} \end{aligned}$$

On suppose que W, X et Y sont des variables indépendantes suivant des lois de Poisson de paramètres respectifs $\lambda_W = 2, \lambda_X = 3, \lambda_Y = 1$.

a) Calculer la probabilité qu'il y ait au moins une crevaison durant la première année.

b) Calculer la probabilité qu'il n'y ait aucune panne due à la chaîne durant la première année.

c) Calculer la probabilité qu'il y ait deux pannes dues aux pneus durant la première année.

d) Calculer le nombre moyen de pannes durant la première année.

e) Calculer la probabilité qu'il y ait au plus deux pannes pendant la première année.

Exercice 4.17 [**] *(extrait du DE 2019-2020)*

Un serrurier (à jeun) tente d'ouvrir une porte à l'aide d'un trousseau de 10 clefs différentes, dont une seule ouvre la porte. On suppose qu'il ne re-essaie pas deux fois la même clef, et soit X la variable aléatoire représentant le nombre d'essais nécessaires pour ouvrir la porte.

a) Donner la loi de probabilité de X. Comment s'appelle-t-elle ?

b) Calculer l'espérance de X.
(On rappelle que $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$.)

c) Calculer la variance de X.
(On rappelle que $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$.)

Le serrurier est maintenant ivre et oublie immédiatement quelle clef il vient d'essayer avant d'en tester une nouvelle. Soit Y la variable aléatoire représentant le nombre d'essais nécessaires pour ouvrir la porte.

d) Donner la probabilité qu'il ouvre la porte au premier essai.

e) Donner la probabilité qu'il ouvre la porte au cinquième essai.

f) Donner la loi de probabilité de Y. Comment s'appelle-t-elle ?

g) Quelle est la probabilité qu'il fasse au moins 100 essais ?

h) Quelle est la probabilité qu'il n'ouvre jamais la porte ?

Exercice 4.18 [***] **Le problème des allumettes de Banach**

Un mathématicien fumant la pipe porte toujours deux boîtes d'allumettes, une dans la poche gauche de son manteau et une dans la poche droite. Chaque fois qu'il allume sa pipe, il pioche au hasard dans une des deux boîtes. Considérez le moment où ce mathématicien découvre pour la première fois qu'une des boîtes d'allumettes est vide. Sachant qu'initialement il y avait N allumettes dans chaque boîte, quelle est la probabilité que la boîte non vide contienne $0 < k \leq N$ allumettes ?

Exercice 4.19 [***] **Preuve de la proposition 4.16**

Soit $X \sim \mathscr{BN}(r,p)$ avec $r \in \mathbb{N}^*$ et $p \in]0,1[$.

a) Calculer $E(X)$ et $V(X)$ dans le cas $r = 1$.

b) Pour $r \neq 1$, expliquer pourquoi l'on peut écrire $X = X_1 + \ldots + X_r$ avec les $X_i \sim \mathscr{BN}(1,p)$ des variables indépendantes.

c) Conclure.

Exercice 4.20 [***] **Preuve du théorème 4.11.**

Soient $X,Y \in \ell_2(\Omega)$ deux variables aléatoires discrètes.

a) Utilisez le fait que $V\left(\frac{X}{\sigma(X)} + \frac{Y}{\sigma(Y}\right) \geq 0$ pour en déduire que $\rho(X,Y) \geq -1$.

b) Considérez maintenant $V\left(\frac{X}{\sigma(X)} - \frac{Y}{\sigma(Y}\right) \geq 0$ pour en déduire que $\rho(X,Y) \leq 1$.

c) Enfin, prouvez que $Y = aX + b \implies |\rho(X,Y)| = 1$ (où $a,b \in \mathbb{R}$ avec $a \neq 0$).

Exercice 4.21 [***] **Preuve de la proposition 4.19**

Prouver que si $X \sim \mathscr{P}(\lambda)$, $Y \sim \mathscr{P}(\theta)$ et X et Y sont indépendantes, alors $X + Y \sim \mathscr{P}(\lambda + \theta)$.

5. Variables aléatoires continues

Jusqu'à présent, nous avons toujours considéré des variables aléatoires ayant un nombre fini ou dénombrable de valeurs possibles. Mais, souvent, une situation est modélisée par une variable aléatoire qui peut prendre *n'importe quelle valeur* dans un intervalle : le temps de désintégration d'un atome, la taille d'une goutte d'eau qui tombe du robinet, la distance à laquelle se trouve la galaxie la plus proche de la Terre dans une direction donnée, la direction dans laquelle pointe une toupie lorsqu'elle est enfin à l'arrêt après avoir tourné, etc.

Lorsque c'est le cas, la situation change radicalement. Mathématiquement, on doit abandonner les sommes et les séries (monde du discret) et utiliser des intégrales (monde du continu). Conceptuellement, on doit abandonner la notion de probabilité en un point (puisque celle-ci est nulle) et introduire une "densité de probabilité" (tout à fait analogue à la densité de charge ou densité de masse en physique).

5.1 Définitions de base

Définition 5.1 (Variable aléatoire continue, densité de probabilité)
Une variable aléatoire X est *continue* s'il existe une fonction $f : \mathbb{R} \longrightarrow \mathbb{R}$ telle que :

- $\star$ $f \geq 0$
- $\star$ $\forall B \subset \mathbb{R}, P(X \in B) = \displaystyle\int_B f(x)dx.$

La fonction f est appelée *fonction de densité de probabilité* associée à la variable aléatoire X.
L'ensemble des variables aléatoires continues sur Ω est noté $\mathscr{L}(\Omega)$.

Remarque : Une densité de probabilité n'est pas une probabilité ! En particulier, rien n'interdit à la fonction f de prendre des valeurs supérieures à 1. En effet, ce ne sont pas les valeurs de f qu'il faut interpréter mais plutôt l'aire sous la courbe.

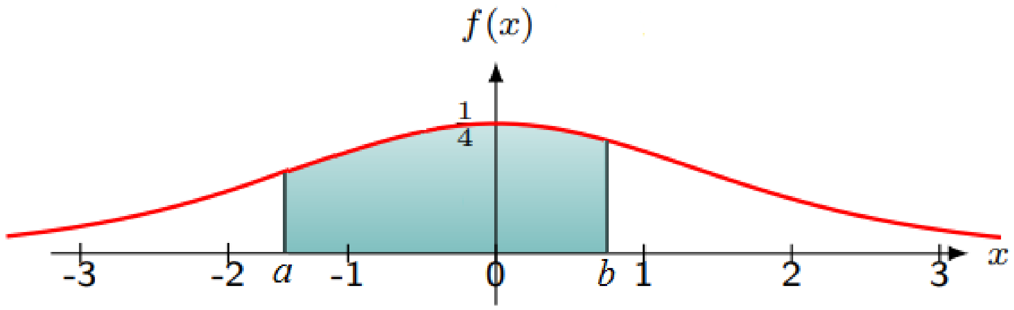


FIGURE 5.1 – La fonction f définie par $\forall x \in \mathbb{R}, f(x) = \frac{e^{-x}}{(1+e^{-x})^2}$ est une densité de probabilité. La probabilité $P(a \leq X \leq b)$ est l'aire sous la courbe sur l'intervalle $[a,b]$.

Proposition 5.1

Soit X une variable aléatoire continue et notons f sa densité de probabilité. Alors :

- ⋆ $P(a \leq X \leq b) = \displaystyle\int_a^b f(x)dx$.
- ⋆ $\displaystyle\int_{-\infty}^{+\infty} f(x)dx = P(X \in \mathbb{R}) = 1$.

Exemples :

- ⋆ La durée de vie d'un ordinateur, en heures d'utilisation, peut être modélisée comme une variable aléatoire continue T dont la densité de probabilité est

$$f(t) = \begin{cases} \frac{1}{1000}e^{-t/1000} \text{ si } t \geq 0 \\ 0 \text{ sinon.} \end{cases}$$

On peut vérifier en effet que $\displaystyle\int_{-\infty}^{+\infty} f(t)dt = 1$. Si l'on veut calculer la probabilité qu'un ordinateur fonctionne encore après 2000 heures d'utilisation, on calcule

$$P(T \geq 2000) = \frac{1}{1000}\int_{2000}^{+\infty} e^{-t/1000}dt = e^{-2} \approx 0.14.$$

- ⋆ Dans le cas d'une désintégration alpha, un atome X émet une particule d'hélium (appelée particule α) et se transforme donc dans un atome Y plus léger :

$${}^{\mathrm{A}}_{\mathrm{Z}}\mathrm{X} \longrightarrow {}^{\mathrm{A}-4}_{\mathrm{Z}-2}\mathrm{Y} + {}^{4}_{2}\mathrm{He}.$$

Dans le référentiel de la particule X, on ne peut pas savoir à l'avance dans quelle direction sera émise la particule α mais toutes les directions sont équiprobables. A deux dimension, on modélise la probabilité d'émission de la particule d'hélium entre 0 et θ par une variable aléatoire Θ dont la densité de probabilité est

$$f(\theta) = \begin{cases} \frac{1}{2\pi} \text{ si } \theta \in [0, 2\pi[\\ 0 \text{ sinon.} \end{cases}$$

Proposition 5.2

Soit $a \in \mathbb{R}$ et X une variable aléatoire continue. Alors, on a $P(X = a) = 0$.

Démonstration. Puisque X est continue, on a

$$P(X = a) = \int_a^a f(x)dx = 0.$$

□

Remarque : Conséquence non triviale de ce résultat qui semble anodin : pour une variable continue, $X(\Omega)$ doit être indénombrable (sinon on aurait $\sum_{x \in X(\Omega)} P(X = x) = \sum_{x \in X(\Omega)} 0 = 0$, ce qui est absurde) et donc l'univers Ω lui-même est indénombrable ! Mais nous ferons l'autruche, ignorant les subtilités des univers indénombrables (voir section 1.4), car en pratique nous considérerons toujours des sous-ensembles $B \subset \mathbb{R}$ qui sont des intervalles ou réunions d'intervalles.

Définition 5.2 (Fonction de répartition)
On appelle *fonction de répartition* de la variable aléatoire X, que l'on note $F_X(x)$ (ou $F(x)$ si la variable aléatoire est évidente), la fonction définie pour tout réel $x \in \mathbb{R}$ par

$$F(x) = P(X \leq x).$$

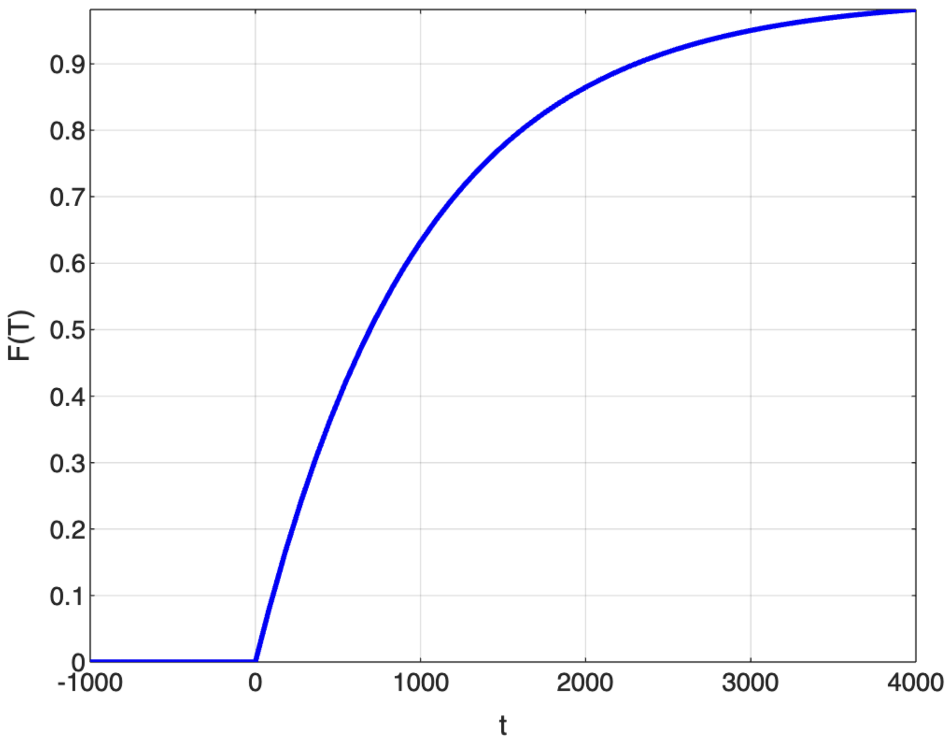
♦Exemple : En suivant l'exemple de la durée de vie d'un ordinateur modélisé par une variable aléatoire T dont la densité de probabilité est

$$f(t) = \begin{cases} \frac{1}{1000} e^{-t/1000} \text{ si } t \leq 0 \\ 0 \text{ sinon,} \end{cases}$$

on a

$$F(t) = \begin{cases} 0 \text{ si } t \leq 0 \\ 1 - e^{-t/1000} \text{ si } t \geq 0. \end{cases}$$

Ici, la fonction de répartition F indique la probabilité qu'à l'instant t l'ordinateur ne marche plus. Graphiquement, cela donne :



Remarque : C'est la même définition que pour les variables aléatoires discrètes. Cela dit, comme on le verra, la fonction de répartition est un objet crucial dans les calculs des variables aléatoires continues alors que F_X n'est pas vraiment utile dans le cas discret.

Théorème 5.3 : Lien entre la fonction de répartition et la densité de probabilité

Soit X une variable aléatoire continue et notons f sa densité de probabilité et F sa fonction de répartition. Alors,

- ⋆ $F(x) = \int_{-\infty}^{x} f(t)dt$.
- ⋆ f est la dérivée de F.
- ⋆ F est l'unique primitive de f vérifiant $\lim_{x \to +\infty} F(x) = 1$.

Démonstration. Le théorème découle directement de la définition. □

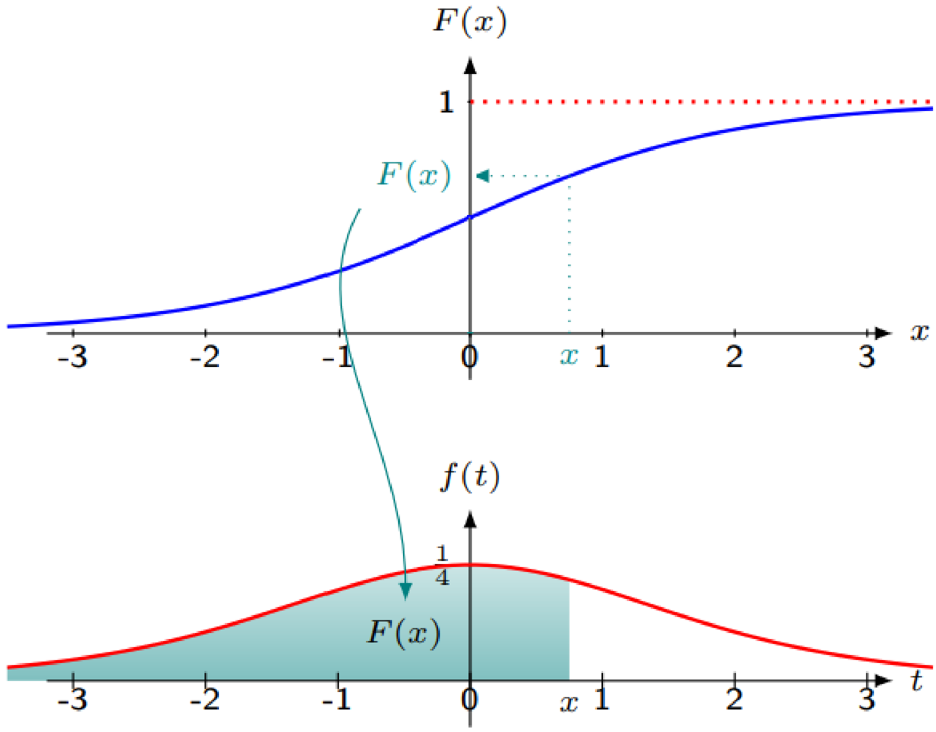


FIGURE 5.2 – Lien entre densité de probabilité et fonction de répartition. Le graphe du bas montre le graphe de la densité de probabilité (en rouge). L'aire sous la courbe correspond à la valeur de la fonction de répartition (graphe du haut).

Proposition 5.4 : Propriétés de la fonction de répartition

Soit X une variable aléatoire continue et notons F sa fonction de répartition. Alors :

- ⋆ F est continue.
- ⋆ F est croissante.
- ⋆ $\lim_{x \to -\infty} F(x) = 0$.
- ⋆ $\lim_{x \to +\infty} F(x) = 1$.
- ⋆ $\forall (a,b) \in \mathbb{R}^2$, $P(a \leq X \leq b) = F(b) - F(a)$.

Démonstration. Les propriétés découlent directement de la définition. □

Tout ce qui précède montre que tous les calculs de probabilités liés à X sont complètement déterminés dès que l'on connaît la fonction de répartition F ou la densité de probabilité f.

Fiche méthode #6 (Loi de probabilité de densité de $Y = g(X)$ dans le cas continu)
Soit X une variable aléatoire continue de support $X(\Omega)$ et dont la loi de densité de probabilité ou la fonction de répartition est connue. Soit g une fonction réelle définie sur $X(\Omega)$ et $Y = g(X)$ une nouvelle variable aléatoire. Pour trouver la densité de probabilité de Y :

a) Déterminer le support $Y(\Omega)$.

b) Exprimer l'évènement $\{Y \leq a\}$ en fonction de la variable aléatoire X.

c) En déduire la fonction de répartition de Y.

d) En déduire la densité de probabilité de Y.

♦Exemple : Soit X une variable aléatoire continue dont la densité de probabilité est

$$f_X(x) = \begin{cases} 1 - |x| \text{ si } -1 \leq x \leq 1 \\ 0 \text{ sinon.} \end{cases}$$

On cherche à trouver la densité de probabilité de $Y = \sqrt{|X|}$.

a) D'abord, puisque $X(\Omega) = [-1,1]$, on a $Y(\Omega) = [0,1]$.

b) Ensuite, soit $a \in \mathbb{R}$. Pour exprimer l'évènement $\{Y \leq a\}$, on distingue trois cas :
- $\star$ Si $a < 0$, l'évènement $\{Y \leq a\}$ est l'évènement impossible.
- $\star$ Si $a > 1$, l'évènement $\{Y \leq a\}$ est l'évènement certain.
- $\star$ Si $a \in [0,1]$, on a $Y \leq a \iff \sqrt{|X|} \leq a \iff |X| \leq a^2 \iff -a^2 \leq X \leq a^2$.

c) On a donc, pour la fonction de répartition de Y :
- $\star$ Si $a < 0$, $F_Y(a) = P(Y \leq a) = 0$.
- $\star$ Si $a > 0$, $F_Y(a) = P(Y \leq a) = 1$.
- $\star$ Si $a \in [0,1]$, on a

$$\begin{aligned} F_Y(a) &= P(Y \leq a) = P(-a^2 \leq X \leq a^2) \\ &= \int_{-a^2}^{a^2} (1 - |x|)dx \\ &= 2\int_0^{a^2} (1-x)dx \\ &= 2a^2 - a^4. \end{aligned}$$

(On peut vérifier que l'on a bien $F(0) = 0$ et $F(1) = 1$.)

d) Enfin, en dérivant F_Y, on obtient la densité de probabilité de Y :

$$f_Y(y) = \begin{cases} 4y(1-y^2) \text{ si } y \in [0,1] \\ 0 \text{ sinon.} \end{cases}$$

(On peut vérifier que l'on a bien $\int_{-\infty}^{+\infty} f_Y(y)dy = 1$.)

5.2 Couples de variables aléatoires continues

Dans cette section, on regarde comment adapter les définitions de la section 4.2 au cas de variables continues.

Définition 5.3 (Fonction de répartition conjointe)
Soient X, Y deux variables aléatoires (continues ou discrètes). La *fonction de répartition conjointe de X et Y* est la fonction $F_{X,Y} : \mathbb{R}^2 \longrightarrow [0,1]$ définie par

$$F_{X,Y}(a,b) = P(X \leq a, Y \leq b).$$

Proposition 5.5

Soient X, Y deux variables aléatoires (continues ou discrètes). Alors, si l'on connaît la fonction de répartition conjointe $F_{X,Y}$, on connaît aussi les fonctions de répartition de X et de Y. (La réciproque est fausse !)

Démonstration. En effet, pour $a \in \mathbb{R}$, on a

$$F_X(a) = P(X \leq a) = P(X \leq a, Y \leq +\infty) = \lim_{b \to +\infty} F_{X,Y}(a,b).$$

□

Remarque : Ainsi, tous les calculs de probabilités sur X et Y peuvent être déterminés à partir de la fonction de répartition conjointe.

♦Exemple : Si l'on cherche à calculer $P(X \geq a, Y \geq b)$ avec $a, b \in \mathbb{R}$, on a

$$\begin{aligned}
P(X \geq a, Y \geq b) &= 1 - P(\overline{\overline{\{X \geq a\}} \cap \overline{\{Y \geq b\}}}) \\
&= 1 - P(\overline{\{X \geq a\}} \cup \overline{\{Y \geq b\}}) \\
&= 1 - P(\overline{\{X \geq a\}}) - P(\overline{\{Y \geq b\}}) + P(\overline{\{X \geq a\}} \cap \overline{\{Y \geq b\}}) \\
&= 1 - P(X < a) - P(Y < b) + P(X < a, Y < b) \\
&= 1 - F_X(a) - F_Y(b) + F_{X,Y}(a,b).
\end{aligned}$$

Définition 5.4 (Couple de variables aléatoires conjointement continues)
Deux variables aléatoires X et Y sont *conjointement continues* s'il existe une fonction $f_{X,Y} : \mathbb{R}^2 \longrightarrow \mathbb{R}$ telle que

- $\star$ $f_{X,Y} \geq 0$
- $\star$ $\forall C \subset \mathbb{R}^2, P\big((X,Y) \in C\big) = \iint\limits_{C} f_{X,Y}(x,y)dxdy.$

$f_{X,Y}$ est appelée la densité de probabilité conjointe de X et Y.

♦Exemple : Un enfant fait tomber un ballon balle dans un mare parfaitement circulaire de rayon R. Ce ballon flotte sur la surface de la mare et bouge au gré des aléas météorologiques (pluie, vent). L'enfant revient le jour d'après pour voir s'il peut récupérer son ballon. Sachant que l'enfant pourra attraper le ballon s'il est à une distance inférieure à $r < R$ du bord de la mare et supposant qu'à ce moment là toutes les positions sur la surface de la mare sont équiprobables, calculer la probabilité que l'enfant puisse récupérer son ballon.

On note (x,y) la position du ballon sur la mare et X et Y les variables aléatoires associées.

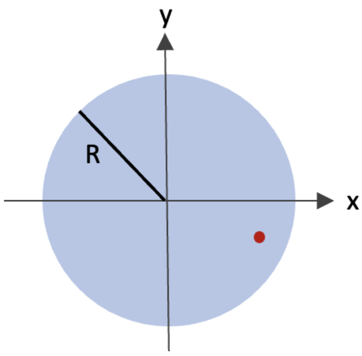


FIGURE 5.3 – Un ballon rouge flotte sur une mare circulaire de rayon R et peut se trouver, après un certain temps, n'importe où sur cette mare.

X et Y sont conjointement continues et l'on a

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi R^2} \text{ si } x^2 + y^2 \leq R^2 \\ 0 \text{ sinon.} \end{cases}$$

L'évènement E ="l'enfant récupère le ballon" correspond au cas où la distance du ballon au centre de la mare est supérieure à $R-r$. Ainsi, la probabilité que l'on cherche à calculer est

$$\begin{aligned} P(\sqrt{x^2+y^2} \geq R-r) &= 1 - P(x^2+y^2 \leq (R-r)^2) \\ &= 1 - \iint\limits_{x^2+y^2\leq(R-r)^2} f_{X,Y}(x,y)dxdy \\ &= 1 - \frac{1}{\pi R^2} \iint\limits_{x^2+y^2\leq(R-r)^2} dxdy \\ &= 1 - \frac{(R-r)^2}{R^2}. \end{aligned}$$

Proposition 5.6

Soient X et Y deux variables aléatoires conjointement continues. Alors X et Y sont toutes les deux des variables aléatoires continues (la réciproque est fausse !).
De plus, les densité de probabilité sont données par

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy \text{ et } f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx.$$

Démonstration. Soit $A \subset \mathbb{R}$. Par définition, la densité de probabilité de X est la fonction telle que $P(X \in A) = \int\limits_A f_X(x)dx$. Or on a

$$P(X \in A) = P(X \in A, Y \in \mathbb{R}) = \int\limits_A \left(\int\limits_{-\infty}^{+\infty} f_{X,Y}(x,y)dy \right) dx.$$

□

♦Exemple : En continuant avec le même exemple, la densité de probabilité de la variable aléatoire X (position sur l'axe X du ballon sur la mare), on a, pour $-R \leq x \leq R$:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy = \frac{1}{\pi R^2} \int\limits_{-\sqrt{R^2-x^2}}^{+\sqrt{R^2-x^2}} dy = \frac{2\sqrt{R^2-x^2}}{\pi R^2}.$$

Définition 5.5 (Couple de variables aléatoires continues indépendantes)
Soient $X, Y \in \mathscr{L}(\Omega)$ deux variables aléatoires conjointement continues. On dit que X et Y sont *indépendantes* si

$$\forall (x,y) \in \mathbb{R}^2, F_{X,Y}(x,y) = F_X(x)F_Y(y).$$

Ceci est équivalent à dire que la densité de probabilité conjointe est le produit des densités de probabilité :

$$\forall (x,y) \in \mathbb{R}^2, f_{X,Y}(x,y) = f_X(x)f_Y(y).$$

♦Exemple : En continuant avec le même exemple, par symétrie on a $f_X = f_Y$ et on voit bien que $f_{X,Y}(x,y) \neq f_X(x)f_Y(y)$ dont la position du ballon selon un axe n'est pas indépendante de la position selon l'autre axe (ce qui est normal puisque la surface est limitée à un cercle).

5.3 Espérance, Variance, Ecart-type

Définition 5.6 (Espérance d'une variable aléatoire continue)
Soit X une variable aléatoire continue. On appelle *espérance de* X, si elle existe, la quantité

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx.$$

L'ensemble des variables aléatoires continues sur Ω pour lesquelles l'espérance existe est noté $\mathscr{L}_1(\Omega)$.

♦Exemples :

⋆ Si l'on reprend l'exemple de la durée de vie d'un ordinateur (pages 102 et 103), on a

$$E(T) = \frac{1}{1000}\int_0^{+\infty} t e^{-t/1000} dt = \int_0^{+\infty} e^{-t/1000} dt = 1000.$$

⋆ Voici l'exemple le plus connu de variable aléatoire continue dont l'espérance n'existe pas : on dit que X suit une *loi de Cauchy standard*[1] si sa densité de probabilité est donnée par

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

1. Malgré ce que le nom suggère, cette loi a été d'abord été étudiée par Siméon Denis Poisson en 1824.

C'est bien une densité de probabilité puisque

$$\int_{-\infty}^{+\infty} f(x)dx = \frac{1}{\pi}[arctan(x)]_{-\infty}^{+\infty} = 1.$$

D'autre part, l'espérance de X est donnée par

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{+\infty} \frac{x}{\pi(1+x^2)}dx$$

et l'intégrale diverge puisque l'on a $xf(x) \underset{+\infty}{\sim} \frac{1}{\pi x}$.

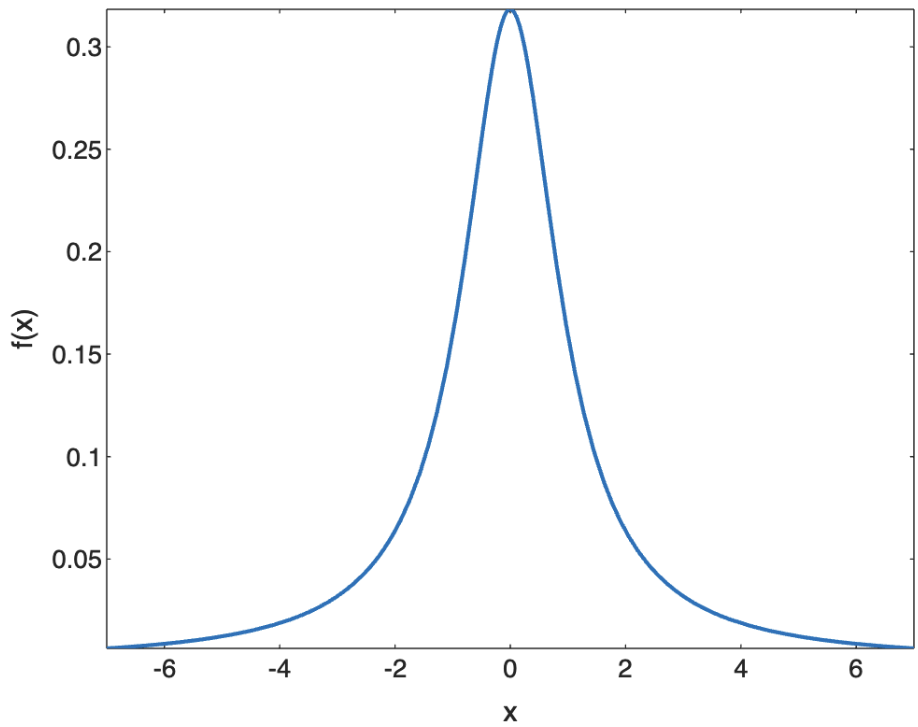


FIGURE 5.4 – Graphe de la densité de probabilité de la loi de Cauchy standard.

Proposition 5.7

L'espérance est une application linéaire sur l'espace vectoriel $\mathscr{L}_1(\Omega)$.
Autrement dit, soient X et Y deux variables aléatoires continues sur Ω pour lesquelles l'espérance existe, et soient $\alpha, \beta \in \mathbb{R}$. Alors, on a

$$E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y).$$

Théorème 5.8 : Théorème du transfert (version continue)

Soit $X \in \mathscr{L}_1(\Omega)$ une variable aléatoire continue et notons f_X la densité de probabilité de X. Soit g une fonction réelle définie sur sur $X(\Omega)$. Alors, si l'espérance de $Y = g(X)$ existe, on a

$$E(g(X)) = \int_{X(\Omega)} g(x) f_X(x) dx.$$

Démonstration. Admis. □

♦Exemple : On tire un nombre au hasard dans l'intervalle $[0,2]$ et on calcule ensuite le carré de ce nombre. Pour trouver l'espérance de cette expérience, on note X le nombre tiré au hasard dans l'intervalle $[0,2]$. Cette variable aléatoire admet comme densité de probabilité

$$f_X(x) = \begin{cases} \frac{1}{2} \text{ si } x \in [0,2] \\ 0 \text{ sinon.} \end{cases}$$

Pour calculer $E(X^2)$, on utilise directement le théorème du transfert :

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \frac{1}{2}\int_0^2 x^2 dx = \frac{4}{3}.$$

Remarque : Dans l'exemple précédent, on peut aussi calculer $E(X^2)$ en utilisant la fiche méthode page 106 pour trouver la densité de probabilité de $Y = X^2$ pour ensuite calculer $E(Y)$:

- ⋆ Puisque $X(\Omega) = [0,2]$, on a $Y(\Omega) = [0,4]$.
- ⋆ Puisque X et Y ne prennent que des valeurs positives, on a, pour $a \in [0,4]$, $\{Y \leq a\} = \{X \leq \sqrt{a}\}$.
- ⋆ Ainsi
$$F_Y(a) = P(Y \leq a) = P(X \leq \sqrt{a}) = \frac{1}{2}\int_0^{\sqrt{a}} dx = \frac{\sqrt{a}}{2}.$$
- ⋆ Donc, on trouve que la densité de probabilité de Y vaut
$$f_Y(y) = \begin{cases} \frac{1}{4\sqrt{y}} \text{ si } y \in [0,4] \\ 0 \text{ sinon.} \end{cases}$$
(On voit au passage que, même si les valeurs de X sont équiprobables, celles de X^2 ne le sont pas.)
- ⋆ Enfin, pour l'espérance, on a
$$E(Y) = \int_0^4 \frac{\sqrt{y}}{4} dy = \frac{4}{3}.$$

On mesure ici à quel point le théorème du transfert nous simplifie les calculs !

Définition 5.7 (Variance et écart-type d'une variable aléatoire continue)
Soit X une variable aléatoire (continue ou discrète). On appelle *variance de X*, si elle existe, la quantité

$$V(X) = E\left((X - E(X))^2\right) = E(X^2) - E(X)^2.$$

De plus, on appelle *écart-type* de X la quantité $\sigma(X) = \sqrt{V(X)}$.
L'ensemble des variables aléatoires continues sur Ω pour lesquelles la variance existe est noté $\mathscr{L}_2(\Omega)$.

Remarque : La définition et l'interprétation de la variance sont exactement les mêmes ici que dans le cas des variables aléatoires discrètes.

♦Exemples :

- ⋆ En continuant l'exemple précédent d'un nombre tiré au hasard dans l'intervalle $[0,2]$. On a clairement $E(X) = 1$ et donc $V(X) = E(X^2) - E(X)^2 = \frac{1}{3}$ et $\sigma(X) = \frac{1}{\sqrt{3}} \approx 0.58$.
- ⋆ Pour l'exemple de la durée de vie d'un ordinateur (pages 68, 72 et 72), on a

$$E(T^2) = \frac{1}{1000}\int_0^{+\infty} t^2 e^{-t/1000} dt = 2\int_0^{+\infty} t e^{-t/1000} dt = 2\times 1000 \int_0^{+\infty} e^{-t/1000} dt = 2\times(1000)^2.$$

(où l'on a effectué deux intégrations par parties). Ainsi, on a donc

$$V(X) = E(X^2) - E(X)^2 = 2\times(1000)^2 - (1000)^2 = 1000^2$$

et

$$\sigma(X) = 1000.$$

Proposition 5.9

Soient $X \in \mathscr{L}_2(\Omega)$ une variable aléatoire continue et $a, b \in \mathbb{R}$. On a

$$V(aX + b) = a^2 V(X).$$

Démonstration. La preuve est exactement la même que celle de la proposition 4.5 (page 75). □

Définition 5.8 (Moment d'ordre n)
Soit X une variable aléatoire continue ou discrète et $n \in \mathbb{N}$. On appelle *moment d'ordre n*, si elle existe, la quantité

$$m_n(X) = E(X^n) = \sum_{x \in X(\Omega)} x^n P(X = x).$$

On appelle *moment centré d'ordre n*, si elle existe, la quantité

$$\mu_n(X) = E\left((X - E(X))^n\right).$$

Remarque : Aide mémoire

Les formules usuelles "discret/continu" se retrouvent avec la transformation suivante :

$$\begin{aligned} \text{Discret} &\longleftrightarrow \text{Continu} \\ x_i &\longleftrightarrow x \\ P(X = x_i) &\longleftrightarrow f(x)dx \\ \sum &\longleftrightarrow \int \end{aligned}$$

	Discret	Continu
Espérance : $E(X)$	$\sum x_i P(X = x_i)$	$\int_{-\infty}^{+\infty} x f(x)dx$
Moment d'ordre 2 : $E(X^2)$	$\sum x_i^2 P(X = x_i)$	$\int x^2 f(x)dx$
Moment d'ordre k : $E(X^k)$	$\sum x_i^k P(X = x_i)$	$\int x^k f(x)dx$
Fonction de répartition : $F(t) = P(X \leq t)$	$\sum_{x_i \leq t} P(X = x_i)$	$\int_{x \leq t} f(x)dx$
$P(a \leq X \leq b)$	$\sum_{a \leq x_i \leq b} P(X = x_i)$	$\int_{a \leq x \leq b} f(x)dx$

Évidemment, cela reste un aide mémoire. Et bien des choses fonctionnent différemment, comme par exemple, la probabilité en une valeur donnée dans le cadre continu qui est toujours nulle, ce qui n'est pas vrai dans le cadre discret.

5.4 Lois usuelles

On va maintenant regarder certaines lois de probabilités de référence que l'on rencontre souvent.

5.4.1 Loi uniforme

Définition 5.9 (Loi uniforme)
Soient X une variable aléatoire continue et $a < b$ deux réels. On dit que X suit une *loi uniforme sur* $[a,b]$, noté $X \sim \mathscr{U}([a,b])$, si sa densité de probabilité est constante sur $[a,b]$ et nulle ailleurs :

$$f(x) = \begin{cases} \dfrac{1}{b-a} \text{ si } x \in [a,b] \\ 0 \text{ sinon.} \end{cases}$$

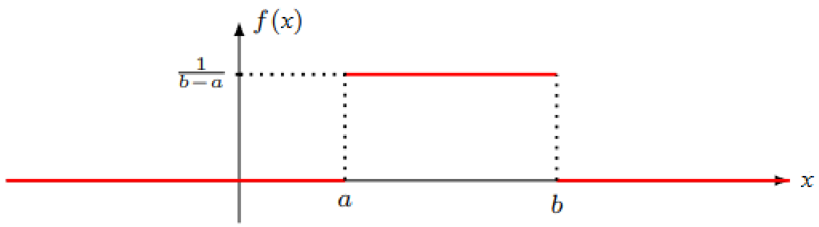


FIGURE 5.5 – Densité de probabilité uniforme $\mathscr{U}([a,b])$

♦<u>Exemples :</u>

⋆ Relire l'exemple de la désintégration alpha (page 102).

⋆ Supposons que le métro de la ligne 8 passe par la station Place d'Italie exactement toutes les six minutes entre 6h et 8h du matin les jours ouvrés. Un étudiant de l'EFREI qui se rend au campus de République arrive à la station de Place d'Italie à un moment donné uniformément distribué entre 7h et 7h30 du matin. Calculer la probabilité que l'étudiant attende moins de deux minutes pour prendre le métro à Place d'Italie.
Soit T le moment d'arrivée de l'étudiant à Place d'Italie. On a $T \sim \mathscr{U}(0,30)$ (où l'on a pris 7h comme temps 0). Soit E l'évènement "l'étudiant attend moins de 2 minutes". On a

$$E = \{4 \leq X \leq 6\} \cup \{10 \leq X \leq 12\} \cup \{16 \leq X \leq 18\} \cup \{22 \leq X \leq 24\} \cup \{28 \leq X \leq 30\}$$

et donc, puisque les différents évènements sont incompatibles,

$$\begin{aligned} P(E) &= P(4 \leq X \leq 6) + P(10 \leq X \leq 12) + P(16 \leq X \leq 18) + P(22 \leq X \leq 24) + P(28 \leq X \leq 30) \\ &= \frac{1}{30}\left(\int_4^6 dt + \int_{10}^{12} dt + \int_{16}^{18} dt + \int_{22}^{24} dt + \int_{28}^{30} dt\right) \\ &= \frac{1}{3}. \end{aligned}$$

Proposition 5.10 : Espérance et variance de la loi uniforme

Si $X \sim \mathscr{U}([a,b])$, alors $E(X) = \dfrac{a+b}{2}$ et $V(X) = \dfrac{(b-a)^2}{12}$.

Démonstration. Si $X \sim ([a,b])$, on a

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \frac{1}{b-a}\int_a^b xdx = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}.$$

De même, par le théorème du transfert, on a

$$E(X^2) = \frac{1}{b-a}\int_a^b x^2dx = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}.$$

Et donc

$$V(X) = E(X^2) - E(X)^2 = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.$$

□

5.4.2 Loi exponentielle

Nous allons maintenant voir une loi importante qui modélise les phénomènes sans mémoire. Par exemple, la durée de vie d'un composant électronique obéit à une loi exponentielle.

Définition 5.10 (Loi exponentielle)
Soient X une variable aléatoire continue et λ un réel strictement positif. On dit que X suit une *loi exponentielle de paramètre* λ, noté $X \sim \mathscr{E}(\lambda)$, si sa densité de probabilité est donnée par :

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

Exemples :

- ⋆ Relire l'exemple sur le temps de vie d'un ordinateur (page 102).
- ⋆ Un exemple plus classique : le temps de vie d'un atome radioactif.
- ⋆ Le temps d'attente dans une file, la durée d'un appel téléphonique ou la distance parcourue par une météorite dans l'atmosphère terrestre avant de se désintégrer complètement peuvent aussi être modélisés par des lois exponentielles.

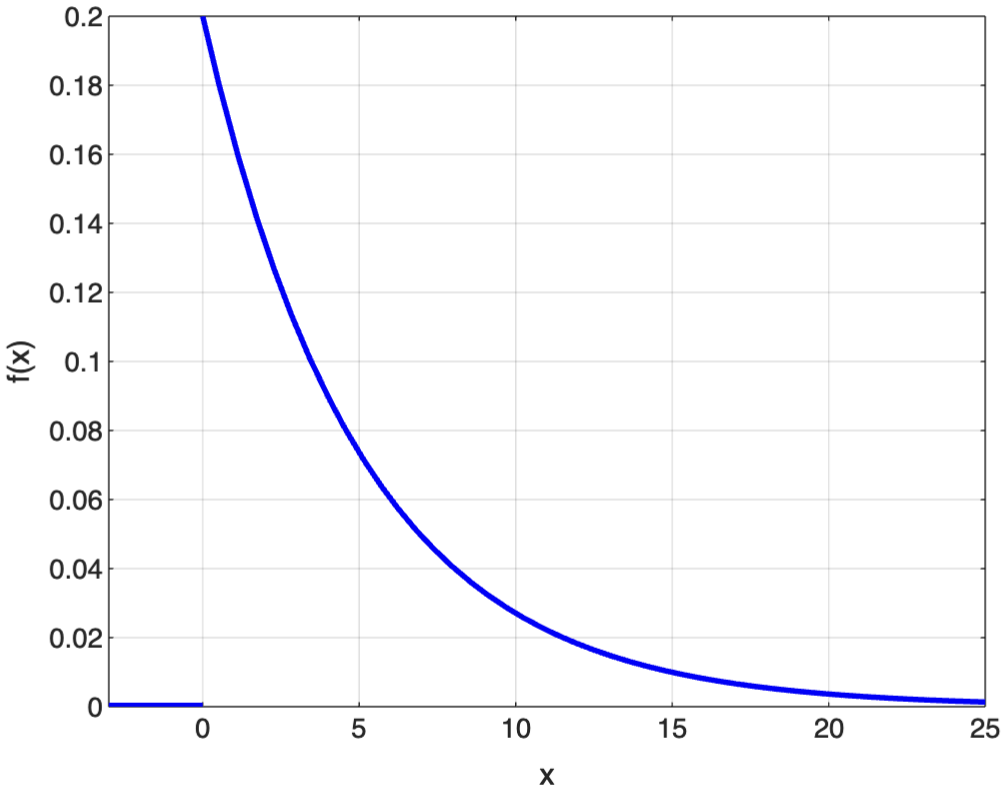


FIGURE 5.6 – Allure de la loi exponentielle de paramètre $\lambda = 1/5$.

Proposition 5.11 : Espérance et variance de la loi exponentielle

Si $X \sim \mathscr{E}(\lambda)$, alors $E(X) = \dfrac{1}{\lambda}$ et $V(X) = \dfrac{1}{\lambda^2}$.

Démonstration. En intégrant par parties, on a

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \lambda \int_0^{+\infty} xe^{-\lambda x}dx = \frac{1}{\lambda}.$$

De même, en intégrant par parties deux fois,

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx = \lambda \int_0^{+\infty} x^2 e^{-\lambda x}dx = \frac{2}{\lambda^2}$$

et donc

$$V(X) = E(X^2) - E(X)^2 = \frac{1}{\lambda^2}.$$

$\square$

La proposition suivante donne une propriété importante qui caractérise la loi exponentielle :

Proposition 5.12 : La loi exponentielle est sans mémoire

On dit qu'une variable aléatoire continue T est *sans mémoire* si elle vérifie

$$\forall s \geq 0, P(T \geq t+s \big| T \geq t) = P(T \geq s).$$

Alors, on a l'équivalence suivante :

$$T \text{ est sans mémoire} \iff T \text{ suit une loi exponentielle.}$$

Démonstration. Voir exercice 5.13. □

Remarque : Il faut comprendre cette propriété simplement : considérons, par exemple, la probabilité qu'un composant électronique fonctionne au moins s jours. Cette probabilité s'écrit $P(T \geq s)$. Le lendemain, si le composant marche encore, c'est à dire "sachant qu'il a tenu au moins un jour" $(T \geq 1)$, cette première probabilité sera exactement la même : $P(T \geq 1+s \big| T \geq 1) = P(T \geq s)$. En répétant le raisonnement, on se rend compte que, peu importe le nombre de jours écoulés, tous les jours, la probabilité que le composant tienne s jours de plus sera la même, jusqu'au moment où il ne fonctionnera plus.

Exemple : Pour un noyau radioactif, le fait qu'il ne se soit pas encore désintégré (et donc qu'il se soit passé un certain temps sans s'être désintégré) n'a aucune influence sur les probabilités que le noyau se désintègre dans le futur.

5.4.3 Loi normale

De toutes les lois usuelles, la loi normale ou loi Gaussienne est de loin la plus importante. Cette loi est omniprésente dans la modélisation des phénomènes physiques : on l'utilise pour décrire, par exemple, la taille des étoiles dans une galaxie, la pression artérielle du corps humain, la taille d'un être humain, l'état fondamental de l'oscillateur harmonique quantique.

Définition 5.11 (Loi normale ou Gaussienne)
Soit X une variable aléatoire continue, $(\mu \times \sigma) \in \mathbb{R} \times \mathbb{R}^{+*}$. On dit que X suit une *loi normale de paramètres μ et σ*, notée $X \sim \mathcal{N}(\mu, \sigma^2)$, si sa densité de probabilité est donnée par

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

En particulier, si $X \sim \mathcal{N}(0,1)$, on dit que X suit une *loi normale centrée réduite*.

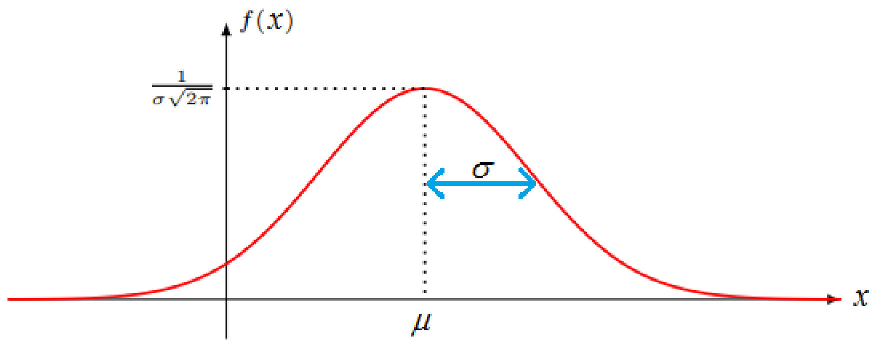
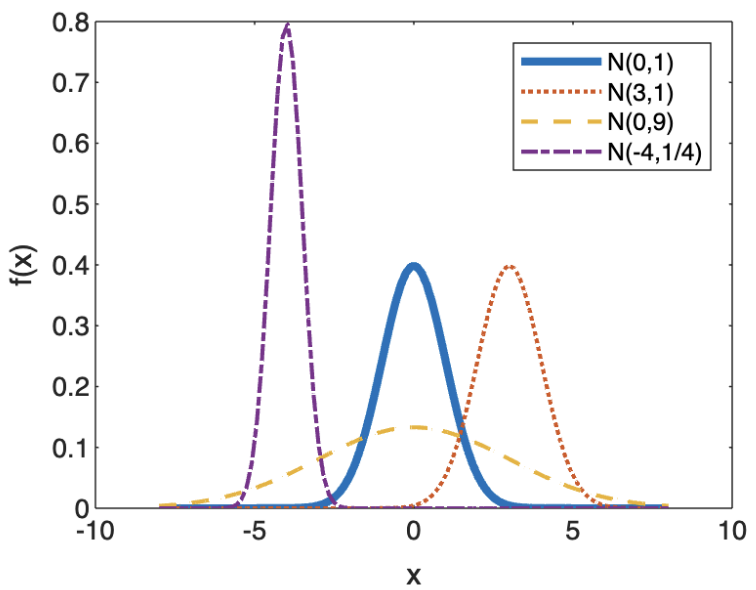


FIGURE 5.7 – Allure de la loi Gaussienne. Le paramètre μ fixe l'axe de symétrie de la loi. Et le paramètre σ représente la distance de cette position au point d'inflexion de la courbe.

Remarque : En partant de la loi normale centrée réduite, si l'on change la valeur de μ, la courbe est translatée vers la gauche ou la droite. Si l'on diminue σ, le pic est plus prononcé. Si l'on augmente σ, la courbe s'étale.



Script Matlab de la figure :

```
x=linspace(-8,8);
y1=normpdf(x,0,1); y2=normpdf(x,3,1);
y3=normpdf(x,0,3); y4=normpdf(x,-4,1/2);
plot(x,y1,'LineWidth',3)
hold on
plot(x,y2,":",'LineWidth',1.5)
plot(x,y3,"-",'LineWidth',1.5)
plot(x,y4,"-.",'LineWidth',1.5)
hold off
legend(["N(0,1)" "N(3,1)" "N(0,9)"
"N(-4,1/4)"])
xlabel('x')
ylabel('f(x)')
```

FIGURE 5.8 – Allure de la loi Gaussienne pour différentes valeurs des paramètres.

En pratique, tous les calculs de probabilités d'une loi normale sont réalisés en changeant de variable pour se ramener à une loi normale centrée réduite, puis en utilisant les valeurs de la fonction de répartition de la loi normale centrée réduite.

Définition 5.12 (Fonction de répartition de la loi normale centrée réduite)
On note Φ la fonction de répartition de la loi normale centrée réduite. Autrement dit, on a

$$\forall x \in \mathbb{R}, \Phi(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2} dx.$$

Proposition 5.13 : Propriétés de la fonction de répartition Φ

Soient $X \sim \mathcal{N}(0,1)$, Φ la fonction de répartition de X et $a \leq b \in \mathbb{R}$. On a

- $\star$ $P(a \leq X \leq b) = \Phi(b) - \Phi(a)$,
- $\star$ $P(X \geq a) = P(X \leq -a) = 1 - \Phi(a)$,
- $\star$ $\Phi(-a) = 1 - \Phi(a)$.

Démonstration. Le mieux est d'avoir en tête les dessins suivants :

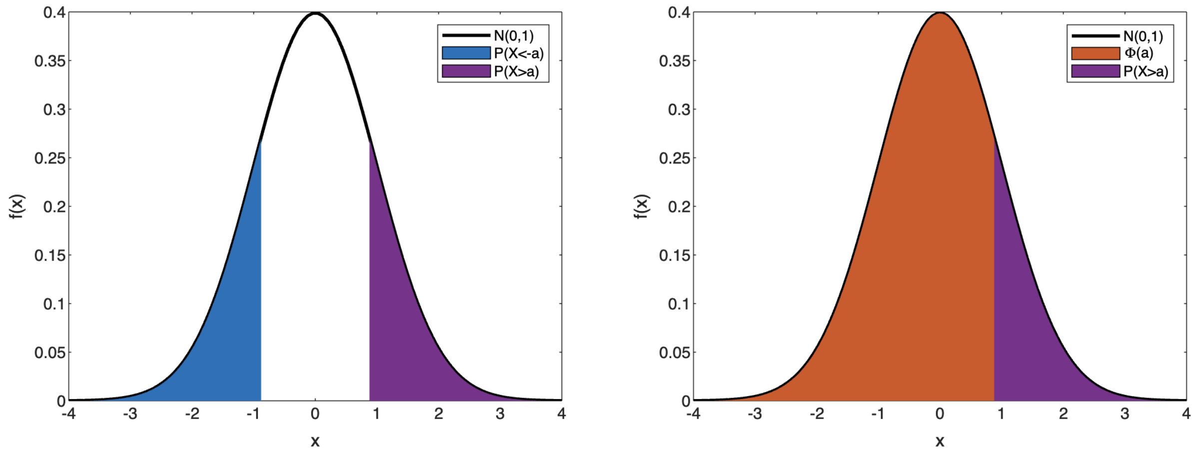


FIGURE 5.9 – Les égalités $P(X \geq a) = P(X \leq -a)$ (à gauche) et $P(X \geq a) + \Phi(a) = 1$ (à droite).
(Les graphiques ont été générés avec $a = 0.88$.)

$\square$

Fiche méthode #7 (Calcul des probabilités d'une variable suivant une loi normale)
Soient $X \sim \mathcal{N}(\mu, \sigma^2)$ et I_1, I_2 deux intervalles de $\mathbb{R}$. Pour calculer $P(X \in I_1 \big| X \in I_2)$, on suit les étapes suivantes :

a) Poser $Z = \frac{X-\mu}{\sigma}$. Ainsi, Z suit une loi normale centrée réduite : $Z \sim \mathcal{N}(0,1)$.

b) Exprimer les évènements $\{X \in I_1\}$ et $\{X \in I_2\}$ avec la nouvelle variable Z.

c) Utiliser les propriétés de la proposition 5.13 pour exprimer la probabilité $P(X \in I_1 \big| X \in I_2) = P(Z \in I_3 \big| Z \in I_4)$ en termes de la fonction Φ.

d) Utiliser la table des valeurs de la fonction Φ (voir page suivante) pour trouver une valeur approchée de $P(X \in I_1 \big| X \in I_2)$.

♦Exemple : Soit $X \sim \mathcal{N}(6,4)$. Calculer $P(|X-4| < 3 \big| X \geq 2)$.

a) On commence par poser $Z = \frac{X-6}{2}$. Z suit alors une loi normale centrée réduite.

b) De plus, on a

$$|X-4| < 3 \iff -3 < X-4 < 3 \iff -\frac{5}{2} < Z < \frac{1}{2}$$

et aussi

$$X \geq 2 \iff Z \geq -2.$$

c) Ainsi, on a

$$\begin{aligned} P(|X-4| < 3 \big| X \geq 2) &= P(-\frac{5}{2} < Z < \frac{1}{2} \big| Z \geq -2) \\ &= \frac{P(\{-\frac{5}{2} < Z < \frac{1}{2}\} \cap \{Z \geq -2\}}{P(Z \geq -2)} \\ &= \frac{P(\{-2 \leq Z < \frac{1}{2}\}}{P(Z \geq -2)} \\ &= \frac{\Phi(\frac{1}{2}) - \Phi(-2)}{1 - \Phi(-2)} \\ &= \frac{\Phi(\frac{1}{2}) + \Phi(2) - 1}{\Phi(2)} \end{aligned}$$

d) En utilisant la table des valeurs de la fonction Φ, on trouve :

$$P(|X-4| < 3 \big| X \geq 2) \approx \frac{0.69146 + 0.97725 - 1}{0.97725} \approx 0.68428.$$

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
+0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
+0.1	.53983	.54380	.54776	.55172	.55567	.55966	.56360	.56749	.57142	.57535
+0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
+0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
+0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
+0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
+0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
+0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
+0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
+0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
+1	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
+1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
+1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
+1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91308	.91466	.91621	.91774
+1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
+1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
+1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
+1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
+1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
+1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
+2	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
+2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
+2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
+2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
+2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
+2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
+2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
+2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
+2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
+2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
+3	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900

FIGURE 5.10 – Table des valeurs de la fonction Φ. Elle donne les valeurs, avec cinq chiffres significatifs, de $\Phi(z)$ pour z compris entre 0 et 3 et défini à deux décimales près. Par exemple, $\Phi(1.73)$ se trouve à la quatrième colonne de la ligne $+1.7$: on lit $\Phi(1.73) = 0.95818$.

Cette table a été générée par le site www.ztable.net.

Proposition 5.14 : Espérance et variance de la loi normale

Si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors $E(X) = \mu$ et $V(X) = \sigma^2$.

Démonstration. Soit $X \sim \mathcal{N}(\mu, \sigma^2)$. Puisque $Z = \dfrac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$ et que, $E(Z) = \dfrac{E(X)-\mu}{\sigma}$ et $V(Z) = \dfrac{V(X)}{\sigma^2}$, il suffit de faire les calculs avec $Z \sim \mathcal{N}(0,1)$ pour prouver le cas général.

On a

$$E(Z) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} z e^{-z^2/2} dz = 0$$

puisque la fonction à l'intérieur de l'intégrale est impaire.

De même, on a

$$\begin{aligned} V(Z) = E(Z^2) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} z^2 e^{-z^2/2} dz \\ &= -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} z \times \left(-z e^{-z^2/2}\right) dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-z^2/2} dz \\ &= P(Z \in \mathbb{R}) = 1 \end{aligned}$$

où, de la deuxième à la troisième ligne, on a fait une intégration par parties en remarquant que $-ze^{-z^2/2}$ est la dérivée de $e^{-z^2/2}$. □

Le théorème suivant, qui est un cas particulier du théorème central limite (voir le point historique ci-dessous), explique en grande partie la prédominance de la loi normale dans la modélisation de phénomènes physiques :

Théorème 5.15 : Lien entre la loi binomiale et la loi normale

Lorsque n tend vers l'infini, la loi binomiale tend vers une loi normale. Plus précisément :

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires discrètes telles que $X_n \sim \mathcal{B}(n,p)$.

Soit $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$ la variable aléatoire centrée réduite associée à X_n. Alors, on a

$$\lim_{n\to+\infty} Z_n \sim \mathcal{N}(0,1).$$

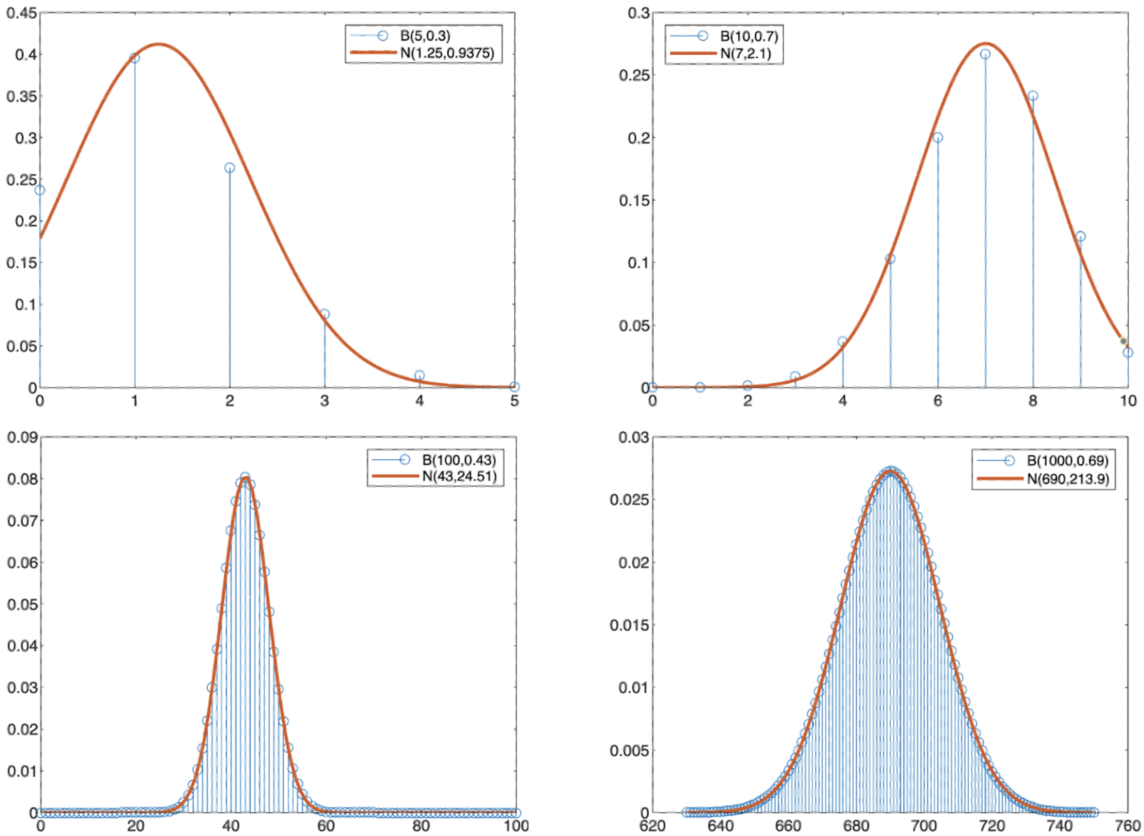


FIGURE 5.11 – Comparaison de la loi binomiale et de la loi normale pour différentes valeurs de n et p. Même pour des petites valeurs de n, l'approximation est déjà bonne.

Proposition 5.16 : Combinaison linéaire de loi normales

Soient X, Y deux variables aléatoires suivant des lois normales. Alors, toute combinaison linéaire de X et Y suit aussi une loi normale.

Plus précisément, supposons que $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ et considérons $Z = aX + bY$ avec $a, b \in \mathbb{R}$:

- $\star$ Si X et Y indépendantes, alors $Z \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$.
- $\star$ Si X et Y non-indépendantes, alors $Z \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2 + 2abCov(X,Y))$.

Démonstration. Pour le cas des variables indépendantes, voir l'exercice 5.19. Le deuxième cas est admis. □

♦Exemple : On note X la taille d'un homme dans la population française et Y la taille d'une femme dans la population française. Sachant que l'on a $X \sim \mathcal{N}(176, 64)$ et $Y \sim \mathcal{N}(164, 49)$, quelle est la probabilité que, si l'on sélectionne au hasard un homme et une femme, la femme soit plus grande que l'homme ?

On veut donc calculer $P(Y > X)$ qui peut aussi s'écrire $P(Y - X > 0)$. D'après la proposition précédente, on a $Y - X \sim \mathcal{N}(-12, 113)$. Pour calculer la probabilité recherchée, on suit la fiche méthode 7 (page 121) :

a) On pose la variable centrée réduite $Z = \dfrac{Y - X + 12}{\sqrt{113}}$.

b) $Y - X > 0 \iff Z > \dfrac{12}{\sqrt{113}}$.

c) Donc $P(Y > X) = P(Z > \dfrac{12}{\sqrt{113}}) = 1 - \Phi(\dfrac{12}{\sqrt{113}})$

d) Puisque $\dfrac{12}{\sqrt{113}} \approx 1.13$, on a, d'après la table des valeurs de la fonction Φ (page 122) :

$$P(Y > X) \approx 1 - \Phi(1.13) \approx 1 - 0.87076 \approx 0.12924.$$

Point historique #3 (La loi normale et le théorème centrale limite)

La loi normale a été d'abord introduite (de façon un peu obscure) par Abraham de Moivre en 1733 pour étudier la limite de lois binomiales. Mais c'est Pierre-Simon Laplace qui, dans son ouvrage *Théorie analytique des probabilités* de 1812, prouve une version améliorée du théorème d'or de Bernoulli (voir page 87) :

$$\forall x \in \mathbb{R}, \lim_{n \to +\infty} P\left(\frac{S_n - np}{\sqrt{npq}} < x\right) = \frac{1}{2\pi} \int_{-\infty}^{x} e^{-y^2/2} dy.$$

Ce théorème montre que, pour des grands nombres, l'expérience (discrète) de Bernoulli peut être décrite par une fonction continue qui prendra de ce fait un rôle central dans la théorie : la *loi normale ou loi Gaussienne* (appelée ainsi pour l'utilisation ultérieure qu'en a fait le mathématicien allemand Carl Friedrich Gauss).

Plus d'un demi-siècle plus tard (en 1887), l'école du mathématicien russe Pafnouti Tchebychev arrive à énoncer rigoureusement et prouver le *théorème central limite*, théorème qui généralise celui de Laplace, que Laplace avait suggéré sans arriver à le prouver et qui constitue probablement le résultat le plus important en théorie des probabilités :

> **Théorème central limite** Soit $\{X_n\}_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et identiques, d'espérance μ et écart-type σ. Alors,

$$\lim_{n \to +\infty} \frac{X_1 + \ldots + X_n}{n} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}).$$

En d'autres mots, *la loi normale est une bonne approximation de la moyenne de n'importe quelle expérience aléatoire qui se répète un grand nombre de fois* ! C'est la raison profonde qui explique l'omniprésence de cette loi.

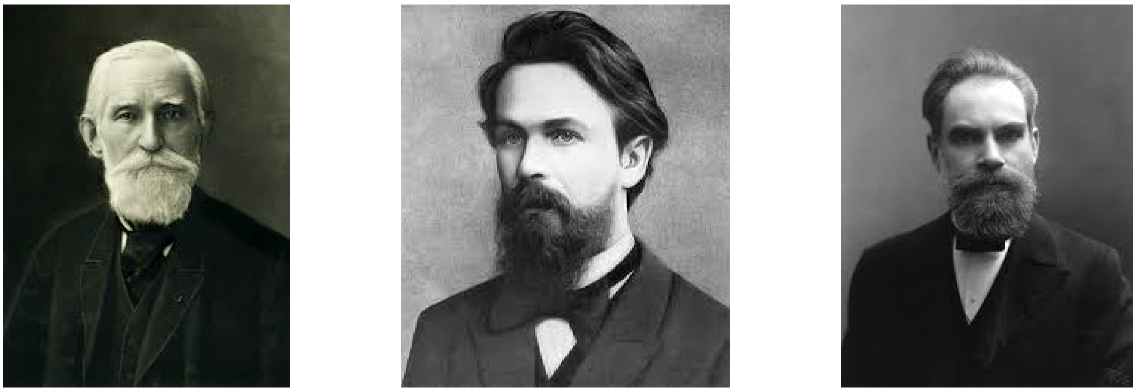


FIGURE 5.12 – L'école de Saint-Pétersbourg qui a prouvé le théorème central limite : Pafnouti Tchebychev et ses étudiants Andrei Markov et Alexandre Liapounov.

<u>Interprétation du théorème central limite</u> : Prenons une population avec une distribution quelconque, par exemple avec deux pics. Disons que cette distribution représente le poids de chaque individu dans cette population. Tirons aléatoirement des échantillons contenant un nombre fixé d'individus à l'intérieur de la population. Les moyennes des poids calculées sur chacun des échantillons suivront alors une loi gaussienne de paramètre $\mathcal{N}(\mu, \frac{\sigma^2}{n})$ avec σ et μ l'écart type et le poids moyen dans la population totale.

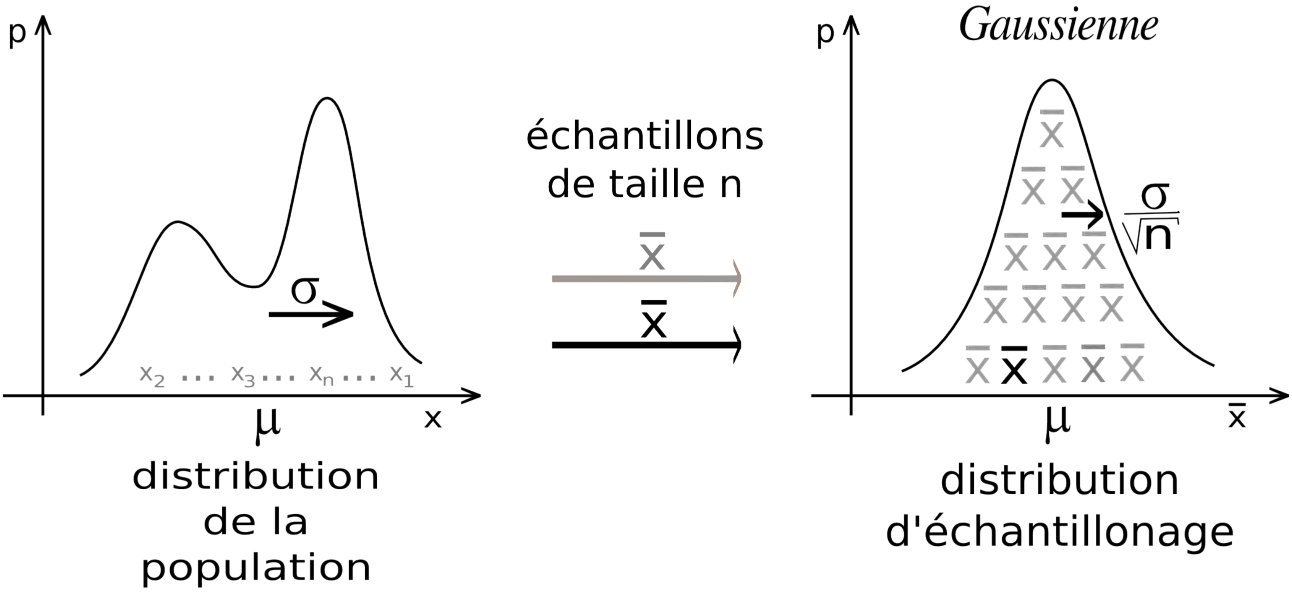


FIGURE 5.13 – La distribution à deux pics de la population est à gauche. A droite, il s'agit de la distribution de la moyenne $\bar{x}$ calculée sur des échantillons de taille n fixée.

Figure tirée de l'ouvrage "Calcul d'incertitudes" de Mathieu Rouaud.

5.5 Acquis d'apprentissage

À la fin de ce chapitre, vous devrez être en mesure de :

- ⋆ Reconnaître si une fonction est une densité de probabilité.
- ⋆ Calculer la probabilité d'un évènement du type $X \in B \subset \mathbb{R}$ pour une variable aléatoire continue.
- ⋆ Calculer la fonction de répartition d'une variable aléatoire continue.
- ⋆ Déterminer la densité de probabilité d'une nouvelle variable Y exprimée en fonction d'une variable X dont on connaît la densité de probabilité.
- ⋆ Calculer l'espérance, la variance et l'écart-type d'une variable aléatoire continue.
- ⋆ Déterminer la densité de probabilité conjointe de deux variables aléatoires conjointement continues.
- ⋆ Déterminer si deux variables aléatoires continues sont indépendantes.
- ⋆ Connaître les caractéristiques principales des lois uniforme, exponentielle et normale.
- ⋆ Comprendre le lien entre la loi normale, la loi binomiale et la loi de Poisson.
- ⋆ Savoir utiliser la table des valeurs de la fonction Φ pour calculer des probabilités sur des variables aléatoires suivant des lois normales.

5.6 Exercices

♦ = Exercice obligatoire.

[*] = Application directe du cours : sert à vérifier que vous avez bien compris les notions et techniques de base.

[**] = Demande plus de réflexion mais vous devez être capables de le résoudre et maîtriser les techniques utilisées dans cet exercice.

[***] = Exercice difficile pour aller plus loin, découvrir de nouvelles idées ou techniques. Pour les plus motivés !

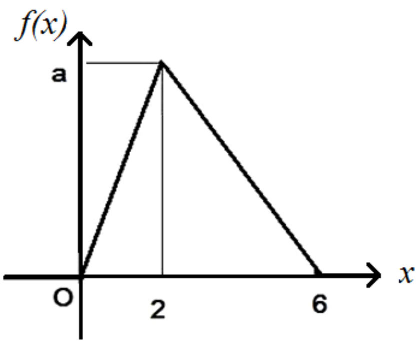
Faire le point en autonomie : QCM sur les notions de base

Soit Z une variable aléatoire suivant la loi normale centrée réduite et X une variable continue réelle quelconque.

Item	Enoncé
A	La formule $Var(X) = E(X^2) - E(X)^2$ ne s'applique pas pour les variables aléatoires continues.
B	Si f est la densité de probabilité de X, alors f est la dérivée de F la fonction de répartition de X.
C	L'espérance de X, si elle existe, s'écrit $E(X) = \sum_{-\infty}^{+\infty} xf(x)dx$.
D	La loi uniforme $\mathscr{U}([a,b])$ a pour espérance $b-a$.
E	La loi exponentielle est une loi dite "sans mémoire".
F	Si $Z \sim \mathscr{N}(1,0)$ alors $P(Z=0) = \frac{1}{2\pi}$.
G	Si $Z \sim \mathscr{N}(1,0)$ alors $P(Z < 1,26)$ est proche de 0,896.
H	Si $Z \sim \mathscr{N}(1,0)$ alors la valeur de a telle que $P(Z < a) = 0,367$ est plus petite que $-0,3$.

♦ **Exercice** 5.1 [*]

Une variable aléatoire X admet une densité f, affine par intervalles, dont la représentation graphique est donnée ci-dessous.



a) Quelle est la valeur de $a = f(2)$?

b) Calculez $P(X=1)$, $P(X<1)$, $P(X<2)$ et $P(2 < X \leq 6)$.

c) Quelle est l'espérance de X ?

Exercice 5.2 [*]

Une variable aléatoire continue X admet une densité de probabilité donnée par

$$f(x) = \begin{cases} \frac{k}{x} \text{ si } x \in [1,4] \\ 0 \text{ sinon.} \end{cases}$$

a) Déterminer la valeur de la constante k.

b) Calculer $P(1 \leq X \leq 2 \big| 1 \leq X \leq 3)$.

c) Déterminer $E(X)$ et $V(X)$.

♦ **Exercice** 5.3 [*]

Une variable aléatoire continue X admet une densité de probabilité donnée par

$$\forall x \in \mathbb{R}, f(x) = \frac{c}{1+x^2}.$$

a) Déterminer la valeur de la constante c.

b) Déterminer la fonction de répartition F_X.

c) Calculer $P(X \geq 2)$ et $P(-3 \leq X \leq 4)$.

d) Calculer $P(\frac{1}{3} \leq X^2 \leq 1)$.

e) Déterminer $E(X)$ et $V(X)$.

Exercice 5.4 [**]

Une variable aléatoire continue X admet une densité de probabilité donnée par

$$f(x) = \begin{cases} kx^2 \text{ si } x \in [-3,6] \\ 0 \text{ sinon.} \end{cases}$$

a) Déterminer la valeur de la constante k.

b) Calculer les densités de probabilité des variables $U = \frac{1}{3}(12 - X)$ et $V = X^2$.

c) Calculer l'espérance et la variance de X, U et V.

Exercice 5.5 [*]

Un joueur de fléchettes trouve que la distance R de l'impact de la fléchette au centre de la cible (de rayon a) suit une loi dont la densité de probabilité est $f(r) = c\left(1 - \frac{r^2}{a^2}\right)$ où c est une constante. Calculer la probabilité de faire mouche dans un rayon b autour du centre de la cible (on fera l'hypothèse que la cible est toujours atteinte).

Exercice 5.6 [**] **Température d'un gaz**

On considère une boîte contenant N particules d'oxygène de masse $m = 2,7.10^{-26}$kg. On suppose que la vitesse des particules est une variable aléatoire V qui suit la loi suivante

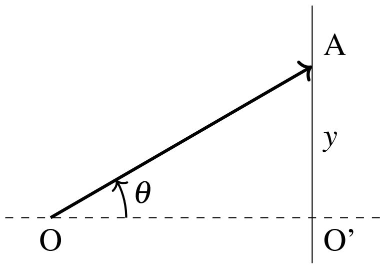
$$f(v) = \begin{cases} ce^{-\frac{|v-v_0|}{v_0}} \text{ si } v \geq 0 \\ 0 \text{ sinon.} \end{cases} \qquad \text{où } c \in \mathbb{R} \text{ et } v_0 = 800\text{km/h}$$

a) Trouver la valeur de la constante c.

b) Soit $K = \frac{1}{2}mV^2$ l'énergie cinétique d'une particule. Sachant que la température du gaz T est liée à l'énergie cinétique moyenne des particules par l'équation $kT = E(K)$ où $k = 1,38.10^{-23}$J/K est la constante de Boltzmann, trouver la température du gaz.

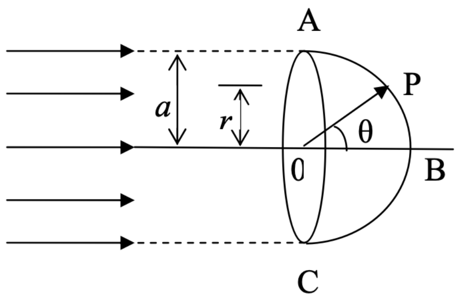
♦ **Exercice** 5.7 [*]

Une source lumineuse émet au point O de façon uniforme dans le plan $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$. Calculer la densité de probabilité de la variable aléatoire Y égale à la valeur algébrique de O'A. On prendra la distance OO' égale à un mètre.



Exercice 5.8 [*]

Un faisceau cylindrique de particules de rayon a est dirigé vers une cible hémisphérique ABC de centre O. Soit $f(r)$ la densité de probabilité uniforme sur $[0,a]$ de la variable aléatoire continue $R =$ "distance de la particule à l'axe OB.



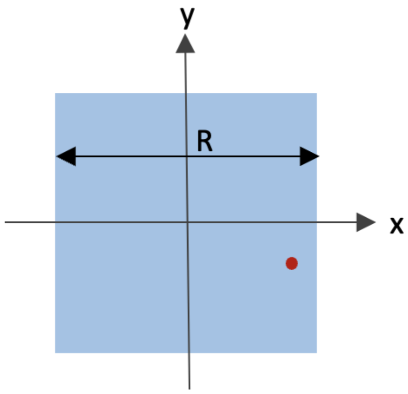
a) Calculer $f(r)$.

b) Calculer la densité de probabilité de l'angle $\theta = BOP$.

c) Calculer la probabilité qu'une particule frappe la cible entre $\theta = 0$ et $\theta = \pi/4$.

♦ **Exercice** 5.9 [**]

Reprendre l'exemple de la section 5.2 mais en considérant une mare carrée de côté R.



En particulier :

a) Trouver la densité de probabilité conjointe de X et Y.

b) Calculer la probabilité que l'enfant puisse récupérer le ballon lorsqu'il revient le jour d'après.

c) Trouver la densité de probabilité de X et celle de Y.

d) Déterminer si X et Y sont indépendantes.

Exercice 5.10 [***] **Le problème de l'aiguille de Buffon**

On jette une aiguille de longueur L sur une feuille de papier lignée dont la séparation entre les lignes horizontales est $D \geq L$. Quelle est la probabilité que l'aiguille intersecte une ligne ?

♦ **Exercice** 5.11 [**] **Temps d'attente**

Cet exercice est tiré de S. Ross, *A first course in Probability*, p. 297.

Un homme et une femme se donnent rendez-vous à la place Furstenberg autour de 12h30. En supposant que les deux personnes sont indépendantes et que l'heure d'arrivée de l'homme suit une loi uniforme entre 12h15 et 12h45 et l'heure d'arrivée de la femme suit une loi uniforme entre 12h et 13h, calculer :

a) La probabilité que le premier arrivé n'attende pas plus de cinq minutes.

b) La probabilité que l'homme arrive en premier.

Exercice 5.12 [**] *(Extrait du DE 2019-2020)*

Un pantalon est retenu par une paire de bretelles et une ceinture.

La paire de bretelles et la ceinture ont des durées de vie données par des variables aléatoires continues T_1 et T_2 qui suivent une loi exponentielle de paramètres λ_1 et λ_2.

On suppose que le pantalon est maintenu tant que l'un des deux accessoires au moins, bretelles ou ceinture, est fonctionnel. Soit T la variable aléatoire continue "durée de maintien du pantalon".

a) Donner les fonctions de répartitions F_1 et F_2 de T_1 et T_2.

b) Calculer $P(T_1 > 2t)$ pour $t > 0$.

c) Calculer $P(t < T_2 < 2t)$ pour $t > 0$.

d) Donner la fonction de répartition de T.

e) En déduire la densité de probabilité de T.

f) Donner les espérances $E(T_1), E(T_2)$ et $E(T)$.

Exercice 5.13 [***] **Variables aléatoires sans mémoire** *(Preuve de la proposition 5.10)*

On dit qu'une variable aléatoire T est *sans mémoire* si elle vérifie

$$\forall s \geq 0, P(T \geq t+s \big| T \geq t) = P(T \geq s).$$

a) Vérifier qu'une variable aléatoire suivant une loi exponentielle est sans mémoire.

b) Réciproquement, on va démontrer que si T est une variable aléatoire continue sans mémoire, alors elle suit une loi exponentielle.

i) Prouver que la fonction $G_T(t) = P(T \geq t)$ satisfait

$$\forall t, s \geq 0, G_T(t+s) = G_T(t) G_T(s).$$

ii) En déduire l'expression de G_T puis celles de la fonction de répartition F_T et de la densité de probabilité de T.

♦ **Exercice** 5.14 [**] **Minimum et maximum de deux lois exponentielles**

Cet exercice est tiré du cours donné par M. Ghassany à l'EFREI en 2022/2023.

L'objectif de cet exercice est de résoudre le problème suivant : deux guichets sont ouverts à une banque. Le temps de service T_1 au premier guichet (respectivement, T_2 au deuxième guichet) suit une loi exponentielle de moyenne 5 minutes (respectivement 8 minutes). On suppose que les deux guichets sont indépendants. Deux clients rentrent simultanément, un choisit le guichet 1 et l'autre le guichet 2. On veut trouver le temps moyen δT au bout duquel sort le premier client et le temps moyen ΔT au bout duquel sort le dernier client.

a) On pose $Z = min(T_1, T_2)$. Exprimer la fonction de répartition de Z en fonction de celles de T_1 et T_2.

b) En déduire que Z suit aussi une loi exponentielle de paramètre $\lambda = 40/13$.

c) Trouver δT et ΔT.

Aide : Pour ΔT, utiliser le fait que $T_1 + T_2 = min(T_1, T_2) + max(T_1, T_2)$.

Exercice 5.15 [**] **Loi exponentielle et loi uniforme**

Soit X une variable aléatoire continue de loi uniforme sur $[0,1]$. Montrer que la variable aléatoire $Y = -ln(X)$ suit une loi exponentielle.

♦ **Exercice** 5.16 [*]

Soit X une variable aléatoire suivant une loi normale de paramètres $\mu = 10$ et $\sigma = 2$.

a) Quelle est la valeur de $P(X \leq 15)$?

b) Quelle est la valeur de $P(|X| \leq 8)$?

c) Quelle est la valeur de $P(8 \leq X \leq 15)$?

♦ **Exercice** 5.17 [*]

On suppose que la répartition du taux de cholestérol dans une population d'enfants (en cg) suit une loi normale de moyenne μ et d'écart-type σ.

On a constaté que 57,9% de ces enfants ont un taux de cholestérol inférieur à 165cg et que 4,5% ont un taux de cholestérol supérieur à 180cg.

a) Calculer μ et σ.

b) Les enfants dont le taux est supérieur à 183cg doivent subir un traitement. Quel pourcentage de la population doit-on soigner ?

♦ **Exercice** 5.18 [*] **Score de Maddrey**

Chez les malades ayant une atteinte sévère, le score de Maddrey est une variable aléatoire X qui est distribuée selon une loi normale $\mathcal{N}(\mu = 54, \sigma^2 = 14^2)$, tandis que chez les malades avec atteinte minime, le score de Maddrey est une variable aléatoire Y qui est distribuée selon une loi normale $\mathcal{N}(\mu = 20, \sigma^2 = 6^2)$.

a) Pour un malade ayant une atteinte sévère, calculez la probabilité qu'il ait un score de Maddrey en dessous de 68.

b) Pour un malade ayant une atteinte minime, calculez la probabilité qu'il ait un score de Maddrey en dessous de 8.

c) Chez les malades ayant une atteinte sévère, quel seuil s devons-nous prendre pour que $P(X > s) = 0,95$?

d) Chez les malades ayant une atteinte minime, quel seuil s devons-nous prendre pour que $P(Y < s) = 0,90$?

On choisit maintenant $s = 32$ pour détecter les deux types d'atteintes.

e) Chez les malades ayant une atteinte sévère, que vaut $P(X < s)$?

f) Chez les malades ayant une atteinte minime, que vaut $P(Y > s)$?

g) Donnez le nombre d'erreurs de diagnostique pour ce seuil s de détection sur un échantillon de 100 personnes.

Exercice 5.19 [***] **Preuve de la proposition 5.16**

Soient X, Y deux variables aléatoires continues et indépendantes. On note $f_{X+Y}, f_X, f_Y, F_{X+Y}, F_X$ et F_Y les densités de probabilité et les fonctions de répartition des variables $X+Y$, X et Y.

a) Soit $a \in \mathbb{R}$. Prouver que si $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, alors $aX \sim \mathcal{N}(a\mu_1, a^2\sigma_1^2)$.

b) Prouver que l'on a, pour $a \in \mathbb{R}$, $F_{X+Y}(a) = \int_{-\infty}^{+\infty} F_X(a-y) f_Y(y) dy$.

c) En déduire que l'on a $f_{X+Y}(a) = \int_{-\infty}^{+\infty} f_X(a-y) f_Y(y) dy$.

d) Conclure que si $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, alors $X+Y \sim \mathcal{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$. Enfin, en déduire la loi de distribution de $aX+bY$, avec $a, b \in \mathbb{R}$.

Exercice 5.20 [**] **Mesure de la masse du boson de Higgs**

Le modèle standard des particules est une des plus belles constructions théoriques de la physique du 20ème siècle. En 1964, le physicien britannique Peter Higgs proposa l'existence d'une particule pas encore observée qui permettrait d'expliquer comment les autres particules acquièrent une masse. Cette particule fut découverte en 2012 par le LHC à Genève, ce qui valut le prix Nobel à Peter Higgs et François Englert.

La masse d'une particule est mesurée en GeV/c^2, avec $1 GeV/c^2 = 1.783 \times 10^{-27} kg$. Dans l'expérience initiale de 2012, on a trouvé $M_H \sim \mathcal{N}(125.06, 0.0841)$. Mais l'expérience a été refaite en 2016 et les résultats on donné $M_H \sim \mathcal{N}(125.46, 0.0289)$ [2] . Si l'on combine ces deux résultats, quelle est la loi de probabilité de la masse du boson de Higgs ?

2. Données extraites de ce rapport du CMS.