

ENS PARIS-SACLAY  
PRÉPARATION À L'AGRÉGATION EXTERNE DE SCIENCES SOCIALES

---

# Probabilités élémentaires



Pierre Montagnon et Simon Coste

---



FIGURE 1 – Kolmogorov.

On situe l'émergence de la théorie moderne des probabilités aux premières décennies du XXème siècle. C'est à partir du XVIIème siècle que les premières questions de probabilités sont étudiées (notamment par Pascal et Fermat); il s'agit alors surtout de problèmes de combinatoire. Pierre-Simon de Laplace énonce la première formulation rigoureuse du théorème central limite vers 1812, mais il faut attendre les travaux d'Émile Borel (1871 - 1956) et surtout ceux d'Andreï Kolmogorov (1903 - 1987) pour que la théorie des probabilités adopte le formalisme rigoureux qu'on lui connaît aujourd'hui, dont ce cours présente une version allégée.

La théorie des probabilités et des statistiques est aujourd'hui devenue un pilier des sciences modernes... mais elle est aussi au cœur des nouvelles pratiques économiques et commerciales.

Nous présentons dans ce cours les bases<sup>1</sup> de la théorie des probabilités : après un chapitre introductif consacré à la notion de probabilité conditionnelle et à celle d'indépendance, nous introduisons la notion de variable aléatoire réelle dans le cadre d'un espace d'états fini ou inclus dans  $\mathbb{N}$  puis dans celui des variables aléatoires réelles à densité. Nous présentons successivement les propriétés générales des variables aléatoires ainsi définies et des  $n$ -uplets de telles variables ainsi qu'un certain nombre de lois de probabilité usuelles dont nous détaillons et illustrons les propriétés. La dernière partie du cours est consacrée aux notions de convergence en probabilité et en loi des suites de variables aléatoires réelles ainsi qu'aux grands théorèmes de convergence, dont on présente des applications à l'estimation paramétrique.

Pour toute question ou remarque sur ce cours, vous pouvez contacter les auteurs aux adresses suivantes :

pierre.montagnon@ens-cachan.fr  
simon@scoste.fr



---

1. ... mais pas les fondements!

---

# Table des matières

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Probabilités conditionnelles et indépendance</b>                  | <b>3</b>  |
| 1.1      | Probabilités conditionnelles et formules classiques . . . . .        | 3         |
| 1.2      | Indépendance . . . . .   | 5         |
| <b>2</b> | <b>Variables aléatoires réelles discrètes</b>                        | <b>7</b>  |
| 2.1      | Généralités sur les variables aléatoires réelles . . . . .           | 7         |
| 2.2      | Variables aléatoires réelles discrètes . . . . .                     | 8         |
| 2.2.1    | Loi d'une variable aléatoire réelle discrète . . . . .               | 8         |
| 2.2.2    | Espérance, moments et variance d'une variable aléatoire discrète .   | 9         |
| 2.2.3    | Couples de variables aléatoires réelles discrètes . . . . .          | 12        |
| 2.3      | Exemples de lois discrètes . . . . .                                 | 17        |
| <b>3</b> | <b>Variables aléatoires à densité</b>                                | <b>20</b> |
| 3.1      | Généralités sur les variables aléatoires réelles à densité . . . . . | 20        |
| 3.2      | Exemples de lois à densité . . . . .                                 | 25        |
| <b>4</b> | <b>Théorèmes de convergence et estimation</b>                        | <b>30</b> |
| 4.1      | Convergence d'une suite de variables aléatoires réelles . . . . .    | 30        |
| 4.1.1    | Convergence en probabilité et inégalités classiques. . . . .         | 30        |
| 4.1.2    | La loi des grands nombres. . . . .                                   | 31        |
| 4.1.3    | Convergence en loi. . . . .  | 33        |
| 4.2      | Estimation . . . . .   | 37        |
| 4.2.1    | Généralités et propriétés des estimateurs . . . . .                  | 37        |
| 4.2.2    | Intervalles de confiance . . . . .                                   | 40        |
| <b>5</b> | <b>Solutions des exercices</b>                                       | <b>45</b> |

# 1 Probabilités conditionnelles et indépendance

On se donne un ensemble des possibles  $\Omega$  fini ou dénombrable que l'on munit d'une fonction de probabilité  $\mathbb{P}$ . On rappelle qu'une telle fonction est définie de la façon suivante :

**DÉFINITION 1.1** (Fonction de probabilité). — On appelle fonction de probabilité une fonction  $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$  telle que :

- $\mathbb{P}(\Omega) = 1$
- pour toute suite d'événements  $(A_i)_{i \in \mathbb{N}}$  deux à deux disjoints :

$$\mathbb{P}\left(\bigsqcup_{i=0}^{+\infty} A_i\right) = \sum_{i=0}^{+\infty} \mathbb{P}(A_i)$$

## 1.1 Probabilités conditionnelles et formules classiques

**DÉFINITION 1.2** (Probabilité conditionnelle). — Soient  $A$  un événement tel que  $\mathbb{P}(A) \neq 0$  et  $B$  un événement quelconque. On appelle probabilité conditionnelle de  $B$  sachant  $A$  la quantité

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

On peut aussi écrire cette quantité  $\mathbb{P}_A(B)$ . Nous conseillons l'usage de la notation proposée dans la définition ci-dessus car elle permet de conditionner par des événements longs à écrire sans que la rédaction perde en élégance — simple question de goût...

**REMARQUE 1.3.** Dans le cadre de la définition ci-dessus, on a  $\mathbb{P}(A)\mathbb{P}(B | A) = \mathbb{P}(A \cap B)$ , ce qui est conforme avec l'idée selon laquelle  $\mathbb{P}(B | A)$  représente, comme son nom l'indique, la probabilité de l'événement  $B$  conditionnellement au fait que  $A$  est réalisé. En effet, il suffit d'écrire les probabilités correspondant à chacune des branches de l'arbre ci-dessous pour obtenir à nouveau cette égalité.

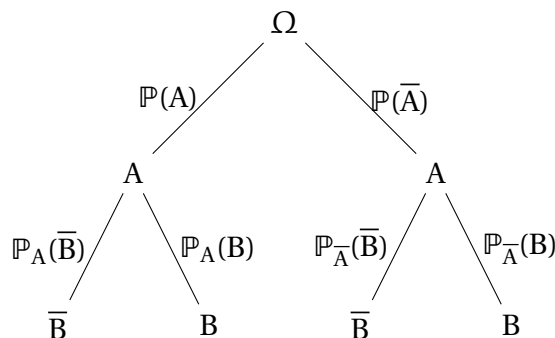


FIGURE 2 – Un arbre de probabilités.

**EXEMPLE 1.4.** On lance deux dés équilibrés à six faces et on sait que le résultat obtenu est supérieur ou égal à 9. Compte tenu de cette information, la probabilité pour que ce

résultat soit supérieur ou égal à 10 est

$$\begin{aligned}\mathbb{P}(B | A) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(\{(5, 5), (4, 6), (6, 4), (5, 6), (6, 5), (6, 6)\})}{\mathbb{P}(\{(5, 4), (6, 3), (4, 5), (3, 6), (5, 5), (4, 6), (6, 4), (5, 6), (6, 5), (6, 6)\})} \\ &= \frac{\frac{6}{36}}{\frac{10}{36}} = \frac{3}{5}\end{aligned}$$

où l'on a choisi  $\Omega = \llbracket 1, 6 \rrbracket^2 = \{(i, j) \mid i, j \in \llbracket 1, 6 \rrbracket\}$  (c'est-à-dire que l'on différencie les deux dés et que l'ensemble des possibles est modélisé par l'ensemble des couples de résultats que l'on peut obtenir), posé A l'événement « le résultat est supérieur ou égal à 9 », B l'événement « le résultat est supérieur ou égal à 10 » et calculé les probabilités respectives de A et de  $A \cap B = B$  en utilisant la situation d'équiprobabilité.

**REMARQUE 1.5** (Formule de Bayes). Dans le cadre de la définition ci-dessus et si  $\mathbb{P}(B) \neq 0$ , on obtient une relation entre  $\mathbb{P}(A | B)$  et  $\mathbb{P}(B | A)$  en écrivant

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

puis

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

La formule suivante est claire sur un arbre, qu'il est conseillé d'écrire en tant qu'exercice :

**THÉORÈME 1.6** (Formule des probabilités composées). — Si  $A_1, \dots, A_n$  sont des événements tels que  $A_1 \cap \dots \cap A_n$  est de probabilité non nulle et si B est un événement quelconque,

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

**Exercice 1.7.** Écrire la formule des probabilités composées pour  $n = 1$  et  $n = 2$ . Établir la formule pour tout  $n$  par récurrence.

**DÉFINITION 1.8** (Système complet d'événements). — On appelle système complet d'événements toute famille d'événements deux à deux incompatibles et dont la réunion vaut  $\Omega$ , c'est-à-dire toute famille  $(A_i)_{i \in I}$  (avec I fini ou non) telle que  $A_i \cap A_j = \emptyset$  dès que  $i \neq j$  et telle que

$$\bigcup_{i \in I} A_i = \Omega$$

**REMARQUE 1.9.** Le terme « système complet d'événements » fait référence au fait que la famille d'événements choisie « couvre tous les cas possibles ».

**EXEMPLE 1.10.** Si A est un événement,  $\{A, \bar{A}\}$  constitue un système complet d'événements.

**EXEMPLE 1.11.** Dans le cas d'une expérience constituée par le lancer successif de deux dés équilibrés à six faces, on modélise l'ensemble des possibles par  $\Omega = \llbracket 1, 6 \rrbracket^2$ . Un exemple de système complet d'événements est donné par la famille  $\{A_i, i \in \llbracket 1, 6 \rrbracket\}$ , où pour tout  $i \in \llbracket 1, 6 \rrbracket$  on écrit  $A_i = \{i\} \times \llbracket 1, 6 \rrbracket = \{(i, 1), (i, 2), (i, 3), (i, 4), (i, 5), (i, 6)\}$  l'événement « le résultat du premier dé est i ».

**THÉORÈME 1.12** (Formule des probabilités totales). — Soit  $(A_i)_{i \in I}$  un système complet d'événements fini ou dénombrable tel que pour tout  $i \in I$  on ait  $\mathbb{P}(A_i) \neq 0$ . Alors pour tout événement  $B$ , on a

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

Dans le cas où  $I$  est infini, la somme ci-dessus est la somme d'une série convergente.

**Exercice 1.13.** Démontrer la formule des probabilités totales.

**EXEMPLE 1.14.** Une classe est constituée de 35 filles et 15 garçons. Deux tiers des garçons et la moitié des filles étudient l'espagnol. Si l'on choisit un élève au hasard dans la classe, la probabilité qu'il ou elle étudie l'espagnol est

$$\mathbb{P}(E) = \mathbb{P}(E \mid G) \mathbb{P}(G) + \mathbb{P}(E \mid F) \mathbb{P}(F) = \frac{2}{3} \frac{15}{50} + \frac{1}{2} \frac{35}{50} = \frac{11}{20} = 55\%$$

où l'on a noté respectivement  $E, F$  et  $G$  les événements « l'élève étudie l'espagnol », « l'élève est une fille » et « l'élève est un garçon » et utilisé la formule des probabilités totales sur le système complet d'événements  $\{G, F\}$ .

**REMARQUE 1.15.** La formule des probabilités totales est appelée ainsi car elle permet d'exprimer la probabilité d'un événement comme la somme des probabilités associées à chacune des branches d'un arbre obtenu en décomposant cet événement sur le système complet d'événements.

**Exercice 1.16.** Deux policiers surveillent le trafic sur une route. Il est de notoriété publique que 1% des automobilistes empruntant la route sont ivres, et les policiers disposent d'un alcootest détectant correctement 99% des personnes ivres et donnant un faux positif que dans le cas de 2% des automobilistes sobres. Sachant que Roger vient de se faire arrêter et que l'alcootest a donné un résultat positif, quelle est la probabilité que Roger soit ivre?

## 1.2 Indépendance

**DÉFINITION 1.17** (Événements indépendants). — On dit que deux événements  $A$  et  $B$  sont indépendants si et seulement si  $\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B)$ .

**REMARQUE 1.18.** Deux événements  $A$  et  $B$  (avec  $\mathbb{P}(B) \neq 0$ ) sont indépendants si et seulement si  $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ , c'est-à-dire que le fait que  $B$  se produise n'influe aucunement sur la probabilité d'occurrence de  $A$ .

**Exercice 1.19.** Montrer que si  $A$  et  $B$  sont deux événements indépendants, alors :

- $A$  et  $\bar{B}$  sont indépendants
- $\bar{A}$  et  $B$  sont indépendants
- $\bar{A}$  et  $\bar{B}$  sont indépendants

**DÉFINITION 1.20.** — On dit que des événements  $(A_i)_{i \in I}$  (avec  $I$  finie ou dénombrable) sont deux à deux indépendants si et seulement si  $A_i$  et  $A_j$  sont indépendants pour tout couple  $(i, j) \in I^2$  tel que  $i \neq j$ .

On dit que des événements  $(A_i)_{i \in I}$  (avec  $I$  finie ou dénombrable) sont mutuellement indépendants si pour toute partie finie  $J \subset I$  on a

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i)$$

**REMARQUE 1.21.** Par convention, lorsqu'un énoncé parle d'« événements indépendants » sans préciser de quel concept d'indépendance il s'agit, il faut comprendre « événements mutuellement indépendants ».

**REMARQUE 1.22.** En prenant  $J = \{(i, j)\} \subset I$  avec  $i \neq j$ , on voit que des événements mutuellement indépendants sont automatiquement deux à deux indépendants. En revanche, la réciproque est fautive comme on pourra s'en convaincre en considérant l'expérience aléatoire réalisée en lançant deux dés équilibrés à six faces et en notant  $A$ , resp.  $B$ , resp.  $C$  les événements « le résultat du premier dé est pair », resp. « le résultat du second dé est pair », resp. « la somme des deux résultats est paire » et en vérifiant que  $\mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ , mais que  $A$ ,  $B$  et  $C$  sont bien indépendants deux à deux.

**PROPOSITION 1.23.** — Si les  $(A_i)_{i \in I}$  sont deux à deux (resp. mutuellement) indépendants, alors les  $(\overline{A_i})_{i \in I}$  le sont, de même que toute famille  $(B_i)_{i \in I}$  où pour tout  $i \in I$  l'ensemble  $B_i$  est égal à  $A_i$  ou à  $\overline{A_i}$ .

Il est très fréquent d'utiliser une hypothèse d'indépendance mutuelle pour transformer la probabilité d'une intersection d'événements en un produit. Cependant, il est indispensable pour cela de s'assurer que les événements en question sont bel et bien mutuellement indépendants (et pas seulement deux à deux indépendants, par exemple).

**Exercice 1.24.** Gonzague jubile : il ne lui manque plus qu'une vignette pour compléter sa collection de 500 vignettes autocollantes représentant les meilleurs mathématiciens de France. Conscient du fait que la probabilité d'obtenir l'autocollant qui lui manque en achetant une vignette au hasard est de  $1/500$ , il décide de commander 500 nouvelles vignettes, dont chacune est choisie de manière équiprobable et indépendante des autres parmi l'ensemble des vignettes existantes par l'entreprise qui les fabrique. Quelle est la probabilité pour que cette commande permette à Gonzague de compléter sa collection ?

## 2 Variables aléatoires réelles discrètes

### 2.1 Généralités sur les variables aléatoires réelles

On suppose toujours que l'ensemble des possibles  $\Omega$  est fini ou dénombrable, et que la fonction  $\mathbb{P}$  est définie sur l'ensemble  $\mathcal{P}(\Omega)$  des parties de  $\Omega$ . Nous présentons ici le modèle formel de base de la théorie des probabilités<sup>2</sup>.

**DÉFINITION 2.1** (Variable aléatoire réelle). — On appelle *variable aléatoire réelle* une fonction  $X : \Omega \rightarrow \mathbb{R}$ .

**REMARQUE 2.2.** On remarquera que  $X$  n'est pas « intrinsèquement » aléatoire mais qu'elle associe à chaque événement élémentaire  $\omega \in \Omega$  (aussi appelé *état du monde*) un réel donné. L'idée centrale de notre modélisation de l'aléatoire étant de considérer que l'état du monde  $\omega$  qui se réalise effectivement est choisi aléatoirement dans  $\Omega$  (c'est le point de vue que l'on adopte en lui associant une probabilité!), la valeur « effective » de  $X(\omega)$  l'est aussi. En notant simplement  $X$  la valeur  $X(\omega)$  effectivement réalisée, on voit que  $X$  est un objet qui prend certaines valeurs réelles avec certaines probabilités : c'est la définition d'une variable aléatoire que l'on trouve au lycée.

Il est néanmoins important de garder à l'esprit le fait que  $X$  désigne, avant tout abus de langage, une fonction, et que l'on peut être amené dans certains cas à travailler avec des notations du type  $X(\omega)$ , qu'il faut savoir interpréter — même si l'on peut affirmer sans risque que rien de très élaboré ne sera attendu sur le sujet.

**REMARQUE 2.3.** Si  $X$  est une variable aléatoire réelle, l'ensemble des valeurs prises par  $X$  est  $X(\Omega)$ . On rencontrera souvent cette notation, qui vient directement du fait que  $X$  est une fonction définie sur  $\Omega$ .

**NOTATION 1.** — Si  $A$  est une partie de  $\mathbb{R}$  et si  $X$  est une variable aléatoire réelle, on note  $(X \in A)$  ou parfois  $\{X \in A\}$  l'événement (éventuellement vide)  $X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}$ , constitué de tous les états du monde dans lesquels la variable  $X$  appartient à l'ensemble  $A$ . De même, si  $x \in \mathbb{R}$ , on note  $(X = x)$  l'événement  $X^{-1}(\{x\})$ .

**REMARQUE 2.4.** Comme on l'a vu plus haut, dans l'écrasante majorité des cas on fait référence à  $X$  comme à une véritable variable; aussi trouve-t-on rarement l'écriture  $X^{-1}(A)$ , à laquelle on préfère  $(X \in A)$ . Rappelons une dernière fois qu'il est commode pour comprendre les notions présentées plus loin et leur motivation de concevoir les variables aléatoires comme des *variables* effectivement aléatoires; dans ce qui suit, on ne cherchera pas à expliciter les variables présentées en tant que fonctions définies sur  $\Omega$  et on admettra l'existence de variables aléatoires suivant effectivement les différentes lois présentées.

**DÉFINITION 2.5** (Indépendance). — Si  $(X_i)_{i \in I}$  est une famille (finie ou non) de variables aléatoires réelles, on dit que les  $X_i$  sont indépendantes deux à deux si et seulement si pour tout  $(i, j) \in I^2$  tel que  $i \neq j$  et tout couple  $(A_i, A_j)$  d'intervalles de  $\mathbb{R}$  on a

$$\mathbb{P}((X_i \in A_i) \cap (X_j \in A_j)) = \mathbb{P}(X_i \in A_i) \mathbb{P}(X_j \in A_j)$$

---

2. Ce formalisme moderne est dû à A. Kolmogorov. Il fut mis en place dans les années 1930, ce qui est relativement tardif!



et que les  $X_i$  sont mutuellement indépendantes si et seulement si pour toute famille finie  $J \subset I$  et toute famille  $(A_i)_{i \in J}$  d'intervalles de  $\mathbb{R}$  on a

$$\mathbb{P}\left(\bigcap_{i \in J} (X_i \in A_i)\right) = \prod_{i \in J} \mathbb{P}(X_i \in A_i)$$

**REMARQUE 2.6.** On remarque que les variables aléatoires réelles  $X_i$  sont indépendantes en un certain sens si et seulement si les événements  $(X_i \in A_i)$  le sont en ce même sens et ce pour tout choix de parties  $A_i \subset \mathbb{R}$ . On en déduit que l'indépendance mutuelle de variables aléatoires réelles implique leur indépendance deux à deux; ici encore, la réciproque est fautive (il suffit d'ailleurs d'introduire les variables aléatoires adéquates pour que le contre-exemple donné plus haut soit à nouveau valable).

**THÉORÈME 2.7.** — Si  $(X_i)_{i \in I}$  est une famille de variables aléatoires réelles mutuellement indépendantes, alors  $(f_i(X_i))_{i \in I}$  l'est aussi pour toute famille  $(f_i)_{i \in I}$  de fonctions à valeurs réelles continues par morceaux sur  $\mathbb{R}$ .

**EXEMPLE 2.8.** Si  $X$  et  $Y$  sont deux variables aléatoires réelles indépendantes, alors  $X^2$  et  $3Y + 1$  le sont aussi.

**DÉFINITION 2.9** (Fonction de répartition). — Si  $X$  est une variable aléatoire réelle, on appelle fonction de répartition de  $X$  la fonction

$$f: \begin{cases} \mathbb{R} \longrightarrow [0, 1] \\ x \rightarrow \mathbb{P}(X \leq x) \end{cases}$$

**PROPOSITION 2.10.** — La fonction de répartition  $F$  d'une variable aléatoire réelle  $X$  est croissante, continue à droite en tout point et tend vers 0 en  $-\infty$  et vers 1 en  $+\infty$ . D'autre part, si  $x \in \mathbb{R}$  on a

$$1 - F(x) = \mathbb{P}(X > x)$$

Si en plus  $a, b \in \mathbb{R}$  vérifient  $a < b$ , alors on a  $\mathbb{P}(X \in ]a, b]) = F(b) - F(a)$

Dans la deuxième partie de la proposition précédente, attention à la forme de l'intervalle  $]a, b]$ ! La proposition devient fautive si on le remplace, par exemple, par l'intervalle  $[a, b]$  dans le cas où  $X$  peut prendre la valeur  $a$  avec une probabilité non nulle.

## 2.2 Variables aléatoires réelles discrètes

### 2.2.1 Loi d'une variable aléatoire réelle discrète

On s'intéresse dans cette section aux variables aléatoires réelles  $X$  prenant un nombre fini ou dénombrable de valeurs, c'est-à-dire telles que  $X(\Omega)$  est fini ou dénombrable.

**DÉFINITION 2.11** (Variable aléatoire réelle discrète). — On appelle variable aléatoire réelle discrète toute variable aléatoire réelle  $X$  prenant un nombre fini ou dénombrable de valeurs.

**DÉFINITION 2.12.** — La loi d'une variable aléatoire réelle discrète  $X$  est la donnée de  $X(\Omega)$  et de  $\mathbb{P}(X = x)$  pour tout  $x \in X(\Omega)$ .

**REMARQUE 2.13.** Il est parfois commode de donner la loi d'une variable aléatoire  $X$  prenant un nombre fini de valeurs sous la forme d'un tableau dont la somme des coefficients de la deuxième ligne vaut 1. Par exemple, le tableau suivant définit une loi de probabilité :

| $\mathbb{P}(X = 0)$ | $\mathbb{P}(X = 1)$ | $\mathbb{P}(X = 2)$ |
|---------------------|---------------------|---------------------|
| 1/8                 | 1/4                 | 5/8                 |

On peut trouver çà et là des questions du type « soit  $(p_i)_{i \in \mathbb{N}}$  une suite de réels définie par ... ; vérifier que cette suite définit une loi de probabilité sur  $\mathbb{N}$  ». Le sens de cette question est le suivant : montrer qu'il est possible de trouver une variable aléatoire réelle  $X$  à valeurs dans  $\mathbb{N}$  telle que  $p_i = \mathbb{P}(X = i)$  pour tout  $i \in \mathbb{N}$ . Pour cela, il suffit de vérifier que  $p_i \geq 0$  pour tout  $i \geq 0$  et que  $\sum_{i \geq 0} p_i$  converge et a pour somme 1. Si les  $p_i$  donnés ne forment pas une suite mais sont en nombre fini, il suffit de vérifier qu'ils sont tous positifs et que leur somme vaut 1.

**EXEMPLE 2.14.** La suite  $(p_i)_{i \in \mathbb{N}}$  définie par  $p_0 = \frac{1}{2}$  et par  $p_i = \frac{1}{2^{k+1}}$  pour tout  $i \geq 1$  définit une loi de probabilité sur  $\mathbb{N}$ .

**THÉORÈME 2.15.** — *Deux variables aléatoires réelles discrètes ont même loi si et seulement si elles ont même fonction de répartition.*

**DÉFINITION 2.16** (Loi conditionnelle). — *Si  $X$  est une variable aléatoire discrète et si  $A$  est un événement de probabilité non nulle, on appelle loi conditionnelle de  $X$  sachant  $A$  la loi de probabilité  $\mathbb{P}_A(X = \cdot)$  définie par :*

$$\forall i \in X(\Omega), \mathbb{P}_A(X = i) = \mathbb{P}(X = i | A) = \frac{\mathbb{P}((X = i) \cap A)}{\mathbb{P}(A)}$$

## 2.2.2 Espérance, moments et variance d'une variable aléatoire discrète

**DÉFINITION 2.17** (Espérance, cas fini). — *Si  $X$  est une variable aléatoire réelle prenant un nombre fini de valeurs, l'espérance de  $X$  est la quantité*

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

**EXEMPLE 2.18.** L'espérance d'une variable aléatoire réelle égale à  $x \in \mathbb{R}$  avec probabilité 1 est  $x$ .

**EXEMPLE 2.19.** L'espérance de la variable aléatoire dont la loi est donnée dans la remarque 2.13 est égale à

$$\mathbb{E}(X) = 0 \times \frac{1}{8} + 1 \times \frac{1}{4} + 2 \times \frac{5}{8} = \frac{3}{2}$$

**REMARQUE 2.20.** On verra plus loin, grâce à la loi des grands nombres, qu'il est possible d'interpréter l'espérance d'une variable aléatoire  $X$  comme le réel vers lequel tend<sup>3</sup> la moyenne des observations de  $X$  en opérant un grand nombre de fois des tirages indépendants de  $X$ . On appelle aussi l'espérance de  $X$  la *moyenne théorique* de  $X$ .

3. en un sens à préciser, bien sûr!

**DÉFINITION 2.21** (Espérance). — Soit  $X$  est une variable aléatoire réelle à valeurs dans  $\mathbb{N}$ . Si la série de terme général  $\mathbb{P}(X = k)$  converge, on définit l'espérance de  $X$  comme étant le réel  $\mathbb{E}(X)$  défini par :

$$\mathbb{E}(X) = \sum_{k=0}^{+\infty} k \mathbb{P}(X = k)$$

Si la série diverge, on dit que  $X$  n'admet pas d'espérance.

**EXEMPLE 2.22.** La variable aléatoire réelle  $X$  définie par  $X(\Omega) = \mathbb{N}^*$  et pour tout  $k \geq 1$ ,  $\mathbb{P}(X = k) = \frac{6}{k^2 \pi^2}$  (on admettra qu'il s'agit bien d'une loi de probabilité) n'admet pas d'espérance puisque la série  $\sum_{k \geq 0} k \mathbb{P}(X = k) = \sum_{k \geq 1} \frac{6}{k \pi^2}$  diverge.

**THÉORÈME 2.23** (Linéarité de l'espérance). — Si  $X$  et  $Y$  sont des variables aléatoires réelles discrètes admettant une espérance et si  $a \in \mathbb{R}$  alors  $aX + Y$  est une variable aléatoire réelle admettant une espérance et

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

$$\mathbb{E}(aX) = a \mathbb{E}(X)$$

Notez qu'aucune hypothèse n'est faite sur  $X$  et  $Y$  en-dehors de l'existence de leurs espérances respectives : en particulier, on ne demande pas à  $X$  et  $Y$  d'être indépendantes ! Cela rend la linéarité de l'espérance très puissante, comme l'illustre l'exercice suivant.

**Exercice 2.24.** Le Père Noël doit livrer  $n$  cadeaux à  $n$  enfants différents. Profitant d'un moment d'inattention de son maître, un lutin maléfique permute au hasard les adresses des enfants. Quelle est l'espérance du nombre de cadeaux correctement livrés à leur destinataire ?

**THÉORÈME 2.25** (Croissance de l'espérance). — Si  $X$  et  $Y$  sont deux variables aléatoires réelles (discrètes) admettant une espérance et si  $X \leq Y$ , alors  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ .

En particulier, si  $X$  est une variable aléatoire réelle discrète positive admettant une espérance, alors  $\mathbb{E}(X) \geq 0$ .

**THÉORÈME 2.26** (de transfert). — Si  $X$  est une variable aléatoire réelle prenant un nombre fini de valeurs et si  $f : X(\Omega) \rightarrow \mathbb{R}$  est une fonction, alors  $f(X)$  est une variable aléatoire réelle prenant un nombre fini de valeurs et

$$\mathbb{E}(f(X)) = \sum_{x \in X(\Omega)} f(x) \mathbb{P}(X = x)$$

On dispose d'un résultat similaire si  $f$  est positive dans le cas où  $X$  prend un nombre infini de valeurs entières sous réserve de la convergence de la série  $\sum_{k \geq 0} f(k) \mathbb{P}(X = k)$ , qui implique alors l'existence de  $\mathbb{E}(f(X))$  en même temps que l'égalité ci-dessus.

**REMARQUE 2.27.** Le théorème de transfert est extrêmement utile et il est très important de bien comprendre ce qu'il apporte. En reprenant les définitions ci-dessus, on pourrait entreprendre de calculer  $\mathbb{E}(f(X))$  en revenant à la définition et en écrivant

$$\mathbb{E}(f(X)) = \sum_{k \in f(X)(\Omega)} k \mathbb{P}(f(X) = k)$$

Toutefois, connaître  $\mathbb{P}(f(X) = k)$  pour tout  $k \in f(X)(\Omega)$  est généralement loin d'être aisé. Dans le cas où  $f$  est injective, en posant  $x = f^{-1}(k)$  on peut toutefois écrire

$$\mathbb{E}(f(X)) = \sum_{k \in f(X)(\Omega)} k \mathbb{P}(f(X) = k) = \sum_{x \in X(\Omega)} x \mathbb{P}(X = x)$$

Le théorème de transfert assure que cette formule est valable même quand  $f$  n'est pas injective, ce qui *a priori* n'a rien d'évident.

**REMARQUE 2.28.** On utilise très souvent le théorème de transfert avec la fonction carré  $f : x \mapsto x^2$ . Comme cette fonction est positive, la deuxième partie du théorème permet d'énoncer le résultat suivant : si  $X$  est une variable aléatoire réelle à valeurs dans  $\mathbb{N}$  telle que  $\sum_{k \geq 0} k^2 \mathbb{P}(X = k)$  converge, alors  $X^2$  admet une espérance, et

$$\mathbb{E}(X^2) = \sum_{k=0}^{+\infty} k^2 \mathbb{P}(X = k)$$

**EXEMPLE 2.29.** Si la loi de  $X$  est définie par le tableau de probabilités

|                      |                     |                     |
|----------------------|---------------------|---------------------|
| $\mathbb{P}(X = -1)$ | $\mathbb{P}(X = 0)$ | $\mathbb{P}(X = 1)$ |
| 1/4                  | 1/4                 | 1/2                 |

alors  $\mathbb{E}(\ln(X^2 + 1)) = \ln((-1)^2 + 1) \times \frac{1}{4} + \ln(0^2 + 1) \times \frac{1}{4} + \ln(1^2 + 1) \times \frac{1}{2} = \frac{3\ln(2)}{4}$ .

**THÉORÈME 2.30.** — Si  $X$  est une variable aléatoire à valeurs dans  $\mathbb{N}$ , alors pour tout  $n \in \mathbb{N}$  on a  $\mathbb{P}(X \leq n) = \sum_{k=0}^n \mathbb{P}(X = k)$  et  $\mathbb{P}(X \geq n) = \sum_{k=n}^{+\infty} \mathbb{P}(X = k)$ . De plus,  $X$  admet une espérance si et seulement si  $\sum_{n \geq 0} \mathbb{P}(X > n)$  converge, et dans ce cas

$$\mathbb{E}(X) = \sum_{n=0}^{+\infty} \mathbb{P}(X > n)$$

**Exercice 2.31.** Démontrer la proposition ci-dessus dans le cas où  $X$  prend un nombre fini de valeurs entières.

**DÉFINITION 2.32 (Moments).** — Si  $k \geq 1$  et si  $X$  est une variable aléatoire réelle discrète telle que  $X^k$  admet une espérance, on dit que  $X$  admet un moment d'ordre  $k$  et ce moment est la quantité  $\mathbb{E}(X^k)$ .

**REMARQUE 2.33.** Dans le cas  $X(\Omega) \subset \mathbb{N}$ , le fait que  $X^k$  admette une espérance équivaut par le théorème de transfert au fait que  $\sum_{n \geq 0} n^k \mathbb{P}(X = n)$  converge. On en déduit par exemple qu'une variable aléatoire réelle prenant un nombre fini de valeurs admet des moments de tous ordres.

**THÉORÈME 2.34.** — Si  $X$  est une variable aléatoire réelle discrète admettant un moment d'ordre 2, alors  $X$  admet une espérance et  $(X - \mathbb{E}(X))^2$  aussi.

Plus généralement, si  $X$  est une variable aléatoire réelle discrète admettant un moment d'ordre  $n \in \mathbb{N}^*$ , alors  $X$  admet un moment d'ordre  $p$  pour tout  $p \in \llbracket 1, n \rrbracket$ .

**DÉFINITION 2.35** (Variance). — Si  $X$  est une variable aléatoire réelle discrète admettant un moment d'ordre 2, on appelle variance de  $X$  le réel

$$V(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

**REMARQUE 2.36** (Positivité de l'espérance). Par croissance de l'espérance, une variance (qui est l'espérance d'une variable aléatoire positive) est toujours positive.

En tant que « moyenne théorique » des carrés des écarts d'une variable à sa valeur moyenne, la variance d'une variable est d'autant plus forte que celle-ci prend souvent des valeurs éloignées de sa moyenne. Il s'agit donc d'un indicateur de *dispersion* de la variable autour de sa moyenne. La proposition suivante est alors tout à fait naturelle.

**PROPOSITION 2.37.** — Une variable aléatoire réelle discrète  $X$  est constante si et seulement si elle admet une variance nulle.

**Exercice 2.38.** Montrer que si  $X$  et  $Y$  sont deux variables aléatoires réelles à valeurs dans  $\mathbb{N}$  admettant un moment d'ordre 2 et si  $a \in \mathbb{R}$ , alors  $XY$  admet une espérance et  $X + Y$  et  $aX$  admettent des moments d'ordre 2 (on utilisera l'inégalité  $2xy \leq x^2 + y^2$  valable pour tout  $(x, y) \in \mathbb{R}^2$ ).

### 2.2.3 Couples de variables aléatoires réelles discrètes

**DÉFINITION 2.39** (Loi jointe). — La loi jointe d'un couple de variables aléatoires réelles discrètes  $(X, Y)$  est la donnée de  $X(\Omega)$ , de  $Y(\Omega)$  et de  $\mathbb{P}((X = x) \cap (Y = y))$  pour tout  $(x, y) \in X(\Omega) \times Y(\Omega)$ .

**REMARQUE 2.40.** La loi jointe du couple  $(X, Y)$  contient bien plus d'informations que les seules lois de variables  $X$  et  $Y$  prises isolément (que l'on appelle *lois marginales du couple*). En effet, elle indique quelle est la propension de  $X$  et  $Y$  à prendre certaines valeurs simultanément, et elle permet de retrouver les lois marginales par les formules suivantes :

$$\begin{aligned} \forall x \in X(\Omega), \quad \mathbb{P}(X = x) &= \sum_{y \in Y(\Omega)} \mathbb{P}((X = x) \cap (Y = y)) \\ \forall y \in Y(\Omega), \quad \mathbb{P}(Y = y) &= \sum_{x \in X(\Omega)} \mathbb{P}((X = x) \cap (Y = y)) \end{aligned}$$

Si l'on représente la loi du couple dans un tableau (avec les lignes indexées par les valeurs de  $X$  et les colonnes par celles de  $Y$ ), la loi marginale de  $X$  se retrouve donc par sommation des coefficients des différentes lignes et celle de  $Y$  par sommation des coefficients des différentes colonnes.

**EXEMPLE 2.41.** La loi jointe du couple  $(X, Y)$  donnée par le tableau suivant :

| $Y \backslash X$ | 0    | 1    | 2    |
|------------------|------|------|------|
| -1               | 1/16 | 1/4  | 1/8  |
| 0                | 1/16 | 1/4  | 0    |
| 1                | 1/8  | 1/16 | 1/16 |

permet de retrouver les lois marginales de  $X$  et  $Y$  :  $X(\Omega) = \{0, 1, 2\}$ ,  $Y(\Omega) = \{-1, 0, 1\}$  et les sommations en ligne et en colonne donnent

$$\mathbb{P}(X=0) = \frac{1}{16} + \frac{1}{16} + \frac{1}{8} = \frac{1}{4} \quad \mathbb{P}(X=1) = \frac{1}{4} + \frac{1}{4} + \frac{1}{16} = \frac{9}{16} \quad \mathbb{P}(X=2) = \frac{1}{8} + \frac{1}{16} = \frac{3}{16}$$

$$\mathbb{P}(Y=-1) = \frac{1}{16} + \frac{1}{4} + \frac{1}{8} = \frac{7}{16} \quad \mathbb{P}(Y=0) = \frac{1}{16} + \frac{1}{4} = \frac{5}{16} \quad \mathbb{P}(Y=1) = \frac{1}{8} + \frac{1}{16} + \frac{1}{16} = \frac{1}{4}$$

Notez qu'il est utile, après avoir déterminé les lois marginales de chacune des variables du couple, de vérifier que la somme des probabilités obtenues pour chaque variable vaut bien 1 pour vérifier qu'une erreur de calcul n'a pas été commise.

**THÉORÈME 2.42.** — *Deux variables aléatoires réelles discrètes  $X$  et  $Y$  sont indépendantes si et seulement si on a*

$$\mathbb{P}((X=x) \cap (Y=y)) = \mathbb{P}(X=x)\mathbb{P}(Y=y)$$

*pour tout  $(x, y) \in X(\Omega) \times Y(\Omega)$ .*

**REMARQUE 2.43.** Cette relation se traduit sur un tableau donnant la loi jointe de  $(X, Y)$  par la colinéarité deux à deux des vecteurs lignes entre eux et des vecteurs colonnes entre eux. Par exemple, la loi jointe donnée par l'exemple 2.41 n'est pas la loi d'un couple de variables aléatoires indépendantes.

**THÉORÈME 2.44** (de transfert pour la loi jointe). — *Si  $X$  et  $Y$  sont deux variables aléatoires réelles prenant un nombre fini de valeurs et si  $f : X(\Omega) \times Y(\Omega) \rightarrow \mathbb{R}$  est une fonction, alors la variable aléatoire réelle  $f(X, Y)$  prend un nombre fini de valeurs et*

$$\mathbb{E}(f(X, Y)) = \sum_{x \in X(\Omega), y \in Y(\Omega)} f(x, y) \mathbb{P}((X=x) \cap (Y=y))$$

**THÉORÈME 2.45.** — *Si  $X$  et  $Y$  sont des variables aléatoires réelles discrètes indépendantes admettant chacune une espérance et telle que  $XY$  admet une espérance, alors  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ .*

*Si  $X$  et  $Y$  sont des variables aléatoires réelles discrètes indépendantes positives admettant chacune une espérance, alors  $XY$  admet une espérance et  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ .*

**EXEMPLE 2.46.** Nabuchodonosor, qui a besoin de cinquante chameaux pour mener à bien une affaire urgente, dit à Assurnasirpal : « si tu me donnes cinquante chameaux aujourd'hui, je lancerai un dé équilibré à six faces trois fois de suite et te donnerai demain autant de chameaux que le produit des résultats obtenus ». La question est de

savoir si Assurnasirpal, dénué d'aversion au risque et de préférence pour le présent, a intérêt à participer au jeu que lui propose Nabuchodonosor.

Il s'agit pour cela de calculer l'espérance de gain associée au jeu proposé par Assurnasirpal et la comparer à la perte certaine de 50 chameaux. Si l'on note  $X_1, X_2$  et  $X_3$  les résultats des trois lancers de dé indépendants qu'effectuera Nabuchodonosor, on a  $\mathbb{E}(X_i) = \frac{3}{2}$  pour tout  $i \in \llbracket 1, 3 \rrbracket$  et donc

$$\mathbb{E}(X_1 X_2 X_3) = \left(\frac{7}{2}\right)^3 = \frac{343}{8} < \frac{400}{8} = 50$$

Assurnasirpal n'a donc pas intérêt à participer au jeu qui lui est proposé.

**DÉFINITION 2.47** (Covariance). — Si  $X$  et  $Y$  sont des variables aléatoires réelles discrètes admettant un moment d'ordre 2, on définit la covariance de  $X$  et  $Y$  par

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

**REMARQUE 2.48.** La covariance de  $X$  et  $Y$  est un indicateur de la propension qu'ont les variables  $X$  et  $Y$  à évoluer « dans le même sens ». En effet, imaginons que  $X$  et  $Y$  aient tendance à prendre simultanément des valeurs élevées ou des valeurs faibles : alors  $X - \mathbb{E}(X)$  (position de  $X$  par rapport à sa « moyenne ») est, avec une grande probabilité, positive lorsque  $Y - \mathbb{E}(Y)$  l'est, et négative lorsque  $Y - \mathbb{E}(Y)$  l'est. La variable aléatoire  $(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))$  est donc positive avec une forte probabilité, et il est raisonnable de supposer que  $\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$  est positive. De la même façon, si  $X$  et  $Y$  ont tendance à évoluer « en sens opposé », la variable aléatoire  $(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))$  a tendance à prendre des valeurs négatives avec une forte probabilité et on a de bonnes raisons de penser que son espérance est négative.

Toutefois, comme on le verra un peu plus bas, l'étude de la covariance nous permet seulement de donner le signe (positif ou négatif) de la corrélation de  $X$  et  $Y$ , et non son degré.

**PROPOSITION 2.49** (Formule de Koenig). — Si  $X$  et  $Y$  sont des variables aléatoires réelles discrètes admettant un moment d'ordre 2, alors

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

En particulier,

$$V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

**Exercice 2.50.** Démontrer la formule de Koenig.

**REMARQUE 2.51.** Si  $X$  et  $Y$  sont deux variables aléatoires réelles discrètes indépendantes (et admettent chacune un moment d'ordre 2), la formule  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$  implique que  $\text{Cov}(X, Y) = 0$ . Toutefois, il est très important de remarquer que deux variables aléatoires peuvent avoir une covariance nulle sans être indépendantes : ce point est une source d'erreurs (et de problèmes de concours) très commune. Un exemple de couple  $(X, Y)$  de variables aléatoires de covariance nulle (on dit alors *non corrélées*) mais non indépendantes est donné par la loi jointe suivante :

| $Y \setminus X$ | -1  | 0   | 1   |
|-----------------|-----|-----|-----|
| -1              | 1/8 | 0   | 1/8 |
| 0               | 0   | 1/2 | 0   |
| 1               | 1/8 | 0   | 1/8 |

En effet :

- X et Y ne sont pas indépendantes puisque les lignes du tableau ci-dessus ne sont pas des vecteurs deux à deux colinéaires. Toutefois, on peut apporter un argument plus direct en revenant à la définition de l'indépendance de deux variables aléatoires réelles discrètes : par exemple, on peut remarquer que  $\mathbb{P}((X = 0) \cap (Y = 1)) = 0 \neq \frac{1}{2} \times \frac{1}{4} = \mathbb{P}(X = 0)\mathbb{P}(Y = 1)$  (on laissera au lecteur le soin de déterminer lui-même les lois marginales de X et Y).
- $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0 - 0 \times 0 = 0$  (on laisse une fois encore au lecteur le soin de déterminer la loi de XY « à la main » à partir du tableau ci-dessus ainsi que de calculer les espérances des différentes variables).

On voit donc qu'il est possible que deux variables soient fortement dépendantes l'une de l'autre (dans notre exemple, le fait que Y prenne certaines valeurs *empêche* purement et simplement X de prendre des valeurs qu'elle peut tout à fait prendre dans d'autres cas) sans pour autant être corrélées. Ce phénomène, dont l'analogue empirique est présent en statistiques, est dû aux diverses influences de Y sur X, qui peuvent se compenser et rendre invisible par une démarche statistique trop sommaire d'importantes dépendances entre les variables en jeu. Bref, la faible corrélation des variables à laquelle s'arrêtent trop facilement certains journalistes/sociologues/économistes est un indicateur à prendre avec précaution... même s'il en est fait usage abondamment dans des outils d'estimation aussi fondamentaux que la régression linéaire, comme nous le verrons dans notre cours sur le sujet.

Rappelons une dernière fois qu'à l'inverse, deux variables corrélées entre elles sont nécessairement non indépendantes<sup>4</sup>.

**PROPOSITION 2.52.** — Soient  $X, Y, Z$  trois variables aléatoires réelles discrètes admettant un moment d'ordre 2 et  $a \in \mathbb{R}$ . Alors

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

$$\text{Cov}(X, aY) = \text{Cov}(aX, Y) = a\text{Cov}(X, Y)$$

$$\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$$

$$\text{Cov}(X, X) = V(X)$$

En utilisant les notations ci-dessus, on remarque que  $\text{Cov}(aX, aY) = a^2\text{Cov}(X, Y)$ . Par exemple, la covariance entre  $2X$  et  $2Y$  est quatre fois plus forte que la covariance entre  $X$  et  $Y$ . Pourtant, le degré de corrélation « intrinsèque » entre  $2X$  et  $2Y$  est le même qu'entre  $X$  et  $Y$  : dire que la covariance permet d'évaluer le degré d'une corrélation entre deux variables est donc erroné. Ceci nous poussera un peu plus loin à définir

4. ... ce qui, dans le cas de variables d'ordre sociologique ou économique, ne veut pas dire non plus — quitte à enfoncer des portes ouvertes — qu'il existe un lien de *causalité* entre elles!



le coefficient de corrélation de deux variables d'une manière plus astucieuse (et l'on vérifiera en tant qu'exercice que le coefficient ainsi défini est bien invariant par multiplication de l'une ou l'autre des variables par un réel positif non nul).

**REMARQUE 2.53.** Les lecteurs familiarisés avec les espaces euclidiens auront reconnu dans la covariance définie sur l'ensemble des variables aléatoires réelles discrètes un produit scalaire<sup>5</sup> et dans la variance la norme euclidienne associée... et ne seront donc pas étonnés des quelques résultats qui vont suivre.

Les autres lecteurs sont tout de même invités à remarquer les similitudes frappantes que possèdent les règles de calcul sur la covariance présentées ci-dessus avec les égalités  $xy = yx$ ,  $x.ay = ax.y = axy$  et  $x(y + z) = xy + xz$  valables pour  $a, x, y, z \in \mathbb{R}$ . Les règles qui suivent peuvent elles aussi être transcrites en termes de produit (la covariance jouant le rôle du produit et la variance celle de la mise au carré) pour être mémorisées plus aisément : par exemple, la proposition suivante est le pendant d'une identité remarquable bien connue.

**THÉORÈME 2.54.** — Si  $X$  et  $Y$  sont deux variables aléatoires réelles discrètes admettant un moment d'ordre 2, alors

$$V(X + Y) = V(X) + 2\text{Cov}(X, Y) + V(Y)$$

De façon plus générale, si  $X_1, \dots, X_n$  sont des variables aléatoires réelles discrètes admettant un moment d'ordre 2, alors

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) + 2 \sum_{\substack{i, j \in [1, n] \\ i < j}} \text{Cov}(X_i, X_j)$$

**Exercice 2.55.** Écrire la formule ci-dessus dans le cas  $n = 3$  et la comparer à son analogue donnant le développement de  $(x_1 + x_2 + x_3)^2$ , où  $x_1, x_2, x_3 \in \mathbb{R}$ .

**THÉORÈME 2.56.** — Si  $X$  et  $Y$  sont deux variables aléatoires réelles discrètes indépendantes admettant un moment d'ordre 2, alors

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

Plus généralement, si  $X_1, \dots, X_n$  sont  $n$  variables aléatoires réelles discrètes mutuellement indépendantes admettant un moment d'ordre 2, alors leur somme admet un moment d'ordre 2 et vérifie

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i)$$

**REMARQUE 2.57.** Contrairement à ce que l'on observe dans le cas de l'espérance, les deux dernières égalités ci-dessus sont fausses en général si  $X$  et  $Y$  ne sont pas indépendantes (comme l'indique le théorème 2.54) : par exemple, en prenant  $X$  non constante

5. ...si l'on accepte la convention, classique, selon laquelle on est en droit d'identifier en tant que fonctions deux variables aléatoires égales avec probabilité 1.

et admettant une variance (par exemple n'importe quelle variable aléatoire prenant exactement deux valeurs de manière équiprobable) et  $Y = -X$ , on obtient  $V(X + Y) = V(0) = 0 \neq 2V(X) = V(X) + V(Y)$ . Attention à bien souligner l'indépendance des variables en jeu lorsque vous utilisez cette proposition!

L'inégalité suivante, appelée *inégalité de Cauchy-Schwarz*<sup>6</sup>, est très importante.

**THÉORÈME 2.58** (Inégalité de Cauchy-Schwarz). — Si  $X$  et  $Y$  sont deux variables aléatoires réelles discrètes admettant un moment d'ordre 2, alors

$$|\text{Cov}(X, Y)| \leq \sqrt{V(X)}\sqrt{V(Y)}$$

De plus, si  $X$  et  $Y$  sont non constantes on a  $\text{Cov}(X, Y) = -\sqrt{V(X)}\sqrt{V(Y)}$  (resp.  $\text{Cov}(X, Y) = \sqrt{V(X)}\sqrt{V(Y)}$ ) si et seulement s'il existe  $a < 0$  (resp.  $a > 0$ ) et  $b \in \mathbb{R}$  tels que  $\mathbb{P}(Y = aX + b) = 1$ , c'est-à-dire qu'il existe une relation de dépendance linéaire négative (resp. positive) entre  $X$  et  $Y$  avec probabilité 1.

Il est fréquent que l'on identifie en tant que fonctions deux variables aléatoires égales l'une à l'autre avec probabilité 1, et vous pouvez sans dommage remplacer dans une copie la relation  $\mathbb{P}(Y = aX + b) = 1$  ci-dessus par la relation  $Y = aX + b$ .

**DÉFINITION 2.59** (Coefficient de corrélation). — Si  $X$  et  $Y$  sont deux variables aléatoires réelles discrètes admettant un moment d'ordre 2 et de variance non nulle, on définit le coefficient de corrélation entre  $X$  et  $Y$  par

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

**REMARQUE 2.60.** D'après l'inégalité de Cauchy-Schwarz, on a alors  $\rho_{X,Y} \in [-1, 1]$ , et  $\rho_{X,Y} = -1$  (resp.  $\rho_{X,Y} = 1$ ) si et seulement s'il existe  $a, b \in \mathbb{R}$  avec  $a < 0$  (resp.  $a > 0$ ) tels que  $\mathbb{P}(X = aY + b) = 1$ .

## 2.3 Exemples de lois discrètes

Tous les exemples qui suivent sont à connaître par cœur; il est primordial de comprendre les interprétations des différentes lois et de savoir en retrouver l'espérance et la variance par le calcul.

**DÉFINITION 2.61** (Loi de Dirac). — Si  $x \in \mathbb{R}$ , on dit qu'une variable  $X$  suit une loi de Dirac  $\delta_x$  si et seulement si  $X(\Omega) = x$  (on a alors  $\mathbb{P}(X = x) = 1$ ). Comme on l'a vu, on s'autorise alors à noter  $X = x$ .

**DÉFINITION 2.62** (Loi uniforme). — Si  $A$  est un sous-ensemble fini de  $\mathbb{R}$ , on dit qu'une variable  $X$  suit une loi uniforme sur  $A$ , et on note  $X \sim \mathcal{U}(A)$ , si et seulement si  $X(\Omega) = A$  et  $\mathbb{P}(X = x) = \frac{1}{\text{card}(A)}$  pour tout  $x \in A$ .

6. Hermann Amandus Schwarz (1843 - 1921), mathématicien allemand. À ne surtout pas confondre avec le mathématicien français Laurent Schwartz (1915 - 2002), dont le nom ne s'écrit pas de la même façon!

Une variable suivant une loi uniforme sur une partie finie  $A \subset \mathbb{R}$  modélise l'issue d'un tirage équiprobable parmi les éléments de  $A$ .

**DÉFINITION 2.63** (Loi de Bernoulli). — Si  $p \in [0, 1]$ , on dit que  $X$  suit la loi de Bernoulli de paramètre  $p$  et on note  $X \sim \text{Bern}(p)$  (ou parfois  $X \sim \mathcal{B}(p)$ ) si et seulement si  $X(\Omega) = \{0, 1\}$  et  $\mathbb{P}(X = 1) = p$  (et alors  $\mathbb{P}(X = 0) = 1 - p$ ).

Notons – c'est important! – que le nom de la famille Bernoulli ne prend qu'un seul « i »...

Une variable suivant une loi de Bernoulli de paramètre  $p \in [0, 1]$  modélise l'issue d'une expérience aléatoire binaire ayant une probabilité  $p$  de succès.

**DÉFINITION 2.64** (Loi binomiale). — Si  $n \in \mathbb{N}^*$  et  $p \in [0, 1]$ , on dit que  $X$  suit la loi binomiale de paramètres  $n$  et  $p$  et on note  $X \sim \mathcal{B}(n, p)$  si et seulement si  $X(\Omega) = \llbracket 0, n \rrbracket$  et pour tout  $k \in \llbracket 0, n \rrbracket$  on a  $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ .

La formule du binôme de Newton assure que la somme des probabilités ci-dessus est bien égale à 1. Vérifiez-le!

**REMARQUE 2.65.** Une variable suivant une loi binomiale de paramètres  $n \in \mathbb{N}^*$  et  $p \in [0, 1]$  modélise le nombre de succès obtenus au cours de  $n$  réalisations indépendantes d'une expérience aléatoire ayant une probabilité  $p$  de succès. La proposition qui suit est la formalisation mathématique de cette remarque.

**PROPOSITION 2.66.** — Une variable aléatoire  $X$  suit une loi binomiale de paramètres  $n \in \mathbb{N}^*$  et  $p \in [0, 1]$  si et seulement si elle s'écrit comme la somme de  $n$  variables aléatoires indépendantes suivant chacune une loi de Bernoulli de paramètre  $p$ .

**PROPOSITION 2.67.** — Soient  $n, m \in \mathbb{N}^*$  et  $p \in [0, 1]$ . Si  $X \sim \mathcal{B}(n, p)$  et  $Y \sim \mathcal{B}(m, p)$  sont indépendantes, alors  $X + Y \sim \mathcal{B}(n + m, p)$ .

**DÉFINITION 2.68** (Loi géométrique). — Si  $p \in ]0, 1]$ , on dit que  $X$  suit la loi géométrique de paramètre  $p$ , et on note  $X \sim \mathcal{G}(p)$ , si et seulement si  $X(\Omega) = \mathbb{N}^*$  et pour tout  $k \in \mathbb{N}^*$  on a  $\mathbb{P}(X = k) = p(1 - p)^{k-1}$ .

Une variable suivant une loi géométrique de paramètre  $p \in ]0, 1]$  modélise le rang d'obtention du premier succès lorsque l'on répète des expériences aléatoires indépendantes ayant chacune une probabilité de succès égale à  $p$ .

**REMARQUE 2.69.** Il arrive que certains auteurs (en particulier anglo-saxons) qualifient de loi géométrique de paramètre  $p$  la loi du nombre d'échecs avant le premier succès (toujours lors de la répétition d'expériences indépendantes ayant une probabilité  $p$  de succès) et non du rang du premier succès. On a alors  $X(\Omega) = \mathbb{N}$  et pour tout  $k \in \mathbb{N}$ ,  $\mathbb{P}(X = k) = p(1 - p)^k$ ; on laisse le soin au lecteur de déterminer l'espérance et la variance de  $X$  dans ce cas. Il est donc recommandé d'être particulièrement attentif à la définition adoptée par l'énoncé si celui-ci introduit la loi géométrique. S'il se contente d'y faire référence comme à un point du programme, on adoptera la définition standard en France qui est celle donnée ici.

**Exercice 2.70** (Absence de mémoire de la loi géométrique). Soit  $p \in ]0, 1]$  et  $X$  une variable aléatoire discrète suivant la loi géométrique de paramètre  $p$ . Montrer que si

$k \in \mathbb{N}$  et  $n \in \mathbb{N}$  vérifient  $k \leq n$ , alors

$$\mathbb{P}(X > n \mid X > k) = \mathbb{P}(X > n - k)$$

et interpréter le résultat obtenu pour justifier le fait que l'on qualifie cette relation de « propriété d'absence de mémoire ».

**DÉFINITION 2.71** (Loi de Poisson). — Si  $\lambda \in \mathbb{R}_+^*$ , on dit que  $X$  suit la loi de Poisson de paramètre  $\lambda$  et on note  $X \sim \mathcal{P}(\lambda)$  si et seulement si  $X(\Omega) = \mathbb{N}$  et pour tout  $k \in \mathbb{N}$  on a  $\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ .

Le fait que la somme des probabilités ci-dessus est égale à 1 est assurée par l'égalité  $e^x = \sum_{k=0}^{+\infty} \frac{x^k}{k!}$  valable pour tout  $x \in \mathbb{R}$ . Vérifiez-le!

L'interprétation de la loi de Poisson est moins évidente; il suffit de savoir qu'elle est souvent utilisée pour modéliser des flux, notamment en raison de la propriété de stabilité exposée dans l'exercice suivant, qui permet de décrire la façon dont deux flux indépendants modélisés par des lois de Poisson s'additionnent.

**Exercice 2.72** (Stabilité de la loi de Poisson). Soient  $\lambda, \mu \in \mathbb{R}_+^*$  et  $X$  et  $Y$  deux variables aléatoires réelles indépendantes de lois respectives  $\mathcal{P}(\lambda)$  et  $\mathcal{P}(\mu)$ . Montrer que  $X + Y \sim \mathcal{P}(\lambda + \mu)$ .

**REMARQUE 2.73.** De manière plus générale, si  $X_1, \dots, X_n$  sont des variables aléatoires réelles mutuellement indépendantes de lois respectives  $\mathcal{P}(\lambda_1), \dots, \mathcal{P}(\lambda_n)$  (avec  $\lambda_1, \dots, \lambda_n \in \mathbb{R}_+^*$ ), alors  $X_1 + \dots + X_n \sim \mathcal{P}(\lambda_1 + \dots + \lambda_n)$ .

Le tableau 1 présente les caractéristiques de base (ensemble des valeurs possibles, définition, espérance et variance) des lois discrètes usuelles que nous venons de présenter. Il est nécessaire de le connaître parfaitement. Il convient aussi de savoir retrouver par le calcul l'intégralité de ses deux dernières colonnes : cela vous aidera à déterminer les différents moments de lois proches des lois de référence qui y sont présentées, vous permettra d'éviter certaines erreurs de mémorisation pénalisantes et peut même faire l'objet de véritables questions de cours!

**SYNTHÈSE 2.74.** Les principales lois des variables aléatoires réelles discrètes sont présentées dans le tableau suivant.

| Loi                    | $X(\Omega)$    | $\mathbb{P}(X = k)$ (avec $k \in X(\Omega)$ ) | $\mathbb{E}(X)$ | $V(X)$             |
|------------------------|----------------|---|-----------------|--------------------|
| $\delta_x$             | $\{x\}$        | 1 si $k = x$ , 0 sinon                        | $x$             | 0                  |
| $\mathcal{U}([1, n])$  | $[1, n]$       | $\frac{1}{n}$                                 | $\frac{n+1}{2}$ | $\frac{n^2-1}{12}$ |
| Bern( $p$ )            | $\{0, 1\}$     | $p$ si $k = 1$ , $1 - p$ si $k = 0$           | $p$             | $p(1 - p)$         |
| $\mathcal{B}(n, p)$    | $[0, n]$       | $\binom{n}{k} p^k (1 - p)^{n-k}$              | $np$            | $np(1 - p)$        |
| $\mathcal{G}(p)$       | $\mathbb{N}^*$ | $p(1 - p)^{k-1}$                              | $\frac{1}{p}$   | $\frac{1-p}{p^2}$  |
| $\mathcal{P}(\lambda)$ | $\mathbb{N}$   | $e^{-\lambda} \frac{\lambda^k}{k!}$           | $\lambda$       | $\lambda$          |

Tableau 1 – Définition, espérance et variance des lois discrètes au programme

**Exercice 2.75.** Retrouvez par le calcul les deux dernières colonnes du tableau ci-dessus.

### 3 Variables aléatoires à densité

On présente dans ce chapitre un type de variables aléatoires qui ne relève pas du cadre formel général que nous avons fixé plus haut. L'enjeu est de modéliser des phénomènes aléatoires pouvant prendre n'importe quelle valeur dans un intervalle de  $\mathbb{R}$ , alors que les variables aléatoires réelles discrètes sont nécessairement à valeurs dans un ensemble plus « petit », c'est-à-dire avec moins d'éléments<sup>7</sup>.

Dans ce qui suit, on fera très rarement référence à  $\Omega$ ; de manière générale, on se dispensera de donner à notre exposé un cadre formel aussi rigoureux que celui que nous avons utilisé dans le cas des variables discrètes. En effet, il est tout à fait suffisant dans un cadre appliqué de concevoir l'objet « variable aléatoire à densité » comme une véritable variable au comportement aléatoire et de ne pas chercher à formaliser cette notion outre mesure<sup>8</sup>.

#### 3.1 Généralités sur les variables aléatoires réelles à densité

**DÉFINITION 3.1** (Densité). — Une fonction  $f : \mathbb{R} \rightarrow \mathbb{R}$  est appelée fonction de densité (ou plus simplement densité) si et seulement si elle est positive, continue par morceaux et vérifie

$$\int_{-\infty}^{+\infty} f(t) dt = 1$$

Une variable aléatoire réelle  $X$  est dite à densité s'il existe une fonction de densité  $f$  telle que si  $a, b \in \mathbb{R}$  sont tels que  $a \leq b$  alors

$$\mathbb{P}(X \in [a, b]) = \int_a^b f(t) dt$$

La fonction  $f$  est alors appelée une densité de la loi de  $X$ .

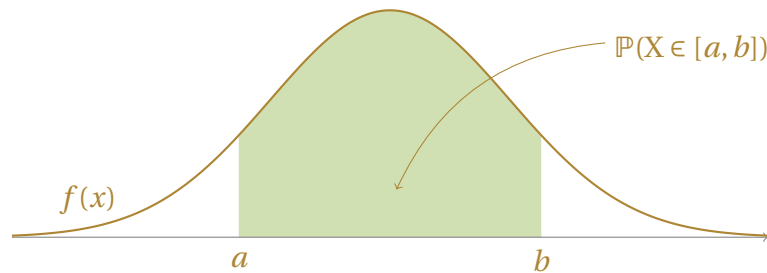


FIGURE 3 – Illustration du calcul de  $\mathbb{P}(X \in [a, b])$  pour une variable aléatoire  $X$  ayant pour densité  $f$ .

7. On peut en effet montrer que  $\mathbb{N}$  contient strictement moins d'éléments que n'importe quel intervalle de longueur non nulle de  $\mathbb{R}$  au sens où il n'existe pas d'injection de cet intervalle dans  $\mathbb{N}$ .

8. ... mais il existe bel et bien un cadre formel rigoureux justifiant tous les résultats présentés ici. Il ressemble fortement à celui que nous avons présenté dans la sous-section 2.1 mais met en jeu des subtilités peu pertinentes dans le cadre de l'agrégation de SES.

**REMARQUE 3.2.** Pour que la condition ci-dessus soit vérifiée, il faut et il suffit que l'on ait, pour tout  $x \in \mathbb{R}$ , l'égalité suivante :

$$\mathbb{P}(X \leq x) = \int_{-\infty}^x f(t) dt$$

On définit la fonction de répartition d'une variable à densité comme dans le cas discret (voir 2.9) ; elle vérifie elle aussi les propriétés exposées en 2.10.

**EXEMPLE 3.3.** La fonction  $x \mapsto \frac{1}{\pi(1+x^2)}$  définie sur  $\mathbb{R}$  est une fonction de densité. En effet, elle est continue, positive et son intégrale sur  $] -\infty, +\infty[$  vaut  $\frac{1}{\pi} \left( \frac{\pi}{2} - \left( -\frac{\pi}{2} \right) \right) = 1$  (on peut admettre ce résultat ou convoquer d'antiques souvenirs de trigonométrie – hors du programme de l'agrégation – pour le démontrer).

Lorsqu'une variable aléatoire admet cette fonction pour densité, on dit qu'elle suit une *loi de Cauchy*<sup>9</sup>.

On voit pourquoi il est nécessaire qu'une fonction de densité soit d'intégrale sur  $\mathbb{R}$  égale à 1 : en effet, cette quantité doit être égale à  $\mathbb{P}(X \in \mathbb{R}) = 1$ .

**DÉFINITION 3.4** (Loi d'une variable à densité). — *La loi d'une variable aléatoire réelle  $X$  à densité est la donnée de  $\mathbb{P}(X \in [a, b])$  pour tout couple  $(a, b) \in \mathbb{R}$  vérifiant  $a \leq b$ .*

**REMARQUE 3.5.** En d'autres termes, deux variables aléatoires réelles à densité  $X$  et  $Y$  ont même loi si et seulement si  $\mathbb{P}(X \in [a, b]) = \mathbb{P}(Y \in [a, b])$  pour tout couple  $(a, b) \in \mathbb{R}$  vérifiant  $a \leq b$ . Le théorème suivant affirme que la fonction de répartition d'une variable aléatoire réelle à densité caractérise sa loi (ce qui, moralement, revient à dire qu'il « suffit de prendre  $a = -\infty$  »).

**THÉORÈME 3.6.** — *Deux variables aléatoires réelles à densité ont même loi si et seulement si elles ont la même fonction de répartition.*

**THÉORÈME 3.7.** — *Si  $X$  est une variable aléatoire réelle admettant une densité  $f$ , la fonction de répartition  $F$  de  $X$  s'écrit*

$$F : x \mapsto \int_{-\infty}^x f(t) dt$$

*En particulier,  $F$  est continue sur  $\mathbb{R}$ . De plus, si  $f$  est continue en  $x \in \mathbb{R}$  alors  $F$  est dérivable en  $x$  et  $F'(x) = f(x)$ .*

En particulier, si  $X$  admet une densité continue  $f$ ,  $F$  est de classe  $\mathcal{C}^1$  sur  $\mathbb{R}$  et on a  $F' = f$ . En fait, on a mieux :

**PROPOSITION 3.8.** — *Si la fonction de répartition  $F$  d'une variable aléatoire réelle  $X$  est de classe  $\mathcal{C}^1$  sur  $\mathbb{R}$ , alors  $X$  admet  $F'$  pour densité.*

9. Cette loi n'est pas au programme mais nous permettra de donner des contre-exemples intéressants.

On utilise souvent cette remarque pour trouver la densité d'une variable aléatoire dont on connaît la fonction de répartition.

**EXEMPLE 3.9.** Soient  $X$  une variable aléatoire suivant une loi de Cauchy et  $a > 0$ . On se propose de montrer que  $aX$  admet une densité et de la calculer (en utilisant la fonction hors-programme arctan, primitive sur  $\mathbb{R}$  de la fonction  $t \mapsto \frac{1}{1+t^2}$  prenant la valeur  $\frac{1}{2}$  en 0).

Pour cela, on écrit que pour tout  $x \in \mathbb{R}$  on a

$$\mathbb{P}(aX \leq x) = \mathbb{P}\left(X \leq \frac{x}{a}\right) = \int_{-\infty}^{\frac{x}{a}} \frac{1}{\pi} \frac{1}{1+t^2} dt = \frac{1}{\pi} \left[ \arctan\left(\frac{x}{a}\right) - \frac{\pi}{2} \right]$$

On connaît donc la fonction de répartition de  $aX$  : il s'agit de  $F_a : x \mapsto \frac{1}{\pi} \left[ \arctan\left(\frac{x}{a}\right) - \frac{\pi}{2} \right]$ . On constate qu'elle est de classe  $\mathcal{C}^1$  sur  $\mathbb{R}$  : en utilisant la proposition précédente, on en déduit que  $aX$  admet pour densité la fonction  $F'_a$ , que nous noterons  $f_a$ . Ainsi,  $f_a(x) = F'_a(x) = \frac{1}{\pi} \frac{1}{a} \frac{1}{1+(\frac{x}{a})^2} = \frac{1}{\pi} \frac{a}{a^2+x^2}$ .

**PROPOSITION 3.10.** — Soit  $X$  une variable aléatoire réelle à densité. Alors pour tout  $(a, b) \in \mathbb{R}^2$  tels que  $a \leq b$ , on a  $\mathbb{P}(X = a) = 0$ , d'où l'on déduit

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

**REMARQUE 3.11.** La relation  $\mathbb{P}(X = a) = 0$  peut sembler étrange. En effet, bien qu'elle soit assez intuitive (par exemple, la probabilité de tomber sur 0 lorsque l'on choisit un réel « au hasard » dans  $[-1, 1]$  est plus petite que n'importe quel réel strictement positif, donc nulle), on peut avoir envie d'écrire ce qui suit :

$$1 = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}\left(\bigcup_{a \in \mathbb{R}} (X = a)\right) = \sum_{a \in \mathbb{R}} \mathbb{P}(X = a) = \sum_{a \in \mathbb{R}} 0 = 0$$

Toutefois, la propriété de la fonction  $\mathbb{P}$  à laquelle la troisième égalité fait référence n'est valable que pour des *suites* (c'est-à-dire un nombre dénombrable) d'événements disjoints, or  $\mathbb{R}$  est indénombrable ; sans compter que l'on n'a pas donné de sens à la notation  $\sum_{x \in \mathbb{R}} \dots$ .

**REMARQUE 3.12.** C'est pour pouvoir formaliser la notion de loi continue malgré la propriété  $\mathbb{P}(X = a) = 0$  que l'on a recours au concept de densité.

Si  $X$  admet une fonction de densité  $f$  continue en un point  $x \in \mathbb{R}$ , il est alors possible d'interpréter  $f(x)$  comme la « probabilité relative » que  $X$  prenne des valeurs voisines de  $x$ . Cependant, il est important de remarquer que  $f$  n'est pas nécessairement à valeurs dans  $[0, 1]$  ! C'est pourquoi nous parlons de « probabilité relative » : si  $x$  et  $y$  sont deux réels, la relation  $f(x) > f(y)$  signifie que  $X$  a une probabilité plus élevée de se trouver dans un voisinage de petite taille de  $x$  que de se trouver dans un voisinage de même taille de  $y$  et la relation  $f(x) \simeq 2f(y)$  implique que le rapport entre ces deux probabilités est proche de 2.

Voici une illustration graphique de cette idée :

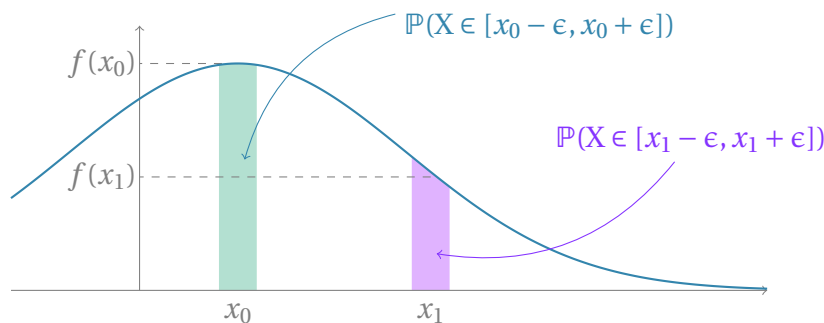


FIGURE 4 – La fonction de densité exprime une probabilité *relative* et non *absolue*.

Supposons que l'on ait  $f(x_0) \approx 2f(x_1)$ . Alors, la probabilité que  $X$  se trouve dans un petit intervalle autour de  $x_0$  est  $\mathbb{P}(X \in ]x_0 - \epsilon, x_0 + \epsilon[)$  (où  $\epsilon$  est petit). Cette quantité est égale à  $\int_{x_0-\epsilon}^{x_0+\epsilon} f(t)dt$  : c'est l'aire du rectangle de gauche sur la figure 4. Or, si  $f$  est continue, on peut s'attendre à ce que  $f(t)$  soit relativement proche de  $f(x_0)$  lorsque  $t$  est proche de  $x_0$  : on devrait donc avoir

$$\mathbb{P}(X \in ]x_0 - \epsilon, x_0 + \epsilon[) \approx 2\epsilon f(x_0)$$

Par le même raisonnement appliqué au point  $x_1$ , on a  $\mathbb{P}(X \in [x_1 - \epsilon, x_1 + \epsilon]) = \int_{x_1-\epsilon}^{x_1+\epsilon} f(t)dt$  (c'est l'aire du rectangle de droite sur la figure). On doit donc avoir l'approximation :

$$\mathbb{P}(X \in ]x_1 - \epsilon, x_1 + \epsilon[) \approx 2\epsilon f(x_1)$$

Comme nous avons fait l'hypothèse  $f(x_0) \sim 2f(x_1)$ , on a donc bien  $\mathbb{P}(X \in ]x_0 - \epsilon, x_0 + \epsilon[) \approx 2\mathbb{P}(X \in [x_1 - \epsilon, x_1 + \epsilon])$  : cela n'est que la traduction du fait que l'aire du rectangle de gauche est à peu près deux fois plus grande que l'aire de celui de droite.

La définition suivante est à comparer avec celle donnée en 2.2.2 dans le cas des variables discrètes :

**DÉFINITION 3.13** (Espérance d'une variable à densité). — Soit  $X$  une variable aléatoire réelle admettant une densité  $f$ . Si l'intégrale

$$\int_{-\infty}^{+\infty} tf(t)dt$$

est convergente<sup>10</sup>, on appelle espérance de  $X$  et on note  $\mathbb{E}(X)$  la quantité

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} tf(t)dt$$

Si  $\int_{-\infty}^{+\infty} tf(t)dt$  ne converge pas, on dit que  $X$  n'admet pas d'espérance.

**EXEMPLE 3.14.** Une variable aléatoire réelle suivant la loi de Cauchy n'admet pas d'espérance : en effet, l'intégrale  $\int_0^{+\infty} \frac{t}{\pi(1+t^2)} dt$  diverge.

10. Rappel : cela signifie que  $\int_0^{+\infty} tf(t)dt$  et  $\int_{-\infty}^0 tf(t)dt$  convergent.



L'espérance définie ci-dessus vérifie les mêmes propriétés de *croissance* et de *linéarité* que l'espérance définie dans le cas des variables discrètes.

**THÉORÈME 3.15** (de transfert). — Soit  $X$  une variable aléatoire réelle admettant une densité  $f$ . Si  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  est une fonction continue, et si l'intégrale

$$\int_{-\infty}^{+\infty} \phi(t) f(t) dt$$

est convergente, alors  $\phi(X)$  admet une espérance donnée par

$$\mathbb{E}(\phi(X)) = \int_{-\infty}^{+\infty} \phi(t) f(t) dt$$

Comme nous l'avons déjà vu dans le cas discret (voir le théorème 2.26 et la remarque 2.27), le théorème de transfert est un outil très puissant pour calculer l'espérance de certaines variables aléatoires sans avoir à calculer leur densité.

**EXEMPLE 3.16.** Soit  $U$  une variable aléatoire de loi uniforme sur  $[0, 1]$ . On pose  $X = U^3$ . Alors,  $\mathbb{E}[X] = \int_0^1 x^3 dx = \frac{1}{4}$ .

Enchaînons sur formule bien pratique et astucieuse, analogue dans le domaine continu de la formule énoncée dans la proposition 2.30 :

**PROPOSITION 3.17.** — Soit  $X$  une variable aléatoire réelle positive admettant une densité. Si  $X$  admet une espérance, alors  $\int_0^{+\infty} \mathbb{P}(X \geq t) dt$  converge et on a la formule suivante :

$$\mathbb{E}(X) = \int_0^{+\infty} \mathbb{P}(X \geq t) dt$$

On définit comme dans le cas des variables discrètes les notions de moments successifs (qui vérifient les mêmes propriétés d'existence que dans le cas discret) et de variance.

On définit enfin la notion de densité jointe pour un couple de variables aléatoires continues.

**DÉFINITION 3.18** (Densité jointe). — Soient  $X$  et  $Y$  deux variables aléatoires réelles. On dit que le couple  $(X, Y)$  admet pour densité jointe une fonction  $f_{(X,Y)} : \mathbb{R}^2 \rightarrow \mathbb{R}_+$  si le fait de considérer l'intégrale de  $f_{(X,Y)}$  sur  $\mathbb{R}^2$  a un sens (par exemple si  $f_{(X,Y)}$  est continue) et si

$$\mathbb{P}(X \in [a, b] \cap Y \in [c, d]) = \int_a^b \left( \int_c^d f_{(X,Y)}(x, y) dy \right) dx$$

pour tous  $(a, b) \in \mathbb{R}^2$  et  $(c, d) \in \mathbb{R}^2$  tels que  $a \leq b$  et  $c \leq d$ .

On dispose alors de résultats analogues à ceux obtenus dans le cas des variables aléatoires discrètes :

**PROPOSITION 3.19** (Marginales et théorème de transfert). — Dans le cadre de la définition précédente, les densités respectives (dites marginales) de  $X$  et  $Y$ , notées respectivement  $f_X$  et  $f_Y$ , sont données par :

$$\forall x \in \mathbb{R}, \quad f_X(x) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x, y) dy$$

et

$$\forall y \in \mathbb{R}, \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x, y) dx$$

Si  $F : \mathbb{R}^2 \rightarrow \mathbb{R}$  est une fonction telle que le fait de considérer l'intégrale de  $(x, y) \mapsto F(x, y)f(x, y)$  sur  $\mathbb{R}^2$  a un sens et si cette intégrale est finie, alors  $F(X, Y)$  admet une espérance donnée par

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

On définit la notion de covariance d'un couple de variables aléatoires comme dans le cas discret.

Tous les résultats, toutes les remarques et toutes les méthodes de calcul donnés en 2.2.3 à l'exception de la caractérisation de l'indépendance par les probabilités ponctuelles<sup>11</sup> sont encore valables dans le cas des variables à densité.

### 3.2 Exemples de lois à densité

**DÉFINITION 3.20** (Loi uniforme). — Si  $a, b \in \mathbb{R}$  sont tels que  $a < b$ , on dit qu'une variable aléatoire réelle suit la loi uniforme sur  $[a, b]$ , et on note  $X \sim \mathcal{U}([a, b])$ , si et seulement si elle admet pour densité la fonction

$$f : \begin{cases} \mathbb{R} \rightarrow \mathbb{R} \\ x \rightarrow \begin{cases} \frac{1}{b-a} \text{ lorsque } x \in [a, b] \\ 0 \text{ sinon.} \end{cases} \end{cases}$$

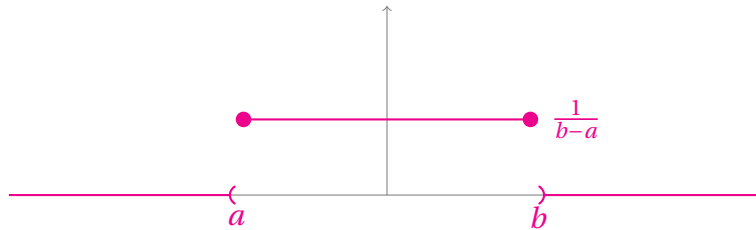


FIGURE 5 – Densité d'une variable aléatoire uniforme sur l'intervalle  $[a, b]$ .

11. En effet,  $\mathbb{P}((X = x) \cap (Y = y)) = 0 = \mathbb{P}(X = x)\mathbb{P}(Y = y)$  pour toutes variables aléatoires réelles à densité  $X$  et  $Y$  et pour tout  $(x, y) \in \mathbb{R}^2$ , ce qui montre que cette condition n'est pas suffisante pour assurer l'indépendance. Il existe cependant une propriété similaire utilisant la notion de densité jointe : les variables  $X$  et  $Y$  sont indépendantes si et seulement si la densité jointe  $f_{(X,Y)}$  du couple  $(X, Y)$  vérifie  $f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$  pour tous  $x, y \in \mathbb{R}$ .

**REMARQUE 3.21.** Une variable  $X$  suivant la loi uniforme sur  $[a, b]$  modélise le résultat d'un tirage aléatoire « uniforme sur  $[a, b]$  » au sens où la probabilité pour que le résultat soit compris dans un sous-intervalle de  $[a, b]$  dépend uniquement de la longueur de ce sous-intervalle :

$$\forall c, d \in \mathbb{R} \mid c < d \text{ et } [c, d] \subset [a, b], \mathbb{P}(X \in [c, d]) = \frac{d - c}{b - a}$$

**DÉFINITION 3.22** (Loi exponentielle). — Soit  $\lambda \in ]0, +\infty[$ . On dit qu'une variable aléatoire réelle suit la loi exponentielle de paramètre  $\lambda$ , et on note  $X \sim \mathcal{E}(\lambda)$ , si et seulement si elle admet pour densité la fonction

$$f : \begin{cases} \mathbb{R} \rightarrow \mathbb{R} \\ x \rightarrow \begin{cases} \lambda e^{-\lambda x} \text{ lorsque } x \geq 0 \\ 0 \text{ sinon.} \end{cases} \end{cases}$$

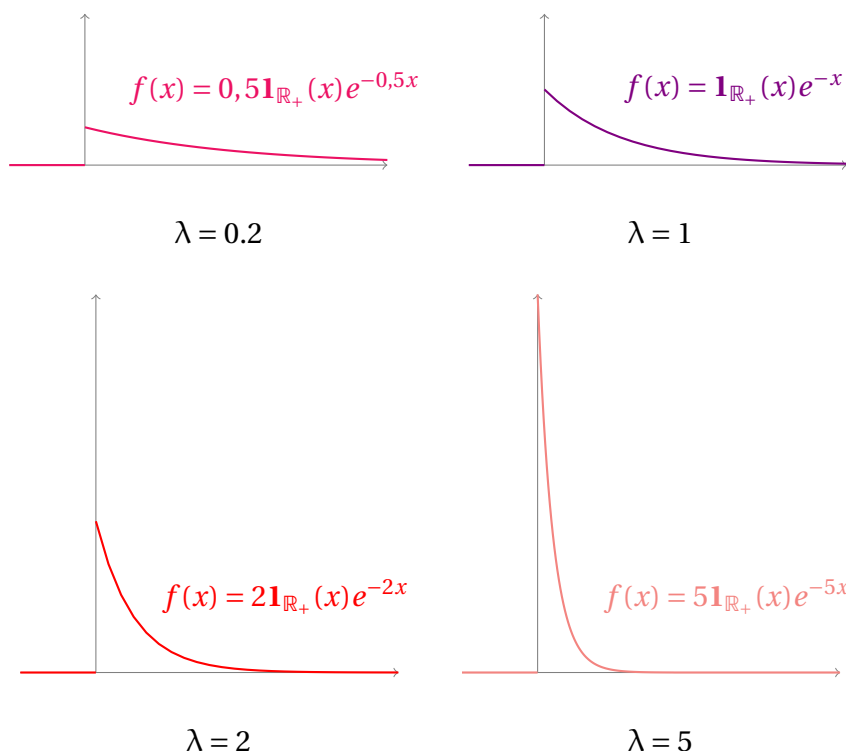


FIGURE 6 – Densité d'une variable aléatoire de loi exponentielle pour différentes valeurs du paramètre  $\lambda$ .

**REMARQUE 3.23.** La loi exponentielle est souvent utilisée pour modéliser des processus ayant une propriété d'absence de mémoire. En effet, on dispose de la propriété suivante, à rapprocher de celle établie pour la loi géométrique dans l'exercice 2.70 et dont la démonstration et l'interprétation sont similaires.

**PROPOSITION 3.24** (Absence de mémoire de la loi exponentielle). — Si  $r, s \in \mathbb{R}_+$  sont tels que  $r \leq s$ , si  $\lambda \in \mathbb{R}_+^*$  et si  $X$  est une variable aléatoire réelle suivant la loi  $\mathcal{E}(\lambda)$ , alors  $\mathbb{P}(X \geq s \mid X \geq r) = \mathbb{P}(X \geq s - r)$ .

**Exercice 3.25.** Soient  $\alpha, \lambda > 0$ . On considère une variable aléatoire réelle  $X$  suivant la loi  $\mathcal{E}(\lambda)$ . Déterminer la loi de  $\alpha X$ .

**DÉFINITION 3.26** (Loi normale centrée réduite). — On dit qu'une variable aléatoire réelle  $X$  suit une loi normale centrée réduite<sup>12</sup> (ou loi normale standard, ou encore loi gaussienne standard), et on note  $X \sim \mathcal{N}(0, 1)$ , lorsque  $X$  admet pour densité la fonction  $f$  définie sur  $\mathbb{R}$  par

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

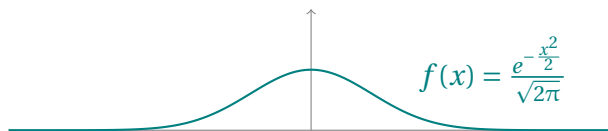


FIGURE 7 – La loi gaussienne centrée réduite.

**REMARQUE 3.27.** Le fait que la fonction  $f$  définie ci-dessus est bien une fonction de densité résulte de la formule de Gauss, admise<sup>13</sup> (et à connaître)

$$\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$$

**REMARQUE 3.28.** On note  $\Phi$  la fonction de répartition de la loi gaussienne standard, c'est-à-dire

$$\Phi : x \mapsto \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Il est important de savoir que  $\Phi$  ne s'exprime pas à l'aide des autres fonctions usuelles. Une propriété commode de  $\Phi$ , dont on laissera la démonstration en exercice<sup>14</sup>, est la relation  $\Phi(-x) = 1 - \Phi(x)$ . Notons enfin que  $\Phi$  est continue et strictement croissante donc bijective, et que l'on notera comme d'habitude  $\Phi^{-1}$  sa fonction réciproque.

**DÉFINITION 3.29** (Loi normale). — Soient  $\mu \in \mathbb{R}$  et  $\sigma \in \mathbb{R}_+^*$ . On dit qu'une variable aléatoire réelle  $X$  suit une loi normale de paramètres  $\mu$  et  $\sigma^2$ , et on note  $X \sim \mathcal{N}(\mu, \sigma^2)$ , si et seulement si  $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$ . En d'autres termes,  $X \sim \mathcal{N}(\mu, \sigma^2)$  si et seulement s'il existe une variable aléatoire réelle de loi gaussienne standard notée  $N$  telle que  $X = \sigma N + \mu$ .

**REMARQUE 3.30.** Il est impératif de bien comprendre la définition ci-dessus, qui implique notamment qu'il est toujours possible lorsque l'on étudie une loi normale de paramètres donnés de se ramener à l'étude de la loi normale standard et donc de faire apparaître la fonction  $\Phi$ .

L'intérêt d'une telle manœuvre apparaîtra dans le dernier chapitre et tient au fait que la fonction réciproque  $\Phi^{-1}$  est tabulée (c'est-à-dire que l'on dispose de tables donnant des valeurs de  $\Phi^{-1}$  et permettant de réaliser des calculs précis sans l'aide d'un outil informatique; vous pouvez d'ores et déjà retenir que  $\Phi^{-1}(0,95) \simeq 1,64$  et  $\Phi^{-1}(0,975) \simeq 1,96$ ).

12. Une loi est dite *centrée* lorsque l'espérance associée est nulle, et *réduite* lorsque la variance associée est égale à 1. On verra un peu plus loin que c'est effectivement le cas de la loi normale standard.

13. Sa démonstration fait l'objet d'un problème corrigé.

14. Un dessin devrait suffire!

**PROPOSITION 3.31.** — Si  $\mu \in \mathbb{R}$  et  $\sigma \in \mathbb{R}_+^*$ , la densité de la loi  $\mathcal{N}(\mu, \sigma^2)$  est la fonction  $f_{\mu, \sigma^2}$  définie sur  $\mathbb{R}$  par

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

**REMARQUE 3.32.** L'expression de  $f_{\mu, \sigma^2}$  ci-dessus n'est pas à apprendre par cœur, mais il faut savoir la retrouver<sup>15</sup> !

**REMARQUE 3.33.** Anticipons sur le tableau qui clôt ce chapitre : une variable aléatoire réelle de loi  $\mathcal{N}(\mu, \sigma^2)$  admet  $\mu$  pour espérance et  $\sigma^2$  pour variance. L'effet de ces différents paramètres sur la forme de la fonction de densité de la loi est à connaître !

Faire varier  $\mu$  translate horizontalement la courbe représentative de la fonction de densité, comme illustré dans la figure 8.

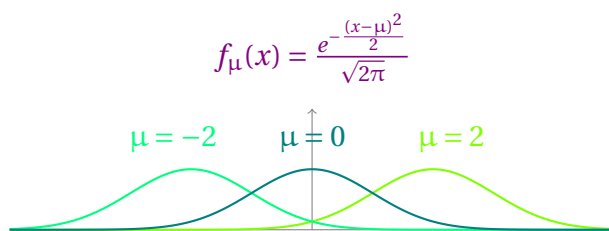


FIGURE 8 – Trois densités gaussiennes réduites, pour différentes valeurs de  $\mu$ .

Faire varier  $\sigma^2$  influe sur la dispersion des valeurs prises par la variable en élargissant d'autant plus la courbe représentative de la densité que  $\sigma^2$  est important. Lorsque  $\sigma^2$  est très faible, la courbe de densité a l'allure d'un pic concentré autour de  $\mu$  (voir figure 9).

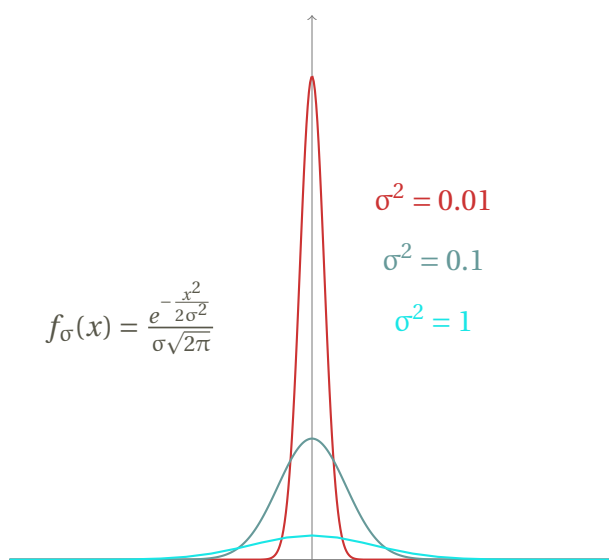


FIGURE 9 – Trois densités gaussiennes centrées, pour différentes valeurs de  $\sigma^2$ .

15. ... par la méthode désormais classique qui consiste à passer par la fonction de répartition et à effectuer un changement de variable dans l'intégrale.

**THÉORÈME 3.34** (Théorème de grande stabilité des lois normales). — Soient  $\mu, \mu' \in \mathbb{R}$  et  $\sigma, \sigma' \in \mathbb{R}_+^*$ . On se donne deux variables aléatoires indépendantes  $X \sim \mathcal{N}(\mu, \sigma^2)$  et  $Y \sim \mathcal{N}(\mu', \sigma'^2)$ . Alors,

$$X + Y \sim \mathcal{N}(\mu + \mu', \sigma^2 + \sigma'^2)$$

De plus, si  $a \in \mathbb{R}^*$ , alors

$$aX \sim \mathcal{N}(a\mu, a^2\sigma^2) \quad \text{et} \quad a + X \sim \mathcal{N}(a + \mu, \sigma^2)$$

**REMARQUE 3.35.** En anticipant sur le tableau 2, notons que les relations ci-dessus se retiennent aisément en se souvenant du fait que les lois normales sont stables par somme de variables indépendantes et par addition et multiplication avec un réel non nul, puis en calculant l'espérance et la variance des expressions  $X + Y$ ,  $aX$  et  $a + X$  pour trouver les paramètres des différentes lois normales obtenues.

**SYNTHÈSE 3.36.** Les principales lois à densité sont représentées dans le tableau suivant.

| Loi                          | Densité  | $\mathbb{E}(X)$     | $V(X)$                |
|------------------------------|--|---------------------|-----------------------|
| $\mathcal{U}([a, b])$        | $\frac{1}{b-a} \mathbf{1}_{[a, b]}(x)$                         | $\frac{a+b}{2}$     | $\frac{(b-a)^2}{12}$  |
| $\mathcal{E}(\lambda)$       | $\lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x)$          | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |
| $\mathcal{N}(0, 1)$          | $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$                     | 0                   | 1                     |
| $\mathcal{N}(\mu, \sigma^2)$ | $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ | $\mu$               | $\sigma^2$            |

Tableau 2 – Définition, espérance et variance des lois à densité au programme

**Exercice 3.37.** Retrouver par le calcul les deux dernières colonnes du tableau ci-dessus.

## 4 Théorèmes de convergence et estimation

On expose dans cette section des résultats de convergence importants dont on présente quelques applications classiques à des problèmes d'estimation. Ces deux parties constituent une introduction au chapitre crucial d'estimation paramétrique (abordé en détail et avec une technicité plus grande dans quelques séances) et font le lien entre le programme de probabilités et celui de statistiques.

### 4.1 Convergence d'une suite de variables aléatoires réelles

Dans le cours d'analyse, nous avons défini la notion de *convergence* des nombres réels. Ici, nous allons faire de même avec les variables aléatoires. C'est plus difficile : on peut définir la convergence de plusieurs manières, et chacune possède ses avantages et ses inconvénients. Seuls deux modes de convergence sont au programme : la convergence en probabilité et la convergence en loi.

#### 4.1.1 Convergence en probabilité et inégalités classiques.

**DÉFINITION 4.1** (Convergence en probabilité). — Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires réelles et  $X$  une variable aléatoire réelle. On dit que  $(X_n)_{n \in \mathbb{N}}$  converge en probabilité vers  $X$ , et on note  $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$ , lorsque pour tout  $\varepsilon > 0$  on a

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \xrightarrow[n \rightarrow +\infty]{} 0$$

**REMARQUE 4.2.** Dire que  $(X_n)_{n \in \mathbb{N}}$  converge en probabilité vers  $X$  signifie que la probabilité pour que les variables aléatoires  $X_n$  et  $X$  prennent des valeurs éloignées l'une de l'autre tend vers 0 lorsque  $n$  tend vers  $+\infty$ . On retiendra que ce type de convergence signifie que, dans un certain sens,  $X_n$  se « rapproche » de  $X$ .

**EXEMPLE 4.3.** Si  $(X_n)_{n \in \mathbb{N}^*}$  est une suite de variables aléatoires telle que pour tout  $n \in \mathbb{N}^*$  on ait  $X \sim \text{Bern}\left(\frac{1}{2^n}\right)$ , alors la suite  $(nX_n)_{n \in \mathbb{N}^*}$  converge en probabilité vers 0. En effet, pour tout  $\varepsilon > 0$  et tout  $n \geq 1$  on a

$$\mathbb{P}(|nX_n| \geq \varepsilon) \leq \mathbb{P}(X_n \neq 0) = \frac{1}{2^n} \xrightarrow[n \rightarrow +\infty]{} 0$$

où l'inégalité résulte du fait que  $nX_n$  ne prend des valeurs non nulles que lorsque c'est le cas de  $X_n$ .

Il est intéressant de noter que les valeurs non nulles prises par la variable  $nX_n$  sont d'autant plus grandes que  $n$  est grand : la convergence en probabilité signifie que ces grandes valeurs ne sont prises qu'avec une probabilité tendant vers 0.

Les deux inégalités suivantes sont très utiles pour démontrer des résultats de convergence en probabilité, ou plus généralement pour étudier les valeurs prises par les variables aléatoires. La première, l'inégalité de Markov<sup>16</sup>, est très simple :

16. 1856 - 1922, mathématicien russe. Il fut l'un des pionniers de la théorie moderne des probabilités.

**THÉORÈME 4.4** (Inégalité de Markov). — Si  $X$  est une variable aléatoire réelle positive admettant une espérance  $\mathbb{E}(X)$ , alors pour tout  $\varepsilon > 0$  on a

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}(X)}{\varepsilon}$$

**REMARQUE 4.5.** Insistons : il est nécessaire que  $X$  soit positive pour que l'inégalité de Markov soit vraie ; le cas d'une variable aléatoire réelle suivant une loi normale centrée réduite fournit un contre-exemple facile dans le cas contraire.

**EXEMPLE 4.6.** Soit  $X \sim \mathcal{N}(0, \sigma^2)$ . Alors, d'après l'inégalité de Markov appliquée à la variable aléatoire positive  $X^2$ , on a  $\mathbb{P}(|X| > \varepsilon) = \mathbb{P}(X^2 > \varepsilon^2) \leq \frac{\mathbb{V}(X)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}$ . La manipulation que nous venons de faire dans cet exemple est en fait à l'origine de l'inégalité suivante.

L'inégalité ci-dessous, dite de Bienaymé-Tchebychev<sup>17</sup>, se prouve à l'aide de l'inégalité de Markov.

**THÉORÈME 4.7** (Inégalité de Bienaymé-Tchebychev). — Si  $X$  est une variable aléatoire réelle (pas forcément positive) admettant un moment d'ordre 2, alors pour tout  $\varepsilon > 0$  on a

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\mathbb{V}(X)}{\varepsilon^2}$$

**REMARQUE 4.8.** L'inégalité ci-dessus permet d'obtenir une majoration de la probabilité qu'une variable  $X$  s'éloigne d'une certaine distance de son espérance, c'est-à-dire de sa valeur moyenne théorique. Le fait que cette probabilité soit contrôlée par la variance de  $X$ , d'autant plus faible que la dispersion de  $X$  l'est, est conforme à l'intuition !

#### 4.1.2 La loi des grands nombres.

Il y a plusieurs lois des grands nombres, mais toutes décrivent en substance le même phénomène : la convergence de la moyenne empirique vers la moyenne théorique. Le point qui différencie ces lois des grands nombres se situe dans les hypothèses et dans le mode de convergence. Nous n'énonçons ici que la loi *faible* des grands nombres, qui n'a de faible que le nom et qui est en fait un résultat très fort !

**DÉFINITION 4.9** (Moyenne empirique). — Si  $(X_n)_{n \in \mathbb{N}^*}$  est une suite de variables aléatoires réelles, on définit pour tout  $n \in \mathbb{N}^*$  la moyenne empirique des  $n$  premiers  $X_i$  comme la variable

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

**THÉORÈME 4.10** (Loi faible des grands nombres). — Si  $(X_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires mutuellement indépendantes admettant une variance bornée par une constante  $M$  et une même espérance  $\mu$ , alors

$$\overline{X}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \mu$$

17. Irénée-Jules Bienaymé (1796 - 1878), statisticien français, fut le premier à formuler cette inégalité, mais c'est son élève Pafnouti Tchebychev (1821-1894) qui l'a démontrée : cette inégalité est parfois simplement appelée « inégalité de Tchebychev », surtout dans la littérature anglo-saxonne.



**REMARQUE 4.11.** En particulier, si  $(X_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires mutuellement indépendantes et suivant chacune la même loi qu'une variable  $X$  (on parle alors de variables *iid*, pour « indépendantes et identiquement distribuées »<sup>18</sup>) admettant une variance, alors on a

$$\overline{X}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \mathbb{E}(X)$$

**REMARQUE 4.12.** La loi faible des grands nombres fut énoncée assez tôt par J. Bernoulli<sup>19</sup> dans son traité de probabilités *Ars conjectandi* (1713), mais ses formulations les plus générales et les démonstrations rigoureuses sont apparues à partir de la fin du XIX<sup>ème</sup> siècle.

La loi faible des grands nombres fait le lien entre moyenne empirique et moyenne théorique d'une suite de variables *iid* admettant une variance : « il y a convergence en probabilité de la moyenne empirique vers la moyenne théorique ».

C'est cette loi, d'une importance historique considérable, qui justifie que l'on introduise dans l'enseignement secondaire la notion d'espérance d'une variable aléatoire comme « la valeur moyenne que l'on obtiendra en répétant le tirage une infinité de fois », et plus généralement la notion de probabilité d'un événement comme celle de « fréquence asymptotique de réalisation d'un événement ». L'exemple ci-dessous formalise ce dernier point.

**EXEMPLE 4.13.** On considère une suite *iid*  $(X_n)_{n \geq 1}$  de variables suivant une loi de Bernoulli de paramètre  $p \in [0, 1]$  ; elles modélisent par exemple la réalisation d'un événement de probabilité  $p$  lié à une expérience aléatoire que l'on répète une infinité de fois, chaque expérience étant indépendante des précédentes. On a alors

$$\overline{X}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} p$$

Une illustration particulièrement claire de ce résultat est donnée par la réalisation d'un sondage : on interroge de manière indépendante  $n$  personnes choisies avec équiprobabilité dans une population dont 55% des membres comptent voter pour le candidat  $X$  aux prochains élections. La loi faible des grands nombres permet alors d'affirmer que lorsque  $n$  est suffisamment grand, la probabilité pour que la proportion de personnes interrogées comptant voter pour le candidat  $X$  soit comprise entre 54% et 56% est inférieure à 0,01, et qu'en augmentant à nouveau suffisamment  $n$  on peut remplacer, par exemple, « 54% et 56% » par « 54,99% et 55,01% » et « 0,01 » par « 0,00001 ».

Un autre exemple d'application de la loi faible des grands nombres est fourni par l'expérience consistant à lancer un grand nombre de fois de suite une pièce équilibrée : on peut affirmer que pour un nombre  $n$  de lancers suffisamment élevé, la proportion de « pile » obtenus se situera entre 49,9999% et 50,0001% avec une probabilité supérieure à 99,9999%.

18. ... mais vous n'êtes pas autorisés à utiliser cette abréviation dans une copie sauf si l'énoncé précise le contraire !

19. Jacques Bernoulli est l'un des innombrables mathématiciens de la dynastie des Bernoulli : cette famille suisse, issue des milieux commerçants belges, compte une dizaine d'hommes de sciences ayant vécu au XVII<sup>ème</sup> et XVIII<sup>ème</sup> siècle. Cette prolifération a permis à la famille de donner son nom à — entre autres — une loi de probabilité, un principe, une équation physique, une équation différentielle, une famille de nombres, un paradoxe, une famille de polynômes, un processus aléatoire, et une revue scientifique de probabilités.

Il est possible de déterminer les valeurs de  $n$  qui permettent de réaliser dans ce cadre une approximation d'une finesse donnée avec une probabilité donnée : ce sera l'objet de la sous-section suivante.

**Exercice 4.14.** Démontrer les inégalités de Markov et de Bienaymé-Tchebychev ainsi que la loi faible des grands nombres.

### 4.1.3 Convergence en loi.

**DÉFINITION 4.15** (Convergence en loi). — Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires réelles et  $X$  une variable aléatoire réelle. On dit que  $(X_n)_{n \in \mathbb{N}}$  converge en loi vers  $X$ , et on note

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$$

lorsque pour tout couple  $(a, b) \in \mathbb{R}^2$  tel que  $a < b$  on a

$$\mathbb{P}(X_n \in ]a, b[) \xrightarrow[n \rightarrow +\infty]{} \mathbb{P}(X \in ]a, b[)$$

**REMARQUE 4.16.** La condition ci-dessus n'implique pas que  $\mathbb{P}(X_n = a) \rightarrow \mathbb{P}(X = a)$  pour tout  $a \in \mathbb{R}$ . Toutefois, on verra tout à l'heure que cette dernière condition est nécessaire et suffisante dans le cas où toutes les variables sont à valeurs dans le même ensemble fini ou inclus dans  $\mathbb{N}$ .

**REMARQUE 4.17.** Contrairement à la convergence en probabilité, la convergence en loi n'indique pas la proximité des valeurs prises par les variables  $X_n$  et  $X$  lorsque  $n$  est grand mais plutôt la proximité des lois de  $X_n$  et  $X$  (d'où, d'ailleurs, le nom de convergence en loi...) : elle exprime le fait que  $X_n$  et  $X$  prennent une valeur donnée avec une fréquence comparable.

Il est clair intuitivement que lorsque la convergence en probabilité a lieu,  $X_n$  et  $X$ , qui ont tendance « à être au même endroit au même moment » lorsque  $n$  est grand, ont *a fortiori* des lois de probabilité proches l'une de l'autre. La proposition suivante, très importante, formalise cette remarque.

**PROPOSITION 4.18.** — Si  $(X_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires réelles convergeant en probabilité vers une variable aléatoire réelle  $X$ , alors la convergence a aussi lieu en loi.

**PROPOSITION 4.19.** — Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires réelles et  $X$  une variable aléatoire réelle. Pour tout  $n \in \mathbb{N}$ , on note  $F_n$  la fonction de répartition de  $X_n$  et on note  $F$  celle de  $X$ . Alors,

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$$

si et seulement si pour tout  $x \in \mathbb{R}$  tel que  $F$  est continue en  $x$ ,

$$F_n(x) \xrightarrow[n \rightarrow +\infty]{} F(x)$$

De plus, s'il existe un ensemble  $A$  fini ou inclus dans  $\mathbb{N}$  tel que toutes les variables  $X_n$  et  $X$  soient à valeurs dans  $A$ , alors on a  $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$  si et seulement si pour tout  $x \in A$  on a  $\mathbb{P}(X_n = x) \xrightarrow[n \rightarrow +\infty]{} \mathbb{P}(X = x)$ .

**REMARQUE 4.20.** On note que si  $X$  et  $Y$  sont deux variables aléatoires réelles de même loi et si  $(X_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires réelles vérifiant  $X_n \xrightarrow{\mathcal{L}} X$ , alors on a aussi  $X_n \xrightarrow{\mathcal{L}} Y$ , et si pour tout  $n \in \mathbb{N}$   $Y_n$  est une variable aléatoire réelle de même loi que  $X_n$  on a aussi  $Y_n \xrightarrow{\mathcal{L}} X$  et  $Y_n \xrightarrow{\mathcal{L}} Y$ .

En d'autres termes, la notion de convergence en loi d'une suite de variables aléatoires vers une variable aléatoire est en fait la convergence d'une suite de lois de probabilité vers une autre loi de probabilité et est indépendante du choix des variables suivant ces différentes lois. Il arrive d'ailleurs que l'on note la convergence en loi, mettons, de la suite  $(X_n)_{n \in \mathbb{N}}$  vers une variable aléatoire réelle suivant la loi normale standard sous la forme :

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

**Exercice 4.21.** Soient  $\lambda \in \mathbb{R}_+^*$  et  $(p_n)_{n \in \mathbb{N}}$  une suite de réels positifs telle que  $np_n \xrightarrow[n \rightarrow +\infty]{} \lambda$ . Montrer que si  $(X_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires réelles telles que pour tout  $n \in \mathbb{N}$  on ait  $X_n \sim \mathcal{B}(n, p_n)$ , alors  $(X_n)_{n \in \mathbb{N}}$  converge en loi vers une variable aléatoire réelle suivant une loi de Poisson de paramètre  $\lambda$ .

**PROPOSITION 4.22.** — Si  $(X_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires réelles et si  $X$  est une variable aléatoire réelle telles que  $X_n \xrightarrow{\mathcal{L}} X$  lorsque  $n \rightarrow +\infty$ , alors pour toute fonction  $f : \mathbb{R} \rightarrow \mathbb{R}$  continue on a  $f(X_n) \xrightarrow{\mathcal{L}} f(X)$ .

La loi faible des grands nombres donne, sous certaines conditions, la convergence en probabilité de la moyenne empirique d'une suite de variables aléatoires vers leur moyenne théorique commune. Toutefois, elle ne dit rien de la vitesse de cette convergence; on aimerait pourtant disposer d'un résultat permettant d'évaluer la distance entre ces deux moyennes pour une valeur donnée de  $n$ , au moins pour  $n$  grand. C'est ce que permet le théorème suivant, qui est le deuxième point crucial de ce chapitre :

**THÉORÈME 4.23** (Théorème central limite). — Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de variables aléatoires mutuellement indépendantes et admettant une espérance commune  $\mu \in \mathbb{R}$  et une variance commune  $\sigma^2 \in \mathbb{R}_+^*$ . Alors on a

$$\sqrt{n} \frac{\overline{X_n} - \mu}{\sigma} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Z$$

où  $Z$  est une variable aléatoire réelle suivant la loi  $\mathcal{N}(0, 1)$ .

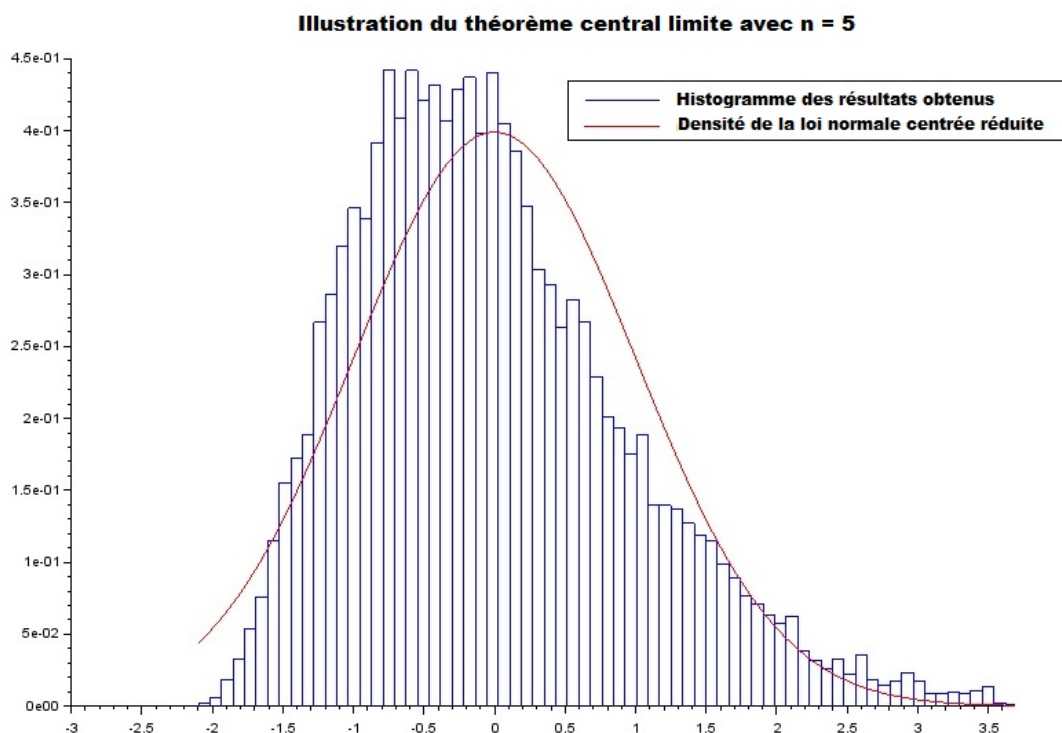
**REMARQUE 4.24.** En d'autres termes, sous les hypothèses précédentes, on a pour tout  $(a, b) \in \mathbb{R}^2$  tel que  $a < b$  :

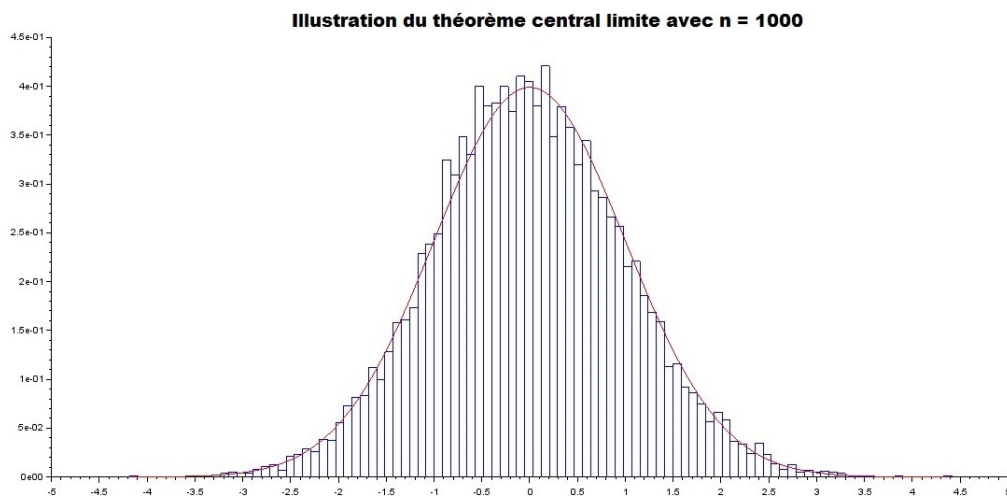
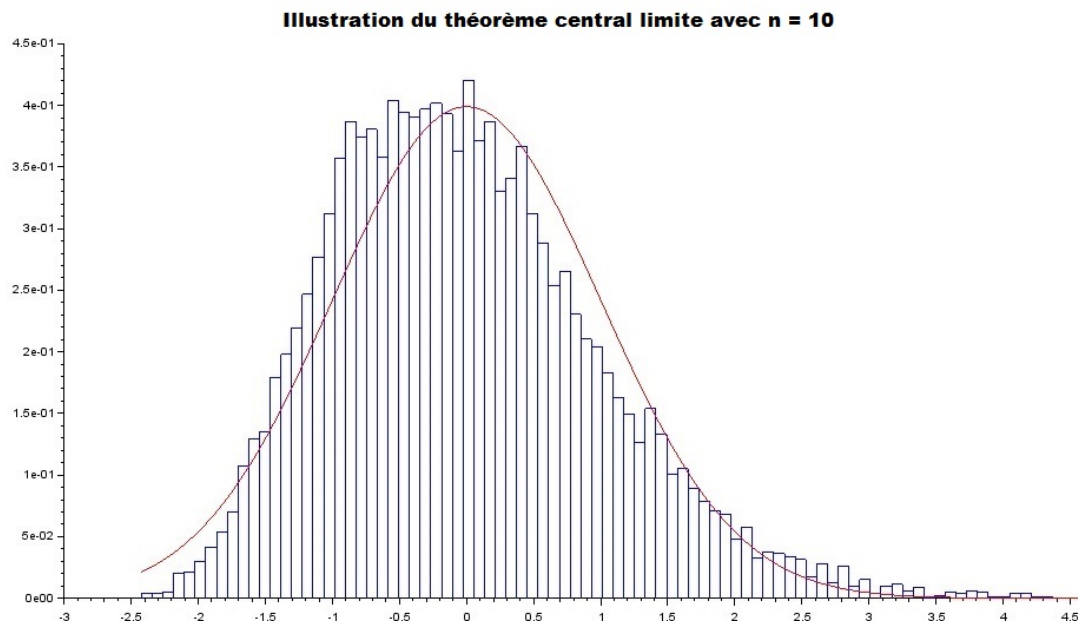
$$\mathbb{P} \left( \frac{\overline{X_n} - \mu}{\sigma} \in [a, b] \right) \xrightarrow[n \rightarrow +\infty]{} \Phi(b) - \Phi(a)$$

Notons que le résultat ci-dessus permet de fournir une approximation précise du comportement de la moyenne empirique de n'importe quelle loi, et en particulier d'une loi discrète, par la loi à densité qu'est la gaussienne standard.

**REMARQUE 4.25.** L'importance historique et scientifique du théorème central limite est immense; aussi est-il indispensable de bien le comprendre. Il signifie que lorsque  $n$  est grand, la loi de la variable  $\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$  se rapproche de la loi normale centrée réduite, c'est-à-dire que si l'on réalise de nombreuses (disons  $N$ ) observations indépendantes de la variable aléatoire réelle  $\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$  (notez qu'il faut alors disposer de  $nN$  observations des  $X_i$  en tout!), l'histogramme des valeurs obtenues aura la forme de la courbe en cloche qu'est la densité de la gaussienne standard.

**EXEMPLE 4.26.** Les tableaux suivants illustrent le théorème central limite appliqué à des variables suivant une loi exponentielle de paramètre 1. On tire indépendamment et 10000 fois de suite  $n$  variables de loi  $\mathcal{E}(1)$  (donc  $\mu = \sigma^2 = 1$ ), puis l'on calcule pour chacun de ces 10000 tirages la valeur de  $\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} = \sqrt{n} \bar{X}_n - 1$  et l'on superpose l'histogramme des valeurs obtenues à la courbe représentative densité de la loi normale standard. Le théorème central limite stipule que l'adéquation entre l'histogramme et la courbe doit être très forte lorsque  $n$  est grand : on s'aperçoit du fait que c'est bien le cas dès que  $n = 1000$ , mais aussi — et c'est un point intéressant — du fait que cette approximation n'est pas raisonnable pour des petites valeurs de  $n$ , cas dans lequel on observe une asymétrie de l'histogramme due à la forme de la densité des lois exponentielles.





**REMARQUE 4.27.** Attention à ne pas écrire  $\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma^2}$  au lieu de  $\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$  dans l'énoncé du théorème central limite : c'est bien la racine carrée de la variance des  $X_i$  (aussi appelée *écart-type* des  $X_i$ ) que l'on retrouve au dénominateur !

Faites le petit exercice suivant pour faciliter votre mémorisation du théorème : montrez en utilisant les propriétés des lois normales que si les variables  $X_i$  de l'énoncé suivent chacune une loi normale de mêmes paramètres  $\mu$  et  $\sigma^2$  et sont mutuellement indépendantes,  $\frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0, 1)$  pour tout  $n \geq 1$  — c'est-à-dire que le résultat du théorème, habituellement asymptotique, est ici exact à n'importe quel ordre. Vous comprendrez en faisant les calculs pourquoi  $\sigma^2$  ne convient pas, et observerez le rôle de  $\mu$  (resp. de  $\sigma$ ) dans le centrage (resp. la réduction) de la variable  $\bar{X}_n$ .

**Exercice 4.28.** Écrire l'énoncé du théorème central limite dans le cas de variables suivant une loi de Poisson.

**REMARQUE 4.29.** Des traductions erronées du terme allemand *zentralen Grenzwertsatz* ont donné lieu à plusieurs appellations alternatives du théorème central limite dans la littérature : bien que celle que nous adoptons plus haut soit, à notre avis, à privilégier, vous trouverez parfois — jusque dans certains sujets — des références au « théorème de la limite centrée » ou encore au « théorème de la limite centrale ». Il ne faut pas vous en étonner ; gardez à l'esprit qu'il s'agit bien du même théorème.

## 4.2 Estimation

Nous présentons pour finir les concepts de base de l'estimation paramétrique (c'est-à-dire de l'estimation d'un paramètre inconnu) ainsi que les méthodes élémentaires de construction d'intervalles de confiance. Toutes ces notions feront l'objet d'une présentation plus technique dans quelques séances.

### 4.2.1 Généralités et propriétés des estimateurs

Un problème d'estimation paramétrique élémentaire auquel nous ferons souvent référence dans cette section et que nous avons déjà évoqué est le suivant : on considère une population constituée d'un grand nombre d'individus et on cherche à estimer la proportion inconnue  $p \in [0, 1]$  de personnes au sein de cette population prêtes à voter pour le candidat X aux prochaines élections. On suppose qu'il est trop coûteux en temps et en moyens matériels d'interroger chaque individu sur ses opinions politiques mais qu'il est possible de disposer, pour un certain  $n \geq 1$ , de  $n$  observations indépendantes réalisées à partir d'individus choisis au hasard dans la population. En codant par 1 le fait d'être prêt à voter pour le candidat X aux prochaines élections et 0 le fait contraire, on dispose donc de  $n$  variables indépendantes et identiquement distribuées selon une loi de Bernoulli de paramètre  $p$ .

L'enjeu est alors de chercher à exprimer en fonction de ces variables aléatoires un estimateur permettant d'approcher  $p$  de manière aussi précise que possible et d'évaluer cette précision en termes probabilistes. On désignera ce problème sous le nom de « problème du sondage ».

**DÉFINITION 4.30.** — Si  $n \geq 1$ , on appelle  $n$ -échantillon tout  $n$ -uplet  $(X_1, \dots, X_n)$  de variables aléatoires réelles indépendantes et identiquement distribuées.

**REMARQUE 4.31.** On supposera toujours dans cette section que les données dont on dispose forment un échantillon. Il va sans dire que cette hypothèse est loin d'être évidente en pratique : songez par exemple au postulat d'indépendance des variables observées, peu crédible dans le problème du sondage si l'on réalise notre enquête au sein d'un quartier ou même d'une ville donnée<sup>20</sup>...

**DÉFINITION 4.32** (Estimateur). — Soit  $n$  un entier et  $(X_1, \dots, X_n)$  un  $n$ -échantillon. On appelle estimateur (ou parfois statistique) toute variable aléatoire réelle  $Z$  s'écrivant

20. ... et l'on ne parle même pas du biais induit par le choix d'un mode d'enquête comme le téléphone ou le démarchage dans la rue, par les non-réponses ou par le mensonge!

en fonction de  $(X_1, \dots, X_n)$ , c'est-à-dire telle qu'il existe une fonction  $f$  définie sur une partie de  $\mathbb{R}^n$  contenant  $X_1(\Omega) \times \dots \times X_n(\Omega)$  et à valeurs dans  $\mathbb{R}$  telle que  $Z = f(X_1, \dots, X_n)$ .

Un estimateur des  $X_i$  est donc simplement « quelque chose qui dépend des  $X_i$  » et non du paramètre à estimer. En général, on cherche à trouver de bons estimateurs, c'est-à-dire des estimateurs qui apportent une information intéressante sur le problème, ou qui donnent une bonne estimation d'une certaine grandeur. Par exemple, la moyenne empirique des observations semble être un « bon » estimateur de la moyenne théorique, en vertu de la loi des grands nombres. Nous allons par la suite rendre ces intuitions plus rigoureuses.

**EXEMPLE 4.33.** Si  $(X_1, \dots, X_n)$  est un  $n$ -échantillon, la moyenne empirique  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  est un estimateur. La fonction  $f$  correspondant à ce problème est ici la fonction *moyenne*:

$$f(x_1, \dots, x_n) = \frac{x_1 + \dots + x_n}{n}$$

De même, le minimum  $\min(X_1, \dots, X_n)$  des  $X_i$ , la variable aléatoire  $X_1$  ou la variable aléatoire constante 0 sont aussi des estimateurs.

Dans ce qui suit, on se place dans le cadre d'un modèle dans lequel on tente d'estimer un paramètre  $\theta \in \mathbb{R}$  grâce à un estimateur  $\hat{\theta}_n$  fonction d'un  $n$ -échantillon  $(X_1, \dots, X_n)$ . Par exemple, dans le cadre du problème du sondage, c'est  $p$  qui joue le rôle de  $\theta$ .

**DÉFINITION 4.34.** — Soit  $\hat{\theta}_n$  un estimateur qui admet une espérance.

- Le biais de  $\hat{\theta}_n$  est la quantité  $\mathbb{E}(\hat{\theta}_n) - \theta$ .
- On dit que  $\hat{\theta}_n$  est un estimateur sans biais lorsque  $\mathbb{E}(\hat{\theta}_n) = \theta$ .
- On dit que  $\hat{\theta}_n$  est un estimateur asymptotiquement sans biais lorsque

$$\lim_{n \rightarrow +\infty} \mathbb{E}(\hat{\theta}_n) - \theta = 0$$

**Exercice 4.35.** Si  $X_1$  admet une variance non nulle, montrer que la moyenne empirique du  $n$ -échantillon  $(X_1, \dots, X_n)$  est un indicateur sans biais de  $\theta = \mathbb{E}(X_1)$  et que la variance empirique

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

est un estimateur biaisé de  $\sigma^2 = V(X_1)$ .

**DÉFINITION 4.36** (Convergence d'un estimateur). — Soit  $\hat{\theta}_n$  un estimateur. On dit que l'estimateur  $\hat{\theta}_n$  de  $\theta$  est convergent (on trouve aussi l'anglicisme consistant) si et seulement si on a

$$\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \theta$$

**REMARQUE 4.37.** Les lecteurs les plus rigoureux (voire tatillons) auront noté que la propriété ci-dessus concerne la suite des estimateurs  $(\hat{\theta}_n)_{n \in \mathbb{N}^*}$  et non le seul estimateur  $\hat{\theta}_n$ , contrairement à ce que la terminologie adoptée (« l'estimateur  $\hat{\theta}_n$  est convergent ») pourrait suggérer. Toutefois, l'usage veut que l'on accepte dans ce cas précis l'abus de langage qui consiste à confondre ces deux objets.

**Exercice 4.38.** Démontrer que si  $\hat{\theta}_n$  est un estimateur sans biais de  $\theta$  admettant une variance  $\sigma_n^2$  et si  $\sigma_n^2 \xrightarrow{n \rightarrow +\infty} 0$ , alors  $\hat{\theta}_n$  est convergent.

**EXEMPLE 4.39.** Dans le cadre du problème du sondage, la moyenne empirique  $\overline{X}_n$  est un estimateur convergent de  $p$ .

**DÉFINITION 4.40** (Risque quadratique). — Si l'estimateur  $\hat{\theta}_n$  admet un moment d'ordre 2, on définit le risque quadratique de  $\hat{\theta}_n$  (en tant qu'estimateur de  $\theta$ ) comme la quantité

$$\text{RQ}(\hat{\theta}_n, \theta) = \mathbb{E}((\hat{\theta}_n - \theta)^2)$$

**REMARQUE 4.41.** Dans le cadre de la définition ci-dessus, si  $\hat{\theta}_n$  est sans biais alors son risque quadratique est tout simplement sa variance. En général, le risque quadratique de  $\hat{\theta}_n$  décrit la dispersion de  $\hat{\theta}_n$  autour de  $\theta$ ; il est raisonnable de penser que l'obtention d'une estimation fine et fiable passe par la construction d'un indicateur dont le risque quadratique est faible.

**PROPOSITION 4.42** (« RQ = biais + variance »). — Si  $\hat{\theta}_n$  admet un moment d'ordre 2, son risque quadratique est donné par

$$\text{RQ}(\hat{\theta}_n, \theta) = (\mathbb{E}(\hat{\theta}_n) - \theta)^2 + V(\hat{\theta}_n)$$

**EXEMPLE 4.43.** Dans le cadre du problème du sondage, le risque quadratique de l'estimateur de  $p$  donné par la moyenne empirique est  $\text{RQ}(\overline{X}_n, p) = V(\overline{X}_n) = \frac{p(1-p)}{n}$ .

**PROPOSITION 4.44.** — Si  $\hat{\theta}_n$  admet un risque quadratique (relativement à  $\theta$ ) tendant vers 0 lorsque  $n$  tend vers l'infini, alors  $\hat{\theta}_n$  est convergent.

La proposition précédente est intéressante, puisqu'elle permet d'étudier la convergence d'un estimateur juste en regardant son risque quadratique, qui est parfois plus simple à étudier. Nous illustrons ceci par un exemple développé en détail.

**EXEMPLE 4.45.** On considère un  $n$ -échantillon de variables aléatoires réelles de loi  $\mathcal{U}([0, \theta + 1])$  (où  $\theta > 0$ ). On cherche à construire un bon estimateur de  $\theta$ . On propose pour cela l'estimateur  $\hat{\theta}_n = \min(X_1, \dots, X_n)$ . On remarque que pour tout  $i \in \llbracket 1, n \rrbracket$  on a  $X_i = \theta + X'_i$  où  $(X'_1, \dots, X'_n)$  est un  $n$ -échantillon de variables aléatoires de loi  $\mathcal{U}([0, 1])$ . On a alors  $Z_n = \theta + Z'_n$ , où  $Z'_n = \min(X'_1, \dots, X'_n)$  donc  $\mathbb{E}(Z_n) = \theta + \mathbb{E}(Z'_n)$ . Cela implique que le biais de  $\hat{\theta}_n = Z_n$  est  $\mathbb{E}(Z'_n)$  et que sa variance<sup>21</sup> est  $V(Z_n) = V(Z'_n)$ . D'après la décomposition « risque quadratique = biais + variance », on a donc :

$$\text{RQ}(\hat{\theta}_n, \theta) = \mathbb{E}(Z'_n)^2 + V(Z'_n) = \mathbb{E}(Z_n'^2)$$

Il faut maintenant calculer  $\mathbb{E}(Z_n'^2)$ . Pour cela, on remarque que  $Z_n'^2$  est positive et admet une densité, et qu'elle est bornée donc admet une espérance. D'après la formule de la proposition 3.17, on a donc :

$$\mathbb{E}(Z_n'^2) = \int_0^{+\infty} \mathbb{P}(Z_n'^2 \geq t) dt$$

21. On rappelle que comme les  $Z'_i$  et les  $Z_i$  sont bornées, ces variances existent toutes.



Mais pour tout  $t \in \mathbb{R}_+$  on a :

$$\begin{aligned}\mathbb{P}(Z_n'^2 \geq t) &= \mathbb{P}(Z_n' \geq \sqrt{t}) \\ &= \mathbb{P}(X_1' \geq \sqrt{t} \cap \dots \cap X_n' \geq \sqrt{t}) \\ &= \mathbb{P}(X_1' \geq \sqrt{t})^n\end{aligned}$$

Où nous avons successivement utilisé la définition de  $Z_n'$ , l'indépendance des  $X_i'$ , et le fait qu'elles sont identiquement distribuées. On a donc  $\mathbb{P}(Z_n'^2 \geq t) = (1 - \sqrt{t})^n$  si  $t \in [0, 1]$  et  $\mathbb{P}(Z_n'^2 \geq t) = 0$  si  $t > 1$ . On obtient alors :

$$\text{RQ}(\hat{\theta}_n, \theta) = \mathbb{E}(Z_n'^2) = \int_0^1 (1 - \sqrt{t})^n dt$$

La valeur de cette intégrale n'est pas évidente à calculer, mais on peut montrer qu'elle tend vers 0 lorsque  $n$  tend vers  $+\infty$  : en effet, comme  $\sqrt{t} \geq t$  pour tout  $t \in [0, 1]$  on a

$$0 \leq \text{RQ}(\hat{\theta}_n, \theta) \leq \int_0^1 (1 - t)^n dt = \left[ -\frac{(1 - t)^{n+1}}{n+1} \right]_0^1 = \frac{1}{n+1}$$

pour tout  $n \geq 1$ , d'où  $\text{RQ}(\hat{\theta}_n) \rightarrow 0$  lorsque  $n \rightarrow +\infty$ . Le risque quadratique de cet estimateur tend donc vers 0 : par la proposition précédente, il est également asymptotiquement sans biais.

#### 4.2.2 Intervalles de confiance

On aborde maintenant le problème de la construction d'intervalles de confiance. Le cas le plus fréquemment rencontré, celui de l'estimation du paramètre d'une loi de Bernoulli, correspond au problème du sondage et est traité *in extenso* ci-dessous.

**DÉFINITION 4.46** (Intervalle de confiance). — Soit  $\alpha \in [0, 1]$ . On appelle intervalle de confiance de niveau  $1 - \alpha$  pour  $\theta$  tout intervalle aléatoire  $[f_n(X_1, \dots, X_n); g_n(X_1, \dots, X_n)]$  (où  $f_n$  et  $g_n$  sont des fonctions à  $n$  variables et à valeurs réelles) dont les bornes dépendent de  $(X_1, \dots, X_n)$  mais pas de  $\theta$  et qui contient  $\theta$  avec une probabilité au moins égale à  $1 - \alpha$ .

Un intervalle de confiance de niveau  $1 - \alpha$  pour  $\theta$  est donc un intervalle construit à partir des observations (et dont on peut donner les valeurs numériques des bornes une fois ces observations réalisées) dans lequel on peut affirmer avec un « niveau de confiance »  $1 - \alpha$  que le paramètre  $\theta$  se trouve.

**REMARQUE 4.47.** N'importe quel intervalle convient dans le cas  $\alpha = 1$ , mais le fait d'obtenir un intervalle de confiance « sûr à 0% » n'est pas très intéressant... Par ailleurs, le cas  $\alpha = 0$  correspond à un encadrement certain qui ne donne généralement pas lieu à des intervalles de confiance très intéressants. On se limitera donc souvent au cas  $\alpha \in ]0, 1[$ .

En pratique, on détermine *a priori* un intervalle de confiance de niveau donné, dont les bornes, rappelons-le encore une fois, sont aléatoires, puis on collecte un échantillon formé de  $n$  observations et on sait que le paramètre  $\theta$  à estimer a une probabilité

au moins égale à  $1 - \alpha$  de se trouver dans l'intervalle (cette fois non aléatoire) obtenu en remplaçant les  $X_i$  dans l'expression de l'intervalle de confiance par les valeurs observées. Notons que  $\alpha$  est souvent pris égal à 5% ou 10%; l'estimation par intervalle de confiance est alors fiable respectivement à 95% et 90%.

**EXEMPLE 4.48.** Si l'on cherche à estimer  $\mu \in \mathbb{R}$  à partir des observations  $(X_1, \dots, X_n)$  indépendantes et identiquement distribuées de la loi  $\mathcal{N}(\mu, 1)$ , on sait par exemple que  $X_1 - \mu$  suit une loi normale centrée réduite, ce qui permet d'écrire pour tout  $\alpha \in ]0, 1[$ , en remarquant que  $\Phi^{-1}(1 - \frac{\alpha}{2}) \geq 0$  :

$$\begin{aligned} \mathbb{P}\left(\mu \in \left[X_1 - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right); X_1 + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right]\right) &= \mathbb{P}\left(|\mu - X_1| \leq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) \\ &= \mathbb{P}\left(X_1 \in \left[\mu - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right); \mu + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right]\right) = \mathbb{P}\left(X_1 - \mu \in \left[-\Phi^{-1}\left(1 - \frac{\alpha}{2}\right); \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right]\right) \end{aligned}$$

Comme expliqué précédemment, nous savons que  $X_1 - \mu$  suit une loi normale centrée réduite. La quantité précédente est donc égale à :

$$\begin{aligned} \Phi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) - \Phi\left(-\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) &= \Phi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) - \left(1 - \Phi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)\right) \\ &= 1 - \frac{\alpha}{2} - \left(1 - \left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha \end{aligned}$$

Nous venons donc d'établir que  $[X_1 - \Phi^{-1}(1 - \frac{\alpha}{2}); X_1 + \Phi^{-1}(1 - \frac{\alpha}{2})]$  est un intervalle de confiance de niveau  $1 - \alpha$  pour  $\mu$ .

Supposons que l'on dispose d'une observation de  $X_1$ , par exemple  $X_1(\omega) = 7$ . Nous ne connaissons pas  $\mu$  et nous cherchons à l'estimer, avec un niveau de confiance de 95% (soit  $\alpha = 5\%$ ). On sait<sup>22</sup> que  $\Phi^{-1}(0.975) = 1.96$ . En utilisant le résultat précédent, on montre qu'il y a 95% de chances pour que  $\mu$  se situe dans l'intervalle  $[7 - 1.96, 7 + 1.96] = [5.04, 8.96]$ . On s'attend donc avec une confiance (95%) à ce que  $\mu$  soit quelque part entre 5.04 et 8.96, ce qui n'est pas très précis!

**Exercice 4.49.** Dans le cadre de l'exemple précédent, montrer que

$$\left[\bar{X}_n - \frac{1}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right); \bar{X}_n + \frac{1}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right]$$

est un intervalle de confiance de niveau  $1 - \alpha$  pour  $\mu$ . Comparer sa longueur avec celle de l'intervalle déterminé précédemment. Ce résultat est-il conforme à l'intuition?

L'exemple suivant est très classique : on peut retenir la formule donnant l'intervalle de confiance par coeur, mais il faut impérativement savoir le reconstruire. On aura l'occasion de s'entraîner longuement sur des exemples numériques dans un cours ultérieur.

**EXEMPLE 4.50** (Intervalle de confiance pour la moyenne de lois normales). De manière générale, si l'on travaille non plus avec une loi  $\mathcal{N}(\mu, 1)$  mais avec une loi  $\mathcal{N}(\mu, \sigma^2)$  (avec  $\sigma > 0$ ), un intervalle de confiance de niveau  $1 - \alpha$  pour  $\mu$  est donné par

$$\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right); \bar{X}_n + \frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right]$$

22. En fait, ce n'est pas facile à calculer, mais les valeurs de  $\Phi$  sont « tabulées » : nous verrons plus tard ce que cela signifie.

On remarquera que l'intervalle est d'autant plus étroit (donc précis) que le nombre d'observations  $n$  augmente, ce qui est là encore conforme à l'intuition donnée par la loi des grands nombres.

Remarquons également qu'il est d'autant plus large (et donc imprécis) que  $\sigma$  est élevé, ce qui est tout à fait conforme à l'intuition puisque la haute variabilité des observations empêche alors une observation fine de leur moyenne théorique. D'ailleurs, la construction de l'intervalle de confiance suppose que l'on connaît  $\sigma$  : dans la pratique, c'est rarement le cas ! Nous verrons plus tard comment pallier ce problème.

**DÉFINITION 4.51** (Intervalle de confiance asymptotique). — Si  $\alpha \in [0, 1]$ , on appelle intervalle de confiance asymptotique pour  $\theta$  de niveau asymptotique  $1 - \alpha$  tout intervalle aléatoire  $[f_n(X_1, \dots, X_n); g_n(X_1, \dots, X_n)]$  (où  $f_n$  et  $g_n$  sont des fonctions à  $n$  variables et à valeurs réelles) dont les bornes dépendent de  $(X_1, \dots, X_n)$  mais pas de  $\theta$  et tel qu'il existe une suite  $(v_n)_{n \in \mathbb{N}^*}$  de limite  $1 - \alpha$  telle que :

$$\mathbb{P}(\theta \in [f_n(X_1, \dots, X_n); g_n(X_1, \dots, X_n)]) \geq v_n$$

pour tout  $n \in \mathbb{N}^*$ .

**REMARQUE 4.52.** Notez que l'on ne demande pas à  $\mathbb{P}(\theta \in [f_n(X_1, \dots, X_n); g_n(X_1, \dots, X_n)])$  d'admettre une limite quand  $n$  tend vers l'infini.

**EXEMPLE 4.53.** On détaille ici les différentes constructions possibles d'un intervalle de confiance dans le cadre du problème du sondage, dont on rappelle qu'il consiste à estimer le paramètre  $p \in [0, 1]$  d'une loi de Bernoulli à partir d'un  $n$ -échantillon  $(X_1, \dots, X_n)$  de cette loi. On suppose que  $p \in ]0, 1[$ , ce qui correspond aux situations pratiques. On se donne  $\alpha \in ]0, 1[$  et on cherche à construire un intervalle de confiance de niveau  $1 - \alpha$  pour  $p$ .

Remarquons tout d'abord que quelle que soit la valeur de  $p$ , on a  $p(1 - p) \leq 1/4$  : en effet, le maximum de la fonction polynomiale  $x \mapsto x(1 - x) = -x^2 + x$  sur  $\mathbb{R}$  est atteint en  $1/2$  et vaut  $1/4$ . On utilisera à plusieurs reprises cette inégalité grossière mais très pratique.

Tentons dans un premier temps d'obtenir un intervalle de confiance non asymptotique pour  $p$  : il s'agit de trouver un événement impliquant un encadrement de  $p$  par des valeurs dépendant des  $X_i$  et qui soit de probabilité au plus  $1 - \alpha$ . L'inégalité de Bienaymé-Tchebytchev semble toute indiquée pour construire notre intervalle, mais il faut encore décider à quelle variable l'appliquer. Parmi les différentes variables d'espérance  $p$  dont on dispose, on distingue notamment  $X_1$  et  $\overline{X}_n$ . Or les exemples précédents suggèrent qu'il est préférable d'utiliser toute l'information disponible pour réaliser notre estimation si l'on veut que celle-ci soit la plus précise possible ; aussi choisit-on d'appliquer l'inégalité à  $\overline{X}_n$ , ce qui permet d'écrire pour tout  $\varepsilon > 0$  :

$$\mathbb{P}(|\overline{X}_n - p| \geq \varepsilon) \leq \frac{V(\overline{X}_n)}{\varepsilon^2} = \frac{p(1 - p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

où l'on a utilisé la majoration obtenue au paragraphe précédent pour se débarrasser de l'expression  $p(1 - p)$  qui contient le paramètre inconnu  $p$  et qui empêcherait de choisir par la suite une valeur de  $\varepsilon$  indépendante de  $p$ .

Par passage au complémentaire, notre inégalité devient :

$$\mathbb{P}\left(\left|\overline{X}_n - p\right| < \varepsilon\right) \geq 1 - \frac{1}{4n\varepsilon^2}$$

c'est-à-dire

$$\mathbb{P}\left(p \in \left[\overline{X}_n - \varepsilon; \overline{X}_n + \varepsilon\right]\right) \geq 1 - \frac{1}{4n\varepsilon^2}$$

Il s'agit ensuite de choisir  $\varepsilon$  pour que le terme de droite soit égal à  $1 - \alpha$ . Un simple calcul permet de voir que la valeur de  $\varepsilon$  qui convient est  $\frac{1}{2\sqrt{n}\sqrt{\alpha}}$ , et donc que l'intervalle aléatoire

$$I_n = \left[\overline{X}_n - \frac{1}{2\sqrt{n}\sqrt{\alpha}}; \overline{X}_n + \frac{1}{2\sqrt{n}\sqrt{\alpha}}\right]$$

est bien un intervalle de confiance de niveau  $1 - \alpha$  pour  $p$ .

On remarque que sa longueur décroît à mesure que  $n$  augmente, c'est-à-dire que le fait de disposer d'un grand nombre d'observations permet de réaliser une estimation plus précise. Elle décroît aussi lorsque  $\alpha$ , qui représente le degré d'incertitude avec lequel on accepte de formuler la proposition «  $p$  est dans l'intervalle de confiance », augmente. Si ces deux mécanismes ne vous semblent pas intuitifs, vous n'avez pas tout à fait compris la définition d'un intervalle de confiance : rendez-vous à la définition 4.46.

Notez enfin que si nous avons choisi de travailler sur l'indicateur  $X_1$  au lieu de l'indicateur  $\overline{X}_n$ , notre intervalle de confiance (construisez-le!) aurait été indépendant de  $n$ , et donc asymptotiquement infiniment moins précis que celui que nous avons construit.

On veut à présent construire un intervalle de confiance asymptotique pour  $p$ , dans l'espoir que l'intervalle obtenu soit meilleur (c'est-à-dire plus petit) que l'intervalle non asymptotique que nous venons de présenter. On utilise à nouveau la moyenne empirique des observations  $\overline{X}_n$ , mais cette fois c'est le théorème central limite qui nous donne notre première relation :

$$\forall \varepsilon > 0, \quad \mathbb{P}\left(-\varepsilon \leq \sqrt{n} \frac{\overline{X}_n - p}{\sqrt{p(1-p)}} \leq \varepsilon\right) \xrightarrow{n \rightarrow +\infty} \Phi(\varepsilon) - \Phi(-\varepsilon) = 2\Phi(\varepsilon) - 1$$

Remarquez que pour des raisons de simplicité et d'esthétique<sup>23</sup> on choisit d'emblée de construire un intervalle de confiance asymptotique centré sur  $\overline{X}_n$ , ce qui explique que l'on s'intéresse à la probabilité d'un événement de la forme  $(-\varepsilon \leq \dots \leq \varepsilon)$ , mais qu'il aurait tout à fait été possible de considérer un événement de la forme  $(a \leq \dots \leq b)$ , ou même de la forme  $(a \leq \dots)$  par exemple, ce dernier ayant alors permis d'obtenir un intervalle de confiance asymptotique unilatéral (c'est-à-dire avec une borne infinie).

La relation ci-dessus se réécrit :

$$\forall \varepsilon > 0, \mathbb{P}\left(p \in \left[\overline{X}_n - \frac{\varepsilon \sqrt{p(1-p)}}{\sqrt{n}}; \overline{X}_n + \frac{\varepsilon \sqrt{p(1-p)}}{\sqrt{n}}\right]\right) \xrightarrow{n \rightarrow +\infty} 2\Phi(\varepsilon) - 1$$

23. ... et sans doute un peu aussi parce que l'on sait que la densité de la gaussienne est concentrée autour de sa moyenne et que centrer l'intervalle de confiance est un bon moyen de réduire sa longueur!

d'où, en remarquant que

$$\left[ \bar{X}_n - \frac{\varepsilon \sqrt{p(1-p)}}{\sqrt{n}}; \bar{X}_n + \frac{\varepsilon \sqrt{p(1-p)}}{\sqrt{n}} \right] \subset \left[ \bar{X}_n - \frac{\varepsilon}{2\sqrt{n}}; \bar{X}_n + \frac{\varepsilon}{2\sqrt{n}} \right]$$

grâce à l'inégalité  $p(1-p) \leq \frac{1}{4}$  démontrée plus haut, le fait que  $\left[ \bar{X}_n - \frac{\varepsilon}{2\sqrt{n}}; \bar{X}_n + \frac{\varepsilon}{2\sqrt{n}} \right]$  est un intervalle de confiance asymptotique pour  $p$  de niveau asymptotique  $2\Phi(\varepsilon) - 1$ .

Il suffit alors de choisir  $\varepsilon$  tel que  $2\Phi(\varepsilon) - 1 = 1 - \alpha$  pour obtenir un intervalle de confiance asymptotique pour  $p$  de niveau asymptotique  $1 - \alpha$ . Un calcul rapide donne  $\varepsilon = \Phi^{-1}(1 - \frac{\alpha}{2})$ , d'où l'intervalle de confiance asymptotique pour  $p$  de niveau  $1 - \alpha$  suivant :

$$I'_n = \left[ \bar{X}_n - \frac{\Phi^{-1}(1 - \frac{\alpha}{2})}{2\sqrt{n}}; \bar{X}_n + \frac{\Phi^{-1}(1 - \frac{\alpha}{2})}{2\sqrt{n}} \right]$$

Il est encore une fois possible d'observer que la longueur de  $I'_n$  est d'autant plus faible (c'est-à-dire, l'estimation d'autant plus précise) que  $n$  est grand et que la tolérance  $\alpha$  est élevée. On peut aussi comparer la longueur de  $I_n$  et celle de  $I'_n$  : pour  $\alpha = 5\%$ , la longueur de  $I_n$  est approximativement égale à  $\frac{10}{\sqrt{n}\sqrt{5}} \approx \frac{4,47}{\sqrt{n}}$  et celle de  $I'_n$  à  $\frac{1,96}{\sqrt{n}}$ . En d'autres termes, pour un même niveau de confiance  $1 - \alpha$ , l'intervalle de confiance asymptotique  $I'_n$  est plus de deux fois plus précis que l'intervalle de confiance non asymptotique  $I_n$  ! Mais bien entendu, il est n'est valable que lorsque  $n$  est grand ; c'est l'objet de la remarque suivante.

**REMARQUE 4.54.** Il peut vous être demandé de commenter la validité de l'utilisation d'un résultat asymptotique pour des valeurs données de  $n$  (par exemple, vous pouvez rencontrer la question « peut-on utiliser le théorème central limite pour construire un intervalle de confiance dans ce cas ? »). On considère d'ordinaire les valeurs de  $n$  plus grandes que 100 (en physique) ou 30 (en sciences sociales) comme suffisamment élevées pour que les résultats asymptotiques puissent être utilisés pour fournir des approximations. Les différentes illustrations du théorème central limite fournies plus haut montrent qu'il n'est pas raisonnable d'utiliser ces approximations pour des valeurs de  $n$  de l'ordre de 10.

Terminons ce chapitre par une application numérique des résultats que nous venons de voir. Deux valeurs de  $\Phi^{-1}$ , déjà données plus haut, sont à connaître :

$$\Phi^{-1}(0,95) \approx 1,64 \text{ et } \Phi^{-1}(0,975) \approx 1,96$$

**Exercice 4.55.** On cherche à savoir quelle est la proportion  $p \in [0,1]$  de personnes convaincues que la Terre est plate dans la population française. On interroge indépendamment 10000 personnes choisies au hasard dans la population, dont 2104 pensent que la Terre est plate. Donner un intervalle de confiance de niveau 95% pour  $p$ .

## 5 Solutions des exercices

**Solution de l'exercice 1.16.** On va utiliser la formule de Bayes pour « inverser le conditionnement » : on connaît la probabilité conditionnelle d'obtenir un résultat positif en cas d'ivresse et on cherche la probabilité d'être ivre sachant que le résultat est positif. On note  $I$  l'événement « Roger est ivre » et  $A$  l'événement « le résultat de l'alcootest est positif ». Les données de l'énoncé se traduisent de la manière suivante :

$$\mathbb{P}(I) = \frac{1}{100}, \quad \mathbb{P}(A | I) = \frac{99}{100} \quad \text{et} \quad \mathbb{P}(A | \bar{I}) = \frac{2}{100}$$

On peut alors écrire :

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

La seule donnée que l'énoncé ne fournit pas directement est  $\mathbb{P}(A)$  ; on la calcule en utilisant la formule des probabilités totales sur le système complet d'événements de probabilités non nulles  $(I, \bar{I})$ , ce qui donne :

$$\mathbb{P}(I | A) = \frac{\mathbb{P}(A | I)\mathbb{P}(I)}{\mathbb{P}(A | I)\mathbb{P}(I) + \mathbb{P}(A | \bar{I})\mathbb{P}(\bar{I})} = \frac{\frac{99}{100} \frac{1}{100}}{\frac{99}{100} \frac{1}{100} + \frac{2}{100} \frac{99}{100}} = \frac{1}{3}$$

Ce résultat plutôt contre-intuitif s'explique par le fait qu'une grande partie de la population est sobre, ce qui fait que le nombre d'erreurs réalisées sur des personnes sobres est assez élevé par rapport au nombre de détections correctes de personnes ivres même si le taux d'erreur lui-même est très faible ; la proportion de personnes sobres parmi les personnes ayant obtenu un résultat positif est donc importante.

**Solution de l'exercice 1.7.** Pour  $n = 1$ , la formule des probabilités composées est simplement la définition de la probabilité conditionnelle  $\mathbb{P}(B | A_1)$ . Pour  $n = 2$ , elle s'écrit  $\mathbb{P}(B \cap A_1 \cap A_2) = \mathbb{P}(B | A_1 \cap A_2)\mathbb{P}(A_2 | A_1)\mathbb{P}(A_1)$  et est vraie comme on le voit en écrivant que  $\mathbb{P}(A_2 | A_1)\mathbb{P}(A_1) = \mathbb{P}(A_1 \cap A_2)$ . Pour tout  $n \geq 1$ , on pose  $P_n$  : « la formule des probabilités composées est vraie au rang  $n$  ». On a montré que  $P_1$  est vraie. Soit maintenant  $n \geq 1$  ; on suppose que  $P_n$  est vraie et on veut montrer que  $P_{n+1}$  l'est. On se donne donc des événements  $A_1, \dots, A_{n+1}, B$  tels que  $A_1 \cap \dots \cap A_{n+1} = (A_1 \cap A_2) \cap \dots \cap A_{n+1}$  est de probabilité non nulle. On peut alors appliquer la formule des probabilités composées à  $B$  et aux  $n$  événements  $A_1 \cap A_2, A_3, \dots, A_{n+1}$  :

$$\mathbb{P}(B \cap A_1 \cap \dots \cap A_{n+1}) = \mathbb{P}(A_1 \cap A_2)\mathbb{P}(A_3 | A_1 \cap A_2) \dots \mathbb{P}(A_{n+1} | A_1 \cap \dots \cap A_n)\mathbb{P}(B | A_1 \cap \dots \cap A_{n+1})$$

et il suffit d'écrire  $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2 | A_1)\mathbb{P}(A_1)$  pour conclure, ce qui établit l'hérédité de la proposition  $P_n$ . On en déduit que  $P_n$  est vraie pour tout  $n \geq 1$ .

**Solution de l'exercice 1.13.** Avec les notations du théorème, on écrit

$$B = B \cap \Omega = B \cap (\sqcup_{i \in I} A_i) = \sqcup_{i \in I} (B \cap A_i)$$

puis par incompatibilité

$$\mathbb{P}(B) = \sum_{i \in I} \mathbb{P}(B \cap A_i) = \sum_{i \in I} \mathbb{P}(B | A_i)\mathbb{P}(A_i)$$

où la dernière égalité vient de la définition de  $\mathbb{P}(B | A_i)$ .

Remarquez que l'utilisation de la notation  $\sqcup$  à la place de  $\cup$  sert uniquement à souligner le fait que l'union d'événements avec laquelle on travaille est disjointe (et donc à accroître la lisibilité de la preuve), ce qui ne dispense pas de mentionner à nouveau ce fait entre les deux lignes de calcul!

**Solution de l'exercice 1.19.** Montrons le premier point : le deuxième en découlera en appliquant le premier avec  $A' = B$  et  $B' = A$  (qui sont bien indépendants), et le troisième en appliquant le premier avec  $A'' = \bar{A}$  et  $B'' = B$  (indépendants d'après le deuxième point).

Il suffit de calculer  $\mathbb{P}(A \cap \bar{B})$  en faisant apparaître  $A \cap B$  : or  $A = A \cap \Omega = A \cap (B \sqcup \bar{B}) = (A \cap B) \sqcup (A \cap \bar{B})$  (faire un dessin si ce n'est pas clair!), donc  $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \bar{B})$  par incompatibilité, donc  $\mathbb{P}(A) = \mathbb{P}(A)\mathbb{P}(B) + \mathbb{P}(A \cap \bar{B})$  d'où enfin

$$\mathbb{P}(A \cap \bar{B}) = \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A)\mathbb{P}(\bar{B})$$

ce qu'il fallait démontrer.

**Solution de l'exercice 1.24.** Chaque vignette a une probabilité égale à  $1/500$  d'être celle que Gonzague attend. Si pour tout  $i \in \llbracket 1, 500 \rrbracket$  on note  $A_i$  l'événement « la vignette  $i$  complète la collection de Gonzague », l'événement « la collection est complète après ouverture des 500 vignettes » s'écrit  $\bigcup_{i=1}^{500} A_i$ . Comme les événements  $A_i$  sont mutuellement indépendants (puisque les vignettes sont sélectionnées indépendamment les unes des autres), on écrit

$$\bigcup_{i=1}^{500} A_i = \overline{\bigcap_{i=1}^{500} \bar{A}_i}$$

et on utilise l'indépendance mutuelle des  $\bar{A}_i$  pour calculer

$$\mathbb{P}\left(\bigcup_{i=1}^{500} A_i\right) = 1 - \mathbb{P}\left(\bigcap_{i=1}^{500} \bar{A}_i\right) = 1 - \prod_{i=1}^{500} (1 - \mathbb{P}(A_i)) = 1 - \left(1 - \frac{1}{500}\right)^{500} \simeq 1 - e^{-1} \simeq 0,63$$

les dernières approximations, non attendues par l'énoncé, résultant du fait que 500 est grand et du fait suivant, vrai pour tout  $x \in \mathbb{R}$  :

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

**Solution de l'exercice 2.24.** Pour tout  $i \in \llbracket 1, n \rrbracket$ , notons  $X_i$  la variable aléatoire qui vaut 1 si le cadeau  $i$  est correctement livré à l'enfant  $i$  et 0 sinon. La permutation des adresses ayant lieu au hasard, il n'y a aucune raison pour que le cadeau  $i$  se retrouve avec l'adresse  $j$  plutôt qu'avec l'adresse  $i$ , ni l'inverse : en d'autres termes, l'adresse à laquelle sera envoyée le cadeau  $i$  est choisie de manière équiprobable parmi  $\llbracket 1, n \rrbracket$ <sup>24</sup>. La probabilité pour que  $X_i$  prenne la valeur 1 est donc égale à  $1/n$ , et l'espérance de  $X_i$  est égale à

$$\mathbb{E}(X_i) = 0 \times \frac{n-1}{n} + 1 \times \frac{1}{n} = \frac{1}{n}$$

24. De manière plus rigoureuse, on peut voir qu'il existe  $n!$  permutations de  $\llbracket 1, n \rrbracket$  et seulement  $(n-1)!$  permutations de  $\llbracket 1, n \rrbracket$  fixant  $i$ , ce qui permet de voir par équiprobabilité que la probabilité que la permutation laisse  $i$  inchangé est bien égale à  $\frac{1}{n}$ .

Le nombre total de cadeaux correctement livrés étant égal à  $\sum_{i=1}^n X_i$ , on en déduit que son espérance est

$$\mathbb{E}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{E}(X_i) = n \times \frac{1}{n} = 1$$

et la réponse attendue est 1.

**Solution de l'exercice 2.31.** Supposons que  $X$  soit à valeurs dans  $\llbracket 0, N \rrbracket$  avec  $N \geq 0$ . Le fait que  $\mathbb{P}(X \leq n) = \sum_{k=0}^n \mathbb{P}(X = k)$  pour tout  $n \in \mathbb{N}$  provient simplement de l'écriture  $(X \leq n) = \sqcup_{k=0}^n (X = k)$  et d'un argument d'incompatibilité. Il en va de même pour le fait que  $\mathbb{P}(X \geq n) = \sum_{k=n}^{+\infty} \mathbb{P}(X = k)$ , la convergence de la série étant, on le rappelle, une conséquence directe de la définition de la fonction de probabilité  $\mathbb{P}$  (voir au besoin le photocopié de rappels).

$X$  prenant un nombre fini de valeurs réelles, elle admet bien une espérance. On a alors

$$\sum_{n=0}^{+\infty} \mathbb{P}(X \geq n) = \sum_{n=0}^N \mathbb{P}(X \geq n) = \sum_{n=1}^{+\infty} \mathbb{P}(X \geq n) = \sum_{n=1}^N \sum_{k=n}^N \mathbb{P}(X = k)$$

qui correspond à la somme des coefficients d'une matrice triangulaire, ce qui permet d'écrire en inversant l'ordre de sommation

$$\sum_{n=0}^{+\infty} \mathbb{P}(X > n) = \sum_{k=1}^N \sum_{n=1}^k \mathbb{P}(X = k) = \sum_{k=1}^N k \mathbb{P}(X = k) = \mathbb{E}(X)$$

**Solution de l'exercice 2.38.** L'inégalité donnée en indication provient bien sûr de l'égalité remarquable  $(x - y)^2 = x^2 + y^2 - 2xy$  valable pour tout  $(x, y) \in \mathbb{R}^2$ . On a donc  $0 \leq XY \leq \frac{1}{2}(X^2 + Y^2)$ , or la variable aléatoire de droite admet une espérance donc  $XY$  aussi.

Alors  $(X + Y)^2 = X^2 + Y^2 + 2XY$  s'écrit comme la somme de variables aléatoires réelles discrètes admettant une espérance, donc en admet elle-même une :  $X + Y$  admet donc un moment d'ordre 2.

Enfin,  $aX$  admet un moment d'ordre 2 par le théorème de transfert :  $\sum_{k \geq 0} k^2 \mathbb{P}(X = k)$  converge donc  $\sum_{k \geq 0} a^2 k^2 \mathbb{P}(X = k)$  aussi, et le théorème de transfert permet de conclure que  $(aX)^2$  admet une espérance, c'est-à-dire que  $aX$  admet un moment d'ordre 2.

**Solution de l'exercice 2.50.** Compte tenu du fait que l'espérance d'une variable aléatoire réelle discrète est une constante réelle et que l'espérance d'une constante réelle est cette même constante, on peut écrire par linéarité de l'espérance :

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY - \mathbb{E}(X)Y - X\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y)) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \end{aligned}$$

**Solution de l'exercice 2.55.** Pour  $n = 3$ , la formule s'écrit  $V(X_1 + X_2 + X_3) = V(X_1) + V(X_2) + V(X_3) + 2\text{Cov}(X_1, X_2) + 2\text{Cov}(X_1, X_3) + 2\text{Cov}(X_2, X_3)$ , à comparer avec la formule  $(x_1 + x_2 + x_3)^2 = x_1^2 + x_2^2 + x_3^2 + 2x_1x_2 + 2x_1x_3 + 2x_2x_3$  valable pour tout  $(x_1, x_2, x_3) \in \mathbb{R}^3$ .

**Solution de l'exercice 2.70.** Si  $N \in \mathbb{N}$ , alors

$$\mathbb{P}(X > N) = \sum_{i=N+1}^{+\infty} \mathbb{P}(X = i) = \sum_{i=N+1}^{+\infty} p(1-p)^{i-1} = p \frac{(1-p)^N}{1 - (1-p)} = (1-p)^N$$



Si  $k$  et  $n$  sont comme dans l'énoncé, comme  $(X > n) \subset (X > k)$  on a bien

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

ce qu'il fallait démontrer.

En interprétant  $X$  comme le rang du premier succès obtenu lors de la répétition d'expériences aléatoires indépendantes ayant chacune une probabilité  $p$  de succès,  $\mathbb{P}(X > N)$  est la probabilité pour que le temps d'attente du premier succès soit strictement supérieur à  $N$ .

La propriété que nous venons de démontrer exprime le fait que conditionnellement au fait que les  $k$  premières tentatives ont échoué, la probabilité pour qu'il faille faire strictement plus de  $n-k$  autres tentatives avant d'obtenir un succès est égale à la probabilité, évaluée avant la première expérience, pour qu'il faille faire strictement plus  $n-k$  tentatives avant d'obtenir un succès : en d'autres termes, les échecs passés n'influencent en aucune façon sur la loi du temps d'attente restant. C'est la raison pour laquelle on appelle « absence de mémoire » cette propriété.

**Solution de l'exercice 2.72.** Le raisonnement suivant, très classique, est à connaître.  $X + Y$  est à valeurs dans  $\mathbb{N}$  comme somme de variables aléatoires à valeurs dans  $\mathbb{N}$ , et pour tout  $n \in \mathbb{N}$  on a

$$\mathbb{P}(X + Y = n) = \mathbb{P}\left(\bigsqcup_{k=0}^n ((Y = k) \cap (X = n - k))\right)$$

Cette écriture revient à décomposer l'événement  $(X + Y = n)$  en différents cas en distinguant les valeurs que peut prendre  $Y$ .

On a donc pour tout  $n \in \mathbb{N}$ , par incompatibilité puis par indépendance de  $X$  et  $Y$  :

$$\begin{aligned} \mathbb{P}(X + Y = n) &= \sum_{k=0}^n \mathbb{P}((Y = k) \cap (X = n - k)) = \sum_{k=0}^n \mathbb{P}(Y = k) \mathbb{P}(X = n - k) \\ &= \sum_{k=0}^n e^{-\lambda} \frac{\lambda^k}{k!} e^{-\mu} \frac{\mu^{n-k}}{(n-k)!} = \sum_{k=0}^n e^{-(\lambda+\mu)} \lambda^k \mu^{n-k} \frac{\binom{n}{k}}{n!} \\ &= e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} \lambda^k \mu^{n-k} = e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^n}{n!} \end{aligned}$$

où la dernière égalité procède de la formule du binôme.  $X + Y$  suit donc bien une loi de Poisson de paramètre  $\lambda + \mu$ .

**Solution de l'exercice 2.75.** On se contente dans ce corrigé de donner quelques indications pour mener à bien les calculs, qu'il est de bon ton de savoir réaliser impeccablement devant un jury. Un corrigé plus complet pourrait donner l'illusion que ces calculs sont faciles, ce qui amènerait à les négliger et serait contre-productif.

La méthode générale de détermination de l'espérance et de la variance d'une variable aléatoire réelle discrète  $X$  de loi donnée est la suivante : on calcule en utilisant la définition de l'espérance donnée dans le cours, et le cas échéant en justifiant la convergence de la série concernée. On détermine ensuite  $\mathbb{E}(X^2)$ , si cette quantité existe (ce

sera toujours le cas ici), en utilisant le théorème de transfert, puis on utilise la formule de Koenig  $V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$  pour déterminer la variance de  $X$ . Pour prendre un exemple simple, si  $X \sim \mathcal{G}(p)$  avec  $p \in ]0, 1]$ , l'espérance de  $X^2$  est la somme de la série convergente  $\sum_{k \geq 1} k^2 p(1-p)^{k-1}$ ; la suite du raisonnement est purement calculatoire.

Voici donc quelques indications pour vous aider à compléter le tableau :

- les lois de Dirac et de Bernoulli ne devraient pas vous poser de problème insurmontable : utilisez la définition de l'espérance et de la variance et appliquez la méthode ci-dessus
- les calculs relatifs à la loi uniforme sur  $[1, n]$  utilisent les résultats bien connus donnant la valeur des sommes des premiers entiers et des premiers carrés
- dans le cas de la loi binomiale, souvenez-vous de l'écriture d'une variable aléatoire suivant une loi binomiale comme somme de variables de Bernoulli indépendantes
- vous pouvez aussi traiter le cas de la loi binomiale directement en utilisant les propriétés des coefficients binomiaux et la méthode ci-dessus, ce qui est plus calculatoire mais constitue un bon exercice
- le cas de la loi géométrique se traite à l'aide des sommes de séries géométriques (voir le cours correspondant); attention à ne pas faire d'erreur dans les indices de sommation!
- le cas de la loi de Poisson se traite par un changement d'indice de sommation et la relation  $k! = k(k-1)!$  valable pour tout  $k \in \mathbb{N}^*$

**Solution de l'exercice 3.25.** La méthode à adopter est classique et doit être maîtrisée. On détermine d'abord la fonction de répartition  $F_{\alpha X}$  de  $\alpha X$  que l'on reconnaît comme la fonction de répartition d'une loi connue.

Pour tout  $x \in \mathbb{R}_+$ , on a  $F_{\alpha X}(x) = \mathbb{P}(\alpha X \leq x) = \mathbb{P}(X \leq \frac{x}{\alpha}) = \int_{-\infty}^{\frac{x}{\alpha}} \lambda e^{-\lambda t} dt = 1 - e^{-\lambda \frac{x}{\alpha}}$ ; pour tout  $x \in \mathbb{R}_-$  on a  $F_{\alpha X}(x) = 0$ . On reconnaît immédiatement<sup>25</sup> la fonction de répartition d'une loi exponentielle de paramètre  $\frac{\lambda}{\alpha}$ , ce qui permet de conclure que  $\alpha X \sim \mathcal{E}\left(\frac{\lambda}{\alpha}\right)$ .

**Solution de l'exercice 3.37.** Cette fois encore, nous nous contenterons de donner quelques indications et astuces de calcul pour résoudre cet exercice.

- le calcul de la variance d'une variable aléatoire à densité à partir de ses moments d'ordre 1 et 2 s'effectue de la même façon que dans le cas discret
- le cas de la loi uniforme ne devrait pas poser problème aux étudiants qui maîtrisent un tant soit peu l'intégration des fonctions polynomiales
- les calculs dans le cas exponentiel demandent une intégration par parties visant à abaisser le degré du monôme présent dans le produit en intégrant
- l'espérance de la loi gaussienne standard se calcule en utilisant l'impairité de l'intégrande après avoir justifié la convergence de l'intégrale considérée, et on détermine son moment d'ordre 2 par passage à la limite après avoir réalisé une intégration par parties sur un segment de  $\mathbb{R}$  dans le but de faire disparaître le terme polynomial (attention à écrire correctement la primitive du terme exponentiel!)

25. Si! En effet, il suffit de prendre  $\alpha = 1$  pour avoir sous les yeux la fonction de répartition générique d'une loi exponentielle; reste alors à remarquer que  $\lambda \frac{x}{\alpha} = \frac{\lambda}{\alpha} x$ ...

- la dernière ligne se traite directement à partir de l'avant-dernière en utilisant la définition des lois normales de paramètres quelconques

**Solution de l'exercice 4.14.** Montrons l'inégalité de Markov. La variable aléatoire réelle  $\mathbf{1}_{\{X \geq \varepsilon\}}$ , qui vaut 1 lorsque  $X \geq \varepsilon$  et 0 sinon, vérifie  $\varepsilon \mathbf{1}_{\{X \geq \varepsilon\}} \leq X$ . On a donc, par croissance de l'espérance :

$$\varepsilon \mathbb{E}(\mathbf{1}_{\{X \geq \varepsilon\}}) \leq \mathbb{E}(X)$$

d'où le résultat en divisant par  $\varepsilon$  puisque  $\mathbb{E}(\mathbf{1}_{\{X \geq \varepsilon\}}) = \mathbb{P}(X \geq \varepsilon)$ .

L'inégalité de Bienaymé-Tchebychev s'en déduit aussitôt en écrivant  $\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) = \mathbb{P}(|X - \mathbb{E}(X)|^2 \geq \varepsilon^2)$  (puisque les deux événements dont on calcule la probabilité sont en fait égaux) puis en appliquant l'inégalité de Markov à la variable positive  $|X - \mathbb{E}(X)|^2$  et à  $\varepsilon^2$ .

Enfin, la loi faible des grands nombres se démontre en appliquant l'inégalité de Bienaymé-Tchebychev à la variable aléatoire  $\overline{X}_n$  pour tout  $n \in \mathbb{N}^*$  et en utilisant le fait que si  $n \geq 1$  on a

$$\mathbb{E}(\overline{X}_n) = \mu$$

et

$$V(\overline{X}_n) = \frac{1}{n^2} \sum_{k=1}^n V(X_k) \leq \frac{M}{n} \xrightarrow{n \rightarrow +\infty} 0$$

par indépendance des  $X_i$  et majoration de leur variance.

Assurez-vous bien que vous comprenez et savez reproduire le raisonnement qui permet d'écrire la ligne de calculs ci-dessus, volontairement peu détaillée : ce calcul fait parfois l'objet d'une question du jury!

**Solution de l'exercice 4.21.** Comme toutes les lois considérées sont à valeurs dans  $\mathbb{N}$ , il suffit de vérifier que pour tout  $k \in \mathbb{N}$  on a  $\mathbb{P}(X_n = k) \xrightarrow{n \rightarrow +\infty} e^{-\lambda} \frac{\lambda^k}{k!}$ . Or on a pour tout  $k \in \mathbb{N}$  :

$$\mathbb{P}(X_n = k) = \frac{n!}{k!(n-k)!} p_n^k (1-p_n)^{n-k} = \frac{n!}{(n-k)!} \frac{1}{k!} \left( \frac{p_n}{1-p_n} \right)^k (1-p_n)^n$$

mais

$$\frac{n!}{(n-k)!} = n(n-1)\dots(n-k+1) \underset{n \rightarrow +\infty}{\sim} n^k$$

et

$$\frac{p_n}{1-p_n} \underset{n \rightarrow +\infty}{\sim} \frac{\lambda}{n}$$

donc

$$\mathbb{P}(X_n = k) \underset{n \rightarrow +\infty}{\sim} \frac{n^k}{k!} \frac{\lambda^k}{n^k} \left( 1 - \frac{np_n}{n} \right)^n \xrightarrow{n \rightarrow +\infty} e^{-\lambda} \frac{\lambda^k}{k!}$$

compte tenu du fait que  $np_n \xrightarrow{n \rightarrow +\infty} \lambda$ , ce qui clôt la preuve.

**Solution de l'exercice 4.28.** Si  $\lambda \in \mathbb{R}_+^*$  et si  $(X_n)_{n \geq 1}$  est une suite de variables aléatoires mutuellement indépendantes de loi  $\mathcal{P}(\lambda)$ , alors

$$\sqrt{n} \frac{\overline{X}_n - \lambda}{\sqrt{\lambda}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

**Solution de l'exercice 4.35.** Si  $X$  admet une variance non nulle, alors il admet également une espérance. Par linéarité de l'espérance on a  $\mathbb{E}(\overline{X}_n) = \frac{1}{n}(\mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)) = \frac{1}{n}n\theta = \theta$ , puisque tous les  $X_i$  sont iid. En revanche, l'estimateur proposé pour la variance est biaisé. Un calcul classique donne d'abord que :

$$\mathbb{E}(\hat{\sigma}_n^2) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}((X_i - \overline{X}_n)^2) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i^2) - \mathbb{E}(\overline{X}_n^2)$$

Or, les  $X_i$  sont iid, donc  $\mathbb{E}[X_i^2] = \mathbb{E}[X_1^2]$  pour tout  $i$ . En utilisant ceci et en développant la deuxième somme, on a donc :

$$\mathbb{E}(\hat{\sigma}_n^2) = \mathbb{E}(X_1^2) - \frac{1}{n^2} \mathbb{E} \left( \sum_{k=1}^n X_k^2 + 2 \sum_{j < i} X_i X_j \right) = \mathbb{E}(X_1^2) - \frac{1}{n} \mathbb{E}(X_1^2) - \frac{2}{n^2} \sum_{i < j} \theta^2$$

La deuxième somme comprend  $n(n-1)/2$  termes, donc on a :

$$\mathbb{E}(\hat{\sigma}_n^2) = \frac{n-1}{n} \mathbb{E}(X_1^2) - \frac{2}{n^2} \frac{n(n-1)}{2} \theta^2$$

En factorisant on obtient donc

$$\mathbb{E}(\hat{\sigma}_n^2) = \frac{n-1}{n} (\mathbb{E}(X_1^2) - \theta^2) = \frac{n-1}{n} \sigma^2 \neq \sigma^2$$

Cet estimateur est donc biaisé. En revanche, l'estimateur  $\frac{n}{n-1} \hat{\sigma}_n^2$  ne l'est plus.

**Solution de l'exercice 4.38.** Il suffit d'écrire l'inégalité de Bienaymé-Tchebychev pour  $\hat{\theta}_n$  (d'espérance  $\theta$  et de variance  $\sigma_n^2$ ) pour tout  $\varepsilon > 0$  et de faire tendre  $n$  vers  $+\infty$ .

**Solution de l'exercice 4.49.** Il suffit de reproduire le raisonnement de l'exemple précédent en utilisant non plus le fait que  $X_1 - \mu \sim \mathcal{N}(0, 1)$  mais le fait que  $\overline{X}_n - \mu$  suit la loi  $\mathcal{N}(0, \frac{1}{n})$  (retrouvez la preuve de ce fait s'il ne vous semble pas évident!) et vérifie donc  $\sqrt{n}(\overline{X}_n - \mu) \sim \mathcal{N}(0, 1)$ . On remarque que la longueur de l'intervalle obtenu est égale à  $\frac{1}{\sqrt{n}}$  fois celle du précédent, et donc que l'estimation est plus précise. C'est tout à fait conforme à l'intuition puisque l'on utilise dans ce dernier cas toutes les informations dont on dispose, tandis que l'exemple précédent faisait seulement usage de la première observation!

**Solution de l'exercice 4.55.** On se ramène bien sûr au résultat donné par l'exemple qui détaille la construction d'intervalles de confiance dans cas du problème du sondage. On peut bien utiliser ce résultat puisque les observations sont réalisées de manière indépendantes les unes des autres. Notez que l'on suppose aussi que le protocole expérimental choisi relève du tirage avec remise, ce qui n'est pas crédible si l'expérimentateur refuse d'interroger deux fois la même personne, mais que l'importance de la population française permet de supposer, même en présence d'un tirage sans remise, que la proportion  $p$  de personnes pensant que la Terre est plate *au sein de la population non encore interrogée* est constante au cours de l'expérience. On peut aussi supposer que  $p \in ]0, 1[$  puisque notre échantillon contient une personne de chaque opinion. Enfin, on est en droit d'utiliser un intervalle de confiance asymptotique puisque l'échantillon dont on dispose est de grande taille (10000).

L'exemple 4.53 donne alors l'intervalle de confiance attendu, que l'on calcule en notant que  $\overline{X}_n = 0,2104$  : il s'agit de

$$\begin{aligned} C'_{10000} &= \left[ \overline{X}_n - \frac{\Phi^{-1}(1 - \frac{0,05}{2})}{2\sqrt{100000}}; \overline{X}_n + \frac{\Phi^{-1}(1 - \frac{0,05}{2})}{2\sqrt{100000}} \right] \\ &= \left[ 0,2104 - \frac{1,96}{200}; 0,2104 + \frac{1,96}{200} \right] = [0,2006; 0,2202] \end{aligned}$$

Remarquez qu'il est de bon ton d'être capable de réaliser de tête ce type de calcul simple<sup>26</sup>.

On peut donc affirmer avec une certitude de 95% qu'une part de la population française comprise entre 20,06% et 22,02% est convaincue que la Terre est plate. Ces chiffres sont bien entendu fictifs...

---

26. Il n'y a aucune honte à s'entraîner à poser des multiplications et des divisions pour retrouver ses réflexes de jeunesse... mais il y en a à perdre des points par incapacité à terminer un calcul élémentaire!