

# Introduction aux probabilités et aux statistiques

Pierre-Loïc Méliot



## Probabilités discrètes et dénombrement

**Contenu :** espace de probabilité discret, événements, indépendance, conditionnement ; cas de la loi uniforme sur un ensemble et problèmes de dénombrement (cardinal d'une union, d'une intersection, d'un ensemble de parties, d'un ensemble de suites ou d'arrangements).

### 1. Probabilités discrètes

On souhaite décrire mathématiquement une expérience aléatoire, c'est-à-dire une expérience dont le résultat est *a priori* inconnu de l'expérimentateur et appartient à un ensemble de plusieurs valeurs possibles, chacune de ces possibilités ayant une certaine chance d'être réalisée. Par exemple, lors d'un lancer de dé à six faces, le résultat de l'expérience est l'un des éléments de l'ensemble

$$\llbracket 1, 6 \rrbracket = \{1, 2, 3, 4, 5, 6\},$$

et chacun de ces éléments a une certaine "chance" d'être obtenu (si le dé est équilibré, on s'attend à ce que toutes ces chances soient égales).

Le cadre des probabilités discrètes rend rigoureux la description de telles expériences, lorsque l'ensemble des possibilités est fini, ou éventuellement infini mais dénombrable (c'est-à-dire en bijection avec l'ensemble  $\mathbb{N}$  des entiers naturels).

**DÉFINITION 1.1.** *Un espace de probabilité discret est la donnée d'un ensemble  $\Omega$  fini ou dénombrable (l'ensemble de toutes les possibilités pour l'expérience aléatoire), et d'une fonction*

$$\mathbb{P} : \mathfrak{P}(\Omega) \rightarrow [0, 1]$$

sur l'ensemble  $\mathfrak{P}(\Omega)$  des parties de  $\Omega$ , telle que :

- (1)  $\mathbb{P}[\emptyset] = 0$ ,  $\mathbb{P}[\Omega] = 1$  ;
- (2) si  $(A_i)_{i \in I}$  est une famille de parties disjointes de  $\Omega$ , alors

$$\mathbb{P} \left[ \bigsqcup_{i \in I} A_i \right] = \sum_{i \in I} \mathbb{P}[A_i].$$

Si  $A \subset \Omega$ , la quantité  $\mathbb{P}[A]$  représente la chance ou **probabilité** que le résultat  $\omega$  de l'expérience aléatoire appartienne à la partie  $A$ .

Les parties  $A \subset \Omega$  sont appelées des **événements**, et un espace de probabilité discret est donné par une fonction  $\mathbb{P}$  qui mesure les probabilités de *tous* les événements possibles.

EXEMPLE 1.2. Considérons de nouveau le lancer d'un dé à six faces, que l'on supposera équilibré. L'espace de probabilité qui décrit cette expérience est

$$\Omega = \llbracket 1, 6 \rrbracket \quad ; \quad \mathbb{P}[A \subset \llbracket 1, 6 \rrbracket] = \frac{\text{card } A}{6},$$

où  $\text{card } A$  désigne le nombre d'éléments de la partie  $A$ . Par exemple, la probabilité d'obtenir un résultat pair est

$$\mathbb{P}[\{2, 4, 6\}] = \frac{3}{6} = \frac{1}{2},$$

et la probabilité d'obtenir un 6 est  $\mathbb{P}[\{6\}] = \frac{1}{6}$ .

EXEMPLE 1.3. Le cas le plus simple d'expérience aléatoire non triviale est celui où  $\Omega = \{0, 1\}$  a deux éléments, le 1 représentant le succès de l'expérience et portant une probabilité  $p$ , et le 0 représentant l'échec de l'expérience et portant une probabilité  $q$ . Comme  $\{0, 1\} = \{0\} \sqcup \{1\}$ ,

$$1 = \mathbb{P}[\{0, 1\}] = \mathbb{P}[\{0\}] + \mathbb{P}[\{1\}] = q + p,$$

donc on a obligatoirement dans ce cadre  $q = 1 - p$ . On parle d'**expérience de Bernoulli** de paramètre  $p$ .

REMARQUE 1.4. Soit  $(\Omega, \mathbb{P})$  un espace de probabilité discret qui est *fini* :  $\Omega$  a un nombre  $n < +\infty$  d'éléments. Le cardinal de  $\mathfrak{P}(\Omega)$  est alors

$$\text{card}(\mathfrak{P}(\Omega)) = 2^{\text{card } \Omega} = 2^n,$$

voir la fin de ce chapitre. Donc, une probabilité sur un ensemble  $\Omega$  à  $n$  éléments est donnée par  $2^n$  valeurs. Toutefois, compte tenu de la condition de compatibilité avec les unions disjointes, il suffit pour décrire une probabilité  $\mathbb{P}$  sur  $\Omega$  de connaître les probabilités de tous les singletons  $\{\omega\}$  avec  $\omega \in \Omega$ . En effet, on a ensuite

$$\mathbb{P}[A] = \mathbb{P}\left[\bigsqcup_{\omega \in A} \{\omega\}\right] = \sum_{\omega \in A} \mathbb{P}[\{\omega\}]$$

pour tout événement  $A$ . On note souvent la probabilité d'un singleton  $\mathbb{P}[\{\omega\}] = \mathbb{P}[\omega]$ . On vient de voir que pour connaître pour une probabilité sur un ensemble fini, il suffit de donner sa valeur sur les singletons ( $n$  informations requises au lieu de  $2^n$ ). Ce résultat est également vrai pour une probabilité sur un ensemble infini dénombrable, mais dans ce cas, il faut tout de même connaître une infinité de valeurs pour déterminer  $\mathbb{P}$ .

REMARQUE 1.5. La condition de compatibilité avec les unions disjointes implique la *croissance* de toute probabilité  $\mathbb{P}$  : si  $A$  et  $B$  sont deux événements d'un espace de probabilité discret  $(\Omega, \mathbb{P})$  avec  $A \subset B$ , alors  $\mathbb{P}[A] \leq \mathbb{P}[B]$ . En effet, notons  $B \setminus A$  l'ensemble des  $\omega$  qui appartiennent à  $B$  et n'appartiennent pas à  $A$ . Alors,  $B = A \sqcup (B \setminus A)$ , donc,

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

Un cas particulier est celui où  $B = \Omega$ . Rappelons que le complémentaire de  $A \subset \Omega$  est la partie notée  $A^c$ , ou  $\bar{A}$ , ou  $\Omega \setminus A$ , qui contient tous les  $\omega \in \Omega$  qui ne sont pas dans  $A$ . On a alors

$$1 = \mathbb{P}[\Omega] = \mathbb{P}[A \sqcup A^c] = \mathbb{P}[A] + \mathbb{P}[A^c],$$

donc  $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$  pour tout événement  $A$ .

REMARQUE 1.6. Attention, la probabilité d'une union d'événements  $(A_i)_{i \in I}$  est égale à la somme des probabilités des événements seulement lorsque ceux-ci sont disjoints, *i.e.*,  $A_i \cap A_j = \emptyset$  si  $i \neq j$ . Supposons donnés deux événements  $A$  et  $B$  non disjoints. Alors, on peut écrire

$$A \cup B = A \sqcup (B \setminus (A \cap B)) = (A \setminus (A \cap B)) \sqcup B,$$

donc

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B].$$

En particulier, on a toujours  $\mathbb{P}[A \cup B] \leq \mathbb{P}[A] + \mathbb{P}[B]$ , et plus généralement,

$$\mathbb{P}\left[\bigcup_{i=1}^r A_i\right] \leq \sum_{i=1}^r \mathbb{P}[A_i].$$

Il existe plus généralement des formules pour la probabilité  $\mathbb{P}[A_1 \cup A_2 \cup \dots \cup A_r]$  d'une union quelconque d'événements, mais elles deviennent vite assez compliquées. Par exemple, pour trois événements quelconques  $A, B, C$ ,

$$\mathbb{P}[A \cup B \cup C] = \mathbb{P}[A] + \mathbb{P}[B] + \mathbb{P}[C] - \mathbb{P}[A \cap B] - \mathbb{P}[A \cap C] - \mathbb{P}[B \cap C] + \mathbb{P}[A \cap B \cap C],$$

comme on peut le voir en appliquant plusieurs fois la formule pour deux événements à  $A \cup B \cup C = A \cup (B \cup C)$ .

## 2. Indépendance et conditionnement

Soit  $(\Omega, \mathbb{P})$  un ensemble de probabilité discret. Si  $B \in \mathfrak{F}(\Omega)$  est un événement, on peut se demander comment les probabilités des réalisations  $\omega \in \Omega$  de l'expérience sont modifiées si l'on a l'information partielle  $\omega \in B$ . Ce problème mène à la notion de probabilité conditionnelle :

DÉFINITION 1.7. Soit  $B$  un événement de probabilité non nulle :  $\mathbb{P}[B] > 0$ . La **probabilité conditionnelle** à  $B$ , ou **probabilité sachant**  $B$ , est la fonction  $\mathbb{P}[\cdot | B] : \mathfrak{F}(\Omega) \rightarrow [0, 1]$  définie par

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

La paire  $(\Omega, \mathbb{P}[\cdot | B])$  est un nouvel espace de probabilité.

EXEMPLE 1.8. Considérons le lancer d'un dé à six faces, et la probabilité sachant que le résultat obtenu est pair, c'est-à-dire sachant

$$B = \{2, 4, 6\}.$$

Si  $A$  est l'événement "le lancer donne un résultat inférieur ou égal à 3", alors

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[\{1, 2, 3\} \cap \{2, 4, 6\}]}{\mathbb{P}[\{2, 4, 6\}]} = \frac{\mathbb{P}[\{2\}]}{\mathbb{P}[\{2, 4, 6\}]} = \frac{1}{3},$$

qui est différent de  $\mathbb{P}[A] = \frac{1}{2}$ . La nouvelle probabilité conditionnelle  $\mathbb{P}[\cdot | B]$  est donc bien différente de  $\mathbb{P}[\cdot]$ ; autrement dit, savoir *a priori* que le résultat du lancer sera pair modifie les probabilités des événements.

Les probabilités conditionnelles permettent souvent de simplifier le calcul de la probabilité d'un événement  $A \subset \Omega$ , en décomposant cet événement en fonction d'une suite d'alternatives. Plus précisément, soit  $(B_i)_{i \in I}$  un **système complet** d'événements disjoints tous de probabilité non nulle, c'est-à-dire que  $\bigsqcup_{i \in I} B_i = \Omega$ . Alors :

PROPOSITION 1.9 (Formule des probabilités totales). *Pour tout événement  $A \subset B$ , on a*

$$\mathbb{P}[A] = \sum_{i \in I} \mathbb{P}[B_i] \mathbb{P}[A|B_i].$$

DÉMONSTRATION. On peut écrire

$$\mathbb{P}[A] = \mathbb{P}[A \cap \Omega] = \mathbb{P}\left[\bigsqcup_{i \in I} (A \cap B_i)\right] = \sum_{i \in I} \mathbb{P}[A \cap B_i],$$

puis,  $\mathbb{P}[A \cap B_i] = \mathbb{P}[B_i] \mathbb{P}[A|B_i]$  pour tout indice  $i$ . □

EXEMPLE 1.10. On considère un tirage uniforme de trois cartes parmi un paquet de 52 cartes (4 couleurs, et pour chaque couleur 13 cartes allant de l'as 1 au roi  $R$ ). Le nombre de possibilités pour un tel tirage est

$$\binom{52}{3} = \frac{52 \times 51 \times 50}{3 \times 2 \times 1} = 22100,$$

voir plus loin les techniques de dénombrement. On cherche à calculer la probabilité de l'événement :

$A$  = "tous les coeurs tirés  $\heartsuit$  parmi les trois cartes sont de valeur supérieure au valet".

Pour faire ce calcul, on peut décomposer l'événement  $A$  en fonction du système complet d'événements  $(B_0, B_1, B_2, B_3)$ , avec

$B_k$  = "le tirage contient  $k$  coeurs  $\heartsuit$ ".

Calculons d'abord la probabilité de chaque événement  $B_k$ . Le nombre d'ensembles de trois cartes avec  $k$  cartes  $\heartsuit$  et  $3 - k$  cartes  $\diamondsuit, \spadesuit, \clubsuit$  est

$$\binom{13}{k} \binom{39}{3-k} = \begin{cases} 9139 & \text{si } k = 0, \\ 9633 & \text{si } k = 1, \\ 3042 & \text{si } k = 2, \\ 286 & \text{si } k = 3. \end{cases}$$

Les probabilités des  $B_k$  sont donc respectivement  $\frac{9139}{22100}$ ,  $\frac{9633}{22100}$ ,  $\frac{3042}{22100}$  et  $\frac{286}{22100}$ . Sachant maintenant qu'il y a  $k$  cartes  $\heartsuit$ , la probabilité que toutes soient supérieures au valet (donc, égales au valet  $V$ , à la dame  $D$  ou au roi  $R$ ) est

$$\frac{\binom{3}{k}}{\binom{13}{k}}.$$

En effet, il y a  $\binom{13}{k}$  possibilités pour les tirages des  $k$  cartes  $\heartsuit$ , et parmi ceux-ci,  $\binom{3}{k}$  ensembles de cartes avec seulement des cartes dans  $\{V, D, R\}$ . Par la formule des probabilités totales, on a donc

$$\mathbb{P}[A] = \sum_{k=0}^3 \frac{\binom{13}{k} \binom{39}{3-k}}{\binom{52}{3}} \frac{\binom{3}{k}}{\binom{13}{k}} = \sum_{k=0}^3 \frac{\binom{3}{k} \binom{39}{3-k}}{\binom{52}{3}} = \frac{574}{1105} = 0.519\dots$$

Si  $A$  et  $B$  sont deux événements de probabilités non nulles, il n'y a pas en général de lien entre  $\mathbb{P}[A|B]$  et  $\mathbb{P}[B|A]$ . Néanmoins, étant donné un système complet d'événements disjoints, on a :

PROPOSITION 1.11 (Formule de Bayes). *Soit  $(B_i)_{i \in I}$  un système complet d'événements disjoints tous de probabilité non nulle, et  $A$  un autre événement de probabilité non nulle. On a pour tout  $i \in I$  :*

$$\mathbb{P}[B_i|A] = \frac{\mathbb{P}[B_i] \mathbb{P}[A|B_i]}{\sum_{j \in I} \mathbb{P}[B_j] \mathbb{P}[A|B_j]}.$$

DÉMONSTRATION. On calcule

$$\mathbb{P}[B_i|A] = \frac{\mathbb{P}[A \cap B_i]}{\mathbb{P}[A]} = \frac{\mathbb{P}[B_i] \mathbb{P}[A|B_i]}{\sum_{j \in I} \mathbb{P}[B_j] \mathbb{P}[A|B_j]}$$

en utilisant la formule des probabilités totales au dénominateur.  $\square$

EXEMPLE 1.12. On considère une population  $\Omega$  de 1000 individus, certains d'entre eux étant porteurs d'une grave maladie. On note  $M \subset \Omega$  l'ensemble des individus malades, et  $S = \Omega \setminus M$  l'ensemble des individus sains. On fait subir à la population un test médical, qui sépare la population en deux parties  $B_+$  et  $B_-$  :  $B_+$  est l'ensemble des individus testés positifs, et  $B_-$  l'ensemble des individus testés négatifs. On tire au hasard uniformément l'un des individus  $\omega \in \Omega$ . Comme le test n'est pas parfait, on supposera :

- (1) Conditionnellement au fait que  $\omega$  soit sain, le test est positif avec probabilité  $\alpha$  proche de 0.
- (2) Conditionnellement au fait que  $\omega$  soit malade, le test est négatif avec probabilité  $\beta$ , avec  $\beta$  proche de 0.

Autrement dit,

$$\mathbb{P}[B_+|M] = 1 - \beta \quad \text{et} \quad \mathbb{P}[B_+|S] = \alpha.$$

Les paramètres  $\alpha$  et  $\beta$  mesurent les erreurs que l'on peut faire durant le test. Soit  $x = \mathbb{P}[M]$  la probabilité pour que  $\omega$  soit malade, c'est-à-dire la proportion de gens malades :

$$x = \frac{\text{card } M}{1000}.$$

Calculons la probabilité pour qu'un individu testé positif soit effectivement malade. Par la formule de Bayes, il s'agit de

$$\mathbb{P}[M|B_+] = \frac{\mathbb{P}[B_+|M] \mathbb{P}[M]}{\mathbb{P}[B_+|M] \mathbb{P}[M] + \mathbb{P}[B_+|S] \mathbb{P}[S]} = \frac{(1-\beta)x}{(1-\beta)x + \alpha(1-x)}.$$

Supposons  $\alpha = 5\%$ ,  $\beta = 1\%$  (il est plus grave de ne pas traiter un individu malade, que de traiter un individu sain, de sorte qu'on s'arrange souvent pour que  $\alpha \gg \beta$ ). Alors, en fonction de  $x$ ,

$$\mathbb{P}[M|B_+] = \frac{99x}{94x + 5}.$$

Si  $x = \frac{1}{100}$  (maladie rare), alors  $\mathbb{P}[M|B_+] = \frac{99}{594} = \frac{1}{6}$  est relativement faible : il faut faire un second test plus spécifique sur les individus testés positifs pour déterminer s'ils sont effectivement malades (mais on n'a laissé passer presque aucun malade au travers du premier test). En revanche, lorsque  $x = \frac{1}{3}$  (épidémie), on a  $\mathbb{P}[M|B_+] = \frac{99}{109}$  qui est proche de 1, donc la plupart des individus testés positifs sont effectivement malades.

On a vu précédemment que  $\mathbb{P}[A]$  et  $\mathbb{P}[A|B]$  ont en général des valeurs différentes. Le cas particulier où ces deux valeurs sont égales, c'est-à-dire où la connaissance de  $B$  ne modifie pas la probabilité de  $A$ , mène à la notion d'**indépendance** :

**DÉFINITION 1.13.** Soient  $A$  et  $B$  deux événements d'un espace de probabilité discret  $(\Omega, \mathbb{P})$ . On dit que  $A$  et  $B$  sont indépendants si

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

Si  $B$  est de probabilité non nulle, ceci revient à demander que  $\mathbb{P}[A] = \mathbb{P}[A|B]$ .

**EXEMPLE 1.14.** Pour le lancer de dé, considérons les trois événements

$$A = \text{"le lancer est pair"} = \{2, 4, 6\}$$

$$B = \text{"le lancer est plus grand que 5"} = \{5, 6\}$$

$$C = \text{"le lancer est plus petit que 5"} = \{1, 2, 3, 4, 5\}.$$

Les événements  $A$  et  $B$  sont indépendants, car

$$\mathbb{P}[A \cap B] = \mathbb{P}[\{6\}] = \frac{1}{6} = \frac{1}{2} \times \frac{1}{3} = \mathbb{P}[A] \mathbb{P}[B].$$

En revanche,  $A$  et  $C$  ne sont pas indépendants, car

$$\mathbb{P}[A \cap C] = \mathbb{P}[\{2, 4\}] = \frac{1}{3} \neq \frac{1}{2} \times \frac{5}{6}.$$

**REMARQUE 1.15.** La notion *mathématique* d'indépendance permet de modéliser la notion *intuitive* d'expériences aléatoires consécutives et sans influence les unes sur les autres. Considérons par exemple l'ensemble  $\Omega$  de tous les lancers de deux dés à six faces, c'est-à-dire que  $\Omega = \llbracket 1, 6 \rrbracket \times \llbracket 1, 6 \rrbracket$ . On équipe  $\Omega$  de la probabilité uniforme :  $\mathbb{P}[(a, b)] = \frac{1}{36}$  pour tout couple  $(a, b) \in \llbracket 1, 6 \rrbracket^2$ . On s'attend à ce que les événements  $A = \{a \text{ est pair}\}$  et  $B = \{b \text{ est plus grand que 5}\}$  soient indépendants dans l'espace de probabilité  $(\Omega, \mathbb{P})$  : en



effet, ces événements dépendent de deux expériences consécutives (le lancer du premier dé et le lancer du second dé) qui sont sans influence l'une sur l'autre. On a bien :

$$\begin{aligned}\mathbb{P}[A] &= \frac{\text{card}(\{2, 4, 6\} \times \llbracket 1, 6 \rrbracket)}{36} = \frac{18}{36} = \frac{1}{2}; \\ \mathbb{P}[B] &= \frac{\text{card}(\llbracket 1, 6 \rrbracket \times \{5, 6\})}{36} = \frac{12}{36} = \frac{1}{3}; \\ \mathbb{P}[A \cap B] &= \frac{\text{card}(\{2, 4, 6\} \times \{5, 6\})}{36} = \frac{6}{36} = \frac{1}{6} = \mathbb{P}[A] \mathbb{P}[B].\end{aligned}$$

On pourra retenir la règle de calcul suivante : des expériences consécutives correspondent à la notion d'indépendance et donc à des *produits* de probabilités (à l'inverse, des événements simultanés et alternatifs, c'est-à-dire disjoints, correspondent à des *sommes* de probabilités).

Pour conclure cette section, détaillons la notion d'indépendance pour plus de deux événements. Si  $A_1, \dots, A_n$  sont  $n$  événements dans un espace de probabilité  $\Omega$ , on dit qu'ils sont indépendants si, pour tout choix de certains de ces événements  $A_{j_1}, A_{j_2}, \dots, A_{j_r}$  avec  $2 \leq r \leq n$ , on a

$$\mathbb{P}[A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_r}] = \mathbb{P}[A_{j_1}] \mathbb{P}[A_{j_2}] \dots \mathbb{P}[A_{j_r}].$$

Attention, c'est beaucoup plus fort que de demander simplement que  $\mathbb{P}[A_i \cap A_j] = \mathbb{P}[A_i] \mathbb{P}[A_j]$  pour tout  $i \neq j$ . Par exemple, avec 3 événements  $A, B, C$ , l'indépendance revient à demander que  $\mathbb{P}[A \cap B] = \mathbb{P}[A] \mathbb{P}[B]$ ,  $\mathbb{P}[A \cap C] = \mathbb{P}[A] \mathbb{P}[C]$  et  $\mathbb{P}[B \cap C] = \mathbb{P}[B] \mathbb{P}[C]$ , mais aussi que

$$\mathbb{P}[A \cap B \cap C] = \mathbb{P}[A] \mathbb{P}[B] \mathbb{P}[C],$$

ce qui n'est pas une conséquence des identités avec les intersections de deux événements.

### 3. Problèmes de dénombrement

Un cas particulier mais courant d'espace de probabilité discret est le suivant :  $\Omega$  est un ensemble fini et chaque  $\omega \in \Omega$  a la même probabilité (**probabilité uniforme** sur  $\Omega$ ). On a alors :

$$1 = \mathbb{P}[\Omega] = \sum_{\omega \in \Omega} \mathbb{P}[\omega] = p(\text{card } \Omega),$$

où  $p$  est cette probabilité commune à tous les singletons ; donc,  $p = \frac{1}{\text{card } \Omega}$ . Alors, pour tout événement  $A \subset \Omega$ ,

$$\mathbb{P}[A] = \sum_{\omega \in A} \frac{1}{\text{card } \Omega} = \frac{\text{card } A}{\text{card } \Omega}.$$

Cette formule est souvent retenue sous la forme

$$\mathbb{P}[A] = \frac{\text{nombre de possibilités } \omega \text{ favorables}}{\text{nombre total de possibilités}}.$$

Dans ce contexte, la seule difficulté est le calcul des **cardinaux** (nombres d'objets) de chaque ensemble, qui peut être assez complexe. On explique dans cette dernière section comment effectuer ces calculs dans diverses configurations classiques.

**3.1. Intervalles d'entiers.** Pour tous entiers  $a \leq b$ , on a

$$\text{card}(\llbracket a, b \rrbracket) = \text{card}(\{a, a+1, a+2, \dots, b-1, b\}) = b - a + 1;$$

(attention au +1). En particulier, on a bien sûr  $\text{card}(\llbracket 1, n \rrbracket) = n$  pour tout  $n \geq 1$ .

**3.2. Cardinal d'une union d'ensembles.** Si  $A$  et  $B$  sont des ensembles finis, pour énumérer leur réunion  $A \cup B$ , on peut d'abord énumérer les éléments de  $A$ , puis ceux de  $B$ ; on a alors compté au moins une fois tous les éléments de  $A \cup B$ , donc,

$$\text{card}(A \cup B) \leq \text{card} A + \text{card} B,$$

et plus généralement, étant donnée une famille d'ensembles finis  $A_1, A_2, \dots, A_n$ , on a toujours

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

En règle générale, il n'y a pas d'égalité, et c'est seulement le cas si les  $A_i$  sont deux à deux disjoints, c'est-à-dire si  $A_i \cap A_j = \emptyset$  dès que  $i \neq j$ . Ainsi :

PROPOSITION 1.16. *Si  $A_1, A_2, \dots, A_n$  est une famille d'ensembles finis disjoints, alors*

$$\text{card}\left(\bigsqcup_{i=1}^n A_i\right) = \sum_{i=1}^n \text{card} A_i.$$

Dans le cas d'ensemble non disjoints, on a également une formule, qui a une forme (et une preuve) analogue à celle donnée pour les probabilités d'événements non disjoints :

$$\text{card}(A \cup B) = \text{card} A + \text{card} B - \text{card}(A \cap B).$$

On trouvera dans les exercices la généralisation de cette formule au cas de  $n \geq 3$  ensembles finis.

**3.3. Cardinal d'un produit d'ensembles.** Si  $A$  et  $B$  sont deux ensembles, on note  $A \times B$  leur **produit cartésien**, qui est l'ensemble des paires ordonnées  $(a, b)$  avec  $a \in A$  et  $b \in B$ . Par exemple, avec  $A = \llbracket 1, 2 \rrbracket$  et  $B = \llbracket 1, 3 \rrbracket$ , le produit  $A \times B$  est l'ensemble des paires

$$\{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3)\}.$$

Notons qu'une paire  $(a, b)$  n'est pas le même objet que la paire  $(b, a)$  (en revanche, c'est le cas pour les ensembles sous-jacents, notés avec des accolades :  $\{a, b\} = \{b, a\}$ ). Il existe une formule très simple pour le cardinal d'un produit :

PROPOSITION 1.17. *Pour tous ensembles  $A$  et  $B$  finis,*

$$\text{card}(A \times B) = \text{card} A \times \text{card} B.$$

DÉMONSTRATION. Pour compter les paires dans  $A \times B$ , si  $A = \{a_1, a_2, \dots, a_n\}$  avec  $n = \text{card} A$ , alors on peut répartir les paires  $(a, b)$  suivant la valeur du premier élément  $a = a_i$  :

$$A \times B = \bigsqcup_{i=1}^n \{(a_i, b), b \in B\}.$$

Par exemple,  $\llbracket 1, 2 \rrbracket \times \llbracket 1, 3 \rrbracket = \{(1, 1), (1, 2), (1, 3)\} \sqcup \{(2, 1), (2, 2), (2, 3)\}$ . On déduit alors de la proposition sur le cardinal des unions disjointes que

$$\text{card}(A \times B) = \sum_{i=1}^n \text{card} B = n \text{card} B = \text{card} A \times \text{card} B.$$

□

EXEMPLE 1.18. Dans l'exemple précédent, on a bien  $6 = 2 \times 3$  paires dans  $\llbracket 1, 2 \rrbracket \times \llbracket 1, 3 \rrbracket$ .

Le résultat se généralise au cas d'un produit cartésien de plus de deux ensembles. Si  $A, B, \dots, Z$  sont des ensembles, on note  $A \times B \times \dots \times Z$  leur produit cartésien, qui est l'ensemble des suites ordonnées  $(a, b, \dots, z)$  avec  $a \in A, b \in B, \text{ etc.}, z \in Z$ . Le cardinal d'un produit cartésien d'un nombre quelconque d'ensembles finis est le produit des cardinaux :

$$\text{card}(A \times B \times \dots \times Z) = \text{card} A \times \text{card} B \times \dots \times \text{card} Z.$$

Un cas particulier est celui où tous les ensembles  $A, B, \dots, Z$  sont les mêmes. On note  $A^r = A \times A \times \dots \times A$  le produit cartésien de  $A$  avec lui-même  $r$  fois ; c'est l'ensemble des suites ordonnées  $(a_1, a_2, \dots, a_r)$  avec chaque  $a_i \in A$ . Le cardinal de  $A^r$  est  $(\text{card} A)^r$ .

EXEMPLE 1.19. Considérons tous les mots constitués de lettres 0 et 1, de longueur  $r$  (code binaire). Par exemple, 011101010 est un tel mot de longueur  $r = 9$ . Le nombre de mots de longueur  $r$  est  $2^r$ , puisqu'un mot de longueur  $r$  avec les lettres 0 ou 1 peut être vu comme une suite dans  $\{0, 1\}^r$ , avec  $\text{card}(\{0, 1\}) = 2$ .

**3.4. Cardinaux des arrangements et permutations.** Fixons un ensemble  $A$ . On rappelle que  $A^r$  est l'ensemble des suites d'éléments de  $A$  de longueur  $r$ . Un **arrangement** de taille  $r$  dans  $A$  est une telle suite  $(a_1, a_2, \dots, a_r)$ , qui ne contient pas deux fois le même élément. Par exemple, si  $A = \llbracket 0, 9 \rrbracket$ , alors  $(5, 8, 2, 4, 1)$  est un arrangement dans  $A$  de taille 5, mais  $(5, 8, 2, 4, 5)$  n'en est pas un, car 5 apparaît deux fois. On note  $\mathfrak{A}(A, r)$  l'ensemble des arrangements de taille  $r$  dans  $A$ . Il existe comme précédemment une formule pour le cardinal de  $\mathfrak{A}(A, r)$  en fonction de celui de  $A$  :

THÉORÈME 1.20. *Si  $A$  est un ensemble fini de cardinal  $n$ , alors le cardinal de  $\mathfrak{A}(A, r)$  est*

$$n^{\downarrow r} = n(n-1) \cdots (n-r+1),$$

*c'est-à-dire le produit de  $r$  entiers consécutifs, en partant de  $n$  et en allant en décroissant. Cette formule donne 0 si  $r > n$ .*

En effet, pour construire une suite  $(a_1, a_2, \dots, a_r)$  d'éléments de  $A$  sans répétition, il y a  $n$  choix possibles pour le premier élément  $a_1$  ; puis,  $n-1$  choix pour le second élément  $a_2$  (tous les éléments de  $A$ , sauf  $a_1$ ) ; puis,  $n-2$  choix pour le troisième élément  $a_3$  (tous les éléments de  $A$ , sauf  $a_1$  et  $a_2$ ) ; *etc.* jusqu'au choix du  $r$ -ième élément.

EXEMPLE 1.21. Si  $A = \llbracket 1, 4 \rrbracket$ , alors il y a bien  $4^{\downarrow 2} = 4 \times 3 = 12$  arrangements de taille 2 dans  $A$ , à savoir,  $(1, 2)$ ,  $(1, 3)$ ,  $(1, 4)$ ,  $(2, 1)$ ,  $(2, 3)$ ,  $(2, 4)$ ,  $(3, 1)$ ,  $(3, 2)$ ,  $(3, 4)$ ,  $(4, 1)$ ,  $(4, 2)$  et  $(4, 3)$ . Par contre, il n'y a pas d'arrangements de taille 5 dans  $A$ , car il n'est pas possible de faire une liste de 5 éléments distincts à partir des 4 éléments de  $A$ .

EXEMPLE 1.22. Lors du tirage des numéros du loto, le présentateur télé donne à la suite 6 numéros distincts et compris entre 1 et 49. Le nombre de telles suites possibles est

$$49^{\downarrow 6} = 49 \times 48 \times 47 \times 46 \times 45 \times 44 \simeq 1 \times 10^{10}.$$

Attention, ce n'est pas le nombre d'ensembles de 6 entiers distincts compris entre 1 et 49, car on fait ici attention à l'ordre des numéros.

Un cas particulier d'arrangement est une liste ordonnée de *tous* les éléments de  $A$ , c'est-à-dire, un arrangement de taille  $n = \text{card } A$ . On parle alors de **permutation** de  $A$ , et si  $n = \text{card } A$ , on note

$$n^{\downarrow n} = n! = n(n-1)(n-2) \cdots 1$$

le nombre de permutations de  $A$ , que l'on lit "factorielle  $n$ ". Par exemple, il y a  $6 = 3 \times 2 \times 1$  permutations de  $\llbracket 1, 3 \rrbracket$ , à savoir, les suites  $(1, 2, 3)$ ,  $(1, 3, 2)$ ,  $(2, 1, 3)$ ,  $(2, 3, 1)$ ,  $(3, 1, 2)$  et  $(3, 2, 1)$ . Notons que si  $r \leq n$ , alors

$$n^{\downarrow r} = n(n-1) \cdots (n-r+1) = \frac{n(n-1) \cdots 1}{(n-r)(n-r-1) \cdots 1} = \frac{n!}{(n-r)!}.$$

**3.5. Cardinaux des ensembles de parties.** Un dernier type de configuration que l'on peut compter est l'**ensemble des parties** de taille donnée dans un ensemble. Soit  $\Omega$  un ensemble de cardinal  $n$ ; on note  $\mathfrak{P}(\Omega)$  l'ensemble des parties  $A \subset \Omega$ , et  $\mathfrak{P}(\Omega, r)$  l'ensemble des parties  $A$  telles que  $\text{card } A = r$ . Par exemple, si  $\Omega = \llbracket 1, 3 \rrbracket$ , alors

$$\begin{aligned} \mathfrak{P}(\Omega, 0) &= \{\emptyset\} \\ \mathfrak{P}(\Omega, 1) &= \{\{1\}, \{2\}, \{3\}\} \\ \mathfrak{P}(\Omega, 2) &= \{\{1, 2\}, \{1, 3\}, \{2, 3\}\} \\ \mathfrak{P}(\Omega, 3) &= \{\{1, 2, 3\}\} \end{aligned}$$

et  $\mathfrak{P}(\Omega)$  est la réunion de ces ensembles, donc, compte 8 éléments. Notons que dans une partie (à l'inverse de ce qui se passe pour les arrangements), l'ordre n'est pas pris en compte : ainsi,  $\{1, 3\}$  et  $\{3, 1\}$  sont la même partie de taille 2 de  $\llbracket 1, 3 \rrbracket$ .

THÉORÈME 1.23. *Si  $\Omega$  est de taille  $n$  et si  $r \in \llbracket 0, n \rrbracket$ , alors l'ensemble  $\mathfrak{P}(\Omega, r)$  de ses parties de taille  $r$  a son cardinal donné par le coefficient binomial*

$$\binom{n}{r} = \frac{n^{\downarrow r}}{r!} = \frac{n!}{r!(n-r)!}.$$

L'ensemble de toutes les parties  $\mathfrak{P}(\Omega)$  a pour cardinal  $2^n$ .

DÉMONSTRATION. Pour la première formule, considérons l'application

$$\begin{aligned} \mathfrak{A}(\Omega, r) &\rightarrow \mathfrak{P}(\Omega, r) \\ (a_1, a_2, \dots, a_r) &\mapsto \{a_1, a_2, \dots, a_r\} \end{aligned}$$

qui oublie l'ordre dans un arrangement, et lui associe la partie sous-jacente. Chaque partie de taille  $r$  est atteinte par cette application, et ce plusieurs fois, puisqu'une même partie  $\{a_1, \dots, a_r\}$  est atteinte par toutes les permutations possibles de la suite  $(a_1, \dots, a_r)$ . Plus précisément, le nombre de suites qui sont envoyées sur une même partie est ce nombre de permutations, c'est-à-dire  $r!$ . On en déduit que

$$\text{card } \mathfrak{A}(\Omega, r) = r! \times \text{card } \mathfrak{P}(\Omega, r),$$

ce qui mène à la première formule puisque  $\text{card } \mathfrak{A}(\Omega, r) = n^{\downarrow r} = \frac{n!}{(n-r)!}$ . Pour la seconde formule, une preuve est donnée par l'argument suivant : pour construire une partie  $A \subset \Omega$ , on doit choisir pour chaque élément  $\omega \in \Omega$  si  $\omega$  appartient ou non à  $A$  (2 choix possibles). Il faut répéter ce choix  $n$  fois consécutivement (le nombre d'objets dans  $\Omega$ ), donc,  $\text{card } \mathfrak{P}(\Omega) = 2^n$ .  $\square$

EXEMPLE 1.24. Les coefficients binomiaux  $\binom{3}{0}$ ,  $\binom{3}{1}$ ,  $\binom{3}{2}$  et  $\binom{3}{3}$  sont respectivement 1, 3, 3 et 1. Ils correspondent bien aux cardinaux des ensembles de parties précédemment présentés lorsque  $\Omega = \llbracket 1, 3 \rrbracket$ . De plus, on a bien  $\text{card } \mathfrak{P}(\Omega) = 1 + 3 + 3 + 1 = 8 = 2^3$ .

Les coefficients binomiaux ont deux propriétés importantes qu'il est très utile de retenir. La première est la formule de récurrence de Pascal :

PROPOSITION 1.25 (Formule de Pascal). *Pour tous entiers  $n, r$ ,*

$$\binom{n}{r} + \binom{n}{r+1} = \binom{n+1}{r+1}.$$

DÉMONSTRATION. On peut démontrer cette formule en réduisant au même dénominateur les fractions mises en jeu, mais on peut aussi donner une preuve sans calcul. Si  $A$  est une partie de  $\llbracket 1, n+1 \rrbracket$  de taille  $r+1$ , alors soit  $A$  ne contient pas  $n+1$ , auquel cas c'est une partie de  $\llbracket 1, n \rrbracket$  de taille  $r+1$  ( $\binom{n}{r+1}$  choix possibles); soit  $A$  contient  $n+1$ , auquel cas  $A \setminus \{n+1\}$  est une partie de  $\llbracket 1, n \rrbracket$  de taille  $r$  ( $\binom{n}{r}$  choix possibles, et ce choix détermine entièrement  $A$  en lui rajoutant l'élément  $n+1$ ). De cette alternative, on en déduit que le nombre de parties de taille  $r+1$  dans  $\llbracket 1, n+1 \rrbracket$  est  $\binom{n+1}{r+1} = \binom{n}{r+1} + \binom{n}{r}$ .  $\square$

La formule de Pascal permet de calculer les coefficients binomiaux à l'aide du **triangle de Pascal** :

$$\begin{array}{ccccccc} & & & & & & 1 \\ & & & & & & 1 & 1 \\ & & & & & & 1 & 2 & 1 \\ & & & & & & 1 & 3 & 3 & 1 \\ & & & & & & 1 & 4 & 6 & 4 & 1 \\ & & & & & & 1 & 5 & 10 & 10 & 5 & 1 \end{array}$$

Les lignes de ce triangle étant numérotées à partir de 0, les nombres de la  $n$ -ième ligne de ce triangle sont les coefficients binomiaux  $\binom{n}{0}, \binom{n}{1}, \dots, \binom{n}{n}$ . La règle de construction du triangle est très simple : les lignes commencent et finissent par 1, et chaque nombre est la somme des deux nombres directement aussi de lui, conformément à la règle donnée par la formule de Pascal.

L'autre règle importante satisfaite par les coefficients binomiaux est celle du binôme de Newton :

PROPOSITION 1.26 (Binôme de Newton). *Pour toutes variables  $x, y$ ,*

$$(x + y)^n = \sum_{r=0}^n \binom{n}{r} x^r y^{n-r}.$$

DÉMONSTRATION. Par récurrence sur  $n$ , en utilisant la formule de Pascal pour réunir les termes de  $x(x + y)^n$  et de  $y(x + y)^n$ .  $\square$

Ainsi, les coefficients binomiaux rentrent en jeu dans le développement des puissances d'une somme. Notons que la formule du binôme implique la seconde partie du théorème 1.23 : si  $\text{card } \Omega = n$ , alors

$$\begin{aligned} \text{card}(\mathfrak{P}(\Omega)) &= \text{card} \left( \bigsqcup_{r=0}^n \mathfrak{P}(\Omega, r) \right) = \sum_{r=0}^n \text{card}(\mathfrak{P}(\Omega, r)) \\ &= \sum_{r=0}^n \binom{n}{r} = \sum_{r=0}^n \binom{n}{r} 1^r 1^{n-r} = (1 + 1)^n = 2^n. \end{aligned}$$

EXEMPLE 1.27. On peut maintenant calculer la probabilité de gagner au loto. Le nombre de parties à 6 éléments dans  $\Omega = \llbracket 1, 49 \rrbracket$  est

$$\binom{49}{6} = \frac{49 \times 48 \times 47 \times 46 \times 45 \times 44}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = 13\,983\,816.$$

Par conséquent, une grille de loto étant justement un choix d'une de ces combinaisons de 6 chiffres, la probabilité pour qu'une grille fixée corresponde au tirage d'une partie aléatoire est  $\frac{1}{13\,983\,816} = 7.15 \cdot 10^{-8}$ , si l'on suppose tous les tirages équiprobables.

## Exercices

- (1) **Warm-up.** Dans un espace de probabilité  $(\Omega, \mathbb{P})$ , à quelles conditions un événement  $A$  peut-il être indépendant de lui-même ?
- (2) **Belote.** Un jeu de belote est constitué de 32 cartes, avec dans chaque couleur les huit cartes du 7 à l'as. Calculer le nombre de mains de huit cartes qui contiennent :
- trois cartes consécutives dans la couleur cœur ;
  - cinq cartes consécutives dans l'une des quatre couleurs ;
  - trois cartes consécutives dans l'une des quatre couleurs ;
  - quatre cartes de même valeur (par exemple 4 as) ;
  - exactement 4 cartes  $\heartsuit$  et 3 valets  $V$  ?
- (3) **Poker.** Au cours d'une partie de poker, on distribue à un joueur 5 cartes d'un jeu de 52 cartes. Rappeler la formule pour le nombre de telles mains de 5 cartes. Calculer les probabilités d'avoir :
- une paire (deux cartes de même niveau) ;
  - deux paires (deux fois deux cartes de même niveau, mais pas un carré) ;
  - un brelan (trois cartes de même niveau, mais pas quatre) ;
  - une suite de 5 cartes consécutives pas toutes de la même couleur (l'as pouvant être à la fois la plus petite et la plus grande des cartes) ;
  - cinq cartes de même couleur ( $\clubsuit$ ,  $\diamondsuit$ ,  $\heartsuit$  ou  $\spadesuit$ ) ;
  - un full (trois cartes de même niveau, et deux autres cartes de même niveau) ;
  - un carré (quatre cartes de même niveau) ;
  - et une suite de 5 cartes consécutives de la même couleur ( $\clubsuit$ ,  $\diamondsuit$ ,  $\heartsuit$  ou  $\spadesuit$ ).
- (4) **Paradoxe des anniversaires.** Calculer la probabilité que  $N$  élèves d'une classe aient tous une date d'anniversaire différente (on ne prend pas en compte le problème des gens nés le 29 février). Donner une valeur approchée de cette probabilité lorsque  $N = 30$ .
- (5) **Nombres de compositions.** Quel est le nombre de vecteurs de solutions entières  $(k_1 \geq 0, \dots, k_r \geq 0)$  de l'équation  $k_1 + k_2 + \dots + k_r = n$ , avec  $n \geq 0$  ?
- (6) **Suites d'événements et probabilités.** Dans cet exercice,  $(\Omega, \mathbb{P})$  est un espace de probabilité discret (fini ou dénombrable), et  $(A_n)_{n \in \mathbb{N}}$  est une suite d'événements.
- On suppose la suite croissante, c'est-à-dire que  $A_n \subset A_{n+1}$  pour tout  $n$ . Rappeler pourquoi  $(\mathbb{P}[A_n])_{n \in \mathbb{N}}$  est une suite croissante de nombres réels. Montrer que dans ce cas,

$$\mathbb{P} \left[ \bigcup_{n=0}^{\infty} A_n \right] = \sup_{n \in \mathbb{N}} \mathbb{P}[A_n].$$

(b) De la même façon, si la suite est décroissante, montrer que

$$\mathbb{P}\left[\bigcap_{n=0}^{\infty} A_n\right] = \inf_{n \in \mathbb{N}} \mathbb{P}[A_n].$$

(c) Lemme de Borel-Cantelli : on ne fait plus d'hypothèse sur les  $A_n$ , et on considère

$$A^* = \bigcap_{n=0}^{\infty} \left( \bigcup_{N \geq n} A_N \right).$$

On suppose que  $\sum_{n=0}^{\infty} \mathbb{P}[A_n] < +\infty$  est une somme finie. Montrer que  $\mathbb{P}[A^*] = 0$  (on pourra montrer que la probabilité est plus petite que tout  $\varepsilon > 0$ ).

(7) **L'inégalité de Bonferroni.** Soit  $(A_1, \dots, A_n)$  une famille d'événements dans un espace de probabilité  $(\Omega, \mathbb{P})$ . Montrer que

$$\mathbb{P}\left[\bigcup_{i=1}^n A_i\right] \geq \sum_{i=1}^n \mathbb{P}[A_i] - \sum_{1 \leq i < j \leq n} \mathbb{P}[A_i \cap A_j].$$

Montrer qu'on a aussi

$$\mathbb{P}\left[\bigcup_{i=1}^n A_i\right] \leq \min_{k \in [1, n]} \left( \sum_{j=1}^n \mathbb{P}[A_j] - \sum_{j \neq k} \mathbb{P}[A_j \cap A_k] \right).$$

(8) **Un problème d'urnes.** On considère deux urnes  $U_1$  et  $U_2$ , la première contenant deux boules blanches et trois boules rouges, et la seconde trois boules blanches et quatre boules rouges. On tire une boule au hasard dans  $U_1$  et on la place dans  $U_2$ . On mélange ensuite  $U_2$  et on y tire une boule au hasard. Calculer la probabilité pour que cette boule soit blanche.

(9) **Un autre problème d'urnes.** On considère deux urnes  $U_{\text{blanche}}$  et  $U_{\text{rouge}}$ , la première contenant une proportion  $r$  de boules blanches et une proportion  $1 - r$  de boules rouges, et la seconde contenant une proportion  $1 - s$  de boules blanches et une proportion  $s$  de boules rouges. En commençant par l'urne blanche, on tire au temps  $n$  une boule  $b_n$  au hasard dans une urne, et la boule tirée au temps  $n + 1$  le sera dans l'urne de la couleur de  $b_n$  (on remet les boules dans les urnes après chaque tirage). Soit  $p_n$  la probabilité que  $b_n$  soit blanche.

(a) Calculer  $p_0$  et  $p_1$ .

(b) Donner une relation de récurrence entre  $p_n$  et  $p_{n+1}$ .

(c) Calculer  $p_n$  pour tout  $n$ , puis  $\lim_{n \rightarrow \infty} p_n$ .

(10) **La formule du crible.** Si  $A \subset \Omega$  est une partie d'un ensemble fini  $\Omega$ , on note  $1_A$  la fonction caractéristique de  $A$  : c'est la fonction  $\Omega \rightarrow \{0, 1\}$  qui vaut 1 sur  $A$  et 0 sur  $A^c$ .



(a) Montrer que

$$1_{A^c} = 1 - 1_A \quad ; \quad 1_{\bigcap_{i=1}^n A_i} = \prod_{i=1}^n 1_{A_i}.$$

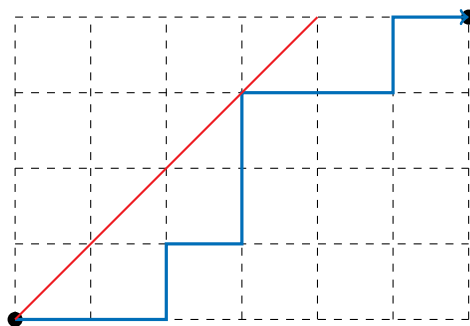
(b) Montrer aussi que pour toute partie  $A \subset \Omega$ ,  $\text{card } A = \sum_{\omega \in \Omega} 1_A(\omega)$ .

(c) En développant  $1_{\bigcup_{i=1}^n A_i} = 1 - 1_{\bigcap_{i=1}^n (A_i)^c}$ , montrer que

$$\text{card} \left( \bigcup_{i=1}^n A_i \right) = \sum_{m=1}^n (-1)^{m-1} \left( \sum_{1 \leq i_1 < i_2 < \dots < i_m \leq n} \text{card} (A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_m}) \right)$$

pour toute famille  $(A_1, A_2, \dots, A_n)$  de parties de  $\Omega$ . Expliciter cette formule lorsque  $n = 3$ .

(11) **Chemins nord-est.** On considère une grille de taille  $n \times m$ , et l'ensemble  $\Omega$  de tous les chemins de longueur  $n + m$  qui relient le coin de coordonnées  $(0, 0)$  au coin de coordonnées  $(n, m)$  avec seulement des pas vers le haut ou vers la droite :



(a) On suppose  $n = 6$  et  $m = 3$ . Calculer  $\text{card } \Omega$ .

(b) Plus généralement, calculer le cardinal  $C_{n,m}$  de  $\Omega = \Omega_{n,m}$  en fonction de  $n$  et  $m$ .

(c) On suppose  $n \geq m$ , et on note  $A$  le sous-ensemble de  $\Omega_{n,m}$  constitué des chemins qui restent sous la diagonale (ainsi, l'exemple ci-dessus de chemin est dans  $A$ ). Montrer qu'il y a autant de chemins :

- dans le complémentaire  $A^c$  de  $A$  dans l'ensemble  $\Omega_{n,m}$  des chemins nord-est dans une grille de taille  $n \times m$  ;
- et dans l'ensemble  $\Omega_{m-1, n+1}$  des chemins nord-est dans une grille de taille  $(m-1) \times (n+1)$ .

On pourra remarquer que si un chemin de  $\Omega_{n,m}$  ne reste pas en dessous de la diagonale, alors il atteint la sur-diagonale  $D$  d'équation  $y = x + 1$ . On peut alors utiliser une réflexion orthogonale par rapport à cette droite pour faire correspondre ce chemin touchant  $D$  dans une grille de taille  $n \times m$ , et un autre chemin nord-est dans une grille de taille  $(m-1, n+1)$ . En déduire une formule pour  $\text{card } A$ .

(d) Si  $n = m$ , montrer que

$$\text{card } A = \frac{1}{n+1} \binom{2n}{n}.$$

Dans ce cas, quelle est la probabilité pour qu'un chemin nord-est aléatoire choisi uniformément reste toujours sous la diagonale ?

(12) **Dérangements.** On demande à  $N$  personnes de fournir chacun sa carte d'identité. On mélange ces cartes, et on les redistribue au hasard à ces  $N$  personnes, chacune des permutations possibles étant équiprobable. On dit qu'on a un dérangement si aucune des personnes ne récupère sa propre carte. L'objectif est de calculer la probabilité  $P_N$  d'avoir un dérangement. On note  $S_N$  le nombre de permutations possibles des cartes, et  $D_N$  le nombre de dérangements.

(a) Rappeler la formule pour  $S_N$ , et donner une formule reliant  $D_N$ ,  $S_N$  et  $P_N$ .

(b) Montrer que  $D_N = (N-1)(D_{N-1} + D_{N-2})$  pour tout  $N \geq 3$ . On pourra interpréter  $D_{N-1} + D_{N-2}$  comme le nombre de dérangements tels que la  $N$ -ième personne récupère la  $i$ -ième carte,  $i \in \llbracket 1, N-1 \rrbracket$ .

(c) En déduire que  $D_N = \sum_{k=0}^N \frac{(-1)^k N!}{k!}$ . Que vaut  $P_N$ , et quelle est sa limite lorsque  $N \rightarrow \infty$  ?

(13) **Tables de couples.** On considère  $N \geq 2$  couples qui s'asseyent autour d'une table de  $2N$  places numérotées, qui est circulaire. Quel est le nombre total de façons possibles de s'asseoir pour ces  $2N$  personnes, étant entendu qu'on identifie deux façons qui différeraient seulement par une rotation de la table ? On suppose que toutes les façons ont la même probabilité. Calculer :

(a) la probabilité pour que chaque homme soit entouré de deux femmes, et chaque femme entourée de deux hommes ;

(b) conditionnellement au fait que les hommes et les femmes soient placés de façon alternée, la probabilité que chaque homme soit placé à côté de sa femme (qui peut être à sa droite ou à sa gauche) ;

(c) conditionnellement au fait que les hommes et les femmes soient placés de façon alternée, la probabilité qu'aucun homme ne soit placé à côté de sa femme.

Pour la dernière question, on doit trouver

$$\frac{1}{N!} \sum_{m=0}^N (-1)^m \frac{2N}{2N-m} \binom{2N-m}{m} (N-m)!.$$

On utilisera la formule du crible, et on pourra d'abord montrer qu'étant fixé  $m \in \llbracket 1, N \rrbracket$  couples parmi les  $N$  couples, le nombre de façons d'asseoir les  $2N$  personnes de façon alternée et telle que dans chacun des  $m$  couples l'homme soit à côté de sa femme est

$$2 \frac{m!(N-m)!^2}{2N-m} \binom{2N-m}{m}$$

(indication : il faut intercaler  $2N - 2m$  personnes entre ces couples, et on peut utiliser l'exercice sur les nombres de composition).

## Variables aléatoires discrètes

**Contenu :** variables aléatoires, espérance, variance, variables indépendantes ; exemples (lois de Bernoulli, binomiale, géométrique, de Poisson).

### 1. Notion de variable aléatoire

DÉFINITION 2.1. Une **variable aléatoire** discrète  $X$  est un nombre aléatoire dont les valeurs possibles sont dans une partie  $\mathfrak{X}$  de  $\mathbb{N}$  ou  $\mathbb{Z}$ . La **loi** de  $X$  est la probabilité  $\mathbb{P}_X$  sur  $\mathfrak{X}$  qui est associée au tirage du nombre aléatoire  $X$  :

$$\forall x \in \mathfrak{X}, \quad \mathbb{P}_X[x] = \mathbb{P}[X = x].$$

Ainsi, se donner une variable aléatoire discrète revient à se donner :

- un espace de probabilité  $(\mathfrak{X}, \mathbb{P}_X)$  avec  $\mathfrak{X} \subset \mathbb{Z}$ ,
- et une réalisation de cet espace par un nombre aléatoire  $X$  dont les probabilités sont données par  $\mathbb{P}_X$ .

REMARQUE 2.2. On a vu dans le chapitre précédent qu'une probabilité discrète était entièrement déterminée par ses valeurs sur les singletons. Par conséquent, la loi d'une variable aléatoire discrète est donnée par les valeurs  $\mathbb{P}_X[k]$ ,  $k \in \mathfrak{X}$ , avec la condition

$$\mathbb{P}[\mathfrak{X}] = \sum_{k \in \mathfrak{X}} \mathbb{P}_X[k] = 1.$$

EXEMPLE 2.3. Le lancer d'un dé équilibré à 6 faces est une variable aléatoire discrète, prenant ses valeurs dans  $\mathfrak{X} = \llbracket 1, 6 \rrbracket$ . Sa loi est la loi uniforme

$$\mathbb{P}_X[k] = \frac{1}{6} \quad \text{pour tout } k \in \llbracket 1, 6 \rrbracket.$$

EXEMPLE 2.4. Une expérience de Bernoulli est une variable aléatoire discrète, d'espace  $\mathfrak{X} = \{0, 1\}$  et de loi

$$\mathbb{P}_X[1] = p \quad ; \quad \mathbb{P}_X[0] = 1 - p$$

avec  $p \in [0, 1]$ .

De nombreuses variables aléatoires discrètes sont construites comme *fonctions* de réalisations d'espaces de probabilité discrets. Plus précisément, soit  $(\Omega, \mathbb{P})$  un espace de probabilité discret, et  $\omega$  une réalisation de cet espace, c'est-à-dire un élément aléatoire  $\omega \in \Omega$  dont les probabilités sont données par  $\mathbb{P}$ . Étant donnée une fonction

$$X : \Omega \rightarrow \mathfrak{X} \subset \mathbb{Z}$$

à valeurs entières, on peut considérer le nombre aléatoire  $X(\omega)$ , qui est une variable aléatoire discrète. Sa loi est

$$\mathbb{P}_X[k] = \mathbb{P}[X(\omega) = k] = \mathbb{P}[\{\omega \in \Omega \mid X(\omega) = k\}] = (\mathbb{P} \circ X^{-1})(k).$$

EXEMPLE 2.5. Considérons le lancer de deux dés à 6 faces, décrit par l'espace de probabilité  $\Omega = \llbracket 1, 6 \rrbracket^2$ ,  $\mathbb{P}[A] = \frac{\text{card}A}{36}$ . Si  $\omega = (a, b)$  est un lancer des deux dés, notons  $X(\omega) = a + b$  la somme : c'est une variable aléatoire discrète, à valeurs dans  $\mathfrak{X} = \llbracket 2, 12 \rrbracket$ . Sa loi est donnée par le tableau suivant :

$k$	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}_X[k]$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

En effet, pour calculer par exemple  $\mathbb{P}_X[5]$ , on regarde toutes les possibilités d'obtenir pour somme  $a + b = 5$  :

$$\mathbb{P}[X = 5] = \mathbb{P}[(a, b) \in \{(1, 4), (2, 3), (3, 2), (4, 1)\}] = \frac{4}{36}.$$

EXEMPLE 2.6. Soit  $\Omega = \mathfrak{P}(\llbracket 1, 49 \rrbracket, 6)$  l'ensemble des parties à 6 éléments de  $\llbracket 1, 49 \rrbracket$ , muni de la probabilité uniforme. On considère la fonction  $X : \Omega \rightarrow \{0, 1\}$

$$X(\omega) = \begin{cases} 1 & \text{si } \omega = \{4, 8, 15, 16, 23, 42\}, \\ 0 & \text{sinon.} \end{cases}$$

C'est une variable aléatoire, de loi

$$\mathbb{P}_X[1] = \frac{1}{\binom{49}{6}} \quad ; \quad \mathbb{P}_X[0] = 1 - \frac{1}{\binom{49}{6}}.$$

EXEMPLE 2.7. On considère le lancer  $X$  d'un seul dé à 6 faces ; on a vu que c'était une variable aléatoire discrète dans  $\mathfrak{X} = \llbracket 1, 6 \rrbracket$ . Comme dans les exemples précédents, on peut considérer  $X$  comme une fonction d'un autre espace de probabilité, à savoir,

$\Omega = \{\text{façons "physiques" de lancer le dé (inclinaison, vitesse, distance à la table, etc.)}\}$ ,

qui serait muni d'une mesure de probabilité  $\mathbb{P}$  modélisant cette expérience aléatoire. Pour un certain lancer  $\omega \in \Omega$ , on peut noter  $X(\omega) \in \llbracket 1, 6 \rrbracket$  le résultat de ce lancer, ce qui fait bien apparaître  $X = X(\omega)$  comme une fonction d'un espace de probabilité discret. Dans cet exemple, il serait très difficile de travailler directement avec  $(\Omega, \mathbb{P})$ , car c'est un ensemble très grand, compte tenu du nombre de paramètres physiques qui régissent le lancer d'un dé. En revanche, on peut tout à fait travailler avec une variable aléatoire  $X$  issue de cet espace, *e.g.* le résultat du lancer.

REMARQUE 2.8. La discussion précédente présente les variables aléatoires comme fonctions numériques d'expériences aléatoires. En particulier, si  $X$  est une variable aléatoire discrète à valeurs dans  $\mathfrak{X} \subset \mathbb{Z}$  et si  $f : \mathfrak{X} \rightarrow \mathfrak{Y} \subset \mathbb{Z}$  est une fonction, alors  $Y = f(X)$  est une nouvelle variable aléatoire discrète, à valeurs dans la partie  $\mathfrak{Y}$  de  $\mathbb{Z}$ . Sa loi est

$$\mathbb{P}_Y[l] = \mathbb{P}_X[f^{-1}(l)] = \mathbb{P}_X[\{k \in \mathfrak{X} \mid f(k) = l\}].$$

## 2. Espérance et variance

Comme les variables aléatoires sont des expériences aléatoires *numériques*, elles donnent lieu à divers calculs que l'on ne pouvait pas mener dans le cadre plus général des espaces de probabilité discrets. En particulier, on peut définir sans difficulté la *moyenne* d'une variable aléatoire discrète :

DÉFINITION 2.9. L'*espérance*, ou *moyenne* d'une variable aléatoire discrète  $X$  de loi  $\mathbb{P}_X$  est la quantité

$$\mathbb{E}[X] = \sum_{k \in \mathfrak{X} \subset \mathbb{Z}} \mathbb{P}_X[k] k.$$

Elle est bien définie si  $\mathfrak{X}$  est un ensemble fini, ou si la série ci-dessus est sommable.

EXEMPLE 2.10. L'espérance d'une variable de Bernoulli de paramètre  $p$  est

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

EXEMPLE 2.11. L'espérance du lancer d'un dé équilibré à 6 faces est

$$\mathbb{E}[X] = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = \frac{7}{2} = 3.5.$$

EXEMPLE 2.12. Comme généralisation de l'exemple précédent, considérons une variable aléatoire de loi uniforme sur un intervalle  $\llbracket a + 1, b \rrbracket$ , c'est-à-dire que

$$\mathbb{P}_X[k] = \frac{1}{b - a} \quad \text{si } k \in \llbracket a + 1, b \rrbracket.$$

On note dans ce cas  $X \sim \mathcal{U}(\llbracket a + 1, b \rrbracket)$ . Le cas précédent est celui où  $\llbracket a + 1, b \rrbracket = \llbracket 1, 6 \rrbracket$ . On a dans le cas général :

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=a+1}^b \frac{k}{b-a} = \frac{1}{b-a} \left( \sum_{k=1}^b k - \sum_{k=1}^a k \right) \\ &= \frac{1}{b-a} \left( \frac{b(b+1)}{2} - \frac{a(a+1)}{2} \right) = \frac{b^2 - a^2 + b - a}{2(b-a)} = \frac{b+a+1}{2}. \end{aligned}$$

On a ici utilisé la formule bien connue  $\sum_{k=1}^n k = \frac{n(n+1)}{2}$ , valable pour tout entier  $n \geq 0$ .

REMARQUE 2.13. La définition d'espérance fait sens lorsque  $X$  n'a pas forcément ses valeurs entières, mais plutôt dans une partie finie ou dénombrable  $\mathfrak{X}$  de l'ensemble des nombres réels  $\mathbb{R}$ . Dans ce qui suit, on utilisera sans autre précaution cette généralisation de la définition de variable aléatoire discrète, tous les résultats précédents restant vrais. D'autre part, dans tout ce qui suit, chaque fois qu'on notera  $\mathbb{E}[X]$ , on supposera implicitement (et sans le mentionner) que la somme ou la série correspondante est bien absolument convergente. La même convention s'appliquera à la variance  $\text{var}(X)$ .

REMARQUE 2.14. Soit  $X$  une variable aléatoire qui est constante, c'est-à-dire qu'il existe une valeur  $k$  telle que  $\mathbb{P}_X[k] = 1$ . Alors,

$$k = X = \mathbb{E}[X],$$

car  $\mathbb{E}[X] = k \times \mathbb{P}_X[k] = k$ .

Si  $X$  et  $Y$  sont deux variables aléatoires discrètes, on peut considérer leur somme  $X + Y$ , qui est encore une variable aléatoire discrète. L'espérance est linéaire par rapport à cette opération :

PROPOSITION 2.15. *Pour toutes variables aléatoires discrètes  $X$  et  $Y$ ,*

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

*si les deux espérances  $\mathbb{E}[X]$  et  $\mathbb{E}[Y]$  sont bien définies.*

DÉMONSTRATION. On peut en effet écrire :

$$\begin{aligned} \mathbb{E}[X + Y] &= \sum_{k \in \mathbb{Z}} k \mathbb{P}[X + Y = k] = \sum_{k \in \mathbb{Z}} k \sum_{a, b | a+b=k} \mathbb{P}[X = a \text{ et } Y = b] \\ &= \sum_{a \in \mathbb{Z}} \sum_{b \in \mathbb{Z}} (a + b) \mathbb{P}[X = a \text{ et } Y = b] \\ &= \sum_{a \in \mathbb{Z}} a \left( \sum_{b \in \mathbb{Z}} \mathbb{P}[X = a \text{ et } Y = b] \right) + \sum_{b \in \mathbb{Z}} b \left( \sum_{a \in \mathbb{Z}} \mathbb{P}[X = a \text{ et } Y = b] \right) \\ &= \sum_{a \in \mathbb{Z}} a \mathbb{P}[X = a] + \sum_{b \in \mathbb{Z}} b \mathbb{P}[Y = b] \\ &= \mathbb{E}[X] + \mathbb{E}[Y]. \end{aligned}$$

□

EXEMPLE 2.16. Considérons la somme  $S$  du lancer de deux dés à 6 faces. On a calculé sa loi précédemment, et on déduit de ce calcul que

$$\begin{aligned} \mathbb{E}[S] &= \frac{1}{36} \times 2 + \frac{2}{36} \times 3 + \frac{3}{36} \times 4 + \frac{4}{36} \times 5 + \frac{5}{36} \times 6 + \frac{6}{36} \times 7 \\ &\quad + \frac{5}{36} \times 8 + \frac{4}{36} \times 9 + \frac{3}{36} \times 10 + \frac{2}{36} \times 11 + \frac{1}{36} \times 12 \\ &= \frac{252}{36} = 7. \end{aligned}$$

On a bien comme prévu  $\mathbb{E}[S] = \mathbb{E}[X] + \mathbb{E}[Y]$ , où  $X$  et  $Y$  sont deux lancers de dés, chacun d'espérance  $\frac{7}{2}$ .

Une autre propriété générale de l'espérance est la positivité : si  $X \geq 0$ , alors  $\mathbb{E}[X] \geq 0$ , et plus généralement, si  $X \geq Y$  avec  $X$  et  $Y$  variables aléatoires, alors  $\mathbb{E}[X] \geq \mathbb{E}[Y]$ .

REMARQUE 2.17. On donnera plus tard une explication précise du fait que  $\mathbb{E}[X]$  est la bonne notion mathématique de moyenne d'une variable aléatoire  $X$ . Pour l'instant, on considère  $\mathbb{E}[X]$  comme un barycentre des valeurs possibles de la variable aléatoire  $X$ , avec des poids choisis suivant les probabilités portées par chaque valeur. Si  $X \geq 0$ , alors  $\mathbb{E}[X]$  donne une bonne idée de l'ordre de grandeur attendu pour la variable  $X$  : avec grande probabilité,  $X$  est du même ordre que  $\mathbb{E}[X]$ , voir l'exercice sur l'inégalité de Markov en fin de chapitre pour des précisions.

REMARQUE 2.18. Soit  $X$  une variable aléatoire discrète à valeurs dans  $\mathfrak{X}$ , et  $f : \mathfrak{X} \rightarrow \mathfrak{Y}$  une fonction ; la quantité  $Y = f(X)$  est une nouvelle variable aléatoire. Son espérance est donnée par la formule suivante :

$$\mathbb{E}[f(X)] = \sum_{k \in \mathfrak{X}} \mathbb{P}_X[k] f(k).$$

C'est donc la même formule que pour  $\mathbb{E}[X]$ , en remplaçant  $k$  par  $f(k)$  mais en gardant les poids  $\mathbb{P}_X[k]$  dans la somme. Pour démontrer ceci, on utilise le fait que la loi de  $Y = f(X)$  est  $\mathbb{P}_Y = \mathbb{P}_X \circ f^{-1}$  :

$$\begin{aligned} \mathbb{E}[f(X)] &= \mathbb{E}[Y] = \sum_{l \in \mathfrak{Y}} \mathbb{P}_Y[l] l = \sum_{l \in \mathfrak{Y}} \mathbb{P}_X[\{k \in \mathfrak{X} \mid f(k) = l\}] l \\ &= \sum_{l \in \mathfrak{Y}} \sum_{k \in \mathfrak{X} \mid f(k)=l} \mathbb{P}_X[k] l = \sum_{(k,l) \in \mathfrak{X} \times \mathfrak{Y} \mid f(k)=l} \mathbb{P}_X[k] f(k). \end{aligned}$$

Pour chaque  $k \in \mathfrak{X}$ , il existe un unique  $l$  tel que  $f(k) = l$ , donc la dernière somme est en fait simplement la somme sur  $k \in \mathfrak{X}$ , ce qui achève la preuve. Ce résultat est parfois appelé le **théorème de transfert**.

Intuitivement, une variable aléatoire discrète  $X$  d'espérance  $\mathbb{E}[X] = m$  est un nombre aléatoire dont les valeurs possibles sont réparties "équitablement" autour de cette valeur moyenne  $m$ . La **variance**  $\text{var}(X)$  est une seconde quantité qui permet d'estimer comment cette répartition s'effectue :

DÉFINITION 2.19. Soit  $X$  une variable aléatoire d'espérance  $m = \mathbb{E}[X]$ . La variance de  $X$  est la quantité positive

$$\text{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[(X - m)^2],$$

l'existence et la finitude de cette quantité étant par exemple garantie lorsque  $X$  prend ses valeurs dans une partie finie  $\mathfrak{X}$  de  $\mathbb{Z}$ .

La variance de  $X$  mesure la dispersion des valeurs possibles de  $X$  autour sa valeur moyenne. Considérons par exemple deux variables aléatoires  $X$  et  $Y$  de lois respectives

$$\begin{aligned}\mathbb{P}_X[0] &= \frac{1}{2} & ; & & \mathbb{P}_X[2] &= \frac{1}{2} \\ \mathbb{P}_Y[-1] &= \frac{2}{3} & ; & & \mathbb{P}_Y[5] &= \frac{1}{3}.\end{aligned}$$

Les variables  $X$  et  $Y$  ont toutes deux pour moyenne

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

En revanche, les valeurs possibles de  $Y$  ( $-1$  et  $5$ ) sont plus éloignées de cette moyenne 1 que les valeurs possibles pour  $X$  ( $0$  et  $2$ ). Ceci est mesuré par les variances, qui valent :

$$\begin{aligned}\text{var}(X) &= \mathbb{E}[(X - 1)^2] = \frac{1}{2}1^2 + \frac{1}{2}1^2 = 1 \\ \text{var}(Y) &= \mathbb{E}[(Y - 1)^2] = \frac{2}{3}2^2 + \frac{1}{3}4^2 = 8 > 1 = \text{var}(X).\end{aligned}$$

Notons que la variance est invariante par translation : si  $a$  est une constante, alors  $\text{var}(a + X) = \text{var}(X)$ . Pour le calcul des variances, on utilisera soit la définition, soit l'autre forme équivalente suivante :

PROPOSITION 2.20. On a  $\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ .

DÉMONSTRATION. Notons  $m = \mathbb{E}[X]$ , qui est une constante, donc d'espérance elle-même. On a :

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

Cette seconde formule  $\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$  pour la variance implique que pour toute variable aléatoire,  $(\mathbb{E}[X])^2 \leq \mathbb{E}[X^2]$  (inégalité de Cauchy-Schwarz).  $\square$

EXEMPLE 2.21. Calculons la variance du lancer d'un dé. On a

$$\mathbb{E}[X^2] = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) = \frac{91}{6},$$

et ainsi

$$\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}.$$

EXEMPLE 2.22. Plus généralement, on peut donner une formule pour la variance d'une variable aléatoire  $X$  de loi uniforme  $\mathcal{U}(\llbracket a + 1, b \rrbracket)$ . En utilisant la formule  $\sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}$ , on calcule

$$\begin{aligned}\mathbb{E}[X^2] &= \frac{1}{b-a} \left( \sum_{k=1}^b k^2 - \sum_{k=1}^a k^2 \right) = \frac{1}{b-a} \left( \frac{b(b+1)(2b+1)}{6} - \frac{a(a+1)(2a+1)}{6} \right) \\ &= \frac{2(b^3 - a^3) + 3(b^2 - a^2) + b - a}{6(b-a)} = \frac{2(a^2 + ab + b^2) + 3(a+b) + 1}{6}.\end{aligned}$$



En retranchant  $\mathbb{E}[X]^2 = \left(\frac{a+b+1}{2}\right)^2$ , on obtient :

$$\text{var}(X) = \frac{(b-a)^2 - 1}{12}.$$

Il n'est pas vrai en général que la variance d'une somme de variables aléatoires  $Y$  et  $Z$  soit égale à la somme des variances de  $Y$  et de  $Z$ . En effet, on peut calculer

$$\begin{aligned} \text{var}(Y + Z) &= \mathbb{E}[(Y + Z)^2] - \mathbb{E}[Y + Z]^2 \\ &= (\mathbb{E}[Y^2] + 2\mathbb{E}[YZ] + \mathbb{E}[Z^2]) - (\mathbb{E}[Y^2] + 2\mathbb{E}[Y]\mathbb{E}[Z] + \mathbb{E}[Z^2]) \\ &= (\mathbb{E}[Y^2] - \mathbb{E}[Y]^2) + (\mathbb{E}[Z^2] - \mathbb{E}[Z]^2) + 2(\mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z]) \\ &= \text{var}(Y) + \text{var}(Z) + 2(\mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z]), \end{aligned}$$

et le dernier terme n'a aucune raison *a priori* d'être nul. On appelle **covariance** de deux variables aléatoires  $Y$  et  $Z$  la quantité  $\text{cov}(Y, Z) = \mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z]$ . D'après ce qui précède, pour toutes variables aléatoires  $Y$  et  $Z$ ,

$$\text{var}(Y + Z) = \text{var}(Y) + \text{var}(Z) + 2\text{cov}(Y, Z).$$

D'autre part, le même calcul que pour la variance montre que la covariance est aussi donnée par la formule

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

Intuitivement, la covariance mesure la propension qu'ont les variables  $Y$  et  $Z$  à dévier de leur moyenne dans la même direction. En effet, supposons  $\text{cov}(Y, Z) > 0$ . Alors, le produit  $(Y - \mathbb{E}[Y])(Z - \mathbb{E}[Z])$  est de moyenne positive, donc si  $Y - \mathbb{E}[Y] > 0$ , c'est-à-dire si la déviation de  $Y$  par rapport à sa moyenne est positive, alors pour que le produit soit de moyenne positive, il faut que  $Z - \mathbb{E}[Z]$  soit également positif (pas forcément tout le temps, mais en moyenne le plus souvent). De même, si  $Y - \mathbb{E}[Y] < 0$ , alors pour que le produit  $(Y - \mathbb{E}[Y])(Z - \mathbb{E}[Z])$  reste de moyenne strictement positive, il faut que le plus souvent  $Z - \mathbb{E}[Z] < 0$ . Donc, si  $\text{cov}(Y, Z) > 0$ , alors les fluctuations des variables aléatoires  $Y$  et  $Z$  ont tendance à être de même signe. Le même raisonnement montre que si  $\text{cov}(Y, Z) < 0$ , alors les fluctuations des variables aléatoires  $Y$  et  $Z$  ont tendance à être de signes opposés.

Soient  $Y$  et  $Z$  deux variables aléatoires discrètes. D'après ce qui précède,  $\text{var}(Y + Z) = \text{var}(Y) + \text{var}(Z)$  si et seulement si  $\text{cov}(Y, Z) = 0$ . On peut alors chercher un critère simple qui garantisse que la covariance de  $Y$  et  $Z$  s'annule. Ceci mène à la notion de **variables aléatoires indépendantes**, qui est dérivée de la notion d'événements indépendants :

**DÉFINITION 2.23.** *On dit que  $Y$  et  $Z$  sont des variables indépendantes si et seulement si, pour tous entiers  $k$  et  $l$ ,*

$$\mathbb{P}[Y = k \text{ et } Z = l] = \mathbb{P}[Y = k] \mathbb{P}[Z = l].$$

*Autrement dit, la loi du couple  $(Y, Z)$  est le produit des lois de  $Y$  et de  $Z$ . Plus généralement, étant données  $n$  variables discrètes  $X_1, \dots, X_n$ , on dit qu'elles sont indépendantes si, pour tous entiers  $k_1, \dots, k_n$ ,*

$$\mathbb{P}[X_1 = k_1 \text{ et } X_2 = k_2 \text{ et } \dots \text{ et } X_n = k_n] = \mathbb{P}[X_1 = k_1] \mathbb{P}[X_2 = k_2] \dots \mathbb{P}[X_n = k_n].$$

Comme dans le cas d'événements indépendants, intuitivement, deux variables  $Y$  et  $Z$  sont indépendantes si la connaissance de la valeur de  $Y$  n'influe pas sur les probabilités des valeurs de  $Z$ , et réciproquement. On étudiera plus tard les couples de variables aléatoires, dans le cas général où elles ne sont pas forcément indépendantes. Dans ce cadre, on disposera comme dans le premier chapitre des outils fournis par les probabilités conditionnelles.

**PROPOSITION 2.24.** *Si  $Y$  et  $Z$  sont deux variables aléatoires discrètes indépendantes, alors  $\mathbb{E}[YZ] = \mathbb{E}[Y]\mathbb{E}[Z]$ , c'est-à-dire que  $\text{cov}(Y, Z) = 0$ .*

**DÉMONSTRATION.** En effet,

$$\begin{aligned} \mathbb{E}[YZ] &= \sum_{m \in \mathbb{Z}} \mathbb{P}[YZ = m] m = \sum_{m \in \mathbb{Z}} \left( \sum_{k, l \mid m=kl} \mathbb{P}[Y = k \text{ et } Z = l] \right) m \\ &= \sum_{k \in \mathbb{Z}, l \in \mathbb{Z}} \mathbb{P}[Y = k \text{ et } Z = l] kl = \sum_{k \in \mathbb{Z}, l \in \mathbb{Z}} \mathbb{P}[Y = k] k \mathbb{P}[Z = l] l \\ &= \left( \sum_{k \in \mathbb{Z}} \mathbb{P}[Y = k] k \right) \left( \sum_{l \in \mathbb{Z}} \mathbb{P}[Z = l] l \right) = \mathbb{E}[Y] \mathbb{E}[Z]. \end{aligned}$$

□

**COROLLAIRE 2.25.** *Si  $Y$  et  $Z$  sont deux variables aléatoires indépendantes, alors  $\text{var}(Y + Z) = \text{var}(Y) + \text{var}(Z)$ .*

**EXEMPLE 2.26.** La variance de la somme des lancers de deux dés indépendants est donc  $\text{var}(S) = 2 \text{var}(X) = \frac{35}{6}$ , où  $X$  est un lancer d'un dé à 6 faces.

### 3. Suites d'expériences de Bernoulli

On conclut de chapitre par l'étude de lois discrètes classiques, toutes construites à partir d'expériences de Bernoulli indépendantes.

**3.1. Loi de Bernoulli.** Une variable aléatoire suit une **loi de Bernoulli** de paramètre  $p$  si  $X$  prend pour valeurs 0 ou 1, avec

$$\mathbb{P}[X = 1] = p \quad ; \quad \mathbb{P}[X = 0] = 1 - p.$$

On note dans ce cas  $X \sim \mathcal{B}(p)$ . Une telle variable modélise une expérience à deux résultats possibles ("succès" et "échec"), et l'on compte 1 en cas de succès. Les espérance et variance de  $X$  sont

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

**3.2. Loi binomiale.** Considérons  $n$  expériences de Bernoulli  $X_1, \dots, X_n$  indépendantes et de même loi  $\mathcal{B}(p)$ . On s'intéresse au nombre  $X$  de succès obtenus, qui est la somme  $X = X_1 + X_2 + \dots + X_n$ . On a, pour  $k \in \llbracket 0, n \rrbracket$ ,

$$\begin{aligned} \mathbb{P}[X = k] &= \sum_{(x_1, \dots, x_n) \in \{0,1\}^n \mid k \text{ entrées } x_i \text{ sont égales à } 1} \mathbb{P}[X_1 = x_1, X_2 = x_2, \dots, X_n = x_n] \\ &= \sum_{(x_1, \dots, x_n) \in \{0,1\}^n \mid k \text{ entrées } x_i \text{ sont égales à } 1} p^k (1-p)^{n-k} \\ &= (\text{nombre de suites dans } \{0,1\}^n \text{ avec } k \text{ entrées égales à } 1) \times p^k (1-p)^{n-k} \\ &= \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

On dit que  $X$  suit une **loi binomiale** de paramètres  $n$  et  $p$ , ce qu'on note  $X \sim \mathcal{B}(n, p)$ . Comme  $X = X_1 + X_2 + \dots + X_n$  est somme de variables indépendantes, son espérance et sa variance sont obtenues en sommant celles calculées pour des variables de Bernoulli :

$$\mathbb{E}[X] = np \quad ; \quad \text{var}(X) = np(1-p).$$

EXEMPLE 2.27. Deux joueurs de même niveau s'affrontent lors de 10 parties de poker. Pour chaque partie, le joueur  $A$  gagne avec probabilité  $\frac{1}{2}$ , et il perd avec cette même probabilité. Le nombre de parties remportées au final suit la loi binomiale  $\mathcal{B}(10, \frac{1}{2})$ , dont les probabilités sont (avec deux chiffres significatifs) :

$k$	0	1	2	3	4	5	6	7	8	9	10
$\mathbb{P}_X[k]$	0.00098	0.0098	0.044	0.12	0.21	0.25	0.21	0.12	0.044	0.0098	0.00098

L'espérance est  $\frac{10}{2} = 5$  et la variance est  $\frac{10}{4} = \frac{5}{2}$ .

REMARQUE 2.28. La loi binomiale est associée à l'expérience aléatoire dite de tirage *avec remise*. On considère une urne avec  $B$  boules blanches et  $N$  boules noires, et on effectue  $n$  tirages dans cette urne, en remettant à chaque fois la boule tirée dans l'urne. Soit  $X$  le nombre de boules blanches obtenues au cours de ces  $n$  tirages. Comme chaque tirage est une expérience de Bernoulli de paramètre  $p = \frac{B}{B+N}$ ,  $X$  suit une loi binomiale  $\mathcal{B}(n, \frac{B}{B+N})$ . Son espérance est  $\frac{nB}{B+N}$ . On étudiera dans les exercices des lois discrètes modélisant les tirages *sans remise*.

**3.3. Loi géométrique.** Si l'on effectue une suite  $X_1, X_2, \dots, X_n, \dots$  d'expérience de Bernoulli indépendantes, on peut se demander combien de temps il faut attendre pour avoir le premier succès, c'est-à-dire, quel est le plus petit  $k \geq 1$  tel que  $X_k = 1$ . Supposons que toutes les variables  $X_i$  sont indépendantes et de même loi  $\mathcal{B}(p)$ . Si  $X$  est la loi du temps du premier succès, alors on a  $X = k$  si et seulement si  $X_1 = X_2 = \dots = X_{k-1} = 0$  et  $X_k = 1$ . Par conséquent,

$$\mathbb{P}[X = k] = \mathbb{P}[X_1 = X_2 = \dots = X_{k-1} = 0 \text{ et } X_k = 1] = (1-p)^{k-1}p.$$

On dit que  $X$  suit une **loi géométrique** de paramètre  $p$ , notée  $\mathcal{G}(p)$ . Le calcul de l'espérance et la variance de  $X$  sont liés à l'identité

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q} \text{ si } |q| < 1,$$

et aux deux premières dérivées de cette identité :

$$\sum_{k=1}^{\infty} k q^{k-1} = \frac{1}{(1-q)^2} \quad ; \quad \sum_{k=2}^{\infty} k(k-1) q^{k-2} = \frac{2}{(1-q)^3}.$$

Avec  $q = 1 - p$ , on obtient

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k q^{k-1} p = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}.$$

D'autre part,

$$\mathbb{E}[X(X-1)] = \sum_{k=2}^{\infty} k(k-1) q^{k-1} p = \frac{2qp}{(1-q)^3} = \frac{2(1-p)p}{p^3} = \frac{2}{p} \left( \frac{1}{p} - 1 \right).$$

On en déduit que  $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \frac{2-p}{p^2}$ , et que

$$\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1-p}{p^2} = \frac{1}{p} \left( \frac{1}{p} - 1 \right).$$

REMARQUE 2.29. La valeur de l'espérance est facile à retenir : si une expérience de Bernoulli a une chance sur  $k \geq 1$  d'être un succès, alors on s'attend à devoir attendre en moyenne  $k$  expériences pour avoir un succès. C'est pourquoi l'espérance d'une loi géométrique est l'inverse du paramètre des expériences de Bernoulli sous-jacentes.

**3.4. Loi de Poisson.** Finalement, il existe une dernière loi classique discrète qu'il est utile de connaître, appelée **loi de Poisson**. Pour tout paramètre  $\lambda$  réel, on rappelle l'identité

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^\lambda,$$

qu'on peut même prendre comme définition de la fonction exponentielle. Par conséquent, si  $\lambda \geq 0$ , alors la formule

$$\mathbb{P}[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}$$

est une mesure de probabilité sur  $\mathbb{N}$ , la loi de Poisson de paramètre  $\lambda$ , notée  $\mathcal{P}(\lambda)$ . Cette loi est couramment utilisée pour le nombre de réalisations d'événements aléatoires rares, tels que : le nombre de pannes d'un outil, le nombre de sinistres couverts par une assurance, le nombre de catastrophes climatiques dans un pays au cours d'une année, le nombre de poissons pêchés dans un grand lac, *etc.* On expliquera dans un instant le lien entre cette

loi et les expériences de Bernoulli. L'espérance et la variance de  $X \sim \mathcal{P}(\lambda)$  sont calculées comme suit :

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=0}^{\infty} k e^{-\lambda} \frac{(\lambda)^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} \sum_{l=0}^{\infty} \frac{\lambda^l}{l!} = \lambda; \\ \mathbb{E}[X(X-1)] &= \sum_{k=0}^{\infty} k(k-1) e^{-\lambda} \frac{(\lambda)^k}{k!} = e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} = \lambda^2 e^{-\lambda} \sum_{l=0}^{\infty} \frac{\lambda^l}{l!} = \lambda^2; \\ \text{var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.\end{aligned}$$

Voyons maintenant plus précisément dans quel contexte les lois de Poisson apparaissent. Pour fixer les idées, considérons un système mécanique, observé durant un intervalle de temps de taille 1, et qui peut subir à tout instant  $t \in [0, 1]$  une panne. On s'intéresse au nombre total  $X$  de pannes qui se produisent. Si l'on regarde le système seulement durant un intervalle de temps  $[a, a + \varepsilon] \subset [0, 1]$ , alors on peut s'attendre à ce que le nombre moyen de pannes  $\mathbb{E}[X_{[a, a + \varepsilon]}]$  observé soit proportionnel à  $\varepsilon$ , donc égal à  $\lambda \varepsilon$  pour un certain  $\lambda > 0$ . De plus, si  $\varepsilon$  est suffisamment petit, alors il est raisonnable de penser que le système ne peut pas subir plus d'une panne dans l'intervalle de temps  $[a, a + \varepsilon]$ , donc,

$$X_{[a, a + \varepsilon]} \sim \mathcal{B}(\lambda \varepsilon).$$

On modélise alors  $X$  par la somme

$$X = X_{[0, \varepsilon]} + X_{[\varepsilon, 2\varepsilon]} + X_{[2\varepsilon, 3\varepsilon]} + \cdots + X_{[(n-1)\varepsilon, n\varepsilon]},$$

où  $n = \frac{1}{\varepsilon}$ ,  $\varepsilon > 0$  est petit et chaque  $X_{[k\varepsilon, (k+1)\varepsilon]}$  suit une loi de Bernoulli  $\mathcal{B}(\lambda \varepsilon)$ . Autrement dit, on a découpé  $[0, 1]$  en petits intervalles, et on écrit le nombre total de pannes  $X$  comme le nombre d'intervalles  $[k\varepsilon, (k+1)\varepsilon]$  où une panne survient.

Supposons que lorsqu'une panne survient, elle est immédiatement réparée et n'influe pas sur le nombre de pannes ultérieures. Alors, les variables  $X_{[k\varepsilon, (k+1)\varepsilon]}$  sont indépendantes les unes des autres, et  $X$  a donc une loi binomiale

$$X \sim \mathcal{B}(n, \lambda \varepsilon) = \mathcal{B}\left(n, \frac{\lambda}{n}\right).$$

Autrement dit, pour tout entier  $k \geq 0$ ,  $\mathbb{P}[X = k] = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$ . La loi de Poisson apparaît alors comme limite lorsque  $n$  tend vers l'infini de cette identité :

**PROPOSITION 2.30.** *Dans le contexte précédent, si  $n \rightarrow \infty$  (c'est-à-dire  $\varepsilon \rightarrow 0$ ), alors pour tout entier positif  $k$ ,*

$$\lim_{\varepsilon \rightarrow 0} \mathbb{P}[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}.$$

**DÉMONSTRATION.** Fixons  $k \geq 0$ , et considérons chacun des termes de la formule  $\mathbb{P}[X = k] = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$  :

— Dans  $\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!}$ , le dénominateur est constant et le numérateur contient  $k$  termes équivalents à  $n$ , donc,  $\binom{n}{k} = \frac{n^k}{k!} (1 + o(1))$ .

— Si l'on multiplie ce qui précède par  $(\frac{\lambda}{n})^k$ , on obtient :

$$\binom{n}{k} \left(\frac{\lambda}{n}\right)^k = \frac{\lambda^k}{k!} (1 + o(1)).$$

— Finalement, regardons le logarithme du terme restant  $(1 - \frac{\lambda}{n})^{n-k}$ . Il s'agit de :

$$\log \left(1 - \frac{\lambda}{n}\right)^{n-k} = (n-k) \log \left(1 - \frac{\lambda}{n}\right) = -\frac{(n-k)\lambda}{n} + O\left(\frac{n-k}{n^2}\right) = -\lambda + o(1).$$

Donc,  $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{n-k} = e^{-\lambda}$ .

En regroupant avec le dernier terme, on obtient le résultat souhaité.  $\square$

Le même raisonnement peut être mené dès que l'on souhaite décrire un nombre aléatoire d'événements "rares" : dans des petits intervalles de temps ou d'espace, ces événements peuvent être décrits par des variables de Bernoulli de petits paramètres, et si l'on suppose que ces variables sont sans corrélations temporelles ou spatiales (indépendance des variables de Bernoulli), alors on se ramène au calcul de la limite des lois  $\mathcal{B}(n, \frac{\lambda}{n})$  lorsque  $n$  tend vers l'infini, qui est la loi de Poisson  $\mathcal{P}(\lambda)$ .

Dans le tableau qui suit, on a résumé tous les calculs sur les lois classiques discrètes :

loi	valeurs possibles	probabilités	espérance	variance
$\mathcal{U}(\llbracket a+1, b \rrbracket)$	$\llbracket a+1, b \rrbracket$	$\frac{1}{b-a}$	$\frac{a+b+1}{2}$	$\frac{(b-a)^2-1}{12}$
$\mathcal{B}(p)$	$\{0, 1\}$	$p$ ou $1-p$	$p$	$p(1-p)$
$\mathcal{B}(n, p)$	$\llbracket 0, n \rrbracket$	$\binom{n}{k} p^k (1-p)^{n-k}$	$np$	$np(1-p)$
$\mathcal{G}(p)$	$\mathbb{N}^*$	$p(1-p)^{k-1}$	$\frac{1}{p}$	$\frac{1}{p} \left(\frac{1}{p} - 1\right)$
$\mathcal{P}(\lambda)$	$\mathbb{N}$	$e^{-\lambda} \frac{\lambda^k}{k!}$	$\lambda$	$\lambda$

## Exercices

- (1) **Autour des expériences de Bernoulli.** On considère une urne contenant  $N$  boules, dont  $M$  boules blanches et  $N - M$  boules rouges. Donner dans chacun des cas suivants la loi, l'espérance et la variance de la variable aléatoire  $X$ .
- (a) On tire au hasard une boule de l'urne;  $X$  est la variable qui vaut 1 si la boule est blanche et  $-1$  si la boule est rouge.
- (b) On tire au hasard des boules de l'urne, en les remettant à chaque fois, jusqu'à ce qu'on obtienne un boule blanche;  $X$  est le nombre d'essais nécessaires.
- (c) On tire au hasard des boules de l'urne, en les remettant à chaque fois;  $X$  est le nombre de boules blanches obtenues après  $n$  essais.
- (d) On tire au hasard des boules de l'urne, en les retirant à chaque fois, jusqu'à ce qu'on obtienne un boule blanche;  $X$  est le nombre d'essais nécessaires.
- (e) On tire au hasard des boules de l'urne, en les retirant de l'urne à chaque fois;  $X$  est le nombre de boules blanches obtenues après  $n \leq \min(M, N - M)$  essais.
- (2) **Une formule pour l'espérance d'une variable positive.** Soit  $X$  une variable aléatoire discrète à valeurs positives, dans  $\mathbb{N}$ .

(a) Montrer que

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} \mathbb{P}[X > k] = \sum_{k=0}^{\infty} \mathbb{P}_X[\{k+1, k+2, \dots\}].$$

(b) Application : dans une urne contenant  $B$  boules blanches et  $N$  boules noires, on retire des boules une à une jusqu'à obtenir une boule blanche. Montrer que si  $X$  est le nombre de tirages nécessaires, alors

$$\mathbb{E}[X] = \frac{N + B + 1}{B + 1}$$

(indication : on pourra remarquer que si  $M = B + N$ , alors la somme à calculer vérifie l'équation de récurrence  $S(N, M) = 1 + \frac{N S(N-1, M-1)}{M}$ .)

- (3) **Fonctions indicatrices.** Soit  $(\Omega, \mathbb{P})$  un espace de probabilité,  $A \subset \Omega$  et  $1_A$  la fonction sur  $\Omega$  définie par :

$$1_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{sinon.} \end{cases}$$

On dit que  $1_A$  est la fonction indicatrice de  $A$ , et elle fournit une variable aléatoire à valeurs dans  $\{0, 1\}$ . Donner la loi de cette variable discrète  $1_A$ . Calculer  $\mathbb{E}[1_A]$ .

- (4) **L'inégalité de Markov.** Soit  $X$  une variable aléatoire discrète, à valeurs positives. Montrer que

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$$

pour tout  $t > 0$ . On pourra comparer  $X$  et  $X 1_{X \geq t}$ , et utiliser l'exercice précédent.

- (5) **L'inégalité de Bienaymé-Chebyshev.** Soit  $X$  une variable aléatoire discrète, telle que  $\text{var}(X)$  soit bien définie. Montrer que pour tout  $t > 0$ ,

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq t] \leq \frac{\text{var}(X)}{t^2}.$$

Commenter. Si  $X \sim \mathcal{B}(n, p)$ , en déduire que pour tout  $\varepsilon > 0$ ,

$$\mathbb{P}[|X - np| \geq n\varepsilon] \rightarrow_{n \rightarrow \infty} 0.$$

- (6) **L'inégalité de Chernov.** On considère encore une variable aléatoire  $X$  discrète.

(a) Montrer que pour tout  $t \in \mathbb{R}$ ,

$$\mathbb{P}[X \geq t] \leq e^{\inf_{\lambda \in \mathbb{R}_+} (\log \mathbb{E}[e^{\lambda X}] - \lambda t)}.$$

(b) Calculer  $\log \mathbb{E}[e^{\lambda X}]$  lorsque  $X \sim \mathcal{B}(n, p)$  (indication : écrire  $X$  comme une somme de  $n$  variables de Bernoulli indépendantes, et factoriser l'espérance  $\mathbb{E}[e^{\lambda X}]$ ).

(c) En déduire que si  $\varepsilon > 0$ , alors il existe  $0 < C(\varepsilon) < 1$  telle que

$$\mathbb{P}[X \geq n(p + \varepsilon)] \leq (C(\varepsilon))^n.$$

Comparer avec le résultat de l'exercice précédent.

- (7) **Échanges entre deux urnes.** On considère deux urnes contenant chacune  $N$  boules : l'urne  $U_1$  contient initialement  $N$  boules blanches et l'urne  $U_2$  contient initialement  $N$  boules rouges. À chaque temps  $n \in \mathbb{N}$ , on tire une boule au hasard dans  $U_1$  et une boule au hasard dans  $U_2$  et on les échange. Soit  $X_n$  le nombre de boules blanches dans l'urne  $U_1$  au temps  $n$ . Calculer par récurrence  $\mathbb{E}[X_n]$ .

- (8) **Somme aléatoire de lois de Bernoulli.** Soit  $N$  une variable aléatoire de loi de Poisson  $\mathcal{P}(\lambda)$ , et  $X_1, X_2, \dots$  une suite de variables de Bernoulli  $\mathcal{B}(p)$  indépendantes entre elles et indépendantes de la variable  $N$ . On pose

$$X = \sum_{i=1}^N X_i.$$

Calculer  $\mathbb{P}[N = n \text{ et } X = k]$  pour  $k \leq n$ . En déduire que  $X$  et  $N - X$  sont des variables indépendantes, et donner leurs lois.

- (9) **Lois puissances.** Soit  $X$  une variable aléatoire à valeurs dans  $\mathbb{N}^*$ , avec

$$\mathbb{P}[X = k] = \frac{C}{k^\alpha}$$

pour tout  $k \geq 1$ ,  $\alpha$  étant une constante strictement positive. Discuter suivant la valeur de  $\alpha$  :

- de l'existence d'une telle variable aléatoire ;
- de l'existence de l'espérance  $\mathbb{E}[X]$ .



- (10) **Lois hypergéométriques.** On considère une population de  $N$  individus, avec deux types d'individus :  $M$  individus de type 1, et  $N - M$  individus de type 2. On tire  $n \leq \min(M, N - M)$  personnes au hasard parmi les  $N$  individus, et on note  $X$  le nombre d'individus de type 1 parmi cette sous-population de  $n$  personnes. Ce nombre peut être n'importe quel entier entre 0 et  $n$ .

- (a) Montrer que la loi de  $X$  est donnée par la formule

$$\mathbb{P}[X = k] = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \quad \text{pour tout } k \in \llbracket 0, n \rrbracket.$$

On dit que c'est la loi hypergéométrique de paramètres  $N$ ,  $M$  et  $n$ .

- (b) En déduire la formule de Vandermonde :

$$\sum_{k=0}^n \binom{M}{k} \binom{N-M}{n-k} = \binom{N}{n}.$$

- (c) Calculer l'espérance et la variance de  $X$ . On utilisera la formule

$$k \binom{n}{k} = n \binom{n-1}{k-1}$$

pour tout  $n \geq k$ .

- (d) Application : on tire deux mains de 2 cartes pour deux joueurs de poker, et 3 cartes que l'on pose faces visibles sur le tapis (*flop*). Le joueur  $A$  a un coeur parmi ses 2 cartes, et il y a 2 coeurs parmi les cartes visibles. La suite du jeu (*turn* et *river*) révèle deux cartes en plus des 3 déjà visibles. Soit  $X$  le nombre de coeurs parmi ces 2 nouvelles cartes. Donner la loi de  $X$ , et calculer la probabilité pour que  $X = 2$  et ainsi que le joueur ait une couleur.
- (e) Étant donnée une variable  $X$  de loi hypergéométrique  $\mathcal{H}(N, M, n)$ , on fixe  $n$  et on suppose que  $M = pN$  et que  $N$  tend vers l'infini,  $p$  étant un paramètre entre 0 et 1. Calculer la limite de  $\mathbb{P}[X = k]$  pour tout  $k \in \llbracket 0, n \rrbracket$ .

- (11) **La marche aléatoire simple.** On considère une particule dont la position  $P_n$  à tout temps  $n$  est un entier  $k \in \mathbb{Z}$ , et qui se déplace par sauts de taille 1. À chaque instant, la particule a une probabilité  $\frac{1}{2}$  de sauter vers le haut, et une probabilité  $\frac{1}{2}$  de sauter vers le bas, tous les sauts étant indépendants. Autrement dit,

$$P_{n+1} = X_{n+1} + P_n,$$

où  $(X_n)_{n \in \mathbb{N}}$  est une suite de variables indépendantes, avec

$$\mathbb{P}[X_i = +1] = \mathbb{P}[X_i = -1] = \frac{1}{2}$$

pour tout  $i$ .

- (a) On suppose  $P_0 = 0$ . En fonction de  $n \geq 0$ , quelles sont les valeurs possibles pour  $P_n$ ? Calculer pour ces valeurs  $\mathbb{P}[P_n = k]$ .

- (b) On fixe un temps impair  $t = 2n + 1$ , et on cherche à calculer la probabilité pour que  $X_{2n+1} = 1$  et pour que tous les  $X_{i \leq 2n}$  soient négatifs ou nuls. Autrement dit, on veut calculer la probabilité de l'événement

"la marche aléatoire atteint  $k = 1$  pour la première fois au temps  $t = 2n + 1$ ".

On note  $C_n$  le nombre d'excursions négatives de taille  $2n$ , c'est-à-dire le nombre de chemins  $x_0, x_1, \dots, x_{2n}$  avec  $x_0 = x_{2n} = 0$  et tous les  $x_{i \leq 2n}$  négatifs ou nuls. Montrer que  $C_n$  vérifie l'équation de récurrence :

$$C_{n+1} = \sum_{k=0}^n C_k C_{n-k}.$$

En déduire que si  $C(z) = \sum_{k=0}^{\infty} C_k z^k$ , alors  $C(z) = z(C(z))^2 + 1$ .

- (c) Résoudre l'équation vérifiée par  $C(z)$ , et montrer que  $C_n = \frac{1}{n+1} \binom{2n}{n}$  pour tout  $n$ .
- (d) Soit  $T_1$  le temps aléatoire nécessaire pour que  $(X_n)_{n \in \mathbb{N}}$  atteigne le niveau 1. Calculer  $\mathbb{P}[T_1 = 2n + 1]$ , puis  $\mathbb{P}[T_1 < +\infty]$ .
- (12) **L'inégalité de Fortuin–Kasteleyn–Ginibre.** Soit  $f, g : \{0, 1\}^n \rightarrow \mathbb{R}$  deux fonctions croissantes, c'est-à-dire que pour tous vecteurs  $x = (x_1, \dots, x_n)$  et  $y = (y_1, \dots, y_n)$  de  $\{0, 1\}^n$  tels que  $x_1 \leq y_1, \dots, x_n \leq y_n$ , on ait

$$f(x) \leq f(y) \quad ; \quad g(x) \leq g(y).$$

On considère  $n$  variables  $X_i$  indépendantes, chacune suivant une loi de Bernoulli  $\mathcal{B}(p)$ . Montrer que les variables aléatoires  $f(X_1, \dots, X_n)$  et  $g(X_1, \dots, X_n)$  ont une covariance positive :

$$\text{cov}(f(X_1, \dots, X_n), g(X_1, \dots, X_n)) \geq 0.$$

On pourra raisonner par récurrence sur  $n$ .

- (13) **Théorème de Turan.** Soit  $G$  un graphe, c'est-à-dire une paire  $(V, E)$  avec  $V$  ensemble de sommets, et  $E$  ensemble d'arêtes  $\{v, w\}$  avec  $v \neq w$  et  $v, w \in V$ . Si  $v \in V$ , on note  $\text{deg}(v)$  le nombre de sommets connectés à  $v$ , c'est-à-dire le nombre de paires  $\{v, w\} \in E$  avec  $w \in V$ . Un ensemble de *sommets indépendants* de  $G$  est un ensemble  $I \subset V$  qui ne contient pas de sommets connectés.

Dans ce qui suit, on ordonne les sommets  $v_1, \dots, v_n$  de  $G$  aléatoirement, et on construit récursivement des ensembles  $\emptyset = I_0, I_1, \dots, I_n$  en ajoutant  $v_k$  à  $I_{k-1}$  si  $v_k$  n'est connecté à aucun sommet de  $I_{k-1}$ .

- (a) Montrer que tous les ensembles  $I_k$  sont des ensembles de sommets indépendants.
- (b) Calculer la probabilité pour que  $I_n$  contienne le sommet  $v_k$ , puis l'espérance du nombre de sommets de  $I_n$ .
- (c) En déduire que  $G$  admet au moins un ensemble de sommets indépendants avec plus de

$$\sum_{v \in V} \frac{1}{\text{deg}(v) + 1}$$

sommets.

## Variables aléatoires continues

**Contenu :** variables aléatoires continues, densité, fonction de répartition ; exemples (lois uniformes, exponentielles, gaussiennes, du  $\chi^2$ ).

### 1. Description des variables continues

Dans le chapitre précédent, on a étudié des variables aléatoires à valeurs entières dans  $\mathbb{Z}$ , ou éventuellement à valeurs dans un ensemble dénombrable de nombres  $A \subset \mathbb{R}$ . On souhaite maintenant étendre cette étude au cas de nombres aléatoires qui peuvent prendre n'importe quelle valeur dans un intervalle  $[a, b]$ , ou même dans tout  $\mathbb{R}$ . Sans rentrer dans les détails techniques, il n'est pas possible de définir des mesures de probabilité  $\mathbb{P} : \mathfrak{P}(\mathbb{R}) \rightarrow [0, 1]$ , car il y a trop de parties  $A \subset \mathbb{R}$  pour que la condition de compatibilité avec les unions disjointes puisse être satisfaite. Une façon de résoudre ce problème est de définir seulement la probabilité de *certaines* parties de  $\mathbb{R}$ , à savoir, tous les intervalles. Dans ce chapitre, on se restreindra au cas des variables **à densité** :

DÉFINITION 3.1. Soit  $I \subset \mathbb{R}$  un intervalle. Une **variable aléatoire continue**, ou à densité, à valeurs dans  $I$  est un nombre aléatoire  $X$  tel qu'existe une fonction  $f_X : I \rightarrow \mathbb{R}_+$  intégrable avec, pour tout  $[a, b] \subset I$ ,

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

On dit que  $f_X$  est la **densité** de la variable  $X$ , et elle vérifie  $\int_I f_X(x) dx = 1$ .

EXEMPLE 3.2. Si  $I = [0, 1]$ , une variable aléatoire uniforme sur  $I$  est un nombre aléatoire  $X$  avec

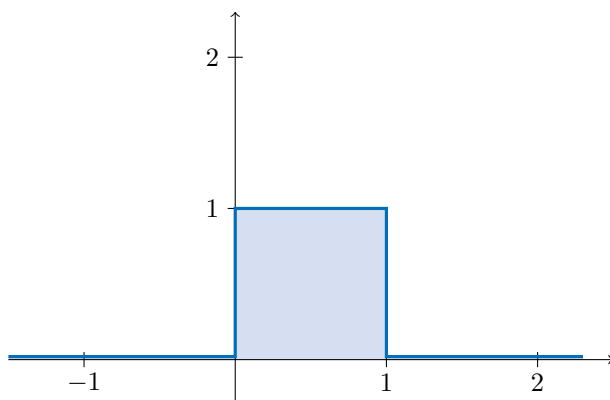
$$\mathbb{P}[X \in [a, b]] = b - a = \int_a^b 1 dx$$

pour tous  $0 \leq a \leq b \leq 1$ . La densité d'une telle variable est la fonction constante  $f(x) = 1$  pour tout  $x \in [0, 1]$ .

EXEMPLE 3.3. Plus généralement, soit  $[A, B]$  un intervalle de taille finie. Pour modéliser un nombre aléatoire à valeurs dans  $[A, B]$  et qui a les mêmes probabilités de tomber à chaque endroit de  $[A, B]$ , on peut prendre une variable de densité

$$f_X(x) = \frac{1}{B - A} \quad \forall x \in [A, B].$$

On a alors, pour tous  $A \leq a \leq b \leq B$ ,  $\mathbb{P}_X[a, b] = \mathbb{P}[X \in [a, b]] = \frac{b-a}{B-A}$ .

FIGURE 1. Densité d'une variable aléatoire uniforme sur  $[0, 1]$ .

REMARQUE 3.4. Pour ne pas surcharger le lecteur avec des détails techniques, on ne précisera pas plus ce que l'on entend par "fonction intégrable" dans la définition de la densité d'une variable aléatoire continue. On peut par exemple supposer que  $f$  est continue par morceaux sur  $I$  : ceci garantit la possibilité de calculer  $\int_a^b f(x) dx$  pour tout intervalle fini  $[a, b]$ , et par passage à la limite aussi pour les intervalles avec des bornes infinies  $a = -\infty$  ou  $b = +\infty$ .

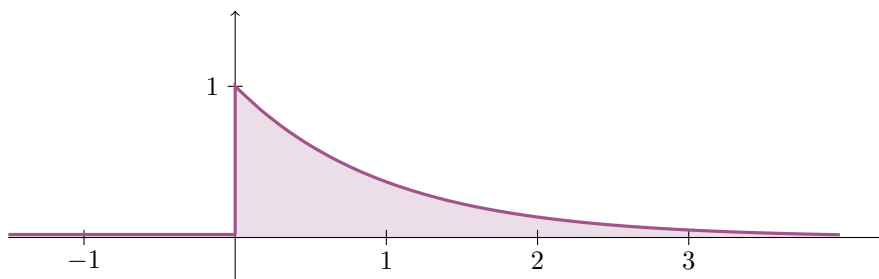
EXEMPLE 3.5. Soit  $f_X(x)$  la fonction égale à  $e^{-x}$  sur  $\mathbb{R}_+$ . C'est la densité d'une variable aléatoire continue à valeurs dans  $\mathbb{R}_+$ , puisque

$$\int_0^{\infty} e^{-x} dx = [-e^{-x}]_0^{\infty} = 1.$$

Pour tous nombres  $a < b$  positifs, si  $X$  est une variable avec la densité  $1_{x \geq 0} e^{-x}$ , alors

$$\mathbb{P}[X \in [a, b]] = \int_a^b e^{-x} dx = e^{-a} - e^{-b}.$$

On parle de variable aléatoire exponentielle, et l'on étudiera plus en détail ces variables dans la section suivante.

FIGURE 2. Densité d'une variable aléatoire exponentielle sur  $\mathbb{R}_+$ .

REMARQUE 3.6. Soit  $X$  une variable à densité à valeurs dans un intervalle  $I$ , avec une densité  $f_X$  qu'on supposera par exemple continue sur  $I$ . Si  $t \in I$  est un point, on peut calculer

$$\mathbb{P}[X = t] \leq \mathbb{P}[X \in [t - \varepsilon, t + \varepsilon]] = \int_{t-\varepsilon}^{t+\varepsilon} f_X(x) dx \underset{\varepsilon \rightarrow 0}{=} f_X(t) (2\varepsilon + o(\varepsilon)).$$

Faisant tendre  $\varepsilon$  vers 0, on constate qu'on a nécessairement  $\mathbb{P}[X = t] = 0$ . Autrement dit, tout singleton a une probabilité nulle d'être obtenu comme résultat d'une variable aléatoire continue. Attention, ça ne veut pas dire qu'on ne peut pas avoir  $X = t$ , mais juste que cet événement a probabilité 0. Plus généralement, en admettant que les probabilités continues ont la même probabilité d'additivité dénombrable que les probabilités discrètes, pour toute partie  $A \subset I$  dénombrable,  $\mathbb{P}[X \in A] = 0$ . Comme  $I$ , ou n'importe quel intervalle de taille non nulle n'est pas dénombrable, ceci n'est pas une contradiction avec la formule  $\mathbb{P}[X \in [a, b]] = \int_a^b f_X(x) dx$ .

Ainsi, si la notion de densité est la bonne notion pour avoir une probabilité "diffuse" sur tout un intervalle, elle force les singletons à avoir une probabilité nulle (autrement dit, on n'a aucune chance d'être *exactement* égal à une valeur donnée  $t$ ). En particulier, ceci implique que si  $X$  est une variable à densité, alors

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

les différences entre ces ensembles étant des singletons.

REMARQUE 3.7. La définition que l'on a donnée aux variables continues implique que l'on ne peut *a priori* mesurer que des probabilités d'intervalles. Dans ce qui suit, on s'autorisera aussi à calculer des probabilités de réunions disjointes d'intervalles, avec les règles d'additivité usuelles.

La **loi** d'une variable aléatoire continue  $X$  sur un intervalle  $I$  est la donnée de toutes les probabilités  $\mathbb{P}[X \in J]$  avec  $J$  intervalle inclus dans  $I$ . D'après ce qui précède, la densité  $f_X$  de  $X$  permet de calculer la loi par intégration. Une façon alternative de décrire la loi d'une variable continue est :

DÉFINITION 3.8. La **fonction de répartition** d'une variable aléatoire continue  $X$  est la fonction

$$F_X : t \mapsto \mathbb{P}[X \leq t] = \mathbb{P}[X \in (-\infty, t]].$$

C'est une fonction croissante, continue, avec  $\lim_{t \rightarrow -\infty} F_X(t) = 0$  et  $\lim_{t \rightarrow +\infty} F_X(t) = 1$ . Si  $X$  a une densité continue  $f_X$ , alors le lien entre la fonction de répartition  $F_X$  et la densité  $f_X$  est

$$F'_X(t) = f_X(t) \quad \text{pour tout } t \in I.$$

On a également  $F_X(t) = \int_{-\infty}^t f_X(x) dx$  pour tout  $t$ , et si  $[a, b]$  est un intervalle, alors

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

EXEMPLE 3.9. Soit  $U$  une variable aléatoire uniforme sur  $[0, 1]$ , donc de densité  $f_U(u) = 1$  si  $u \in [0, 1]$ , et 0 sinon. La fonction de répartition de  $U$  est

$$F_U(t) = \begin{cases} 0 & \text{si } t < 0, \\ t & \text{si } 0 \leq t \leq 1, \\ 1 & \text{si } t > 1. \end{cases}$$

En effet, si l'on veut par exemple calculer  $F_U(t)$  pour  $t \in [0, 1]$ , on a :

$$F_U(t) = \int_{-\infty}^t f_U(u) du = \int_{-\infty}^0 0 du + \int_0^t 1 du = t.$$

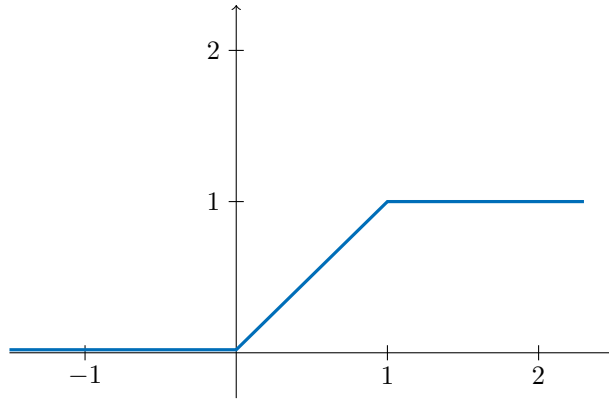


FIGURE 3. Fonction de répartition d'une variable aléatoire uniforme sur  $[0, 1]$ .

EXEMPLE 3.10. Soit  $X$  la variable aléatoire continue de fonction de répartition  $F_X(t) = \frac{\arctan t}{\pi} + \frac{1}{2}$ . C'est bien une fonction continue croissante qui a pour limites respectivement 0 et 1 en  $-\infty$  et  $+\infty$ . La densité de cette variable aléatoire est  $\frac{1}{\pi}$  fois la dérivée de la fonction arctan, c'est-à-dire,

$$f_X(x) = \frac{1}{\pi(1+x^2)}.$$

Dans le chapitre précédent, on a défini l'espérance d'une variable aléatoire discrète par la formule  $\mathbb{E}[X] = \sum_{k \in \mathbb{Z}} \mathbb{P}[X = k] k$ . Par analogie, si  $X$  est une variable aléatoire à densité, on définit son espérance par la formule

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

Cette formule fait sens si la densité de  $X$  est à support compact (dans un intervalle  $I = [A, B]$  de taille finie), ou si l'intégrale est absolument convergente. Plus généralement, si  $h : \mathbb{R} \rightarrow \mathbb{R}$  est une fonction, on définit

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

sous réserve que cette intégrale soit finie. Comme dans le cas discret, on supposera dans tout ce qui suit que les espérances manipulées sont bien associées à des intégrales *absolument* convergentes, et on ne le précisera plus forcément.

REMARQUE 3.11. On convient que  $\mathbb{E}[X]$  (ou  $\mathbb{E}[h(X)]$ ) n'est pas correctement définie si elle est donnée par une intégrale simplement mais pas absolument convergente. Autrement dit,  $\mathbb{E}[h(X)]$  est correctement définie si et seulement si  $\int_{\mathbb{R}} f_X(x) |h(x)| dx < +\infty$ .

EXEMPLE 3.12. Soit  $U$  une variable aléatoire uniforme sur  $[0, 1]$ , de densité  $f_U(u) = 1_{u \in [0,1]}$ . Son espérance est

$$\mathbb{E}[U] = \int_0^1 u \, du = \frac{1}{2}.$$

On obtient une moyenne de  $\frac{1}{2}$ , ce qui est bien la valeur attendue pour un nombre aléatoire uniformément distribué entre 0 et 1.

EXEMPLE 3.13. Soit  $X$  une variable aléatoire de densité  $f_X(x) = 1_{x \geq 0} e^{-x}$ . Son espérance est

$$\mathbb{E}[X] = \int_0^{\infty} e^{-x} x \, dx = [-e^{-x} x]_{x=0}^{\infty} + \int_0^{\infty} e^{-x} \, dx = 1.$$

Plus généralement, calculons les **moments** de  $X$ , c'est-à-dire les espérances  $\mathbb{E}[X^n]$  avec  $n \in \mathbb{N}^*$ . Une intégration par parties montre que :

$$\mathbb{E}[X^n] = \int_0^{\infty} e^{-x} x^n \, dx = [-e^{-x} x^n]_{x=0}^{\infty} + n \int_0^{\infty} e^{-x} x^{n-1} \, dx = n \mathbb{E}[X^{n-1}].$$

Par récurrence, ceci implique  $\mathbb{E}[X^n] = n!$  pour tout entier  $n$ .

Comme dans le cas discret, l'espérance est linéaire et positive :

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

si  $X, Y$  sont deux variables continues et  $\lambda, \mu$  sont deux scalaires, et d'autre part,  $\mathbb{E}[X] \geq 0$  si  $X \geq 0$ , c'est-à-dire si la densité  $f_X$  est nulle sur  $\mathbb{R}_-$ . Finalement, on définit comme dans le cas discret la variance d'une variable continue par la formule

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

Cette quantité, lorsqu'elle est bien définie, est toujours positive ou nulle, et de plus, si  $a, b$  sont des scalaires, alors  $\text{var}(aX + b) = a^2 \text{var}(X)$ .

EXEMPLE 3.14. Si  $U$  est une variable uniforme sur  $[0, 1]$ , alors on a calculé  $\mathbb{E}[U] = \frac{1}{2}$ , et d'autre part,

$$\mathbb{E}[U^2] = \int_0^1 u^2 \, du = \left[ \frac{u^3}{3} \right]_0^1 = \frac{1}{3},$$

donc  $\text{var}(U) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$ .

Pour conclure notre présentation des variables aléatoires continues, expliquons comment la notion d'indépendance s'adapte à ce nouveau cadre. Si  $X$  et  $Y$  sont deux variables aléatoires continues, on dira qu'elles sont indépendantes si, pour tous intervalles  $[a, b]$  et  $[c, d]$ ,

$$\mathbb{P}[a \leq X \leq b \text{ et } c \leq Y \leq d] = \mathbb{P}[a \leq X \leq b] \times \mathbb{P}[c \leq Y \leq d].$$

Ceci est équivalent à demander que pour tous nombres réels  $s$  et  $t$ ,

$$\mathbb{P}[X \leq s \text{ et } Y \leq t] = F_X(s) F_Y(t).$$

On donnera dans le chapitre suivant une définition équivalente avec les densités, qui sera bien plus commode à manipuler. Si  $X$  et  $Y$  sont indépendantes et si l'on suppose que les espérances  $\mathbb{E}[X]$  et  $\mathbb{E}[Y]$  sont bien définies, alors on a aussi  $\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y]$ . En effet, supposons pour commencer  $X$  et  $Y$  à valeurs positives. Alors,

$$\mathbb{E}[X] = \int_0^\infty t f_X(t) dt = \int_0^\infty F_X(t) dt$$

par intégration par parties. De plus,

$$F_{XY}(t) = \mathbb{P}[XY \leq t] = \sum_{k \in \mathbb{N}} \mathbb{P}\left[k\varepsilon \leq X \leq (k+1)\varepsilon \text{ et } Y \leq \frac{t}{X}\right].$$

Estimons chaque terme  $p_k = \mathbb{P}[k\varepsilon \leq X \leq (k+1)\varepsilon \text{ et } Y \leq \frac{t}{X}]$ . Cette probabilité est encadrée par

$$\mathbb{P}\left[k\varepsilon \leq X \leq (k+1)\varepsilon \text{ et } Y \leq \frac{t}{(k+1)\varepsilon}\right] \leq p_k \leq \mathbb{P}\left[k\varepsilon \leq X \leq (k+1)\varepsilon \text{ et } Y \leq \frac{t}{k\varepsilon}\right],$$

donc, en utilisant l'indépendance de  $X$  et  $Y$ , on obtient :

$$\begin{aligned} \sum_{k \in \mathbb{N}} (F_X((k+1)\varepsilon) - F_X(k\varepsilon)) F_Y\left(\frac{t}{(k+1)\varepsilon}\right) &\leq F_{XY}(t) \\ &\leq \sum_{k \in \mathbb{N}} (F_X((k+1)\varepsilon) - F_X(k\varepsilon)) F_Y\left(\frac{t}{k\varepsilon}\right). \end{aligned}$$

Si  $\varepsilon$  est suffisamment petit, alors on peut approximer chaque terme par  $f_X(k\varepsilon) F_X\left(\frac{t}{k\varepsilon}\right) \varepsilon$ . On reconnaît alors des sommes de Riemann, et en négligeant les problèmes d'interversion de somme et de limite, on obtient :

$$F_{XY}(t) = \int_{s=0}^\infty f_X(s) F_Y\left(\frac{t}{s}\right) ds.$$

Une autre preuve plus simple de cette identité sera donnée dans le chapitre suivant. On peut maintenant écrire, en faisant le changement de variables  $u = \frac{t}{s}$  :

$$\begin{aligned} \mathbb{E}[XY] &= \int_{t=0}^\infty F_{XY}(t) dt = \int_{s=0}^\infty \int_{t=0}^\infty f_X(s) F_Y\left(\frac{t}{s}\right) ds dt = \int_{s=0}^\infty \int_{u=0}^\infty f_X(s) s F_Y(u) ds du \\ &= \mathbb{E}[Y] \int_{s=0}^\infty f_X(s) s ds = \mathbb{E}[X] \mathbb{E}[Y]. \end{aligned}$$



Ensuite, pour traiter le cas de variables aléatoires à signe quelconque, on peut écrire toute variable continue  $X$  comme la différence de deux variables positives :

$$X = X 1_{X \geq 0} - (-X 1_{X < 0}).$$

Un raisonnement analogue permet d'établir le résultat suivant, qu'on admettra :

**THÉORÈME 3.15.** *Soient  $X$  et  $Y$  deux variables aléatoires continues indépendantes. Si  $f$  et  $g$  sont deux fonctions telles que les espérances  $\mathbb{E}[f(X)]$  et  $\mathbb{E}[g(Y)]$  soient bien définies, alors la variable aléatoire  $f(X)g(Y)$  admet pour espérance*

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

**COROLLAIRE 3.16.** *Si  $X$  et  $Y$  sont deux variables aléatoires continues indépendantes, alors  $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$ .*

**DÉMONSTRATION.** Comme dans le cas discret, on peut introduire pour deux variables continues  $X$  et  $Y$  leur covariance  $\text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y]$ , et on a le développement

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2 \text{cov}(X, Y).$$

Or, d'après la discussion précédente, si  $X$  et  $Y$  sont deux variables continues indépendantes, alors  $\text{cov}(X, Y) = 0$ .  $\square$

## 2. Lois uniformes et lois exponentielles

Après une présentation générale des variables aléatoires à densité, étudions en détail deux exemples importants : les variables aléatoires uniformes et les variables aléatoires exponentielles.

**2.1. Variables uniformes.** Si  $[a, b]$  est un intervalle de taille finie  $b - a$ , on dit qu'une variable aléatoire continue a la **loi uniforme** sur  $[a, b]$  si sa densité est la fonction qui vaut  $\frac{1}{b-a}$  sur le segment  $[a, b]$ , et 0 ailleurs :

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

On note dans ce cas  $X \sim \mathcal{U}([a, b])$ . Cette notion est l'analogue continu des lois discrètes  $\mathcal{U}([a + 1, b])$ . Dans la section précédente, on a étudié en exemple le cas où  $X \sim \mathcal{U}([0, 1])$ . Dans le cas plus général où  $X \sim \mathcal{U}([a, b])$ , la fonction de répartition est

$$F_X(t) = \begin{cases} 0 & \text{si } t \leq a, \\ \frac{t-a}{b-a} & \text{si } a \leq t \leq b, \\ 1 & \text{si } t \geq b. \end{cases}$$

Sa moyenne est

$$\mathbb{E}[X] = \frac{1}{b-a} \int_a^b x dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2},$$

et sa variance est

$$\begin{aligned}\operatorname{var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} \\ &= \frac{a^2 - 2ab + b^2}{12} = \frac{(b-a)^2}{12}.\end{aligned}$$

Notons que si  $X$  est une variable uniforme sur un segment, alors toute fonction affine  $\alpha X + \beta$  est encore une variable uniforme sur un (autre) segment. En effet, en supposant par exemple  $\alpha > 0$ , et en remarquant qu'alors  $\mathbb{P}[X \leq t] = \mathbb{P}[\alpha X + \beta \leq \alpha t + \beta]$ , on montre l'identité des fonctions de répartition de  $\alpha X + \beta$  et de  $Y$  si  $X \sim \mathcal{U}([a, b])$  et  $Y \sim \mathcal{U}([\alpha a + \beta, \alpha b + \beta])$ .

**2.2. Variables exponentielles.** Si  $\lambda > 0$  est un paramètre positif, on dit qu'une variable continue  $X$  a la **loi exponentielle** de paramètre  $\lambda$  si sa densité est la fonction qui vaut  $\lambda e^{-\lambda x}$  sur  $\mathbb{R}_+$ , et 0 sur  $\mathbb{R}_-^*$  :

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

On note dans ce cas  $X \sim \mathcal{E}(\lambda)$ , et la fonction de répartition de  $X$  est

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

En particulier, si  $t \geq 0$ , alors  $\mathbb{P}[X \geq t] = 1 - F_X(t) = e^{-\lambda t}$ . L'espérance de  $X$  est

$$\mathbb{E}[X] = \int_0^\infty \lambda e^{-\lambda x} x dx = \frac{1}{\lambda} \int_0^\infty e^{-u} u du = \frac{1}{\lambda}$$

en faisant d'abord le changement de variables  $u = \lambda x$ , puis en utilisant le calcul de la section précédente pour la loi  $\mathcal{E}(1)$ . De même, on calcule

$$\mathbb{E}[X^2] = \int_0^\infty \lambda e^{-\lambda x} x^2 dx = \frac{1}{\lambda^2} \int_0^\infty e^{-u} u^2 du = \frac{2}{\lambda^2},$$

donc la variance de  $X$  est  $\operatorname{var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$ . Les calculs précédents montrent que si  $X \sim \mathcal{E}(\lambda)$  suit une loi exponentielle, alors pour tout  $\mu > 0$ ,  $\mu X$  suit également une loi exponentielle, à savoir la loi  $\mathcal{E}(\frac{\lambda}{\mu})$ . En effet,

$$\mathbb{P}[\mu X \leq t] = \mathbb{P}\left[X \leq \frac{t}{\mu}\right] = 1 - e^{-\frac{\lambda}{\mu}t}$$

est bien la fonction de répartition de  $\mathcal{E}(\frac{\lambda}{\mu})$ .

**EXEMPLE 3.17.** Soit  $T$  le temps que l'on doit attendre avant la première panne d'une voiture. On peut imaginer que, si à un instant  $t$  la voiture n'a pas encore eu de panne, alors elle est "comme neuve" à cet instant, de sorte que pour tous  $s, t \geq 0$ ,

$$\frac{\mathbb{P}[T \geq t + s]}{\mathbb{P}[T \geq t]} = \mathbb{P}[T \geq t + s \mid T \geq t] = \mathbb{P}[T \geq s],$$

puisque conditionnellement au fait que  $T \geq t$ , on regarde à l'instant  $t$  une voiture qui a les mêmes temps de panne qu'une voiture neuve. Soit  $F_T(t)$  la fonction de répartition de  $T$ , et  $G_T(t) = 1 - F_T(t) = \mathbb{P}[T \geq t]$ . On vient de voir que

$$G_T(s+t) = G_T(s)G_T(t)$$

pour tous  $s, t \geq 0$ . Fixons  $t$  et dérivons cette équation par rapport à  $s$  :

$$G'_T(s+t) = G'_T(s)G_T(t) \quad ; \quad G'_T(t) = G'_T(0)G_T(t)$$

la seconde partie étant obtenue en prenant  $s = 0$ . Si  $-\lambda = G'_T(0)$ , alors l'unique solution de cette équation différentielle avec  $G_T(0) = \mathbb{P}[T \geq 0] = 1$  est  $G_T(t) = e^{-\lambda t}$ , d'où  $F_T(t) = 1 - e^{-\lambda t}$ . Par conséquent,  $T$  suit une loi exponentielle avec un certain paramètre  $\lambda > 0$ , qui est relié au temps moyen de panne par l'équation  $\mathbb{E}[T] = \frac{1}{\lambda}$ .

REMARQUE 3.18. L'exemple précédent montre que le *temps* aléatoire nécessaire pour observer une panne d'un système mécanique est correctement modélisé par une variable continue de loi exponentielle. D'autre part, on a expliqué dans le chapitre précédent que le *nombre* de pannes observées durant un intervalle de temps fixé est correctement modélisé par une loi de Poisson. Le lien entre ces deux lois (respectivement continue et discrète) est étudié dans les exercices du chapitre.

REMARQUE 3.19. Les lois exponentielles sont les analogues continues des lois géométriques. En particulier, on peut montrer que si  $X \sim \mathcal{E}(\lambda)$  suit une loi exponentielle, alors la partie entière  $\lfloor X \rfloor$  suit une loi géométrique, voir les exercices.

### 3. Lois gaussiennes et lois du $\chi^2$

On a étudié dans la section précédentes deux classes de lois continues classiques, qui sont les analogues à densité des lois uniformes discrètes et des lois géométriques. Dans la même direction, on peut se demander s'il existe un analogue continu des lois binomiales, qui sont des exemples importants de lois discrètes. Un tel analogue va apparaître par renormalisation de lois  $\mathcal{B}(n, p)$ , lorsque le paramètre  $n$  tend vers l'infini. Pour simplifier, on étudiera le cas particulier où  $p = \frac{1}{2}$ , le cas général étant traité par le théorème central limite, voir le chapitre suivant.

On commence par démontrer la célèbre formule asymptotique suivante :

PROPOSITION 3.20 (Formule de Stirling). *Lorsque  $n$  tend vers l'infini,*

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n} (1 + o(1)).$$

DÉMONSTRATION. On a calculé dans la première section  $\mathbb{E}[X] = \int_0^\infty e^{-x} x^n dx = n!$ , où  $X \sim \mathcal{E}(1)$ . Ceci fournit une expression intégrale pour  $n!$ , qui est adaptée à une étude asymptotique. Soit  $f_n(x) = \log x - \frac{x}{n}$ ; cette fonction admet son maximum en  $x = n$ , égal

à  $(\log n) - 1$ . On décompose alors l'intégrale en plusieurs morceaux :  $n! = A + B + C$ , avec

$$A = \int_0^{n(1-\varepsilon)} e^{-x} x^n dx \quad ; \quad B = \int_{n(1-\varepsilon)}^{n(1+\varepsilon)} e^{-x} x^n dx \quad ; \quad C = \int_{n(1+\varepsilon)}^{\infty} e^{-x} x^n dx.$$

Concentrons-nous d'abord sur

$$B = \int_{n(1-\varepsilon)}^{n(1+\varepsilon)} e^{nf_n(x)} dx = n^{n+1} e^{-n} \int_{-\varepsilon}^{\varepsilon} e^{n(\log(1+u)-u)} du.$$

Si  $0 \leq |u| \leq \varepsilon \leq \frac{1}{2}$ , alors le développement de Taylor-Lagrange de  $\log(1+u) - u$  montre que

$$\log(1+u) - u + \frac{u^2}{2} = \frac{u^3}{6} \frac{2}{(1+v)^3}$$

avec  $|v| \leq |u| \leq \varepsilon$ , donc

$$\log(1+u) - u = -\frac{u^2}{2}(1 + C(u)\varepsilon)$$

avec  $|C(u)|$  borné par exemple par 6. Ainsi,

$$\int_{-\varepsilon}^{\varepsilon} e^{-\frac{n(1+6\varepsilon)u^2}{2}} du \leq \frac{B}{n^{n+1} e^{-n}} \leq \int_{-\varepsilon}^{\varepsilon} e^{-\frac{n(1-6\varepsilon)u^2}{2}} du$$

si  $\varepsilon \leq \frac{1}{2}$ . Par changement de variables, on calcule l'asymptotique du terme de gauche :

$$G = \int_{-\varepsilon}^{\varepsilon} e^{-\frac{n(1+6\varepsilon)u^2}{2}} du = \frac{1}{\sqrt{n(1+6\varepsilon)}} \int_{-\varepsilon\sqrt{n(1+6\varepsilon)}}^{\varepsilon\sqrt{n(1+6\varepsilon)}} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{n}} \left( \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt \right) (1 + o(1))$$

si  $\varepsilon = o(1)$  et  $\varepsilon \gg \frac{1}{\sqrt{n}}$  (la seconde condition garantit que les bornes de l'intégrale tendent vers  $\pm\infty$ ). On obtient sous les mêmes hypothèses la même asymptotique pour le terme de droite, donc,

$$B = \left(\frac{n}{e}\right)^n \sqrt{n} \left( \int_{\mathbb{R}} e^{-\frac{t^2}{2}} dt \right) (1 + o(1)).$$

Pour calculer l'intégrale  $I = \int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx$ , on utilise un changement de variables polaire en dimension 2 :

$$I^2 = \int_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dx dy = \int_{r=0}^{\infty} \int_{\theta=0}^{2\pi} e^{-\frac{r^2}{2}} r dr d\theta = 2\pi \int_{r=0}^{\infty} r e^{-\frac{r^2}{2}} dr = 2\pi.$$

Par conséquent,  $B = \left(\frac{n}{e}\right)^n \sqrt{2\pi n} (1 + o(1))$ . Il suffit maintenant de montrer que les deux autres intégrales  $A$  et  $C$  sont négligeables devant cette quantité. Comme  $f_n(x)$  est croissante entre 0 et  $n$ ,

$$A \leq \int_0^{n(1-\varepsilon)} e^{nf_n(n(1-\varepsilon))} dx \leq n \left(\frac{n}{e}\right)^n e^{n\varepsilon + n \log(1-\varepsilon)}.$$

Prenons  $\varepsilon = \frac{1}{n^{1/4}}$ , qui est bien un  $o(1)$  et beaucoup plus grand que  $\frac{1}{\sqrt{n}}$ . Alors,

$$n\varepsilon + n \log(1-\varepsilon) \leq -\frac{n\varepsilon^2}{2} = -\frac{\sqrt{n}}{2},$$

donc  $A$  est bien négligeable devant  $B$  (et même exponentiellement plus petite). De même, pour l'intégrale  $C$ , on utilise le fait que

$$f'_n(x) = \frac{1}{x} - \frac{1}{n} \leq \frac{1}{n} \left( \frac{1}{1+\varepsilon} - 1 \right);$$

$$f_n(x) \leq f_n(n(1+\varepsilon)) - \frac{x - n(1+\varepsilon)}{n} \left( \frac{\varepsilon}{1+\varepsilon} \right)$$

si  $x \geq n(1+\varepsilon)$ . Ceci implique :

$$C \leq e^{nf_n(n(1+\varepsilon))} \int_{x=n(1+\varepsilon)}^{\infty} e^{-(x-n(1+\varepsilon))\frac{\varepsilon}{1+\varepsilon}} dx = \frac{1+\varepsilon}{\varepsilon} e^{n(\log(1+\varepsilon)-\varepsilon)} \left(\frac{n}{e}\right)^n \leq 2n^{1/4} \left(\frac{n}{e}\right)^n$$

puisque  $\log(1+\varepsilon)-\varepsilon \leq 0$  pour tout  $\varepsilon > 0$ . On obtient de nouveau une quantité négligeable devant  $B$ , ce qui achève la preuve.  $\square$

Considérons maintenant une variable discrète  $G_n = \frac{2X_n - n}{\sqrt{n}}$ , où  $X_n \sim \mathcal{B}(n, \frac{1}{2})$ . C'est une variable aléatoire dont les valeurs possibles sont comprises entre  $-\sqrt{n}$  et  $\sqrt{n}$ , et qui a pour espérance 0 et pour variance  $\frac{4}{n} \text{var}(X_n) = 1$ .

THÉORÈME 3.21. *Pour tous  $a$  et  $b$  nombres réels,*

$$\lim_{n \rightarrow \infty} \mathbb{P}[a \leq G_n \leq b] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

DÉMONSTRATION. Soit  $x$  un nombre réel, tel que  $\frac{n}{2} + \frac{x\sqrt{n}}{2} = k$  soit un entier compris entre 0 et  $n$ . On utilise la formule de Stirling pour évaluer

$$\begin{aligned} \binom{n}{k} &= \frac{2}{\sqrt{2\pi}} \sqrt{\frac{n}{n^2 - x^2}} \frac{n^n}{\left(\frac{n}{2} + \frac{x\sqrt{n}}{2}\right)^{\frac{n}{2} + \frac{x\sqrt{n}}{2}} \left(\frac{n}{2} - \frac{x\sqrt{n}}{2}\right)^{\frac{n}{2} - \frac{x\sqrt{n}}{2}}} (1 + o(1)) \\ &= \frac{2^{n+1}}{\sqrt{2\pi(n-x^2)}} e^{-\frac{1}{2}((n+x\sqrt{n}) \log(1+\frac{x}{\sqrt{n}}) + (n-x\sqrt{n}) \log(1-\frac{x}{\sqrt{n}}))} (1 + o(1)), \end{aligned}$$

en supposant que  $x = o(\sqrt{n})$ . On peut alors faire un développement limité du terme  $f(n, x)$  dans l'exponentielle :

$$\begin{aligned} f(n, x) &= (n + x\sqrt{n}) \log\left(1 + \frac{x}{\sqrt{n}}\right) + (n - x\sqrt{n}) \log\left(1 - \frac{x}{\sqrt{n}}\right) \\ &= (n + x\sqrt{n}) \left(\frac{x}{\sqrt{n}} - \frac{x^2}{2n}\right) + (n - x\sqrt{n}) \left(-\frac{x}{\sqrt{n}} - \frac{x^2}{2n}\right) + o(1) = x^2 + o(1). \end{aligned}$$

On en déduit que si  $x = o(\sqrt{n})$ , alors

$$\frac{1}{2^n} \binom{n}{k} = \frac{2}{\sqrt{2\pi n}} e^{-\frac{x^2}{2}} (1 + o(1)).$$

Fixons maintenant  $a \leq b$  dans  $\mathbb{R}$ . On a :

$$\begin{aligned} \mathbb{P}[a \leq G_n \leq b] &= \mathbb{P}\left[\frac{n}{2} + \frac{a\sqrt{n}}{2} \leq X_n \leq \frac{n}{2} + \frac{b\sqrt{n}}{2}\right] \\ &= \sum_{k=\lceil \frac{n}{2} + \frac{a\sqrt{n}}{2} \rceil}^{\lfloor \frac{n}{2} + \frac{b\sqrt{n}}{2} \rfloor} \frac{1}{2^n} \binom{n}{k} = \sum_{\substack{x=a_+ + \frac{2m}{\sqrt{n}} \\ m \in \mathbb{N}, x \leq b_-}} \frac{2}{\sqrt{2\pi n}} e^{-\frac{x^2}{2}} (1 + o(1)) \end{aligned}$$

où  $a_+ \geq a$  et  $b_- \leq b$  sont définis par les identités :

$$\left\lceil \frac{n}{2} + \frac{a\sqrt{n}}{2} \right\rceil = \frac{n}{2} + \frac{a_+\sqrt{n}}{2} \quad ; \quad \left\lfloor \frac{n}{2} + \frac{b\sqrt{n}}{2} \right\rfloor = \frac{n}{2} + \frac{b_-\sqrt{n}}{2}.$$

Notons que  $a_+ = a + O(\frac{1}{\sqrt{n}})$  et  $b_- = b + O(\frac{1}{\sqrt{n}})$ . On reconnaît alors une somme de Riemann de pas  $\frac{2}{\sqrt{n}}$ , donc,

$$\lim_{n \rightarrow \infty} \mathbb{P}[a \leq G_n \leq b] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

□

REMARQUE 3.22. On peut illustrer le théorème en dressant l'histogramme des probabilités  $\mathbb{P}[X_n = k]$ , comparé au graphe de la fonction  $x \mapsto (2\pi)^{-1/2} e^{-x^2/2}$  correctement renormalisé, voir la figure suivante.

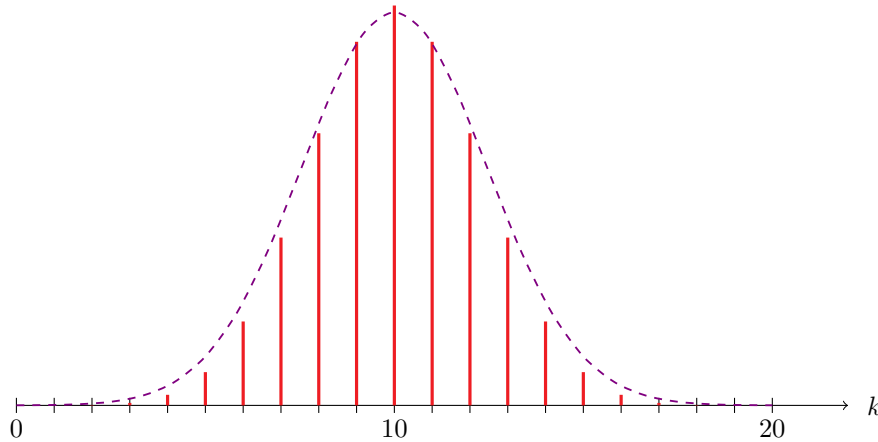


FIGURE 4. Histogramme de la loi binomiale  $\mathcal{B}(n, \frac{1}{2})$  avec  $n = 20$ .

La limite qui apparaît dans le théorème précédent met en jeu la densité  $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$  d'une variable aléatoire continue, appelée variable normale ou gaussienne. Plus généralement, si  $m \in \mathbb{R}$  et  $\sigma^2 \in \mathbb{R}_+^*$ , la **loi normale** ou **loi gaussienne** de moyenne  $m$  et de variance  $\sigma^2$  est la densité de probabilité

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

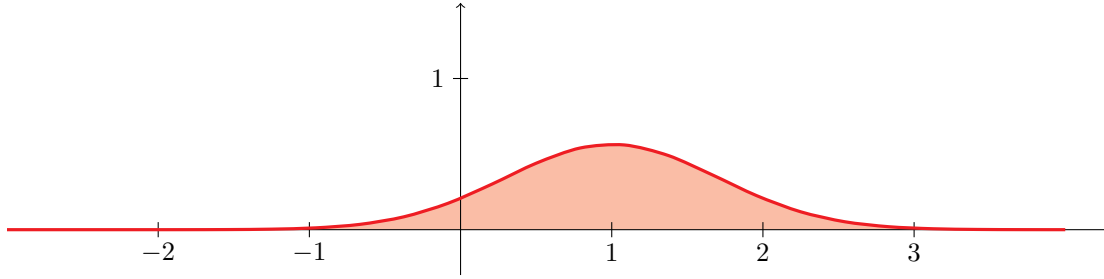


FIGURE 5. Densité d'une variable aléatoire gaussienne, ici avec  $m = 1$  et  $\sigma^2 = \frac{1}{2}$ .

On note dans ce cas  $X \sim \mathcal{N}(m, \sigma^2)$ , et le cas où  $m = 0$  et  $\sigma^2 = 1$  (variable gaussienne *standard*) est celui du théorème précédent. Si  $X \sim \mathcal{N}(0, 1)$ , alors on a bien

$$\mathbb{E}[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \left[ -e^{-\frac{x^2}{2}} \right]_{-\infty}^{\infty} = 0,$$

puis, par intégration par parties,

$$\begin{aligned} \text{var}(X) = \mathbb{E}[X^2] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \left[ -x e^{-\frac{x^2}{2}} \right]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \\ &= 0 + 1 = 1. \end{aligned}$$

D'autre part, si  $X \sim \mathcal{N}(0, 1)$ , alors  $Y = \sigma X + m \sim \mathcal{N}(m, \sigma^2)$ , car

$$\begin{aligned} \mathbb{P}[\sigma X + m \leq t] &= \mathbb{P}\left[X \leq \frac{t-m}{\sigma}\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{t-m}{\sigma}} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^t e^{-\frac{(u-m)^2}{2\sigma^2}} du \quad \text{en posant } u = \sigma x + m. \end{aligned}$$

On en déduit que  $\mathbb{E}[Y] = m$  et  $\text{var}(Y) = \sigma^2$  si  $Y \sim \mathcal{N}(m, \sigma^2)$ . On a ainsi calculé l'espérance et la variance d'une loi normale de paramètres arbitraires. Il n'existe en revanche pas de formule simple (autre que la formule intégrale) pour la fonction de répartition.

On verra dans le chapitre suivant que les lois gaussiennes apparaissent naturellement comme limites continues de lois de sommes de variables aléatoires indépendantes, ce qui généralisera le théorème vu précédemment avec des sommes de variables de Bernoulli. Une dernière famille de lois continues classiques qui apparaîtra dans certains résultats de statistiques est la famille des **lois du  $\chi^2$** . Soit  $k$  un entier positif, et

$$Z = \sum_{i=1}^k (X_i)^2,$$

où les  $X_i$  sont des variables indépendantes, continues, toutes de loi gaussienne standard  $\mathcal{N}(0, 1)$ . On dit que  $Z$  est une variable  $\chi^2$  à  $k$  degrés de liberté, ce qu'on notera  $Z \sim \chi^2(k)$ .

PROPOSITION 3.23. *La densité de la variable  $Z$  est*

$$f_Z(x) = \frac{1}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} 1_{x \geq 0},$$

où  $\Gamma$  est la fonction définie sur  $\mathbb{R}_+^*$  par  $\Gamma(t) = \int_{s=0}^{\infty} s^{t-1} e^{-s} ds$ .

LEMME 3.24. Soit  $k \in \mathbb{N}$ . Les valeurs entières de la fonction  $\Gamma$  sont  $\Gamma(k+1) = k!$ , et les valeurs demi-entières de la fonction  $\Gamma$  sont données par

$$\Gamma\left(k + \frac{1}{2}\right) = \sqrt{\pi} \frac{(2k-1)(2k-3)\cdots 3 \cdot 1}{2^k}.$$

DÉMONSTRATION. On a déjà calculé par récurrence sur  $k$  la valeur de

$$\Gamma(k+1) = \int_0^{\infty} x^k e^{-x} dx = k!;$$

c'est la valeur du moment  $\mathbb{E}[X^k]$  avec  $X \sim \mathcal{E}(1)$ . Pour les valeurs demi-entières, on a par intégration par parties :

$$\Gamma\left(k + \frac{1}{2}\right) = \int_0^{\infty} x^{k-\frac{1}{2}} e^{-x} dx = \left(k - \frac{1}{2}\right) \int_0^{\infty} x^{k-\frac{3}{2}} e^{-x} dx = \left(k - \frac{1}{2}\right) \Gamma\left(k - \frac{1}{2}\right).$$

Par récurrence sur  $k$ , on en déduit que

$$\Gamma\left(k + \frac{1}{2}\right) = \frac{(2k-1)(2k-3)\cdots 3 \cdot 1}{2^k} \Gamma\left(\frac{1}{2}\right),$$

de sorte qu'il suffit maintenant de calculer la valeur de  $\Gamma(\frac{1}{2})$ . Elle s'obtient par changement de variables :

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{\infty} e^{-x} \frac{dx}{\sqrt{x}} = 2\sqrt{2} \int_0^{\infty} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2}} \int_{\mathbb{R}} e^{-\frac{t^2}{2}} dt = \sqrt{\pi}.$$

□

LEMME 3.25. La densité de la loi  $\chi^2(1)$  est  $\sqrt{\frac{1}{2\pi x}} e^{-\frac{x}{2}} 1_{x \geq 0}$ .

DÉMONSTRATION. On calcule la fonction de répartition  $F_Z(t)$ , avec  $Z \sim \chi^2(1)$ , qu'on écrit comme le carré d'une variable  $X$  de loi  $\mathcal{N}(0, 1)$  :

$$\begin{aligned} \mathbb{P}[Z \leq t] &= \mathbb{P}[X^2 \leq t] = \mathbb{P}[-\sqrt{t} \leq X \leq \sqrt{t}] = 2\mathbb{P}[0 \leq X \leq \sqrt{t}] \\ &= 2 \int_{s=0}^{\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds = \int_{x=0}^t \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} dx \end{aligned}$$

en faisant le changement de variables  $s = \sqrt{x}$  dans la dernière formule. En dérivant par rapport à  $t$ , on conclut que

$$f_Z(x) = F'_Z(x) = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} 1_{x \geq 0}.$$

□

LEMME 3.26. La densité de la loi  $\chi^2(2)$  est  $\frac{1}{2} e^{-\frac{x}{2}} 1_{x \geq 0}$ , c'est-à-dire que  $\chi^2(2) = \mathcal{E}(\frac{1}{2})$ .



DÉMONSTRATION. On veut calculer la fonction de répartition de  $Z = X + Y$ , où  $X$  et  $Y$  sont les carrés de variables gaussiennes standards indépendantes, c'est-à-dire des variables indépendantes de loi  $\chi^2(1)$ . On a vu que pour toutes fonctions  $g$  et  $h$ , on avait dans le cadre de variables indépendantes

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)] \mathbb{E}[h(Y)] = \int_{\mathbb{R}^2} g(x)h(y) f_X(x) f_Y(y) dx dy.$$

Une généralisation de cette identité, qui a essentiellement la même preuve, est la suivante : si  $X$  et  $Y$  sont indépendantes, alors pour toute fonction  $G : \mathbb{R}^2 \rightarrow \mathbb{R}$ ,

$$\mathbb{E}[G(X, Y)] = \int_{\mathbb{R}^2} G(x, y) f_X(x) f_Y(y) dx dy.$$

Appliquons cette identité au cas où  $G$  est la fonction  $G(x, y) = 1_{x+y \leq t}$ , avec  $t \geq 0$ . On obtient :

$$\begin{aligned} F_{X+Y}(t) &= \mathbb{E}[G(X, Y)] = \int_{x=0}^{\infty} \int_{y=0}^{\infty} 1_{x+y \leq t} f_X(x) f_Y(y) dx dy \\ &= \int_{x=0}^t \int_{y=0}^{t-x} f_X(x) f_Y(y) dx dy = \int_{x=0}^t f_X(x) F_Y(t-x) dx \end{aligned}$$

En dérivant cette identité par rapport à  $t$ , on obtient :

$$f_{X+Y}(t) = f_X(x) F_Y(0) + \int_{x=0}^t f_X(x) f_Y(t-x) dx = \int_{x=0}^t f_X(x) f_Y(t-x) dx.$$

Cette formule, appelée **formule du produit de convolution**, sera réexpliquée en détail dans le chapitre suivant. En l'utilisant avec  $f_X = f_Y$  égale à la densité de la loi  $\chi^2(1)$ , on obtient :

$$f_Z(t) = \int_{x=0}^t \frac{1}{2\pi \sqrt{x(t-x)}} e^{-\frac{t}{2}} dx = \frac{e^{-\frac{t}{2}}}{2\pi} \int_{x=0}^t \frac{1}{\sqrt{x(t-x)}} dx.$$

Posons  $x = t \sin^2 \theta$ , de sorte que  $dx = 2t \sin \theta \cos \theta d\theta$  et  $\sqrt{x(t-x)} = t \sin \theta \cos \theta$  pour  $\theta \in [0, \frac{\pi}{2}]$ . On obtient par ce changement de variables

$$f_Z(t) = \frac{e^{-\frac{t}{2}}}{\pi} \int_0^{\frac{\pi}{2}} d\theta = \frac{1}{2} e^{-\frac{t}{2}}$$

pour tout  $t \geq 0$ , ce que l'on voulait démontrer.  $\square$

LEMME 3.27. La densité de la loi  $\chi^2(2k)$  est  $\frac{1}{2^k (k-1)!} x^{k-1} e^{-\frac{x}{2}} 1_{x \geq 0}$ .

DÉMONSTRATION. On montre le résultat par récurrence sur  $k$ , étant entendu que la formule vient d'être démontrée lorsque  $k = 1$ . Supposons le résultat vrai au rang  $k-1 \geq 1$ , et considérons une variable  $Z$  de loi  $\chi^2(2k)$ , qu'on écrit comme somme  $X + Y$  avec  $X$  et  $Y$  indépendants,  $X \sim \chi^2(2k-2)$  et  $Y \sim \chi^2(2)$ . La formule du produit de convolution donnée précédemment s'applique de nouveau, de sorte que

$$f_Z(t) = \int_{x=0}^t f_X(x) f_Y(t-x) dx = \frac{1}{2^k (k-2)!} e^{-\frac{t}{2}} \int_{x=0}^t x^{k-2} dx = \frac{1}{2^k (k-1)!} e^{-\frac{t}{2}} t^{k-1},$$

ce que l'on voulait démontrer.  $\square$

La formule pour la densité de  $\chi^2(2k)$  est donc établie dans le cas d'un nombre pair de degrés de liberté, et le cas impair est traité par le dernier lemme suivant :

LEMME 3.28. *La densité de la loi  $\chi^2(2k+1)$  est  $\frac{1}{2^{k+\frac{1}{2}} \Gamma(k+\frac{1}{2})} x^{k-\frac{1}{2}} e^{-\frac{x}{2}} 1_{x \geq 0}$ .*

DÉMONSTRATION. On raisonne par récurrence sur  $k$ , le cas  $k=0$  ayant déjà été traité. Supposons le résultat vrai au rang  $k-1 \geq 0$ , et soit  $Z$  de loi  $\chi^2(2k+1)$ , qu'on écrit sous la forme  $Z = X + Y$  avec  $X \sim \chi^2(2k-1)$  et  $Y \sim \chi^2(2)$ . En utilisant de nouveau la formule du produit de convolution, on obtient

$$\begin{aligned} f_Z(t) &= \int_{x=0}^t f_X(x) f_Y(t-x) dx = \frac{e^{-\frac{t}{2}}}{2^{k+\frac{1}{2}} \Gamma(k-\frac{1}{2})} \int_{x=0}^t x^{k-\frac{3}{2}} dx \\ &= \frac{e^{-\frac{t}{2}}}{2^{k+\frac{1}{2}} (k-\frac{1}{2}) \Gamma(k-\frac{1}{2})} t^{k-\frac{1}{2}} = \frac{e^{-\frac{t}{2}}}{2^{k+\frac{1}{2}} \Gamma(k+\frac{1}{2})} t^{k-\frac{1}{2}} \end{aligned}$$

puisque  $(k-\frac{1}{2}) \Gamma(k-\frac{1}{2}) = \Gamma(k+\frac{1}{2})$ , comme on l'a vu lors du calcul des valeurs demi-entières de la fonction  $\Gamma$  (de façon plus générale, on a l'équation fonctionnelle  $\Gamma(s+1) = s \Gamma(s)$  valable pour tout  $s > 0$ ).  $\square$

Un calcul nettement plus simple est celui de la moyenne et de la variance de la loi  $\chi^2(k)$ . En effet, si  $X \sim \chi^2(k)$ , alors c'est la somme de  $k$  variables indépendantes de loi  $\chi^2(1)$ , donc

$$\mathbb{E}[X] = k \mathbb{E}[Y] \quad ; \quad \text{var}(X) = k \text{var}(Y)$$

avec  $Y \sim \chi^2(1)$ . Avec  $G \sim \mathcal{N}(0,1)$ , on calcule aisément  $\mathbb{E}[Y] = \mathbb{E}[G^2] = 1$ , puis par intégration par parties

$$\mathbb{E}[Y^2] = \mathbb{E}[G^4] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^4 e^{-\frac{x^2}{2}} dx = \frac{3}{\sqrt{2\pi}} \int_{\mathbb{R}} x^2 e^{-\frac{x^2}{2}} dx = 3$$

et donc  $\text{var}(Y) = 3 - 1 = 2$ . Ainsi :

PROPOSITION 3.29. *Si  $X \sim \chi^2(k)$ , alors  $\mathbb{E}[X] = k$  et  $\text{var}(X) = 2k$ .*

REMARQUE 3.30. Dans la formule pour la densité des lois  $\chi^2$ , on peut prendre un paramètre  $k > 0$  qui n'est pas un entier : on obtient alors une mesure de probabilité  $\chi^2(k)$  sur  $\mathbb{R}_+$ , qui correspond bien à une variable d'espérance  $k$  et de variance  $2k$ . En revanche, cette variable aléatoire n'a plus la loi de la somme de carrés de gaussiennes indépendantes.

Dans le tableau qui suit, on a résumé tous les calculs sur les lois classiques continues :

loi	intervalle des valeurs possibles	densité	espérance	variance
$\mathcal{U}([a, b])$	$[a, b]$	$\frac{1}{b-a} 1_{x \in [a, b]}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$\mathcal{E}(\lambda)$	$\mathbb{R}_+$	$\lambda e^{-\lambda x} 1_{x \geq 0}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$\mathcal{N}(m, \sigma^2)$	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$	$m$	$\sigma^2$
$\chi^2(k)$	$\mathbb{R}_+$	$\frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} 1_{x \geq 0}$	$k$	$2k$

## Exercices

- (1) **Un calcul de densité.** Soit  $X$  une variable aléatoire à densité, dont la fonction de répartition est :

$$F_X(t) = \begin{cases} 0 & \text{si } t < 0, \\ t^3 & \text{si } t \in [0, 1], \\ 1 & \text{si } t > 1. \end{cases}$$

- (a) Calculer la densité de  $X$ .  
 (b) Calculer l'espérance et la variance de  $X$ .  
 (c) Soit  $Y = \frac{1}{X}$ . Donner la fonction de répartition, puis la densité de  $Y$ .
- (2) **Un calcul d'intégrales.** Soit  $X$  une variable aléatoire réelle, dont la densité  $f_X$  a pour forme  $f_X(x) = \frac{C}{(1+x^2)^2}$ .

- (a) Trouver la valeur de la constante  $C$ . On pourra utiliser l'identité

$$\frac{1}{(1+x^2)^2} = \frac{1}{1+x^2} - \frac{x^2}{(1+x^2)^2}.$$

- (b) Calculer l'espérance et la variance de  $X$ . Que vaut  $\mathbb{E}[X^3]$  ?
- (3) **Moments des lois gaussiennes et du  $\chi^2$ .**
- (a) Soit  $X \sim \mathcal{N}(0, 1)$ . Montrer que  $\mathbb{E}[X^{2n}] = (2n-1)(2n-3) \cdots 3 \cdot 1$  et  $\mathbb{E}[X^{2n+1}] = 0$  pour tout  $n \in \mathbb{N}$ .  
 (b) Soit  $X \sim \chi^2(k)$ . Montrer que  $\mathbb{E}[X^n] = k(k+2) \cdots (k+2n-2)$  pour tout  $n \in \mathbb{N}$ .

- (4) **Lois gamma.** On rappelle que la fonction  $\Gamma$  est définie par l'équation intégrale  $\Gamma(t) = \int_{s=0}^{\infty} s^{t-1} e^{-s} ds$ . La loi  $\gamma$  de paramètres  $a, b > 0$ , notée  $\gamma(a, b)$ , est donnée par la densité

$$f_X(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} 1_{x \geq 0}.$$

- (a) Vérifier que  $\int_{\mathbb{R}} f_X(x) dx = 1$ .  
 (b) Calculer l'espérance et la variance de  $X \sim \gamma(a, b)$ .  
 (c) Quelle loi retrouve-t-on lorsque  $a = 1$  ? lorsque  $a = 2$  ?

- (5) **Lois beta.**

- (a) Si  $a, b > 0$ , montrer que

$$\int_0^1 x^{a-1} (1-x)^{b-1} dx = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a+b)}.$$

On pourra faire un changement de variables dans le calcul de l'intégrale double  $\Gamma(a) \Gamma(b)$ .

(b) En déduire que

$$f_X(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \mathbf{1}_{x \in [0,1]}$$

est la densité d'une variable aléatoire continue, et qu'elle correspond à la loi d'un rapport  $\frac{Y}{Y+Z}$ , où  $Y$  et  $Z$  sont deux variables indépendantes de lois  $\gamma(a, c)$  et  $\gamma(b, c)$  avec  $c > 0$  paramètre arbitraire. On dit que  $f_X$  est la densité d'une loi  $\beta$  de paramètres  $a$  et  $b$ , notée  $\beta(a, b)$ .

(c) Quelle loi retrouve-t-on lorsque  $a = b = 1$  ?

(d) Calculer la moyenne et la variance de  $X \sim \beta(a, b)$ .

(6) **Simulation de variables aléatoires.** Soit  $F_X$  une fonction strictement croissante sur un intervalle  $I$ , à valeurs dans  $[0, 1]$ ; et  $F_X^{-1}$  la bijection réciproque  $[0, 1] \rightarrow I$ . On note  $U$  une variable aléatoire de loi  $\mathcal{U}([0, 1])$ .

(a) Montrer que  $X = F_X^{-1}(U)$  est une variable aléatoire qui a pour fonction de répartition  $F_X$ .

(b) En déduire comment simuler n'importe quelle variable aléatoire à partir d'une variable aléatoire uniforme.

(7) **Densité d'une fonction d'une variable aléatoire.** Soit  $X$  une variable aléatoire à valeurs dans un intervalle  $I$ , et de densité  $f_X(x)$ .

(a) Soit  $g : I \rightarrow J$  une fonction dérivable et strictement croissante, et  $Y = g(X)$ . Montrer que la densité de  $Y$  est donnée par la formule

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}.$$

(b) Que dire si  $g$  est strictement décroissante ?

(c) Si  $X \sim \mathcal{E}(1)$ , donner la densité de la variable aléatoire  $Y = \frac{1}{X}$ .

(d) Soit  $X$  une variable aléatoire continue à valeurs dans  $\mathbb{R}$ , de densité  $f_X(x)$ . Donner une formule pour la fonction de répartition, puis pour la densité de  $X^2$ .

(8) **Lois exponentielles et lois géométriques.** Soit  $X$  une variable aléatoire continue, de loi  $\mathcal{E}(\lambda)$ .

(a) Calculer pour tout entier  $k \geq 1$  la probabilité pour que  $X$  soit compris entre  $k-1$  et  $k$ .

(b) En déduire que  $\lceil X \rceil$ , l'entier approchant  $X$  par valeurs supérieures, suit une loi géométrique de paramètre à préciser.

(9) **Variables exponentielles indépendantes et lois de Poisson.** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires indépendantes, chaque variable  $X_n$  suivant une loi exponentielle  $\mathcal{E}(\lambda)$ .

(a) Montrer que pour tout  $n \in \mathbb{N}$ , la somme  $S_n = X_1 + X_2 + \dots + X_n$  suit une loi  $\gamma(n, \lambda)$ . On pourra utiliser la formule du produit de convolution.

- (b) Calculer la probabilité de l'événement  $(S_n \leq 1 \text{ et } S_{n+1} > 1)$  (indication : comparer  $\mathbb{P}[S_{n+1} \leq 1]$  et  $\mathbb{P}[S_n \leq 1]$ ).
- (c) En déduire la loi de  $N = \max\{n \in \mathbb{N} \mid S_n \leq 1\}$ .
- (d) Application : on considère une pièce mécanique, qui subit des pannes espacées par des intervalles de temps aléatoires  $X_i$ , qui sont des variables aléatoires indépendantes exponentielles de même paramètre  $\lambda$ . Calculer l'espérance et la variance du nombre  $N$  de pannes observées durant l'intervalle de temps  $[0, 1]$ .
- (10) **Loi du maximum de deux variables aléatoires indépendantes.** Soient  $X$  et  $Y$  deux variables aléatoires continues indépendantes, de densités respectives  $f_X$  et  $f_Y$ .
- (a) On pose  $V = \max(X, Y)$ . Donner un lien entre les fonctions de répartition  $F_V$ ,  $F_X$  et  $F_Y$ . En déduire la valeur de la densité de  $V$ .
- (b) On pose  $U = \min(X, Y)$ . En remarquant que  $\min(x, y) = -\max(-x, -y)$ , donner la densité de  $U$ .
- (c) Application : on suppose que  $X$  et  $Y$  sont des variables exponentielles de paramètres  $\lambda$  et  $\mu$ . Donner la loi de  $U = \min(X, Y)$ .

- (11) **Loi de Cauchy.** Soit  $U$  et  $V$  deux variables aléatoires indépendantes, toutes deux de loi gaussienne  $\mathcal{N}(0, 1)$ . On note  $X = \frac{U}{V}$ .

- (a) En faisant un changement de variables polaires, montrer que

$$\mathbb{P}[X \leq t] = \frac{1}{2\pi} \int_{\theta=0}^{2\pi} \int_{r=0}^{\infty} r e^{-\frac{r^2}{2}} \mathbf{1}_{\tan(\theta) \leq t} dr d\theta$$

- (b) En déduire la valeur de la fonction de répartition de la variable  $X$ .
- (c) Montrer que la densité de  $X$  est  $f_X(x) = \frac{1}{\pi(1+x^2)}$ . On dit que  $X$  suit une loi de Cauchy.
- (d) Que vaut  $\mathbb{E}[X]$  ?
- (e) Quelle est la loi de  $\frac{1}{X}$  ?

- (12) **Distance entre deux points aléatoires.** On note  $X$  et  $Y$  deux variables indépendantes de même loi  $\mathcal{U}([0, a])$ , et  $Z = |X - Y|$  la distance entre ces deux points.

- (a) Écrire  $\mathbb{P}[Z \leq t] = F_Z(t)$  comme une intégrale double.
- (b) On voit  $F_Z(t) = F_Z(t, a)$  comme une fonction de  $t$  et du paramètre  $a$ . Montrer que

$$\frac{\partial F_Z(t, a)}{\partial a} = -\frac{2F_Z(t, a)}{a} + \frac{2t}{a^2}.$$

- (c) En déduire la fonction de répartition de  $Z$  s'écrit sous la forme  $F_Z(t, a) = \frac{2t}{a} + \frac{K(t)}{a^2}$ , où  $K(t)$  est une fonction qui ne dépend que de  $t$  (et pas de  $a$ ).
- (d) Montrer que  $F_Z(t, a)$  est une fonction du rapport  $\frac{t}{a}$ . En déduire la valeur de  $K(t)$ , et de  $F_Z(t, a)$ . Si  $a$  est fixé, quelle est la densité de  $Z$  ? son espérance ?

(13) **Le paradoxe de Bertrand.** On étudie dans cet exercice deux façons différentes de choisir une corde reliant deux points aléatoires d'un cercle de rayon 1.

- (a) Soit  $U$  et  $V$  deux variables aléatoires indépendantes de même loi uniforme sur le segment  $[0, 2\pi]$ . On note  $L$  la longueur de la corde reliant les deux points d'angle  $U$  et  $V$  sur le cercle unité. Montrer que

$$L = 2 \sin \left| \frac{V - U}{2} \right|.$$

- (b) Donner la loi de  $\left| \frac{V-U}{2} \right|$ , puis celle de  $L$  (on demande les densités dans les deux cas).
- (c) Pour choisir une corde au hasard sur un cercle, on peut également faire comme suit : on tire au hasard une direction  $\theta \in [0, 2\pi]$ , puis un point  $P$  au hasard sur le rayon reliant l'origine au point d'angle  $\theta$  sur le cercle, la distance de ce point à l'origine suivant une loi  $\mathcal{U}([0, 1])$ . On choisit alors l'unique corde du cercle qui a pour milieu  $P$ . Donner dans ce cas la loi de  $L'$ , la longueur de la corde.
- (d) Dans quel cas a-t-on la plus longue corde en moyenne ?

## Sommes de variables indépendantes

**Contenu :** couples de variables aléatoires, loi jointe, loi marginale ; loi de la somme de deux variables indépendantes, convolution ; loi des grands nombres, théorème central limite.

### 1. Généralités sur les familles de variables aléatoires

Dans les chapitres précédents, on a rencontré à plusieurs reprises des couples  $(X, Y)$  ou des familles  $(X_1, \dots, X_d)$  de variables aléatoires indépendantes. Plus généralement, on peut essayer de décrire des vecteurs  $(X_1, \dots, X_d)$  dont les coordonnées sont des variables discrètes à valeurs dans  $\mathbb{Z}$  ou des variables continues à valeurs dans  $\mathbb{R}$ , sans imposer que ces coordonnées soient indépendantes. Ce point de vue mènera *in fine* à une meilleure compréhension de l'indépendance de variables aléatoires, particulièrement dans le cas continu.

Pour simplifier un peu l'exposé, on se concentrera sur le cas où  $d = 2$ , qui contient déjà toutes les difficultés. Un **couple de variables aléatoires discrètes** est un vecteur aléatoire  $(X, Y) \in \mathbb{Z}^2$ , dont la **loi jointe** est la fonction

$$\mathbb{P}_{(X,Y)}[k, l] = \mathbb{P}[X = k \text{ et } Y = l]$$

définie pour tous entiers  $k$  et  $l$ . La série double  $\sum_{(k,l) \in \mathbb{Z}^2} \mathbb{P}_{(X,Y)}[k, l]$  a pour somme 1, et pour toute partie  $A \subset \mathbb{Z}^2$ ,

$$\mathbb{P}[(X, Y) \in A] = \sum_{(k,l) \in A} \mathbb{P}_{(X,Y)}[k, l]$$

EXEMPLE 4.1. Soit  $\mathbb{P}_{(X,Y)}$  la fonction sur  $\mathbb{Z}^2$  dont les valeurs non nulles sont données par le tableau suivant :

$k \backslash l$	1	2	3
1	0	$\frac{1}{12}$	$\frac{1}{6}$
2	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{3}$

Comme  $0 + \frac{1}{12} + \frac{1}{6} + \frac{1}{6} + \frac{1}{4} + \frac{1}{3} = 1$ ,  $\mathbb{P}_{(X,Y)}$  est la loi jointe d'un couple de variables aléatoires discrètes  $(X, Y)$ , avec, par exemple,  $\mathbb{P}[X = 1 \text{ et } Y = 2] = \frac{1}{12}$ .

EXEMPLE 4.2. Soit  $X$  une variable aléatoire qui suit une loi géométrique décalée de  $-1$ , de paramètre  $p$  :

$$\mathbb{P}[X = k] = (1 - p)^k p \quad \text{pour tout } k \geq 0.$$

On considère alors  $Y$ , somme de  $X$  variables aléatoires indépendantes de Bernoulli de paramètre  $q$ , ces variables étant indépendantes de  $X$  :

$$Y = \sum_{j=1}^X \xi_j, \quad \text{avec les } \xi_j \text{ indépendantes entre elles et indépendantes de } X.$$

Sachant  $X = k$ , la probabilité pour que  $Y = l$  avec  $l \in \llbracket 0, k \rrbracket$  est donnée par une loi binomiale :

$$\mathbb{P}[Y = l | X = k] = \binom{k}{l} q^l (1 - q)^{k-l}.$$

Par conséquent, la loi jointe de  $(X, Y)$  porte sur les couples d'entiers  $(k, l)$  avec  $0 \leq k \leq l$ , et elle est donnée par la formule

$$\mathbb{P}[X = k \text{ et } Y = l] = \mathbb{P}[X = k] \mathbb{P}[Y = l | X = k] = \binom{k}{l} (1 - p)^k p q^l (1 - q)^{k-l}.$$

Étant donné un couple de variables aléatoires discrètes  $(X, Y)$ , chaque coordonnée  $X$  ou  $Y$  constitue une variable aléatoire discrète à valeurs dans  $\mathbb{Z}$ , dont on peut considérer la loi  $\mathbb{P}_X$  ou  $\mathbb{P}_Y$ . On dit que ces deux lois  $\mathbb{P}_X$  et  $\mathbb{P}_Y$  sont les **lois marginales** du couple  $(X, Y)$ , et elles sont données par les formules :

$$\mathbb{P}_X[k] = \sum_{l \in \mathbb{Z}} \mathbb{P}_{(X,Y)}[k, l] \quad ; \quad \mathbb{P}_Y[l] = \sum_{k \in \mathbb{Z}} \mathbb{P}_{(X,Y)}[k, l].$$

EXEMPLE 4.3. Considérons de nouveau le couple  $(X, Y)$  de variables aléatoires à valeurs dans  $\{1, 2\} \times \{1, 2, 3\}$ , dont la loi jointe donnée par le tableau de l'exemple 4.1. La loi marginale de  $X$  est calculée en faisant une somme sur chaque ligne, et la loi marginale de  $Y$  est calculée en faisant une somme sur chaque colonne :

$k \backslash l$	1	2	3	$\mathbb{P}_X$
1	0	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$
2	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{3}{4}$
$\mathbb{P}_Y$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	1

EXEMPLE 4.4. Soit  $(X, Y)$  le couple de variables aléatoires de l'exemple 4.2, de loi jointe

$$\mathbb{P}[X = k \text{ et } Y = l] = \binom{k}{l} (1 - p)^k p q^l (1 - q)^{k-l}$$



pour  $k \geq 0$  et  $l \in \llbracket 0, k \rrbracket$ . La loi marginale de  $X$  est

$$\mathbb{P}_X[k] = \sum_{l=0}^k \binom{k}{l} (1-p)^k p^l (1-q)^{k-l} = (1-p)^k p,$$

c'est-à-dire une loi géométrique de paramètre  $p$ , décalée de  $-1$ . La loi marginale de  $Y$  est donnée par

$$\begin{aligned} \mathbb{P}_Y[l] &= \sum_{k=l}^{\infty} \binom{k}{l} p(1-p)^k q^l (1-q)^{k-l} \\ &= p(1-p)^l q^l \sum_{k=l}^{\infty} \binom{k}{l} [(1-p)(1-q)]^{k-l} \\ &= p(1-p)^l q^l \sum_{j=0}^{\infty} \binom{j+l}{l} [(1-p)(1-q)]^j \end{aligned}$$

pour tout  $l \geq 0$ . Remarquons que le développement de Taylor de  $\frac{1}{(1-x)^{l+1}}$  en  $x = 0$  est

$$\begin{aligned} \frac{1}{(1-x)^{l+1}} &= 1 + (l+1)x + \frac{(l+1)(l+2)}{2}x^2 + \frac{(l+1)(l+2)(l+3)}{6}x^3 + \dots \\ &= \sum_{j=0}^{\infty} \binom{j+l}{l} x^j. \end{aligned}$$

Par conséquent,

$$\mathbb{P}_Y[l] = \frac{p(1-p)^l q^l}{(1-(1-p)(1-q))^{l+1}} = \left(1 - \frac{p}{p+q-pq}\right)^l \frac{p}{p+q-pq},$$

ce qui est de nouveau une loi géométrique décalée.

REMARQUE 4.5. Attention, les lois marginales  $\mathbb{P}_X$  et  $\mathbb{P}_Y$  ne déterminent pas la loi jointe  $\mathbb{P}_{(X,Y)}$ . Par exemple, le tableau suivant

$k \setminus l$	1	2	3	$\mathbb{P}_X$
1	$\frac{1}{24}$	$\frac{1}{12}$	$\frac{1}{8}$	$\frac{1}{4}$
2	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{3}{4}$
$\mathbb{P}_Y$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	1

correspond à un couple de variables aléatoires discrètes  $(X, Y)$  qui a les mêmes lois marginales que dans l'exemple 4.1, mais pas la même loi jointe.

Avec les définitions données précédemment, on peut réinterpréter la notion d'indépendance définie dans le second chapitre. Ainsi :

PROPOSITION 4.6. Soit  $(X, Y)$  un couple de variables aléatoires discrètes. Les variables  $X$  et  $Y$  sont indépendantes si et seulement si la loi jointe  $\mathbb{P}_{(X,Y)}$  est le produit des lois marginales  $\mathbb{P}_X$  et  $\mathbb{P}_Y$  :

$$\forall (k, l) \in \mathbb{Z}^2, \quad \mathbb{P}_{(X,Y)}[k, l] = \mathbb{P}_X[k] \mathbb{P}_Y[l].$$

EXEMPLE 4.7. Dans l'exemple 4.1, les variables  $X$  et  $Y$  ne sont pas indépendantes, car  $\mathbb{P}_{(X,Y)}[1, 1] = 0 \neq \mathbb{P}_X[1] \mathbb{P}_Y[1]$ . En revanche, elles le sont dans la remarque 4.5.

REMARQUE 4.8. Si l'on suppose  $X$  et  $Y$  indépendantes, alors les lois marginales  $\mathbb{P}_X$  et  $\mathbb{P}_Y$  déterminent entièrement la loi jointe  $\mathbb{P}_{(X,Y)}$ . D'après la remarque précédente, on ne peut pas se passer de l'indépendance pour avoir ce résultat.

Traisons maintenant le cas continu. Un **couple de variables aléatoires à densité** est un vecteur aléatoire  $(X, Y)$  à valeurs dans  $\mathbb{R}^2$ , tel qu'existe une fonction  $f_{(X,Y)} : \mathbb{R}^2 \rightarrow \mathbb{R}_+$  intégrable, avec pour tout rectangle  $[a, b] \times [c, d]$ ,

$$\mathbb{P}[(X, Y) \in [a, b] \times [c, d]] = \mathbb{P}[a \leq X \leq b \text{ et } c \leq Y \leq d] = \int_{x=a}^b \int_{y=c}^d f_{(X,Y)}(x, y) dx dy.$$

On dit alors que  $f_{(X,Y)}$  est la **densité jointe** du couple  $(X, Y)$ , et on a la condition de normalisation

$$\iint_{\mathbb{R}^2} f_{(X,Y)}(x, y) dx dy = 1.$$

EXEMPLE 4.9. Soit  $f_{(X,Y)}(x, y)$  la fonction qui vaut 0 en dehors de  $[0, 1] \times [0, 1]$ , et qui est donnée par la formule

$$f_{(X,Y)}(x, y) = \frac{2(x^2 + y^2)}{1 + xy}$$

pour  $(x, y) \in [0, 1]^2$ .

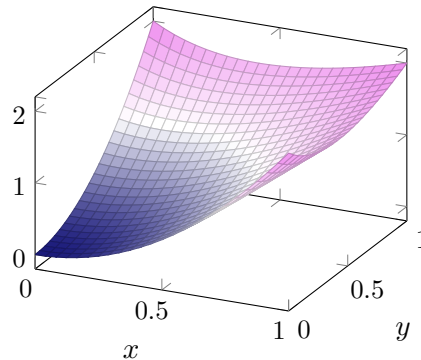


FIGURE 1. La densité jointe d'un couple de variables  $(X, Y)$  à valeurs dans  $[0, 1]^2$ .

C'est la densité de probabilité jointe d'un couple  $(X, Y)$  de variables aléatoires continues à valeurs dans  $[0, 1]^2$ , car  $\iint_{[0,1]^2} f_{(X,Y)}(x, y) dx dy = 1$ . En effet,

$$\begin{aligned} I &= \int_{x=0}^1 \int_{y=0}^1 \frac{x^2 + y^2}{1 + xy} dx dy = \int_{x=0}^1 \left( \int_{y=0}^1 \frac{xy - 1}{x^2} + \left(x^2 + \frac{1}{x^2}\right) \frac{1}{1 + xy} dy \right) dx \\ &= \int_{x=0}^1 \left( \frac{1}{2x} - \frac{1}{x^2} + x \log(1 + x) + \frac{\log(1 + x)}{x^3} \right) dx. \end{aligned}$$

Pour résoudre la singularité de la fonction en  $x = 0$ , on utilise la formule de Taylor avec reste intégral à l'ordre 3 pour  $f(x) = \log(1 + x)$  :

$$\log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{2} \int_0^1 (1 - x^2) f^{(3)}(ux) du = x - \frac{x^2}{2} + x^3 \int_0^1 \frac{(1 - u)^2}{(1 + ux)^3} du.$$

On a donc bien :

$$\begin{aligned} I &= \int_{x=0}^1 x \log(1 + x) dx + \int_{u=0}^1 \int_{x=0}^1 \frac{(1 - u)^2}{(1 + ux)^3} du dx \\ &= \left[ \frac{x^2 - 1}{2} \log(1 + x) + \frac{2x - x^2}{4} \right]_{x=0}^1 + \frac{1}{2} \int_{u=0}^1 \frac{(2 + u)(1 - u)^2}{(1 + u)^2} du \\ &= \frac{1}{4} + \frac{1}{2} \left[ \frac{u^2}{2} - 2u - \frac{4}{1 + u} \right]_{u=0}^1 = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}. \end{aligned}$$

Comme dans le cas discret, on définit les **densités marginales** de  $X$  et  $Y$  par les formules :

$$\begin{aligned} f_X(x) &= \int_{y \in \mathbb{R}} f_{(X,Y)}(x, y) dy; \\ f_Y(y) &= \int_{x \in \mathbb{R}} f_{(X,Y)}(x, y) dx. \end{aligned}$$

**PROPOSITION 4.10.** *Soit  $(X, Y)$  un couple de variables aléatoires à densité, de densité jointe  $f_{(X,Y)}$ . Les variables  $X$  et  $Y$  sont des variables continues à valeurs dans  $\mathbb{R}$  et de densités respectives  $f_X$  et  $f_Y$ , données par les formules ci-dessus.*

**DÉMONSTRATION.** Si  $[a, b]$  est un intervalle, alors

$$\mathbb{P}[X \in [a, b]] = \mathbb{P}[X \in [a, b] \text{ et } Y \in \mathbb{R}] = \int_{x=a}^b \left( \int_{y \in \mathbb{R}} f_{(X,Y)}(x, y) dy \right) dx,$$

donc la densité de la variable  $X$  en  $x$  est bien  $\int_{y \in \mathbb{R}} f_{(X,Y)}(x, y) dy$ , et le cas de  $Y$  est symétrique.  $\square$

EXEMPLE 4.11. Dans l'exemple 4.9, on a été amené à calculer la densité marginale de  $X$  :

$$f_X(x) = \begin{cases} \int_0^1 \frac{2(x^2+y^2)}{1+xy} dy = \frac{1}{x} - \frac{2}{x^2} + 2x \log(1+x) + \frac{2 \log(1+x)}{x^3} & \text{si } x \in [0, 1], \\ 0 & \text{sinon.} \end{cases}$$

Par symétrie, on a aussi

$$f_Y(y) = \begin{cases} \frac{1}{y} - \frac{2}{y^2} + 2y \log(1+y) + \frac{2 \log(1+y)}{y^3} & \text{si } y \in [0, 1], \\ 0 & \text{sinon.} \end{cases}$$

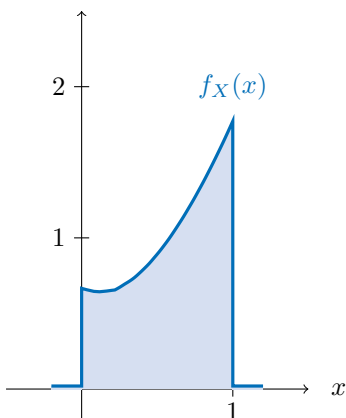


FIGURE 2. Loi marginale de  $X$  lorsque  $f_{(X,Y)}(x,y) = 1_{(x,y) \in [0,1]^2} \frac{2(x^2+y^2)}{1+xy}$ .

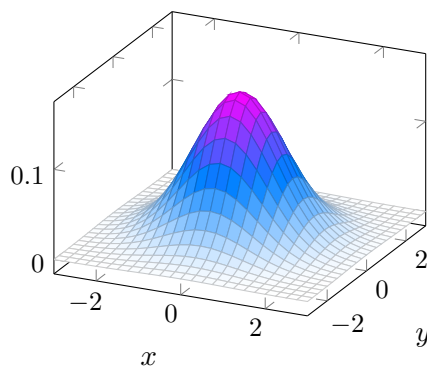


FIGURE 3. Densité de probabilité d'un couple de variables  $(X, Y)$  avec  $X \sim \mathcal{N}(0, 1)$  et  $Y \sim \mathcal{N}(0, 1)$ .

EXEMPLE 4.12. Soit  $f_{(X,Y)}(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$ . L'intégrale de  $f_{(X,Y)}(x,y)$  sur  $\mathbb{R}^2$  est

$$\iint_{\mathbb{R}^2} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dx dy = \left( \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx \right) \left( \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{y^2}{2}} dy \right) = 1 \times 1 = 1,$$

car les deux intégrales sont celles des densités de variables gaussiennes de paramètres  $m = 0$  et  $\sigma^2 = 1$ . Par conséquent,  $f_{(X,Y)}(x, y)$  est la densité jointe d'un couple de variables  $(X, Y)$  à valeurs dans  $\mathbb{R}^2$ . De plus, les densités marginales  $f_X$  et  $f_Y$  sont les densités de la loi normale  $\mathcal{N}(0, 1)$  :

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\frac{x^2+y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left( \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{y^2}{2}} dy \right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

et de même pour la variable  $Y$ .

Les densités jointes et marginales permettent de donner une nouvelle définition de l'indépendance de variables aléatoires continues :

**DÉFINITION 4.13.** *Soit  $(X, Y)$  un couple de variables aléatoires à densité. On dit que les variables  $X$  et  $Y$  sont indépendantes si la densité jointe  $f_{(X,Y)}$  est le produit des densités marginales  $f_X$  et  $f_Y$  :*

$$\forall (x, y) \in \mathbb{R}^2, f_{(X,Y)}(x, y) = f_X(x) f_Y(y).$$

Cette définition est nettement plus pratique à manipuler que celle donnée dans le chapitre précédent, et elle l'implique :

$$\begin{aligned} \mathbb{P}[(X, Y) \in [a, b] \times [c, d]] &= \int_{x=a}^b \int_{y=c}^d f_{(X,Y)}(x, y) dx dy = \int_{x=a}^b \int_{y=c}^d f_X(x) f_Y(y) dx dy \\ &= \left( \int_{x=a}^b f_X(x) dx \right) \left( \int_{y=c}^d f_Y(y) dy \right) = \mathbb{P}[X \in [a, b]] \mathbb{P}[Y \in [c, d]]. \end{aligned}$$

Si l'on suppose que la densité jointe  $f_{(X,Y)}$  est continue, alors on peut montrer l'équivalence des deux définitions. Dans tout ce qui suit, on utilisera la nouvelle définition d'indépendance pour les couples de variables continues.

**EXEMPLE 4.14.** Dans l'exemple 4.9, les variables  $X$  et  $Y$  ne sont pas indépendantes : par exemple,

$$f_{(X,Y)}\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{4}{5} \neq \left(17 \log\left(\frac{3}{2}\right) - 6\right)^2 = f_X\left(\frac{1}{2}\right) f_Y\left(\frac{1}{2}\right).$$

En revanche, dans l'exemple 4.12, les variables  $X$  et  $Y$  sont deux variables gaussiennes standards indépendantes.

Après ces généralités sur les lois de couples de variables aléatoires discrètes ou à densité, expliquons comment calculer l'espérance d'une fonction d'un tel couple. Si  $G : \mathbb{Z}^2 \rightarrow \mathbb{R}$  et si  $(X, Y)$  est un couple de variables aléatoires discrètes, on définit  $\mathbb{E}[G(X, Y)]$  par la formule

$$\mathbb{E}[G(X, Y)] = \sum_{(k,l) \in \mathbb{Z}^2} G(k, l) \mathbb{P}_{(X,Y)}[k, l],$$

sous réserve que cette somme soit bien sommable ( $\sum_{k,l} |G(k, l)| \mathbb{P}_{(X,Y)}[k, l] < \infty$ ). C'est l'analogie avec deux variables de la formule usuelle pour l'espérance d'une fonction d'une variable aléatoire discrète.

EXEMPLE 4.15. Calculons  $\mathbb{E}[XY]$  lorsque  $(X, Y)$  suit la loi discrète donnée par le tableau de l'exemple 4.1. La fonction  $G(k, l)$  est

$$\begin{aligned} G : \mathbb{Z}^2 &\rightarrow \mathbb{R} \\ (k, l) &\mapsto kl, \end{aligned}$$

et son espérance est

$$\begin{aligned} \mathbb{E}[XY] &= 0 \times (1 \times 1) + \frac{1}{12} \times (1 \times 2) + \frac{1}{6} \times (1 \times 3) \\ &\quad + \frac{1}{6} \times (2 \times 1) + \frac{1}{4} \times (2 \times 2) + \frac{1}{3} \times (2 \times 3) \\ &= 4. \end{aligned}$$

De façon similaire, si  $(X, Y)$  est un couple de variables aléatoires à densité et si  $G : \mathbb{R}^2 \rightarrow \mathbb{R}$  est une fonction *continue*, on définit  $\mathbb{E}[G(X, Y)]$  par la formule

$$\mathbb{E}[G(X, Y)] = \iint_{\mathbb{R}^2} G(x, y) f_{(X, Y)}(x, y) dx dy,$$

sous réserve que cette intégrale soit absolument convergente, c'est-à-dire que

$$\iint_{\mathbb{R}^2} |G(x, y)| f_{(X, Y)}(x, y) dx dy < \infty.$$

EXEMPLE 4.16. Calculons  $\mathbb{E}[X^2(3Y^2 + 1)]$  lorsque  $(X, Y)$  a densité jointe  $\frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$ . La fonction  $G$  est  $G(x, y) = x^2(3y^2 + 1)$ , et

$$\begin{aligned} \mathbb{E}[X^2(3Y^2 + 1)] &= \iint_{(x, y) \in \mathbb{R}^2} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} x^2(3y^2 + 1) dx dy \\ &= \left( \frac{1}{\sqrt{2\pi}} \int_{x=-\infty}^{\infty} e^{-\frac{x^2}{2}} x^2 dx \right) \left( \frac{1}{\sqrt{2\pi}} \int_{y=-\infty}^{\infty} e^{-\frac{y^2}{2}} (3y^2 + 1) dy \right) \\ &= \mathbb{E}[X^2] \mathbb{E}[3Y^2 + 1] = 1 \times 4 = 4, \end{aligned}$$

car  $X$  et  $Y$  suivent des lois gaussiennes  $\mathcal{N}(0, 1)$ , donc  $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 1$ .

L'exemple précédent rentre dans le cadre plus général de la proposition suivante :

PROPOSITION 4.17. *Soit  $(X, Y)$  un couple de variables aléatoires indépendantes discrètes ou à densité. Pour toute fonction  $G(X, Y)$  qui se factorise sous la forme  $G(X, Y) = g(X)h(Y)$ , on a*

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

DÉMONSTRATION. Traitons par exemple le cas continu :

$$\begin{aligned} \mathbb{E}[g(X)h(Y)] &= \iint_{(x, y) \in \mathbb{R}^2} f_{(X, Y)}(x, y) g(x) h(y) dx dy \\ &= \iint_{(x, y) \in \mathbb{R}^2} f_X(x) f_Y(y) g(x) h(y) dx dy \\ &= \left( \int_{x \in \mathbb{R}} f_X(x) g(x) dx \right) \left( \int_{y \in \mathbb{R}} f_Y(y) h(y) dy \right) = \mathbb{E}[g(X)] \mathbb{E}[h(Y)], \end{aligned}$$

l'exemple précédent correspondant au cas où  $g(x) = x^2$  et  $h(y) = 3y^2 + 1$ . Le cas discret est tout à fait analogue, avec des séries au lieu d'intégrales.  $\square$

Pour conclure cette section, évoquons une généralisation des calculs d'espérances précédents, qui mettra hélas en exergue certaines limites de notre présentation des variables aléatoires à densité. Souvent, la formule pour l'espérance d'une fonction continue  $G(X, Y)$  de variables aléatoires à densité peut également être employée lorsque  $G$  n'est pas une fonction continue. En particulier, on peut prendre  $G$  égale à l'indicatrice d'une partie  $A$  du plan  $\mathbb{R}^2$ . Dans ce cas,  $\mathbb{E}[1_A(X, Y)] = \mathbb{P}[(X, Y) \in A]$ , et la partie  $A$  peut être choisie essentiellement arbitrairement. Commençons par étudier deux exemples :

EXEMPLE 4.18. Soient  $X$  et  $Y$  deux variables continues indépendantes, à valeurs positives ; leur densité jointe  $f_{X,Y}(x, y) = f_X(x) f_Y(y)$  charge donc le quart de plan  $(\mathbb{R}_+)^2$ . On peut calculer la fonction de répartition du produit  $XY$  comme suit : si  $t \geq 0$ , alors

$$\begin{aligned} \mathbb{P}[XY \leq t] &= \mathbb{E}[1_{XY \leq t}] = \mathbb{E}[1_{\{(x,y) \in (\mathbb{R}_+)^2 \mid xy \leq t\}}(X, Y)] \\ &= \iint_{(\mathbb{R}_+)^2} f_{X,Y}(x, y) 1_{xy \leq t} dx dy \\ &= \int_{x=0}^{\infty} f_X(x) \left( \int_0^{\frac{t}{x}} f_Y(y) dy \right) dx \\ &= \int_{x=0}^{\infty} f_X(x) F_Y\left(\frac{t}{x}\right) dx \end{aligned}$$

Ainsi,  $F_{XY}(t) = \int_{x=0}^{\infty} f_X(x) F_Y\left(\frac{t}{x}\right) dx$ . On en déduit en dérivant par rapport à  $t$  la valeur de la densité de  $Z = XY$  :

$$f_Z(t) = \int_{x=0}^{\infty} \frac{f_X(x) f_Y\left(\frac{t}{x}\right)}{x} dx.$$

Avec la nouvelle définition d'indépendance basée sur les densités jointes et marginales, la formule pour  $F_{XY}(t)$  s'obtient beaucoup plus naturellement que dans le chapitre précédent.

EXEMPLE 4.19. Soit  $A = D_{(0,R)}$  le disque de centre l'origine et de rayon  $R$  dans  $\mathbb{R}^2$ , et  $(X, Y)$  un couple de variables à densité suivant la loi gaussienne bivariée de l'exemple 4.12. On a :

$$\begin{aligned} \mathbb{P}[(X, Y) \in A] &= \frac{1}{2\pi} \iint_{\mathbb{R}^2} 1_A(x, y) e^{-\frac{x^2+y^2}{2}} dx dy \\ &= \frac{1}{2\pi} \int_{r=0}^{\infty} 1_{r \leq R} \int_{\theta=0}^{2\pi} e^{-\frac{r^2}{2}} r dr d\theta \\ &= \int_{r=0}^R r e^{-\frac{r^2}{2}} dr = \left[ -e^{-\frac{r^2}{2}} \right]_{r=0}^R = 1 - e^{-\frac{R^2}{2}}. \end{aligned}$$

On en déduit en dérivant par rapport à  $R$  que la variable aléatoire continue  $\sqrt{X^2 + Y^2} = Z$  a pour densité  $f_Z(z) = z e^{-\frac{z^2}{2}} 1_{z \geq 0}$ .

Détaillons maintenant le cadre de validité des calculs d'espérances précédents. La formule qui exprime  $\mathbb{P}[(X, Y) \in A]$  comme intégrale double  $\iint 1_A(x, y) f_{(X, Y)}(x, y) dx dy$  est vraie pour toute partie  $A$  qui peut être "approchée" par une réunion de rectangles. C'est par exemple le cas de l'intérieur d'une courbe continue fermée :

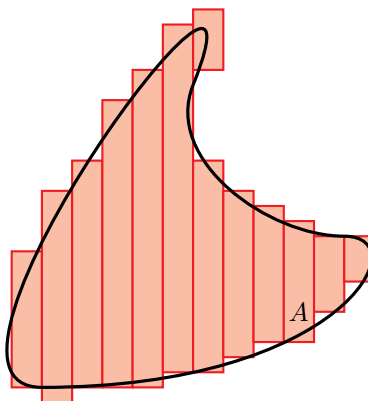


FIGURE 4. Partie  $A \subset \mathbb{R}^2$  qui peut être approchée par une réunion de rectangles.

Malheureusement, il existe aussi des parties  $A$  de  $\mathbb{R}^2$ , telles que l'intégrale double

$$\iint 1_A(x, y) f_{(X, Y)}(x, y) dx dy$$

ne soit pas correctement définie. Nous n'expliquerons pas exactement quelles parties sont exclues, car il faudrait pour cela introduire toute la *théorie de la mesure* ; le lecteur peut néanmoins deviner que si  $A$  est une partie "fractale" avec un bord très irrégulier, alors l'intégrale double ne peut pas être calculée. Heureusement, de telles parties sont relativement difficiles à construire, et dans tous nos exemples on ne rencontrera pas ces parties exceptionnelles.

## 2. Convolution de deux lois

Dans le chapitre précédent, on a rencontré lors de l'étude des variables de loi  $\chi^2(k)$  des sommes de variables positives indépendantes, dont la densité était calculée à l'aide de la formule du produit de convolution. Ce paragraphe redétaille cette formule, en traitant également le cas de variables pouvant prendre des valeurs positives et négatives, et le cas des variables discrètes.

Commençons par le cas des variables discrètes. Soit  $(X, Y)$  un couple de variables discrètes indépendantes, et  $S = X + Y$  leur somme. La loi de  $S$  est calculée en décomposant



en fonction des valeurs possibles de  $X$  :

$$\begin{aligned}\mathbb{P}[S = k] &= \sum_{l \in \mathbb{Z}} \mathbb{P}[S = k \text{ et } X = l] = \sum_{l \in \mathbb{Z}} \mathbb{P}[X + Y = k \text{ et } X = l] \\ &= \sum_{l \in \mathbb{Z}} \mathbb{P}[Y = k - l \text{ et } X = l] \\ &= \sum_{l \in \mathbb{Z}} \mathbb{P}[X = l] \mathbb{P}[Y = k - l].\end{aligned}$$

Notons que l'indépendance entre  $X$  et  $Y$  est seulement utilisée à la dernière ligne du calcul. On obtient la **formule du produit de convolution** pour des lois discrètes.

Un cas particulier est celui où  $X$  et  $Y$  sont deux variables positives, c'est-à-dire à valeurs dans  $\mathbb{N}$ . Dans ce cas, la somme précédente est restreinte à  $l \in \mathbb{N}$ , puisque si  $l < 0$ , alors  $\mathbb{P}[X = l] = 0$ . D'autre part, pour que  $\mathbb{P}[Y = k - l]$  soit non nulle, il faut aussi dans ce cas  $k - l \geq 0$ , donc  $l \leq k$ . Ainsi :

**PROPOSITION 4.20.** *Soit  $X$  et  $Y$  deux variables aléatoires discrètes indépendantes, à valeurs positives dans  $\mathbb{N}$ , de lois  $\mathbb{P}_X$  et  $\mathbb{P}_Y$ . La loi de  $S = X + Y$  est*

$$\mathbb{P}_S[k] = \sum_{l=0}^k \mathbb{P}_X[l] \mathbb{P}_Y[k-l] \quad \text{pour tout } k \geq 0.$$

**EXEMPLE 4.21.** Soit  $X$  et  $Y$  deux variables indépendantes suivant des lois de Poisson  $\mathcal{P}(\lambda)$  et  $\mathcal{P}(\mu)$ . La loi de  $S = X + Y$  est calculée comme suit : pour  $k \geq 0$ ,

$$\mathbb{P}_S[k] = \sum_{l=0}^k e^{-\lambda} \frac{\lambda^l}{l!} e^{-\mu} \frac{\mu^{k-l}}{(k-l)!} = \frac{e^{-\lambda-\mu}}{k!} \sum_{l=0}^k \frac{k!}{l!(k-l)!} \lambda^l \mu^{k-l} = e^{-\lambda-\mu} \frac{(\lambda + \mu)^k}{k!}$$

en utilisant à la fin la formule du binôme de Newton. On obtient donc pour  $S$  une loi de Poisson de paramètre  $\lambda + \mu$ .

**EXEMPLE 4.22.** Soit  $X$  et  $Y$  deux variables indépendantes suivant des lois binomiales  $\mathcal{B}(m, p)$  et  $\mathcal{B}(n, p)$ , avec le même paramètre  $p$  pour les deux variables. La loi de  $S = X + Y$  porte sur  $\llbracket 0, m + n \rrbracket$ , et elle est donnée par la formule

$$\begin{aligned}\mathbb{P}_S[k] &= \sum_{l=0}^k \binom{m}{l} p^l (1-p)^{m-l} \binom{n}{k-l} p^{k-l} (1-p)^{n-k+l} \\ &= p^k (1-p)^{m+n-k} \sum_{l=0}^k \binom{m}{l} \binom{n}{k-l}.\end{aligned}$$

La somme de produits de coefficients binomiaux peut être simplifiée à l'aide de la **formule de Vandermonde** (voir les exercices du chapitre sur les variables aléatoires discrètes) :

$$\sum_{l=0}^k \binom{m}{l} \binom{n}{k-l} = \binom{m+n}{k}.$$

Par conséquent,

$$\mathbb{P}_S[k] = \binom{m+n}{k} p^k (1-p)^{m+n-k},$$

donc  $S$  suit une loi binomiale de paramètres  $(m+n, p)$ .

Les mêmes calculs peuvent être menés avec des variables indépendantes à densité. Soit  $(X, Y)$  un couple de variables continues indépendantes, de densité jointe  $f_{(X,Y)}(x, y) = f_X(x) f_Y(y)$ . Pour tout  $t \in \mathbb{R}$ ,

$$\begin{aligned} F_{X+Y}(t) &= \mathbb{E}[1_{X+Y \leq t}] = \iint_{\mathbb{R}^2} 1_{x+y \leq t} f_X(x) f_Y(y) dx dy \\ &= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{t-x} f_X(x) f_Y(y) dx dy \\ &= \int_{x=-\infty}^{\infty} f_X(x) F_Y(t-x) dx. \end{aligned}$$

En dérivant par rapport à  $t$ , on en déduit la densité de  $S = X + Y$  :

$$f_{X+Y}(t) = \int_{\mathbb{R}} f_X(x) f_Y(t-x) dx.$$

C'est la formule du produit de convolution pour des variables continues. Si  $X$  et  $Y$  sont des variables positives, on peut effectuer la même réduction que dans le cas discret, et ainsi :

**PROPOSITION 4.23.** *Soit  $X$  et  $Y$  deux variables aléatoires continues indépendantes, à valeurs positives dans  $\mathbb{R}_+$ , de densités  $f_X$  et  $f_Y$ . La densité de  $S = X + Y$  est*

$$f_S(t) = \int_{x=0}^t f_X(x) f_Y(t-x) dx \quad \text{pour tout } t \geq 0.$$

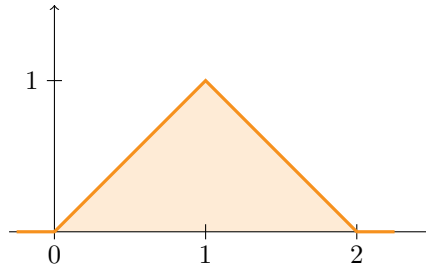


FIGURE 5. Densité de la somme de deux variables indépendantes de loi  $\mathcal{U}([0, 1])$ .

**EXEMPLE 4.24.** Soient  $X$  et  $Y$  deux variables indépendantes de loi uniforme sur  $[0, 1]$ . La somme  $S = X + Y$  a pour densité  $f_S(t) = \int_{x=0}^t f_X(x) f_Y(t-x) dx$ . Comme  $f_X(x)$  s'annule en dehors de  $[0, 1]$ , l'intégrale peut être restreinte à l'intervalle  $[0, \min(1, t)]$  :

$$f_S(t) = \int_{x=0}^{\min(1,t)} f_Y(t-x) dx.$$

Comme la densité  $f_Y$  est nulle en dehors de  $[0, 1]$ , on peut restreindre encore l'intervalle d'intégration : il faut  $t - x \leq 1$ , donc  $x \geq t - 1$  pour avoir un terme non nul, et ainsi,

$$f_S(t) = \int_{x=\max(0, t-1)}^{\min(1, t)} 1 \, dx = \min(1, t) - \max(0, t-1).$$

La fonction obtenue peut être réécrite comme suit : elle vaut  $t$  sur  $[0, 1]$ ,  $2 - t$  sur  $[1, 2]$ , et 0 partout ailleurs.

### 3. Loi des grands nombres

L'objectif des deux dernières sections de ce chapitre est de présenter deux résultats *asymptotiques* concernant les suites de variables aléatoires indépendantes. Plus précisément, soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires indépendantes, continues ou discrètes, et qui ont toutes la même loi

$$\mathbb{P}_{X_1}[k] = \mathbb{P}_{X_2}[k] = \dots = \mathbb{P}_{X_n}[k] = \dots$$

dans le cas discret, et toutes la même densité

$$f_{X_1}(x) = f_{X_2}(x) = \dots = f_{X_n}(x) = \dots$$

dans le cas continu. On supposera dans ce paragraphe que chaque  $X_i$  admet une espérance finie :  $\forall i, \mathbb{E}[|X_i|] < +\infty$ . Dans ce cas,

$$\mathbb{E}[X_1] = \mathbb{E}[X_2] = \dots = \mathbb{E}[X_n] = \dots$$

compte tenu des hypothèses précédentes.

**THÉORÈME 4.25.** *Dans le contexte précédent (suite de variables aléatoires indépendantes de même loi), on pose*

$$M_n = \frac{X_1 + \dots + X_n}{n},$$

qui est la moyenne empirique des  $X_i$ . On a la convergence presque sûre :

$$M_n \xrightarrow{\text{presque sûrement}} \mathbb{E}[X_1].$$

Ceci signifie la chose suivante : pour tout  $\varepsilon > 0$ , il existe un rang  $n_0$  tel que

$$\mathbb{P}[\forall n \geq n_0, |M_n - \mathbb{E}[X_1]| \leq \varepsilon] \geq 1 - \varepsilon.$$

On démontrera ce résultat dans le cas particulier où le quatrième moment

$$\mathbb{E}[(X_i)^4] = \begin{cases} \sum_{k \in \mathbb{Z}} k^4 \mathbb{P}_{X_i}[k] & \text{dans le cas discret,} \\ \int_{\mathbb{R}} x^4 f_{X_i}(x) \, dx & \text{dans le cas continu} \end{cases}$$

est fini, bien qu'il soit valable en toute généralité (la preuve est alors sensiblement plus difficile). On commence par établir :

LEMME 4.26. Soient  $X$  et  $Y$  deux variables continues ou discrètes, telles que  $\mathbb{E}[|X|^4] < \infty$  et  $\mathbb{E}[|Y|^4] < \infty$ . On introduit la quantité

$$\kappa^{(4)}(X) = \mathbb{E}[X^4] - 4\mathbb{E}[X^3]\mathbb{E}[X] - 3\mathbb{E}[X^2]^2 + 12\mathbb{E}[X^2](\mathbb{E}[X])^2 - 6(\mathbb{E}[X])^4,$$

et de même pour  $\kappa^{(4)}(Y)$  et  $\kappa^{(4)}(S)$  avec  $S = X + Y$ . Si  $X$  et  $Y$  sont indépendants, alors

$$\kappa^{(4)}(X + Y) = \kappa^{(4)}(X) + \kappa^{(4)}(Y).$$

DÉMONSTRATION. Notons pour commencer que si  $\mathbb{E}[X] = 0$ , alors

$$\kappa^{(4)}(X) = \mathbb{E}[X^4] - 3(\mathbb{E}[X^2])^2,$$

tous les autres termes s'annulant dans la formule donnant  $\kappa^{(4)}(X)$ . Par conséquent, dans tous les cas,  $\kappa^{(4)}(X) = \kappa^{(4)}(\tilde{X})$ , où  $\tilde{X} = X - \mathbb{E}[X]$ . En effet, par la formule du binôme de Newton, en utilisant le fait que  $\mathbb{E}[\tilde{X}] = 0$  :

$$\begin{aligned} \kappa^{(4)}(\tilde{X}) &= \mathbb{E}[(X - \mathbb{E}[X])^4] - 3\mathbb{E}[(X - \mathbb{E}[X])^2]^2 \\ &= (\mathbb{E}[X^4] - 4\mathbb{E}[X^3]\mathbb{E}[X] + 6\mathbb{E}[X^2](\mathbb{E}[X])^2 - 3(\mathbb{E}[X])^4) \\ &\quad - (3(\mathbb{E}[X^2])^2 - 6\mathbb{E}[X^2](\mathbb{E}[X])^2 + 3(\mathbb{E}[X])^4) \\ &= \mathbb{E}[X^4] - 4\mathbb{E}[X^3]\mathbb{E}[X] - 3(\mathbb{E}[X^2])^2 + 12\mathbb{E}[X^2](\mathbb{E}[X])^2 - 6(\mathbb{E}[X])^4 \\ &= \kappa^{(4)}(X). \end{aligned}$$

Quitte à remplacer  $X$  et  $Y$  par  $\tilde{X}$  et  $\tilde{Y}$ , il suffit donc de montrer le résultat du lemme lorsque  $\mathbb{E}[X] = \mathbb{E}[Y] = \mathbb{E}[X + Y] = 0$ . Dans ce cas,

$$\begin{aligned} \kappa^{(4)}(X + Y) &= \mathbb{E}[(X + Y)^4] - 3(\mathbb{E}[(X + Y)^2])^2 \\ &= \mathbb{E}[X^4] + 6\mathbb{E}[X^2]\mathbb{E}[Y^2] + \mathbb{E}[Y^4] - 3(\mathbb{E}[X^2])^2 - 3(\mathbb{E}[Y^2])^2 - 6\mathbb{E}[X^2]\mathbb{E}[Y^2] \\ &= \kappa^{(4)}(X) + \kappa^{(4)}(Y). \end{aligned}$$

□

PREUVE DU THÉORÈME 4.25. Calculons sous l'hypothèse de quatrième moment la quantité  $\mathbb{E}[(M_n - \mathbb{E}[X_1])^4]$ . On note  $\tilde{X}_i = X_i - \mathbb{E}[X_i] = X_i - \mathbb{E}[X_1]$ , de sorte que

$$M_n - \mathbb{E}[X_1] = \frac{1}{n} \sum_{i=1}^n \tilde{X}_i.$$

Notons que chaque variable  $\tilde{X}_i$  est d'espérance nulle, et il en va de même pour  $M_n - \mathbb{E}[X_1]$ . On a alors, d'après le lemme précédent,

$$\kappa^{(4)}(M_n) = \kappa^{(4)}(M_n - \mathbb{E}[X_1]) = \sum_{i=1}^n \kappa^{(4)}\left(\frac{\tilde{X}_i}{n}\right) = \frac{1}{n^4} \sum_{i=1}^n \kappa^{(4)}(\tilde{X}_i) = \frac{\kappa^{(4)}(X_1)}{n^3}.$$

Par conséquent,

$$\begin{aligned} \mathbb{E}[(M_n - \mathbb{E}[X_1])^4] &= \kappa^{(4)}(M_n - \mathbb{E}[X_1]) + 3\mathbb{E}[(M_n - \mathbb{E}[X_1])^2]^2 \\ &= \frac{\kappa^{(4)}(X_1)}{n^3} + 3(\text{var}(M_n))^2 = \frac{\kappa^{(4)}(X_1)}{n^3} + 3\frac{(\text{var}(X_1))^2}{n^2}, \end{aligned}$$

car  $\text{var}(M_n) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) = \frac{1}{n} \text{var}(X_1)$ . On en déduit que :

$$\begin{aligned} \mathbb{P}[|M_n - \mathbb{E}[X_1]| \geq \varepsilon] &= \mathbb{E}[1_{|M_n - \mathbb{E}[X_1]| \geq \varepsilon}] = \frac{\mathbb{E}[1_{|M_n - \mathbb{E}[X_1]| \geq \varepsilon} (M_n - \mathbb{E}[X_1])^4]}{\varepsilon^4} \\ &\leq \frac{\mathbb{E}[(M_n - \mathbb{E}[X_1])^4]}{\varepsilon^4} \leq \frac{1}{\varepsilon^4} \left( \frac{\kappa^{(4)}(X_1)}{n^3} + 3 \frac{(\text{var}(X_1))^2}{n^2} \right) \\ &\leq \frac{1}{n^2 \varepsilon^4} (\kappa^{(4)}(X_1) + 3 (\text{var}(X_1))^2). \end{aligned}$$

Comme  $\mathbb{E}[(\widetilde{X}_1)^4] = \kappa^{(4)}(\widetilde{X}_1) + 3 (\text{var}(\widetilde{X}_1))^2 = \kappa^{(4)}(X_1) + 3 (\text{var}(X_1))^2$ , on obtient finalement :

$$\forall \varepsilon > 0, \forall n \in \mathbb{N}^*, \quad \mathbb{P}[|M_n - \mathbb{E}[X_1]| \geq \varepsilon] \leq \frac{\mathbb{E}[(X_1 - \mathbb{E}[X_1])^4]}{n^2 \varepsilon^4}.$$

Il s'ensuit :

$$\begin{aligned} \mathbb{P}[\exists n \geq n_0, |M_n - \mathbb{E}[X_1]| \geq \varepsilon] &= \mathbb{P}\left[ \bigcup_{n=n_0}^{\infty} \{|M_n - \mathbb{E}[X_1]| \geq \varepsilon\} \right] \\ &\leq \frac{\mathbb{E}[(X_1 - \mathbb{E}[X_1])^4]}{\varepsilon^4} \left( \sum_{n=n_0}^{\infty} \frac{1}{n^2} \right) \\ &\leq \frac{\mathbb{E}[(X_1 - \mathbb{E}[X_1])^4]}{\varepsilon^4 (n_0 - 1)} \leq \varepsilon \end{aligned}$$

si  $n_0$  est pris assez grand. Ceci achève la preuve du théorème sous l'hypothèse d'existence de moments d'ordre 4 (on peut même donner une valeur pour  $n_0$  en fonction de  $\varepsilon$  et de  $\mathbb{E}[(X_1 - \mathbb{E}[X_1])^4]$ .  $\square$

REMARQUE 4.27. Attention, les hypothèses de la loi des grands nombres sont très fortes : il faut à la fois que les variables  $X_n$  soient toutes indépendantes, et qu'elles aient toutes la même loi. Il existe néanmoins des versions du théorème 4.25 avec des hypothèses plus faibles, par exemple lorsque les  $X_n$  sont indépendantes mais pas toutes de même loi, ou lorsque les variables  $X_n$  sont toutes de même loi mais avec une certaine dépendance entre elles. On peut remarquer que la preuve que l'on a donnée de la loi des grands nombres reste valable sans l'hypothèse de loi identique, et avec uniquement l'hypothèse suivante : les variables  $X_n$  sont indépendantes et ont les mêmes moments  $\mathbb{E}[(X_n)^k] = \mathbb{E}[(X_1)^k]$  pour tout  $k \in \{1, 2, 3, 4\}$ .

REMARQUE 4.28. Le théorème 4.25 est ce qu'on appelle la loi *forte* des grands nombres. Il existe également une version *faible* dont l'énoncé est le suivant : sous les mêmes hypothèses que pour la loi forte des grands nombres,

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbb{P}[|M_n - \mathbb{E}[X_1]| \geq \varepsilon] = 0.$$

On renvoie aux exercices pour une preuve très simple de cet énoncé, qui donne une information sur la probabilité pour que la moyenne empirique  $M_n$  soit éloignée de  $\mathbb{E}[X_1]$ , mais qui ne dit pas que la suite aléatoire  $(M_n)_{n \in \mathbb{N}}$  converge toujours vers  $\mathbb{E}[X_1]$ . La différence entre ces deux types d'énoncés est également illustrée dans les exercices.

REMARQUE 4.29. La loi des grands nombres justifie que l'espérance donnée par une somme  $\sum_{k \in \mathbb{Z}} \mathbb{P}_X[k] k$  ou par une intégrale  $\int_{\mathbb{R}} f_X(x) x dx$  soit la bonne définition mathématique de la notion intuitive de moyenne. En effet, elle affirme que la moyenne *empirique* d'une série d'expériences indépendantes répétant la loi de  $X$  s'approche toujours de la moyenne *théorique*  $\mathbb{E}[X]$ .

EXEMPLE 4.30. Soit  $(X_n)_{n \in \mathbb{N}}$  une suite d'expériences de Bernoulli indépendantes et de même paramètre  $p$ , et  $S_n = X_1 + X_2 + \dots + X_n$ , qui suit donc une loi binomiale  $\mathcal{B}(n, p)$ . La loi des grands nombres assure que

$$M_n = \frac{S_n}{n} \xrightarrow{n \rightarrow \infty} p$$

presque sûrement. Ce résultat justifie la méthode des sondages. Considérons en effet une très grande population de taille  $N$  (disons  $N = 60\,000\,000$ ), dont  $pN$  individus ont une opinion positive au sujet d'une question, et  $(1-p)N$  ont une opinion négative. On souhaite savoir si  $p \geq \frac{1}{2}$ . Pour ce faire, on tire au hasard  $n$  personnes parmi les  $N$ , avec  $n$  beaucoup plus petit que  $N$ , mais néanmoins assez grand, disons  $n = 5000$ ; et on leur demande à chacun leur opinion, en notant  $X_1, \dots, X_n$  les résultats. Chaque variable  $X_i$  a la loi de Bernoulli  $\mathcal{B}(p)$ , et si l'on fait suffisamment attention à la façon dont on tire au hasard les personnes, alors on peut supposer que les variables  $X_i$  sont indépendantes. Dans ce cas, d'après ce qui précède, on est sûr que  $\frac{X_1 + \dots + X_n}{n} \rightarrow p$ , ce qui permet de décider si  $p \geq \frac{1}{2}$  ou non.

Le problème est alors de savoir quelle valeur de  $n$  prendre. Supposons qu'on veuille estimer  $p$  à 2% près, avec une probabilité pour se tromper dans l'estimation également inférieure à 2%. Autrement dit, on veut une variable aléatoire  $M_n$  telle que

$$\mathbb{P} \left[ |M_n - p| \geq \frac{2}{100} \right] \leq \frac{2}{100}.$$

Dans la preuve de la loi forte des grands nombres, on a montré l'inégalité :

$$\mathbb{P}[|M_n - p| \geq \varepsilon] \leq \frac{\mathbb{E}[(X_1 - p)^4]}{n^2 \varepsilon^4}.$$

Pour une variable de Bernoulli,  $\mathbb{E}[(X_1 - p)^4] = p(1-p)^4 + (1-p)p^4$ , qui atteint sa valeur maximale sur le segment  $[0, 1]$  lorsque  $6p^2 - 6p + 1 = 0$ , c'est-à-dire lorsque  $p = p_{\pm} = \frac{1}{2} \pm \frac{1}{\sqrt{12}}$ . Cette valeur maximale est donc  $p_+(1-p_+)^4 + (1-p_+)p_+^4 = \frac{1}{12}$ , et par conséquent,

$$\mathbb{P}[|M_n - p| \geq \varepsilon] \leq \frac{1}{12 n^2 \varepsilon^4}$$

quelque soit la valeur de  $p$ . Avec  $\varepsilon = \frac{2}{100}$ , pour que le terme de droite soit inférieur à  $\frac{2}{100}$ , il faut que

$$n \geq \frac{1}{\sqrt{12}} \frac{1}{(2\%)^{5/2}}.$$

On obtient comme valeur numérique pour le terme de droite de la dernière inégalité  $n_0 = 5103$  : il est donc bien suffisant d'interroger environ  $n = 5000$  personnes, ce qui est un grand nombre, mais nettement moins grand que  $N = 60\,000\,000$ .

#### 4. Théorème central limite

Considérons toujours une suite  $(X_n)_{n \in \mathbb{N}}$  de variables aléatoires discrètes ou continues, qui sont indépendantes et de même loi. On a vu dans le paragraphe précédent que lorsque  $n$  grandit, la moyenne  $M_n = \frac{X_1 + \dots + X_n}{n}$  de ces variables s'approche avec très grande probabilité de la moyenne théorique  $\mathbb{E}[X_1]$ . Une question naturelle est alors de savoir quelle est la taille attendue pour la différence  $M_n - \mathbb{E}[X_1]$  : on sait qu'elle tend presque sûrement vers 0, et on souhaiterait savoir à quelle vitesse.

Le résultat que nous allons démontrer est le suivant :  $M_n - \mathbb{E}[X_1]$  est de l'ordre de  $\sqrt{n}$ , et la variable renormalisée  $\sqrt{n}(M_n - \mathbb{E}[X_1])$  a une loi qui s'approche d'une limite universelle, ne dépendant pas de la loi de  $X_1$  (ou plus précisément, ne dépendant que de la valeur de la variance  $\text{var}(X_1)$ ). Avant de pouvoir énoncer le résultat, il nous faut expliquer ce que l'on entend par la convergence de la loi d'une suite de variables aléatoires.

**DÉFINITION 4.31.** *Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires à valeurs réelles, ces variables pouvant être discrètes ou continues. On dit que  $(X_n)_{n \in \mathbb{N}}$  converge en loi s'il existe une fonction croissante  $F : \mathbb{R} \rightarrow [0, 1]$  telle que  $\lim_{x \rightarrow -\infty} F(x) = 0$  et  $\lim_{x \rightarrow +\infty} F(x) = 1$ , et telle, que pour tous réels  $a$  et  $b$  tels que  $F$  soit continue en  $a$  et en  $b$ ,*

$$\lim_{n \rightarrow \infty} \mathbb{P}[X_n \in [a, b]] = F(b) - F(a).$$

Cette définition a l'avantage de s'adapter simultanément aux cas où la limite est discrète ou continue. Plus précisément, détaillons deux cas particuliers :

- (1) Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables à valeurs dans  $\mathbb{Z}$ , et  $X$  une autre variable aléatoire à valeurs dans  $\mathbb{Z}$ . On note  $\mathbb{P}_X$  la loi de  $X$  ; elle est entièrement déterminée par les valeurs  $\mathbb{P}_X[k]$  avec  $k \in \mathbb{Z}$ , et aussi par les valeurs

$$F_X(k) = \sum_{l=-\infty}^k \mathbb{P}_X[l], \quad k \in \mathbb{Z},$$

qui déterminent une fonction croissante. Supposons que  $\lim_{n \rightarrow \infty} \mathbb{P}_{X_n}[k] = \mathbb{P}_X[k]$  pour tout entier  $k$ . Alors,  $(X_n)_{n \in \mathbb{N}}$  converge en loi avec pour limite  $F = F_X$ . En effet, soit  $a$  et  $b$  des points de continuité de  $F_X$  : ce sont donc des points qui ne sont pas entiers, ou tels que  $\mathbb{P}_X[a] = 0$  ou  $\mathbb{P}_X[b] = 0$ . Pour simplifier la présentation, on supposera que  $\mathbb{P}_X$  charge tout  $\mathbb{Z}$ , de sorte que  $a$  et  $b$  sont nécessairement dans  $\mathbb{R} \setminus \mathbb{Z}$ . Alors,

$$\mathbb{P}[a \leq X_n \leq b] = \sum_{k=\lceil a \rceil}^{\lfloor b \rfloor} \mathbb{P}_{X_n}[k] \rightarrow \sum_{k=\lceil a \rceil}^{\lfloor b \rfloor} \mathbb{P}_X[k] = F_X(b) - F_X(a),$$

d'où le résultat. La réciproque est aisée, de sorte qu'on a le résultat suivant : une suite de variables  $(X_n)_{n \in \mathbb{N}}$  à valeurs dans  $\mathbb{Z}$  converge en loi vers (la fonction de répartition de)  $X$  variable à valeurs dans  $\mathbb{Z}$  si et seulement si

$$\lim_{n \rightarrow \infty} \mathbb{P}_{X_n}[k] = \mathbb{P}_X[k]$$

pour tout  $k \in \mathbb{Z}$ .

- (2) Attention, une suite de variables à *valeurs discrètes* dans une partie dénombrable de  $\mathbb{R}$  peut converger en loi vers (la fonction de répartition d')une variable aléatoire à *densité*; on verra dans un instant des exemples de ce phénomène. Considérons en toute généralité une variable  $X$  à densité, dont la fonction de répartition  $F_X$  est donc continue croissante de  $\mathbb{R}$  vers  $[0, 1]$ . Dans ce cas, la condition sur les points de continuité est tout le temps vérifiée, et une suite de variables discrètes ou continues  $(X_n)_{n \in \mathbb{N}}$  converge en loi vers  $X$  si et seulement si, pour tous réels  $a$  et  $b$ ,  $\mathbb{P}[X_n \in [a, b]] \rightarrow F_X(b) - F_X(a)$ . Comme  $[a, b] = (-\infty, b] \setminus (-\infty, a]$ , cet énoncé est équivalent à :

$$\forall t \in \mathbb{R}, \quad \lim_{n \rightarrow \infty} \mathbb{P}[X_n \leq t] = \mathbb{P}[X \leq t].$$

EXEMPLE 4.32. Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires de lois respectives  $\mathcal{B}(p_n)$ , où  $(p_n)_{n \in \mathbb{N}}$  est une suite à valeurs dans  $[0, 1]$ . Si  $\lim_{n \rightarrow \infty} p_n = p$ , alors  $(X_n)_{n \in \mathbb{N}}$  converge en loi vers une variable aléatoire de loi  $\mathcal{B}(p)$ .

EXEMPLE 4.33. Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires indépendantes et de même loi, et  $(M_n = \frac{X_1 + \dots + X_n}{n})_{n \in \mathbb{N}}$  la suite de leurs moyennes empiriques. Pour tout  $\varepsilon > 0$ , par la loi des grands nombres (sous sa forme faible),

$$\lim_{n \rightarrow \infty} \mathbb{P}[M_n \leq \mathbb{E}[X_1] - \varepsilon] = 0 \quad ; \quad \lim_{n \rightarrow \infty} \mathbb{P}[M_n \leq \mathbb{E}[X_1] + \varepsilon] = 1.$$

Par conséquent,  $(M_n)_{n \in \mathbb{N}}$  converge en loi vers la constante déterministe  $m = \mathbb{E}[X_1]$ , qui en tant que variable aléatoire a pour fonction de répartition

$$F_m(x) = \begin{cases} 0 & \text{si } x < m, \\ 1 & \text{si } x \geq m. \end{cases}$$

En effet, on a vérifié la définition de la convergence en loi en tout point  $x \neq m$ , c'est-à-dire en tout point de continuité de  $F_m$ . Attention, ce résultat de convergence en loi est moins fort que le théorème 4.25.

On peut maintenant énoncer le résultat principal de cette section, connu sous le nom de **théorème central limite** :

THÉORÈME 4.34. Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires indépendantes et de même loi, telles que  $\mathbb{E}[(X_1)^2] < +\infty$ , de sorte que  $\text{var}(X_1)$  soit bien définie. On suppose également  $\text{var}(X_1) \neq 0$ , c'est-à-dire que les variables  $X_i$  ne sont pas des constantes. On pose

$$Y_n = \sqrt{\frac{n}{\text{var}(X_1)}} (M_n - \mathbb{E}[X_1]) = \frac{1}{\sqrt{n \text{var}(X_1)}} \sum_{k=1}^n (X_k - \mathbb{E}[X_k]).$$

La suite de variables aléatoires  $(Y_n)_{n \in \mathbb{N}}$  converge en loi vers une variable gaussienne standard de loi  $\mathcal{N}(0, 1)$ . Autrement dit, pour tout  $t \in \mathbb{R}$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[ \sqrt{\frac{n}{\text{var}(X_1)}} (M_n - \mathbb{E}[X_1]) \leq t \right] = \frac{1}{\sqrt{2\pi}} \int_{x=-\infty}^t e^{-\frac{x^2}{2}} dx.$$



REMARQUE 4.35. Le théorème 4.34 est une généralisation importante du résultat du chapitre précédent sur les limites de lois binomiales de paramètre  $\frac{1}{2}$ , qui nous avait permis d'introduire les lois gaussiennes.

REMARQUE 4.36. Fixons deux réels positifs  $a$  et  $b$ , et considérons la probabilité pour que

$$\frac{a}{\sqrt{n}} \leq \left| \frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \right| \leq \frac{b}{\sqrt{n}}.$$

Le théorème central limite assure que la probabilité de cet événement tend vers une valeur strictement positive, à savoir,

$$\frac{1}{\sqrt{2\pi}} \int_{-\frac{b}{\sqrt{\text{var}(X_1)}}}^{-\frac{a}{\sqrt{\text{var}(X_1)}}} e^{-\frac{x^2}{2}} dx + \frac{1}{\sqrt{2\pi}} \int_{\frac{a}{\sqrt{\text{var}(X_1)}}}^{\frac{b}{\sqrt{\text{var}(X_1)}}} e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{2\pi}} \int_{\frac{a}{\sqrt{\text{var}(X_1)}}}^{\frac{b}{\sqrt{\text{var}(X_1)}}} e^{-\frac{x^2}{2}} dx.$$

Ce résultat confirme l'affirmation donnée au début du paragraphe : la taille de  $M_n - \mathbb{E}[X_1]$  est bien avec grande probabilité de l'ordre de  $\sqrt{n}$ .

Pour démontrer le théorème central limite, on utilisera le résultat suivant, que l'on admettra :

LEMME 4.37 (Théorème de continuité de Lévy). *Soit  $(X_n)_{n \in \mathbb{N}}$  et  $X$  des variables aléatoires réelles, discrètes ou continues. Les variables  $X_n$  convergent en loi vers  $X$  si et seulement si, pour tout  $\xi \in \mathbb{R}$ ,*

$$\lim_{n \rightarrow \infty} \mathbb{E}[e^{i\xi X_n}] = \mathbb{E}[e^{i\xi X}].$$

Dans ce lemme, on manipule des espérances de variables aléatoires complexes  $Y = e^{iX}$  : elles sont définies de la même façon que dans le cas réel, et la partie réelle de l'espérance est l'espérance de la partie réelle (et de même pour la partie imaginaire).

LEMME 4.38. *Soit  $X$  une variable aléatoire de loi  $\mathcal{N}(0, 1)$ . Pour tout  $\xi \in \mathbb{R}$ ,*

$$\mathbb{E}[e^{i\xi X}] = e^{-\frac{\xi^2}{2}}.$$

DÉMONSTRATION. Notons

$$f(\xi) = \mathbb{E}[e^{i\xi X}] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\xi x} e^{-\frac{x^2}{2}} dx.$$

Dans l'intégrale, la fonction de deux variables  $g(\xi, x)$  est infiniment dérivable, et dominée par  $e^{-\frac{x^2}{2}}$ . Sa dérivée par rapport à  $\xi$  est

$$\frac{\partial g(\xi, x)}{\partial \xi} = ix e^{i\xi x} e^{-\frac{x^2}{2}},$$

qui est dominée uniformément par la fonction intégrable  $|x| e^{-\frac{x^2}{2}}$ . Ces hypothèses permettent de dériver  $f(\xi)$  sous le signe intégral :

$$\begin{aligned} f'(\xi) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} ix e^{i\xi x} e^{-\frac{x^2}{2}} dx \\ &= \left[ -i e^{i\xi x} e^{-\frac{x^2}{2}} \right]_{-\infty}^{\infty} - \int_{\mathbb{R}} \xi e^{i\xi x} e^{-\frac{x^2}{2}} dx \\ &= -\xi f(\xi). \end{aligned}$$

On en déduit une équation différentielle vérifiée par  $f(\xi)$ , dont la solution est  $f(\xi) = K e^{-\frac{\xi^2}{2}}$ . Comme  $f(0) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx = 1$ ,  $K = 1$  et le lemme est démontré.  $\square$

LEMME 4.39. *Soit  $X$  une variable aléatoire telle que  $\mathbb{E}[X^2] < \infty$ . On pose  $h(\xi) = \log \mathbb{E}[e^{i\xi X}]$ . Cette fonction est deux fois dérivable dans un voisinage de  $\xi = 0$ , de dérivées première et seconde données par :*

$$\begin{aligned} h'(\xi) &= i \frac{\mathbb{E}[X e^{i\xi X}]}{\mathbb{E}[e^{i\xi X}]}; \\ h''(\xi) &= -\frac{\mathbb{E}[X^2 e^{i\xi X}] \mathbb{E}[e^{i\xi X}] - (\mathbb{E}[X e^{i\xi X}])^2}{\mathbb{E}[e^{i\xi X}]^2}. \end{aligned}$$

DÉMONSTRATION. Traitons par exemple le cas où  $X$  est une variable à densité continue  $f_X(x)$ , le cas discret étant tout à fait analogue avec des séries au lieu d'intégrales. On pose

$$H(\xi) = \mathbb{E}[e^{i\xi X}] = \int_{\mathbb{R}} f_X(x) e^{i\xi x} dx,$$

et dans cette intégrale, la fonction de deux variables  $g(\xi, x) = f_X(x) e^{i\xi x}$  est infiniment dérivable par rapport à  $\xi$  et continue par rapport à  $x$ . Les deux premières dérivées par rapport à  $\xi$  sont

$$\frac{\partial g(\xi, x)}{\partial \xi} = ix f_X(x) e^{i\xi x} \quad ; \quad \frac{\partial^2 g(\xi, x)}{\partial \xi^2} = -x^2 f_X(x) e^{i\xi x},$$

et elles sont dominées respectivement par

$$\left| \frac{\partial g(\xi, x)}{\partial \xi} \right| \leq x f_X(x) \quad ; \quad \left| \frac{\partial^2 g(\xi, x)}{\partial \xi^2} \right| \leq x^2 f_X(x),$$

qui sont deux fonctions intégrables car  $\mathbb{E}[|X|^2] < \infty$ . Des dominations permettent de dériver  $H(\xi)$  sous le signe intégral, et ainsi,

$$\begin{aligned} H'(\xi) &= i \int_{\mathbb{R}} x f_X(x) e^{i\xi x} dx = i \mathbb{E}[X e^{i\xi X}]; \\ H''(\xi) &= - \int_{\mathbb{R}} x^2 f_X(x) e^{i\xi x} dx = -\mathbb{E}[X^2 e^{i\xi X}]. \end{aligned}$$

Comme  $h(\xi) = \log H(\xi)$ , on en déduit que  $h$  est également deux fois dérivable sur tout intervalle où l'on peut prendre le logarithme (complexe). Par continuité de  $H$ , c'est le

cas sur un petit intervalle autour de 0, car  $H(0) = 1$ . Les formules pour  $h'(\xi)$  et  $h''(\xi)$  découlent maintenant de celles pour  $H'(\xi)$  et  $H''(\xi)$ , par composition des dérivées.  $\square$

PREUVE DU THÉORÈME 4.34. Compte tenu du lemme admis, il suffit de montrer que  $\mathbb{E}[e^{i\xi Y_n}]$  est proche de  $e^{-\frac{\xi^2}{2}}$  pour tout  $\xi$ , lorsque  $n$  tend vers l'infini. De façon équivalente, il faut montrer que

$$\log \mathbb{E}[e^{i\xi Y_n}] \rightarrow -\frac{\xi^2}{2}.$$

Or, si  $\widetilde{X}_i = X_i - \mathbb{E}[X_i]$ , alors on peut écrire :

$$\begin{aligned} \log \mathbb{E}[e^{i\xi Y_n}] &= \log \mathbb{E}\left[e^{\frac{i\xi}{\sqrt{n \operatorname{var}(X_1)}} \sum_{i=1}^n \widetilde{X}_i}\right] = \log \mathbb{E}\left[\prod_{i=1}^n e^{\frac{i\xi \widetilde{X}_i}{\sqrt{n \operatorname{var}(X_1)}}}\right] \\ &= \log \left(\prod_{i=1}^n \mathbb{E}\left[e^{\frac{i\xi \widetilde{X}_i}{\sqrt{n \operatorname{var}(X_1)}}}\right]\right) = n \log \mathbb{E}\left[e^{\frac{i\xi \widetilde{X}_1}{\sqrt{n \operatorname{var}(X_1)}}}\right], \end{aligned}$$

la seconde ligne découlant de l'indépendance des  $\widetilde{X}_i$  et du fait qu'elles ont toutes même loi. Ainsi,

$$\log \mathbb{E}[e^{i\xi Y_n}] = n h\left(\frac{\xi}{\sqrt{n \operatorname{var}(X_1)}}\right),$$

où  $h = h_{\widetilde{X}_1}$  est la fonction du lemme précédent pour la variable  $\widetilde{X}_1$ . En combinant le lemme précédent et la formule de Taylor-Lagrange à l'ordre 2 pour la fonction  $h$ , on en déduit que si  $\xi$  est fixé, alors il existe  $v \in \mathbb{R}$  avec  $|v| \leq \frac{|\xi|}{\sqrt{n \operatorname{var}(X_1)}}$  telle que,

$$\begin{aligned} \log \mathbb{E}[e^{i\xi Y_n}] &= n \left( h(0) + h'(0) \frac{\xi}{\sqrt{n \operatorname{var}(X_1)}} + \frac{h''(v)}{2} \frac{\xi^2}{n \operatorname{var}(X_1)} \right) \\ &= n \left( \mathbb{E}[\widetilde{X}_1] \frac{i\xi}{\sqrt{n \operatorname{var}(X_1)}} + \frac{h''(v)}{2} \frac{\xi^2}{n \operatorname{var}(X_1)} \right) \\ &= \frac{h''(v)}{\operatorname{var}(X_1)} \frac{\xi^2}{2} \quad \text{car } \mathbb{E}[\widetilde{X}_1] = 0. \end{aligned}$$

Lorsque  $n$  tend vers l'infini,  $\frac{|\xi|}{\sqrt{n \operatorname{var}(X_1)}}$  tend vers 0, donc  $v$  tend vers 0 et  $h''(v)$  tend vers

$$h''(0) = -\mathbb{E}[(\widetilde{X}_1)^2] = -\operatorname{var}(X_1).$$

Ainsi,  $\lim_{n \rightarrow \infty} \mathbb{E}[e^{i\xi Y_n}] = e^{-\frac{\xi^2}{2}}$ , et c'est ce que l'on voulait montrer.  $\square$

EXEMPLE 4.40. On considère de nouveau une population de  $N = 60\,000\,000$  personnes, dont on interroge  $n$  personnes au sujet d'une question pour laquelle  $pN$  personnes ont une opinion positive, et  $(1-p)N$  personnes ont une opinion négative. On suppose que les résultats  $X_1, \dots, X_n$  des questions posées aux personnes sondées sont des variables indépendantes de Bernoulli, de paramètre  $p$ . On cherche comme précédemment à estimer

$p$  à l'aide de  $M_n = \frac{X_1 + \dots + X_n}{n}$ , mais on s'autorise maintenant à utiliser le théorème central limite. La variance de  $X_1$  est  $p(1-p)$ , donc

$$\begin{aligned} \mathbb{P}[|M_n - p| \geq \varepsilon] &= \mathbb{P}\left[\sqrt{\frac{n}{p(1-p)}} |M_n - p| \geq \varepsilon \sqrt{\frac{n}{p(1-p)}}\right] \\ &\simeq_{n \rightarrow \infty} \frac{2}{\sqrt{2\pi}} \int_{\varepsilon \sqrt{\frac{n}{p(1-p)}}}^{\infty} e^{-\frac{x^2}{2}} dx. \end{aligned}$$

Comme  $M = \varepsilon \sqrt{\frac{n}{p(1-p)}}$  tend vers l'infini, il s'agit alors d'estimer la queue de distribution

$$\mathbb{P}[|X| \geq M] = \frac{2}{\sqrt{2\pi}} \int_M^{\infty} e^{-\frac{x^2}{2}} dx$$

d'une variable  $X$  gaussienne standard. On procède par intégration par parties :

$$\begin{aligned} I_M &= \int_M^{\infty} e^{-\frac{x^2}{2}} dx = \left[-\frac{1}{x} e^{-\frac{x^2}{2}}\right]_M^{\infty} - \int_M^{\infty} \frac{1}{x^2} e^{-\frac{x^2}{2}} dx \\ &= \frac{e^{-\frac{M^2}{2}}}{M} - R_M \quad \text{avec } 0 \leq R_M \leq \frac{I_M}{M^2}. \end{aligned}$$

Par conséquent,  $\frac{M I_M}{e^{-\frac{M^2}{2}}} \in \left[\frac{M^2}{1+M^2}, 1\right]$ . En particulier, pour  $n$  assez grand,

$$\mathbb{P}[|M_n - p| \geq \varepsilon] \leq \frac{2}{\sqrt{2\pi}} \frac{\sqrt{p(1-p)} e^{-\frac{n\varepsilon^2}{2p(1-p)}}}{\sqrt{n\varepsilon^2}}.$$

La fonction  $M \mapsto \frac{1}{M} e^{-\frac{M^2}{2}}$  étant décroissante en  $M$ , elle est bornée par sa valeur lorsque  $M$  est le plus petit possible, c'est-à-dire lorsque  $p(1-p)$  est le plus grand possible. C'est lorsque  $p = \frac{1}{2}$ , ce qui implique

$$\mathbb{P}[|M_n - p| \geq \varepsilon] \leq \frac{1}{\sqrt{2\pi}} \frac{e^{-2n\varepsilon^2}}{\sqrt{n\varepsilon^2}}$$

pour  $n$  assez grand. Supposons  $\varepsilon = 2\%$ . Alors, pour que le terme de droite dans l'inégalité soit aussi inférieur à  $2\%$ , il faut que  $n\varepsilon^2 \geq 1.42$ , et donc que  $n \geq n_0 = \frac{1.42}{(2\%)^2}$ . Ceci donne une valeur numérique  $n_0 = 3550$ , qui est sensiblement inférieure à la valeur  $n_0 = 5103$  précédemment calculée : ainsi, il suffit d'interroger environ 3500 personnes.

L'étude générale d'estimations semblables à celles de cet exemple sera l'un des objets d'étude principaux du dernier chapitre.

REMARQUE 4.41. Il y a une petite imprécision dans le raisonnement précédent : on remplace  $\mathbb{P}[|M_n - p| \geq \varepsilon]$  par la valeur donnée par le théorème central limite, mais ceci n'est valable que pour  $n$  très grand, et on ne sait pas a priori pour quelles valeurs  $n \geq n_0$  cette estimation est valable. Des résultats plus avancés (estimées de Berry-Esseen) permettent de résoudre ce problème.

## Exercices

- (1) **Couple de variables de Bernoulli.** Soit  $X$  et  $Y$  deux variables aléatoires de Bernoulli, dont la loi jointe est donnée par le tableau suivant :

$X \setminus Y$	0	1
0	$\frac{1}{2} - a$	$a + \frac{1}{3}$
1	$b$	$\frac{1}{6} - 2a$

- (a) Quelles valeurs sont autorisées pour  $a$  et  $b$  ?  
 (b) Calculer en fonction de  $a$  et  $b$  les lois marginales de  $X$  et de  $Y$ .  
 (c) Quelles valeurs de  $a$  et  $b$  correspondent à un couple  $(X, Y)$  de variables indépendantes ?
- (2) **Un couple de variables à densité.** Soit  $(X, Y)$  un couple de variables aléatoires à densité, avec

$$f_{(X,Y)}(x, y) = \begin{cases} C y e^{-x} & \text{si } 0 \leq y \leq x, \\ 0 & \text{sinon.} \end{cases}$$

- (a) Trouver la valeur de la constante  $C$ .  
 (b) Calculer les densités marginales de  $X$  et de  $Y$ . Quelles lois retrouve-t-on ?  
 (c) Calculer la covariance de  $X$  et  $Y$ . Les variables  $X$  et  $Y$  sont-elles indépendantes ?
- (3) **Lois conditionnelles.** Soit  $X$  et  $Y$  deux variables aléatoires à valeurs dans  $\mathbb{Z}$ . La loi conditionnelle de  $X$  sachant  $Y$  est la fonction aléatoire  $\mathbb{P}_{X|Y} : \mathbb{Z} \rightarrow [0, 1]$ , dont les valeurs sont données par

$$\mathbb{P}_{X|Y}[k] = \mathbb{P}[X = k | Y = l] = \frac{\mathbb{P}[X = k \text{ et } Y = l]}{\mathbb{P}[Y = l]} \quad \text{si } Y = l.$$

- (a) Expliciter  $\mathbb{P}_{X|Y}$  dans le cas où  $X$  et  $Y$  sont indépendantes.  
 (b) Soit  $W$  et  $X$  deux variables aléatoires indépendantes suivant des lois de Poisson de paramètres  $w$  et  $x$ . On pose  $Y = W + X$ , et on rappelle que  $Y$  suit une loi de Poisson de paramètre  $w + x$ . Calculer la loi conditionnelle  $\mathbb{P}_{X|Y}$ .
- (4) **Somme de deux loi géométriques.** Calculer la loi de la somme de deux variables aléatoires indépendantes  $X$  et  $Y$  de lois géométriques  $\mathcal{G}(p)$  et  $\mathcal{G}(p')$  (on distinguera les cas  $p = p'$  et  $p \neq p'$ ).
- (5) **Somme de deux lois gaussiennes.** Calculer la loi de la somme de deux variables aléatoires indépendantes  $X$  et  $Y$ , avec  $X \sim \mathcal{N}(m_1, \sigma_1^2)$  et  $Y \sim \mathcal{N}(m_2, \sigma_2^2)$ . On pourra se ramener au cas où  $m_1 = m_2 = 0$ .
- (6) **Loi faible des grands nombres.** Dans cet exercice,  $(X_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires indépendantes, de même loi et centrées :  $\mathbb{E}[X_n] = 0$  pour tout  $n$ . On suppose aussi  $\mathbb{E}[(X_n)^2] = \mathbb{E}[(X_1)^2] = \sigma^2 \in (0, +\infty)$  : la variance est finie et non nulle.

- (a) On pose  $M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ . Calculer  $\mathbb{E}[(M_n)^2]$ .  
 (b) Montrer que

$$\mathbb{P}[|M_n| \geq \varepsilon] \leq \frac{\sigma^2}{n\varepsilon^2},$$

et en déduire que cette probabilité tend vers 0 pour tout  $\varepsilon > 0$ .

- (c) On suppose plus généralement que les variables  $X_n$  sont indépendantes centrées, et que  $\mathbb{E}[(X_n)^2] = (\sigma_n)^2$ , les valeurs  $(\sigma_n)^2$  des variances pouvant maintenant dépendre de  $n$ . Donner une condition simple sur ces variances pour que  $\lim_{n \rightarrow \infty} \mathbb{P}[|M_n| \geq \varepsilon] = 0$  quelque soit la valeur de  $\varepsilon > 0$ .

- (7) **Cumulants d'une variable aléatoire.** Soit  $X$  une variable aléatoire dont les valeurs possibles restent bornées dans un intervalle  $[-A, A]$ . On admet que dans ce cas, on peut développer en série

$$\mathbb{E}[e^{zX}] = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{(zX)^k}{k!}\right] = \sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k]}{k!} z^k,$$

cette expansion étant valable pour tout  $z \in \mathbb{C}$ . On pose  $h_X(z) = \log \mathbb{E}[e^{zX}]$ .

- (a) Montrer que

$$h'_X(0) = \mathbb{E}[X] \quad ; \quad h''_X(0) = \text{var}(X) \quad ; \quad h'''_X(0) = \kappa^{(4)}(X).$$

- (b) En déduire une preuve simple des identités  $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$  et  $\kappa^{(4)}(X + Y) = \kappa^{(4)}(X) + \kappa^{(4)}(Y)$  lorsque  $X$  et  $Y$  sont deux variables indépendantes.

- (8) **Moyenne de variables aléatoires indépendantes mais pas de même loi.** Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de variables aléatoires indépendantes, telles que les lois des variables  $X_n$  soient discrètes et données par les formules suivantes :

$$\mathbb{P}[X_n = 0] = 1 - \frac{1}{n \log(n+1)};$$

$$\mathbb{P}[X_n = n] = \mathbb{P}[X_n = -n] = \frac{1}{2n \log(n+1)}$$

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

- (a) On pose  $M_n = \frac{X_1 + \dots + X_n}{n}$ . Calculer  $\mathbb{E}[X_n]$ ,  $\text{var}(X_n)$ ,  $\mathbb{E}[M_n]$  et  $\text{var}(M_n)$ .  
 (b) Montrer que  $\mathbb{P}[|M_n| \geq \varepsilon] \rightarrow 0$  pour tout  $\varepsilon > 0$ .  
 (c) Soit  $(A_n)_{n \in \mathbb{N}}$  une suite d'événements indépendants :

$$\mathbb{P}[A_{n_1} \cap A_{n_2} \cap \dots \cap A_{n_k}] = \prod_{i=1}^k \mathbb{P}[A_{n_i}]$$

pour tout choix d'entiers  $n_1 < n_2 < \dots < n_k$ . On s'intéresse à l'événement  $B_{n_0} = \bigcup_{n \geq n_0} A_n$ . Montrer que

$$\mathbb{P}[B_{n_0}] = 1 - \prod_{n=n_0}^{\infty} (1 - \mathbb{P}[A_n]) \geq 1 - \exp\left(-\sum_{n=n_0}^{\infty} \mathbb{P}[A_n]\right).$$

En déduire la probabilité de  $B_{n_0}$  si  $\sum_{n=1}^{\infty} \mathbb{P}[A_n] = +\infty$ .

(d) Montrer que pour tout  $n_0$  et tout  $\varepsilon < 1$ ,  $\mathbb{P}[\forall n \geq n_0, |M_n| \leq \varepsilon] = 0$ . Commenter.

(9) **Méthode de Monte Carlo.** Soit  $f$  une fonction continue sur  $[0, 1]$ . On se donne une suite de variables aléatoires  $(X_n)_{n \in \mathbb{N}}$  indépendantes et de loi uniforme sur  $[0, 1]$ . Expliquer comment calculer de façon approchée l'intégrale  $\int_0^1 f(x) dx$ , en utilisant uniquement les nombres  $f(X_1), f(X_2), \dots, f(X_n)$ .

(10) **Limites de variables de Poisson.** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires de lois de Poisson  $\mathcal{P}(\lambda_n)$ , avec  $\lambda_n \rightarrow +\infty$  et les  $\lambda_n$  entiers.

(a) Écrire  $X_n$  comme somme de  $\lambda_n$  variables aléatoires indépendantes et de même loi.

(b) Quelle est la limite en loi des variables  $Y_n$ , où

$$Y_n = \frac{X_n - \lambda_n}{\sqrt{\lambda_n}} ?$$

(c) Traiter aussi le cas où les paramètres  $\lambda_n$  ne sont plus supposés entiers, mais tendent toujours vers  $+\infty$ .

(11) **Convergence vers une loi de Poisson.** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires de lois  $X_n \sim \mathcal{B}(n, p_n)$ . On rappelle le critère sur les transformées de Fourier pour la convergence en loi :  $X_n$  converge en loi vers une variable  $X$  si et seulement si  $\mathbb{E}[e^{i\xi X_n}] \rightarrow \mathbb{E}[e^{i\xi X}]$  pour tout  $\xi \in \mathbb{R}$ .

(a) Calculer  $\mathbb{E}[e^{i\xi X_n}]$ .

(b) On suppose que  $p_n = \frac{\lambda}{n}$ . Montrer que  $(X_n)_{n \in \mathbb{N}}$  converge en loi vers une variable  $X \sim \mathcal{P}(\lambda)$ .

(c) Étendre le résultat précédent au cas où  $X_n = \sum_{i=1}^n A_{n,i}$ , avec les  $A_{n,i} \sim \mathcal{B}(p_{n,i})$  variables de Bernoulli indépendantes. On demande dans ce cas des conditions suffisantes simples sur les paramètres  $(p_{n,i})_{n \geq i \geq 1}$  pour avoir la convergence vers une loi de Poisson  $\mathcal{P}(\lambda)$ .

(12) **Distance entre deux probabilités discrètes.** Si  $X$  et  $Y$  sont deux variables discrètes à valeurs dans  $\mathbb{Z}$ , on définit leur *distance en variation totale* par la formule suivante :

$$d(X, Y) = \frac{1}{2} \sum_{k \in \mathbb{Z}} |\mathbb{P}_X[k] - \mathbb{P}_Y[k]|.$$

(a) Montrer qu'on a toujours  $d(X, Y) \leq 1$ .

(b) Si  $X \sim \mathcal{B}(p)$  et  $Y \sim \mathcal{P}(p)$ , montrer que  $d(X, Y) = p(1 - e^{-p})$ , puis que  $d(X, Y) \leq p^2$ .

(c) Soient  $(X_n)_{n \in \mathbb{N}}$  et  $X$  des variables à valeurs dans  $\mathbb{Z}$ . On suppose que  $(X_n)_{n \in \mathbb{N}}$  converge en loi vers  $X$ . Montrer que  $d(X_n, X)$  tend vers 0. La réciproque est elle vraie ?

- (d) Soient  $X_1$  et  $X_2$  deux variables indépendantes, et  $Y_1$  et  $Y_2$  deux variables indépendantes. Montrer que

$$d(X_1 + X_2, Y_1 + Y_2) \leq d(X_1, X_2) + d(Y_1, Y_2).$$

- (e) Application : soit  $(X_n)_{n \in \mathbb{N}}$  une famille de variables indépendantes de lois de Bernoulli de paramètres  $(p_n)_{n \in \mathbb{N}}$ , et  $(Y_n)_{n \in \mathbb{N}}$  une famille de variables indépendantes de lois de Poisson de mêmes paramètres  $(p_n)_{n \in \mathbb{N}}$ . Montrer que

$$d(X_1 + X_2 + \cdots + X_n, Y_1 + Y_2 + \cdots + Y_n) \leq \sum_{i=1}^n (p_i)^2.$$

Retrouver le résultat de l'exercice précédent (cette fois-ci sans utiliser le critère de Fourier).

- (13) **Approximation de fonctions par des polynômes.** Soit  $f$  une fonction continue sur  $[0, 1]$ . On définit une suite de polynômes

$$P_n^f(x) = \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

- (a) Soit  $X_n$  une variable aléatoire de loi  $\mathcal{B}(n, p)$ . Calculer  $\mathbb{E}[f(\frac{X_n}{n})]$ .  
 (b) Quelle est la limite de  $\mathbb{E}[f(\frac{X_n}{n})]$ ? On pourra décomposer l'espérance en deux parties :

$$\mathbb{E}\left[1_{|\frac{X_n}{n}-p| \leq \varepsilon} f\left(\frac{X_n}{n}\right)\right] + \mathbb{E}\left[1_{|\frac{X_n}{n}-p| > \varepsilon} f\left(\frac{X_n}{n}\right)\right]$$

et utiliser la loi des grands nombres.

- (c) Montrer que pour tout  $x \in [0, 1]$ ,  $\lim_{n \rightarrow \infty} P_n^f(x) = f(x)$ . Commenter.



## Introduction aux statistiques

**Contenu :** estimateurs, biais, erreur quadratique moyenne ; estimateurs de moyenne et de variance, intervalles de confiance ; tests d'hypothèses, test de la moyenne.

### 1. Estimateurs paramétriques

Dans les chapitres précédents, on a introduit et étudié divers modèles de variables aléatoires, en fixant *a priori* la loi de ces variables, et en en déduisant des résultats tels que la loi des grands nombres ou le théorème central limite. Ainsi, on a à chaque fois fixé une loi (binomiale  $\mathcal{B}(n, p)$ , de Poisson  $\mathcal{P}(\lambda)$ , gaussienne  $\mathcal{N}(m, \sigma^2)$ , etc.), et on en a déduit des résultats concernant une variable  $X$  suivant cette loi, ou concernant une suite de variables  $X_1, X_2, \dots, X_n, \dots$  munies de cette loi.

Lorsqu'on modélise des phénomènes réels, on se retrouve souvent dans la situation inverse : on observe une série de données  $X_1, X_2, \dots, X_n, \dots$ , dont la loi n'est pas encore déterminée (c'est justement le problème posé dans la modélisation), et on recherche donc cette loi. Les exemples suivants précisent ce point de vue "inverse", qui est exactement celui de la théorie des statistiques.

EXEMPLE 5.1. Considérons comme dans le chapitre précédent une population de  $N$  personnes, qu'on interroge un à un au sujet d'une question pour laquelle  $pN$  individus ont une opinion favorable, et  $(1 - p)N$  individus ont une opinion défavorable. Les résultats de ces questions sont des variables  $X_1, X_2, \dots$  indépendantes suivant des lois de Bernoulli  $\mathcal{B}(p)$  de paramètre  $p$  inconnu, et l'objectif d'un sondage est d'estimer la valeur de  $p$ , pour savoir par exemple si  $p \geq \frac{1}{2}$ . La loi des grands nombres fournit un estimateur raisonnable de ce paramètre en fonction des données  $X_1, X_2, \dots, X_n$ , à savoir, la moyenne empirique  $M_n = \frac{1}{n} \sum_{i=1}^n X_i$ . On verra plus loin comment mesurer la "qualité" de cet estimateur.

EXEMPLE 5.2. On découvre des poteries lors d'une fouille archéologique, toutes provenant d'un même four, mais ce four étant inconnu. On suppose que l'on connaît l'ensemble des fours  $(F_1, F_2, \dots, F_N)$  ayant pu servir à la réalisation de ces poteries, et que l'on connaît parfaitement les propriétés physiques de ces fours. Par exemple, on peut supposer que pour chaque four  $F_i$ , une poterie qui y est réalisée contient une quantité aléatoire de fer suivant une loi à densité

$$f_i(x) = 1_{x \in [0,1]} a_i x^{a_i-1} dx,$$

où  $a_i > 0$  est un paramètre spécifique au four  $F_i$ . Alors, si l'on dose les quantités de fer  $X_1, X_2, \dots$  dans les poteries découvertes, on observera des variables indépendantes de loi

$\gamma(a, 1)$  (voir les exercices du chapitre sur les variables continues), où  $a$  est un paramètre inconnu. Si l'on parvient à estimer  $a$  à partir des données  $(X_n)_{n \in \mathbb{N}}$  et à identifier le four  $F_i$  tel que  $a = a_i$ , alors on pourra retrouver le four ayant servi à la confection des poteries, et conclure quant à leur origine. En remarquant que

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f(x) dx = \int_{x=0}^1 a x^a dx = \frac{a}{a+1},$$

on déduit de la loi des grands nombres un estimateur de la quantité  $\frac{a}{a+1}$  : la moyenne empirique  $M_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Elle converge presque sûrement vers  $\frac{a}{a+1}$ , et par conséquent,

$$\hat{a}_n = \frac{M_n}{1 - M_n} = \frac{X_1 + \dots + X_n}{n - X_1 - \dots - X_n}$$

est une quantité aléatoire qui converge presque sûrement vers  $a$ , et qui permet donc de trouver l'origine des poteries.

Précisons maintenant la méthodologie statistique que nous employerons dans tout ce chapitre. On suppose donnée une famille de variables aléatoires indépendantes et de même loi  $(X_n)_{n \in \mathbb{N}}$ , ces variables pouvant être discrètes ou continues, et leur loi étant inconnue ou partiellement connue. Par exemple, on pourra supposer que les variables  $X_n$  suivent une loi gaussienne  $\mathcal{N}(m, \sigma^2)$ , mais ne pas préciser la valeur de  $m$ , ou alors ne préciser ni la valeur de  $m$  ni la valeur de  $\sigma^2$ . L'objectif sera alors d'estimer la valeur du ou des paramètres inconnus dans la loi, et on s'autorisera pour cela à manipuler n'importe quelle **statistique** des données  $X_1, X_2, \dots, X_n$ , c'est-à-dire, n'importe quelle fonction  $f(X_1, X_2, \dots, X_n)$ .

REMARQUE 5.3. On pourrait plus généralement ne rien supposer du tout sur la loi des variables, mais dans ce cas, l'ensemble des possibilités pour la loi  $\mathbb{P}_X$  serait un espace de dimension infini (estimation dite *non paramétrique*), et le problème serait trop difficile à résoudre avec les outils développés précédemment.

Dans le contexte précédent, notons de façon générique  $\theta$  le paramètre à estimer dans la loi des variables  $X_1, \dots, X_n$ ; ce peut être  $\theta = p$  si les variables sont supposées suivre une loi de Bernoulli  $\mathcal{B}(p)$ , ou  $\theta = \lambda$  si les variables sont supposées suivre une loi de Poisson  $\mathcal{P}(\lambda)$ , etc.

DÉFINITION 5.4. Un **estimateur** du paramètre  $\theta$  qui est construit à partir des données  $X_1, \dots, X_n$  est une variable aléatoire  $\hat{\theta}$  qui s'écrit sous la forme

$$\hat{\theta} = f(X_1, \dots, X_n),$$

c'est-à-dire une statistique. Le **biais** de cet estimateur est la quantité

$$b(\hat{\theta}) = \mathbb{E}[\hat{\theta} - \theta] = \mathbb{E}[\hat{\theta}] - \theta.$$

On dit que l'estimateur est sans biais si  $b(\hat{\theta}) = 0$ .

EXEMPLE 5.5. Soit  $m$  la moyenne de la loi des variables  $X_1, X_2, \dots, X_n$ , qui sont indépendantes et de même loi :  $m = \mathbb{E}[X_1] = \dots = \mathbb{E}[X_n]$ . Chaque variable  $X_i$  est un estimateur sans biais du paramètre  $m$  :  $b(X_i) = \mathbb{E}[X_i - m] = 0$ . Il en va de même pour la moyenne empirique  $\widehat{m} = M_n = \frac{1}{n} \sum_{i=1}^n X_i$  :

$$\mathbb{E}[\widehat{m} - m] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i - m] = 0.$$

La moyenne empirique  $\widehat{m}$  est un bien meilleur estimateur de  $m$  qu'une variable  $X_i$ , au sens de l'**erreur moyenne au carré**  $\mathbb{E}[(\widehat{m} - m)^2]$ . En effet,

$$\mathbb{E}[(\widehat{m} - m)^2] = \text{var}(\widehat{m}) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) = \frac{\text{var}(X_1)}{n} = \frac{\mathbb{E}[(X_i - m)^2]}{n}.$$

Ainsi, l'erreur moyenne au carré de l'estimateur  $\widehat{m}$  par rapport au paramètre  $m$  est  $n$  fois moins grande que celle de l'estimateur  $X_i$ .

De façon générale, si  $\widehat{\theta}$  est un estimateur d'un paramètre  $\theta$ , on appelle **erreur moyenne au carré** de  $\widehat{\theta}$  la quantité  $e^2(\widehat{\theta}) = \mathbb{E}[(\widehat{\theta} - \theta)^2]$ . L'objectif est alors de trouver des estimateurs de  $\theta$  construits à partir des données  $X_1, \dots, X_n$ , et tels que l'erreur moyenne au carré  $e^2(\widehat{\theta})$  soit la plus petite possible. Le résultat très simple suivant établit une relation entre  $e^2(\widehat{\theta})$  et le biais  $b(\widehat{\theta})$ .

PROPOSITION 5.6. *Pour tout estimateur  $\widehat{\theta}$ , on a*

$$e^2(\widehat{\theta}) = \text{var}(\widehat{\theta}) + \left(b(\widehat{\theta})\right)^2.$$

DÉMONSTRATION. Si  $m = \mathbb{E}[\widehat{\theta}]$ , alors

$$\begin{aligned} e^2(\widehat{\theta}) &= \mathbb{E}[(\widehat{\theta} - \theta)^2] = \mathbb{E}[(\widehat{\theta} - m)^2] + 2\mathbb{E}[(\widehat{\theta} - m)(m - \theta)] + \mathbb{E}[(m - \theta)^2] \\ &= \text{var}(\widehat{\theta}) + 0 + (m - \theta)^2 = \text{var}(\widehat{\theta}) + \left(b(\widehat{\theta})\right)^2. \end{aligned}$$

□

Cette proposition assure que pour avoir une erreur moyenne au carré la plus petite possible, il faut à la fois un faible biais, mais également une faible variance de l'estimateur. Pour avoir le meilleur estimateur possible, il faut souvent trouver un équilibre entre ces deux contraintes, comme l'illustre l'exemple important suivant.

EXEMPLE 5.7. Étant donnée une loi de probabilité discrète ou continue, on souhaite en estimer la variance  $V$  à partir de réalisations indépendantes  $X_1, X_2, \dots, X_n$  de cette loi. Comme  $V = \mathbb{E}[(X - \mathbb{E}[X])^2]$ , si l'on connaissait l'espérance  $m$  de la loi, alors un estimateur raisonnable de  $V$  serait

$$\widehat{W} = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2,$$

en considérant la variance comme une espérance, et en utilisant la discussion précédente sur les estimateurs de moyennes. Si l'on ne suppose pas connue  $m$ , alors une idée est de remplacer  $m$  dans  $\widehat{W}$  par son estimateur, ce qui donne l'estimateur suivant pour la variance :

$$\widehat{V} = \frac{1}{n} \sum_{i=1}^n (X_i - M_n)^2 = \frac{1}{n} \sum_{i=1}^n \left( X_i - \frac{X_1 + \dots + X_n}{n} \right)^2.$$

Calculons le biais de cet estimateur  $\widehat{V}$ . On a :

$$\mathbb{E}[\widehat{V}] = \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \left( X_i - \frac{X_1 + \dots + X_n}{n} \right)^2 \right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \left( \widetilde{X}_i - \frac{\widetilde{X}_1 + \dots + \widetilde{X}_n}{n} \right)^2 \right],$$

où  $\widetilde{X}_i = X_i - \mathbb{E}[X_i] = X_i - \mathbb{E}[X_1]$ . Fixons l'indice  $i$  et développons l'espérance du carré :

$$\begin{aligned} \mathbb{E} \left[ \left( \widetilde{X}_i - \frac{\widetilde{X}_1 + \dots + \widetilde{X}_n}{n} \right)^2 \right] &= \text{var}(\widetilde{X}_i) - \frac{2}{n} \sum_{j=1}^n \mathbb{E}[\widetilde{X}_i \widetilde{X}_j] + \frac{1}{n^2} \text{var}(\widetilde{X}_1 + \dots + \widetilde{X}_n) \\ &= \text{var}(X_1) - \frac{2}{n} \text{var}(X_1) + \frac{1}{n} \text{var}(X_1) = \frac{n-1}{n} \text{var}(X_1), \end{aligned}$$

en remarquant que dans la somme  $\sum_{j=1}^n \mathbb{E}[\widetilde{X}_i \widetilde{X}_j]$ , seul le terme  $j = i$  est non nul par indépendance de  $\widetilde{X}_i$  et  $\widetilde{X}_j$ . On conclut que  $\mathbb{E}[\widehat{V}] = \frac{n-1}{n} \text{var}(X_1)$ , et donc que

$$b(\widehat{V}) = -\frac{1}{n} \text{var}(X_1).$$

Si l'on veut un estimateur sans biais, on peut prendre

$$\widehat{V}' = \frac{n}{n-1} \widehat{V} = \frac{1}{n-1} \sum_{i=1}^n \left( X_i - \frac{X_1 + \dots + X_n}{n} \right)^2.$$

Néanmoins,  $\widehat{V}'$  n'est pas forcément meilleur que  $\widehat{V}$ . Supposons dans tout ce qui suit que la loi à estimer est une loi gaussienne  $\mathcal{N}(0, \sigma^2)$ . Le vecteur  $\vec{X} = (X_1, \dots, X_n)$  est un vecteur aléatoire dans  $\mathbb{R}^n$ , de densité jointe

$$f_{\vec{X}}(\vec{x}) = f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{(x_1)^2 + \dots + (x_n)^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\|\vec{x}\|^2}{2\sigma^2}},$$

où  $\vec{x} = (x_1, \dots, x_n)$  et  $\|\vec{x}\|$  est sa norme euclidienne, c'est-à-dire  $\sqrt{\sum_{i=1}^n (x_i)^2}$ . On fixe une base orthonormale  $(e_1, \dots, e_n)$  de  $\mathbb{R}^n$ , telle que  $(e_1, \dots, e_{n-1})$  soit une base de l'hyperplan

$$H = \{\vec{x} \in \mathbb{R}^n \mid x_1 + x_2 + \dots + x_n = 0\}.$$

On peut prendre  $e_n = (\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})$ . Si  $\vec{x} \in \mathbb{R}^n$ , notons

$$p_H(\vec{x}) = \text{projeté orthogonal de } \vec{x} \text{ sur } H = \sum_{i=1}^{n-1} \langle \vec{x} \mid e_i \rangle e_i;$$

$$p_D(\vec{x}) = \vec{x} - p_H(\vec{x}) = \langle \vec{x} \mid e_n \rangle e_n.$$

Les vecteurs  $p_H(\vec{x})$  et  $p_D(\vec{x})$  sont orthogonaux, donc  $\|\vec{x}\|^2 = \|p_H(\vec{x})\|^2 + \|p_D(\vec{x})\|^2$ . Alors,

$$f_{\vec{X}}(\vec{x}) = \left( \frac{1}{(2\pi\sigma^2)^{\frac{n-1}{2}}} e^{-\frac{\|p_H(\vec{x})\|^2}{2\sigma^2}} \right) \left( \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-\frac{\|p_D(\vec{x})\|^2}{2\sigma^2}} \right),$$

et cette factorisation prouve que les vecteurs  $p_H(\vec{X})$  et  $p_D(\vec{X})$  sont indépendants, avec

- $p_H(\vec{X})$  suivant la loi d'un vecteur gaussien dans un espace de dimension  $n-1$ , avec toutes les coordonnées orthogonales indépendantes de variance  $\sigma^2$ ,
- et  $p_D(\vec{X})$  ayant une loi gaussienne  $\mathcal{N}(0, \sigma^2)$  sur la droite  $D = H^\perp$ .

Remarquons alors que

$$p_H(\vec{X}) = \vec{X} - p_D(\vec{X}) = \vec{X} - (M_n, M_n, \dots, M_n) = (X_1 - M_n, \dots, X_n - M_n),$$

et donc que  $\frac{n\widehat{V}}{\sigma^2}$  est la norme au carré de  $p_H(\frac{\vec{X}}{\sigma})$ , qui a la loi de la somme de  $(n-1)$  carrés de variables gaussiennes standards indépendantes. Ce raisonnement géométrique montre que

$$\frac{n\widehat{V}}{\sigma^2} = \frac{(n-1)\widehat{V}'}{\sigma^2} \sim \chi^2(n-1).$$

On obtient une loi du  $\chi^2$  à  $n-1$  degrés de liberté. Sa densité a été calculée précédemment :

$$f_n \widehat{v}(t) = 1_{t>0} \frac{1}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} t^{\frac{n-1}{2}-1} e^{-\frac{t}{2}}.$$

On calcule alors aisément :

$$\mathbb{E}[n\widehat{V}] = \sigma^2 \int_{t=0}^{\infty} \frac{1}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} t^{\frac{n-1}{2}} e^{-\frac{t}{2}} dt = \sigma^2 \frac{2^{\frac{n+1}{2}} \Gamma(\frac{n+1}{2})}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} = \sigma^2 (n-1);$$

$$\mathbb{E}[(n\widehat{V})^2] = \sigma^4 \int_{t=0}^{\infty} \frac{1}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} t^{\frac{n-1}{2}+1} e^{-\frac{t}{2}} dt = \sigma^4 \frac{2^{\frac{n+3}{2}} \Gamma(\frac{n+3}{2})}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} = \sigma^4 (n+1)(n-1)$$

de sorte que  $\text{var}(n\widehat{V}) = \text{var}((n-1)\widehat{V}') = 2\sigma^4(n-1)$ . On en déduit finalement :

$$e^2(\widehat{V}) = \text{var}(\widehat{V}) + b(\widehat{V})^2 = \frac{2(n-1)\sigma^4}{n^2} + \frac{\sigma^4}{n^2} = \frac{(2n-1)\sigma^4}{n^2}$$

$$e^2(\widehat{V}') = \text{var}(\widehat{V}') + b(\widehat{V}')^2 = \frac{2\sigma^4}{n-1}.$$

En particulier,  $e^2(\widehat{V}) \leq \frac{2\sigma^4}{n} < \frac{2\sigma^4}{n-1} = e^2(\widehat{V}')$ . Ainsi, l'estimateur  $\widehat{V}$  a un biais plus grand que le biais nul de  $\widehat{V}'$ , mais cet estimateur a une meilleure erreur moyenne au carré.

En pratique, on ne manipulera pas seulement un estimateur  $\widehat{\theta}$  du paramètre  $\theta$ , mais une suite  $(\widehat{\theta}_n)_{n \in \mathbb{N}}$  d'estimateurs, chaque  $\widehat{\theta}_n$  étant une fonction des données  $X_1, X_2, \dots, X_n$ . Dans ce cadre, la définition suivante explique ce que l'on attend d'une "bonne" suite d'estimateurs :

DÉFINITION 5.8. Une suite d'estimateurs  $(\widehat{\theta}_n)_{n \in \mathbb{N}}$  du paramètre  $\theta$  sera dite **consistante** si pour tout  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[ |\widehat{\theta}_n - \theta| \geq \varepsilon \right] = 0.$$

Dans ce cas, si l'on dispose d'un grand nombre  $n$  de données, alors avec très grande probabilité,  $\widehat{\theta}_n$  sera très proche de  $\theta$ , et c'est ce que l'on souhaite.

EXEMPLE 5.9. Par la loi (faible) des grands nombres, la suite des moyennes empiriques  $(M_n = \frac{X_1 + \dots + X_n}{n})_{n \in \mathbb{N}}$  d'une suite de variables aléatoires indépendantes, de même loi et ayant une espérance finie est une suite consistante d'estimateurs de la moyenne théorique  $\mathbb{E}[X_1]$ .

Établissons finalement le lien entre erreur moyenne au carré et consistance :

PROPOSITION 5.10. Soit  $(\widehat{\theta}_n)_{n \in \mathbb{N}}$  une suite d'estimateurs d'un paramètre  $\theta$  d'une mesure de probabilité. Si  $\lim_{n \rightarrow \infty} e^2(\widehat{\theta}_n) = 0$ , alors la suite est consistante.

DÉMONSTRATION. On écrit l'inégalité de Markov (ou Bienaymé-Chebyshev) :

$$\mathbb{P} \left[ |\widehat{\theta}_n - \theta| \geq \varepsilon \right] \leq \frac{\mathbb{E}[(\widehat{\theta}_n - \theta)^2]}{\varepsilon^2} = \frac{e^2(\widehat{\theta}_n)}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

Notons que compte tenu de la relation entre erreur moyenne au carré, variance et biais,  $e^2(\widehat{\theta}_n)$  tend vers 0 si et seulement si le biais  $b(\widehat{\theta}_n)$  et la variance  $\text{var}(\widehat{\theta}_n)$  tendent vers 0.  $\square$

EXEMPLE 5.11. Si  $X_1, \dots, X_n$  sont des variables gaussiennes de loi  $\mathcal{N}(0, \sigma^2)$ , un estimateur sans biais de la variance  $V = \sigma^2$  est

$$\widehat{V}_n' = \frac{1}{n-1} \sum_{i=1}^n \left( X_i - \frac{X_1 + \dots + X_n}{n} \right)^2,$$

et on a vu que son erreur moyenne au carré était  $e^2(\widehat{V}_n') = 2\sigma^4/(n-1)$ . Elle tend donc vers 0, et par conséquent,  $(\widehat{V}_n')_{n \in \mathbb{N}}$  est une suite consistante d'estimateurs de  $V$ . Le même raisonnement montre que

$$\widehat{V}_n = \frac{1}{n} \sum_{i=1}^n \left( X_i - \frac{X_1 + \dots + X_n}{n} \right)^2,$$

est aussi une suite consistante d'estimateurs de  $V$ . De façon générale, si  $X_1, X_2, \dots$  sont des variables indépendantes et de même loi admettant un moment d'ordre 4, on peut calculer

$$\begin{aligned} \text{var}(\widehat{V}_n') &= \frac{1}{n-1} \left( \frac{n-1}{n} \mathbb{E}[(X_1 - \mathbb{E}[X_1])^4] - \frac{n-3}{n} (\mathbb{E}[(X_1 - \mathbb{E}[X_1])^2])^2 \right); \\ \text{var}(\widehat{V}_n) &= \frac{n-1}{n^2} \left( \frac{n-1}{n} \mathbb{E}[(X_1 - \mathbb{E}[X_1])^4] - \frac{n-3}{n} (\mathbb{E}[(X_1 - \mathbb{E}[X_1])^2])^2 \right). \end{aligned}$$

Par conséquent, les deux variances tendent vers 0, et comme il en va de même pour les biais, les deux suites d'estimateurs  $(\widehat{V}_n')_{n \in \mathbb{N}}$  et  $(\widehat{V}_n)_{n \in \mathbb{N}}$  sont consistants pour la variance  $\text{var}(X_1)$ .

## 2. Intervalles de confiance

On a introduit dans la section précédente la notion d'estimateur d'un paramètre  $\theta$  d'une loi d'une variable aléatoire : c'est une variable aléatoire  $\widehat{\theta}$  qui est construite à partir de données  $X_1, \dots, X_n$ , et qui approche  $\theta$ , la qualité de cette approximation étant par exemple contrôlée par l'erreur moyenne au carré  $e^2(\widehat{\theta})$ .

Pour contrôler la qualité de l'approximation d'un estimateur, on peut plus généralement se donner une marge d'erreur  $\varepsilon$ , et un niveau de confiance  $\gamma < 1$  proche de 1 (par exemple 95%) ; on souhaite alors que

$$\mathbb{P}\left[\widehat{\theta} - \varepsilon \leq \theta \leq \widehat{\theta} + \varepsilon\right] \geq \gamma.$$

La notion d'**intervalle de confiance** généralise encore cette approche :

**DÉFINITION 5.12.** Soit  $\gamma = 1 - \alpha$  un réel plus petit que 1 (par exemple  $\gamma = 95\%$ ), et  $\theta$  un paramètre de la loi commune de variables indépendantes  $X_1, \dots, X_n$ . Un intervalle de confiance de **niveau**  $\gamma$  pour le paramètre  $\theta$  est la donnée de deux variables aléatoires  $U$  et  $V$ , construites à partir des données  $X_1, \dots, X_n$ , telles que

$$\mathbb{P}[U \leq \theta \leq V] = \gamma.$$

**EXEMPLE 5.13.** Considérons la loi gaussienne  $\mathcal{N}(m, 1)$ , et une variable  $X$  suivant cette loi. On a :

$$\begin{aligned} \mathbb{P}[|X - m| \leq t] &= \int_{m-t}^{m+t} e^{-\frac{(x-m)^2}{2}} \frac{dx}{\sqrt{2\pi}} = \int_{-t}^t e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \\ &= F_G(t) - F_G(-t) = 1 - 2F_G(-t) = 1 - 2 \int_t^\infty e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}, \end{aligned}$$

où  $F_G$  est la fonction de répartition d'une gaussienne. Notons  $e(t) = \int_t^\infty e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}$  ; c'est une fonction décroissante de  $t$ , qui tend vers 0 lorsque  $t$  tend vers l'infini. Si  $\gamma < 1$  est un niveau fixé, soit  $t$  l'unique réel positif tel que  $1 - 2e(t) = \gamma$ . On a alors

$$\mathbb{P}[X - t \leq m \leq X + t] = \mathbb{P}[|X - m| \leq t] = 1 - 2e(t) = \gamma,$$

donc  $[X - t, X + t]$  est un intervalle de confiance de niveau  $\gamma$  pour le paramètre de moyenne  $m$  de la loi.

**EXEMPLE 5.14.** Considérons toujours la loi gaussienne  $\mathcal{N}(m, 1)$ , et des variables indépendantes  $X_1, \dots, X_n$  suivant cette loi. La loi de la somme  $X_1 + X_2 + \dots + X_n$  est  $\mathcal{N}(nm, n)$ , voir les exercices du chapitre précédent. Par conséquent,

$$\begin{aligned} \mathbb{P}\left[\left|\frac{X_1 + \dots + X_n}{n} - m\right| \leq t\right] &= \mathbb{P}[|(X_1 + \dots + X_n) - nm| \leq nt] \\ &= \int_{nm-nt}^{nm+nt} e^{-\frac{(x-nm)^2}{2n}} \frac{dx}{\sqrt{2\pi n}} = \int_{-nt}^{nt} e^{-\frac{x^2}{2n}} \frac{dx}{\sqrt{2\pi n}} \\ &= \int_{-\sqrt{nt}}^{\sqrt{nt}} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}. \end{aligned}$$

Par conséquent, compte tenu des calculs de l'exemple précédent, si  $t$  est le réel positif tel que  $1 - 2e(t) = \gamma$ , alors

$$\begin{aligned} & \mathbb{P}\left[\frac{X_1 + \cdots + X_n}{n} - \frac{t}{\sqrt{n}} \leq m \leq \frac{X_1 + \cdots + X_n}{n} + \frac{t}{\sqrt{n}}\right] \\ &= \mathbb{P}[|X_1 + \cdots + X_n - nm| \leq \sqrt{nt}] = \int_{-t}^t e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \\ &= 1 - 2e(t) = \gamma \end{aligned}$$

donc  $[M_n - \frac{t}{\sqrt{n}}, M_n + \frac{t}{\sqrt{n}}]$  est un intervalle de confiance de niveau  $\gamma$  pour le paramètre  $m$ . Si  $\gamma = 95\%$ , on trouve une valeur de  $t$  égale environ à 1.96, donc

$$\mathbb{P}\left[M_n - \frac{1.96}{\sqrt{n}} \leq m \leq M_n + \frac{1.96}{\sqrt{n}}\right] \simeq 95\%.$$

Lorsque  $n$  augmente, on obtient un intervalle de confiance de taille qui décroît en  $\frac{1}{\sqrt{n}}$ .

Le théorème central limite permet de construire facilement des intervalles de confiance pour la moyenne d'une variable aléatoire. Soit  $m$  la moyenne de la loi de variables aléatoires indépendantes  $X_1, X_2, \dots, X_n, \dots$ , et  $\sigma^2$  leur variance. On sait que

$$\mathbb{P}\left[\sqrt{\frac{n}{\sigma^2}} \left|\frac{X_1 + \cdots + X_n}{n} - m\right| \leq t\right] \xrightarrow{n \rightarrow \infty} \int_{-t}^t e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}.$$

Par conséquent, si  $\gamma$  est un niveau de confiance fixé, et si  $t > 0$  est le **quantile** de la loi gaussienne  $\mathcal{N}(0, 1)$  tel que

$$F_{\mathcal{N}(0,1)}(t) = \frac{1}{2} + \frac{F_{\mathcal{N}(0,1)}(t) - F_{\mathcal{N}(0,1)}(-t)}{2} = \frac{1}{2} + \frac{\gamma}{2},$$

alors

$$\left[\frac{X_1 + \cdots + X_n}{n} - \frac{t\sigma}{\sqrt{n}}, \frac{X_1 + \cdots + X_n}{n} + \frac{t\sigma}{\sqrt{n}}\right]$$

est un intervalle de confiance pour  $m$  qui est *asymptotiquement* de niveau  $\gamma$ , c'est-à-dire que

$$\mathbb{P}\left[\frac{X_1 + \cdots + X_n}{n} - \frac{t\sigma}{\sqrt{n}}, \frac{X_1 + \cdots + X_n}{n} + \frac{t\sigma}{\sqrt{n}}\right] \xrightarrow{n \rightarrow \infty} \gamma.$$

On prend souvent un niveau  $\gamma = 95\%$ , qui donne un quantile  $t = 1.96$ .

**EXEMPLE 5.15.** Considérons une loi exponentielle  $\mathcal{E}(\lambda)$ , de paramètre  $\lambda$  à estimer. Sa moyenne  $m$  est  $\frac{1}{\lambda}$ , et il en va de même pour  $\sigma = \sqrt{\text{var}(X)} = \frac{1}{\lambda}$ . D'après ce qui précède,

$$\left[\frac{X_1 + \cdots + X_n}{n} - \frac{1.96}{\sqrt{n}\lambda}, \frac{X_1 + \cdots + X_n}{n} + \frac{1.96}{\sqrt{n}\lambda}\right]$$

est un intervalle de confiance de niveau asymptotiquement 95% pour le paramètre  $\frac{1}{\lambda}$ . Ceci peut se réécrire sous la forme suivante :

$$\lim_{n \rightarrow \infty} \mathbb{P}\left[\frac{1}{M_n} \left(1 - \frac{1.96}{\sqrt{n}}\right) \leq \lambda \leq \frac{1}{M_n} \left(1 + \frac{1.96}{\sqrt{n}}\right)\right] = 95\%,$$



où  $M_n = \frac{X_1 + \dots + X_n}{n}$  est l'estimateur usuel de la moyenne. Donc, un intervalle de confiance asymptotiquement de niveau 95% pour le paramètre  $\lambda$  est

$$\left[ \frac{1}{M_n} \left( 1 - \frac{1.96}{\sqrt{n}} \right), \frac{1}{M_n} \left( 1 + \frac{1.96}{\sqrt{n}} \right) \right].$$

REMARQUE 5.16. La notion d'intervalle de confiance est très employée dans les métiers de l'industrie, où l'on souhaite fabriquer des pièces avec une certaine tolérance pour leurs propriétés (dimension, poids, *etc.*), et où l'on s'autorise sur les chaînes de fabrication un certain taux de pertes (pièces non conformes, que l'on devra remplacer). Dans ce contexte, le taux de pertes est lié au niveau de confiance de l'intervalle à construire, typiquement 99% ou plus ; et la tolérance autorisée est liée à la taille  $V - U$  de l'intervalle de confiance  $[U, V]$ .

### 3. Tests d'hypothèses statistiques

À partir des estimateurs et intervalles de confiance construits dans les sections précédentes, on peut mettre au point des processus de décision statistiques, qui sont appelés **tests statistiques** et qui permettent de confirmer ou d'infirmer des hypothèses. De façon générale, on suit le programme suivant :

- (1) On considère deux hypothèses  $H_0$  et  $H_1$ , la première étant appelée **hypothèse nulle** et la seconde **hypothèse alternative**. La personne chargée de décider entre ces deux hypothèses suppose *a priori* que  $H_0$  est vraie.
- (2) En fonction d'observations  $X_1, \dots, X_n$  considérées comme des variables aléatoires, la personne chargée de la décision va :
  - changer d'avis, et *rejeter* l'hypothèse  $H_0$  en faveur de l'hypothèse  $H_1$  ;
  - ou, *ne pas rejeter* l'hypothèse nulle  $H_0$  (ce qui n'est pas exactement équivalent à l'*accepter*, voir les exemples qui suivent).

Lors de ce processus, le décideur peut se tromper de deux façons possibles : il peut rejeter l'hypothèse  $H_0$  alors qu'elle est vraie (**erreur de première espèce**), et il peut ne pas rejeter l'hypothèse  $H_0$  alors qu'elle est fautive (**erreur de seconde espèce**). La théorie statistique des tests permet :

- de construire des tests standards d'hypothèses à partir d'observations aléatoires  $X_1, \dots, X_n$ , par exemple des variables aléatoires indépendantes et de même loi ;
- et d'estimer les deux types d'erreur, en cherchant un compromis entre elles.

EXEMPLE 5.17. Supposons qu'un jury soit chargé de décider de la culpabilité d'une personne interpellée par la police. Dans la plupart des pays développés, l'hypothèse de départ  $H_0$  est que la personne est innocente, et à partir des investigations menées par la justice (des observables  $X_1, \dots, X_n$ ), le jury doit décider :

- (1) soit que la personne est coupable, et donc condamnée (hypothèse alternative  $H_1$ ). Il faut pour cela qu'il y ait suffisamment de preuves issues des observations  $X_1, \dots, X_n$ .

- (2) soit que la personne n'est pas condamnée. Notons que les jurés n'affirment pas que la personne est non coupable : les hypothèses  $H_0$  ou  $H_1$  peuvent être vraies, mais le jury constate simplement qu'il *ne peut pas rejeter* l'hypothèse  $H_0$ , faute de preuves.

Dans ce contexte, les erreurs possibles sont résumées dans le tableau suivant :

décision du jury	relaxe	condamnation
individu innocent ( $H_0$ )	bonne décision	erreur de première espèce
individu coupable ( $H_1$ )	erreur de seconde espèce	bonne décision

Selon le type de régime politique, on cherchera à avoir l'erreur de première espèce la plus faible possible (justice humaine, mais peut-être laxiste), ou l'erreur de seconde espèce la plus faible possible (justice ferme, au risque que des innocents soient emprisonnés).

EXEMPLE 5.18. Considérons le problème du dépistage d'une maladie chez une personne. L'hypothèse de base  $H_0$  est que la personne est saine ; on ne la considérera malade que si une biopsie ou une prise de sang met en évidence la maladie. Dans ce contexte, le résultat du test médical et les erreurs possibles sont présentés dans le tableau suivant :

résultat du test	négatif	positif
individu sain ( $H_0$ )	bonne décision	erreur de première espèce
individu malade ( $H_1$ )	erreur de seconde espèce	bonne décision

Cet exemple illustre bien la dissymétrie des deux hypothèses et des deux types d'erreur. En effet, il est ici presque toujours considéré que l'erreur de seconde espèce est beaucoup plus grave que l'erreur de première espèce : on ne veut pas laisser des individus malades passer au travers du dépistage, tandis que l'on peut s'autoriser des "faux positifs" sans trop de risques (on peut par la suite retester avec une autre procédure les individus positifs pour vérifier qu'ils sont bien malades). De façon générale, si l'hypothèse alternative  $H_1$  est une hypothèse "sensible", c'est-à-dire une source de problèmes importants si elle est avérée (accident dans un processus industriel, maladie pouvant se répandre dans une population, etc.), alors on cherchera en premier lieu à minimiser l'erreur de seconde espèce.

On expliquera dans un instant comment construire une variable de décision à partir d'observations aléatoires de lois connues sous l'hypothèse  $H_0$  et sous l'hypothèse  $H_1$ . Donnons avant une définition plus précise des erreurs de première et de seconde espèce :

DÉFINITION 5.19. *L'erreur de première espèce dans un test statistique est la probabilité conditionnelle :*

$$\alpha = \mathbb{P}[\text{rejeter l'hypothèse } H_0 \mid H_0 \text{ est vraie}].$$

*L'erreur de seconde espèce dans un test statistique est la probabilité conditionnelle :*

$$\beta = \mathbb{P}[\text{ne pas rejeter l'hypothèse } H_0 \mid H_1 \text{ est vraie}].$$

On appelle **niveau** du test la quantité  $1 - \alpha$ , et **puissance** du test la quantité  $1 - \beta$ .

REMARQUE 5.20. D'après la discussion et les exemples précédents, *accepter* l'hypothèse nulle lorsqu'elle n'est *pas rejetée* peut être une décision particulièrement dangereuse, sauf si le test est très puissant (avec une valeur  $1 - \beta$  très proche de 1).

Voyons maintenant comment construire en pratique un test. Supposons donnée une variable aléatoire réelle  $X$ , dont on connaît la loi sous l'hypothèse  $H_0$ . On fixe un seuil  $\alpha$ , qui est l'erreur de première espèce que l'on s'autorise. Étant donnée une réalisation  $x$  de la variable  $X$ , on calcule la probabilité  $p(x)$  pour que sous l'hypothèse  $H_0$ ,  $X$  prenne des valeurs au moins aussi extrêmes que  $x$  (en général, ceci veut dire au moins aussi grandes, ou de valeur absolue au moins aussi grande).

- Si cette **p-valeur**  $p(x)$  est plus petite que le seuil  $\alpha$  fixé, alors l'hypothèse nulle est extrêmement improbable, et on la rejette. La probabilité pour que l'on se trompe est alors exactement  $\alpha$ , donc on a construit un test de niveau  $1 - \alpha$ .
- Sinon, on ne rejette pas l'hypothèse  $H_0$ . On peut alors se tromper avec une erreur de seconde espèce, et le calcul de cette erreur (ou de la puissance du test) est possible si l'on connaît aussi la loi de  $X$  sous l'hypothèse alternative  $H_1$ .

EXEMPLE 5.21 (Test de la moyenne). Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires indépendantes de même loi  $\mathcal{N}(m, \sigma^2)$ . On souhaite tester l'hypothèse nulle :

$$H_0 : m = m_0,$$

contre l'hypothèse alternative  $H_1$  définie par  $m \neq m_0$ . On fixe un seuil  $\alpha$  et on souhaite construire à partir des variables  $X_1, \dots, X_n$  un test de niveau  $1 - \alpha$  pour ce schéma d'hypothèses. Si  $\alpha = 5\%$ , on rappelle qu'on a un intervalle de confiance de niveau exactement  $1 - \alpha$  pour le paramètre  $m$  :

$$\mathbb{P} \left[ M_n - \frac{1.96 \sigma}{\sqrt{n}} \leq m \leq M_n + \frac{1.96 \sigma}{\sqrt{n}} \right] = 1 - \alpha,$$

car la moyenne empirique  $M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$  suit sous l'hypothèse  $H_0$  une loi gaussienne  $\mathcal{N}(m, \frac{\sigma^2}{n})$ .

Supposons donnée une valeur  $x$  de la moyenne. La probabilité pour que sous  $H_0$ , la différence  $|M_n - m_0|$  soit au moins égale à  $|x - m_0|$  est la  $p$ -valeur

$$\begin{aligned} \mathbb{P}[|M_n - m_0| > |x - m_0|] &= 1 - \mathbb{P}[m_0 - |x - m_0| \leq M_n \leq m_0 + |x - m_0|] \\ &= 2 \int_{|x - m_0|}^{\infty} e^{-\frac{ny^2}{2\sigma^2}} \frac{\sqrt{n} dy}{\sqrt{2\pi\sigma^2}} \\ &= 2 \int_{\frac{\sqrt{n}|x - m_0|}{\sigma}}^{\infty} e^{-\frac{y^2}{2}} \frac{dy}{\sqrt{2\pi}}. \end{aligned}$$

Pour que cette  $p$ -valeur soit plus petite que le seuil  $\alpha = 5\%$ , il faut que  $\frac{\sqrt{n}|x - m_0|}{\sigma}$  soit au moins égal à 1.96, donc que  $|x - m_0| \geq \frac{1.96 \sigma}{\sqrt{n}}$ . On peut donc construire un test de niveau  $1 - \alpha = 95\%$  de la façon suivante :

- (1) On rejette l'hypothèse  $H_0$  si  $|M_n - m_0| \geq \frac{1.96\sigma}{\sqrt{n}}$ , ce qui revient à demander que  $m_0$  ne tombe pas dans l'intervalle de confiance

$$\left[ M_n - \frac{1.96\sigma}{\sqrt{n}}, M_n + \frac{1.96\sigma}{\sqrt{n}} \right]$$

qui est de niveau  $1 - \alpha$ .

- (2) Sinon, on ne rejette pas cette hypothèse nulle.

Si l'on ne précise pas plus l'hypothèse alternative  $H_1$ , on ne pas évaluer la puissance du test. Supposons maintenant que  $H_1$  est l'hypothèse  $m = m_1$ , avec  $m_1 \neq m_0$ . Alors, l'erreur de seconde espèce est :

$$\mathbb{P} \left[ |M_n - m_0| \leq \frac{1.96\sigma}{\sqrt{n}} \mid m = m_1 \right].$$

Supposons par exemple  $m_0 > m_1$ . Alors, l'événement ci-dessus a pour probabilité

$$\begin{aligned} & \mathbb{P} \left[ m_0 - m_1 - \frac{1.96\sigma}{\sqrt{n}} \leq M_n - m_1 \leq m_0 - m_1 - \frac{1.96\sigma}{\sqrt{n}} \right] \\ &= \sqrt{\frac{n}{2\pi\sigma^2}} \int_{m_0 - m_1 - \frac{1.96\sigma}{\sqrt{n}}}^{m_0 - m_1 + \frac{1.96\sigma}{\sqrt{n}}} e^{-\frac{ny^2}{2\sigma^2}} dy = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{n} \frac{m_0 - m_1}{\sigma} - 1.96}^{\sqrt{n} \frac{m_0 - m_1}{\sigma} + 1.96} e^{-\frac{y^2}{2}} dy \\ &\leq 1 - F_{\mathcal{N}(0,1)} \left( \sqrt{n} \frac{m_0 - m_1}{\sigma} - 1.96 \right), \end{aligned}$$

où  $F_{\mathcal{N}(0,1)}$  est la fonction de répartition d'une variable gaussienne standard. À niveau  $1 - \alpha$  fixé, la puissance du test tend donc vers 1 lorsque  $n$  tend vers l'infini.

## Exercices

- (1) **Estimateurs de variables de Bernoulli.** Soit  $X_1, \dots, X_n$  des variables aléatoires indépendantes de loi  $\mathcal{B}(p)$ . On considère les estimateurs classiques de la moyenne et de la variance

$$\hat{m} = \frac{1}{n} \sum_{i=1}^n X_i \quad ; \quad \hat{v} = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{m})^2.$$

- (a) Calculer le biais et l'erreur moyenne au carré de  $\hat{m}$ , vu comme estimateur de  $p$  (indication : quelle est la loi de la variable  $n\hat{m}$  ?).
- (b) Montrer que  $\hat{m}$  donne une suite consistante d'estimateurs de  $p$ .
- (c) Calculer le biais et l'erreur moyenne au carré de  $\hat{v}$ , vu comme estimateur du paramètre  $p(1-p)$ . On pourra remarquer que  $(X_i)^k = X_i$  pour tout  $k \geq 1$ .
- (2) **Meilleur estimateur de la variance d'une gaussienne.** Soit  $X_1, \dots, X_n$  des variables aléatoires indépendantes gaussiennes, de loi  $\mathcal{N}(m, \sigma^2)$ . On pose

$$T = c \sum_{i=1}^n \left( X_i - \frac{X_1 + \dots + X_n}{n} \right)^2,$$

avec  $c > 0$ .

- (a) Rappeler quelle est la loi de la variable  $\frac{1}{c\sigma} T$ .
- (b) Calculer en fonction de  $c$  le biais, puis l'erreur moyenne au carré de  $T$  en tant qu'estimateur de la variance  $\sigma^2$ .
- (c) Quelle valeur de  $c$  donne un estimateur sans biais ?
- (d) Quelle valeur de  $c$  donne l'estimateur avec la plus petite erreur moyenne au carré ?
- (3) **Estimateur du maximum de vraisemblance : cas de la loi de Poisson.** Soit  $X_1, \dots, X_n$  des variables indépendantes de loi  $\mathcal{P}(\lambda)$ . On souhaite estimer le paramètre  $\lambda$ . Étant donné un paramètre  $t > 0$  et des valeurs entières  $x_1, \dots, x_n$ , on note

$$h(x_1, \dots, x_n; t) = \log \left( \prod_{i=1}^n \mathcal{P}(t)[x_i] \right)$$

avec  $\mathcal{P}(t)[x] = e^{-t} \frac{t^x}{x!}$  probabilité de la loi de Poisson  $\mathcal{P}(t)$ .

- (a) Montrer que

$$h(x_1, \dots, x_n; t) = (x_1 + \dots + x_n) \log t - nt - \sum_{i=1}^n \log(x_i)!.$$

- (b) Les nombres entiers  $x_1, \dots, x_n$  étant fixés, calculer la valeur de  $t$  qui maximise la quantité  $h(x_1, \dots, x_n; t)$  ci-dessus. On note cette valeur  $t = \text{EMV}(x_1, \dots, x_n)$ , et on dit que  $t$  est l'*estimateur du maximum de vraisemblance*. Commenter cette terminologie.

- (c) On pose maintenant  $T_n = \text{EMV}(X_1, \dots, X_n)$ . Quel estimateur retrouve-t-on ? Montrer que  $(T_n)_{n \in \mathbb{N}}$  est une suite d'estimateurs sans biais du paramètre  $\lambda$ , qui est consistante.

- (4) **Estimateur du maximum de vraisemblance : cas de la loi normale.** Soit  $X_1, \dots, X_n$  des variables indépendantes de loi  $\mathcal{N}(m, \sigma^2)$ . On souhaite estimer les deux paramètres  $m$  et  $\sigma^2$ . Étant donnés des paramètres  $u \in \mathbb{R}$  et  $v > 0$  et des valeurs réelles  $x_1, \dots, x_n$ , on note

$$h(x_1, \dots, x_n; u, v) = \log \left( \prod_{i=1}^n f_{\mathcal{N}(u,v)}(x_i) \right)$$

avec  $f_{\mathcal{N}(u,v)}(x) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{(x-u)^2}{2v}}$  densité de la loi gaussienne  $\mathcal{N}(u, v)$ .

- (a) On fixe  $v$  et les nombres réels  $x_1, \dots, x_n$ . Montrer que la valeur  $u$  qui maximise la quantité  $h(x_1, \dots, x_n; u, v)$  est  $u = \frac{x_1 + \dots + x_n}{n}$ .
- (b) Les nombres réels  $x_1, \dots, x_n$  étant fixés, trouver les deux paramètres  $u$  et  $v$  qui maximise la quantité  $h(x_1, \dots, x_n; u, v)$ . On pose  $(u, v) = \text{EMV}(x_1, \dots, x_n)$ .
- (c) On pose maintenant  $(U_n, V_n) = \text{EMV}(X_1, \dots, X_n)$ . Quels estimateurs classiques retrouve-t-on ? Montrer que  $(U_n)_{n \in \mathbb{N}}$  et  $(V_n)_{n \in \mathbb{N}}$  sont des suites consistantes d'estimateurs des paramètres  $m$  et  $\sigma^2$ .
- (5) **Intervalles de confiance pour des variables de Bernoulli.** Soit  $X_1, \dots, X_n$  des variables indépendantes de loi  $\mathcal{B}(p)$ . On note  $M_n = \frac{1}{n} \sum_{i=1}^n X_i$ .

- (a) Montrer que

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[ M_n - 1.96 \sqrt{\frac{p(1-p)}{n}} \leq p \leq M_n + 1.96 \sqrt{\frac{p(1-p)}{n}} \right] = 95\%.$$

- (b) Quelle est la valeur maximale possible pour  $\sqrt{p(1-p)}$  ? En déduire que pour  $n$  assez grand,  $[M_n - \frac{1}{\sqrt{n}}, M_n + \frac{1}{\sqrt{n}}]$  est un intervalle de confiance pour le paramètre  $p$  de niveau supérieur à 95%.
- (c) Que peut-on dire de la quantité  $M_n(1 - M_n)$  lorsque  $n$  tend vers l'infini ? En déduire que

$$\left[ M_n - \frac{1.96 \sqrt{M_n(1 - M_n)}}{\sqrt{n}}, M_n + \frac{1.96 \sqrt{M_n(1 - M_n)}}{\sqrt{n}} \right]$$

est un intervalle de confiance asymptotiquement de niveau 95% pour le paramètre  $p$ .

- (6) **Pièce équilibrée.** Un joueur lance une pièce au hasard, et note  $X_1, \dots, X_n$  les résultats pile ou face de ces lancers, le pile étant noté 1 et le face étant noté 0. On suppose donc que  $X_1, \dots, X_n$  sont des variables de Bernoulli  $\mathcal{B}(p)$ , et l'on souhaite savoir si la pièce est équilibrée, c'est-à-dire si l'hypothèse

$$H_0 : p = \frac{1}{2}$$

est vérifiée.

- (a) Quelle hypothèse alternative  $H_1$  est raisonnable ?  
 (b) On utilise le résultat de l'exercice précédent, qui affirme que l'intervalle

$$\left[ \frac{1}{n} \sum_{i=1}^n X_i - \frac{1.96\sqrt{p(1-p)}}{\sqrt{n}}, \frac{1}{n} \sum_{i=1}^n X_i + \frac{1.96\sqrt{p(1-p)}}{\sqrt{n}} \right]$$

est un intervalle de confiance pour le paramètre  $p$  de niveau asymptotique 95%.  
 En déduire un test pour la paire d'hypothèses  $(H_0, H_1)$ , qui est asymptotiquement de niveau 95%. Peut-on évaluer sa puissance ?

- (7) **Tout-terrain.** Un fabricant de pédaliers de vélo veut produire des boîtiers dont le diamètre est de 41mm, avec une tolérance autorisée égale à  $+0.00\text{mm}/-0.10\text{mm}$  (standard Pressfit GXP). Autrement dit, pour pouvoir être utilisées, les pièces qu'il produit doivent avoir un diamètre au moins égal à 40.9mm, et au plus égal à 41mm. Il étudie deux procédés de fabrication, qui donnent des pièces dont le diamètre  $X$  ou  $Y$  suit :

— dans le premier cas, une loi de densité

$$f_X(x) = \begin{cases} 40 e^{-40(41-x)} & \text{si } x \leq 41, \\ 0 & \text{sinon.} \end{cases}$$

— dans le second cas, une loi de densité

$$f_Y(y) = \frac{40}{\sqrt{2\pi}} e^{-800(y-40.95)^2}.$$

On ne fera pas attention au fait que dans ces modèles, le diamètre  $X$  ou  $Y$  pourrait être négatif (ceci arrive avec une probabilité quasiment nulle).

- (a) Identifier la loi de  $41 - X$ , et celle de  $Y$ .  
 (b) Quelles sont les espérance et variance de  $X$  ? celles de  $Y$  ? Commenter.  
 (c) Pour  $t \leq 41$ , que vaut  $\mathbb{P}[X \leq t]$  ? On donne  $e^{-4} \leq 1.9\%$ . Construire à partir de la variable  $X$  un intervalle de confiance de niveau au moins 98% pour le paramètre  $m_X = \mathbb{E}[X]$ .  
 (d) On donne  $\int_2^\infty e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \geq 2\%$ . Construire à partir de la variable  $Y$  un intervalle de confiance de niveau au plus 96% pour le paramètre  $m_Y = \mathbb{E}[Y]$ .  
 (e) Quel procédé permet de produire avec moins de 2% de pertes des boîtiers de pédalier utilisables ?





## Table des matières

Chapitre 1. Probabilités discrètes et dénombrement	1
1. Probabilités discrètes	1
2. Indépendance et conditionnement	3
3. Problèmes de dénombrement	7
Exercices	13
Chapitre 2. Variables aléatoires discrètes	17
1. Notion de variable aléatoire	17
2. Espérance et variance	19
3. Suites d'expériences de Bernoulli	24
Exercices	29
Chapitre 3. Variables aléatoires continues	33
1. Description des variables continues	33
2. Loix uniformes et loix exponentielles	39
3. Loix gaussiennes et loix du $\chi^2$	41
Exercices	49
Chapitre 4. Sommes de variables indépendantes	53
1. Généralités sur les familles de variables aléatoires	53
2. Convolution de deux loix	62
3. Loi des grands nombres	65
4. Théorème central limite	69
Exercices	75
Chapitre 5. Introduction aux statistiques	79
1. Estimateurs paramétriques	79
2. Intervalles de confiance	85
3. Tests d'hypothèses statistiques	87
Exercices	91