

# Chapitre VIII

## Probabilités

La théorie mathématique des probabilités naît au XVI<sup>e</sup> siècle sous l'impulsion de Jérôme Cardan puis de Blaise Pascal qui analysent les jeux de hasard. Des avancées majeures sont ensuite réalisées par Kolmogorov au début du XX<sup>e</sup> siècle, qui fait la connexion entre la théorie de la mesure de Borel, l'intégration de Lebesgue et les probabilités, donnant à ces dernières des fondements incontestés.

Dans tout ce chapitre nous supposons importée la bibliothèque `numpy.random` sous la forme :

```
import numpy.random as rd
```

## 1. Espaces probabilisés

### 1.1 Ensembles dénombrables

Le cours de première année s'est restreint aux univers finis ; cette année, nous allons étendre nos connaissances à certains univers infinis, *mais pas à tous* : seuls les plus « simples » des univers infinis seront abordés, les univers dits *dénombrables* c'est-à-dire ceux qui peuvent être mis en bijection avec  $\mathbb{N}$ .

**DÉFINITION.** — On dit d'un ensemble  $X$  qu'il est :

- fini lorsqu'il existe un entier  $n \in \mathbb{N}$  tel que  $X$  est en bijection avec  $\llbracket 1, n \rrbracket$  ;
- dénombrable lorsqu'il est en bijection avec  $\mathbb{N}$  ;
- discret s'il est fini ou dénombrable.

Si  $X$  est un ensemble dénombrable, il existe donc une bijection  $\phi : \mathbb{N} \rightarrow X$ . En posant pour tout  $n \in \mathbb{N}$ ,  $x_n = \phi(n)$  il devient possible de définir  $X$  en *extension*, c'est-à-dire sous la forme :  $X = \{x_n \mid n \in \mathbb{N}\}$ .

**Exemple.** L'ensemble  $2\mathbb{N}$  des entiers pairs est dénombrable, puisqu'il peut être défini en extension :  $2\mathbb{N} = \{2n \mid n \in \mathbb{N}\}$ , ce qui correspond à la bijection  $\phi : \mathbb{N} \rightarrow 2\mathbb{N}$ ,  $\phi(n) = 2n$ .

Il en est bien entendu de même de l'ensemble  $2\mathbb{N} + 1$  des entiers impairs :  $2\mathbb{N} + 1 = \{2n + 1 \mid n \in \mathbb{N}\}$ .

Plus généralement, on dispose du résultat suivant :

**PROPOSITION 1.1** — Toute partie d'un ensemble dénombrable est finie ou dénombrable.

Par exemple, l'ensemble  $\mathcal{P}$  des nombres premiers est infini (vous avez du démontrer ceci en première année) donc dénombrable puisqu'inclus dans  $\mathbb{N}$ . Il existe donc une suite  $(p_n)$  telle que  $\mathcal{P} = \{p_n \mid n \in \mathbb{N}\}$ .

**PROPOSITION 1.2** — Soit  $X$  un ensemble dénombrable et  $Y$  un ensemble fini ou dénombrable. Alors  $X \cup Y$  est dénombrable.

**COROLLAIRE** —  $\mathbb{Z}$  est un ensemble dénombrable.

**COROLLAIRE** — La réunion d'un nombre fini de parties discrètes est discrète.

**PROPOSITION 1.3** — Soit  $X$  un ensemble dénombrable et  $Y$  un ensemble non vide fini ou dénombrable. Alors le produit cartésien  $X \times Y$  est dénombrable.

**COROLLAIRE** —  $\mathbb{Q}$  est un ensemble dénombrable.

**COROLLAIRE** — Le produit cartésien d'un nombre fini de parties discrètes est discret.

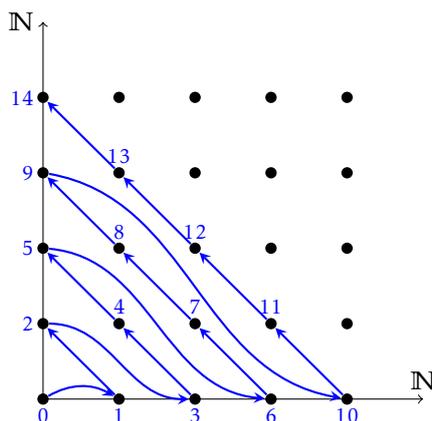


FIGURE 1 –  $\mathbb{N} \times \mathbb{N}$  est dénombrable car on peut énumérer ses éléments.

Jusqu'à présent, nous n'avons vu que des ensembles dénombrables, pour la bonne et simple raison qu'il est plus facile de prouver qu'un ensemble est dénombrable que de prouver qu'il ne l'est pas. C'est Cantor qui le premier a donné des exemples d'ensembles non dénombrables, en utilisant une méthode qui maintenant porte son nom : *l'argument de la diagonale de Cantor*. Nous admettrons le résultat suivant :

**THÉORÈME 1.4** — *L'ensemble  $\mathcal{P}(\mathbb{N})$  des parties de  $\mathbb{N}$  n'est pas dénombrable.  $\mathbb{R}$  n'est pas dénombrable. L'ensemble  $\{0, 1\}^{\mathbb{N}}$  des suites à valeurs dans  $\{0, 1\}$  n'est pas dénombrable.*

Les ensembles cités sont d'une certaine manière « trop gros » pour être dénombrables.

Terminons par un résultat qui nous sera utile par la suite :

**PROPOSITION 1.5** — *Soit  $(x_n)$  une suite de réels positifs telle que  $\sum x_n$  converge, et  $\phi : \mathbb{N} \rightarrow \mathbb{N}$  une bijection. Alors la série  $\sum x_{\phi(n)}$  converge et  $\sum_{n=0}^{+\infty} x_{\phi(n)} = \sum_{n=0}^{+\infty} x_n$ .*

Cette proposition montre que pour une série convergente à terme général positif, l'ordre dans lequel on somme les éléments n'influe pas sur la convergence ni sur la valeur de la somme. Ceci nous permettra désormais, lorsque  $X = \{x_i \mid i \in I\}$  est un ensemble fini ou dénombrable de réels positifs, de noter  $\sum_{i \in I} x_i$  la somme de ces éléments sans avoir besoin de préciser l'ordre de sommation, avec la convention  $\sum_{i \in I} x_i = +\infty$  dans le cas d'une somme infinie divergente.

## 1.2 Expérience aléatoire et univers

**DÉFINITION.** — *On appelle expérience aléatoire une expérience qui, reproduite dans des conditions identiques, peut conduire à des résultats différents non prévisibles à l'avance. L'ensemble des résultats possibles de cette expérience est appelé univers et est classiquement noté  $\Omega$ .*

**Exemples.** Examinons tout d'abord quelques expériences aléatoires et l'univers qui leur est associé :

- on lance trois dés à 6 faces. Dans ce cas, on choisira pour univers  $\Omega = \llbracket 1, 6 \rrbracket^3$ .
- on lance une pièce de monnaie jusqu'à obtenir Face. Ici pourra choisir  $\Omega = \mathbb{N} \cup \{+\infty\}$  si on choisit de représenter une expérience par le nombre d'essais infructueux.
- on casse une baguette de bois en trois et on mesure les longueurs des trois morceaux. En fixant à 1 la longueur de la baguette, l'univers peut être représenté par  $\Omega = \{(x, y, z) \in ]0, 1[^3 \mid x + y + z = 1\}$ .

Le premier exemple correspond à un univers fini, le second à un univers dénombrable, le troisième à un univers non dénombrable. Nous travaillerons essentiellement avec des univers discrets, autrement dit finis ou

dénombrables.

On observera que la description de l'univers ne nous indique pas la façon dont l'expérience est réalisée : les dés, la pièce, sont-ils pipés ou non ? Suivant quel protocole la baguette est-elle brisée ? Ce sont ces informations qui vont conditionner le choix de la probabilité que nous allons associer à cet univers.

### 1.3 Tribu et événements

Considérons un univers  $\Omega$ . Lorsque ce dernier est fini, on appelle *événement* toute partie de  $\Omega$ . En conjonction avec le vocabulaire de la théorie des ensembles, ont été définies les notions suivantes :

- un événement est dit *élémentaire* si c'est un singleton ;
- l'événement *certain* est l'événement  $\Omega$  ;
- l'événement *impossible* est l'événement  $\emptyset$  ;
- l'événement *NON A*, *contraire* de l'événement  $A$ , est l'événement  $\bar{A} = \Omega \setminus A$  (le résultat de l'expérience n'appartient pas à  $A$ ) ;
- si  $A$  et  $B$  sont des événements, l'événement *A ET B* est l'événement  $A \cap B$  (le résultat de l'expérience se trouve dans  $A$  et dans  $B$ ) ;
- si  $A$  et  $B$  sont des événements, l'événement *A OU B* est l'événement  $A \cup B$  (le résultat de l'expérience se trouve dans  $A$  ou dans  $B$ ) ;
- les événements  $A$  et  $B$  sont dits *incompatibles* lorsque  $A \cap B = \emptyset$  (le résultat de l'expérience ne peut se trouver à la fois dans  $A$  et dans  $B$ ) ;
- Si  $A$  et  $B$  sont deux événements, on dit que *A entraîne B* lorsque  $A \subset B$  (si le résultat de l'expérience se trouve dans  $A$ , il se trouve aussi dans  $B$ ).

**Exemple.** Considérons le lancer de trois dés associé à l'univers  $\Omega = \llbracket 1, 6 \rrbracket^3$ .

$A = \{(x, y, z) \in \Omega \mid x + y + z \leq 10\}$  est l'événement : « la somme des trois dés est inférieure ou égale à 10 ».

$B = \{(x, y, z) \in \Omega \mid x + y + z \geq 10\}$  est l'événement : « la somme des trois dés est supérieure ou égale à 10 ».

$A \cup B$  est l'événement certain ;  $A \cap B$  est l'événement : « la somme des trois dés est égale à 10 ». Enfin,  $\text{NON } A$  est l'événement « la somme des trois dés est strictement supérieure à 10 » donc  $\text{NON } A$  entraîne  $B$ .

Une fois la notion d'événement définie, l'étape suivante dans la construction d'un espace probabilisé consiste à définir une probabilité  $\mathbb{P}(A)$  mesurant la chance de réalisation d'un événement  $A$ . Or lorsque l'univers  $\Omega$  est infini, il n'est en général pas possible de définir cette probabilité pour toutes les parties de  $\Omega$  ; il faut se restreindre à un sous-ensemble  $\mathcal{A}$  de  $\mathcal{P}(\Omega)$  qu'on appelle une *tribu*, et qui en quelque sorte contient les événements dont on pourra mesurer la probabilité de réussite.

Plus formellement nous adopterons la définition suivante :

**DÉFINITION.** — Si  $\Omega$  est un ensemble, on appelle *tribu* sur  $\Omega$  une partie  $\mathcal{A}$  de  $\mathcal{P}(\Omega)$  vérifiant :

- $\Omega \in \mathcal{A}$  (l'événement certain appartient à la tribu) ;
- pour tout  $A \in \mathcal{A}$ , l'événement contraire  $\bar{A}$  appartient à  $\mathcal{A}$  ;
- $\mathcal{A}$  est stable par réunion dénombrable, c'est-à-dire que si  $(A_n)_{n \in \mathbb{N}}$  est une suite d'éléments de  $\mathcal{A}$  alors  $\bigcup_{n \in \mathbb{N}} A_n$  appartient à  $\mathcal{A}$ .

Désormais, le terme d'événement désignera un élément d'une tribu  $\mathcal{A}$ , supposée définie précédemment.

**PROPOSITION 1.6** — Si  $\mathcal{A}$  est une tribu sur l'univers  $\Omega$ , alors :

- $\emptyset \in \mathcal{A}$  (l'événement impossible appartient à la tribu) ;
- si  $A$  et  $B$  sont deux événements de la tribu  $\mathcal{A}$ , il en est de même de  $A \cup B$  et de  $A \cap B$  ;
- $\mathcal{A}$  est stable par intersection dénombrable, c'est-à-dire que si  $(A_n)_{n \in \mathbb{N}}$  est une suite d'éléments de  $\mathcal{A}$  alors  $\bigcap_{n \in \mathbb{N}} A_n$  appartient à  $\mathcal{A}$ .

**Exemple.**  $\{\emptyset, \Omega\}$  est une tribu, appelée *tribu triviale* puisqu'elle ne mesure que deux événements : l'événement certain et l'événement impossible.

**Exemple.** À l'inverse,  $\mathcal{P}(\Omega)$  est la tribu la plus fine qui soit. Cependant, à l'exception des univers finis ou dénombrables, cette tribu ne peut engendrer que des espaces probabilisés sans intérêt.

**Exemple.** Considérons de nouveau l'expérience consistant à jeter une pièce jusqu'à obtenir Face, mais choisissons cette fois l'univers  $\Omega = \{P, F\}^{\mathbb{N}^*}$  (autrement dit, dans l'univers des possibles on joue à Pile ou Face indéfiniment). Cet univers n'est pas dénombrable, il est donc nécessaire de définir une tribu sur laquelle on pourra ensuite définir une probabilité. Compte tenu du problème qui nous intéresse on admet l'existence d'une tribu  $\mathcal{A}$  dans laquelle « Face apparaît pour la première fois au  $n^e$  tirage » est un événement noté  $A_n$ .

Compte tenu des propriétés des tribus, l'événement  $A = \bigcup_{n \in \mathbb{N}^*} A_n$  appartient à  $\mathcal{A}$  (il s'agit de l'événement « Face apparaît au moins une fois ») ainsi que l'événement contraire  $\bar{A}$  (« la pièce tombe indéfiniment sur Pile »). Tous les événements nécessaires à l'étude de l'expérience sont bien présents dans la tribu.

### EXERCICE 1

Soit  $\mathcal{A}$  une tribu de  $\mathbb{R}$  contenant toutes les demi-droites  $[a, +\infty[$ ,  $a \in \mathbb{R}$ . Montrer que cette tribu contient tous les intervalles de  $\mathbb{R}$ .

## 1.4 Définition d'une probabilité

Nous sommes maintenant en mesure de donner la définition générale d'une probabilité.

**DÉFINITION.** — Soit  $\Omega$  un univers et  $\mathcal{A}$  une tribu sur  $\Omega$ . On appelle probabilité sur  $(\Omega, \mathcal{A})$  une application  $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$  vérifiant :

- $\mathbb{P}(\Omega) = 1$  ;
- pour toute suite dénombrable  $(A_n)_{n \in \mathbb{N}}$  d'événements de  $\mathcal{A}$  deux-à-deux incompatibles la série  $\sum \mathbb{P}(A_n)$  converge, et  $\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n)$ .

On appelle espace probabilisé le triplet  $(\Omega, \mathcal{A}, \mathbb{P})$  constitué d'un univers, d'une tribu sur  $\Omega$  et d'une probabilité sur  $(\Omega, \mathcal{A})$ .

Commençons par observer que les propriétés sur les univers finis qui ont été établies dans le cours de première année restent vérifiées :

**PROPOSITION 1.7** — Une probabilité vérifie les propriétés suivantes :

- $\mathbb{P}(\emptyset) = 0$  ;
- si  $A$  et  $B$  sont deux événements incompatibles,  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$  ;
- si  $A$  est un événement,  $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$  ;
- si  $A$  et  $B$  sont deux événements,  $\mathbb{P}(A \cap B) + \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$  ;
- si  $A \subset B$  sont deux événements, alors  $\mathbb{P}(A) \leq \mathbb{P}(B)$ .

**Exemple.** Lorsque l'univers  $\Omega$  est fini et  $\mathcal{A} = \mathcal{P}(\Omega)$  il existe une unique probabilité, appelée probabilité uniforme telle que pour tout  $\omega \in \Omega$ ,  $\mathbb{P}(\{\omega\}) = \frac{1}{\text{card } \Omega}$ . Dans ce cas, pour tout événement  $A$ ,  $\mathbb{P}(A) = \frac{\text{card } A}{\text{card } \Omega}$ .

Cette expression montre que dans le cas de la probabilité uniforme, le calcul des probabilités se ramène à du calcul combinatoire. Par exemple, dans le cas du problème du lancer de trois dés non pipés associé à l'univers  $\Omega = \llbracket 1, 6 \rrbracket^3$ , notons  $A$  l'événement « la somme des trois dés est inférieure ou égale à 10 » :

$$A = \{(x, y, z) \in \Omega \mid x + y + z \leq 10\}.$$

Observons l'équivalence :  $x + y + z \leq 10 \iff (7-x) + (7-y) + (7-z) \geq 11 \iff (7-x) + (7-y) + (7-z) > 10$ . Elle traduit le fait que l'application  $\begin{pmatrix} A & \longrightarrow & \bar{A} \\ (x, y, z) & \longmapsto & (7-x, 7-y, 7-z) \end{pmatrix}$  est une bijection. Ainsi,  $\text{card } A = \text{card } \bar{A}$  et donc  $\mathbb{P}(A) = 1/2$ .

**Exemple.** Revenons maintenant sur l'expérience consistant à jeter une pièce de monnaie jusqu'à obtenir Face. Nous avons vu qu'on pouvait définir une tribu  $\mathcal{A}$  sur l'univers  $\Omega = \{P, F\}^{\mathbb{N}^*}$  qui contient tous les événements  $A_n$  : « Face apparaît pour la première fois au  $n^e$  tirage ».

Si on note  $p \in ]0, 1[$  la probabilité pour la pièce de tomber sur Face, les éléments de l'univers sont des suites d'épreuves de Bernoulli de paramètre  $p$ , et les éléments de  $A_n$  les suites qui débutent par  $n-1$  échecs suivis d'une réussite donc  $\mathbb{P}(A_n) = (1-p)^{n-1}p$ .

Les événements  $A_n$  étant deux à deux incompatibles ( $i \neq j \implies A_i \cap A_j = \emptyset$ ) on a  $\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n) =$

$\sum_{n=1}^{+\infty} (1-p)^{n-1}p = p \times \frac{1}{1-(1-p)} = 1$ . L'événement  $A = \bigcup_{n \in \mathbb{N}^*} A_n$  (« Face apparaît au moins une fois ») vérifie  $\mathbb{P}(A) = 1$ ,

l'événement  $\bar{A}$  (« la pièce tombe indéfiniment sur Pile ») vérifie  $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A) = 0$ .

L'événement  $\bar{A}$  est dit « quasi-impossible » : bien qu'il soit un événement envisageable (il n'est pas égal à l'événement impossible  $\emptyset$ ) sa probabilité est nulle. À l'inverse, l'événement  $A$  est dit « quasi-certain ».

Voyons maintenant quelques résultats propres aux univers infinis :

**THÉORÈME 1.8 (limite monotone)** — Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé. Alors :

– pour toute suite d'événements  $(A_n)$  croissante au sens de l'inclusion ( $A_n \subset A_{n+1}$ ), la suite  $(\mathbb{P}(A_n))$  converge, et

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim \mathbb{P}(A_n);$$

– pour toute suite d'événements  $(A_n)$  décroissante au sens de l'inclusion ( $A_{n+1} \subset A_n$ ), la suite  $(\mathbb{P}(A_n))$  converge, et

$$\mathbb{P}\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim \mathbb{P}(A_n).$$

**PROPOSITION 1.9 (sous-additivité)** — Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé. Pour toute suite d'événements  $(A_n)$ ,

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n=0}^{+\infty} \mathbb{P}(A_n) \text{ (cette somme peut éventuellement être égale à } +\infty \text{)}.$$

**Remarque.** Lorsque la suite  $(A_n)$  n'est pas monotone au sens de l'inclusion, on peut néanmoins appliquer le théorème de la limite monotone à la suite des « union partielles » ou la suite des « intersections partielles ».

En effet, la suite  $B_n = \bigcup_{k=0}^n A_k$  est croissante donc  $\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \lim \mathbb{P}(B_n)$ .

De même la suite  $C_n = \bigcap_{k=0}^n A_k$  est décroissante donc  $\mathbb{P}\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \mathbb{P}\left(\bigcap_{n \in \mathbb{N}} C_n\right) = \lim \mathbb{P}(C_n)$ .

## EXERCICE 2

Soit  $(A_n)$  une suite d'événements quasi-certains d'un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ . Montrer que l'événement  $\bigcap_{n \in \mathbb{N}} A_n$  est quasi-certain.

## ■ Probabilité sur un univers dénombrable

Considérons maintenant un univers dénombrable  $\Omega$ , que l'on peut donc décrire par extension :  $\Omega = \{\omega_n \mid n \in \mathbb{N}\}$ . Nous allons prouver le résultat suivant, qui montre qu'il est toujours possible de définir une probabilité sur  $(\Omega, \mathcal{P}(\Omega))$  à partir de la valeur de  $\mathbb{P}$  sur les singletons :

**THÉORÈME 1.10** — Soit  $(p_n)$  une suite de réels positifs telle que la série  $\sum p_n$  converge et  $\sum_{n=0}^{+\infty} p_n = 1$ . Alors il existe une unique probabilité  $\mathbb{P}$  sur  $(\Omega, \mathcal{P}(\Omega))$  telle que pour tout  $n \in \mathbb{N}$ ,  $\mathbb{P}(\{\omega_n\}) = p_n$ .

**Exemple.** Soit  $\theta > 0$  et  $p_n = e^{-\theta} \frac{\theta^n}{n!}$ . Il est facile de vérifier que  $0 \leq p_n \leq 1$  et que  $\sum_{n=0}^{+\infty} p_n = e^{-\theta} \sum_{n=0}^{+\infty} \frac{\theta^n}{n!} = 1$ . La suite  $(p_n)$  définit donc une probabilité sur  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$  en posant  $\mathbb{P}(\{n\}) = e^{-\theta} \frac{\theta^n}{n!}$ , appelée *loi de Poisson de paramètre  $\theta$* . Nous aurons l'occasion d'y revenir.

### EXERCICE 3

- Soit  $\mathbb{P}$  une probabilité sur  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ . Montrer que  $\lim \mathbb{P}(\{n\}) = 0$ .
- Soit  $(a_n)$  une suite strictement décroissante de réels positifs de limite nulle. Déterminer une constante  $\lambda > 0$  pour qu'il existe une probabilité  $\mathbb{P}$  sur  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$  vérifiant :  $\mathbb{P}(\llbracket n, +\infty \rrbracket) = \lambda a_n$ .

## 1.5 Conditionnement et indépendance

Dans toute la suite du cours,  $(\Omega, \mathcal{A}, \mathbb{P})$  désigne un espace probabilisé.

### ■ Probabilité conditionnelle

**DÉFINITION.** — Si  $A$  et  $B$  sont deux événements tels que  $\mathbb{P}(B) > 0$ , on appelle probabilité conditionnelle de  $A$  sachant  $B$  le réel  $\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ , réel qu'on pourra aussi noter  $\mathbb{P}(A | B)$ .

**THÉORÈME 1.11** —  $\mathbb{P}_B$  est une probabilité sur  $(\Omega, \mathcal{A})$ .

**Remarque.** On dispose donc de l'égalité  $\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A | B)$  lorsque  $\mathbb{P}(B) \neq 0$ . Lorsque  $\mathbb{P}(B) = 0$ , on peut observer que cette égalité garde un sens (celui de «  $0 = 0$  ») même si  $\mathbb{P}(A | B)$  n'est pas formellement défini puisque  $A \cap B \subset B \Rightarrow 0 \leq \mathbb{P}(A \cap B) \leq \mathbb{P}(B) = 0$ .

Si  $A$  et  $B$  sont deux événements quelconques, la formule  $\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A | B)$  est appelée *formule des probabilités composées*.

**DÉFINITION.** — On appelle système complet d'événements toute famille  $(B_i)_{i \in I}$  finie ou dénombrable d'événements deux-à-deux incompatibles ( $i \neq j \Rightarrow B_i \cap B_j = \emptyset$ ) et telle que  $\bigcup_{i \in I} B_i = \Omega$ .

En d'autres termes, la famille  $(B_i)_{i \in I}$  constitue une partition finie ou dénombrable de  $\Omega$ .

**THÉORÈME 1.12 (formule des probabilités totales)** — Soit  $A$  un événement et  $(B_i)_{i \in I}$  un système complet d'événements. Alors

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(B_i)\mathbb{P}(A | B_i).$$

**Remarque.** La formule reste valable lorsque  $\sum_{i \in I} \mathbb{P}(B_i) = 1$ , autrement dit lorsque l'événement  $\bigcup_{i \in I} B_i$  est quasi-certain.

**PROPOSITION 1.13 (Formule de Bayes)** — Soit  $(B_i)_{i \in I}$  un système complet d'événements tel que pour tout  $i \in I$ ,

$$\mathbb{P}(B_i) > 0, \text{ et } A \text{ un événement tel que } \mathbb{P}(A) > 0. \text{ Alors } \mathbb{P}(B_i | A) = \frac{\mathbb{P}(B_i)\mathbb{P}(A | B_i)}{\sum_{j \in I} \mathbb{P}(B_j)\mathbb{P}(A | B_j)}.$$

**Remarque.** Cette formule est souvent utilisée lorsque le système complet est constitué des deux seuls événements  $B$  et  $\bar{B}$ . Dans ce cas, la formule devient :  $\mathbb{P}(B | A) = \frac{\mathbb{P}(B)\mathbb{P}(A | B)}{\mathbb{P}(B)\mathbb{P}(A | B) + \mathbb{P}(\bar{B})\mathbb{P}(A | \bar{B})}$ .

**EXERCICE 4**

On dispose de quatre dés à six faces, l'un d'entre eux étant pipé : il tombe sur 6 avec une probabilité égale à  $5/6$ .

On choisit au hasard un dé parmi les quatre et on le lance  $2n$  fois. On obtient  $n$  fois 6. Avec quelle probabilité le dé choisi est-il pipé ?

**Remarque.** La formule de Bayes a longtemps été appelée formule de probabilité des causes. Elle permet en effet de calculer la probabilité d'une cause (ici le fait d'avoir pris le dé pipé) connaissant celle de sa conséquence (le nombre de 6 obtenus).

**EXERCICE 5**

On dépose dans une urne vide une boule blanche puis on joue à Pile ou Face avec une pièce non pipée. Tant que la pièce retombe sur Pile, on ajoute une boule noire dans l'urne. Lorsqu'on obtient Face pour la première fois on tire au hasard une boule de l'urne. Celle-ci est blanche. Quelle est la probabilité qu'il n'y ait aucune boule noire dans l'urne ?

**■ Indépendance**

De manière informelle, deux événements  $A$  et  $B$  sont indépendants lorsque le fait de savoir que  $A$  est réalisé ne donne aucune information sur la réalisation de  $B$ , et réciproquement. Ainsi, lorsque  $\mathbb{P}(A) > 0$  et  $\mathbb{P}(B) > 0$  on souhaite que  $\mathbb{P}(A | B) = \mathbb{P}(A)$  et  $\mathbb{P}(B | A) = \mathbb{P}(B)$ , ce qui se traduit par  $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A)$  et  $\frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \mathbb{P}(B)$ . Ces deux égalités sont identiques, et pour pouvoir s'abstraire des hypothèses  $\mathbb{P}(A) > 0$  et  $\mathbb{P}(B) > 0$  on adoptera la définition suivante :

**DÉFINITION.** — Deux événements  $A$  et  $B$  sont dits indépendants lorsque  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ .

**PROPOSITION 1.14** — Si  $A$  et  $B$  sont indépendants, il en est de même de  $\bar{A}$  et  $B$ , de  $A$  et  $\bar{B}$ , de  $\bar{A}$  et  $\bar{B}$ .

La notion d'indépendance se généralise à une suite finie ou infinie d'événements de la manière suivante :

**DÉFINITION.** — Une famille finie ou dénombrable  $(A_i)_{i \in I}$  d'événements est dite indépendante lorsque pour tout entier  $p \leq \text{card } I$ , pour toute  $p$ -liste  $(i_1, \dots, i_p) \in I^p$  d'indices deux-à-deux distincts,  $\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_p}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_p})$  (on dit aussi que les événements  $A_i$  sont mutuellement indépendants).

On observera que cette définition est très délicate à mettre en œuvre. Ne serait-ce que pour trois événements  $A$ ,  $B$  et  $C$  il faut vérifier *chacune* des égalités :

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

En particulier, les trois dernières égalités, qui traduisent le fait que ces trois événements sont deux-à-deux indépendants, *ne sont pas suffisantes* pour s'assurer que les trois événements sont indépendants.

**EXERCICE 6**

Soit  $(A_n)$  une suite d'événements indépendants ; on pose  $A^* = \bigcap_{p \in \mathbb{N}} \bigcup_{n \geq p} A_n$ .

- Montrer que si la série  $\sum \mathbb{P}(A_n)$  converge alors  $\mathbb{P}(A^*) = 0$ .
- À l'inverse, montrer que si la série  $\sum \mathbb{P}(A_n)$  diverge alors  $\mathbb{P}(A^*) = 1$ .

**Remarque.** Le résultat ci-dessus, appelé Lemme de Borel-Cantelli, possède une conséquence amusante. Imaginons un singe tapant sur un clavier d'ordinateur (on suppose que la suite de ses frappes est indépendante) et considérons un mot quelconque, par exemple POLYTECHNIQUE. Notons  $A_1$  l'événement « ce mot est écrit lors des 13 premières frappes »,  $A_2$  l'événement : « ce mot est écrit lors des 13 frappes suivantes », etc. Par hypothèse ces événements sont indépendants et ont même probabilité  $p > 0$  de se réaliser. Ainsi, la suite  $\sum \mathbb{P}(A_n)$  diverge et d'après le lemme de Borel-Cantelli l'événement  $A^*$  se réalise quasi-certainement, ce qui signifie que le singe tapera une infinité de fois ce mot au cours des temps.

Bien entendu ceci est vrai pour tout mot de longueur finie, donc il est aussi quasi-certain que ce singe écrira à un moment ou un autre le cours que vous lisez en ce moment (ce qui ne signifie pas, bien entendu, que son auteur soit lui-même un singe).

## 2. Variables aléatoires

### 2.1 Définition d'une variable aléatoire

Jusqu'à présent, nous avons beaucoup parlé des événements, autrement dit adopté un point de vue *ensembliste* sur les probabilités. Nous allons maintenant changer de point de vue en choisissant un point de vue *fonctionnel* à l'aide de la notion de *variable aléatoire* qui, contrairement à ce que pourrait laisser supposer son nom, n'est pas une variable mais une fonction. De manière informelle, une variable aléatoire est une grandeur qui dépend du résultat de l'expérience ; ce peut être par exemple :

- le nombre de 6 obtenus dans un lancé de trois dés ;
- le temps d'attente avant d'obtenir Face dans un lancer de pièce ;
- la longueur du plus grand des deux morceaux lorsqu'on brise une baguette de bois en deux.

Pour chacun des trois exemples cités ci-dessus, nous allons réaliser un grand nombre d'expériences et calculer la moyenne de la variable aléatoire correspondante.

**Exemple.** Pour le lancer de trois dés :

```
n = 100000 # nombre d'expériences
v = 0      # somme des variables aléatoires

for _ in range(n):
    (a, b, c) = rd.randint(1, 7, 3)
    if a == 6:
        v += 1
    if b == 6:
        v += 1
    if c == 6:
        v += 1
print(v / n)
```

0.49998

En moyenne, on obtient un 6 tous les deux lancers.

**Exemple.** Pour le lancer d'une pièce pipée qui possède une probabilité  $p = 0,1$  de tomber sur Face :

```
def experience():
    s = 1
    while True:
        if rd.random() < .1:
            return s
        s += 1

n = 100000 # nombre d'expériences
v = 0      # somme des variables aléatoires

for _ in range(n):
    v += experience()
print(v / n)
```

10.02955

Il faut en moyenne 10 lancers avant que l'expérience se termine.

**Exemple.** Pour la longueur de plus grand morceau d'une baguette cassée en deux :

```
n = 100000 # nombre d'expériences
v = 0      # somme des variables aléatoires

for _ in range(n):
    x = rd.rand()
    v += max(x, 1 - x)
print(v / n)
```

```
0.7502040108780781
```

En moyenne, la longueur du plus grand des deux morceaux est égal aux  $3/4$  de la longueur du morceau initial.

Nous démontrerons avec le théorème 2.24 (la loi faible des grands nombres) que dans chacun des cas le résultat de ces expériences donne un résultat proche de ce qu'on appellera l'espérance de la variable aléatoire.

**DÉFINITION.** — Si  $(\Omega, \mathcal{A})$  est un espace probabilisable et  $E$  un ensemble, on appelle variable aléatoire toute fonction  $X : \Omega \rightarrow E$  telle que pour tout  $e \in E$ ,  $X^{-1}(\{e\}) \in \mathcal{A}$  (autrement dit,  $X^{-1}(\{e\})$  est un événement).

Lorsque  $E = \mathbb{R}$ , la variable aléatoire  $X$  sera dite réelle.

Lorsque  $X(\Omega)$  (l'ensemble des valeurs que peut prendre  $X$ ) est fini ou dénombrable, la variable aléatoire  $X$  sera dite discrète.

**Rappel.** La notation  $X^{-1}(\{e\})$  désigne l'image réciproque de  $e$ , c'est-à-dire l'ensemble des antécédents de  $e$  :  $X^{-1}(\{e\}) = \{\omega \in \Omega \mid X(\omega) = e\}$ .

**Exemples.**

– Pour l'expérience consistant à lancer trois dés et à compter le nombre de 6, nous pouvons choisir  $\Omega = \llbracket 1, 6 \rrbracket^3$ ,  $E = \mathbb{N}$  et  $X : \Omega \rightarrow \mathbb{N}$  définie par  $X(e_1, e_2, e_3) = \text{card}\{i \in \llbracket 1, 3 \rrbracket \mid e_i = 6\}$ .

$X(\Omega) = \{0, 1, 2, 3\}$  donc la variable aléatoire  $X$  est discrète (finie).

– Pour l'expérience consistant à lancer une pièce jusqu'à obtenir Face, nous pouvons choisir  $\Omega = \{P, F\}^{\mathbb{N}^*}$ ,  $E = \mathbb{N} \cup \{+\infty\}$  et  $X : \Omega \rightarrow E$  définie par  $X((u_n)) = \min\{n \in \mathbb{N}^* \mid u_n = F\}$ . Ici  $X$  est une variable aléatoire discrète (dénombrable).

– Pour l'expérience consistant à casser une baguette de deux pour mesurer le plus grand des deux morceaux, nous avons  $\Omega = ]0, 1[$ ,  $E = ]0, 1[$  et  $X(x) = \max(x, 1 - x)$ . Dans cet exemple,  $X$  n'est pas une variable aléatoire discrète car  $X(\Omega) = ]1/2, 1[$  n'est pas dénombrable.

Dans la suite de ce cours nous ne prendrons en considération que des variables aléatoires discrètes.

**PROPOSITION 2.1** — Lorsque  $X$  est une variable aléatoire discrète, pour tout  $U \subset X(\Omega)$ ,  $X^{-1}(U) \in \mathcal{A}$  (autrement dit,  $X^{-1}(U)$  est un événement).

**Remarque.** On introduit la notion de variable aléatoire pour s'intéresser aux chances de réalisation des valeurs de  $X$  plutôt qu'aux chances de réalisation des résultats de l'expérience. Autrement dit, cette notion permet d'une certaine façon d'« oublier » l'espace probabilisable  $(\Omega, \mathcal{A})$  (qui reste présent, mais dont on se contentera le plus souvent d'admettre son existence) au profit des valeurs prises par  $X$ .

Par la suite, l'événement  $X^{-1}(U)$  sera noté plus simplement  $\{X \in U\}$ , voire même  $(X \in U)$ .

Par exemple, pour le jeté de trois dés,  $\{X = 2\}$  désigne l'événement « deux des trois dés ont donné un 6 ». Pour le lancer d'une pièce jusqu'à obtenir Face,  $\{X \geq 3\}$  désigne l'événement « il a fallu au moins trois lancers avant d'obtenir un Face ».

L'intérêt du résultat précédent est que puisque  $\{X \in U\}$  est un événement, il est possible de lui associer une probabilité. Il s'agit du résultat suivant :

**THÉORÈME 2.2** — Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé et  $X : \Omega \rightarrow E$  une variable aléatoire discrète. Alors l'application  $\mathbb{P}_X : \mathcal{P}(X(\Omega)) \rightarrow [0, 1]$  définie par  $\mathbb{P}_X(U) = \mathbb{P}(X^{-1}(U)) = \mathbb{P}(\{X \in U\})$  est une probabilité sur  $(X(\Omega), \mathcal{P}(X(\Omega)))$ , appelée loi de la variable  $X$ , ou encore distribution de  $X$ .

**EXERCICE 7**

Une urne contient initialement une boule blanche et une boule noire. On tire au hasard une de ces boules, on note sa couleur et on la replace dans l'urne accompagnée d'une seconde boule de la même couleur. On réalise ce processus  $n$  fois, et on note  $X_n$  la variable aléatoire égale au nombre de boules blanches tirées durant ce processus. Déterminer la loi de  $X_n$ .

Ce résultat définit la loi de la variable aléatoire discrète  $X$  à partir de la loi de probabilité sur  $\Omega$ . Il existe une réciproque de ce résultat, que nous admettrons : il est possible de choisir a priori la loi de  $X$  et d'en déduire une probabilité  $\mathbb{P}$  sur  $\Omega$  de sorte que la loi de  $X$  soit celle qui découle de  $\mathbb{P}$ . De manière plus formelle :

**THÉORÈME 2.3** — Soit  $(\Omega, \mathcal{A})$  un espace probabilisable et  $X : \Omega \rightarrow E$  une variable aléatoire discrète. On note  $X(\Omega) = \{x_i \mid i \in I\}$  et on considère une famille discrète  $(p_i)_{i \in I}$  de réels positifs telle que  $\sum_{i \in I} p_i = 1$ . Alors il existe une probabilité  $\mathbb{P}$  sur  $(\Omega, \mathcal{A})$  telle que pour tout  $i \in I$ ,  $\mathbb{P}(\{X = x_i\}) = p_i$ .

L'intérêt de ce résultat est qu'il sera souvent suffisant de raisonner directement à partir de  $\mathbb{P}_X$  sans véritablement avoir besoin d'explicitier formellement l'espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ .

**■ Fonction de répartition d'une variable aléatoire**

**DÉFINITION.** — Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé et  $X : \Omega \rightarrow \mathbb{R}$  une variable aléatoire réelle. La fonction de répartition de  $X$  est la fonction  $F_X : \mathbb{R} \rightarrow [0, 1]$  définie par :  $\forall x \in \mathbb{R}, F_X(x) = \mathbb{P}(X \leq x)$ .

**PROPOSITION 2.4** — La fonction  $F_X$  possède les propriétés suivantes :

- $F_X$  est une fonction croissante ;
- $\lim_{x \rightarrow -\infty} F_X(x) = 0$  et  $\lim_{x \rightarrow +\infty} F_X(x) = 1$  ;
- pour tout  $a \in \mathbb{R}$ ,  $F_X$  est continue à droite et possède une limite à gauche en  $a$  ;
- pour tout  $a \in \mathbb{R}$ ,  $F_X(a^+) - F_X(a^-) = \mathbb{P}(X = a)$ .

**2.2 Lois discrètes classiques**

Certaines lois interviennent régulièrement dans les problèmes de probabilités ; il est donc intéressant de les connaître afin de ne pas refaire à chaque fois les mêmes calculs. Nous allons maintenant passer en revue celles que vous devez connaître.

**■ Loi uniforme**

L'expérience type consiste à considérer une urne contenant  $n$  boules numérotées de 1 à  $n$  et à effectuer un tirage équiprobable. La variable aléatoire  $X$  est le numéro de la boule obtenue.

**DÉFINITION.** — Soit  $n \in \mathbb{N}^*$ . On dit qu'une variable aléatoire réelle  $X$  suit une loi uniforme de paramètre  $n$  lorsque

$$X(\Omega) = \llbracket 1, n \rrbracket \text{ et si pour tout } k \in \llbracket 1, n \rrbracket, \mathbb{P}(X = k) = \frac{1}{n}. \text{ On note dans ce cas } X \hookrightarrow \mathcal{U}(n).$$

**■ Loi de Bernoulli**

L'expérience type consiste à tirer dans une urne contenant une proportion  $p$  de boules blanches. On note  $X$  la variable aléatoire égale à 1 si on tire une boule blanche, et 0 sinon. On peut aussi tirer à pile ou face avec une pièce truquée ayant la probabilité  $p$  de tomber sur Face et poser  $X = 0$  lorsque la pièce tombe sur Pile, et  $X = 1$  lorsque la pièce tombe sur Face.

**DÉFINITION.** — Soit  $p \in ]0, 1[$ . On dit qu'une variable aléatoire réelle  $X$  suit une loi de Bernoulli de paramètre  $p$  lorsque  $X(\Omega) = \{0, 1\}$  et  $\mathbb{P}(X = 0) = 1 - p$ ,  $\mathbb{P}(X = 1) = p$ . On note dans ce cas  $X \hookrightarrow \mathcal{B}(p)$ .

**Remarque.** Pour des raisons de symétrie il est fréquent d'introduire la quantité  $q = 1 - p$ .

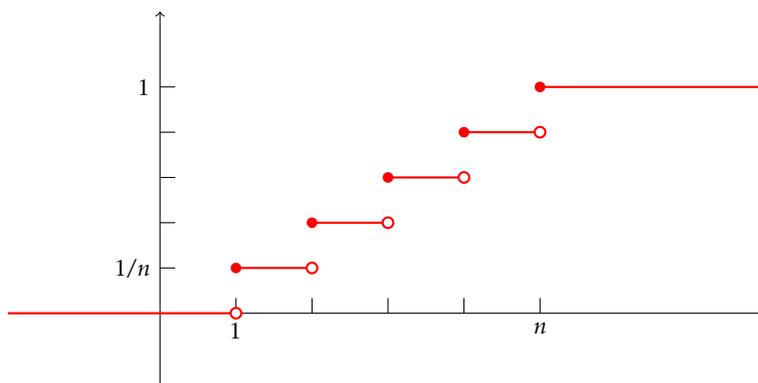


FIGURE 2 – La fonction de répartition d’une variable suivant une loi uniforme.

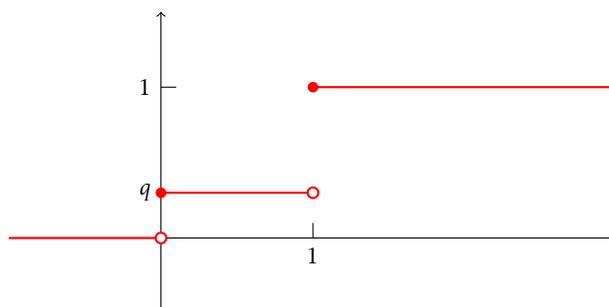
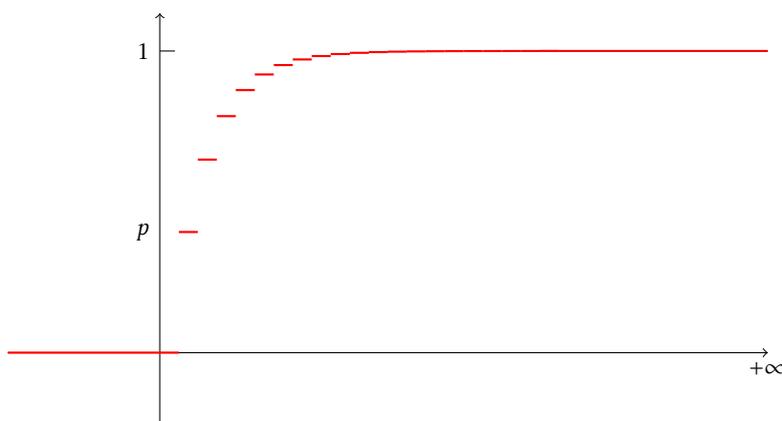


FIGURE 3 – La fonction de répartition d’une variable suivant une loi de Bernoulli.

### ■ Loi géométrique

L’expérience type consiste en une succession infinie d’expériences de Bernoulli de paramètre  $p$ . On note  $X$  le rang du premier succès.

**DÉFINITION.** — Soit  $p \in ]0, 1[$ . On dit qu’une variable aléatoire réelle  $X$  suit une loi géométrique de paramètre  $p$  lorsque  $X(\Omega) = \mathbb{N}^*$  et pour tout  $k \in \mathbb{N}^*$ ,  $\mathbb{P}(X = k) = pq^{k-1}$  avec  $q = 1 - p$ . On note dans ce cas  $X \hookrightarrow \mathcal{G}(p)$ .

FIGURE 4 – La fonction de répartition d’une variable suivant une loi géométrique ( $p = 0,4$ ).

**PROPOSITION 2.5** — Soit  $X \hookrightarrow \mathcal{G}(p)$ . Alors pour tout  $(m, n) \in (\mathbb{N}^*)^2$ ,  $\mathbb{P}(X > m + n \mid X > n) = \mathbb{P}(X > m)$ .

Ce résultat traduit le fait qu'une loi géométrique est *sans mémoire* : après  $n$  expériences les variables  $X - n$  et  $X$  suivent la même loi : les expériences passées n'influent pas sur les succès futurs. C'est la raison pour laquelle le fait qu'un nombre ne soit pas sorti depuis longtemps au loto n'augmente pas la probabilité qu'il sorte au tirage suivant.

### EXERCICE 8

Soit  $X$  une variable aléatoire sans mémoire à valeurs dans  $\mathbb{N}^*$ . Montrer que  $X$  suit une loi géométrique.

## ■ Loi binomiale

L'expérience type consiste à effectuer  $n$  fois une expérience de Bernoulli et à noter  $X$  le nombre de succès.

**DÉFINITION.** — Soit  $n \in \mathbb{N}^*$  et  $p \in ]0, 1[$ . On dit qu'une variable aléatoire réelle  $X$  suit une loi binomiale de paramètres  $(n, p)$  lorsque  $X(\Omega) = \llbracket 0, n \rrbracket$  et pour tout  $k \in \llbracket 0, n \rrbracket$ ,  $\mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k}$  avec  $q = 1 - p$ . On note dans ce cas  $X \hookrightarrow \mathcal{B}(n, p)$ .

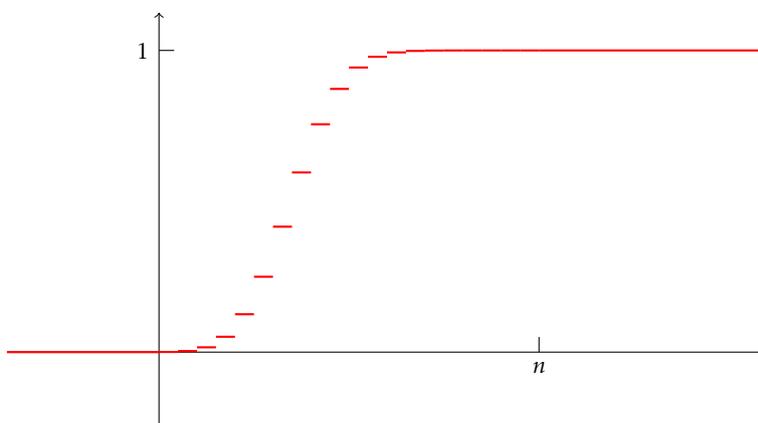


FIGURE 5 – La fonction de répartition d'une variable suivant une loi binomiale ( $n = 40$ ,  $p = 0,4$ ).

## ■ Loi de Poisson

La dernière loi que nous allons définir est un peu différente des précédentes, dans le sens où elle ne correspond pas à la modélisation d'une expérience précise mais apparaît (dans un certain sens) comme limite des lois binomiales.

**THÉORÈME 2.6 (loi des événements rares)** — Soit  $(X_n)$  une suite de variables aléatoires réelles telle que pour tout  $n \in \mathbb{N}$ ,  $X_n \hookrightarrow \mathcal{B}(n, p_n)$ . On suppose  $p_n \sim \frac{\lambda}{n}$  avec  $\lambda > 0$ . Alors pour tout  $k \in \mathbb{N}$ ,  $\lim_{n \rightarrow +\infty} \mathbb{P}(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ .

**DÉFINITION.** — Soit  $\lambda > 0$ . On dit qu'une variable aléatoire réelle  $X$  suit une loi de Poisson de paramètre  $\lambda$  lorsque

$$X(\Omega) = \mathbb{N} \text{ et pour tout } k \in \mathbb{N}, \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}. \text{ On note dans ce cas } X \hookrightarrow \mathcal{P}(\lambda).$$

**Remarque.** Concrètement, ce résultat affirme que si des événements indépendants ont une très faible probabilité d'apparition, alors la fonction de répartition, lors d'expériences répétées, est distribuée selon une loi de Poisson. Ceci explique pourquoi les lois de Poisson sont souvent associées à la modélisation d'événements rares. Dans la pratique, on estime souvent qu'on peut utiliser l'approximation de  $\mathcal{B}(n, p)$  par  $\mathcal{P}(\lambda)$  (avec  $\lambda = np$ ) dès lors que  $n \geq 50$  et  $np < 10$ . Dans le cadre de cette approximation les calculs numériques s'en trouvent grandement simplifiés.

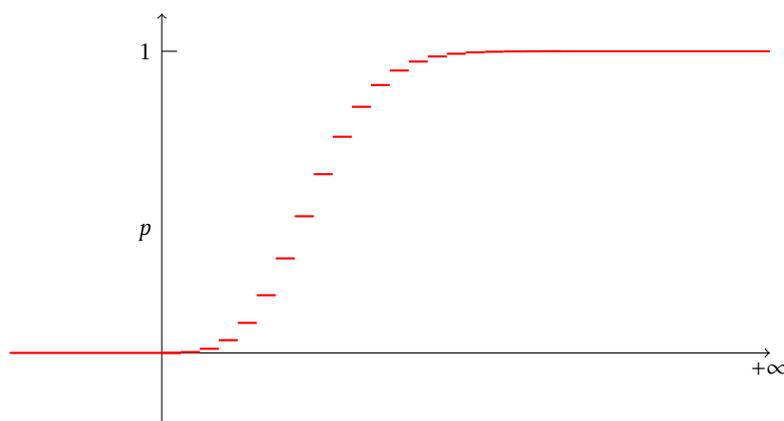


FIGURE 6 – La fonction de répartition d'une variable suivant une loi de Poisson ( $\lambda = 10$ ).

**Exemple.** Un central téléphonique possède 5 lignes. On estime à  $n = 1\,200$  le nombre de personnes susceptibles d'appeler le standard sur une journée de huit heures, les appels étant répartis uniformément durant la journée et d'une durée de deux minutes en moyenne.

On souhaite calculer la probabilité que le standard soit saturé à un instant donné. Pour cela, on note  $X$  la variable aléatoire égale au nombre de personnes en train de téléphoner à un instant donné et on cherche à

calculer  $\mathbb{P}(X > 5) = 1 - \sum_{k=0}^5 \mathbb{P}(X = k)$ .

Un appel au standard à un instant donné est une éventualité de probabilité  $p = \frac{1}{8 \times 30} = \frac{1}{240}$ . La variable aléatoire  $X$  suit donc une loi binomiale de paramètres  $(n, p)$ , et on est dans le cadre de l'approximation par une loi de Poisson de paramètre  $\lambda = np = 5$ . Effectuons le calcul avec ces deux lois :

```
from scipy.stats import binom, poisson

print(1-sum([binom.pmf(k, 1200, 1/240) for k in range(6)]))
print(1-sum([poisson.pmf(k, 5) for k in range(6)]))
```

```
0.384039090245462
0.38403934516693705
```

Les deux formules donnent effectivement des réponses très proches : de l'ordre de 38,4%.

### EXERCICE 9

Soit  $X$  une variable aléatoire suivant une loi de Poisson de paramètre  $\lambda > 0$ . Est-il plus probable que la valeur de  $X$  soit paire ou impaire ?

## 2.3 Couple de variables aléatoires

**DÉFINITION.** — Si  $X$  et  $Y$  sont deux variables aléatoires sur un même espace probabilisable  $(\Omega, \mathcal{A})$ , on note  $(X, Y)$  la variable aléatoire  $\omega \mapsto (X(\omega), Y(\omega))$ . On appelle loi conjointe de  $X$  et de  $Y$  la loi de  $(X, Y)$ , autrement dit la loi  $\mathbb{P}_{(X, Y)}$  définie par :

$$\forall (x, y) \in X(\Omega) \times Y(\Omega), \quad \mathbb{P}_{(X, Y)}(x, y) = \mathbb{P}(X = x \text{ et } Y = y).$$

À l'inverse, si  $(X, Y)$  est un couple de variables aléatoires, on appelle lois marginales de  $(X, Y)$  les lois de  $X$  et de  $Y$ .

Connaissant la loi conjointe de  $X$  et  $Y$  il est facile de retrouver les lois marginales :

$$\mathbb{P}(X = x) = \sum_{y \in Y(\Omega)} \mathbb{P}(X = x \text{ et } Y = y) \quad \text{et} \quad \mathbb{P}(Y = y) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x \text{ et } Y = y).$$

À l'inverse, la connaissance des lois marginales ne permet pas en général de déterminer la loi conjointe, car en général les événements  $\{X = x\}$  et  $\{Y = y\}$  n'ont aucune raison d'être indépendants. C'est la raison pour laquelle on adopte la définition suivante :

**DÉFINITION.** — Deux variables aléatoires  $X$  et  $Y$  sur un même espace probabilisable  $(\Omega, \mathcal{A})$  sont dites indépendantes lorsque pour tout  $x \in X(\Omega)$  et tout  $y \in Y(\Omega)$  les événements  $\{X = x\}$  et  $\{Y = y\}$  sont indépendants. On a dans ce cas :  $\mathbb{P}(X = x \text{ et } Y = y) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y)$ .

**PROPOSITION 2.7** — Soient  $X$  et  $Y$  deux variables aléatoires indépendantes d'un même espace probabilisable  $(\Omega, \mathcal{A})$ . Alors pour toutes parties  $A$  dans  $X(\Omega)$  et  $B$  dans  $Y(\Omega)$  on a :  $\mathbb{P}(X \in A \text{ et } Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B)$ .

**PROPOSITION 2.8** — Soient  $X$  et  $Y$  deux variables aléatoires indépendantes d'un même espace probabilisable  $(\Omega, \mathcal{A})$ , et  $f$  et  $g$  deux fonctions de  $\mathbb{R}$  dans  $\mathbb{R}$ . Alors les variables aléatoires  $f(X)$  et  $g(Y)$  sont indépendantes.

**PROPOSITION 2.9** — Soit  $X \hookrightarrow \mathcal{P}(\lambda)$  et  $Y \hookrightarrow \mathcal{P}(\mu)$  deux variables aléatoires indépendantes. Alors  $X + Y \hookrightarrow \mathcal{P}(\lambda + \mu)$ .

**Remarque.** Lorsque deux variables  $X$  et  $Y$  ne sont pas indépendantes, on utilise une probabilité conditionnelle pour calculer la probabilité de l'événement  $\{X = x \text{ et } Y = y\}$  :  $\mathbb{P}(X = x \text{ et } Y = y) = \mathbb{P}(X = x \mid Y = y) \cdot \mathbb{P}(Y = y)$ .

### EXERCICE 10

Soit  $X$  une variable aléatoire qui suit une loi de Poisson de paramètre  $\lambda$ ,  $Y$  une variable aléatoire qui suit une loi binomiale  $\mathcal{B}(n, p)$  lorsque  $X = n$ , et  $Z = X - Y$ .

Déterminer les lois de  $Y$  et de  $Z$ . Les variables  $Y$  et  $Z$  sont-elles indépendantes ?

### Indépendance mutuelle

**DÉFINITION.** — Soit  $(X_i)_{i \in I}$  une famille finie ou dénombrable de variables aléatoires d'un même espace probabilisable  $(\Omega, \mathcal{A})$ . On dit que ces variables sont mutuellement indépendantes si et seulement si pour toute  $p$ -liste  $(i_1, \dots, i_p) \in I^p$  d'indices deux-à-deux distincts, et toute  $p$ -liste  $(x_{i_1}, \dots, x_{i_p}) \in X_{i_1}(\Omega) \times \dots \times X_{i_p}(\Omega)$ , les événements  $\{X_{i_k} = x_{i_k}\}$  sont indépendants.

À l'instar de l'indépendance d'une famille finie ou dénombrable d'événements, cette définition est particulièrement malcommode à vérifier. En particulier, on notera qu'il n'est pas équivalent de se contenter de vérifier que les variables sont deux-à-deux indépendantes.

**Exemple.** Un jeu de pile ou face infini peut être modélisé par une suite de variables aléatoires de Bernoulli mutuellement indépendantes.

Pour finir, nous admettrons le résultat suivant, très utile bien qu'il soit en principe hors programme :

**THÉORÈME 2.10 (lemme des coalitions)** — Soient  $X_1, \dots, X_n$  une famille de  $n$  variables aléatoires mutuellement indépendantes, et  $p \in \llbracket 1, n \rrbracket$ . Soit  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  et  $g : \mathbb{R}^{n-p} \rightarrow \mathbb{R}$  deux fonctions. Alors les variables  $f(X_1, \dots, X_p)$  et  $g(X_{p+1}, \dots, X_n)$  sont indépendantes.

## 2.4 Espérance

Lorsque l'univers  $\Omega$  est fini, l'espérance d'une variable aléatoire réelle est la moyenne des valeurs qu'elle est susceptible de prendre pondérées par la probabilité d'apparition de ces valeurs :  $\mathbb{E}(X) = \sum_{\omega \in \Omega} \mathbb{P}(\omega)X(\omega) =$

$\sum_{x \in X(\Omega)} x\mathbb{P}(X = x)$ . Lorsque  $\Omega$  est infini et  $X(\Omega) = \{x_n \mid n \in \mathbb{N}\}$  dénombrable, cette somme devient  $\sum_{n=0}^{+\infty} x_n \mathbb{P}(X = x_n)$ ,

et il peut se poser un problème de convergence. Aussi adopterons-nous la définition suivante :

**DÉFINITION.** — Soit  $X$  une variable aléatoire discrète :  $X(\Omega) \subset \{x_n \mid n \in \mathbb{N}\}$ . On dit que  $X$  est d'espérance finie

lorsque la série  $\sum x_n \mathbb{P}(X = x_n)$  converge absolument, et dans ce cas on appelle espérance de  $X$  la quantité

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

**Remarque.** Nous admettons que cette quantité est indépendante de l'ordre dans lequel on énumère les valeurs prises par  $X$  (ceci résulte de l'hypothèse de convergence absolue).

**Exemple.** Considérons une fois de plus le problème du lancer de pièce jusqu'à obtenir Face. Nous avons montré que si  $X$  désigne la variable aléatoire qui compte le nombre de lancers nécessaires nous avons  $\mathbb{P}(X = n) = p(1-p)^{n-1}$  où  $p$  désigne la probabilité d'obtenir un Face lors d'un lancer.

Puisque  $1-p \in ]0, 1[$  la série  $\sum_{n \geq 1} n(1-p)^{n-1}$  converge donc  $E$  est d'espérance finie, et  $\mathbb{E}(X) = \sum_{n=1}^{+\infty} np(1-p)^{n-1} = \frac{p}{(1-(1-p))^2} = \frac{1}{p}$ . Nous pouvons observer que lorsque nous avons réalisé un grand nombre d'expériences pour calculer le nombre moyen de lancers nécessaires (voir page 8), nous avons obtenu un résultat proche de cette valeur. La justification théorique de cette constatation sera donnée par la *loi faible des grands nombres* (théorème 2.24).

Notons qu'il existe une formule équivalente pour calculer l'espérance d'une variable aléatoire à valeurs dans  $\mathbb{N}$  :

**PROPOSITION 2.11** — Si  $X$  est une variable aléatoire à valeurs dans  $\mathbb{N}$  et d'espérance finie alors  $\mathbb{E}(X) = \sum_{n=1}^{+\infty} \mathbb{P}(X \geq n)$ .

On admettra les résultats suivants :

**PROPOSITION 2.12** — Soient  $X$  et  $Y$  deux variables aléatoires d'espérances finies. Alors :

- (i) pour tout  $\lambda \in \mathbb{R}$ ,  $\lambda X + Y$  est d'espérance finie, et  $\mathbb{E}(\lambda X + Y) = \lambda \mathbb{E}(X) + \mathbb{E}(Y)$  (linéarité de l'espérance);
- (ii) et si, de plus,  $X$  et  $Y$  sont indépendantes, alors  $XY$  est d'espérance finie et  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ .

**COROLLAIRE** — Soient  $X$  et  $Y$  deux variables aléatoires d'espérances finies. Alors :

- (i) si  $X$  est à valeurs positives,  $\mathbb{E}(X) \geq 0$  (positivité de l'espérance);
- (ii) si  $X \leq Y$  alors  $\mathbb{E}(X) \leq \mathbb{E}(Y)$  (croissance de l'espérance).

Enfin, nous admettons le

**THÉORÈME 2.13 (de transfert)** — Si  $X(\Omega) \subset \{x_n \mid n \in \mathbb{N}\}$  et  $f$  une application à valeurs réelles définie sur  $X(\Omega)$ , alors  $f(X)$  est d'espérance finie si et seulement si la série  $\sum f(x_n) \mathbb{P}(X = x_n)$  converge absolument, et dans ce cas,

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

## ■ Espérances des lois usuelles

Toutes les lois que nous avons étudiées à la section 2.2 sont d'espérance finie, et la valeur de leur espérance doit être connue.

**Loi uniforme** Si  $X \hookrightarrow \mathcal{U}(n)$ , alors  $\mathbb{E}(X) = \frac{n+1}{2}$ .

**Loi de Bernoulli** Si  $X \hookrightarrow \mathcal{B}(p)$ , alors  $\mathbb{E}(X) = p$ .

**Loi géométrique** Si  $X \hookrightarrow \mathcal{G}(p)$  alors  $\mathbb{E}(X) = \frac{1}{p}$ .

**Loi binomiale** Si  $X \hookrightarrow \mathcal{B}(n, p)$  alors  $\mathbb{E}(X) = np$ .

**Loi de Poisson** Si  $X \hookrightarrow \mathcal{P}(\lambda)$  alors  $\mathbb{E}(X) = \lambda$ .

**Remarque.** La loi binomiale étant la somme de  $n$  loi de Bernoulli (indépendantes) on a bien  $\mathbb{E}(\mathcal{B}(n, p)) = n \times \mathbb{E}(\mathcal{B}(p))$ .

## 2.5 Variance et écart type

**THÉORÈME 2.14** — Si la variable aléatoire  $X^2$  est d'espérance finie, il en est de même de  $X$ . Dans ce cas, on appelle variance de  $X$  la quantité  $\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$ , et écart type la quantité  $\sigma(X) = \sqrt{\mathbb{V}(X)}$ .

**PROPOSITION 2.15 (Formule de Koenig-Huyghens)** — Lorsque  $X^2$  est d'espérance finie,  $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ .

**PROPOSITION 2.16** — Si  $a$  et  $b$  sont deux réels et  $X$  une variable aléatoire réelle telle que  $X^2$  soit d'espérance finie, alors  $\mathbb{V}(aX + b) = a^2\mathbb{V}(X)$  et  $\sigma(aX + b) = |a|\sigma(X)$ .

### EXERCICE 11

Soit  $X$  une variable aléatoire réelle telle que  $X^2$  admette une espérance. Quelle est la valeur minimale de la fonction  $t \mapsto \mathbb{E}((X - t)^2)$ ?

### ■ Variance des lois usuelles

Toutes les lois que nous avons étudiées à la section 2.2 possèdent une variance et une espérance dont les valeurs doivent être connues.

**Loi uniforme** Si  $X \hookrightarrow \mathcal{U}(n)$ , alors  $\mathbb{V}(X) = \frac{n^2 - 1}{12}$ .

**Loi de Bernoulli** Si  $X \hookrightarrow \mathcal{B}(p)$ , alors  $\mathbb{V}(X) = pq = p(1 - p)$ .

**Loi géométrique** Si  $X \hookrightarrow \mathcal{G}(p)$  alors  $\mathbb{V}(X) = \frac{q}{p^2} = \frac{1 - p}{p^2}$ .

**Loi binomiale** Si  $X \hookrightarrow \mathcal{B}(n, p)$  alors  $\mathbb{V}(X) = npq$ .

**Loi de Poisson** Si  $X \hookrightarrow \mathcal{P}(\lambda)$  alors  $\mathbb{V}(X) = \lambda$ .

### ■ Moment d'une variable aléatoire

Espérance et variance se généralisent avec la notion de *moment* : étant donné un entier  $r \in \mathbb{N}$ , on dit qu'une variable aléatoire réelle  $X$  possède un moment d'ordre  $r$  lorsque  $\mathbb{E}(X^r)$  existe, et un moment centré d'ordre  $r$  lorsque  $\mathbb{E}((X - \mathbb{E}(X))^r)$  existe. Ainsi, l'espérance est un moment d'ordre 1 et la variance un moment centré d'ordre 2.

**Remarque (Variable centrée réduite).** Si  $X$  est une variable aléatoire possédant un moment d'ordre 2, la variable aléatoire  $Y = \frac{X - \mathbb{E}(X)}{\sigma(X)}$  possède une espérance nulle (on dit qu'elle est *centrée*) et un écart type égal à 1 (on dit qu'elle est *réduite*).

Il existe de nombreuses analogies entre l'ensemble des variables aléatoires possédant un moment d'ordre 2 et les fonctions de carré intégrables. En particulier, on admettra que si  $X$  et  $Y$  possèdent des moments d'ordre 2, alors  $XY$  possède un moment d'ordre 1. Ce qui nous conduit à :

**THÉORÈME 2.17 (Inégalité de Cauchy-Schwarz)** — Si  $X$  et  $Y$  possèdent des moments d'ordre 2, alors  $XY$  possède un moment d'ordre 1, et  $\mathbb{E}(XY) \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$ .

**Remarque.** Il y a égalité dans l'inégalité de Cauchy-Schwarz si et seulement s'il existe  $(\lambda, \mu) \in \mathbb{R}^2$  tel que  $\lambda X + \mu Y$  est quasi-sûrement nul, autrement dit lorsque  $\mathbb{P}(\lambda X + \mu Y = 0) = 1$ .

## 2.6 Covariance

Nous avons démontré à la proposition 2.12 que lorsque  $X$  et  $Y$  sont deux variables aléatoires indépendantes,  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ . Lorsque  $X$  et  $Y$  ne sont pas indépendantes, on peut considérer que la quantité  $\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$  mesure le « défaut d'indépendance » de ces deux variables. Pour des raisons pratiques (liées à l'inégalité de Cauchy-Schwarz, voir le théorème 2.25), nous allons introduire cette quantité sous une forme légèrement différente. En effet,

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

Ceci conduit à la définition suivante :

**DÉFINITION.** — Soient  $X$  et  $Y$  deux variables aléatoires réelles. Sous réserve d'existence on appelle covariance de  $X$  et de  $Y$  la quantité  $\boxed{\text{cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))}$ .

Le théorème 2.17 nous permet d'énoncer :

**PROPOSITION 2.18** — Si  $X$  et  $Y$  sont deux variables aléatoires réelles possédant un moment d'ordre 2 alors  $\text{cov}(X, Y)$  existe.

et le calcul réalisé ci-dessus nous permet d'affirmer :

**PROPOSITION 2.19** — Lorsque  $X$  et  $Y$  sont deux variables aléatoires réelles indépendantes possédant un moment d'ordre 2, alors  $\text{cov}(X, Y) = 0$ . On dira que  $X$  et  $Y$  ne sont pas corrélées.

**Attention.** La réciproque de ce résultat est fautive : deux variables aléatoires peuvent ne pas être corrélées sans pour autant être indépendantes.

**PROPOSITION 2.20 (propriétés de la covariance)** — Soient  $X$ ,  $Y$  et  $Z$  trois variables aléatoires réelles possédant des moments d'ordre 2. Alors :

- $\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ ;
- $\text{cov}(X, Y) = \text{cov}(Y, X)$ ;
- $\text{cov}(X, 1) = 0$ ;
- $\forall (a, b) \in \mathbb{R}^2, \text{cov}(X, aY + bZ) = a\text{cov}(X, Y) + b\text{cov}(X, Z)$ .

**THÉORÈME 2.21** — Soient  $X$  et  $Y$  deux variables aléatoires réelles possédant des moments d'ordre 2. Alors

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

En particulier, lorsque ces deux variables aléatoires sont indépendantes,  $\boxed{\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)}$ .

**Remarque.** Cette formule se généralise au cas de  $n$  variables aléatoires :  $\mathbb{V}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \mathbb{V}(X_k) + 2 \sum_{i < j} \text{cov}(X_i, X_j)$ .

En particulier on retiendra le :

**COROLLAIRE** — Lorsque  $X_1, \dots, X_n$  sont des variables aléatoires possédant des moments d'ordre 2 et deux-à-deux indépendantes,  $\mathbb{V}(X_1 + \dots + X_n) = \mathbb{V}(X_1) + \dots + \mathbb{V}(X_n)$ .

## 2.7 Inégalités de concentration

Dans la théorie des probabilités, les inégalités de concentration fournissent des bornes sur la probabilité qu'une variable aléatoire dévie d'une certaine valeur (généralement l'espérance de cette variable aléatoire).

**THÉORÈME 2.22 (inégalité de Markov)** — Soit  $X$  une variable aléatoire positive possédant une espérance, et  $a > 0$ .

Alors  $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$ .

**THÉORÈME 2.23 (inégalité de Bienaymé-Tchebychev)** — Soit  $X$  une variable aléatoire réelle admettant un moment

d'ordre 2, et  $\alpha > 0$ . Alors  $\mathbb{P}(|X - \mathbb{E}(X)| \geq \alpha) \leq \frac{\mathbb{V}(X)}{\alpha^2} = \frac{\sigma(X)^2}{\alpha^2}$ .

Que signifie cette inégalité? La probabilité calculée mesure le risque de s'écarter de l'espérance d'une quantité supérieure à  $\alpha$ . La majorant obtenu montre que plus l'écart type est faible, plus ce risque est négligeable. Ainsi, un écart type faible caractérise une faible dispersion autour de l'espérance. À l'inverse, un écart type important dénote une grande dispersion des valeurs.

## ■ Loi faible des grands nombres

Lorsque nous avons voulu évaluer expérimentalement l'espérance du nombre de lancers nécessaire pour obtenir Face lors d'un lancer de pièce (voir page 8), nous avons utilisé une simulation informatique : nous avons réalisé un grand nombre d'expériences indépendantes en calculant à chaque fois la valeur prise par  $X$ , puis nous avons calculé la moyenne. Nous avons effectivement obtenu une valeur très proche de l'espérance théorique. Le théorème qui suit donne une justification à cet état de fait :

**THÉORÈME 2.24 (loi faible des grands nombres)** — Soit  $(X_n)_{n \geq 1}$  une succession de variables aléatoires deux-à-deux indépendantes et de même loi admettant un moment d'ordre 2. On pose  $S_n = X_1 + X_2 + \dots + X_n$ ,  $m = \mathbb{E}(X_k)$  et  $\sigma = \sigma(X_k)$  (puisque les  $X_i$  suivent la même loi ces quantités ne dépendent pas de  $k$ ). Alors pour tout  $\epsilon > 0$ ,

$$\mathbb{P}\left(\left|\frac{1}{n}S_n - m\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}.$$

**Remarque.** Avec les mêmes hypothèses on en déduit que :  $\lim_{n \rightarrow +\infty} \mathbb{P}\left(\left|\frac{1}{n}S_n - m\right| \geq \epsilon\right) = 0$ .

En d'autres termes, plus on réalise un grand nombre d'expériences, plus le risque que la moyenne s'écarte de l'espérance de plus de  $\epsilon$  est faible.

**Exemple.** Dans l'exemple numérique de la page 8 nous avons pris  $n = 100\,000$  et nous avons  $\mathbb{E}(X) = \frac{1}{p} = 10$  et  $\sigma(X) = \frac{\sqrt{q}}{p} = \sqrt{90}$  (car  $X \hookrightarrow \mathcal{G}(p)$ ). Pour  $\epsilon = 0,1$  nous avons  $\frac{\sigma^2}{n\epsilon^2} = 0,09$  donc il y a plus de 91% de chance que le résultat obtenu diffère de l'espérance théorique de moins de 0,1.

## 2.8 Coefficient de corrélation

**THÉORÈME 2.25** — Soient  $X$  et  $Y$  deux variables aléatoires possédant des moments d'ordre 2. Alors

$$|\text{cov}(X, Y)| \leq \sigma(X)\sigma(Y)$$

**DÉFINITION.** — Lorsque  $\sigma(X) > 0$  et  $\sigma(Y) > 0$ , on appelle coefficient de corrélation linéaire la quantité

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

D'après le résultat précédent on a  $\rho(X, Y) \in [-1, 1]$ .

### Interprétation du coefficient de corrélation

Nous l'avons déjà dit, la covariance mesure d'une certaine façon le « défaut d'indépendance » de deux variables aléatoires. cette mesure est absolue, tandis que le coefficient de corrélation est une mesure relative. Plus précisément, le coefficient de corrélation marque la tendance des deux variables à posséder une relation de dépendance *linéaire*. Plus  $|\rho(X, Y)|$  est proche de 1, plus les variables  $X$  et  $Y$  auront tendance à évoluer conjointement de manière linéaire.

**Exemple.** Supposons qu'il existe  $(a, b) \in \mathbb{R}^2$ ,  $a \neq 0$ , tel que  $Y = aX + b$ . Supposons de plus  $\mathbb{V}(X) > 0$ . Alors  $\mathbb{V}(Y) = a^2\mathbb{V}(X)$  et  $\text{cov}(X, Y) = a\text{cov}(X, X) + b\text{cov}(X, 1) = a\mathbb{V}(X)$  donc :

$$\rho(X, Y) = \frac{a\mathbb{V}(X)}{\sqrt{a^2\mathbb{V}(X)^2}} = \frac{a}{|a|} = \begin{cases} 1 & \text{si } a > 0 \\ -1 & \text{si } a < 0 \end{cases}$$

le signe de  $\rho(X, Y)$  indique donc le sens (identique ou inverse) de variation conjoint.

À l'inverse, si  $X$  et  $Y$  sont indépendantes, nous avons déjà dit que  $\rho(X, Y) = 0$ . Mais attention, deux variables peuvent être dépendantes et avoir un coefficient de corrélation nul ! Ce peut être le cas par exemple lorsque la dépendance existe mais n'est pas linéaire. Ainsi, un coefficient de corrélation nul indique simplement que la dépendance des variables, si elle existe, est très faiblement linéaire.

### EXERCICE 12

Soient  $X$  et  $Y$  deux variables aléatoires linéaires possédant un moment d'ordre 2. On suppose  $\mathbb{V}(X) > 0$ . Déterminer  $(a, b) \in \mathbb{R}^2$  minimisant la quantité  $\mathbb{E}((Y - aX - b)^2)$ .

**Remarque.** La droite d'équation  $y = ax + b$  est appelée la droite de régression linéaire de  $X$  et  $Y$  ; elle donne la « meilleure » expression linéaire de  $Y$  en fonction de  $X$ .

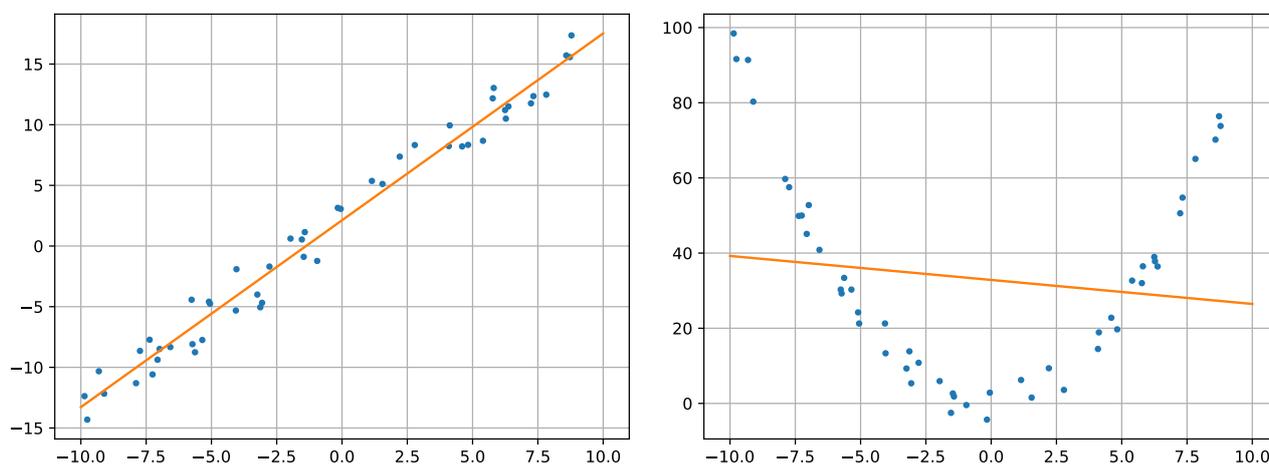


FIGURE 7 – Deux variables fortement corrélées à gauche, faiblement corrélées à droite

J'ai représenté figure 7 le résultat de deux séries d'expériences. À gauche, deux variables aléatoires fortement corrélées ( $\rho(X, Y) \approx 0,99$ ) ont été simulées et la droite de régression linéaire tracée. À droite, la même chose mais avec deux variables aléatoires très faiblement corrélées ( $\rho(X, Y) \approx -0,14$ ). La droite de régression linéaire n'a pas grand sens dans ce deuxième cas. Cependant, il apparaît clairement que bien que non corrélées, ces deux variables ne sont pas indépendantes. Leur relation de dépendance n'est tout simplement pas linéaire.

## 2.9 Séries génératrices

**DÉFINITION.** — Soit  $X$  une variable aléatoire à valeurs dans  $\mathbb{N}$ . On appelle série génératrice de  $X$  la série entière  $G_X(t) = \mathbb{E}(t^X) = \sum_{n \in \mathbb{N}} \mathbb{P}(X = n)t^n$ .

Pourquoi s'intéresser à cette série entière ? Nous savons que les coefficients d'une série entière de rayon de convergence  $R > 0$  sont définis de manière unique, aussi pouvons-nous affirmer que si  $R > 0$  la série génératrice d'une variable aléatoire caractérise cette dernière. On peut donc espérer utiliser la souplesse d'utilisation des séries entières pour calculer plus facilement certaines caractéristiques de  $X$ , telles l'espérance ou la variance.

**LEMME** — La série génératrice  $G_X$  de  $X$  est au moins définie sur  $[-1, 1]$ .

**COROLLAIRE** — Si deux variables aléatoires ont même série génératrice sur  $]-1, 1[$  alors ces deux variables aléatoires suivent la même loi.

Supposons maintenant  $R > 1$ .  $G_X$  est de classe  $\mathcal{C}^\infty$  sur  $]-R, R[$  et  $G'_X(t) = \sum_{n=1}^{+\infty} n\mathbb{P}(X = n)t^{n-1}$ . On voit immédiatement qu'en posant  $t = 1$  on obtient  $G'_X(1) = \mathbb{E}(X)$ . Nous admettrons que ce résultat reste vrai lorsque  $R = 1$  à condition que  $G_X$  soit dérivable en 1, ce qui nous permet d'énoncer le

**THÉORÈME 2.26** —  $X$  admet un moment d'ordre 1 (une espérance) si et seulement si  $G_X$  est dérivable en 1, et dans ce cas,  $\mathbb{E}(X) = G'_X(1)$ .

Voyons comment obtenir la variance. Si on suppose toujours  $R > 1$  nous avons  $G''_X(1) = \sum_{n=1}^{+\infty} n(n-1)\mathbb{P}(X = n)$ .

$$\begin{aligned} \text{Par ailleurs, } \mathbb{V}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \sum_{n=1}^{+\infty} n^2\mathbb{P}(X = n) - \mathbb{E}(X)^2 = \sum_{n=1}^{+\infty} n(n-1)\mathbb{P}(X = n) + \sum_{n=1}^{+\infty} n\mathbb{P}(X = n) - \mathbb{E}(X)^2 \\ &= G''_X(1) + G'_X(1) - G'_X(1)^2. \end{aligned}$$

Nous admettrons que sous réserve d'existence cette formule reste vraie lorsque  $R = 1$ , ce qui donne le

**THÉORÈME 2.27** —  $X$  admet un moment d'ordre 2 si et seulement si  $G_X$  est deux fois dérivable en 1, et dans ce cas,  $\mathbb{V}(X) = G''_X(1) + G'_X(1) - G'_X(1)^2$ .

## ■ Séries génératrices les lois usuelles

**Loi uniforme** Si  $X \hookrightarrow \mathcal{U}(n)$ , alors  $G_X(t) = \frac{1}{n}(t + t^2 + \dots + t^n)$ .

**Loi de Bernoulli** Si  $X \hookrightarrow \mathcal{B}(p)$ , alors  $G_X(t) = pt + q$  avec  $q = 1 - p$ .

**Loi géométrique** Si  $X \hookrightarrow \mathcal{G}(p)$  alors  $G_X(t) = \frac{pt}{1 - qt}$  avec  $q = 1 - p$ .

**Loi binomiale** Si  $X \hookrightarrow \mathcal{B}(n, p)$  alors  $G_X(t) = (pt + q)^n$  avec  $q = 1 - p$ .

**Loi de Poisson** Si  $X \hookrightarrow \mathcal{P}(\lambda)$  alors  $G_X(t) = e^{\lambda t - \lambda}$ .

### EXERCICE 13

À l'aide des séries génératrices ci-dessus, retrouver l'espérance et la variance des lois usuelles.

## ■ Série génératrice d'une somme de deux variables aléatoires indépendantes

Considérons pour finir deux variables aléatoires indépendantes  $X$  et  $Y$  à valeurs dans  $\mathbb{N}$ .

Nous avons  $G_X(t) = \sum_{n=0}^{+\infty} \mathbb{P}(X = n)t^n$  et  $G_Y(t) = \sum_{n=0}^{+\infty} \mathbb{P}(Y = n)t^n$  et

$$G_{X+Y}(t) = \sum_{n=0}^{+\infty} \mathbb{P}(X + Y = n)t^n = \sum_{n=0}^{+\infty} \left( \sum_{k=0}^n \mathbb{P}(X = k \text{ et } Y = n - k) \right) t^n = \sum_{n=0}^{+\infty} \left( \sum_{k=0}^n \mathbb{P}(X = k)\mathbb{P}(Y = n - k) \right) t^n.$$

On reconnaît un produit de Cauchy donc

$$G_{X+Y}(t) = \left( \sum_{n=0}^{+\infty} \mathbb{P}(X = n)t^n \right) \cdot \left( \sum_{n=0}^{+\infty} \mathbb{P}(Y = n)t^n \right) = G_X(t)G_Y(t).$$

Nous avons prouvé le :

**THÉORÈME 2.28** — Si  $X$  et  $Y$  sont deux variables aléatoires indépendantes à valeur entières, alors

$$G_{X+Y}(t) = G_X(t)G_Y(t).$$

**Exemple.**

- Si  $X_i \hookrightarrow \mathcal{B}(p)$  ( $1 \leq i \leq n$ ) sont des variables aléatoires indépendantes et  $S$  leur somme, alors  $S \hookrightarrow \mathcal{B}(n, p)$ .
- Si  $X \hookrightarrow \mathcal{B}(m, p)$  et  $Y \hookrightarrow \mathcal{B}(n, p)$  sont deux variables aléatoires indépendantes, alors  $X + Y \hookrightarrow \mathcal{B}(m + n, p)$ .  
En effet,  $(pt + q)^m \cdot (pt + q)^n = (pt + q)^{m+n}$ .
- Si  $X \hookrightarrow \mathcal{P}(\lambda)$  et  $Y \hookrightarrow \mathcal{P}(\mu)$  sont deux variables aléatoires indépendantes, alors  $X + Y \hookrightarrow \mathcal{P}(\lambda + \mu)$ .  
En effet,  $e^{\lambda t - \lambda} \cdot e^{\mu t - \mu} = e^{(\lambda + \mu)t - (\lambda + \mu)}$ .