Chapitre 9

Probabilités

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

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

1. Espaces probabilisés

1.1 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* Ω.

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

Exemple 1. On lance trois dés à 6 faces non pipés. Dans ce cas, $\Omega = [\![1,6]\!]^3$. Le script suivant simule 5 expériences consécutives :

```
In [1]: rd.randint(1, 7, (5, 3))
Out[1]: array([[6, 3, 2],
               [4, 5, 3],
               [2, 4, 4],
               [6, 5, 6],
               [4, 5, 2]])
```

Chacune des lignes de ce tableau est le résultat d'une expérience aléatoire.

Remarque. Dans cet univers les trois dés sont *discernables* : les triplets $(6,3,2)$ et $(6,2,3)$ sont deux résultats distincts. Quel univers faudrait-il choisir dans le cas de trois dés indiscernables ?

Exemple 2. On lance une pièce de monnaie jusqu'à obtenir Face. Ici on peut choisir $\Omega = \{\mathrm{P}^n\mathrm{F} \mid n \in \mathbb{N}\} \cup \{\mathrm{P}^\infty\}$, où $\mathrm{P}^n\mathrm{F}$ représente l'expérience durant laquelle n lancers ont été infructueux avant de voir apparaître Face, et P^∞ l'expérience durant laquelle Face n'apparaît jamais. Mais on peut aussi choisir $\Omega = \mathbb{N} \cup \{+\infty\}$ si on choisit de représenter une expérience par le nombre d'essais infructueux.

Le script qui suit modélise quelques expériences utilisant une pièce pipée qui possède une probabilité $p = 0{,}1$ de tomber sur Face :

```
def experience():
    while True:
        if rd.random() < .1:
            print('F')
            break
        else:
            print('P', end='')
```

```
In [1]: experience()
PPPPPPPPPPPPPPF
In [2]: experience()
PPF
In [3]: experience()
PPPPPPPPF
```

Exemple 3. 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 $\Omega = \left\{(x,y,z) \in \left]0,1\right[^3 \;\middle|\; x+y+z=1\right\}$. La simulation de cette expérience dépend du protocole utilisé pour choisir les deux points de rupture. On peut choisir indépendamment les abscisses de ces deux points :

```python
def protocole1():
    (x, y) = rd.rand(2)
    if x > y:
        x, y = y, x
    return x, y - x, 1 - y
```

mais on peut aussi commencer par briser la baguette en deux, choisir arbitrairement l'un des deux morceaux et briser ce dernier en deux :

```python
def protocole2():
    x = rd.rand()
    if rd.randint(2) == 0:  # on brise le morceau de longueur x
        y = rd.rand() * x
        return y, x - y, 1 - x
    else:                   # on brise le morceau de longueur 1 - x
        y = rd.rand() * (1 - x)
        return x, y, 1 - x - y
```

```
In [1]: protocole1()
Out[1]: (0.2688354135233324, 0.21486900687116939, 0.51629557960549821)
In [2]: protocole2()
Out[2]: (0.6120125719029146, 0.13890758563818661, 0.2490798424588988)
```

Ces deux protocoles ne sont pas identiques, comme on peut le constater en cherchant à assembler ces trois morceaux pour former un triangle. Nous admettrons (ce n'est pas difficile à démontrer) que trois morceaux de longueurs $a < b < c$ forment un triangle si et seulement si $c < a + b$ (illustration figure 1).

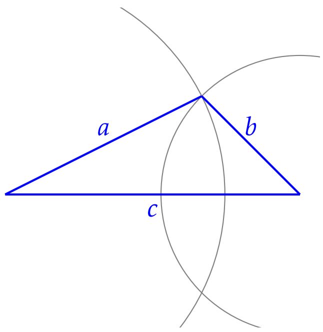

Figure 1 – Trois morceaux formant un triangle.

Réalisons maintenant un grand nombre d'expériences avec chacun des deux protocoles pour calculer la proportion d'expériences conduisant à la réalisation de l'événement « les trois morceaux forment un triangle ».

```python
def teste(protocole):
    n = 100000  # nombre d'expériences réalisées
    s = 0       # nombre d'expériences réussies
    for _ in range(n):
        (a, b, c) = sorted(protocole())
        if a + b > c:
            s += 1
    return s / n
```

```
In [1]: teste(protocole1)
Out[1]: 0.24918
In [2]: teste(protocole2)
Out[2]: 0.19276
```

On peut raisonnablement penser qu'avec le premier protocole la proportion d'expériences réussies est de l'ordre de $\frac{1}{4}$ mais est plus faible avec le second protocole (la valeur exacte est $\ln 2 - \frac{1}{2} \approx 0{,}193$).
Cet exemple illustre la nécessité qu'il y a à être le plus précis possible pour décrire le protocole conduisant à la réalisation de l'expérience. En l'espèce, « casser une baguette en trois » est un énoncé ambigu qui ne peut servir de support à une étude probabiliste.

1.2 Ensembles dénombrables

Le cours de première année s'est restreint aux univers finis : seul le premier des trois exemples de la section précédente rentre dans ce cadre. 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 *dénombrables* c'est-à-dire ceux qui peuvent être mis en bijection avec $\mathbb{N}$. L'univers choisi dans le deuxième exemple est dénombrable, mais pas celui du troisième exemple.

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* $[\![1, n]\!]$;
- 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 $\phi : \mathbb{N} \to \mathrm{X}$. 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 : $\mathrm{X} = \{x_n \mid n \in \mathbb{N}\}$.

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 $\phi : \mathbb{N} \to 2\mathbb{N}$, $\phi(n) = 2n$.
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.*

Proposition 1.2 — *Soit* X *un ensemble dénombrable et* Y *un ensemble fini ou dénombrable. Alors* $\mathrm{X} \cup \mathrm{Y}$ *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* $\mathrm{X} \times \mathrm{Y}$ *est dénombrable.*

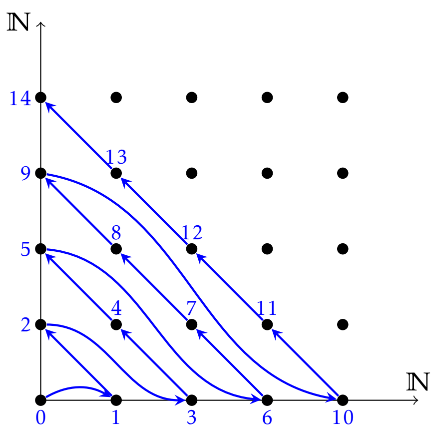

Figure 2 – $\mathbb{N} \times \mathbb{N}$ est dénombrable car on peut énumérer ses éléments.

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

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

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* $\mathscr{P}(\mathbb{N})$ *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* $\phi : \mathbb{N} \to \mathbb{N}$ *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.

1.3 Tribu et événements

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

- un événement et dit *élémentaire* si c'est un singleton ;
- l'événement *certain* est l'événement Ω ;
- l'événement *impossible* est l'événement $\varnothing$;
- l'événement non A, *contraire* de l'événement A, est l'événement $\overline{A} = \Omega \setminus A$ (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 $A \cap B = \varnothing$ (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 de nouveau le lancer de trois dés discernables associé à l'univers $\Omega = [\![1,6]\!]^3$.
$A = \left\{(x,y,z) \in \Omega \;\middle|\; x+y+z \leqslant 10\right\}$ est l'événement : « la somme des trois dés est inférieure ou égale à 10 ».
$B = \left\{(x,y,z) \in \Omega \;\middle|\; x+y+z \geqslant 10\right\}$ est l'événement : « la somme des trois dés est supérieure ou égale à 10 ».
A ou B est l'événement certain ; A et B est l'événement : « la somme des trois dés est égale à 10 ». Enfin, non A est l'événement « la somme des trois dés est strictement supérieure à 10 » donc 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 Ω est infini, il n'est en général pas possible de définir cette probabilité pour toutes les parties de Ω ; il faut se restreindre à un sous-ensemble $\mathscr{A}$ de $\mathscr{P}(\Omega)$ 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* Ω *est un ensemble, on appelle* tribu *sur* Ω *une partie* $\mathscr{A}$ *de* $\mathscr{P}(\Omega)$ *vérifiant :*

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

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

Proposition 1.6 — *Si* $\mathscr{A}$ *est une tribu sur l'univers* Ω*, alors :*

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

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

Exemple. À l'inverse, $\mathscr{P}(\Omega)$ 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 $\mathscr{A}$ 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 $\overline{A}$ (« 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 $\mathscr{A}$ 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* Ω *un univers et* $\mathscr{A}$ *une tribu sur* Ω*. On appelle* probabilité *sur* $(\Omega, \mathscr{A})$ *une application* $\mathbb{P} : \mathscr{A} \to [0, 1]$ *vérifiant :*

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

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

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 :*

- $\mathbb{P}(\varnothing) = 0$ *;*
- *si* A *et* B *sont deux événements incompatibles,* $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ *;*

– *si* A *est un événement,* $\mathbb{P}(\overline{\mathrm{A}}) = 1 - \mathbb{P}(\mathrm{A})$ *;*
– *si* A *et* B *sont deux événements ,* $\mathbb{P}(\mathrm{A} \cap \mathrm{B}) + \mathbb{P}(\mathrm{A} \cup \mathrm{B}) = \mathbb{P}(\mathrm{A}) + \mathbb{P}(\mathrm{B})$ *;*
– *si* $\mathrm{A} \subset \mathrm{B}$ *sont deux événements, alors* $\mathbb{P}(\mathrm{A}) \leqslant \mathbb{P}(\mathrm{B})$.

Exemple. Lorsque l'univers Ω est fini (et $\mathscr{A} = \mathscr{P}(\Omega)$) il existe une unique probabilité, appelée *probabilité uniforme* telle que pour tout $\omega \in \Omega$, $\mathbb{P}(\{\omega\}) = \dfrac{1}{\operatorname{card}\Omega}$. Dans ce cas, pour tout événement A, $\mathbb{P}(\mathrm{A}) = \dfrac{\operatorname{card}\mathrm{A}}{\operatorname{card}\Omega}$.
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 $\Omega = [\![1,6]\!]^3$, notons A l'événement « la somme des trois dés est inférieure ou égale à 10 » :

$$\mathrm{A} = \Big\{(x,y,z) \in \Omega \;\Big|\; x+y+z \leqslant 10\Big\}.$$

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

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 $\mathscr{A}$ sur l'univers $\Omega = \{\mathrm{P},\mathrm{F}\}^{\mathbb{N}^*}$ 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 $\mathbb{P}(\mathrm{A}_n) = (1-p)^{n-1}p$.

Les événements A_n étant deux à deux incompatibles ($i \neq j \implies \mathrm{A}_i \cap \mathrm{A}_j = \varnothing$) on a $\mathrm{P}\Big(\bigcup_{n\in\mathbb{N}} \mathrm{A}_n\Big) = \sum_{n=1}^{+\infty} \mathbb{P}(\mathrm{A}_n) = \sum_{n=1}^{+\infty} (1-p)^{n-1}p = p \times \dfrac{1}{1-(1-p)} = 1$. L'événement $\mathrm{A} = \bigcup_{n\in\mathbb{N}^*} \mathrm{A}_n$ (« Face apparaît au moins une fois ») vérifie $\mathbb{P}(\mathrm{A}) = 1$, l'événement $\overline{\mathrm{A}}$ (« la pièce tombe indéfiniment sur Pile ») vérifie $\mathbb{P}(\overline{\mathrm{A}}) = 1 - \mathbb{P}(\mathrm{A}) = 0$.
L'événement $\overline{\mathrm{A}}$ est dit « quasi-impossible » : bien qu'il soit un événement envisageable (il n'est pas égal à l'événement impossible $\varnothing$) 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* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé. Alors :*

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

$$\mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} \mathrm{A}_n\Big) = \lim \mathbb{P}(\mathrm{A}_n)\ ;$$

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

$$\mathbb{P}\Big(\bigcap_{n\in\mathbb{N}} \mathrm{A}_n\Big) = \lim \mathbb{P}(\mathrm{A}_n).$$

Proposition 1.9 (sous-additivité) — *Soit* $(\Omega, \mathscr{A}, \mathrm{P})$ *un espace probabilisé. Pour toute suite d'événements* (A_n), $\mathrm{P}\Big(\bigcup_{n\in\mathbb{N}} \mathrm{A}_n\Big) \leqslant \sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{A}_n)$ *(cette somme peut éventuellement être égale à* $+\infty$*).*

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 $\mathrm{B}_n = \bigcup_{k=0}^{n} \mathrm{A}_k$ est croissante donc $\mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} \mathrm{A}_n\Big) = \mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} \mathrm{B}_n\Big) = \lim \mathbb{P}(\mathrm{B}_n)$.
De même la suite $\mathrm{C}_n = \bigcap_{k=0}^{n} \mathrm{A}_k$ est décroissante donc $\mathbb{P}\Big(\bigcap_{n\in\mathbb{N}} \mathrm{A}_n\Big) = \mathbb{P}\Big(\bigcap_{n\in\mathbb{N}} \mathrm{C}_n\Big) = \lim \mathbb{P}(\mathrm{C}_n)$.

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 Ω, que l'on peut donc écrire : $\Omega = \{\omega_n \mid n \in \mathbb{N}\}$. Nous allons prouver le résultat suivant, qui montre qu'il est toujours possible de définir une probabilité sur $(\Omega, \mathscr{P}(\Omega))$ à partir de la valeur de 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 $(\Omega, \mathscr{P}(\Omega))$ telle que pour tout $n \in \mathbb{N}$, $\mathbb{P}(\{\omega_n\}) = p_n$.*

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

Exercice 3.

a) Soit $\mathbb{P}$ une probabilité sur $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$. Montrer que $\lim \mathbb{P}(\{n\}) = 0$.

b) 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 $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$ vérifiant : $\mathbb{P}([\![n, +\infty[\![) = \lambda a_n$.

1.5 Conditionnement et indépendance

Dans toute la suite du cours, $(\Omega, \mathscr{A}, \mathbb{P})$ 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* $\mathbb{P}_B(A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$, *réel qu'on pourra aussi noter* $\mathbb{P}(A \mid B)$.

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

Remarque. On dispose donc de l'égalité $\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A \mid B)$ 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 $\mathbb{P}(A \mid B)$ n'est pas formellement défini puisque $A \cap B \subset B \Rightarrow 0 \leqslant \mathbb{P}(A \cap B) \leqslant \mathbb{P}(B) = 0$.
Si A est B sont deux événements quelconques, la formule $\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A \mid B)}$ 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* ($i \neq j \Rightarrow B_i \cap B_j = \varnothing$) *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 Ω.

Théorème 1.12 (formule des probabilités totales) — *Soit* A *un événement et* $(B_i)_{i\in I}$ *un système complet d'événements. Alors* $\boxed{\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(B_i)\mathbb{P}(A \mid B_i).}$

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

Exercice 4. Neymar a neuf chances sur dix de marquer un penalty ; vous n'avez qu'une chance sur trois de marquer un penalty. Néanmoins, vous décidez de lancer un défi au brésilien : chacun tire un penalty à tour de rôle, le premier qui marque a gagné. Sportivement, Neymar vous laisse commencer. Quelle est la probabilité que vous l'emportiez ?

Proposition 1.13 (Formule de Bayes) — *Soit* $(\mathrm{B}_i)_{i\in\mathrm{I}}$ *un système complet d'événements tel que pour tout* $i \in \mathrm{I}$, $\mathbb{P}(\mathrm{B}_i) > 0$, *et* A *un événement tel que* $\mathbb{P}(\mathrm{A}) > 0$. *Alors* $\boxed{\mathbb{P}(\mathrm{B}_i \mid \mathrm{A}) = \dfrac{\mathbb{P}(\mathrm{B}_i)\mathbb{P}(\mathrm{A} \mid \mathrm{B}_i)}{\sum\limits_{j\in\mathrm{I}} \mathbb{P}(\mathrm{B}_j)\mathbb{P}(\mathrm{A} \mid \mathrm{B}_j)}}$.

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

Exercice 5. Un individu est tiré au hasard dans une population où l'on trouve une proportion 10^{-4} de séropositifs. On lui fait passer un test de détection de la séropositivité. Ce test fournit une réponse exacte à 99% si l'individu est séropositif, et à 99,9% s'il ne l'est pas.
Sachant que le test est positif, quelle est la probabilité pour cet individu soit effectivement séropositif ?

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 la séropositivité de l'individu) connaissant celle de sa conséquence (la positivité du test).

Exercice 6. 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 $\mathbb{P}(\mathrm{A}) > 0$ et $\mathbb{P}(\mathrm{B}) > 0$ on souhaite que $\mathbb{P}(\mathrm{A} \mid \mathrm{B}) = \mathbb{P}(\mathrm{A})$ et $\mathbb{P}(\mathrm{B} \mid \mathrm{A}) = \mathbb{P}(\mathrm{B})$, ce qui se traduit par $\dfrac{\mathbb{P}(\mathrm{A}\cap\mathrm{B})}{\mathbb{P}(\mathrm{B})} = \mathbb{P}(\mathrm{A})$ et $\dfrac{\mathbb{P}(\mathrm{B}\cap\mathrm{A})}{\mathbb{P}(\mathrm{A})} = \mathbb{P}(\mathrm{B})$. Ces deux égalités sont identiques, et pour pouvoir s'abstraire des hypothèses $\mathbb{P}(\mathrm{A}) > 0$ et $\mathbb{P}(\mathrm{B}) > 0$ on adoptera la définition suivante :

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

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

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

Définition. — *Une famille finie ou dénombrable* $(\mathrm{A}_i)_{i\in\mathrm{I}}$ *d'événements est dite* indépendante *lorsque pour tout entier* $p \leqslant \operatorname{card}\mathrm{I}$, *pour toute p-liste* $(i_1,\dots i_p) \in \mathrm{I}^p$ *d'indices deux-à-deux distincts*, $\mathbb{P}(\mathrm{A}_{i_1}\cap\cdots\cap\mathrm{A}_{i_p}) = \mathbb{P}(\mathrm{A}_{i_1})\cdots\mathbb{P}(\mathrm{A}_{i_p})$.

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 :

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

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

a) Montrer que si la série $\sum \mathbb{P}(A_n)$ converge alors $\mathbb{P}(A^*) = 0$.

b) 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 :

```python
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 :

```python
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.

Remarque. Nous démontrerons avec le théorème 2.23 (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 $(\Omega, \mathscr{A})$ est un espace probabilisable et E un ensemble, on appelle* variable aléatoire *toute fonction $X : \Omega \to E$ telle que pour tout $e \in E$, $X^{-1}(\{e\}) \in \mathscr{A}$ (autrement dit, $X^{-1}(\{e\})$ est un événement).*

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

Lorsque $X(\Omega)$ 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 : $X^{-1}(\{e\}) = \left\{\omega \in \Omega \,\middle|\, X(\omega) = e\right\}$.

Exemple. Pour l'expérience consistant à lancer trois dés et à compter le nombre de 6, nous pouvons choisir $\Omega = [\![1,6]\!]^3$, $E = \mathbb{N}$ et $X : \Omega \to \mathbb{N}$ définie par $X(e_1, e_2, e_3) = \operatorname{card}\{i \in [\![1,3]\!] \mid e_i = 6\}$.
$X(\Omega) = \{0, 1, 2, 3\}$ donc la variable aléatoire X est discrète (finie).

Exemple. 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 $X : \Omega \to E$ définie par $X((u_n)) = \min\{n \in \mathbb{N}^* \mid u_n = F\}$. Ici X est une variable aléatoire discrète (dénombrable).

Exemple. 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 $U \subset X(\Omega)$, $X^{-1}(U) \in \mathscr{A}$ (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, c'est pourquoi 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, $\{X \geqslant 3\}$ 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 $(\Omega, \mathscr{A}, P)$ un espace probabilisé et $X : \Omega \to E$ une variable aléatoire discrète. Alors l'application $\mathbb{P}_X : \mathscr{P}(X(\Omega)) \to [0,1]$ définie par $\mathbb{P}_X(U) = \mathbb{P}(X^{-1}(U)) = \mathbb{P}(\{X \in U\})$ est une probabilité sur $(X(\Omega), \mathscr{P}(X(\Omega)))$, appelée* loi de la variable X, *ou encore* distribution de X.

Exercice 8. 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 recommence 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 Ω. 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 Ω de sorte que la loi de X soit celle qui découle de $\mathbb{P}$. De manière plus formelle :

Théorème 2.3 — *Soit $(\Omega, \mathscr{A})$ un espace probabilisable et $X : \Omega \to E$ 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 $(\Omega, \mathscr{A})$ 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'expliciter formellement Ω et $\mathbb{P}$.

Exemple. Il existe une probabilité sur $[\![1,6]\!]^3$ tel que si X désigne le nombre de 6 obtenus lors d'un lancer, alors $\mathbb{P}(X = 0\} = 0$, $\mathbb{P}(X = 1) = 1/2$, $\mathbb{P}(X = 2) = 1/3$, $\mathbb{P}(X = 3) = 1/6$.

Remarque. Voici un résultat qui peut sembler bien étrange, mais souvenez-vous que la probabilité dépend du protocole utilisé pour réaliser l'expérience. Par exemple, la probabilité définie ci-dessus est obtenue en suivant le protocole :

(i) On lance le premier dé jusqu'à obtenir un 6.

(ii) On lance ensuite une pièce ; si elle tombe sur Face, on lance les deux autres dés jusqu'à obtenir deux valeurs différentes de 6, et l'expérience se termine. Sinon on lance le second dé jusqu'à obtenir un 6.

(iii) On lance enfin le dernier dé, et s'il tombe sur 5 on le relance jusqu'à obtenir un 6.

■ Fonction de répartition d'une variable aléatoire

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

Proposition 2.4 — *La fonction F_X possède les propriétés suivantes :*

- *F_X est une fonction croissante ;*
- $\lim_{x \to -\infty} F_X(x) = 0$ *et* $\lim_{x \to +\infty} F_X(x) = 1$;
- *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* $\boxed{X(\Omega) = [\![1, n]\!] \text{ et si pour tout } k \in [\![1, n]\!],\ \mathbb{P}(X = k) = \frac{1}{n}.}$ *On note dans ce cas* $X \hookrightarrow \mathscr{U}(n)$.

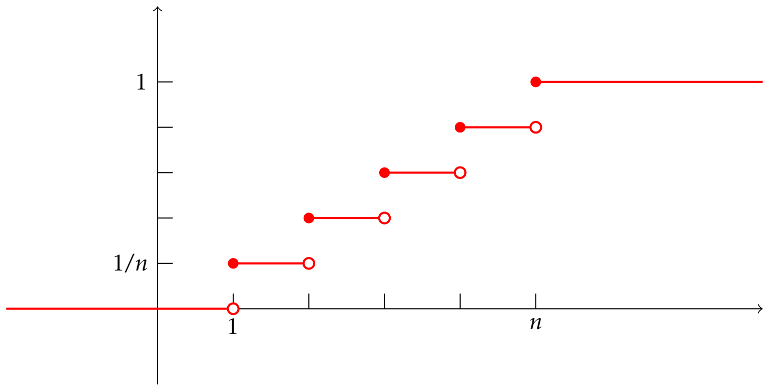

Figure 3 – La fonction de répartition d'une variable suivant une loi uniforme.

■ 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* $\boxed{X(\Omega) = \{0, 1\} \text{ et } \mathbb{P}(X = 0) = 1 - p,\ \mathbb{P}(X = 1) = p.}$ *On note dans ce cas* $X \hookrightarrow \mathscr{B}(p)$.

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

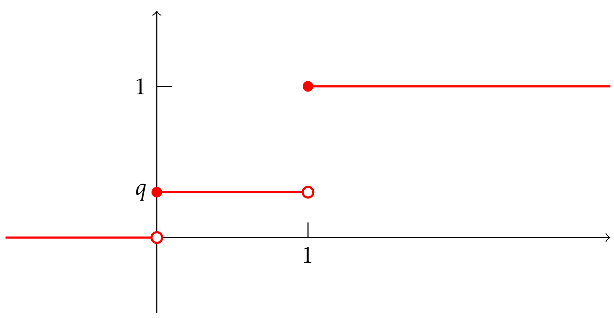

Figure 4 – La fonction de répartition d'une variable suivant une loi de Bernoulli.

■ 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* $\boxed{X(\Omega) = \mathbb{N}^* \text{ et pour tout } k \in \mathbb{N}^*,\ \mathbb{P}(X = k) = pq^{k-1}}$ *avec $q = 1 - p$. On note dans ce cas* $X \hookrightarrow \mathscr{G}(p)$.

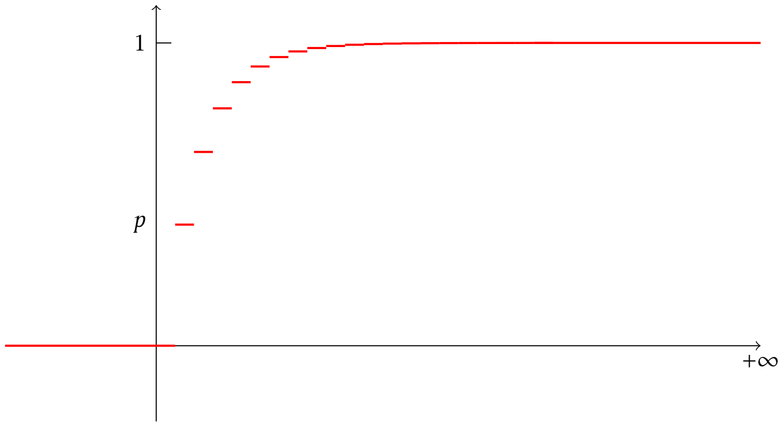

Figure 5 – La fonction de répartition d'une variable suivant une loi géométrique ($p = 0{,}4$).

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

Ce résultat traduit le fait qu'une loi géométrique est *sans mémoire* : après n expériences les variables $\mathrm{X} - n$ et X suivent le 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 9. 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* $\boxed{\mathrm{X}(\Omega) = [\![0,n]\!] \text{ et pour tout } k \in [\![0,n]\!],\ \mathbb{P}(\mathrm{X} = k) = \binom{n}{k} p^k q^{n-k}}$ *avec* $q = 1 - p$. *On note dans ce cas* $\mathrm{X} \hookrightarrow \mathscr{B}(n,p)$.

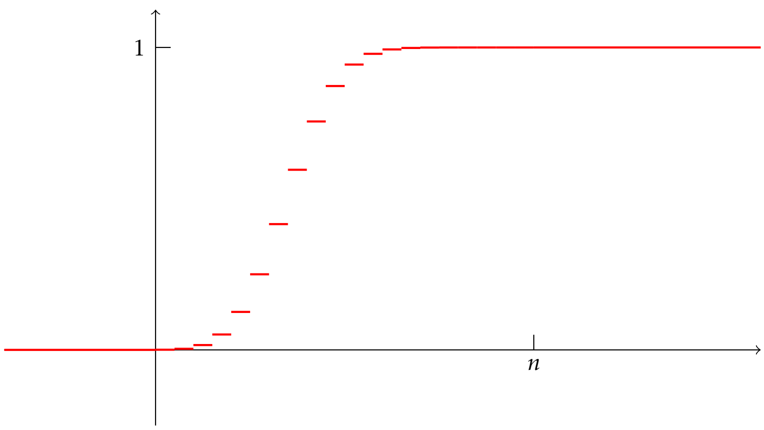

Figure 6 – La fonction de répartition d'une variable suivant une loi binomiale ($n = 40$, $p = 0{,}4$).

■ 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 $n \in \mathbb{N}$, $X_n \hookrightarrow \mathscr{B}(n, p_n)$. Si, de plus, $\lim np_n = \lambda$, alors pour tout $k \in \mathbb{N}$, $\lim\limits_{n \to +\infty} \mathbb{P}(X_n = k) = \dfrac{\lambda^k}{k!} e^{-\lambda}$.*

Définition. — *Soit $\lambda > 0$. On dit qu'une variable aléatoire réelle X suit une loi de Poisson de paramètre λ lorsque* $\boxed{X(\Omega) = \mathbb{N} \text{ et pour tout } k \in \mathbb{N},\ \mathbb{P}(X = k) = \dfrac{\lambda^k}{k!} e^{-\lambda}.}$ *On note dans ce cas $X \hookrightarrow \mathscr{P}(\lambda)$.*

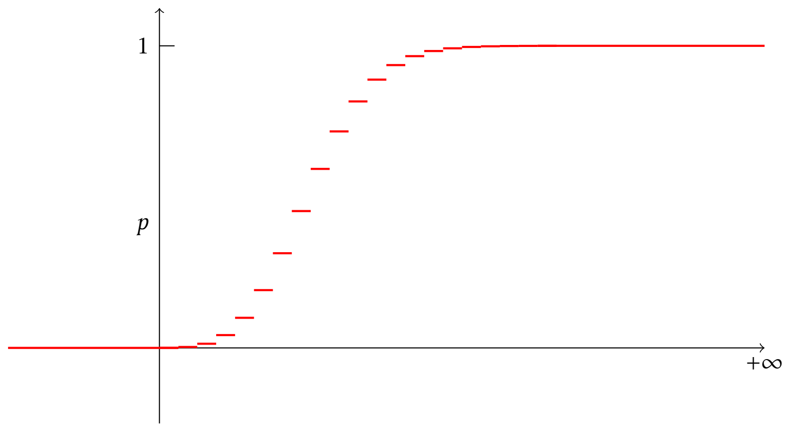

Figure 7 – La fonction de répartition d'une variable suivant une loi de Poisson ($\lambda = 10$).

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 $\mathscr{B}(n, p)$ par $\mathscr{P}(\lambda)$ (avec $\lambda = np$) dès lors que $n \geqslant 50$ et $np < 10$. Dans le cadre de cette approximation les calculs numériques s'en trouvent grandement simplifiés.

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

Un appel au standard à un instant donné est une éventualité de probabilité $p = \dfrac{1}{8 \times 30} = \dfrac{1}{240}$. 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 :

```python
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 10. 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* $(\Omega, \mathscr{A})$, *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 :

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

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* $(\Omega, \mathscr{A})$ *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* $(\Omega, \mathscr{A})$. *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* $(\Omega, \mathscr{A})$, *et* f *et* g *deux fonctions de* $\mathbb{R}$ *dans* $\mathbb{R}$. *Alors les variables aléatoires* $f(X)$ *et* $g(Y)$ *sont indépendantes.*

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

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

Exercice 11. Soit X une variable aléatoire qui suit une loi de Poisson de paramètre λ, Y une variable aléatoire qui suit une loi binomiale $\mathscr{B}(n, p)$ 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* $\Omega, \mathscr{A})$. *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* $(x_{i_1}, \dots, x_{i_p}) \in X_{i_1}(\Omega) \times \cdots \times X_{i_p}(\Omega)$, *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* $p \in [\![1, n]\!]$. *Soit* $f : \mathbb{R}^p \to \mathbb{R}$ *et* $g : \mathbb{R}^{n-p} \to \mathbb{R}$ *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 Ω 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 Ω 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 admettrons 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 $\sum_{n \geqslant 1} n(1-p)^{n-1}$ 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 9), nous avions obtenu un résultat proche de cette valeur. La justification théorique de cette constatation sera donnée par la *loi faible des grands nombres* (théorème 2.23).

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* $\mathbb{E}(X) = \sum_{n=1}^{+\infty} \mathbb{P}(X \geqslant n)$.

On dispose des résultats élémentaires 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) si X *est à valeurs positives,* $\mathbb{E}(X) \geqslant 0$ *(*positivité de l'espérance*);*

(iii) si $X \leqslant Y$ *alors* $\mathbb{E}(X) \leqslant \mathbb{E}(Y)$ *(*croissance de l'espérance*);*

(iv) et si, de plus, X *et* Y *sont indépendantes, alors* XY *est d'espérance finie et* $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Plus généralement, nous admettrons 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,* $\mathbb{E}(f(X)) = \sum_{n=0}^{+\infty} f(x_n) \mathbb{P}(X = x_n)$.

Exercice 12. On considère un jeu de Pile ou Face infini avec une pièce possédant la probabilité $p = \dfrac{2}{3}$ de tomber sur Face, et on note X la variable aléatoire égale au nombre de lancers nécessaires pour obtenir deux Face consécutifs pour la première fois.
Pour $n \geqslant 1$ on note $p_n = \mathbb{P}(X = n)$.

a) Calculer p_1 et p_2.

b) Pour $n \geqslant 3$, exprimer p_n en fonction de p_{n-1} et p_{n-2}, et en déduire l'expression de p_n.

c) Calculer enfin $\mathbb{E}(X)$.

■ 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 $X \hookrightarrow \mathscr{U}(n)$, alors $\mathbb{E}(X) = \dfrac{n+1}{2}$.

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

Loi géométrique Si $X \hookrightarrow \mathscr{G}(p)$ alors $\mathbb{E}(X) = \dfrac{1}{p}$.

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

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

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

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 $\mathbb{V}(X) = \mathbb{E}\big((X-\mathbb{E}(X))^2\big)$, 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,* $\boxed{\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.}$

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

• **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 $Y = \dfrac{X - \mathbb{E}(X)}{\sigma(X)}$ possède une espérance nulle (on dit qu'elle est *centrée*) et un écart type égal à 1 (on dit qu'elle est *réduite*).

■ 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 $X \hookrightarrow \mathscr{U}(n)$, alors $\mathbb{V}(X) = \dfrac{n^2-1}{12}$.

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

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

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

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

■ Inégalités de concentration

Théorème 2.17 (inégalité de Markov) — *Soit* X *une variable aléatoire positive possédant une espérance, et* $a > 0$. *Alors* $\boxed{\mathbb{P}(X \geqslant a) \leqslant \dfrac{\mathbb{E}(X)}{a}.}$

Théorème 2.18 (inégalité de Bienaymé-Tchebychev) — *Soit* X *une variable aléatoire réelle admettant un moment d'ordre 2, et* $\alpha > 0$. *Alors* $\boxed{\mathrm{P}\Big(\big|X - \mathbb{E}(X)\big| \geqslant \alpha\Big) \leqslant \dfrac{\mathbb{V}(X)}{\alpha^2} = \dfrac{\sigma(X)^2}{\alpha^2}.}$

Que signifie cette inégalité ? La probabilité calculée mesure le risque de s'écarter de l'espérance d'une quantité supérieure à α. 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.

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, 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* $\mathrm{cov}(X,Y) = \mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y)))$.

Le calcul précédent nous permet déjà d'énoncer :

Théorème 2.19 — *Lorsque* X *et* Y *sont deux variables aléatoires réelles indépendantes et possédant une espérance,* $\mathrm{cov}(X,Y) = 0$. *On dira que* X *et* Y ne sont pas corrélées.

Dans le cas général, pour établir l'existence de la covariance, nous nous contenterons d'une condition suffisante (que nous admettrons) :

Proposition 2.20 — *Si* X *et* Y *sont deux variables aléatoires réelles possédant un moment d'ordre 2 (autrement dit une variance) alors* $\mathrm{cov}(X,Y)$ *existe.*

Proposition 2.21 (propriétés de la covariance) — *Soient* X Y *et* Z *trois variables aléatoires réelles possédant des moments d'ordre 2. Alors :*

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

Théorème 2.22 — *Soient* X *et* Y *deux variables aléatoires réelles possédant des moments d'ordre 2. Alors*

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

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

Remarque. Cette formule se généralise au cas de n variables aléatoires : $\mathbb{V}\Big(\sum_{k=1}^{n} X_k\Big) = \sum_{k=1}^{n} \mathbb{V}(X_k) + 2\sum_{i<j} \mathrm{cov}(X_i, X_j)$.
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, $\mathbb{V}(X_1 + \cdots + X_n) = \mathbb{V}(X_1) + \cdots + \mathbb{V}(X_n)$.

■ 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 9), 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 théorique à cet état de fait :

Théorème 2.23 (loi faible des grands nombres) — *Soit* $(X_n)_{n\geqslant 1}$ *une succession de variables aléatoires deux-à-deux indépendantes et de même loi admettant un moment d'ordre 2. On pose* $S_n = X_1 + X_2 + \cdots + X_n$, $m = \mathbb{E}(X_k)$ *et* $\sigma = \sigma(X_k)$ *(puisque les* X_i *suivent la même loi ces quantités ne dépendent pas de k). Alors pour tout* $\epsilon > 0$,

$$\boxed{\mathbb{P}\Big(\Big|\frac{1}{n}S_n - m\Big| \geqslant \epsilon\Big) \leqslant \frac{\sigma^2}{n\epsilon^2}.}$$

Remarque. Avec les mêmes hypothèses on en déduit que : $\lim_{n\to+\infty} \mathbb{P}\Big(\Big|\frac{1}{n}S_n - m\Big| \geqslant \epsilon\Big) = 0$.
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 ϵ est faible.

Exemple. Dans l'exemple numérique de la page 9 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 $X \hookrightarrow \mathscr{G}(p)$). 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.

■ Coefficient de corrélation

Théorème 2.24 — *Soient* X *et* Y *deux variables aléatoires possédant des moments d'ordre 2. Alors*

$$|\mathrm{cov}(X,Y)| \leqslant \sigma(X)\sigma(Y)$$

Définition. — *Lorsque* $\sigma(X) > 0$ *et* $\sigma(Y) > 0$, *on appelle* coefficient de correlation *la quantité* $\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma(X)\sigma(Y)}$.
D'après ce qui précède, $\rho(X,Y) \in [-1,1]$.

• Interprétation du coefficient de correlation
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 $\operatorname{cov}(X,Y) = a\operatorname{cov}(X,X) + b\operatorname{cov}(X,1) = a\mathbb{V}(X)$ 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 correlation nul ! Ce peut être le cas par exemple lorsque la dépendance existe mais n'est pas linéaire. Ainsi, un coefficient de correlation nul indique simplement que la dépendance des variables, si elle existe, est très faiblement linéaire.

Exercice 14. 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 $\mathbb{E}\big((Y - aX - b)^2\big)$.

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.

2.7 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 caractéristique 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 $\mathscr{C}^\infty$ 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 2.25 — *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 $G_X''(1) = \sum_{n=1}^{+\infty} n(n-1)\mathbb{P}(X = n)$.

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

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

Théorème 2.26 — *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 $X \hookrightarrow \mathscr{U}(n)$, alors $G_X(t) = \dfrac{1}{n}(t + t^2 + \cdots + t^n)$.

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

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

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

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

Exercice 15. 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 $G_X(t) = \displaystyle\sum_{n=0}^{+\infty} \mathbb{P}(X = n)t^n$ et $G_Y(t) = \displaystyle\sum_{n=0}^{+\infty} \mathbb{P}(Y = n)t^n$ 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 2.27 — *Si X et Y sont deux variables aléatoires indépendantes à valeur entières, alors*

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

Exemple.

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

3. Exercices

Espaces probabilisés

Exercice 16 Soit $\mathscr{A}$ une tribu de $[0, 1]$ contenant tous les segments $[a, b]$ avec $0 \leqslant a < b \leqslant 1$.

a) Montrer que $\mathscr{A}$ contient tous les intervalles ouverts $]a, b[$ avec $0 \leqslant a < b \leqslant 1$, ainsi que tous les singletons $\{\alpha\}$ avec $0 \leqslant \alpha \leqslant 1$.

b) On suppose que $\mathbb{P}$ est une probabilité sur $([0, 1], \mathscr{A})$ vérifiant : $\mathbb{P}([a, b]) = b - a$ pour $0 \leqslant a < b \leqslant 1$. Calculer $\mathbb{P}(]a, b[)$ et $\mathbb{P}(\{\alpha\})$.

Exercice 17 Soit $(\Omega, \mathscr{A})$ un espace probabilisé, et (A_n) une suite d'événements. Traduire en notations ensemblistes les événements suivants :

a) l'un au moins des événements A_n est réalisé ;

b) tous les événements A_n sont réalisés ;

c) il existe un rang à partir duquel tous les événements A_n sont réalisés ;

d) une infinité d'événements parmi les A_n sont réalisés ;

e) seul un nombre fini d'événements A_n sont réalisés ;

f) une infinité d'événements parmi les A_n ne sont pas réalisés.

Exercice 18 À l'occasion de la coupe de France on organise un tirage au sort entre n équipes de football de 1re division et n équipes de 2e division (chaque équipe joue un seul match).

a) Calculer la probabilité p_n que tous les matchs opposent une équipe de première division à une équipe de seconde division.

b) Calculer la probabilité q_n que tous les matchs opposent deux équipes de la même division.

Exercice 19 Une urne contient une boule blanche et une boule noire. On tire dans cette urne une boule, puis on la remet accompagnées de deux autres boules de la même couleur. On répète ensuite cette opération indéfiniment.

a) Quelle est la probabilité que les n premières boules tirées soient blanches ?

b) Quelle est la probabilité de tirer indéfiniment des boules blanches ?

Conditionnement et indépendance

Exercice 20 Soit $(\Omega, \mathscr{A}, P)$ un espace probabilisé, et $(A, B) \in \mathscr{A}^2$. Montrer que A et B sont indépendants si et seulement si $\mathbb{P}(A \cap B)\mathbb{P}(\overline{A} \cap \overline{B}) = \mathbb{P}(A \cap \overline{B})\mathbb{P}(\overline{A} \cap B)$.

Exercice 21 Soit $(\Omega, \mathscr{A}, P)$ un espace probabilisé, et $(A, B) \in \mathscr{A}^2$.

a) Montrer que $\mathbb{P}(A \cap B)^2 \leqslant \mathbb{P}(A)\mathbb{P}(B)$.

b) Si $\mathbb{P}(A) > 0$, montrer que $\mathbb{P}(A \cap B \mid A \cup B) \leqslant \mathbb{P}(A \cap B \mid A)$.

c) Si $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$, montrer que $\mathbb{P}(A) < \mathbb{P}(A \mid B) \Longrightarrow \mathbb{P}(B) < \mathbb{P}(B \mid A)$.

Exercice 22 On considère deux urnes U_1, contenant n_1 boules noires et b_1 boules blanches, et U_2, contenant n_2 boules noires et b_2 boules blanches. On choisit de façon équiprobable l'une des deux urnes puis on y effectue deux tirages successifs avec remise. Soit N_1 l'événement « tirer une boule noire au premier tirage » et N_2 l'événement « tirer une boule noire au second tirage ».

a) Quelle est la probabilité de N_1 ? de N_2 ? de $N_1 \cap N_2$?

b) Les événements N_1 et N_2 sont-ils indépendants ?

Exercice 23 Alice porte le gène de l'hémophilie avec une probabilité de 0,5. Si elle est porteuse, chacun de ses fils aura une chance sur deux de souffrir de cette maladie. Sachant qu'Alice a eu trois fils et qu'aucun n'est hémophile, quelle est la probabilité qu'elle soit porteuse du gène ? Et s'il nait un quatrième fils, avec quelle probabilité sera-t-il hémophile ?

Exercice 24 Un sac contient 100 dés dont 25 sont pipés. Un dé pipé a une chance sur deux de donner le chiffre 6.

a) On prend un dé au hasard et on le lance. On obtient le chiffre 6. Quelle est la probabilité que ce dé soit pipé ?

b) On prend un dé au hasard et on le lance n fois. On obtient n fois le chiffre 6. Quelle est la probabilité p_n que ce dé soit pipé ?

c) Déterminer $\lim p_n$ et interpréter ce résultat.

Exercice 25 Soit $n \in \mathbb{N}^*$. On dispose de $n+1$ urnes $U_0, \dots, U_n$, l'urne U_k contenant k boules blanches et $n-k$ boules noires.

a) On choisit une urne au hasard et on tire une boule dans l'urne. Quelle est la probabilité de tirer une boule blanche ?

b) On choisit une urne au hasard et on tire une boule. Celle-ci est blanche. Quelle est la probabilité d'avoir choisi l'urne U_k ?

c) On choisit une urne au hasard et on tire successivement avec remise deux boules dans l'urne. Quelle est la probabilité d'obtenir deux boules blanches ?

d) On suppose $n \geqslant 2$. On choisit une urne au hasard et on tire successivement sans remise deux boules dans l'urne. Quelle est la probabilité d'obtenir deux boules blanches ?

Exercice 26 Une puce se trouve à l'instant 0 au point d'abscisse $a \in [\![0, N]\!]$. À chaque instant elle fait un bond de $+1$ avec la probabilité $p \in]0, 1/2[$ ou un bond de -1 avec la probabilité $q = 1-p$. Le processus se termine lorsque la particule atteint une des extrémités de l'intervalle 0 ou N.

a) Écrire en Python une fonction qui simule cette marche aléatoire, en renvoyant la valeur d'arrivée (O ou N) et le nombre de pas utilisés.

On note u_a la probabilité pour que la puce, partant de a, s'arrête en 0.

b) Que vaut u_0 ? u_N ?

c) Lorsque $0 < a < N$, exprimer u_a en fonction de u_{a-1} et u_{a+1}.

d) En déduire l'expression générale de u_a.

e) Calculer la probabilité v_a pour que la puce, partant de a, s 'arrête en N.

f) Calculer $u_a + v_a$. Comment interpréter ce résultat ?

Exercice 27 Pour $s > 1$ et $\lambda \in \mathbb{R}$, on pose $P\{n\}) = \dfrac{\lambda}{n^s}$ pour tout $n \in \mathbb{N}^*$.

a) Pour quelle valeur de λ l'application P détermine-t-elle une probabilité sur $(\mathbb{N}^*, \mathcal{P}(\mathbb{N}^*))$? On suppose désormais cette valeur choisie.

b) Pour $p \in \mathbb{N}^*$ on considère l'événement $A_p = \{n \in \mathbb{N}^* \mid p \text{ divise } n\}$. Calculer $\mathbb{P}(A_p)$.

c) On note $\mathcal{P}$ l'ensemble des nombres premiers. Vérifier que la famille $(A_p)_{p\in\mathcal{P}}$ est indépendante.

d) En calculant de deux manière $\mathbb{P}(\{1\})$ en déduire que $\displaystyle\sum_{n=1}^{+\infty} \frac{1}{n^s} = \frac{1}{\displaystyle\prod_{p\in\mathcal{P}} \left(1 - \frac{1}{p^s}\right)}$.

Variables aléatoires

Exercice 28 Soient X et Y deux variables aléatoires à valeurs dans $\mathbb{N}$. On suppose que la loi conjointe de X et Y vérifie $\mathbb{P}(X = j \text{ et } Y = k) = a\dfrac{j+k}{2^{j+k}}$ avec $a \in \mathbb{R}$.

a) Déterminer la valeur de a.

b) Reconnaître les lois marginales de X et Y.

c) Les variables X et Y sont-elles indépendantes ?

d) Calculer $\mathbb{P}(X = Y)$.

Exercice 29 Soient X et Y deux variables aléatoires indépendantes qui suivent la même loi géométrique de paramètre p. Déterminer la loi de $X + Y$.

Exercice 30 Soient X et Y deux variables aléatoires indépendantes suivant des lois géométriques de paramètres respectifs p_1 et p_2. Calculer $\mathbb{P}(X < Y)$.

Exercice 31 Soient X et Y deux variables aléatoires indépendantes suivant des lois géométriques de paramètres respectifs p_1 et p_2. Quelle est la probabilité que la matrice $A = \begin{pmatrix} X & 1 \\ 0 & Y \end{pmatrix}$ soit diagonalisable ?

Espérance et variance

Exercice 32 Un joueur joue à une partie et gagne avec une probabilité p dans $]0,1[$. On note L_1 la longueur de la première liste de parties toutes gagnées ou toutes perdues et L_2 la longueur de la deuxième liste.
Par exemple, pour la séquence suivante : GGGPPG on a $L_1 = 3$ et $L_2 = 2$.

a) Déterminer la loi de L_1 et son espérance.

b) Déterminer la loi de L_2 et son espérance.

c) L_1 et L_2 sont-elles indépendantes ?

Exercice 33 On dispose d'une urne contenant N boules numérotées de 1 à N. On effectue n tirages successifs avec remise, et on note X le plus grand nombre obtenu.

a) Que vaut $\mathbb{P}(X \geqslant k)$? En déduire la valeur de $\mathbb{E}(X)$.

b) Calculer $\lim\limits_{N\to+\infty} \dfrac{\mathbb{E}(X)}{N}$.

Exercice 34

a) Calculer $\mathbb{E}(1/X)$ lorsque $X \hookrightarrow \mathscr{G}(p)$.

b) Calculer $\mathbb{E}\left(\dfrac{1}{X+1}\right)$ lorsque $X \hookrightarrow \mathscr{P}(\lambda)$.

Exercice 35 On dispose de k dés équilibrés. On les lance et on ne garde que ceux qui n'ont pas amené un 6. On recommence avec les dés restants jusqu'à ce qu'ils soient tous retirés.

a) On note X la variable aléatoire égale au nombre de lancers qui seront effectués. Pour $n \in \mathbb{N}^*$, déterminer $\mathbb{P}(X \leqslant n)$, et en déduire $\mathbb{P}(X = n)$.

b) Déterminer l'espérance de X lorsque $k = 2$, et rédiger en Python un script pour vérifier expérimentalement cette valeur.

Exercice 36 Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées, à valeurs strictement positives. On suppose que X et $1/X$ possèdent une espérance. On pose $S_n = X_1 + X_2 + \cdots + X_n$.

a) Calculer $\mathbb{E}\left(\dfrac{X_k}{S_n}\right)$ pour tout $n \in \mathbb{N}^*$ et $k \in [\![1,n]\!]$.

b) Calculer $\mathbb{E}\left(\dfrac{S_m}{S_n}\right)$ pour tout $(m,n) \in (\mathbb{N}^*)^2$.

Exercice 37 Soient X et Y deux variables aléatoires discrètes à valeurs réelles, admettant toutes deux un moment d'ordre 2.

a) Montrer que $\mathbb{E}(XY)^2 \leqslant \mathbb{E}(X^2)\mathbb{E}(Y^2)$.

b) Montrer que $\mathbb{E}(|X|)^2 \leqslant \mathbb{E}(X^2)\mathbb{P}(|X| > 0)$.

c) On suppose que $\mathbb{E}(Y) \geqslant 0$ et que $\mathbb{E}(Y^2) > 0$. Montrer que $\mathbb{P}(Y > 0) \geqslant \dfrac{\mathbb{E}(Y)^2}{\mathbb{E}(Y^2)}$.

d) En déduire l'inégalité de Cantelli : $\forall \epsilon > 0,\ \mathbb{P}\Big(|X - \mathbb{E}(X)| \geqslant \epsilon\Big) \leqslant \dfrac{2\mathbb{V}(X)}{\mathbb{V}(X) + \epsilon^2}$.

Séries génératrice

Exercice 38 Soient $p \in]0,1[$ et $r \in]0,1[$. On considère deux variables aléatoires réelles indépendantes X et Y telles que $X \hookrightarrow \mathscr{B}(p)$ et $Y \hookrightarrow \mathscr{G}(r)$. On définit une nouvelle variable aléatoire U en posant $U = 0$ si $X = 0$ et $U = Y$ si $X = 1$.
Déterminer la série génératrice de U et en déduire $\mathbb{E}(U)$ et $\mathbb{V}(U)$.

Exercice 39 Dans une population donnée on suppose que la probabilité p_n pour qu'une famille ait exactement n enfants est définie par :

$$p_0 = p_1 = a \in \left]0, 1/2\right[\qquad \text{et} \qquad p_n = \frac{1-2a}{2^{n-1}} \quad \text{si } n \geqslant 2$$

a) Vérifier que la famille (p_n) définit bien une loi de probabilité et déterminer sa série génératrice.

b) Démontrer que cette loi de probabilité admet une espérance et la calculer.

On suppose que la naissance d'un garçon ou d'une fille sont des événements équiprobables.

c) Quelle est la probabilité pour qu'une famille ayant deux garçons ait exactement deux enfants ?

d) Quelle est la probabilité pour qu'une famille ait deux filles sachant qu'elle a deux garçons ?

Exercice 40 Soit X une variable aléatoire à valeurs dans $\mathbb{N}$ dont la loi est donnée par $\mathbb{P}(\mathrm{X} = k) = a\binom{n+k}{k}p^k$ avec $a > 0$ et $p \in \left]0, 1\right[$.

a) Calculer la fonction génératrice G_X de X et en déduire la valeur de a.

b) Calculer espérance et variance de X.