Chapitre VII
Probabilités

La théorie mathématique des probabilités nait au XVI^{e} 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^{e} 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.

1. Ensembles dénombrables et familles sommables

1.1 Ensembles dénombrables

Le cours de première année s'est restreint aux variables aléatoires à valeurs dans un ensemble fini ; cette année, nous allons étendre nos connaissances aux variables aléatoires à valeurs dans un ensemble infini, *mais pas à n'importe lesquels* : seuls les plus « simples » des ensembles infinis seront abordés, les ensembles dits *dénombrables* c'est-à-dire ceux qui peuvent être mis en bijection avec $\mathbb{N}$.

Définition. — *On dit d'un ensemble* E *qu'il est :*

- fini *lorsqu'il existe un entier* $n \in \mathbb{N}$ *tel que* E *est en bijection avec* $[\![1, n]\!]$;
- dénombrable *lorsqu'il est en bijection avec* $\mathbb{N}$;
- discret *s'il est fini ou dénombrable. On dira aussi que* E *est* au plus *dénombrable.*

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

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.*

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* E *un ensemble dénombrable et* F *un ensemble au plus dénombrable. Alors* $E \cup F$ *est dénombrable.*

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

Proposition 1.3 — *Soit* E *un ensemble dénombrable et* F *un ensemble non vide au plus dénombrable. Alors le produit cartésien* $E \times F$ *est dénombrable.*

■ Réunion et intersection dénombrables

Définition. — *Soit* $(A_n)_{n\in\mathbb{N}}$ *une suite de parties d'un même ensemble* E. *On définit les ensembles* $\bigcup_{n\in\mathbb{N}} A_n$ *et* $\bigcap_{n\in\mathbb{N}} A_n$ *par :*

$$x \in \bigcup_{n\in\mathbb{N}} A_n \iff \exists n \in \mathbb{N} \,\Big|\, x \in A_n \qquad \text{et} \qquad x \in \bigcap_{n\in\mathbb{N}} A_n \iff \forall n \in \mathbb{N}, x \in A_n$$

Proposition 1.4 — *Une union finie ou dénombrable d'ensembles dénombrables est dénombrable.*

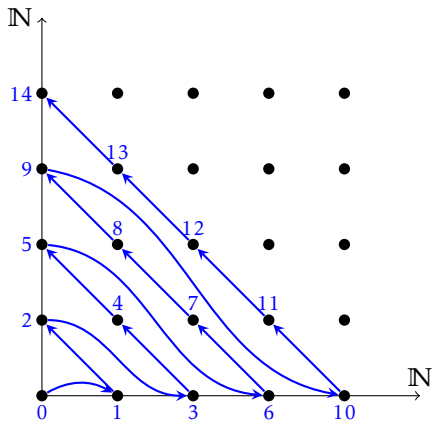

Figure 1 – $\mathbb{N} \times \mathbb{N}$ est dénombrable car on peut énumérer ses éléments.

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

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 1.5 — *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.

1.2 Familles sommables de réels positifs

Étant donné une famille de réels positifs $(x_i)_{i\in I}$ indexée par un ensemble I, nous allons chercher à déterminer s'il est possible de donner un sens à la somme $\sum_{i\in I} x_i$ de ces derniers.

Lorsque I est un ensemble fini : $I = \{i_1, \dots, i_n\}$, cette question ne parait guère intéressante : la commutativité de l'addition implique que la somme $x_{i_1} + \cdots + x_{i_n}$ reste inchangée quel que soit la manière de permuter ses éléments. *Mais ceci n'a plus rien d'évident lorsque* I *est un ensemble infini.*

Attention. Il ne s'agit pas ici du concept de série numérique : dans le cas d'une série numérique, les éléments de la famille ont été *ordonnés au préalable*, et on détermine ensuite la convergence ou la divergence de la suite des sommes partielles. Nous cherchons ici à nous affranchir de l'ordre des éléments dans la somme.

Définition. — *Soit $(x_i)_{i\in I}$ une famille de réels positifs. Cette famille est dite* sommable *lorsque l'ensemble*

$$\left\{ \sum_{i\in J} x_i \;\middle|\; J \subset I \text{ et } J \text{ finie} \right\}$$

est majoré. Dans ce cas, on note $\sum_{i\in I} x_i$ la borne supérieure de cet ensemble.

Remarque. Lorsque la famille de réels positifs $(x_i)_{i\in I}$ n'est pas sommable, on notera par commodité $\sum_{i\in I} x_i = +\infty$.

Proposition 1.6 — *Si la famille $(x_i)_{i\in I}$ est sommable, l'ensemble $\left\{ i \in I \;\middle|\; x_i \neq 0 \right\}$ est au plus dénombrable.*

Il résulte de la proposition ci-dessus que dans la pratique, on pourra toujours supposer, lorsque la famille est sommable, que I est un ensemble fini ou dénombrable.

■ Lien avec les séries numériques

Théorème 1.7 — *Soit $(x_n)_{n\in\mathbb{N}}$ une suite de nombres réels positifs. Alors la famille $(x_n)_{n\in\mathbb{N}}$ est sommable si et seulement si la série numérique $\sum x_n$ converge, et dans ce cas, $\sum_{n\in\mathbb{N}} x_n = \sum_{n=0}^{+\infty} x_n$.*

Ce résultat nous donne une manière très simple d'étudier la sommabilité d'une famille dénombrable de réels positifs : il suffit de les ordonner *d'une manière arbitraire* puis d'étudier la convergence de la série numérique afférente.

Maintenant que nous avons trouvé un moyen de nous affranchir de l'ordre de sommation d'une famille de réels positifs, il nous reste à énoncer deux formules couramment utilisées dans les calculs :

Théorème 1.8 (sommation par paquets) — *Soit $(\mathrm{I}_n)_{n\in\mathbb{N}}$ une partition dénombrable de I, et $(x_i)_{i\in\mathrm{I}}$ une famille sommable de réels positifs. Alors $\sum_{i\in\mathrm{I}} x_i = \sum_{n=0}^{+\infty}\left(\sum_{i\in\mathrm{I}_n} x_i\right)$.*

Théorème 1.9 (Fubini) — *Soit $(x_{i,j})_{(i,j)\in\mathrm{I}\times\mathrm{J}}$ une famille sommable de réels positifs. Alors*

$$\sum_{(i,j)\in\mathrm{I}\times\mathrm{J}} x_{i,j} = \sum_{i\in\mathrm{I}}\sum_{j\in\mathrm{J}} x_{i,j} = \sum_{j\in\mathrm{J}}\sum_{i\in\mathrm{I}} x_{i,j}$$

Remarque. Nous l'avons dit, lorsqu'une famille $(x_i)_{i\in\mathrm{I}}$ de réels positifs n'est pas sommable, on s'autorisera à écrire $\sum_{i\in\mathrm{I}} x_i = +\infty$. Ceci a pour conséquence que les deux théorèmes ci-dessus peuvent s'appliquer *sans justification préalable de la sommabilité*. Obtenir à la fin des calculs une somme finie justifiera *a posteriori* la sommabilité de la famille, et au contraire obtenir une somme divergente prouvera la non sommabilité de cette famille.

1.3 Familles sommables de réels quelconques

Définition. — *Une famille $(x_i)_{i\in\mathrm{I}}$ est dite* sommable *lorsque la famille de réels positifs $(|x_i|)_{i\in\mathrm{I}}$ est sommable.*

Une fois cette définition posée, on définit la somme à l'aide du résultat suivant :

Théorème 1.10 — *Soit $(x_i)_{i\in\mathrm{I}}$ une famille sommable de réels quelconques. Pour tout $i\in\mathrm{I}$ on pose $x_i^+ = \max(x_i, 0)$ et $x_i^- = \max(0, -x_i)$. Alors les familles de réels positifs $(x_i^+)_{i\in\mathrm{I}}$ et $(x_i^-)_{i\in\mathrm{I}}$ sont sommables, et on pose par définition :*

$$\sum_{i\in\mathrm{I}} x_i = \sum_{i\in\mathrm{I}} x_i^+ - \sum_{i\in\mathrm{I}} x_i^-$$

On dispose alors des résultats suivants, que nous admettrons. Notez cependant que contrairement aux familles sommables de réels positifs pour lesquels l'obtention d'un résultat fini à la fin des calculs justifie a posteriori la sommabilité de la famille, il est indispensable, dans le cas d'une famille de réels quelconques, de justifier la sommabilité *en préalable à tout calcul*.

Proposition 1.11 — *Soit $(x_i)_{i\in\mathrm{I}}$ une famille sommable. Alors $\left|\sum_{i\in\mathrm{I}} x_i\right| \leqslant \sum_{i\in\mathrm{I}} |x_i|$.*

Proposition 1.12 — *soient $(x_i)_{i\in\mathrm{I}}$ et $(y_i)_{i\in\mathrm{I}}$ deux familles de nombres réels telles que pour tout $i\in\mathrm{I}$, $|x_i| \leqslant y_i$. Alors la sommabilité de $(y_i)_{i\in\mathrm{I}}$ entraîne celle de $(x_i)_{i\in\mathrm{I}}$.*

Proposition 1.13 — *Soit $(x_i)_{i\in\mathrm{I}}$ et $(y_i)_{i\in\mathrm{I}}$ deux familles sommables, et $\lambda\in\mathbb{R}$. Alors la famille $(\lambda x_i + y_i)_{i\in\mathrm{I}}$ est sommable, et $\sum_{i\in\mathrm{I}}(\lambda x_i + y_i) = \lambda\sum_{i\in\mathrm{I}} x_i + \sum_{i\in\mathrm{I}} y_i$.*

Théorème 1.14 (sommation par paquets) — *Soit* $(I_n)_{n\in\mathbb{N}}$ *une partition dénombrable de* I, *et* $(x_i)_{i\in I}$ *une famille sommable. Alors* $\displaystyle\sum_{i\in I} x_i = \sum_{n=0}^{+\infty}\left(\sum_{i\in I_n} x_i\right)$.

Théorème 1.15 (Fubini) — *Soit* $(x_{i,j})_{(i,j)\in I\times J}$ *une famille sommable. Alors* $\displaystyle\sum_{(i,j)\in I\times J} x_{i,j} = \sum_{i\in I}\sum_{j\in J} x_{i,j} = \sum_{j\in J}\sum_{i\in I} x_{i,j}$.

2. Espaces probabilisés

2.1 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.

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 $\Omega = [\![1,6]\!]^3$.

– 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 $\Omega = \left\{(x,y,z)\in\,]0,1[^3 \;\middle|\; x+y+z=1\right\}$.

Le premier exemple correspond à un univers fini, le second à un univers dénombrable, le troisième à un univers non dénombrable.

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.

2.2 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 est 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 le lancer de trois dés 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 2.1 — *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 $\Omega = \{0,1\}^{\mathbb{N}}$ (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 $\mathbb{R}$.

2.3 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 $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n)$.*

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 2.2 — *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*, $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$;
- *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* $\mathbb{P}(A) \leqslant \mathbb{P}(B)$.

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, $\mathbb{P}(A) = \dfrac{\operatorname{card} A}{\operatorname{card}\Omega}$.

Exemple. Revenons maintenant sur l'expérience consistant à jeter une pièce de monnaie jusqu'à obtenir Face. Nous avons admis qu'on pouvait définir une tribu $\mathscr{A}$ sur l'univers $\Omega = \{0,1\}^{\mathbb{N}}$ 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 indépendantes 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 ($i \neq j \implies A_i \cap A_j = \varnothing$) on a $\mathrm{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n) = \sum_{n=1}^{+\infty} (1-p)^{n-1}p = p \times \dfrac{1}{1-(1-p)} = 1$. L'événement $A = \bigcup_{n\in\mathbb{N}^*} A_n$ (« Face apparaît au moins une fois ») vérifie $\mathbb{P}(A) = 1$, l'événement $\overline{A}$ (« la pièce tombe indéfiniment sur Pile ») vérifie $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A) = 0$.
L'événement $\overline{A}$ est dit « quasi-impossible », ou « négligeable » : 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 », ou « presque sûr ».

Voyons maintenant quelques résultats propres aux univers infinis :

Théorème 2.3 (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* ($A_n \subset A_{n+1}$), *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*

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

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

Proposition 2.4 (sous-additivité) — *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé. Pour toute suite d'événements* (A_n), $\mathrm{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leqslant \sum_{n=0}^{+\infty} \mathbb{P}(A_n)$ *(cette somme peut éventuellement être égale à* $+\infty$*).*

Exercice 2 [Lemme de Borel-Cantelli]

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

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

■ Probabilité sur un univers dénombrable

Considérons maintenant un univers dénombrable Ω, 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 $(\Omega, \mathscr{P}(\Omega))$ à partir de la valeur de $\mathbb{P}$ sur les singletons :

Théorème 2.5 — *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 : $\mathbb{P}(\llbracket n, +\infty \llbracket) = \lambda a_n$.

2.4 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 2.6 — *$\mathbb{P}_B$ 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 2.7 (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 $\sum_{i\in I}\mathbb{P}(B_i)=1$, autrement dit lorsque l'événement $\bigcup_{i\in I}B_i$ est presque sûr. On parle alors de système *quasi-complet* d'événements.

Proposition 2.8 (Formule de Bayes) — *Soit* $(B_i)_{i\in I}$ *un système complet d'événements tel que pour tout* $i\in I$, $\mathbb{P}(B_i)>0$, *et* A *un événement tel que* $\mathbb{P}(A)>0$. *Alors* $\boxed{\mathbb{P}(B_i\mid A)=\dfrac{\mathbb{P}(B_i)\mathbb{P}(A\mid B_i)}{\sum_{j\in I}\mathbb{P}(B_j)\mathbb{P}(A\mid B_j)}.}$

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

Exercice 4

Un QCM propose 4 réponses pour chaque question. Soit p la probabilité qu'un étudiant connaisse la bonne réponse à une question donnée. S'il ignore la réponse, il choisit au hasard l'une des réponses proposées. Quel est la probabilité qu'un étudiant connaisse vraiment la bonne réponse lorsqu'il a correctement répondu à une question ?

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 $\mathbb{P}(A\mid B)=\mathbb{P}(A)$ et $\mathbb{P}(B\mid A)=\mathbb{P}(B)$, ce qui se traduit par $\dfrac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}=\mathbb{P}(A)$ et $\dfrac{\mathbb{P}(B\cap A)}{\mathbb{P}(A)}=\mathbb{P}(B)$. 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 2.9 — *Si* A *et* B *sont indépendants, il en est de même de* $\overline{A}$ *et* B, *de* A *et* $\overline{B}$, *de* $\overline{A}$ *et* $\overline{B}$.

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* $p\leqslant \operatorname{card} I$, *pour toute* p*-liste* $(i_1,\dots i_p)\in I^p$ *d'indices deux-à-deux distincts,* $\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_p})=\mathbb{P}(A_{i_1})\cdots\mathbb{P}(A_{i_p})$ *(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 ; pour tout $p \in \mathbb{N}$ on pose $B_p = \bigcup_{n \geqslant p} A_n$ puis $A^* = \bigcap_{p \in \mathbb{N}} B_p$.

a. Justifier que $\mathbb{P}(A^*) = \lim_{p \to +\infty} \mathbb{P}(B_p)$ et que $\mathbb{P}(B_p) = 1 - \lim_{n \to +\infty} \prod_{k=p}^{n} \big(1 - \mathbb{P}(A_k)\big)$.

b. Montrer que $\prod_{k=p}^{n} \big(1 - \mathbb{P}(A_k)\big) \leqslant \exp\Big(-\sum_{k=p}^{n} \mathbb{P}(A_k)\Big)$ et en déduire que si la série $\sum \mathbb{P}(A_n)$ diverge, $\mathbb{P}(A^*) = 1$.

Remarque. Le résultat ci-dessus, associé au lemme de Borel-Cantelli (voir page 7) constitue la *loi du zéro-un de Borel* : si (A_n) est une suite d'événements indépendants, la probabilité qu'une infinité d'entre eux se réalise est :

- égale à 0 si la série $\sum \mathbb{P}(A_n)$ converge ;
- égale à 1 si la série $\sum \mathbb{P}(A_n)$ diverge.

3. Variables aléatoires

3.1 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.

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 $X(\Omega)$ (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 : $X^{-1}(\{e\}) = \big\{\omega \in \Omega \;\big|\; X(\omega) = e\big\}$.

Exemples.

– 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 $X(e_1, e_2, e_3) = \text{card}\{i \in [\![1,3]\!] \mid e_i = 6\}$.
$X(\Omega) = \{0,1,2,3\}$ donc la variable aléatoire X est discrète (finie).

– Pour l'expérience consistant à lancer une pièce jusqu'à obtenir Face, nous pouvons choisir $\Omega = \{0,1\}^{\mathbb{N}}$, $E = \mathbb{N} \cup \{+\infty\}$ et $X : \Omega \to E$ définie par $X((u_n)) = \min\{n \in \mathbb{N}^* \mid u_n = 0\}$. 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 3.1 — *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. Autrement dit, cette notion permet d'une certaine façon d'« oublier » l'espace probabilisable $(\Omega, \mathscr{A})$ (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 $[X \in U]$.
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 3.2 — *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ 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 $\big(X(\Omega), \mathscr{P}(X(\Omega))\big)$, 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 Ω. Il existe une réciproque de ce résultat : il est possible de choisir a priori la loi de X et d'en déduire une probabilité sur $X(\Omega)$. De manière plus formelle :

Théorème 3.3 — *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 $\mathbb{P}_X$ sur $\big(X(\Omega), \mathscr{P}(X(\Omega))\big)$ telle que pour tout $i \in I$, $\mathbb{P}_X(X = x_i) = p_i$.*

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 l'espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$.

3.2 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.

Remarque. Deux variables aléatoires X et Y qui suivent la même loi seront notées $X \sim Y$. On notera que si $X \sim Y$ alors pour toute fonction f on a $f(X) \sim f(Y)$.

■ 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 $X \sim \mathscr{U}(n)$.*

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

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

Variable indicatrice associée à un événement
À tout événement $A \in \mathcal{A}$ on peut associer la variable aléatoire $\mathbb{1}_A$ définie par :

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

Cette variable aléatoire $\mathbb{1}_A$ est appelée l'*indicatrice* de A ; elle suit une loi de Bernoulli de paramètre $p = \mathbb{P}(A)$.

■ Loi géométrique

L'expérience type consiste en une succession infinie d'expériences de Bernoulli indépendantes 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* $X \sim \mathscr{G}(p)$.

Proposition 3.4 — *Soit $X \sim \mathscr{G}(p)$. 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 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 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* $\boxed{X(\Omega) = [\![0,n]\!] \text{ et pour tout } k \in [\![0,n]\!],\ \mathbb{P}(X=k) = \binom{n}{k} p^k q^{n-k}}$ *avec $q = 1-p$. On note dans ce cas* $X \sim \mathscr{B}(n,p)$.

■ 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 3.5 (loi des événements rares) — *Soit (X_n) une suite de variables aléatoires réelles telle que pour tout $n \in \mathbb{N}$, $X_n \sim \mathscr{B}(n,p_n)$. On suppose $p_n \sim \dfrac{\lambda}{n}$ avec $\lambda > 0$. 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* $X \sim \mathscr{P}(\lambda)$.

Remarque. Concrètement, ce résultat affirme que si des événements indépendants ont une très faible probabilité d'apparition, leur distribution, qui suit en principe une loi binomiale, est dans la pratique très voisine d'une loi de Poisson. 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 $\mathbb{P}(X > 5) = 1 - \sum_{k=0}^{5} \mathbb{P}(X = k)$.

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 :

```
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 ?

3.3 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 :* $\mathbb{P}(X = x \text{ et } Y = y) = \mathbb{P}(X = x)\cdot\mathbb{P}(Y = y)$. *L'indépendance des deux variables aléatoires* X *et* Y *sera notée* $X \perp\!\!\!\perp Y$.

Proposition 3.6 — *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 3.7 — *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. Autrement dit,*

$$X \perp\!\!\!\perp Y \implies f(X) \perp\!\!\!\perp g(Y)$$

Exercice 10

Soient $X \sim \mathscr{P}(\lambda)$ et $Y \sim \mathscr{P}(\mu)$ deux variables aléatoires indépendantes. Montrer que $X + Y \sim \mathscr{P}(\lambda + \mu)$.

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, lorsque $X = n$, suit une loi binomiale $\mathscr{B}(n, p)$. On pose enfin $Z = X - Y$.
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. Si $(X_k)_{1\leqslant k\leqslant n}$ est une famille finie de variables aléatoires mutuellement indépendantes suivant toutes la même loi de Bernoulli de paramètre p, alors $S_n = \sum_{k=1}^{n} X_k$ suit une loi binomiale de paramètre (n, p).

Plus généralement, nous rencontrerons fréquemment des familles de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ indépendantes suivant toutes la même loi. Une telle suite de variables aléatoires sera dite *identiquement distribuée*.

Pour finir, nous admettrons le résultat suivant :

Théorème 3.8 (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.*

On peut bien entendu étendre ce résultat à plus de deux coalitions.

3.4 Espérance

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

Lorsque $X(\Omega)$ est infini, nous avons vu dans la première partie de ce chapitre que pour pouvoir donner un sens à cette expression, il fallait pouvoir s'assurer que cette expression ne dépend pas de l'ordre d'indexation choisi pour $X(\Omega)$. Ceci nous conduit à la :

Définition. — *On dit qu'une variable aléatoire réelle et discrète* X *est d'*espérance finie *lorsque la famille de nombres réels* $\big(x\mathbb{P}(X = x)\big)_{x\in X(\Omega)}$ *est sommable, et on appelle dans ce cas on appelle* espérance *de* X *la quantité*

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

Remarque. Lorsqu'on décrit par compréhension l'ensemble $X(\Omega) = \left\{x_n \;\middle|\; n \in \mathbb{N}\right\}$, X admet une espérance si et seulement si la série $\sum x_n \mathbb{P}(X = x_n)$ est *absolument* convergente.

Remarque. Lorsque la variable aléatoire X est à valeurs positives, nous avons vu que l'on pouvait noter $\sum_{x \in X(\Omega)} x\mathbb{P}(X = x) = +\infty$ lorsque cette famille n'est pas sommable. On notera alors $\mathbb{E}(X) = +\infty$ dans ce cas de figure (attention, ceci n'est pas valable lorsque X n'est pas à valeurs positives).

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

$$\mathbb{E}(X) = \sum_{n=1}^{+\infty} np(1-p)^{n-1} = \frac{p}{(1-(1-p))^2} = \frac{1}{p}$$

Proposition 3.9 — *Soit* X *une variable aléatoire presque sûrement bornée (autrement dit, il existe* $M > 0$ *et que* $\mathbb{P}\left(|X| \leqslant M\right) = 1$*). Alors* X *est d'espérance finie.*

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

Proposition 3.10 — *Si* X *est une variable aléatoire à valeurs dans* $\mathbb{N} \cup \{+\infty\}$ *alors* $\mathbb{E}(X) = \sum_{n=1}^{+\infty} \mathbb{P}(X \geqslant n)$.

Les principaux résultats de l'espérance sont les suivants :

Théorème 3.11 (de transfert) — *Si* f *une application à valeurs réelles définie sur* $X(\Omega)$*, alors* $f(X)$ *est d'espérance finie si et seulement si la famille* $\left(f(x)\mathbb{P}(X = x)\right)_{x \in X(\Omega)}$ *est sommable, et dans ce cas,* $\mathbb{E}(f(X)) = \sum_{x \in X(\Omega)} f(x)\mathbb{P}(X = x)$.

Corollaire — X *est d'espérance finie il en est de même de* $|X|$*, et* $\left|\mathbb{E}(X)\right| \leqslant \mathbb{E}(|X|)$.

Proposition 3.12 — *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,* $\mathbb{E}(X) \geqslant 0$ *(positivité de l'espérance) ;*

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

■ Espérances des lois usuelles

Toutes les lois que nous avons étudiées à la section 3.2 sont d'espérance finie, et la valeur de leur espérance doit être connue.

Loi uniforme Si $X \sim \mathcal{U}(n)$, alors $\mathbb{E}(X) = \dfrac{n+1}{2}$.
Loi de Bernoulli Si $X \sim \mathcal{B}(p)$, alors $\mathbb{E}(X) = p$.
Loi géométrique Si $X \sim \mathcal{G}(p)$ alors $\mathbb{E}(X) = \dfrac{1}{p}$.
Loi binomiale Si $X \sim \mathcal{B}(n,p)$ alors $\mathbb{E}(X) = np$.
Loi de Poisson Si $X \sim \mathcal{P}(\lambda)$ alors $\mathbb{E}(X) = \lambda$.

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

3.5 Variance et écart type

Définition. — *On dit qu'une variable aléatoire réelle* X *possède* un moment d'ordre 1 *lorsque* X *est d'espérance finie, et un moment d'ordre 2 lorsque* X^2 *est d'espérance finie.*

Théorème 3.13 — *Si la variable aléatoire* X *possède un moment d'ordre 2 alors* X *possède un moment d'ordre 1, et dans ce cas, on appelle* variance *de* X *la quantité* $\mathbb{V}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$, *et* écart type *la quantité* $\sigma(X) = \sqrt{\mathbb{V}(X)}$.

Proposition 3.14 (Formule de Koenig-Huyghens) — *Lorsque* X *possède un moment d'ordre 2,*

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

Proposition 3.15 — *Si* a *et* b *sont deux réels et* X *une variable aléatoire réelle possédant un moment d'ordre 2, alors* $\mathbb{V}(aX + b) = a^2\mathbb{V}(X)$ *et* $\sigma(aX + b) = |a|\sigma(X)$.

Théorème 3.16 — *Si* X *et* Y *sont deux variables aléatoires* indépendantes *admettant un moment d'ordre 2, il en est de même de* X + Y, *et* $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$.

Exercice 12

Soit X une variable aléatoire possédant un moment d'ordre 2. 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 3.2 possèdent une variance (à connaître) :

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

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

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

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

Loi de Poisson Si $X \sim \mathcal{P}(\lambda)$ 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 $Y = \dfrac{X - \mathbb{E}(X)}{\sigma(X)}$ possède une espérance nulle (on dit qu'elle est *centrée*) et un écart type égal à 1 (on dit qu'elle est *réduite*).

L'application $(X, Y) \mapsto \mathbb{E}(XY)$ est une application bilinéaire, symétrique et positive. En conséquence de quoi il est possible d'établir le résultat suivant :

Théorème 3.17 (Inégalité de Cauchy-Schwarz) — *Si* X *et* Y *possèdent des moments d'ordre 2, alors* XY *possède un moment d'ordre 1, et* $\boxed{\mathbb{E}(XY)^2 \leqslant \mathbb{E}(X^2)\mathbb{E}(Y^2).}$

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 $\mathbb{P}(\lambda X + \mu Y = 0) = 1$.

3.6 Covariance

Nous avons démontré à la proposition 3.12 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 plus loin), 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* $\boxed{\operatorname{cov}(X,Y) = \mathbb{E}\Big((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\Big).}$

Le théorème 3.17 nous permet d'énoncer :

Proposition 3.18 — *Si* X *et* Y *sont deux variables aléatoires réelles possédant un moment d'ordre 2 alors* $\operatorname{cov}(X,Y)$ *existe.*

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

Proposition 3.19 — *Lorsque* X *et* Y *sont deux variables aléatoires réelles indépendantes possédant un moment d'ordre 2, alors* $\operatorname{cov}(X,Y) = 0$. *On dira que* X *et* Y ne sont pas corrélées.

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

Proposition 3.20 (propriétés de la covariance) — *Soient* X Y *et* Z *trois variables aléatoires réelles possédant des moments d'ordre 2. Alors :*

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

Théorème 3.21 — *Soient* X *et* Y *deux variables aléatoires réelles possédant des moments d'ordre 2. Alors*

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

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

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

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, $\mathbb{V}(X_1 + \cdots + X_n) = \mathbb{V}(X_1) + \cdots + \mathbb{V}(X_n)$.

3.7 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 3.22 (inégalité de Markov) — *Soit* X *une variable aléatoire positive possédant un moment d'ordre 1, et* $a > 0$. *Alors* $\boxed{\mathbb{P}(X \geqslant a) \leqslant \dfrac{\mathbb{E}(X)}{a}.}$

Théorème 3.23 (inégalité de Bienaymé-Tchebychev) — *Soit* X *une variable aléatoire réelle admettant un moment d'ordre 2, et* $\alpha > 0$. *Alors* $\boxed{\mathbb{P}\Big(\big|X - \mathbb{E}(X)\big| \geqslant \alpha\Big) \leqslant \dfrac{\mathbb{V}(X)}{\alpha^2} = \dfrac{\sigma(X)^2}{\alpha^2}.}$

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.

■ Loi faible des grands nombres

Le théorème que nous allons énoncer ensuite va justifier la démarche expérimentale qu'on utilise pour estimer empiriquement une espérance : on réalise un grand nombre d'expérience (en général par le biais d'une simulation numérique) puis on calcule la moyenne des valeurs qu'a prise la variable aléatoire X. Par exemple, pour estimer l'espérance d'une loi géométrique de paramètre 1/10 on réalise le script suivant :

```
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
```

Nous obtenons effectivement une valeur proche de l'espérance théorique égale à 10. Le théorème qui suit donne une justification à cet état de fait :

Théorème 3.24 (loi faible des grands nombres) — *Soit* $(X_n)_{n\geqslant 1}$ *une suites de variables aléatoires deux-à-deux indépendantes et de même loi admettant un moment d'ordre 2. On pose* $S_n = X_1 + X_2 + \cdots + X_n$. *Alors pour tout* $\epsilon > 0$,

$$\boxed{\mathbb{P}\left(\left|\frac{1}{n}S_n - \mathbb{E}(X)\right| \geqslant \epsilon\right) \leqslant \frac{\mathbb{V}(X)}{n\epsilon^2}.}$$

Remarque. Avec les mêmes hypothèses on en déduit que : $\lim\limits_{n\to+\infty} \mathbb{P}\left(\left|\dfrac{1}{n}S_n - \mathbb{E}(X)\right| \geqslant \epsilon\right) = 0$.
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 ϵ est faible.

Exemple. Dans l'exemple numérique ci-dessus nous avons pris $n = 100\,000$ et nous avons $\mathbb{E}(X) = 10$ et $V(X) = 90$. Pour $\epsilon = 0{,}1$ nous avons $\dfrac{\mathbb{V}(X)}{n\epsilon^2} = 0{,}09$ donc il y a plus de 91% de chance que le résultat obtenu diffère de l'espérance théorique de moins de 0,1.

3.8 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 $\mathscr{C}^\infty$ 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 3.25 — 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, $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \sum_{n=1}^{+\infty} n^2\mathbb{P}(X=n) - \mathbb{E}(X)^2 = \sum_{n=1}^{+\infty} n(n-1)\mathbb{P}(X=n) + \sum_{n=1}^{+\infty} n\mathbb{P}(X=n) - \mathbb{E}(X)^2$

$$= G''_X(1) + G'_X(1) - G'_X(1)^2.$$

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

Théorème 3.26 — 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 $X \sim \mathscr{U}(n)$, alors $G_X(t) = \frac{1}{n}(t + t^2 + \cdots + t^n)$.

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

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

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

Loi de Poisson Si $X \sim \mathscr{P}(\lambda)$ 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 3.27 — *Si X et Y sont deux variables aléatoires indépendantes à valeur entières, alors*

$$\forall t \in]-1,1[, \quad \boxed{G_{X+Y}(t) = G_X(t)G_Y(t).}$$

Exemple.

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