PROBABILITÉS

B. Landelle

Table des matières

I Espaces probabilisés

1 Tribu, événements

Définition 1. *Soit* Ω *un ensemble non vide appelé* univers. *Une tribu sur* Ω *est une partie* $\mathscr{A}$ *de* $\mathcal{P}(\Omega)$ *telle que :*

1. $\Omega \in \mathscr{A}$ *; (événement certain dans la tribu)*
2. *Pour tout* $\mathrm{A} \in \mathscr{A}$*, on a* $\bar{\mathrm{A}} \in \mathscr{A}$ *; (stabilité par complémentation)*
3. *Pour toute suite* $(\mathrm{A}_n)_{n\in\mathbb{N}} \in \mathscr{A}^{\mathbb{N}}$*, on a* $\displaystyle\bigcup_{n=0}^{+\infty} \mathrm{A}_n \in \mathscr{A}$*. (stabilité par union dénombrable)*

Le couple $(\Omega, \mathscr{A})$ *est dit* espace probabilisable.

Exemples : La famille $\{\varnothing, \Omega\}$ est une tribu dite *grossière*.
La famille $\mathcal{P}(\Omega)$ est une tribu dite *discrète* (celle sous-jacente au cas d'un univers fini).

Remarques : (1) On a $\varnothing \in \mathscr{A}$ par stabilité par complémentation.
(2) Si Ω est fini ou dénombrable, la tribu discrète est la tribu naturelle pour travailler dans le cadre probabiliste. Il existe en revanche des situations plus élaborées ($\Omega = \mathbb{R}$ par exemple) où le choix de tribu adaptée n'est plus $\mathcal{P}(\Omega)$ mais ceci dépasse le cadre de ce cours.

Définition 2. *Soit* $(\Omega, \mathscr{A})$ *un espace probabilisable. On appelle* événement *un élément de la tribu* $\mathscr{A}$*.*

Définition 3. *Soit* $(\Omega, \mathscr{A})$ *un espace probabilisable. Deux événements* A *et* B *sont dits* incompatibles *ou* disjoints *si*

$$\mathrm{A} \cap \mathrm{B} = \varnothing$$

Définition 4. *Soit* $(\Omega, \mathscr{A})$ *un espace probabilisable. On appelle* système complet d'événements *une famille d'événements* $(\mathrm{A}_n)_{n\in\mathbb{N}}$ *vérifiant*

$$\bigcup_{n=0}^{+\infty} \mathrm{A}_n = \Omega \quad \textit{et} \quad \forall (i,j) \in \mathbb{N}^2 \quad \textit{avec} \quad i \neq j \qquad \mathrm{A}_i \cap \mathrm{A}_j = \varnothing$$

Notation : On note $\displaystyle\bigsqcup_{n=0}^{+\infty} \mathrm{A}_n$ l'union disjointe des A_n, notation non officielle mais bien commode.

Remarque : Le cas d'une famille finie d'événements est couvert par la définition en prenant $\mathrm{A}_n = \varnothing$ pour n supérieur à un certain rang.

Exemple : On lance une pièce indéfiniment. Soit l'événement A_k : obtenir pile en k lancers exactement et A_∞ : ne pas obtenir pile. La famille $(\mathrm{A}_k)_{k\in\mathbb{N}^*\cup\{\infty\}}$ est un système complet d'événements. Formellement, on a $\Omega = \{0,1\}^{\mathbb{N}^*}$, pour k entier non nul $\mathrm{A}_k = \{0\}^{[\![1\,;\,k-1]\!]} \times \{1\}^{\{k\}} \times \{0,1\}^{[\![k+1\,;\,+\infty [\![}$ et $\mathrm{A}_\infty = \{0\}^{\mathbb{N}^*}$. Cette situation, très simple en apparence, est délicate : l'univers Ω n'est pas dénombrable (argument diagonale de Cantor).

2 Probabilité

Définition 5. *Soit* $(\Omega, \mathscr{A})$ *un espace probabilisable. On appelle* probabilité *sur* $(\Omega, \mathscr{A})$ *une application* $\mathbb{P} : \mathscr{A} \to [0\,;1]$ *vérifiant :*

1. $\mathbb{P}(\Omega) = 1$
2. *pour toute suite* $(\mathrm{A}_n)_{n\in\mathbb{N}}$ *d'événements (deux à deux) incompatibles, on a*

$$\mathbb{P}\left(\bigsqcup_{n=0}^{+\infty} \mathrm{A}_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{A}_n) \qquad \textit{σ-additivité}$$

Remarque : La propriété de σ-additivité donne implicitement la convergence de la série $\sum \mathbb{P}(\mathrm{A}_n)$ pour des A_n incompatibles. Comme précédemment, le cas d'une famille finie est couvert en considérant $\mathrm{A}_n = \varnothing$ pour n plus grand qu'un certain rang ($\mathbb{P}(\varnothing) = 0$ puisque $\varnothing = \bigsqcup_{n=0}^{+\infty} \varnothing$).

Définition 6. *Soit* $(\Omega, \mathscr{A})$ *un espace probabilisable et* $\mathbb{P}$ *une probabilité sur* $(\Omega, \mathscr{A})$*. Le triplet* $(\Omega, \mathscr{A}, \mathbb{P})$ *est appelé* espace probabilisé.

Remarque : Cette définition étend celle du cas d'un univers Ω fini muni de la tribu $\mathcal{P}(\Omega)$.

Proposition 1. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé. On a*

1. $$\forall (\mathrm{A}, \mathrm{B}) \in \mathscr{A}^2 \qquad \mathbb{P}(\mathrm{A} \cup \mathrm{B}) = \mathbb{P}(\mathrm{A}) + \mathbb{P}(\mathrm{B}) - \mathbb{P}(\mathrm{A} \cap \mathrm{B})$$
2. $$\forall \mathrm{A} \in \mathscr{A} \qquad \mathbb{P}(\bar{\mathrm{A}}) = 1 - \mathbb{P}(\mathrm{A})$$
3. $$\forall (\mathrm{A}, \mathrm{B}) \in \mathscr{A}^2 \qquad \mathrm{A} \subset \mathrm{B} \Longrightarrow \mathbb{P}(\mathrm{A}) \leqslant \mathbb{P}(\mathrm{B}) \qquad \textit{Croissance}$$

Démonstration. 1. On a $\mathrm{A} \cup \mathrm{B} = \mathrm{A} \sqcup (\mathrm{B} \cap \bar{\mathrm{A}})$ et cette union est disjointe d'où

$$\mathbb{P}(\mathrm{A} \cup \mathrm{B}) = \mathbb{P}(\mathrm{A}) + \mathbb{P}(\mathrm{B} \cap \bar{\mathrm{A}})$$

Puis, avec l'union disjointe $\mathrm{B} = (\mathrm{B} \cap \mathrm{A}) \sqcup (\mathrm{B} \cap \bar{\mathrm{A}})$, on obtient

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

d'où le résultat.
2. On a $\Omega = \mathrm{A} \sqcup \bar{\mathrm{A}}$ et cette union est disjointe. Le résultat suit.
3. Comme $\mathrm{A} \subset \mathrm{B}$, on a $\mathrm{B} = \mathrm{A} \sqcup (\mathrm{B} \cap \bar{\mathrm{A}})$, union disjointe d'où

$$\mathbb{P}(\mathrm{B}) = \mathbb{P}(\mathrm{A}) + \mathbb{P}(\mathrm{B} \cap \bar{\mathrm{A}}) \geqslant \mathbb{P}(\mathrm{A})$$

□

Proposition 2 (Inégalité de Boole finie ou sous-additivité). *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* $(\mathrm{A}_k)_{k\in[\![0\,;n]\!]}$ *une suite d'événements. On a*

$$\mathbb{P}\left(\bigcup_{k=0}^{n} \mathrm{A}_k\right) \leqslant \sum_{k=0}^{n} \mathbb{P}(\mathrm{A}_k)$$

Démonstration. Par récurrence. □

Proposition 3. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé. On a*

$$\forall (\mathrm{A}, \mathrm{B}) \in \mathscr{A}^2 \qquad \mathbb{P}(\mathrm{A} \smallsetminus \mathrm{B}) = \mathbb{P}(\mathrm{A}) - \mathbb{P}(\mathrm{A} \cap \mathrm{B})$$

En particulier $\quad \forall (\mathrm{A}, \mathrm{B}) \in \mathscr{A}^2 \quad \mathrm{B} \subset \mathrm{A} \Longrightarrow \mathbb{P}(\mathrm{A} \smallsetminus \mathrm{B}) = \mathbb{P}(\mathrm{A}) - \mathbb{P}(\mathrm{B})$

Démonstration. On a l'union disjointe

$$\mathrm{A} = (\mathrm{A} \cap \mathrm{B}) \sqcup (\mathrm{A} \cap \bar{\mathrm{B}}) = (\mathrm{A} \cap \mathrm{B}) \cup (\mathrm{A} \smallsetminus \mathrm{B})$$

Le résultat suit. □

Théorème 1. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé. Pour toute suite* $(\mathrm{A}_n)_{n \in \mathbb{N}}$ *d'événements, on a*

1. $\displaystyle\bigcap_{n=0}^{+\infty} \mathrm{A}_n \in \mathscr{A}$ *; (stabilité par intersection dénombrable)*
2. *continuité croissante : si* $\mathrm{A}_n \subset \mathrm{A}_{n+1}$ *pour tout* n*, alors*
$$\lim_{n \to +\infty} \mathbb{P}(\mathrm{A}_n) = \mathbb{P}\left(\bigcup_{n=0}^{+\infty} \mathrm{A}_n\right)$$
3. *continuité décroissante : si* $\mathrm{A}_{n+1} \subset \mathrm{A}_n$ *pour tout* n*, alors*
$$\lim_{n \to +\infty} \mathbb{P}(\mathrm{A}_n) = \mathbb{P}\left(\bigcap_{n=0}^{+\infty} \mathrm{A}_n\right)$$

Démonstration. 1. On a $\overline{\mathrm{A}_n} \in \mathscr{A}$ pour tout n entier (stabilité par complémentation). Puis, par union dénombrable, $\displaystyle\bigcup_{n=0}^{+\infty} \overline{\mathrm{A}_n} \in \mathscr{A}$ et par stabilité par complémentation, $\displaystyle\Omega \smallsetminus \bigcup_{n=0}^{+\infty} \overline{\mathrm{A}_n} = \bigcap_{n=0}^{+\infty} \mathrm{A}_n \in \mathscr{A}$.

2. Posons $\mathrm{A}_{-1} = \varnothing$ puis $\mathrm{B}_n = \mathrm{A}_n \smallsetminus \mathrm{A}_{n-1}$ pour tout $n \in \mathbb{N}$. Pour n entier non nul, on a $\mathrm{B}_n \subset \mathrm{A}_n$ et $\mathrm{B}_n \cap \mathrm{A}_{n-1} = \varnothing$ d'où $\mathrm{B}_n \cap \mathrm{B}_k = \varnothing$ pour tout $k < n$. Par suite, les B_n sont incompatibles et par construction, on a $\displaystyle\mathrm{A}_n = \bigsqcup_{k=0}^{n} \mathrm{B}_k$ pour tout $n \in \mathbb{N}$ (par récurrence ou pour $x \in \mathrm{A}_n$, considérer $k = \min\{i \in [\![0\,;\, n]\!] \mid x \in \mathrm{A}_i\}$) et $\displaystyle\bigcup_{n=0}^{+\infty} \mathrm{A}_n = \bigsqcup_{n=0}^{+\infty} \mathrm{B}_n$. Par suite

$$\mathbb{P}(\mathrm{A}_n) = \mathbb{P}\left(\bigsqcup_{k=0}^{n} \mathrm{B}_k\right) = \sum_{k=0}^{n} \mathbb{P}(\mathrm{B}_k) \xrightarrow[n \to \infty]{} \sum_{k=0}^{+\infty} \mathbb{P}(\mathrm{B}_k) = \mathbb{P}\left(\bigsqcup_{n=0}^{+\infty} \mathrm{B}_n\right) = \mathbb{P}\left(\bigcup_{n=0}^{+\infty} \mathrm{A}_n\right)$$

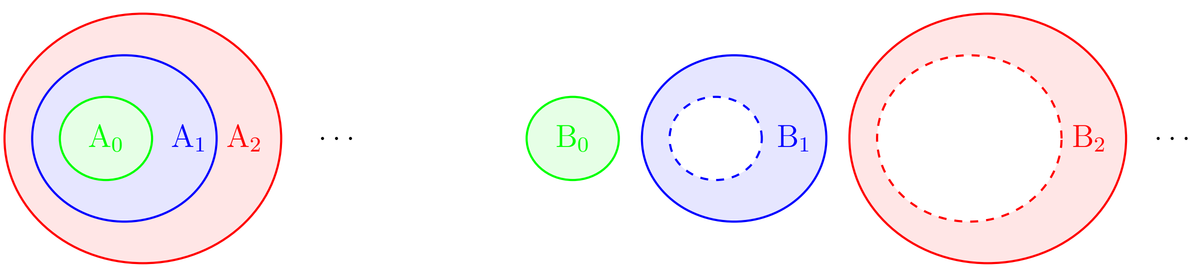

Figure 1 – Famille croissante $(\mathrm{A}_n)_n$, famille disjointe $(\mathrm{B}_n)_n$

3. Il suffit d'appliquer le résultat précédent sur les ensembles $\overline{\mathrm{A}_n}$.

$\square$

Corollaire 1. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé. Pour toute suite $(\mathrm{A}_n)_{n\in\mathbb{N}}$ d'événements, on a*

$$\mathbb{P}\left(\bigcap_{k=0}^{n} \mathrm{A}_k\right) \xrightarrow[n\to\infty]{} \mathbb{P}\left(\bigcap_{k=0}^{+\infty} \mathrm{A}_k\right) \quad \textit{et} \quad \mathbb{P}\left(\bigcup_{k=0}^{n} \mathrm{A}_k\right) \xrightarrow[n\to\infty]{} \mathbb{P}\left(\bigcup_{k=0}^{+\infty} \mathrm{A}_k\right)$$

Démonstration. La suite $\left(\bigcap_{k=0}^{n} \mathrm{A}_k\right)_n$ décroît avec $\bigcap_{n=0}^{+\infty}\bigcap_{k=0}^{n} \mathrm{A}_k = \bigcap_{k=0}^{+\infty} \mathrm{A}_k$ et la suite $\left(\bigcup_{k=0}^{n} \mathrm{A}_k\right)_n$ croît avec $\bigcup_{n=0}^{+\infty}\bigcup_{k=0}^{n} \mathrm{A}_k = \bigcup_{k=0}^{+\infty} \mathrm{A}_k$. Ainsi, par continuité décroissante et croissante

$$\mathbb{P}\left(\bigcap_{k=0}^{n} \mathrm{A}_k\right) \xrightarrow[n\to\infty]{} \mathbb{P}\left(\bigcap_{k=0}^{+\infty} \mathrm{A}_k\right) \quad \text{et} \quad \mathbb{P}\left(\bigcup_{k=0}^{n} \mathrm{A}_k\right) \xrightarrow[n\to\infty]{} \mathbb{P}\left(\bigcup_{k=0}^{+\infty} \mathrm{A}_k\right)$$

$\square$

Proposition 4 (Inégalité de Boole ou sous-additivité). *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé. Pour toute suite $(\mathrm{A}_n)_{n\in\mathbb{N}}$ d'événements, on a*

$$\mathbb{P}\left(\bigcup_{n=0}^{+\infty} \mathrm{A}_n\right) \leqslant \sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{A}_n)$$

Démonstration. D'après l'inégalité de Boole finie, on a

$$\forall \mathrm{N} \in \mathbb{N} \qquad \mathbb{P}\left(\bigcup_{n=0}^{\mathrm{N}} \mathrm{A}_n\right) \leqslant \sum_{n=0}^{\mathrm{N}} \mathbb{P}(\mathrm{A}_n) \leqslant \underbrace{\sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{A}_n)}_{\in[0\,;+\infty]}$$

Faisant tendre $\mathrm{N} \to +\infty$ d'après le résultat du corollaire précédent, l'inégalité suit. $\square$

Définition 7. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé. Un événement A est dit* négligeable *si* $\mathbb{P}(\mathrm{A}) = 0$.

Proposition 5. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé.*

1. *Un événement inclus dans un événement négligeable est négligeable.*
2. *Une réunion finie ou dénombrable d'événements négligeables est négligeable.*

Démonstration. 1. Immédiate par croissance de $\mathbb{P}$.
2. Conséquence de l'inégalité de Boole. $\square$

Définition 8. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé. Un événement A est dit* presque sûr *si* $\mathbb{P}(\mathrm{A}) = 1$. *Une propriété $\mathscr{P}$ est dite* presque sûre *ou* réalisée presque sûrement *si l'événement $\{\mathscr{P}$ vraie$\}$ est presque sûr.*

Proposition 6. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé.*

1. *Un événement contenant un événement presque sûr est presque sûr.*
2. *Une intersection finie ou dénombrable d'événements presque sûrs est un événement presque sûr.*

Démonstration. Par complémentation avec le résultat de la proposition 5. □

Définition 9. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé. On appelle* système quasi-complet d'événements *une famille* $(\mathrm{A}_n)_n$ *d'événements incompatibles telle que* $\bigsqcup_{n\in\mathbb{N}} \mathrm{A}_n$ *est un événement presque sûr.*

Remarque : Un système complet est quasi-complet.

Définition 10. *Soit* Ω *un ensemble. Une* distribution de probabilité discrète *sur* Ω *est une famille* $(\mathbb{P}(\{\omega\}))_{\omega\in\Omega}$ *à valeurs dans* $\mathbb{R}_+$ *telle que* $\sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = 1$. *Le* support *d'une distribution de probabilité discrète est l'ensemble*

$$\{\omega \in \Omega \ : \ \mathbb{P}(\{\omega\}) > 0\}$$

Remarque : L'existence d'une distribution de probabilité discrète sur Ω impose $\Omega \neq \varnothing$ sans quoi on aurait $\sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = 0$.

Proposition 7. *Soit* Ω *un ensemble et* $(\mathbb{P}(\{\omega\}))_{\omega\in\Omega}$ *une distribution de probabilité discrète sur* Ω. *Son support est au plus dénombrable.*

Démonstration. Résultat établi dans le chapitre **Familles sommables**. □

Proposition 8. *Soit* Ω *un ensemble,* $\mathscr{A} = \mathscr{P}(\Omega)$ *et* $(\mathbb{P}(\{\omega\}))_{\omega\in\Omega}$ *une distribution de probabilité discrète. On définit une probabilité* $\mathbb{P}$ *sur* $(\Omega, \mathscr{A})$ *par*

$$\forall \mathrm{A} \in \mathscr{A} \qquad \mathbb{P}(\mathrm{A}) = \sum_{\omega\in\mathrm{A}} \mathbb{P}(\{\omega\})$$

Démonstration. Par construction de $\mathbb{P}$, on a $\mathbb{P}$ à valeurs dans $[0\,;1]$, $\mathbb{P}(\Omega) = 1$ et pour $(\mathrm{A}_n)_n$ une suite d'événements incompatibles, il vient par sommation par paquets pour une famille à termes positifs

$$\mathbb{P}\left(\bigsqcup_{n=0}^{+\infty} \mathrm{A}_n\right) = \sum_{\omega\in\bigsqcup_{n=0}^{+\infty}\mathrm{A}_n} \mathbb{P}(\{\omega\}) = \sum_{n=0}^{+\infty}\left(\sum_{\omega\in\mathrm{A}_n} \mathbb{P}(\{\omega\})\right) = \sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{A}_n)$$

□

Remarques : (1) Si Ω est au plus dénombrable, toute probabilité sur $(\Omega, \mathscr{A})$ avec $\mathscr{A} = \mathscr{P}(\Omega)$ est obtenue à partir d'une distribution de probabilités discrètes selon la construction ci-avant. En effet, étant donnée $\mathbb{P}$ une probabilité sur $(\Omega, \mathscr{A})$, pour $\mathrm{A} \in \mathscr{A}$, l'événement A est au plus dénombrable et par σ-additivité

$$\mathbb{P}(\mathrm{A}) = \mathbb{P}\left(\bigsqcup_{\omega\in\mathrm{A}} \{\omega\}\right) = \sum_{\omega\in\mathrm{A}} \mathbb{P}(\{\omega\})$$

(2) Cette notion de distribution de probabilité discrète est très limitée. Par exemple, elle ne couvre pas le cas du jeu de pile/face infini : on a $\Omega = \{0,1\}^{\mathbb{N}^*}$ non dénombrable, univers pour lequel on ne choisit pas $\mathscr{P}(\Omega)$ comme tribu afin d'éviter des situations paradoxales.

3 Probabilité conditionnelle

Définition 11. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* B *un événement vérifiant* $\mathbb{P}(\mathrm{B}) > 0$. *Pour* A *événement, on définit la* probabilité conditionnelle *de* A *sachant* B *notée* $\mathbb{P}(\mathrm{A}|\mathrm{B})$ *ou* $\mathbb{P}_{\mathrm{B}}(\mathrm{A})$ *par*

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

Théorème 2. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* B *un événement vérifiant* $\mathbb{P}(\mathrm{B}) > 0$. *L'application* $\mathbb{P}_{\mathrm{B}}$ *est une probabilité sur* $(\Omega, \mathscr{A})$.

Démonstration. On a $\mathbb{P}_{\mathrm{B}}$ à valeurs dans $[\,0\,;1\,]$ puisque $\mathbb{P}(\mathrm{A} \cap \mathrm{B}) \leqslant \mathbb{P}(\mathrm{B})$ pour $\mathrm{A} \in \mathscr{A}$ par croissance de $\mathbb{P}$ puis $\mathbb{P}_{\mathrm{B}}(\Omega) = \mathbb{P}(\Omega \cap \mathrm{B})/\mathbb{P}(\mathrm{B}) = \mathbb{P}(\mathrm{B})/\mathbb{P}(\mathrm{B}) = 1$ et la propriété de σ-additivité est clairement héritée. □

Vocabulaire : La probabilité $\mathbb{P}_{\mathrm{B}}$ est dite probabilité *a priori*.

Proposition 9 (Formules des probabilités composées). *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* $(\mathrm{A}_i)_{i \in [\![1\,;n]\!]}$ *des événements tels que* $\mathbb{P}\left(\mathrm{A}_1 \cap \ldots \cap \mathrm{A}_{n-1}\right) > 0$. *On a*

$$\mathbb{P}\left(\mathrm{A}_1 \cap \ldots \cap \mathrm{A}_n\right) = \mathbb{P}(\mathrm{A}_1) \times \mathbb{P}(\mathrm{A}_2|\mathrm{A}_1) \times \ldots \times \mathbb{P}(\mathrm{A}_n|\mathrm{A}_1 \cap \ldots \cap \mathrm{A}_{n-1})$$

Démonstration. Comme $\mathrm{A}_1 \cap \ldots \cap \mathrm{A}_{n-1} \subset \mathrm{A}_1 \cap \ldots \cap \mathrm{A}_{n-2} \subset \ldots \subset \mathrm{A}_1$ et $\mathbb{P}\left(\mathrm{A}_1 \cap \ldots \cap \mathrm{A}_{n-1}\right) > 0$, on peut conditionner par chacun de ces événements. Il s'agit ensuite d'un simple produit téléscopique :

$$\begin{aligned}
&\mathbb{P}(\mathrm{A}_1) \times \mathbb{P}(\mathrm{A}_2|\mathrm{A}_1) \times \ldots \times \mathbb{P}(\mathrm{A}_n|\mathrm{A}_1 \cap \ldots \cap \mathrm{A}_{n-1}) \\
&\quad = \mathbb{P}(\mathrm{A}_1) \times \frac{\mathbb{P}(\mathrm{A}_1 \cap \mathrm{A}_2)}{\mathbb{P}(\mathrm{A}_1)} \times \frac{\mathbb{P}(\mathrm{A}_1 \cap \mathrm{A}_2 \cap \mathrm{A}_3)}{\mathbb{P}(\mathrm{A}_1 \cap \mathrm{A}_2)} \times \ldots \times \frac{\mathbb{P}(\mathrm{A}_1 \cap \ldots \cap \mathrm{A}_n)}{\mathbb{P}(\mathrm{A}_1 \cap \ldots \cap \mathrm{A}_{n-1})} = \mathbb{P}(\mathrm{A}_1 \cap \ldots \cap \mathrm{A}_n)
\end{aligned}$$

□

Remarque : La situation typique d'utilisation des probabilités composées est celle de tirages successifs dans une urne avec évolution de la composition de l'urne (sans remise, ou avec remise selon résultat du tirage).

Théorème 3 (Formules des probabilités totales). *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* $(\mathrm{A}_n)_{n \in \mathbb{N}}$ *un système quasi-complet d'événements. Pour* $\mathrm{B} \in \mathscr{A}$, *on a*

$$\mathbb{P}(\mathrm{B}) = \sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{B} \cap \mathrm{A}_n) = \sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{B}|\mathrm{A}_n)\mathbb{P}(\mathrm{A}_n)$$

avec pour convention $\mathbb{P}(\mathrm{B}|\mathrm{A}_n)\mathbb{P}(\mathrm{A}_n) = 0$ *si* $\mathbb{P}(\mathrm{A}_n) = 0$.

Démonstration. Notons $\mathrm{A} = \displaystyle\bigsqcup_{n=0}^{+\infty} \mathrm{A}_n$. Considérant le système complet $\left\{\mathrm{A}, \bar{\mathrm{A}}\right\}$, il vient

$$\mathbb{P}(\mathrm{B}) = \mathbb{P}(\mathrm{B} \cap \mathrm{A}) + \underbrace{\mathbb{P}(\mathrm{B} \cap \bar{\mathrm{A}})}_{\leqslant \mathbb{P}(\bar{\mathrm{A}}) = 1 - \mathbb{P}(\mathrm{A}) = 0} = \mathbb{P}(\mathrm{B} \cap \mathrm{A}) = \mathbb{P}\left(\bigsqcup_{n=0}^{+\infty} (\mathrm{B} \cap \mathrm{A}_n)\right) = \sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{B} \cap \mathrm{A}_n)$$

d'où le résultat annoncé. □

Théorème 4 (Formules de Bayes). *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé.*

1. *Soient* A, B *des événements tels que* $\mathbb{P}(\mathrm{B}) > 0$. *On a*

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

2. *Soit* $(\mathrm{A}_n)_{n\in\mathbb{N}}$ *un système quasi-complet d'événements et* B *un événement tel que* $\mathbb{P}(\mathrm{B}) > 0$. *On a*

$$\forall k \in \mathbb{N} \qquad \mathbb{P}(\mathrm{A}_k|\mathrm{B}) = \frac{\mathbb{P}(\mathrm{B}|\mathrm{A}_k)\mathbb{P}(\mathrm{A}_k)}{\sum\limits_{n=0}^{+\infty} \mathbb{P}(\mathrm{B}|\mathrm{A}_n)\mathbb{P}(\mathrm{A}_n)}$$

avec la convention mentionnée dans le théorème 3.

Démonstration. 1. Par définition, on a

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

2. On procède comme au 1 et on applique en plus la formule des probabilités totales au dénominateur. □

4 Indépendance

Définition 12. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé. Des événements* A *et* B *sont dits* indépendants *si*

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

Proposition 10. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et deux événements* A *et* B. *Si* A *et* B *sont indépendants, alors* A *et* $\bar{\mathrm{B}}$ *le sont aussi.*

Démonstration. On a

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

□

Proposition 11. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé. Les événements presque sûrs et négligeables sont indépendants de tout autre événement.*

Démonstration. Soit $\mathrm{A} \in \mathscr{A}$ et B un événement négligeable. L'événement $\mathrm{A}\cap\mathrm{B}$ est négligeable car contenu dans B et

$$\mathbb{P}(\mathrm{A}\cap\mathrm{B}) = 0 = \mathbb{P}(\mathrm{A})\mathbb{P}(\mathrm{B})$$

Puis, soit B un événement presque sûr. On a $\bar{\mathrm{B}}$ négligeable donc indépendant de A d'où B indépendant de A d'après le résultat de la proposition précédente. □

Remarque : En particulier, les événements $\varnothing$ et Ω sont indépendants de tout autre événement.

Proposition 12. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et deux événements* A *et* B *avec* $\mathbb{P}(\mathrm{B}) > 0$. *On a*

$$\mathrm{A}, \mathrm{B} \textit{ indépendants} \iff \mathbb{P}(\mathrm{A}|\mathrm{B}) = \mathbb{P}(\mathrm{A})$$

Démonstration. On a

$$\mathrm{A},\mathrm{B}\text{ indépendants} \iff \mathbb{P}(\mathrm{A}\cap\mathrm{B}) = \mathbb{P}(\mathrm{A})\times\mathbb{P}(\mathrm{B}) \iff \frac{\mathbb{P}(\mathrm{A}\cap\mathrm{B})}{\mathbb{P}(\mathrm{B})} = \mathbb{P}(\mathrm{A})$$

□

Définition 13. *Soit* $(\Omega,\mathscr{A},\mathbb{P})$ *un espace probabilisé. La famille* $(\mathrm{A}_n)_{n\in\mathbb{N}}$ *est constituée d'événements dits* indépendants *si*

$$\forall \mathrm{I} \textit{ fini } \subset \mathbb{N} \qquad \mathbb{P}\left(\bigcap_{i\in\mathrm{I}} \mathrm{A}_i\right) = \prod_{i\in\mathrm{I}} \mathbb{P}(\mathrm{A}_i)$$

Remarque : Cette définition couvre le cas d'une famille finie d'événements en prenant $\mathrm{A}_n = \Omega$ pour n supérieur à un certain rang.

Proposition 13. *Soit* $(\Omega,\mathscr{A},\mathbb{P})$ *un espace probabilisé et* $(\mathrm{A}_n)_{n\in\mathbb{N}}$ *une famille d'événements indépendants. Alors les événements* A_n *sont deux à deux indépendants.*

Démonstration. Il suffit de considérer $\mathrm{I} = \{i,j\}$ avec $(i,j)\in\mathbb{N}^2$ et $i\neq j$ pour établir l'indépendance de A_i et A_j. □

Remarque importante : La réciproque est fausse : des événements peuvent être deux à deux indépendants sans être indépendants.

On lance deux fois de suite une pièce. On note P_i l'événement pile au i-ème lancer et F_i l'événement face au i-ème lancer. Les événements $\mathrm{P}_1, \mathrm{P}_2$ et $\mathrm{A} = \mathrm{P}_1\mathrm{P}_2 \cup \mathrm{F}_1\mathrm{F}_2$ sont deux à deux indépendants mais non mutuellement indépendants (pour alléger, on note $\mathrm{P}_1\mathrm{P}_2$ au lieu de $\mathrm{P}_1\cap\mathrm{P}_2$ et de même avec les F_i).

Proposition 14. *Soit* $(\Omega,\mathscr{A},\mathbb{P})$ *un espace probabilisé et* $(\mathrm{A}_n)_{n\in\mathbb{N}}$ *une famille d'événements indépendants. Considérons la famille d'événements* $(\mathrm{B}_n)_{n\in\mathbb{N}}$ *telle que* $\mathrm{B}_n = \mathrm{A}_n$ *ou* $\overline{\mathrm{A}_n}$ *pour tout* $n\in\mathbb{N}$*, alors les* B_n *sont indépendants.*

Démonstration. Pour p entier, on note

$$\mathscr{P}(p) :\ \forall (\mathrm{B}_n)_n \in \left\{\mathrm{A}_n, \overline{\mathrm{A}_n}\right\}^{\mathbb{N}} \quad \forall \mathrm{I}_{\text{ fini}} \subset \mathbb{N} \mid \operatorname{Card}\left\{i\in\mathrm{I} \mid \mathrm{B}_i = \overline{\mathrm{A}_i}\right\} = p \quad \mathbb{P}\left(\bigcap_{i\in\mathrm{I}} \mathrm{B}_i\right) = \prod_{i\in\mathrm{I}} \mathbb{P}(\mathrm{B}_i)$$

L'initialisation $\mathscr{P}(0)$ est vraie puisque pour toute suite $(\mathrm{B}_n)_n$ avec $\mathrm{B}_n = \mathrm{A}_n$ ou $\overline{\mathrm{A}_n}$ pour tout n entier, pour $\mathrm{I}\subset\mathbb{N}$ tel que $\operatorname{Card}\left\{i\in\mathrm{I} \mid \mathrm{B}_i = \overline{\mathrm{A}_i}\right\} = 0$, on a $\mathrm{A}_i = \mathrm{B}_i$ pour tout $i\in\mathrm{I}$ d'où

$$\mathbb{P}\left(\bigcap_{i\in\mathrm{I}} \mathrm{B}_i\right) = \mathbb{P}\left(\bigcap_{i\in\mathrm{I}} \mathrm{A}_i\right) = \prod_{i\in\mathrm{I}} \mathbb{P}(\mathrm{A}_i) = \prod_{i\in\mathrm{I}} \mathbb{P}(\mathrm{B}_i)$$

On suppose le résultat vrai au rang p entier fixé. Soit $\mathrm{I}\subset\mathbb{N}$ tel que $\operatorname{Card}\left\{i\in\mathrm{I} \mid \mathrm{B}_i = \overline{\mathrm{A}_i}\right\} = p+1$. Soit $k\in\mathrm{I}$ tel que $\mathrm{B}_k = \overline{\mathrm{A}_k}$. On a

$$\mathbb{P}\left(\bigcap_{i\in\mathrm{I}} \mathrm{B}_i\right) = \mathbb{P}\left(\overline{\mathrm{A}_k} \cap \bigcap_{i\in\mathrm{I}\smallsetminus\{k\}} \mathrm{B}_i\right) = \mathbb{P}\left(\bigcap_{i\in\mathrm{I}\smallsetminus\{k\}} \mathrm{B}_i\right) - \mathbb{P}\left(\mathrm{A}_k \cap \bigcap_{i\in\mathrm{I}\smallsetminus\{k\}} \mathrm{B}_i\right)$$

Par hypothèse de récurrence avec la famille $(\mathrm{B}_n)_n$ et la partie $\mathrm{I}\smallsetminus\{k\}$, on a

$$\mathbb{P}\left(\bigcap_{i\in \mathrm{I}\smallsetminus\{k\}} \mathrm{B}_i\right) = \prod_{i\in \mathrm{I}\smallsetminus\{k\}} \mathbb{P}(\mathrm{B}_i)$$

On pose $\mathrm{C}_i = \mathrm{B}_i$ pour tout $i \in \mathbb{N} \smallsetminus \{k\}$ et $\mathrm{C}_k = \mathrm{A}_k$. On a

$$\mathrm{Card}\left\{i \in \mathrm{I} \mid \mathrm{C}_i = \overline{\mathrm{A}_i}\right\} = \mathrm{Card}\left\{i \in \mathrm{I} \smallsetminus \{k\} \mid \mathrm{B}_i = \overline{\mathrm{A}_i}\right\} = p$$

Ainsi, par hypothèse de récurrence appliquée avec la famille $(\mathrm{C}_n)_n$ et la partie I, il vient

$$\mathbb{P}\left(\mathrm{A}_k \cap \bigcap_{i\in \mathrm{I}\smallsetminus\{k\}} \mathrm{B}_i\right) = \mathbb{P}\left(\bigcap_{i\in\mathrm{I}} \mathrm{C}_i\right) = \prod_{i\in\mathrm{I}} \mathbb{P}(\mathrm{C}_i) = \mathbb{P}(\mathrm{A}_k) \prod_{i\in \mathrm{I}\smallsetminus\{k\}} \mathbb{P}(\mathrm{B}_i)$$

Il vient ensuite

$$\mathbb{P}\left(\bigcap_{i\in\mathrm{I}} \mathrm{B}_i\right) = \prod_{i\in \mathrm{I}\smallsetminus\{k\}} \mathbb{P}(\mathrm{B}_i) - \mathbb{P}(\mathrm{A}_k) \prod_{i\in \mathrm{I}\smallsetminus\{k\}} \mathbb{P}(\mathrm{B}_i) = (1-\mathbb{P}(\mathrm{A}_k)) \prod_{i\in \mathrm{I}\smallsetminus\{k\}} \mathbb{P}(\mathrm{B}_i) = \prod_{i\in\mathrm{I}} \mathbb{P}(\mathrm{B}_i)$$

ce qui clôt la récurrence. □

II Variables aléatoires discrètes

1 Définitions

Pour une application $\mathrm{X} : \Omega \to \mathrm{E}$, l'ensemble $\mathrm{X}(\Omega)$ désigne l'ensemble image par X avec

$$\mathrm{X}(\Omega) = \{\mathrm{X}(\omega),\ \omega \in \Omega\}$$

On appelle *aléa* un élément $\omega \in \Omega$.

Définition 14. *Soit $(\Omega, \mathscr{A})$ un espace probabilisable. On appelle* variable aléatoire discrète *sur $(\Omega, \mathscr{A})$ une application X définie sur Ω à valeurs dans un ensemble E telle que $\mathrm{X}(\Omega)$ est fini ou dénombrable et*

$$\forall x \in \mathrm{X}(\Omega) \qquad \mathrm{X}^{-1}(\{x\}) = \{\mathrm{X} = x\} = \{\omega \in \Omega \mid \mathrm{X}(\omega) = x\} \in \mathscr{A}$$

Si $\mathrm{E} \subset \mathbb{R}$, on parle de variable aléatoire réelle discrète.

Remarques : (1) Cette définition généralise le cas de Ω fini puisque la condition $\mathrm{X}^{-1}(\{x\}) \in \mathscr{A}$ est nécessairement réalisée si $\mathscr{A} = \mathcal{P}(\Omega)$.
(2) Pour $x \in \mathrm{E} \smallsetminus \mathrm{X}(\Omega)$, on a aussi $\{\mathrm{X} = x\} \in \mathscr{A}$ puisque $\{\mathrm{X} = x\} = \varnothing$.

Exemple : Soit $\mathrm{A} \in \mathscr{A}$. L'application $\mathbb{1}_{\mathrm{A}}$ est une variable aléatoire discrète : $\mathbb{1}_{\mathrm{A}}(\Omega) \subset \{0, 1\}$ avec $\mathbb{1}_{\mathrm{A}}^{-1}(\{1\}) = \mathrm{A}$ et $\mathbb{1}_{\mathrm{A}}^{-1}(\{0\}) = \bar{\mathrm{A}}$.

Proposition 15. *Soit $(\Omega, \mathscr{A})$ un espace probabilisable et $\mathrm{X} : \Omega \to \mathrm{E}$ une variable aléatoire discrète. On a*

$$\forall \mathrm{A} \subset \mathrm{E} \qquad \mathrm{X}^{-1}(\mathrm{A}) = \{\omega \in \Omega \mid \mathrm{X}(\omega) \in \mathrm{A}\} \in \mathscr{A}$$

Démonstration. On a

$$\mathrm{X}^{-1}(\mathrm{A}) = \bigsqcup_{x\in \mathrm{A}\cap\mathrm{X}(\Omega)} \{\mathrm{X} = x\}$$

avec $\mathrm{A} \cap \mathrm{X}(\Omega)$ au plus dénombrable. Par stabilité par union dénombrable, le résultat suit. □

Notations : Soit $X : \Omega \to E$ variable aléatoire discrète. Pour $A \subset E$, on note

$$X^{-1}(A) = \{X \in A\} \quad \text{ou} \quad (X \in A)$$

Si X est une variable aléatoire réelle discrète, on note pour x réel

$$X^{-1}(]-\infty\,;x]) = \{X \leqslant x\} \quad \text{ou} \quad (X \leqslant x), \qquad X^{-1}(]-\infty\,;x[) = \{X < x\} \quad \text{ou} \quad (X < x) \quad \textit{etc.}$$

Dans $(\Omega, \mathscr{A}, \mathbb{P})$ espace probabilisé, on note (abusivement) $\mathbb{P}(X \in A)$ au lieu de $\mathbb{P}(\{X \in A\})$ et de même avec $\mathbb{P}(X \leqslant x)$ au lieu de $\mathbb{P}(\{X \leqslant x\})$ pour x réel si X à valeurs réelles (omission d'accolades délibérée).

Exemple : On lancé un dé indéfiniment. On note T le rang de première obtention de 6. On admet que le résultat X_k du k-ième lancer est une variable aléatoire discrète. Alors, la fonction T est une variable aléatoire discrète. En effet, on a T à valeurs $\mathbb{N}^* \cup \{+\infty\}$ puis

$$\forall n \in \mathbb{N}^* \qquad \{T = n\} = \{X_n = 6\} \cap \bigcap_{k=1}^{n-1} \{X_k < 6\} \quad \text{et} \quad \{T = +\infty\} = \bigcap_{k=1}^{+\infty} \{X_k < 6\}$$

Soit $(\Omega, \mathscr{A}, \mathbb{P})$ espace probabilisé et X une variable aléatoire discrète. On note $\operatorname{supp} X$ (notation non officielle) son *support* défini par

$$\operatorname{supp} X = \{x \in X(\Omega) \mid \mathbb{P}(X = x) > 0\}$$

Proposition 16. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire discrète. L'événement $\{X \in \operatorname{supp} X\}$ est presque sûr.*

Démonstration. L'ensemble $X(\Omega) \setminus \operatorname{supp} X$ est au plus dénombrable car inclus dans $X(\Omega)$. L'événement $\{X \notin \operatorname{supp} X\} = \bigsqcup_{x \in X(\Omega) \setminus \operatorname{supp} X} \{X = x\}$ est négligeable comme union au plus dénombrable d'événements négligeables d'où le résultat sur $\{X \in \operatorname{supp} X\}$ par complémentation. □

Proposition 17. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire discrète à valeurs dans E. Pour $D \subset E$ avec D au plus dénombrable contenant $\operatorname{supp} X$ (comme $D = X(\Omega)$ par exemple), la famille $(\mathbb{P}(X = x))_{x \in D}$ est une distribution de probabilité discrète dont le support est $\operatorname{supp} X$.*

Démonstration. C'est une famille à valeurs dans $\mathbb{R}_+$ et par σ-additivité, l'ensemble D étant au plus dénombrable

$$\begin{aligned}\sum_{x \in D} \mathbb{P}(X = x) = \mathbb{P}\left(X \in \bigsqcup_{x \in D} \{X = x\}\right) &= \mathbb{P}(X \in D) \\ &= \mathbb{P}(X \in D, X \in \operatorname{supp} X) = \mathbb{P}(X \in \operatorname{supp} X) = 1\end{aligned}$$

Son support est clairement $\operatorname{supp} X$. □

Proposition 18. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire discrète. L'application notée $\mathbb{P}_X$ définie sur $\mathcal{P}(\operatorname{supp} X)$ par $\mathbb{P}_X : A \mapsto \mathbb{P}(X \in A)$ est une probabilité sur $(\operatorname{supp} X, \mathcal{P}(\operatorname{supp} X))$.*

Démonstration. La famille $(\mathbb{P}(X = x))_{x \in \operatorname{supp} X}$ est une distribution de probabilité discrète. Pour $A \subset \operatorname{supp} X$ qui est donc au plus dénombrable, on a par σ-additivité

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \sum_{x \in A} \mathbb{P}(X = x)$$

D'après la proposition 8, l'application $\mathbb{P}_X$ est une probabilité sur $(\operatorname{supp} X, \mathcal{P}(\operatorname{supp} X))$. □

Définition 15. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* X *une variable aléatoire discrète. On appelle* loi de la variable aléatoire X *la probabilité* $\mathbb{P}_X$ *sur* $(\operatorname{supp} X, \mathcal{P}(\operatorname{supp} X))$.

Notation : Si $\mathscr{L}$ désigne une loi usuelle et que X suit cette loi $\mathscr{L}$, on note $X \sim \mathscr{L}$.

Exemple : Pour $A \in \mathscr{A}$, on a $\mathbb{1}_A \sim \mathscr{B}(\mathbb{P}(A))$.

Proposition 19. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé,* X *une variable aléatoire discrète à valeurs dans* E *et* D *partie de* E *au plus dénombrable contenant* $\operatorname{supp} X$. *La famille* $(\mathbb{P}(X = x))_{x \in D}$ *caractérise la loi de* X.

Démonstration. La loi de X est la probabilité construite à partir de la distribution de probabilité discrète $(\mathbb{P}(X = x))_{x \in \operatorname{supp} X}$ qui est une sous-famille de $(\mathbb{P}(X = x))_{x \in D}$ de support $\operatorname{supp} X$. Le résultat suit. $\square$

Remarque importante : Si l'ensemble D est au plus dénombrable et contient $X(\Omega)$, il convient pour caractériser la loi. En pratique, on a plus souvent accès à un tel ensemble D qu'au support de X ou $X(\Omega)$.

Définition 16. *Soient* $(\Omega_1, \mathscr{A}_1, \mathbb{P}_1)$, $(\Omega_2, \mathscr{A}_2, \mathbb{P}_2)$ *des espaces probabilisés et* X, Y *des variables aléatoires discrètes respectivement sur l'un et l'autre de ces espaces probabilisés et vérifiant* $\operatorname{supp} X = \operatorname{supp} Y$. *On dit que* X *et* Y *suivent la même loi si* $\mathbb{P}_{1,X} = \mathbb{P}_{2,Y}$ *et on note* $X \sim Y$.

Exemple : Deux personnes lancent chacune un dé équilibré à 6 faces. On note X et Y les résultats pour chaque dé. Alors, on a $X \sim Y \sim \mathscr{U}_{[\![1;6]\!]}$. L'égalité en loi n'est pas l'égalité ! Il n'y aucune raison que les deux dés fournissent le même résultat.

Proposition 20. *Soient* $(\Omega_1, \mathscr{A}_1, \mathbb{P}_1)$, $(\Omega_2, \mathscr{A}_2, \mathbb{P}_2)$ *des espaces probabilisés et* X, Y *des variables aléatoires discrètes respectivement sur l'un et l'autre de ces espaces probabilisés. S'il existe* D *au plus dénombrable tel que* $\operatorname{supp} X \cup \operatorname{supp} Y \subset D$ *et*

$$\forall x \in D \qquad \mathbb{P}_1(X = x) = \mathbb{P}_2(Y = x)$$

alors on a $X \sim Y$.

Démonstration. Conséquence immédiate de la proposition 19. $\square$

Remarque : Si D au plus dénombrable contient $X(\Omega_1) \cup Y(\Omega_2)$, on a le résultat. En particulier si $X(\Omega_1) = Y(\Omega_2)$ et $D = X(\Omega_1)$ ou $D \supset X(\Omega_1)$, le résultat vaut.

Proposition 21. *Soient* $(\Omega_1, \mathscr{A}_1, \mathbb{P}_1)$, $(\Omega_2, \mathscr{A}_2, \mathbb{P}_2)$ *des espaces probabilisés et* X, Y *des variables aléatoires discrètes respectivement sur l'un et l'autre de ces espaces probabilisés et à valeurs dans* E. *Si on a* $X \sim Y$, *alors*

$$\forall A \subset E \qquad \mathbb{P}_1(X \in A) = \mathbb{P}_2(Y \in A)$$

Démonstration. On a

$$\forall A \subset E \qquad \mathbb{P}_1(X \in A) = \mathbb{P}_1(X \in A \cap \operatorname{supp} X) = \sum_{x \in A \cap \operatorname{supp} X} \mathbb{P}_1(X = x) = \ldots = \mathbb{P}_2(Y \in A)$$

$\square$

Proposition 22. *Soit* $(\Omega, \mathscr{A})$ *un espace probabilisable,* X *une variable aléatoire discrète à valeurs dans* E *et* $f : E \to F$ *une fonction. Alors* $f(X)$ *est une variable aléatoire discrète sur* $(\Omega, \mathscr{A})$.

Remarque : La notation $f(\mathrm{X})$ est abusive. En toute rigueur, on devrait noter $f \circ \mathrm{X}$.

Démonstration. On a $f(\mathrm{X}(\Omega)) = \{f(x), x \in \mathrm{X}(\Omega)\}$ au plus dénombrable. Puis, soit $y \in f(\mathrm{X}(\Omega))$. Il vient

$$(f \circ \mathrm{X})^{-1}(\{y\}) = \{\omega \in \Omega \mid f \circ \mathrm{X}(\omega) \in \{y\}\} = \{\omega \in \Omega \mid \mathrm{X}(\omega) \in f^{-1}(\{y\})\} = \mathrm{X}^{-1}(f^{-1}(\{y\}))$$

Comme $f^{-1}(\{y\}) \subset \mathrm{E}$, on conclut grace à la proposition 15. □

Proposition 23. *Soient $(\Omega_1, \mathscr{A}_1, \mathbb{P}_1)$, $(\Omega_2, \mathscr{A}_2, \mathbb{P}_2)$ des espaces probabilisés et X, Y des variables aléatoires discrètes respectivement sur l'un et l'autre de ces espaces probabilisés à valeurs dans E et $f : \mathrm{E} \to \mathrm{F}$. Si on a $\mathrm{X} \sim \mathrm{Y}$, il s'ensuit $f(\mathrm{X}) \sim f(\mathrm{Y})$.*

Démonstration. Soit $a \in \operatorname{supp} f(\mathrm{X}) \cup \operatorname{supp} f(\mathrm{Y})$. On a

$$\mathbb{P}_1(f(\mathrm{X}) = a) = \mathbb{P}_1\left(\mathrm{X} \in f^{-1}(\{a\})\right) = \mathbb{P}_2\left(\mathrm{Y} \in f^{-1}(\{a\})\right) = \mathbb{P}_2(f(\mathrm{Y}) = a)$$

□

Définition 17. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé, X une variable aléatoire discrète et B un événement vérifiant $\mathbb{P}(\mathrm{B}) > 0$. On appelle* loi conditionnelle *de X sachant B la probabilité sur $(\operatorname{supp} \mathrm{X}, \mathcal{P}(\operatorname{supp} \mathrm{X}))$ notée $\mathbb{P}_{\mathrm{X}|\mathrm{B}}$ définie par*

$$\forall \mathrm{A} \in \mathcal{P}(\operatorname{supp} \mathrm{X}) \qquad \mathbb{P}_{\mathrm{X}|\mathrm{B}}(\mathrm{A}) = \mathbb{P}(\mathrm{X} \in \mathrm{A} | \mathrm{B})$$

Remarque : Cette définition est valide d'après la proposition 18 et le théorème 2 ($\mathbb{P}_{\mathrm{B}}$ est une probabilité et on regarde la loi de X pour cette probabilité).

Proposition 24. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé, $\mathrm{X} : \Omega \to \mathrm{E}$ une variable aléatoire discrète et B un événement vérifiant $\mathbb{P}(\mathrm{B}) > 0$. Pour $\mathrm{D} \subset \mathrm{E}$ avec D au plus dénombrable contenant $\operatorname{supp} \mathrm{X}$, la famille $(\mathbb{P}(\mathrm{X} = x | \mathrm{B}))_{x \in \mathrm{D}}$ caractérise la loi de X sachant B.*

Démonstration. On applique la proposition 19 à $\mathbb{P}_{\mathrm{X}|\mathrm{B}}$. □

Définition 18. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire réelle discrète. On définit la* fonction de répartition *de X notée F_{X} sur $\mathbb{R}$ par*

$$\forall x \in \mathbb{R} \qquad \mathrm{F}_{\mathrm{X}}(x) = \mathbb{P}(\mathrm{X} \leqslant x)$$

Proposition 25 (À savoir refaire). *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé, X une variable aléatoire réelle discrète de fonction de répartition F_{X}. On a*

1. *F_{X} est croissante ;*
2. *$\mathrm{F}_{\mathrm{X}}(x) \xrightarrow[x \to +\infty]{} 1$ et $\mathrm{F}_{\mathrm{X}}(x) \xrightarrow[x \to -\infty]{} 0$.*

Démonstration. 1. Soit $x \leqslant y$. On a $\{\mathrm{X} \leqslant x\} \subset \{\mathrm{X} \leqslant y\}$ d'où le résultat par croissance de $\mathbb{P}$.
2. La fonction F_{X} est croissante bornée donc admet une limite finie en $+\infty$ et $-\infty$ par limite monotone. Ainsi

$$\lim_{x \to +\infty} \mathrm{F}_{\mathrm{X}}(x) = \lim_{n \to +\infty} \mathrm{F}_{\mathrm{X}}(n) \quad \text{et} \quad \lim_{x \to -\infty} \mathrm{F}_{\mathrm{X}}(x) = \lim_{n \to +\infty} \mathrm{F}_{\mathrm{X}}(-n)$$

La famille $(\{\mathrm{X} \leqslant n\})_{n \in \mathbb{N}}$ est croissante pour l'inclusion d'où, par continuité croissante,

$$\mathrm{F}_{\mathrm{X}}(n) = \mathbb{P}(\mathrm{X} \leqslant n) \xrightarrow[n \to \infty]{} \mathbb{P}\left(\bigcup_{n=0}^{+\infty} \{\mathrm{X} \leqslant n\}\right) = \mathbb{P}(\mathrm{X} \in \mathbb{R}) = 1$$

La famille $(\{X \leqslant -n\})_{n\in\mathbb{N}}$ est décroissante pour l'inclusion d'où, par continuité décroissante,

$$F_X(-n) = \mathbb{P}(X \leqslant -n) \xrightarrow[n\to\infty]{} \mathbb{P}\left(\bigcap_{n=0}^{+\infty} \{X \leqslant -n\}\right) = \mathbb{P}(\varnothing) = 0$$

□

Illustration : Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé, $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes de même loi $\mathscr{B}(x)$ avec $x \in]\,0\,;1\,[$ et $S_n = \sum_{i=1}^{n} X_i$ pour n entier. On admet l'égalité $\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}\, dt = \sqrt{2\pi}$ et on rappelle le résultat du *théorème de Moivre-Laplace* :

$$\forall \alpha \in \mathbb{R} \qquad \mathbb{P}\left(\frac{S_n - nx}{\sqrt{nx(1-x)}} \leqslant \alpha\right) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-\frac{t^2}{2}}\, dt$$

On peut donc observer la convergence annoncée par des tracés de fonctions de répartition.

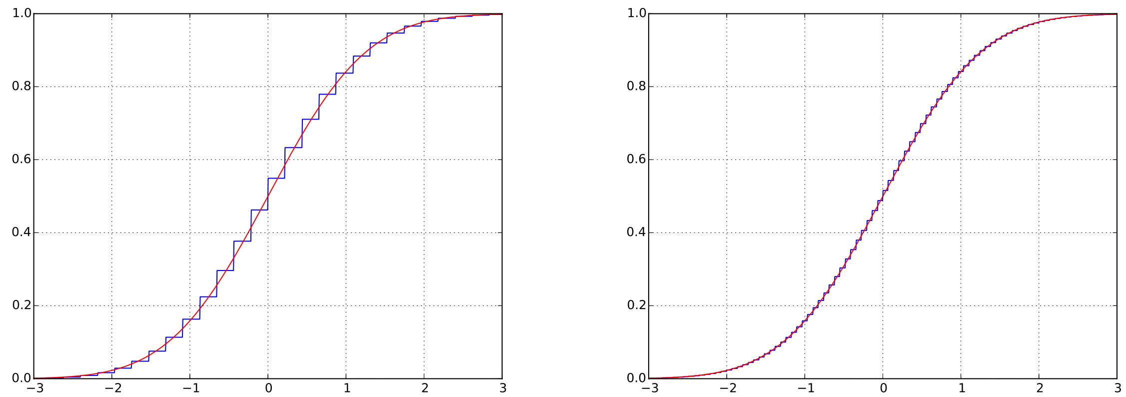

FIGURE 2 – Théorème de Moivre-Laplance, fonctions de répartition pour $n = 100$ et $n = 1000$

La fonction $\alpha \mapsto \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-\frac{t^2}{2}}\, dt$ est la fonction de répartition de la loi normale centrée réduite.

2 Couples de variables aléatoires

Définition 19. *Soit $(\Omega, \mathscr{A})$ un espace probabilisable. On appelle* couple de variables aléatoires discrètes *un couple (X, Y) avec X et Y des variables aléatoires discrètes sur $(\Omega, \mathscr{A})$.*

Proposition 26. *Soit $(\Omega, \mathscr{A})$ un espace probabilisable et (X, Y) un couple de variables aléatoires discrètes. Le couple (X, Y) est une variable aléatoire discrète sur $(\Omega, \mathscr{A})$.*

Démonstration. On a $(X, Y)(\Omega) \subset X(\Omega) \times Y(\Omega)$ au plus dénombrable comme produit fini d'ensembles au plus dénombrables. Puis, pour $(x, y) \in (X, Y)(\Omega)$, on a

$$(X, Y)^{-1}(\{(x, y)\}) = X^{-1}(\{x\}) \cap Y^{-1}(\{y\}) \in \mathscr{A}$$

d'où le résultat. □

Proposition 27. *Soit $(\Omega, \mathscr{A})$ un espace probabilisable. L'ensemble des variables aléatoires discrètes sur $(\Omega, \mathscr{A})$ à valeurs dans $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$ est un $\mathbb{K}$-ev.*

Démonstration. La fonction nulle $\Omega \to \mathbb{K}, \omega \mapsto 0$ est une variable aléatoire scalaire discrète. Soient X, Y des variables aléatoires scalaires discrètes et λ scalaire. Posant $f : \mathbb{K}^2 \to \mathbb{K}, (x, y) \mapsto x + \lambda y$, on a $\mathrm{X} + \lambda \mathrm{Y} = f(\mathrm{X}, \mathrm{Y})$ qui est une variable aléatoire discrète en tant que fonction d'une variable aléatoire discrète. On en déduit que l'ensemble des variables aléatoires réelles est un $\mathbb{K}$-ev en tant que sev de $\mathbb{K}^\Omega$. □

Définition 20. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* (X, Y) *un couple de variables aléatoires discrètes. On appelle* loi conjointe *de* X *et* Y *la loi du couple* (X, Y) *et* lois marginales *du couple* (X, Y) *les lois de* X *et* Y.

Remarque : En général, les lois marginales ne déterminent pas la loi conjointe.

Par exemple, pour $\mathrm{X} \sim \mathscr{B}(1/2)$, les lois marginales des couples (X, X) et $(\mathrm{X}, 1 - \mathrm{X})$ sont identiques mais les couples n'ont pas même loi puisque, par exemple, on a

$$\mathbb{P}((\mathrm{X}, \mathrm{X}) = (0, 1)) = 0 \neq \mathbb{P}((\mathrm{X}, 1 - \mathrm{X}) = (0, 1)) = \mathbb{P}(\mathrm{X} = 0) = \frac{1}{2}$$

Remarque : Étant donné un couple de variables aléatoires discrètes (X, Y), on peut définir la *loi conditionnelle* de X sachant $\{\mathrm{Y} = y\}$ avec $y \in \operatorname{supp} \mathrm{Y}$ selon la définition 17.

Définition 21. *Soit* $(\Omega, \mathscr{A})$ *un espace probabilisable. On appelle* vecteur aléatoire discret *un* n*-uplet* $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ *avec les* X_i *des variables aléatoires discrètes sur* $(\Omega, \mathscr{A})$.

Proposition 28. *Soit* $(\Omega, \mathscr{A})$ *un espace probabilisable et* $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ *un vecteur aléatoire discret. Il s'agit d'une variable aléatoire discrète sur* $(\Omega, \mathscr{A})$.

Démonstration. Identique à celle de la proposition 26. □

Exemple : Soit $(\Omega, \mathscr{A})$ un espace probabilisable, $(\mathrm{X}_n)_{n \geqslant 1}$ une suite infinie de variables aléatoires discrètes avec $\mathrm{X}_n(\Omega) = \{0, 1\}$ pour tout $n \geqslant 1$ et N une variable aléatoire discrète avec $\mathrm{N}(\Omega) = \mathbb{N}$. On pose

$$\forall \omega \in \Omega \qquad \mathrm{T}(\omega) = \sum_{i=1}^{\mathrm{N}(\omega)} \mathrm{X}_i(\omega)$$

Alors, la fonction T est une variable aléatoire discrète. En effet, l'application T est à valeurs dans $\mathbb{N}$ et

$$\forall k \in \mathbb{N} \qquad \{\mathrm{T} = k\} = \bigsqcup_{n=0}^{+\infty} \{\mathrm{N} = n\} \cap \left\{\sum_{i=1}^{n} \mathrm{X}_i = k\right\}$$

et $\sum_{i=1}^{n} \mathrm{X}_i$ est une variable aléatoire comme fonction du vecteur aléatoire $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$.

3 Variables aléatoires indépendantes

Définition 22. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé. Des variables aléatoires discrètes* X *et* Y *sont dites* indépendantes *si*

$$\forall (x, y) \in \mathrm{X}(\Omega) \times \mathrm{Y}(\Omega) \qquad \mathbb{P}(\mathrm{X} = x, \mathrm{Y} = y) = \mathbb{P}(\mathrm{X} = x)\mathbb{P}(\mathrm{Y} = y)$$

Notations : On note $X \perp\!\!\!\perp Y$.
L'événement $\{X = x, Y = y\}$ désigne l'événement $\{X = x\} \cap \{Y = y\}$.

Remarque : L'indépendance de X et Y équivaut à l'égalité entre distributions de probabilité discrète suivantes :

$$(\mathbb{P}(X = x, Y = y))_{(x,y)\in(X,Y)(\Omega)} = (\mathbb{P}(X = x)\mathbb{P}(Y = y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$$

Proposition 29. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé. Les variables aléatoires discrètes X et Y à valeurs respectives dans E et F sont indépendantes si et seulement si*

$$\forall (A, B) \in \mathcal{P}(E) \times \mathcal{P}(F) \qquad \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

Démonstration. Le sens indirect est immédiat. Soit $(A, B) \in \mathcal{P}(E) \times \mathcal{P}(F)$. On a les égalités

$$\{X \in A\} = \{X \in A \cap X(\Omega)\} \quad \text{et} \quad \{Y \in B\} = \{Y \in B \cap Y(\Omega)\}$$

On peut donc considérer $A \subset X(\Omega)$ et $Y \subset Y(\Omega)$ sans perte de généralité et on a alors A et B au plus dénombrables. L'ensemble $A \times B$ est au plus dénombrable comme produit d'ensembles au plus dénombrables. D'après le théorème de Fubini pour des familles à termes positifs, il vient

$$\begin{aligned}\mathbb{P}(X \in A, Y \in B) &= \sum_{(x,y)\in A\times B} \mathbb{P}(X = x, Y = y)\\ &= \sum_{(x,y)\in A\times B} \mathbb{P}(X = x)\mathbb{P}(Y = y) = \sum_{x\in A}\mathbb{P}(X = x)\sum_{y\in B}\mathbb{P}(Y = y)\end{aligned}$$

□

Définition 23. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé. Des variables aléatoires discrètes $X_1, \ldots, X_n$ sont dites* indépendantes

$$\forall (x_i)_{i\in[\![1\,;\,n]\!]} \in \prod_{i=1}^{n} X_i(\Omega) \qquad \mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i = x_i\}\right) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i)$$

Proposition 30. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé. Les variables aléatoires discrètes $X_1, \ldots, X_n$ avec $X_i : \Omega \to E_i$ sont indépendantes si et seulement si*

$$\forall (A_i)_{i\in[\![1\,;\,n]\!]} \in \prod_{i=1}^{n} \mathcal{P}(E_i) \qquad \mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \in A_i\}\right) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i)$$

ou de manière équivalente

$$\forall I \subset [\![1\,;\,n]\!] \qquad \forall (A_i)_{i\in I} \in \prod_{i\in I} \mathcal{P}(E_i) \qquad \mathbb{P}\left(\bigcap_{i\in I} \{X_i \in A_i\}\right) = \prod_{i\in I} \mathbb{P}(X_i \in A_i)$$

Autrement dit, les événements $\{X_i \in A_i\}_{i\in[\![1\,;\,n]\!]}$ sont indépendants.

Lemme 1. *Soient $X_1, \ldots, X_n$ indépendantes, alors $X_1, \ldots, X_{n-1}$ le sont aussi.*

Il suffit en effet de considérer le système complet $\{X_n = x_n\}_{x_n\in X_n(\Omega)}$.

Démonstration. Le sens indirect est immédiat. Pour le sens direct, on procède par récurrence. Les variables $(X_1, \ldots, X_{n-1})$ et X_n indépendantes. En effet, pour $(x_1, \ldots, x_{n-1}) \in (X_1, \ldots, X_n)(\Omega)$

et $x_n \in \mathrm{X}_n(\Omega)$, on a

$$\mathbb{P}\left((\mathrm{X}_1,\ldots,\mathrm{X}_{n-1})=(x_1,\ldots,x_{n-1}),\mathrm{X}_n=x_n\right)=$$
$$\mathbb{P}\left(\bigcap_{i=1}^{n}\{\mathrm{X}_i=x_i\}\right)=\prod_{i=1}^{n}\mathbb{P}(\mathrm{X}_i=x_i)=\mathbb{P}\left((\mathrm{X}_1,\ldots,\mathrm{X}_{n-1})=(x_1,\ldots,x_{n-1})\right)\mathbb{P}(\mathrm{X}_n=x_n)$$

Soit $(\mathrm{A}_i)_{i\in[\![1\,;\,n]\!]} \in \prod_{i=1}^{n}\mathcal{P}(\mathrm{E}_i)$. D'après la proposition 29 (on considère $(\mathrm{X}_1,\ldots,\mathrm{X}_{n-1})$ à valeurs dans $\prod_{i=1}^{n-1}\mathrm{E}_i$ et on a bien $\prod_{i=1}^{n-1}\mathrm{A}_i \in \mathcal{P}(\prod_{i=1}^{n-1}\mathrm{E}_i)$), il vient

$$\begin{aligned}\mathbb{P}\left(\bigcap_{i=1}^{n}\{\mathrm{X}_i\in\mathrm{A}_i\}\right)&=\mathbb{P}\left((\mathrm{X}_1,\ldots,\mathrm{X}_{n-1})\in\prod_{i=1}^{n-1}\mathrm{A}_i,\mathrm{X}_n\in\mathrm{A}_n\right)\\&=\mathbb{P}\left((\mathrm{X}_1,\ldots,\mathrm{X}_{n-1})\in\prod_{i=1}^{n-1}\mathrm{A}_i\right)\mathbb{P}(\mathrm{X}_n\in\mathrm{A}_n)\end{aligned}$$

et le résultat suit par hypothèse de récurrence. Pour la dernière équivalence, le sens indirect est immédiat en prenant $\mathrm{I}=[\![1\,;\,n]\!]$ et pour le sens direct, on choisit $\mathrm{A}_j=\mathrm{E}_j$ pour $j\notin\mathrm{I}$. □

Remarque : La deuxième caractérisation garantit que toute sous-famille $(\mathrm{X}_i)_{i\in\mathrm{I}}$ est constituée de variables indépendantes.

Proposition 31. *Soit $(\Omega,\mathscr{A},\mathbb{P})$ un espace probabilisé. Si $\mathrm{X}_1,\ldots,\mathrm{X}_n$ sont des variables aléatoires discrètes indépendantes, alors elles sont deux à deux indépendantes.*

Démonstration. Il suffit de choisir $\mathrm{I}=\{i,j\}$ avec $i\neq j$ dans la dernière caractérisation de la proposition précédente. □

Remarque : La réciproque est fausse.
On peut reprendre le contre-exemple fourni dans la remarque faisant suite à la proposition 13 et poser $\mathrm{X}_1=\mathbb{1}_{\mathrm{P}_1}$, $\mathrm{X}_2=\mathbb{1}_{\mathrm{P}_2}$ et $\mathrm{X}_3=\mathbb{1}_{\mathrm{A}}$. Les variables $\mathrm{X}_1,\mathrm{X}_2,\mathrm{X}_3$ sont indépendantes deux à deux mais non indépendantes.

Définition 24. *Soit $(\Omega,\mathscr{A},\mathbb{P})$ un espace probabilisé et $(\mathrm{X}_n)_{n\geqslant1}$ une suite de variables aléatoires discrètes. On dit que les variables aléatoires $(\mathrm{X}_n)_{n\geqslant1}$ sont* indépendantes *si toute sous-famille finie de $(\mathrm{X}_n)_{n\geqslant1}$ est formée de variables aléatoires indépendantes.*

Proposition 32. *Soit $(\Omega,\mathscr{A},\mathbb{P})$ un espace probabilisé.*

1. *Si X et Y sont des variables aléatoires indépendantes et f et g sont des applications définies respectivement sur $\mathrm{X}(\Omega)$ et $\mathrm{Y}(\Omega)$, alors $f(\mathrm{X})$ et $g(\mathrm{Y})$ sont indépendantes.*
2. *Si $\mathrm{X}_1,\ldots,\mathrm{X}_n$ sont des variables aléatoires discrètes indépendantes et $f_1,\ldots,f_n$ définies respectivement sur $\mathrm{X}_1(\Omega),\ldots,\mathrm{X}_n(\Omega)$, alors $f_1(\mathrm{X}_1),\ldots,f_n(\mathrm{X}_n)$ sont indépendantes ;*
3. *Plus généralement, si $(\mathrm{X}_n)_n$ est une suite de variables aléatoires discrètes indépendantes et f_n une application définie sur $\mathrm{X}_n(\Omega)$ pour tout n entier, alors $(f_n(\mathrm{X}_n))_n$ est une suite de variables aléatoires indépendantes.*

Démonstration. 1. Soit $(a,b)\in f(\mathrm{X}(\Omega))\times g(\mathrm{Y}(\Omega))$. On a

$$\mathbb{P}(f(\mathrm{X})=a,g(\mathrm{Y})=b)=\mathbb{P}\left(\mathrm{X}\in f^{-1}(\{a\}),\mathrm{Y}\in g^{-1}(\{b\})\right)=\mathbb{P}(f(\mathrm{X})=a)\mathbb{P}(g(\mathrm{Y})=b)$$

2. Soit $(a_i)_{1\leqslant i\leqslant n} \in \prod_{i=1}^{n} f_i(\mathrm{X}_i(\Omega))$. On a

$$\begin{aligned}\mathbb{P}\left(\bigcap_{i=1}^{n}\{f_i(\mathrm{X}_i)=a_i\}\right) &= \mathbb{P}\left(\bigcap_{i=1}^{n}\mathrm{X}_i \in f_i^{-1}(\{a_i\})\right)\\ &= \prod_{i=1}^{n}\mathbb{P}\left(\mathrm{X}_i \in f_i^{-1}(\{a_i\})\right) = \prod_{i=1}^{n}\mathbb{P}(f_i(\mathrm{X}_i)=a_i)\end{aligned}$$

3. Découle du cas précédent.

□

Proposition 33 (Lemme des coalitions). *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé.*

1. *Si $\mathrm{X}_1, \ldots, \mathrm{X}_n$ sont des variables aléatoires indépendantes et f et g sont des applications définies respectivement sur $\prod_{i=1}^{p}\mathrm{X}_i(\Omega)$ et $\prod_{i=p+1}^{n}\mathrm{X}_i(\Omega)$, alors $f(\mathrm{X}_1, \ldots, \mathrm{X}_p)$ et $g(\mathrm{X}_{p+1}, \ldots, \mathrm{X}_n)$ sont indépendantes.*
2. *Plus généralement, si $(\mathrm{X}_n)_n$ est une suite de variables aléatoires indépendantes , alors $f_1(\mathrm{X}_1, \ldots, \mathrm{X}_{n_1})$, $f_2(\mathrm{X}_{n_1+1}, \ldots, \mathrm{X}_{n_2}), \ldots, f_p(\mathrm{X}_{n_{p-1}+1}, \ldots, \mathrm{X}_{n_p}), \ldots$ forment une suite de variables aléatoires indépendantes.*

Démonstration. 1. On pose $\mathrm{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_p)$ et $\mathrm{Y} = (\mathrm{X}_{p+1}, \ldots, \mathrm{X}_n)$. Pour $x = (x_1, \ldots, x_p) \in \mathrm{X}(\Omega)$ et $y = (x_{p+1}, \ldots, x_n) \in \mathrm{Y}(\Omega)$, on a sans difficulté

$$\mathbb{P}(\mathrm{X}=x, \mathrm{Y}=y) = \mathbb{P}\left(\bigcap_{i=1}^{n}\{\mathrm{X}_i = x_i\}\right) = \prod_{i=1}^{n}\mathbb{P}(\mathrm{X}_i = x_i) = \mathbb{P}(\mathrm{X}=x)\mathbb{P}(\mathrm{Y}=y)$$

Il suffit ensuite d'appliquer le résultat du 1. de la proposition 32.
2. On procède à l'identique avec p coalitions puisqu'il suffit de vérifier l'indépendance de toute sous-famille finie. □

Théorème 5. *Il existe un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$ sur lequel existe une suite $(\mathrm{X}_n)_{n\geqslant 1}$ de variables aléatoires indépendantes avec $\mathrm{X}_n \sim \mathscr{L}_n$ pour tout $n \geqslant 1$, la famille $(\mathscr{L}_n)_{n\geqslant 1}$ étant une famille de lois donnée.*

[Admis]

Application : Il existe un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$ sur lequel existe une suite $(\mathrm{X}_n)_{n\geqslant 1}$ de variables aléatoires indépendantes de loi $\mathscr{B}(p)$ avec $p \in [\,0\,;1\,]$. Pour $p = 1/2$, on peut donc considérer le jeu de pile ou face infini. Le choix d'une tribu est une autre affaire !

III Espérance et variance

1 Espérance

Définition 25. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire réelle discrète à valeurs dans $[\,0\,;+\infty\,]$. On définit l'espérance de X notée $\mathbb{E}(\mathrm{X})$ à valeurs dans $[\,0\,;+\infty\,]$ par*

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

Proposition 34 (Antirépartition). *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* X *une variable aléatoire réelle discrète à valeurs dans* $\mathbb{N} \cup \{+\infty\}$. *On a dans* $[\,0\,;+\infty\,]$

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

Démonstration. D'après le théorème de Fubini pour des familles à termes positifs, il vient

$$\begin{aligned}\sum_{n=1}^{+\infty} \mathbb{P}(\mathrm{X} \geqslant n) &= \sum_{n\in\mathbb{N}^*} \sum_{k\in[\![\, n\,;+\infty\,[\![\cup\{+\infty\}} \mathbb{P}(\mathrm{X} = k) \\ &= \sum_{n\in\mathbb{N}^*} \sum_{k\in\mathbb{N}^*\cup\{+\infty\}} \mathbb{1}_{[\![\,1\,;\,k\,]\!]}(n)\mathbb{P}(\mathrm{X} = k)\end{aligned}$$

$$\sum_{n=1}^{+\infty} \mathbb{P}(\mathrm{X} \geqslant n) = \sum_{k\in\mathbb{N}^*\cup\{+\infty\}} \sum_{n\in\mathbb{N}^*} \mathbb{1}_{[\![\,1\,;\,k\,]\!]}(n)\mathbb{P}(\mathrm{X} = k) = \sum_{k\in\mathbb{N}^*\cup\{+\infty\}} k\mathbb{P}(\mathrm{X} = k)$$

d'où le résultat. □

Définition 26. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* X *une variable aléatoire réelle ou complexe discrète. On dit que* X *est* d'espérance finie *si la famille* $(x\mathbb{P}(\mathrm{X} = x))_{x\in\mathrm{X}(\Omega)}$ *est sommable. Dans ce cas, on définit l'*espérance *de* X *notée* $\mathbb{E}(\mathrm{X})$ *par*

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

La condition de sommabilité et l'égalité valent toujours en remplaçant $\mathrm{X}(\Omega)$ *par* D *au plus dénombrable qui contient* $\mathrm{X}(\Omega)$ *(puisque* $\mathbb{P}(\mathrm{X} = x) = 0$ *pour* $x \in \mathrm{D} \smallsetminus \mathrm{X}(\Omega)$*).*

Notation : On note L^1 l'ensemble des variables complexes discrètes d'espérance finie.

Commentaire : L'espérance est la somme des valeurs prises par la variable aléatoire X pondérées par la probabilité que X prenne ces valeurs. Il s'agit d'un moyenne en probabilité de X.

Remarques : (1) Si l'univers Ω est fini, on retrouve la même définition.
(2) Dans la pratique, il n'est pas pertinent de chercher $\mathrm{X}(\Omega)$ si celui-ci n'est pas donné. Un ensemble au plus dénombrable qui le contient suffit.

Définition 27. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* $\mathrm{X} \in \mathrm{L}^1$. *On dit que* X *est* centrée *si* $\mathbb{E}(\mathrm{X}) = 0$.

Proposition 35. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* X *une variable aléatoire constante égale à un réel ou complexe* a. *Alors on a* $\mathrm{X} \in \mathrm{L}^1$ *et* $\mathbb{E}(\mathrm{X}) = a$.

Démonstration. On a $\mathrm{X}(\Omega) = \{a\}$ et le résultat suit. □

Remarque : Le résultat vaut aussi pour une variable constante presque sûrement puisque si $\mathrm{X} = a$ presque sûrement, la famille $(x\mathbb{P}(\mathrm{X} = x))_{x\in\mathrm{X}(\Omega)}$ contient comme seul terme éventuellement non nul $a\mathbb{P}(\mathrm{X} = a)$ c'est-à-dire a.

Théorème 6 (Théorème de transfert). *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé,* X *une variable aléatoire discrète et* $f : \mathrm{X}(\Omega) \to \mathbb{C}$. *On a* $f(\mathrm{X})$ *d'espérance finie si et seulement si* $(f(x)\mathbb{P}(\mathrm{X} = x))_{x\in\mathrm{X}(\Omega)}$ *est sommable et dans ce cas*

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

L'équivalence et l'égalité valent aussi en remplaçant $\mathrm{X}(\Omega)$ *par* D *au plus dénombrable qui contient* $\mathrm{X}(\Omega)$ *(puisque* $\mathbb{P}(\mathrm{X} = x) = 0$ *pour* $x \in \mathrm{D} \smallsetminus \mathrm{X}(\Omega)$*).*

Démonstration. On considère $f : \mathrm{D} \to \mathbb{C}$ avec D au plus dénombrable qui contient $\mathrm{X}(\Omega)$. On dispose du recouvrement disjoint de $\mathrm{X}(\Omega)$:

$$\left(f^{-1}(\{y\}) \cap \mathrm{X}(\Omega)\right)_{y \in f(\mathrm{X})(\Omega)}$$

En effet, on a

$$\bigsqcup_{y \in f(\mathrm{X})(\Omega)} f^{-1}(\{y\}) \cap \mathrm{X}(\Omega) = \mathrm{X}(\Omega) \cap f^{-1}\left(\bigsqcup_{y \in f(\mathrm{X})(\Omega)} \{y\}\right) = \mathrm{X}(\Omega) \cap \underbrace{f^{-1}(f(\mathrm{X}(\Omega)))}_{\supset \mathrm{X}(\Omega)} = \mathrm{X}(\Omega)$$

Par sommation par paquets pour une famille à termes positifs, il vient dans $[\,0\,;+\infty\,]$

$$\begin{aligned}\sum_{x \in \mathrm{D}} |f(x)|\,\mathbb{P}(\mathrm{X}=x) &= \sum_{x \in \mathrm{X}(\Omega)} |f(x)|\,\mathbb{P}(\mathrm{X}=x)\\ &= \sum_{y \in f(\mathrm{X})(\Omega)} \left(\sum_{x \in f^{-1}(\{y\}) \cap \mathrm{X}(\Omega)} |y|\,\mathbb{P}(\mathrm{X}=x)\right) = \sum_{y \in f(\mathrm{X})(\Omega)} |y|\,\mathbb{P}(f(\mathrm{X})=y)\end{aligned}$$

On en déduit l'équivalence des sommabilités. Ainsi, quand cette condition est réalisée, on obtient toujours par sommation par paquets avec le même recouvrement disjoint que précédemment

$$\begin{aligned}\sum_{x \in \mathrm{D}} f(x)\mathbb{P}(\mathrm{X}=x) &= \sum_{x \in \mathrm{X}(\Omega)} f(x)\mathbb{P}(\mathrm{X}=x)\\ &= \sum_{y \in f(\mathrm{X})(\Omega)} \left(\sum_{x \in f^{-1}(\{y\}) \cap \mathrm{X}(\Omega)} y\mathbb{P}(\mathrm{X}=x)\right) = \sum_{y \in f(\mathrm{X})(\Omega)} y\mathbb{P}(f(\mathrm{X})=y)\end{aligned}$$

d'où le résultat annoncé. □

Remarque : Dans le cas particulier (fréquent) où $\mathrm{X}(\Omega) \subset \mathbb{N}$, on a

$$\begin{aligned}f(\mathrm{X}) \in \mathrm{L}^1 &\iff (f(n)\mathbb{P}(\mathrm{X}=n))_{n \in \mathbb{N}} \in \ell^1(\mathbb{N})\\ &\iff \sum f(n)\mathbb{P}(\mathrm{X}=n) \text{ converge absolument}\end{aligned}$$

Théorème 7 (Linéarité de l'espérance). *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé. L'ensemble $\mathbb{K}^\Omega \cap \mathrm{L}^1$ est un $\mathbb{K}$-ev et l'application définie sur cet espace par $\mathrm{X} \mapsto \mathbb{E}(\mathrm{X})$ est linéaire.*

Démonstration. La variable aléatoire nulle est clairement d'espérance finie. On pose $\mathrm{Z} = \mathrm{X} + \lambda\mathrm{Y} = f(\mathrm{X}, \mathrm{Y})$ avec X et Y variables aléatoires à valeurs dans $\mathbb{K}$ d'espérance finie et $\lambda \in \mathbb{K}$. On a $(\mathrm{X}, \mathrm{Y})(\Omega) \subset \mathrm{X}(\Omega) \times \mathrm{Y}(\Omega)$ et le produit $\mathrm{X}(\Omega) \times \mathrm{Y}(\Omega)$ est au plus dénombrable. On travaille sur cet ensemble pour la suite, les calculs s'en trouvant grandement simplifiés. Par transfert (vers la loi du couple (X, Y)), on a Z d'espérance finie si et seulement si la famille $(f(x,y)\mathbb{P}(\mathrm{X}=x, \mathrm{Y}=y))_{(x,y) \in \mathrm{X}(\Omega) \times \mathrm{Y}(\Omega)}$ sommable. Or

$$\forall (x,y) \in \mathrm{X}(\Omega) \times \mathrm{Y}(\Omega) \qquad |f(x,y)| \leqslant |x| + |\lambda|\,|y|$$

On a la sommabilité de $(|x|\,\mathbb{P}(\mathrm{X}=x, \mathrm{Y}=y))_{(x,y) \in \mathrm{X}(\Omega) \times \mathrm{Y}(\Omega)}$ et $(|y|\,\mathbb{P}(\mathrm{X}=x, \mathrm{Y}=y))_{(x,y) \in (\mathrm{X},\mathrm{Y})(\Omega)}$. En effet, d'après le théorème de Fubini pour une famille à termes positifs et par probabilités totales, on a

$$\begin{aligned}\sum_{(x,y) \in \mathrm{X}(\Omega) \times \mathrm{Y}(\Omega)} |x|\,\mathbb{P}(\mathrm{X}=x, \mathrm{Y}=y) &= \sum_{x \in \mathrm{X}(\Omega)} |x| \left(\sum_{y \in \mathrm{Y}(\Omega)} \mathbb{P}(\mathrm{X}=x, \mathrm{Y}=y)\right)\\ &= \sum_{x \in \mathrm{X}(\Omega)} |x|\,\mathbb{P}(\mathrm{X}=x) < +\infty\end{aligned}$$

On procède de même pour l'autre famille. Ainsi, par transfert, il vient que Z est d'espérance finie puis, par transfert, linéarité du symbole somme, théorème de Fubini et probabilités totales :

$$\mathbb{E}(X+\lambda Y)=\mathbb{E}(Z)=\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} f(x,y)\mathbb{P}(X=x,Y=y)$$
$$=\sum_{x\in X(\Omega)} x\mathbb{P}(X=x)+\lambda\sum_{y\in Y(\Omega)} y\mathbb{P}(Y=y)=\mathbb{E}(X)+\lambda\mathbb{E}(Y)$$

□

Remarque : On a utilisé avec succès l'ensemble $D = X(\Omega)\times Y(\Omega)$ au plus dénombrable qui contient $(X,Y)(\Omega)$ pour appliquer le théorème de transfert.

Proposition 36 (Positivité, croissance de l'espérance). *Soit* $(\Omega,\mathscr{A},\mathbb{P})$ *un espace probabilisé.*

1. *Soit* $X\in\mathbb{R}_+^\Omega\cap L^1$, *alors* $\mathbb{E}(X)\geqslant 0$.
2. *Soient* X, Y *dans* $\mathbb{R}^\Omega\cap L^1$. *On a*

$$X\leqslant Y\Longrightarrow \mathbb{E}(X)\leqslant\mathbb{E}(Y)$$

Démonstration. 1. Immédiate.
2. Conséquence de la positivité et de la linéarité de l'espérance appliquée à $Y-X$. □

Proposition 37 (Inégalité triangulaire). *Soit* $(\Omega,\mathscr{A},\mathbb{P})$ *un espace probabilisé et* $X\in L^1$. *On a*

$$|\mathbb{E}(X)|\leqslant\mathbb{E}(|X|)$$

Démonstration. Par inégalité triangulaire appliquée à la famille sommable $(x\mathbb{P}(X=x))_{x\in X(\Omega)}$,

il vient $$\left|\sum_{x\in X(\Omega)} x\mathbb{P}(X=x)\right|\leqslant\sum_{x\in X(\Omega)}|x|\,\mathbb{P}(X=x)$$

c'est-à-dire, après transfert $$|\mathbb{E}(X)|\leqslant\mathbb{E}(|X|)$$

□

Théorème 8. *Soit* $(\Omega,\mathscr{A},\mathbb{P})$ *un espace probabilisé. Si* X *et* Y *sont dans* L^1 *et indépendantes, alors on a* $XY\in L^1$ *et*

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

Démonstration. On a $(X,Y)(\Omega)\subset X(\Omega)\times Y(\Omega)$ et le produit $X(\Omega)\times Y(\Omega)$ est au plus dénombrable. On travaille sur cet ensemble pour la suite, les calculs s'en trouvant grandement simplifiés. On pose $Z = XY$. Par transfert (vers la loi du couple (X,Y)), on a Z d'espérance finie si et seulement si $(xy\mathbb{P}(X=x,Y=y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$ sommable. On a

$$\forall (x,y)\in X(\Omega)\times Y(\Omega)\qquad xy\mathbb{P}(X=x,Y=y)=x\mathbb{P}(X=x)y\mathbb{P}(Y=y)$$

Or, les familles $(x\mathbb{P}(X=x))_{x\in X(\Omega)}$ et $(y\mathbb{P}(Y=y))_{y\in Y(\Omega)}$ sont sommables d'où, par théorème de Fubini, la sommabilité de $(xy\mathbb{P}(X=y)\mathbb{P}(Y=y))_{(x,y)\in(X,Y)(\Omega)}$. On conclut par transfert et théorème de Fubini

$$\mathbb{E}(Z)=\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} xy\mathbb{P}(X=x,Y=y)=\sum_{x\in X(\Omega)} x\mathbb{P}(X=x)\sum_{y\in Y(\Omega)} y\mathbb{P}(Y=y)=\mathbb{E}(X)\mathbb{E}(Y)$$

□

Remarques : (1) La réciproque est fausse.
Soient X, Y deux variables aléatoires indépendantes de même loi $\mathscr{U}_{\{-1,1\}}$ et posons $\mathrm{U}=\mathrm{X}+\mathrm{Y}$, $\mathrm{V}=\mathrm{X}-\mathrm{Y}$. On a

$$\mathbb{E}(\mathrm{U})\mathbb{E}(\mathrm{V})=0 \quad \text{et} \quad \mathbb{E}(\mathrm{UV})=\mathbb{E}(\mathrm{X}^2)-\mathbb{E}(\mathrm{Y}^2)=0$$

Pourtant, les variables U et V ne sont pas indépendantes puisque par exemple

$$\mathbb{P}(\mathrm{U}=2,\mathrm{V}=2)=\mathbb{P}(\mathrm{X}+\mathrm{Y}=2,\mathrm{X}-\mathrm{Y}=2)=\mathbb{P}(\mathrm{X}=2,\mathrm{Y}=0)=0$$

et $$\mathbb{P}(\mathrm{U}=2)=\mathbb{P}(\mathrm{X}=1,\mathrm{Y}=1)=\left(\frac{1}{2}\right)^2, \quad \mathbb{P}(\mathrm{V}=2)=\mathbb{P}(\mathrm{X}=1,\mathrm{Y}=-1)=\left(\frac{1}{2}\right)^2$$

Si on veut un contre-exemple avec des variables non centrées, il suffit de considérer $a+\mathrm{U}$ et $b+\mathrm{V}$ avec a et b non nuls. On a

$$\mathbb{E}(\mathrm{U}+a)\mathbb{E}(\mathrm{V}+b)=ab \quad \text{et} \quad \mathbb{E}((\mathrm{U}+a)(\mathrm{V}+b))=\mathbb{E}(\mathrm{UV}+a\mathrm{V}+b\mathrm{U}+ab)=ab$$

(2) Pour X, Y variables indépendantes, on a l'équivalence :

$$\mathrm{X},\mathrm{Y} \text{ dans } \mathrm{L}^1 \iff \mathrm{XY}\in\mathrm{L}^1$$

En effet, on a

$$\begin{aligned}\mathrm{XY}\in\mathrm{L}^1 &\iff (x\mathbb{P}(\mathrm{X}=x)y\mathbb{P}(\mathrm{Y}=y))_{(x,y)\in\mathrm{X}(\Omega)\times\mathrm{Y}(\Omega)} \text{ sommable}\\ &\iff (x\mathbb{P}(\mathrm{X}=x))_{x\in\mathrm{X}(\Omega)} \quad \text{et} \quad (y\mathbb{P}(\mathrm{Y}=y))_{y\in\mathrm{Y}(\Omega)} \text{ sommables}\end{aligned}$$

la dernière équivalence résultant du théorème de Fubini.

Corollaire 2. *Soit* $(\Omega,\mathscr{A},\mathbb{P})$ *un espace probabilisé. Soient* $\mathrm{X}_1,\ldots,\mathrm{X}_n$ *dans* L^1 *et indépendantes. Alors on a* $\prod_{i=1}^n \mathrm{X}_i\in\mathrm{L}^1$ *et*

$$\mathbb{E}\left(\prod_{i=1}^n \mathrm{X}_i\right)=\prod_{i=1}^n \mathbb{E}(\mathrm{X}_i)$$

Démonstration. On procède par récurrence. Le cas $n=1$ est immédiat et s'il a lieu au rang $n-1\geqslant 1$ fixé, on applique le théorème 8 à $\prod_{i=1}^{n-1}\mathrm{X}_i$ et X_n qui sont indépendantes par coalition. L'hérédité suit. □

Proposition 38. *Soit* $(\Omega,\mathscr{A},\mathbb{P})$ *un espace probabilisé et* X,Y *des variables aléatoires discrètes telles que* $\mathrm{Y}\in\mathbb{R}_+^\Omega\cap\mathrm{L}^1$ *et* $|\mathrm{X}|\leqslant\mathrm{Y}$. *Alors on a* $\mathrm{X}\in\mathrm{L}^1$ *et* $\mathbb{E}(|\mathrm{X}|)\leqslant\mathbb{E}(\mathrm{Y})$.

Démonstration. Considérons $\mathrm{Z}=(|\mathrm{X}|,\mathrm{Y})$ et p_1,p_2 définies sur $\mathrm{Z}(\Omega)$ par $p_1:(x,y)\mapsto x$, $p_2:(x,y)\mapsto y$. La variable $\mathrm{Y}=p_2(\mathrm{Z})$ est d'espérance finie d'où, par transfert, la sommabilité de $(p_2(z)\mathbb{P}(\mathrm{Z}=z))_{z\in\mathrm{Z}(\Omega)}$. Or, on a l'encadrement $0\leqslant p_1(\mathrm{Z})\leqslant p_2(\mathrm{Z})$ par hypothèse. On en déduit

$$\forall z\in\mathrm{Z}(\Omega) \qquad 0\leqslant p_1(\mathrm{Z})\mathbb{1}_{\{\mathrm{Z}=z\}}\leqslant p_2(\mathrm{Z})\mathbb{1}_{\{\mathrm{Z}=z\}}$$

Et passant à l'espérance

$$\forall z\in\mathrm{Z}(\Omega) \qquad 0\leqslant p_1(z)\mathbb{P}(\mathrm{Z}=z)\leqslant p_2(z)\mathbb{P}(\mathrm{Z}=z)$$

La sommabilité de $(p_2(z)\mathbb{P}(\mathrm{Z}=z))_{z\in\mathrm{Z}(\Omega)}$ implique celle de $(p_1(z)\mathbb{P}(\mathrm{Z}=z))_{z\in\mathrm{Z}(\Omega)}$. Par comparaison et transfert, on conclut

$$\sum_{z\in\mathrm{Z}(\Omega)} p_1(z)\mathbb{P}(\mathrm{Z}=z)=\mathbb{E}(p_1(\mathrm{Z}))=\mathbb{E}(|\mathrm{X}|)\leqslant\sum_{z\in\mathrm{Z}(\Omega)} p_2(z)\mathbb{P}(\mathrm{Z}=z)=\mathbb{E}(p_2(\mathrm{Z}))=\mathbb{E}(\mathrm{Y})$$

Variante. Par transfert, on peut aussi annoncer que $|\mathrm{X}|$ est d'espérance finie et conclure par croissance de l'espérance. □

Remarque : Si X est une variable aléatoire réelle ou complexe discrète bornée, alors elle est d'espérance finie. En effet, il existe $C \geqslant 0$ tel que $|X| \leqslant C$ et C est d'espérance finie d'où le résultat.

Proposition 39. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé et X variable aléatoire réelle discrète positive. On a*

$$\mathbb{E}(X) = 0 \iff X = 0 \text{ p.s.}$$

Démonstration. On a $\mathbb{E}(X) = \sum_{x \in X(\Omega)} \underbrace{x\mathbb{P}(X = x)}_{\geqslant 0}$ d'où

$$\begin{aligned}\mathbb{E}(X) = 0 &\iff \forall x \in X(\Omega) \qquad x\mathbb{P}(X = x) = 0 \\ &\iff \forall x \in X(\Omega) \smallsetminus \{0\} \qquad \mathbb{P}(X = x) = 0 \iff \mathbb{P}(X = 0) = 1\end{aligned}$$

la dernière équivalence résultant de $1 = \mathbb{P}(X = 0) + \sum_{x \in X(\Omega) \smallsetminus \{0\}} \mathbb{P}(X = x)$. □

2 Variance et écart-type

Théorème 9. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire réelle discrète. Si X^2 est d'espérance finie, alors X est également d'espérance finie.*

Démonstration. On a $$(|X| - 1)^2 \geqslant 0 \iff |X| \leqslant \frac{1}{2}(1 + X^2)$$

Le résultat suit d'après le théorème 7 et la proposition 38. □

Notation : On note $X \in L^2$ pour signifier que X est réelle avec X^2 d'espérance finie.
Ainsi, on a $L^2 \subset L^1$.

Remarque : L'inclusion réciproque est fausse. Considérer X variable aléatoire avec $X(\Omega) = \mathbb{N}^*$ et $\mathbb{P}(X = n) = \dfrac{1}{\zeta(3)n^3}$ pour n entier non nul.

Définition 28. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé et $X \in L^2$. on définit la* variance *de X notée $\mathbb{V}(X)$ et l'*écart-type *de X noté $\sigma(X)$ par*

$$\mathbb{V}(X) = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right] \quad \text{et} \quad \sigma(X) = \sqrt{\mathbb{V}(X)}$$

Remarque : D'après les théorèmes 7 et 9, $(X - \mathbb{E}(X))^2 = X^2 - 2X \times \mathbb{E}(X) + \mathbb{E}(X)^2$ est d'espérance finie.

Commentaire : La variance de X est la moyenne en probabilité des écarts quadratiques de X par rapport à sa moyenne en probabilité. Cette grandeur mesure la dispersion de X autour de son espérance.

Proposition 40 (Relation de König-Huygens). *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé et $X \in L^2$. On a*

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

Démonstration. On a par linéarité

$$\mathbb{V}(X) = \mathbb{E}\left[X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2\right] = \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

□

Proposition 41. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* $X \in L^2$. *On a*

$$\forall (a, b) \in \mathbb{R}^2 \qquad aX + b \in L^2 \quad \text{et} \quad \mathbb{V}(aX + b) = a^2\mathbb{V}(X)$$

Démonstration. On a $(aX + b)^2 = a^2X^2 + 2abX + b^2$ d'espérance finie d'après les théorèmes 7 et 9. On a par propriété sur l'espérance

$$\mathbb{V}(aX + b) = \mathbb{E}\left[(aX + b - \mathbb{E}(aX + b))^2\right] = \mathbb{E}\left[(aX + b - (a\mathbb{E}(X) + b))^2\right] = a^2\mathbb{V}(X)$$

□

Définition 29. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* $X \in L^2$. *On dit que* X *est* réduite *si* $\mathbb{V}(X) = 1$.

Proposition 42. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* $X \in L^2$. *Si* $\sigma(X) > 0$, *alors la variable aléatoire* $\dfrac{X - \mathbb{E}(X)}{\sigma(X)}$ *est centrée réduite.*

Démonstration. Immédiate. □

Proposition 43. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* X, Y *dans* L^2. *Alors, on a* $XY \in L^1$.

Démonstration. On a

$$(|X| - |Y|)^2 \geqslant 0 \iff |XY| \leqslant \frac{1}{2}(X^2 + Y^2)$$

D'après le théorème 7, la variable aléatoire $\dfrac{1}{2}(X^2 + Y^2)$ est d'espérance finie puis on conclut avec la proposition 38. □

Théorème 10. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé . L'ensemble* L^2 *est un* $\mathbb{R}$*-ev.*

Démonstration. La variable aléatoire nulle est dans L^2. Soient X, Y dans L^2 et λ réel. On a $(X + \lambda Y)^2 = X^2 + 2\lambda XY + \lambda^2 Y^2$ et, d'après la proposition précédente, chaque terme est d'espérance finie d'où $(X + \lambda Y)^2 \in L^1$ d'après le théorème 7. Il en résulte que l'ensemble L^2 est un sev de $\mathbb{R}^\Omega \cap L^1$. □

Théorème 11 (Inégalité de Cauchy-Schwarz). *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* X, Y *dans* L^2. *Alors*

$$\mathbb{E}(XY)^2 \leqslant \mathbb{E}(X^2)\mathbb{E}(Y^2)$$

Démonstration. L'application $(X, Y) \mapsto \mathbb{E}(XY)$ est une forme bilinéaire, symétrique, positive sur L^2. L'inégalité de Cauchy-Schwarz s'applique donc (voir cours **Espaces préhilbertiens réels**). □

Définition 30. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* X, Y *dans* L^2. *On définit la* covariance *de* X *et* Y *notée* $\mathrm{Cov}(X, Y)$ *par*

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right]$$

Commentaire : La covariance de X et Y mesure partiellement la manière dont X et Y sont dépendantes.

Remarques : (1) Les variables $\mathrm{X}-\mathbb{E}(\mathrm{X})$ et $\mathrm{Y}-\mathbb{E}(\mathrm{Y})$ sont dans L^2 (proposition 41) et on applique la proposition 43 pour justifier $(\mathrm{X}-\mathbb{E}(\mathrm{X}))(\mathrm{Y}-\mathbb{E}(\mathrm{Y})) \in \mathrm{L}^1$.
(2) La covariance est une forme bilinéaire symétrique positive sur L^2. On peut donc écrire l'inégalité de Cauchy-Schwarz pour la covariance et on obtient :

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

Proposition 44 (Relation de König-Huygens). *Soit* $(\Omega,\mathscr{A},\mathbb{P})$ *un espace probabilisé et* X,Y *dans* L^2. *On a*

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

Démonstration. On développe le produit puis on utilise la linéarité de l'espérance, chaque terme étant d'espérance finie d'après le théorème 9. □

Proposition 45. *Soit* $(\Omega,\mathscr{A},\mathbb{P})$ *un espace probabilisé et* X,Y *dans* $\mathbb{R}^\Omega \cap \mathrm{L}^1$ *indépendantes. Alors, on peut définir* $\mathrm{Cov}(\mathrm{X},\mathrm{Y})$ *et on a* $\mathrm{Cov}(\mathrm{X},\mathrm{Y}) = 0$.

Démonstration. D'après le théorème 8, la variable $(\mathrm{X}-\mathbb{E}(\mathrm{X}))(\mathrm{Y}-\mathbb{E}(\mathrm{Y}))$ est d'espérance finie et on a

$$\mathbb{E}\left[(\mathrm{X}-\mathbb{E}(\mathrm{X}))(\mathrm{Y}-\mathbb{E}(\mathrm{Y}))\right] = \mathbb{E}(\mathrm{X}-\mathbb{E}(\mathrm{X}))\mathbb{E}(\mathrm{Y}-\mathbb{E}(\mathrm{Y})) = 0$$

□

Remarque : La réciproque est fausse.
Reprenons le contre-exemple du théorème 8. Soient X,Y deux variables aléatoires indépendantes de même loi $\mathscr{U}_{\{-1,1\}}$ et posons $\mathrm{U}=\mathrm{X}+\mathrm{Y}$, $\mathrm{V}=\mathrm{X}-\mathrm{Y}$. On a $\mathbb{E}(\mathrm{U})=\mathbb{E}(\mathrm{V})=\mathbb{E}(\mathrm{UV})=0$ d'où $\mathrm{Cov}(\mathrm{U},\mathrm{V})=0$ mais U et V ne sont pas indépendantes.

Théorème 12. *Soit* $(\Omega,\mathscr{A},\mathbb{P})$ *un espace probabilisé.*

1. *Soient* X,Y *dans* L^2. *On a*

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

2. *Soient* $\mathrm{X}_1,\ldots,\mathrm{X}_n$ *dans* L^2. *On a*

$$\begin{aligned}\mathbb{V}\left(\sum_{i=1}^n \mathrm{X}_i\right) &= \sum_{1\leqslant i,j\leqslant n} \mathrm{Cov}(\mathrm{X}_i,\mathrm{X}_j)\\ &= \sum_{i=1}^n \mathbb{V}(\mathrm{X}_i) + 2\sum_{1\leqslant i<j\leqslant n} \mathrm{Cov}(\mathrm{X}_i,\mathrm{X}_j)\end{aligned}$$

Démonstration. 1. Notons $\mathrm{U}=\mathrm{X}-\mathbb{E}(\mathrm{X})$ et $\mathrm{V}=\mathrm{Y}-\mathbb{E}(\mathrm{Y})$. Les variable U, V sont dans L^2 et par linéarité de l'espérance

$$\mathbb{V}(\mathrm{X}+\mathrm{Y}) = \mathbb{E}\left[(\mathrm{U}+\mathrm{V})^2\right] = \mathbb{E}(\mathrm{U}^2) + 2\mathbb{E}(\mathrm{UV}) + \mathbb{E}(\mathrm{V}^2)$$

Autrement dit $$\mathbb{V}(\mathrm{X}+\mathrm{Y}) = \mathbb{V}(\mathrm{X}) + 2\mathrm{Cov}(\mathrm{X},\mathrm{Y}) + \mathbb{V}(\mathrm{Y})$$

2. C'est juste une généralisation de ce qui précède. Notons $\mathrm{U}_i = \mathrm{X}_i - \mathbb{E}(\mathrm{X}_i)$ pour $i \in [\![1\,;n]\!]$. Les U_i sont dans L^2 et par linéarité de l'espérance

$$\mathbb{V}\left(\sum_{i=1}^{n}\mathrm{X}_i\right) = \mathbb{E}\left[\left(\sum_{i=1}^{n}\mathrm{U}_i\right)^2\right] = \mathbb{E}\left(\sum_{1\leqslant i,j\leqslant n}\mathrm{U}_i\mathrm{U}_j\right) = \sum_{1\leqslant i,j\leqslant n}\mathbb{E}(\mathrm{U}_i\mathrm{U}_j)$$
$$= \sum_{1\leqslant i,j\leqslant n}\mathrm{Cov}(\mathrm{X}_i,\mathrm{X}_j) = \sum_{i=1}^{n}\mathbb{V}(\mathrm{X}_i) + 2\sum_{1\leqslant i<j\leqslant n}\mathrm{Cov}(\mathrm{X}_i,\mathrm{X}_j)$$

□

Corollaire 3. *Soit* $(\Omega,\mathscr{A},\mathbb{P})$ *un espace probabilisé. Soient* $\mathrm{X}_1,\dots,\mathrm{X}_n$ *dans* L^2 *et* décorrelées, *c'est-à-dire* $\mathrm{Cov}(\mathrm{X}_i,\mathrm{X}_j)=0$ *pour* $i\neq j$. *Alors*

$$\mathbb{V}\left(\sum_{i=1}^{n}\mathrm{X}_i\right) = \sum_{i=1}^{n}\mathbb{V}(\mathrm{X}_i)$$

Démonstration. Conséquence immédiate du théorème précédent. □

Commentaire : Si les variables $\mathrm{X}_1,\dots,\mathrm{X}_n$ sont indépendantes ou même simplement indépendantes deux à deux, alors elles sont décorrelées.

3 Inégalités en probabilités

Théorème 13 (Inégalité de Markov). *Soit* $(\Omega,\mathscr{A},\mathbb{P})$ *un espace probabilisé et* $\mathrm{X}\in\mathrm{L}^1$ *positive. On a*

$$\forall\varepsilon>0 \qquad \mathbb{P}(\mathrm{X}\geqslant\varepsilon)\leqslant\frac{\mathbb{E}(\mathrm{X})}{\varepsilon}$$

Démonstration. Soit $\varepsilon>0$. On a les inégalités

$$\mathrm{X} = \mathrm{X}\mathbb{1}_{\{\mathrm{X}<\varepsilon\}} + \mathrm{X}\mathbb{1}_{\{\mathrm{X}\geqslant\varepsilon\}} \geqslant \mathrm{X}\mathbb{1}_{\{\mathrm{X}\geqslant\varepsilon\}} \geqslant \varepsilon\mathbb{1}_{\{\mathrm{X}\geqslant\varepsilon\}}$$

la première résultant de la positivité de X. Par croissance et linéarité de l'espérance, il s'ensuit

$$\mathbb{E}(\mathrm{X}) \geqslant \mathbb{E}(\mathrm{X}\mathbb{1}_{\{\mathrm{X}\geqslant\varepsilon\}}) \geqslant \mathbb{E}(\varepsilon\mathbb{1}_{\{\mathrm{X}\geqslant\varepsilon\}}) = \varepsilon\mathbb{P}(\mathrm{X}\geqslant\varepsilon)$$

d'où le résultat. □

Remarques : (1) On peut assouplir les hypothèses en exigeant seulement que X soit une variable aléatoire discrète positive même si ça ne présente pas réellement d'intérêt de travailler avec $\mathrm{X}\notin\mathrm{L}^1$.
(2) On peut majorer de même $\mathbb{P}(\mathrm{X}>\varepsilon)$ avec $\varepsilon>0$ en utilisant $\{\mathrm{X}>\varepsilon\}\subset\{\mathrm{X}\geqslant\varepsilon\}$.
(3) Pour le même effort, on peut écrire une inégalité plus fine que celle de Markov :

$$\forall\varepsilon>0 \qquad \mathbb{P}(\mathrm{X}\geqslant\varepsilon)\leqslant\frac{\mathbb{E}(\mathrm{X}\mathbb{1}_{\{\mathrm{X}\geqslant\varepsilon\}})}{\varepsilon}$$

Par double-limite, on peut en déduire

$$\mathbb{P}(\mathrm{X}\geqslant\varepsilon) \underset{\varepsilon\to+\infty}{=} \mathrm{o}\left(\frac{1}{\varepsilon}\right)$$

Application importante : La méthode de Chernoff

Soient $\mathrm{X}_1,\dots,\mathrm{X}_n$ des variables aléatoires réelles discrètes indépendantes et $t>0$ tel que $\mathrm{e}^{t\mathrm{X}_i}\in\mathrm{L}^1$ pour tout $i\in[\![1\,;n]\!]$. Soit m réel. On a par croissance stricte de $u\mapsto\mathrm{e}^{tu}$ sur $\mathbb{R}$

$$\left\{\sum_{i=1}^{n}\mathrm{X}_i\geqslant m\right\} = \left\{\exp\left(t\sum_{i=1}^{n}\mathrm{X}_i\right)\geqslant\mathrm{e}^{tm}\right\} = \left\{\prod_{i=1}^{n}\mathrm{e}^{t\mathrm{X}_i}\geqslant\mathrm{e}^{tm}\right\}$$

La variable $\prod_{i=1}^{n} \mathrm{e}^{t\mathrm{X}_i}$ est positive, dans L^1 en tant que produit de variables aléatoires indépendantes dans L^1 et il vient d'après l'inégalité de Markov

$$\mathbb{P}\left(\sum_{i=1}^{n} \mathrm{X}_i \geqslant m\right) \leqslant \mathrm{e}^{-tm}\mathbb{E}\left(\prod_{i=1}^{n} \mathrm{e}^{t\mathrm{X}_i}\right) = \mathrm{e}^{-tm} \prod_{i=1}^{n} \mathbb{E}(\mathrm{e}^{t\mathrm{X}_i})$$

Cette méthode est très utilisée pour obtenir des inégalités de type *grandes déviations.*

Théorème 14 (Inégalité de Bienaymé-Tchebychev). *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* $\mathrm{X} \in \mathrm{L}^2$*. On a*

$$\forall \varepsilon > 0 \qquad \mathbb{P}(|\mathrm{X} - \mathbb{E}(\mathrm{X})| \geqslant \varepsilon) \leqslant \frac{\mathbb{V}(\mathrm{X})}{\varepsilon^2}$$

Démonstration. On a $\mathrm{X} \in \mathrm{L}^2$ d'où $(\mathrm{X} - \mathbb{E}(\mathrm{X}))^2 \in \mathrm{L}^1$ (proposition 41) et on applique l'inégalité de Markov en remarquant par croissance stricte de $u \mapsto u^2$ sur $\mathbb{R}_+$ l'égalité pour $\varepsilon > 0$

$$\{|\mathrm{X} - \mathbb{E}(\mathrm{X})| \geqslant \varepsilon\} = \{(\mathrm{X} - \mathbb{E}(\mathrm{X}))^2 \geqslant \varepsilon^2\}$$

$\square$

IV Fonctions génératrices

1 Définition

Théorème 15. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* X *une variable aléatoire discrète à valeurs dans* $\mathbb{N}$*. Le rayon de convergence de la série entière* $\sum t^n \mathbb{P}(\mathrm{X} = n)$ *est supérieur ou égal à 1 et la série converge normalement sur* $[-1\,;1]$*.*

Démonstration. On a $\mathbb{P}(\mathrm{X} = n) = \mathrm{O}(1)$ pour tout n entier d'où un rayon de convergence supérieure ou égal à 1 (rayon de $\sum t^n$). Notant $u_n : t \mapsto t^n \mathbb{P}(\mathrm{X} = n)$ pour n entier, on a $\|u_n\|_{\infty,[-1\,;1]} = \mathbb{P}(\mathrm{X} = n)$ et la convergence normale suit par σ-additivité. $\square$

Remarque : Si $\mathrm{X}(\Omega)$ est fini, la série entière est polynomiale et son rayon de convergence est $+\infty$.

Définition 31. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* X *une variable aléatoire discrète à valeurs dans* $\mathbb{N}$*. On définit la* fonction génératrice *de* X *notée* G_X *par*

$$\forall t \in [0\,;1] \qquad \mathrm{G}_\mathrm{X}(t) = \mathbb{E}(t^\mathrm{X}) = \sum_{n=0}^{+\infty} t^n \mathbb{P}(\mathrm{X} = n)$$

Remarques : (1) L'égalité entre les deux écritures résulte du théorème de transfert. La fonction est bien définie sur $[0\,;1]$ d'après le théorème précédent.
(2) On pourrait étendre la définition au segment $[-1\,;1]$ mais travailler sur $[0\,;1]$ présente un intérêt majeur : on ne manipule que des séries à termes positifs ce qui s'avère, en pratique, extrêmement confortable.

Corollaire 4. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* X *une variable aléatoire discrète à valeurs dans* $\mathbb{N}$*. La fonction génératrice* G_X *est continue sur* $[0\,;1]$*, à valeurs dans* $[0\,;1]$*.*

Démonstration. La série de fonctions continues $\sum u_n$ converge normalement donc uniformément sur $[\,0\,;1\,]$ et on a

$$\forall t \in [\,0\,;1\,] \qquad 0 \leqslant \mathrm{G_X}(t) = \sum_{n=0}^{+\infty} t^n \mathbb{P}(\mathrm{X}=n) = \sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{X}=n) = 1$$

□

Exemples : 1. Si $\mathrm{X} \sim \mathscr{B}(p)$, on a $\mathrm{G_X}(t) = pt+1-p$ pour $t \in [\,0\,;1\,]$.
2. Si $\mathrm{X} \sim \mathscr{B}(n,p)$, on a $\mathrm{G_X}(t) = (pt+1-p)^n$ pour $t \in [\,0\,;1\,]$.

2 Propriétés

Théorème 16. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire discrète à valeurs dans $\mathbb{N}$. La fonction génératrice $\mathrm{G_X}$ caractérise la loi de X.*

Démonstration. La fonction $\mathrm{G_X}$ coïncide sur $[\,0\,;1\,[$ avec la somme de la somme de la série entière $\sum t^n \mathbb{P}(\mathrm{X}=n)$ qui est de classe $\mathscr{C}^\infty$ sur $]-1\,;1\,[$. Ainsi, la fonction $\mathrm{G_X}$ admet des dérivées en 0 à tout ordre avec

$$\forall n \in \mathbb{N} \qquad \mathrm{G_X}^{(n)}(0) = n!\,\mathbb{P}(\mathrm{X}=n)$$

Le résultat suit. □

Proposition 46. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire discrète à valeurs dans $\mathbb{N}$. On a*

$$\mathrm{X} \in \mathrm{L}^1 \iff \mathrm{G_X} \textit{ dérivable en } 1$$

et pour $\mathrm{X} \in \mathrm{L}^1$ $$\mathbb{E}(\mathrm{X}) = \mathrm{G_X}'(1)$$

Démonstration. Les fonctions u_n sont de classe $\mathscr{C}^1$ avec $u_n'(t) = nt^{n-1}\mathbb{P}(\mathrm{X}=n)$ pour $(n,t) \in \mathbb{N}^* \times [\,0\,;1\,]$ et $\|u_n'\|_\infty = n\mathbb{P}(\mathrm{X}=n)$. Si $\mathrm{X} \in \mathrm{L}^1$, alors la série $\sum n\mathbb{P}(\mathrm{X}=n)$ converge d'où la convergence normale donc uniforme de $\sum u_n'$ sur $[\,0\,;1\,]$ et on a la convergence simple de $\sum u_n$. Il s'ensuit que $\mathrm{G_X}$ est dérivable sur $[\,0\,;1\,]$ et

$$\mathrm{G_X}'(1) = \sum_{n=1}^{+\infty} u_n'(1) = \sum_{n=0}^{+\infty} n\mathbb{P}(\mathrm{X}=n) = \mathbb{E}(\mathrm{X})$$

Réciproquement, on a pour $t \in [\,0\,;1\,[$

$$\frac{\mathrm{G_X}(t) - \mathrm{G_X}(1)}{t-1} = \sum_{n=0}^{+\infty} \frac{t^n-1}{t-1}\mathbb{P}(\mathrm{X}=n) = \sum_{n=1}^{+\infty} (1+t+\ldots+t^{n-1})\,\mathbb{P}(\mathrm{X}=n)$$

D'où, pour N entier

$$\sum_{n=0}^{\mathrm{N}} n\mathbb{P}(\mathrm{X}=n) = \lim_{t\to 1} \underbrace{\sum_{n=0}^{\mathrm{N}} (1+t+\ldots+t^{n-1})\,\mathbb{P}(\mathrm{X}=n)}_{\leqslant \sum_{n=0}^{+\infty} \cdots} \leqslant \lim_{t\to 1} \frac{\mathrm{G_X}(t) - \mathrm{G_X}(1)}{t-1} = \mathrm{G_X}'(1)$$

Le résultat suit. □

Proposition 47. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire discrète à valeurs dans $\mathbb{N}$. Si $\mathrm{X} \in \mathrm{L}^2$, alors $\mathrm{G_X}$ est dérivable deux fois en 1 et on*

$$\mathbb{V}(\mathrm{X}) = \mathrm{G_X}''(1) + \mathrm{G_X}'(1) - \mathrm{G_X}'(1)^2$$

Démonstration. Les fonctions u_n sont de classe $\mathscr{C}^2$ avec $u_n''(t) = n(n-1)t^{n-2}\mathbb{P}(X = n)$ pour $n \geqslant 2$, $t \in [0;1]$ et $\|u_n''\|_\infty = n(n-1)\mathbb{P}(X = n) = O(n^2\mathbb{P}(X = n))$. Si $X \in L^2$, alors la série $\sum n^2\mathbb{P}(X = n)$ et donc aussi $\sum n(n-1)\mathbb{P}(X = n)$ converge d'où la convergence normale donc uniforme de $\sum u_n''$ sur $[0;1]$ et on a la convergence simple des séries dérivées d'ordre inférieur (celle de $\sum u_n'$ résultant de $\|u_n'\|_\infty = O(n^2\mathbb{P}(X = n))$ ou par X d'espérance finie) d'où G_X est deux fois dérivable et

$$G_X''(1) = \sum_{n=0}^{+\infty} u_n''(1) = \sum_{n=0}^{+\infty} n(n-1)\mathbb{P}(X = n) = \mathbb{E}(X^2) - \mathbb{E}(X)$$

□

Remarque : On peut établir le résultat plus fort

$$X \in L^2 \iff G_X \text{ dérivable deux fois en } 1$$

en suivant une démarche identique à celle de la proposition 46.

Proposition 48. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé et $X_1, \ldots, X_n$ des variables aléatoires discrètes indépendantes à valeurs dans $\mathbb{N}$. On a*

$$\forall t \in [0;1] \qquad G_{S_n}(t) = \prod_{i=1}^{n} G_{X_i}(t) \quad \text{avec} \quad S_n = \sum_{i=1}^{n} X_i$$

Démonstration. Soit $t \in [0;1]$. On a

$$G_{S_n}(t) = \mathbb{E}\left(\prod_{i=1}^{n} t^{X_i}\right)$$

Le résultat suit d'après le corollaire 2. □

Exemple : Soit $Y \sim \mathscr{B}(n,p)$ et $X_1, \ldots, X_n$ indépendantes de loi $\mathscr{B}(p)$ avec n entier non nul et $p \in [0;1]$. On a $Y \sim S_n = \sum_{i=1}^{n} X_i$ ce qui explique le résultat observé :

$$G_Y = G_{S_n} = \prod_{i=1}^{n} G_{X_i} = G_{X_1}^n$$

V Lois usuelles

1 Loi géométrique

Définition 32. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire discrète. On dit que X suit une* loi géométrique *de paramètre $p \in]0;1[$ noté $X \sim \mathscr{G}(p)$ si $X(\Omega) = \mathbb{N}^*$ et*

$$\forall k \in \mathbb{N}^* \qquad \mathbb{P}(X = k) = p(1-p)^{k-1}$$

Remarque : Il s'agit bien d'une loi de probabilité puisque $p(1-p)^{k-1} \geqslant 0$ pour $k \in \mathbb{N}^*$ et, d'après le résultat sur les séries géométriques , $\sum p(1-p)^{k-1}$ converge puisque $1 - p \in]0;1[$ avec

$$\sum_{k=1}^{+\infty} p(1-p)^{k-1} = \frac{p}{1-(1-p)} = 1$$

Interprétation : Une loi géométrique modélise le rang du premier succès d'une suite illimitée d'épreuves de Bernoulli indépendantes de même paramètre p, par exemple un jeu de pile ou face où l'on s'arrête dès qu'on obtient pile. Pour $(\mathrm{X}_k)_{k\geqslant 1}$ suite de variables aléatoires indépendantes de loi $\mathscr{B}(p)$ et

$$\mathrm{X} = \inf\{k \in \mathbb{N}^* \mid \mathrm{X}_k = 1\}$$

on a $\mathrm{X}(\Omega) \subset \mathbb{N}^* \cup \{+\infty\}$ avec $\mathbb{P}(\mathrm{X} = +\infty) = \lim\limits_{n\to+\infty} \mathbb{P}\left(\bigcap\limits_{k=1}^{n} \{\mathrm{X}_k = 0\}\right) = 0$ d'où $\operatorname{supp}\mathrm{X} = \mathbb{N}^*$ et l'interprétation est valide.

Proposition 49. *Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé et $\mathrm{X} \sim \mathscr{G}(p)$ avec $p \in]\,0\,;1\,[$. On a $\mathrm{X} \in \mathrm{L}^2$ et*

$$\mathbb{E}(\mathrm{X}) = \frac{1}{p}, \qquad \mathbb{V}(\mathrm{X}) = \frac{1-p}{p^2}, \qquad \forall t \in [\,0\,;1\,] \qquad \mathrm{G}_\mathrm{X}(t) = \frac{pt}{1-t(1-p)}$$

Démonstration. Les séries entières $\sum k^2 x^k$ et $\sum x^k$ ont même rayon de convergence égal à 1 d'où $\mathrm{X} \in \mathrm{L}^2$. Par dérivation d'une série entière, on a pour $x \in]-1\,;1\,[$

$$\sum_{k=1}^{+\infty} kx^{k-1} = \frac{1}{(1-x)^2}, \qquad \sum_{k=2}^{+\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3}$$

Par linéarité dans l'intervalle de convergence

$$\sum_{k=1}^{+\infty} k(k-1)x^{k-1} + \sum_{k=1}^{+\infty} kx^{k-1} = \sum_{k=1}^{+\infty} k^2 x^{k-1} = \frac{2x}{(1-x)^3} + \frac{1}{(1-x)^2} = \frac{1+x}{(1-x)^3}$$

Prenant $x = 1-p$, on trouve

$$\mathbb{E}(\mathrm{X}) = p\sum_{k=1}^{+\infty} k(1-p)^{k-1} = \frac{p}{(1-(1-p))^2} = \frac{1}{p} \qquad \mathbb{E}(\mathrm{X}^2) = p\sum_{k=1}^{+\infty} k^2(1-p)^{k-1} = \frac{p(2-p)}{(1-(1-p))^3}$$

puis

$$\mathbb{V}(\mathrm{X}) = \mathbb{E}(\mathrm{X}^2) - \mathbb{E}(\mathrm{X})^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

Enfin, pour $t \in [\,0\,;1\,]$, il vient

$$\mathrm{G}_\mathrm{X}(t) = \sum_{k=1}^{+\infty} t^k \mathbb{P}(\mathrm{X} = k) = tp\sum_{k=1}^{+\infty} (t(1-p))^{k-1} = \frac{pt}{1-t(1-p)}$$

□

Exemple : Déterminer la complexité en moyenne du tri suivant d'argument `L` une liste de n nombres distincts :

```
def tri(L):
    while not est_triee(L):
        rd.shuffle(L)
```

où `est_triee` est une fonction qui renvoie `True` si la liste `L` est triée et `False` sinon. L'instruction `rd.shuffle` vient du module `numpy.random` importé en tant tant qu'alias `rd`. Elle agit en place en mélangeant la liste aléatoirement. Tant que la liste n'est pas triée, on la mélange. Notant $\sigma \in \mathrm{S}_n$ la permutation permettant de trier la liste `L` de taille n, on effectue des tirages uniformes dans S_n jusqu'à obtenir σ. Soit $(\mathrm{X}_k)_{k\geqslant 1}$ une suite de variables indépendantes de loi $\mathscr{U}_{\mathrm{S}_n}$ et

$$\mathrm{T} = \inf\{k \in \mathbb{N}^* \mid \mathrm{X}_k = \sigma\}$$

La variable aléatoire T modélise le nombre de mélanges effectués par la fonction `tri`. On a

$$\mathrm{T} \sim \mathscr{G}(p) \quad \text{avec} \quad p = \frac{1}{\operatorname{Card} \mathrm{S}_n}$$

Le nombre moyen de passages dans la boucle `while` est donné par $\mathbb{E}(\mathrm{T})$. Les complexités de `est_triee` et `rd.shuffle` sont en $\mathrm{O}(n)$. Ainsi, la complexité en moyenne est donnée par

$$\mathbb{E}(\mathrm{T})\,\mathrm{O}(n) = \frac{1}{p}\mathrm{O}(n) = n!\,\mathrm{O}(n)$$

ce qui est catastrophique.

2 Loi de Poisson

Définition 33. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* X *une variable aléatoire discrète. On dit que* X *suit une* loi de Poisson *de paramètre* $\lambda > 0$ *noté* $\mathrm{X} \sim \mathscr{P}(\lambda)$ *si* $\mathrm{X}(\Omega) = \mathbb{N}$ *et*

$$\forall k \in \mathbb{N} \qquad \mathbb{P}(\mathrm{X} = k) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$$

Remarque : Il s'agit bien d'une loi de probabilité puisque $\dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda} \geqslant 0$ pour $k \in \mathbb{N}$ et, d'après le résultat sur la série exponentielle, $\sum \dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$ converge avec

$$\sum_{k=0}^{+\infty} \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} = \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda} = 1$$

Proposition 50. *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé et* $\mathrm{X} \sim \mathscr{P}(\lambda)$ *avec* $\lambda > 0$. *On a* $\mathrm{X} \in \mathrm{L}^2$ *et*

$$\mathbb{E}(\mathrm{X}) = \lambda, \qquad \mathbb{V}(\mathrm{X}) = \lambda, \qquad \forall t \in [\,0\,;1\,] \qquad \mathrm{G_X}(t) = \mathrm{e}^{\lambda(t-1)}$$

Démonstration. Les séries entières $\sum k^2 \dfrac{x^k}{k!}$ et $\sum \dfrac{x^k}{k!}$ ont même rayon de convergence égal à $+\infty$ d'où $\mathrm{X} \in \mathrm{L}^2$. Par dérivation d'une série entière, on a pour x réel

$$\sum_{k=1}^{+\infty} k\frac{x^{k-1}}{k!} = \mathrm{e}^x, \qquad \sum_{k=2}^{+\infty} k(k-1)\frac{x^{k-2}}{k!} = \mathrm{e}^x$$

Par linéarité car convergence

$$\sum_{k=0}^{+\infty} k(k-1)\frac{x^k}{k!} + \sum_{k=0}^{+\infty} k\frac{x^k}{k!} = \sum_{k=0}^{+\infty} k^2\frac{x^k}{k!} = x^2\mathrm{e}^x + x\mathrm{e}^x = x(1+x)\mathrm{e}^x$$

Prenant $x = \lambda$, on trouve $\qquad \mathbb{E}(\mathrm{X}) = \displaystyle\sum_{k=0}^{+\infty} k\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} = \lambda\mathrm{e}^{\lambda}\mathrm{e}^{-\lambda} = \lambda$

et $\qquad \mathbb{V}(\mathrm{X}) = \mathbb{E}(\mathrm{X}^2) - \mathbb{E}(\mathrm{X})^2 = \displaystyle\sum_{k=0}^{+\infty} k^2\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} - \mathbb{E}(\mathrm{X})^2 = \lambda(1+\lambda)\mathrm{e}^{\lambda}\mathrm{e}^{-\lambda} - \lambda^2 = \lambda$

Enfin, pour $t \in [\,0\,;1\,]$, il vient

$$\mathrm{G_X}(t) = \sum_{k=0}^{+\infty} t^k\mathbb{P}(\mathrm{X} = k) = \sum_{k=0}^{+\infty} \frac{(\lambda t)^k}{k!}\mathrm{e}^{-\lambda} = \mathrm{e}^{\lambda(t-1)}$$

□

VI Résultats asymptotiques

1 Approximation de la loi binomiale par la loi de Poisson

Théorème 17 (Loi des événements rares). *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé. Soit* $(\mathrm{X}_n)_{n\geqslant 1}$ *une suite de variables aléatoires discrètes avec* $\mathrm{X}_n \sim \mathscr{B}(n, p_n)$ *et* $np_n \xrightarrow[n\to\infty]{} \lambda > 0$*, alors*

$$\forall k \in \mathbb{N} \qquad \mathbb{P}(\mathrm{X}_n = k) \xrightarrow[n\to\infty]{} \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$$

Vocabulaire : On dit que la suite $(\mathrm{X}_n)_{n\geqslant 1}$ *converge en loi* vers la loi de Poisson $\mathscr{P}(\lambda)$ (notion de convergence hors-programme).

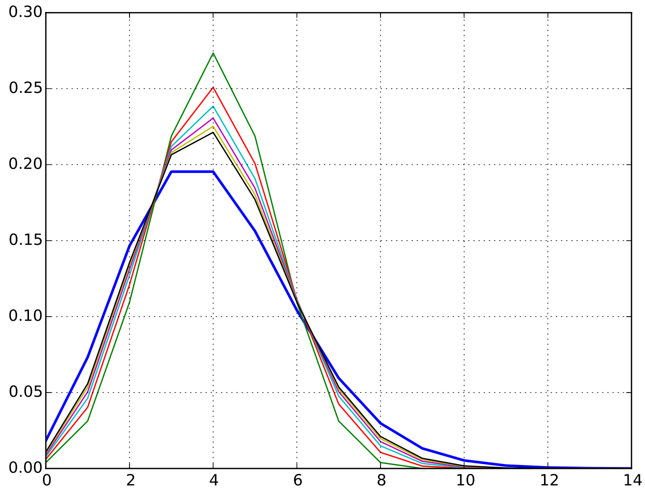

FIGURE 3 – Approximation de la loi binomiale par la loi de Poisson

Démonstration. On a $np_n = \lambda + \mathrm{o}(1)$ d'où $p_n \xrightarrow[n\to\infty]{} 0$ puis pour $n \geqslant k$

$$\begin{aligned}
\mathbb{P}(\mathrm{X}_n = k) &= \tbinom{n}{k} p_n^k (1-p_n)^{n-k} \\
&= \frac{1}{k!}(np_n)^k \underbrace{\left(1-\frac{1}{n}\right)\times\ldots\times\left(1-\frac{k-1}{n}\right)}_{=1+\mathrm{o}(1)} \exp\left[n\ln(1-p_n)\right] \underbrace{\frac{1}{(1-p_n)^k}}_{=1+\mathrm{o}(1)} \\
&= \frac{1}{k!}(\lambda+\mathrm{o}(1))^k \exp\left[n(-p_n+\mathrm{o}(p_n))\right](1+\mathrm{o}(1)) \\
\mathbb{P}(\mathrm{X}_n = k) &= \frac{1}{k!}(\lambda+\mathrm{o}(1))^k \exp\left[-\lambda+\mathrm{o}(1)\right](1+\mathrm{o}(1)) \xrightarrow[n\to\infty]{} \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}
\end{aligned}$$

□

Interprétation : Sur un intervalle de temps $[\,0\,;\mathrm{T}\,]$ qu'on subdivise en $0 < \Delta\mathrm{T} < 2\Delta\mathrm{T} < \ldots < n\Delta\mathrm{T}$, on observe sur chaque sous-intervalle des résultats d'expériences aléatoires succès/échec. Le succès a lieu proportionnellement à la durée du sous intervalle à savoir $\mu\Delta\mathrm{T}$. Ainsi, sur $[\,0\,;\mathrm{T}\,]$, le nombre total de succès suit une loi $\mathscr{B}(n, p_n)$ avec $p_n = \mu\Delta\mathrm{T}$ où $n\Delta\mathrm{T} = \mathrm{T}$. Faire tendre $n \to +\infty$ équivaut à faire tendre $\Delta\mathrm{T} \to 0$, autrement dit une subdivision de plus en plus fine. La probabilité de succès $\mu\Delta\mathrm{T}$ tend vers zéro d'où cette interprétation de la loi de Poisson comme *loi des événements rares*.

2 Loi faible des grands nombres

On note i.i.d. pour : indépendantes identiquement distribuées, c'est-à-dire indépendantes et de même loi.

Théorème 18 (Loi faible des grands nombres). *Soit* $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé. Soit* $(\mathrm{X}_n)_{n\geqslant 1}$ *une suite de variables aléatoires i.i.d. dans* L^2*. Notant* $\mathrm{S}_n = \sum_{i=1}^{n} \mathrm{X}_i$*,* $m = \mathbb{E}(\mathrm{X}_1)$*, on a*

$$\forall \varepsilon > 0 \qquad \mathbb{P}\left(\left|\frac{\mathrm{S}_n}{n} - m\right| \geqslant \varepsilon\right) \xrightarrow[n\to\infty]{} 0$$

Vocabulaire : On dit que la suite $\left(\frac{1}{n}\sum_{i=1}^{n}\mathrm{X}_i\right)_{n\geqslant 1}$ *converge en probabilité* vers m (notion de convergence hors-programme).

Remarque : La déclinaison *forte* de ce résultat (convergence presque sûre) est le fondement des *méthodes de Monte-Carlo.*

Démonstration. Notons $\sigma = \sigma(\mathrm{X}_1)$. Soit $\varepsilon > 0$ et n entier non nul. D'après l'inégalité de Bienaymé-Tchebychev, on a

$$\mathbb{P}\left(\left|\frac{\mathrm{S}_n}{n} - m\right| \geqslant \varepsilon\right) \leqslant \frac{\mathbb{V}(\mathrm{S}_n)}{n^2\varepsilon^2}$$

puis, par indépendance deux à deux des X_i

$$\mathbb{V}(\mathrm{S}_n) = \mathbb{V}(\sum_{i=1}^{n}\mathrm{X}_i) = \sum_{i=1}^{n}\mathbb{V}(\mathrm{X}_i) = n\sigma^2$$

Ainsi

$$\mathbb{P}\left(\left|\frac{\mathrm{S}_n}{n} - m\right| \geqslant \varepsilon\right) \leqslant \frac{\sigma^2}{n\varepsilon^2}$$

et le résultat suit. □

Remarque : On n'utilise en réalité que la décorrélation des variables aléatoires X_i.

Commentaire : La démonstration fournit une inégalité permettant d'exhiber ce qu'on appelle en statistique un *intervalle de confiance.* Si $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ est un échantillon de mesures et m une grandeur d'intérêt (durée de vie d'un produit par exemple), on veut pouvoir estimer cette grandeur. En utilisant l'inclusion

$$\left]\frac{\mathrm{S}_n}{n} - \varepsilon\,; \frac{\mathrm{S}_n}{n} + \varepsilon\right[\subset \left[\frac{\mathrm{S}_n}{n} - \varepsilon\,; \frac{\mathrm{S}_n}{n} + \varepsilon\right]$$

il vient par complémentation

$$\mathbb{P}\left(m \in \left[\frac{\mathrm{S}_n}{n} - \varepsilon\,; \frac{\mathrm{S}_n}{n} + \varepsilon\right]\right) \geqslant 1 - \frac{\sigma^2}{n\varepsilon^2}$$

Ainsi, la grandeur d'intérêt m est localisée dans un intervalle qui est fonction de l'échantillon $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ avec un niveau de confiance au moins $1 - \frac{\sigma^2}{n\varepsilon^2}$. Pour un niveau de confiance fixé (proche de 1 idéalement), on est face à un compromis : un choix de ε petit assure un petit intervalle de confiance mais il faut choisir un grand échantillon ce qui est coûteux. On perçoit l'intérêt d'avoir ici une minoration aussi fine que possible pour optimiser ce compromis.

Illustration :

On représente pour des lois uniformes $\mathscr{U}_{[\![7\,;13]\!]}$ et $\mathscr{U}_{[\![1\,;19]\!]}$ le tracé de plusieurs réalisations des moyennes empiriques, à savoir les termes de la suite $\left(\dfrac{S_n(\omega)}{n}\right)_{n\geqslant 1}$ pour différentes réalisations de ω.

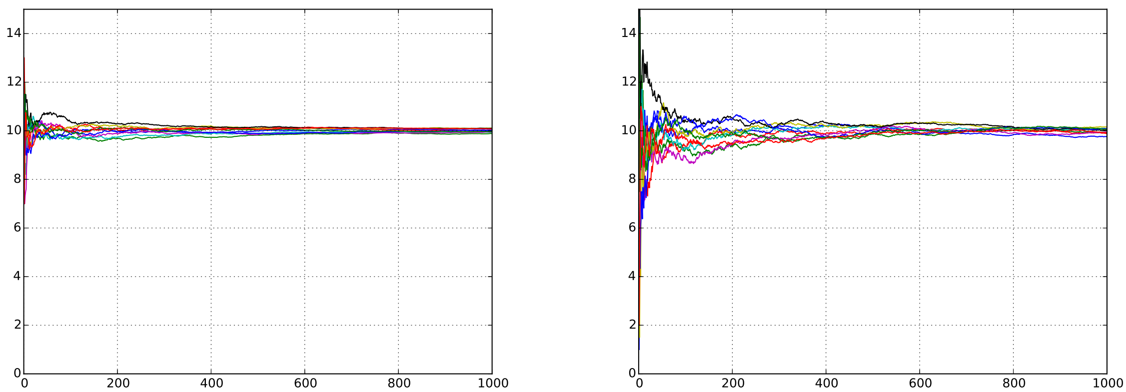

Figure 4 – Tracé des moyennes empiriques

On rappelle que pour $X \sim \mathscr{U}_{[\![a\,;b]\!]}$, on a

$$\mathbb{E}(X) = \frac{a+b}{2} \qquad \mathbb{V}(X) = \frac{(b-a)(2+b-a)}{12}$$

Dans la simulation illustrée ci-avant, les lois ont même espérance mais des variances différentes ce qui apparaît clairement dans la dispersion des tracés : plus la variance est grande, plus la dispersion des trajectoires est élevée.