# Chapitre 9

# Probabilités et variables aléatoires discrètes

### Révisions MP2I

Revoir les chapitres 43, 44 et 45.

La notion de probabilité a été introduite en MP2I dans le cas d'un univers $\Omega$ fini. Toute partie de $\Omega$ est alors considérée comme un événement, et une probabilité sur $\Omega$ définie comme une application $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$ vérifiant :

- $\mathbb{P}(\Omega) = 1$
- $\forall A, B \in \mathcal{P}(\Omega),\ A \cap B = \varnothing \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Cette propriété d'*additivité*, fondamentale, s'étend facilement par récurrence à toute famille *finie* $(A_k)_{1 \leqslant k \leqslant n}$ d'événements deux à deux incompatibles.

L'hypothèse d'un univers fini est cependant trop restrictive : impossible de modéliser correctement par exemple une variable aléatoire pouvant prendre une infinité de valeurs avec une probabilité non nulle. Sans hypothèse sur la "taille" de $\Omega$, on est conduit naturellement à vouloir définir une probabilité vérifiant une propriété d'additivité plus générale, pouvant concerner une famille infinie $(A_i)_{i \in I}$ d'événements. Cela n'est malheureusement pas toujours possible de le faire de façon satisfaisante sur l'ensemble $\mathcal{P}(\Omega)$ des parties de $\Omega$. On doit donc définir la notion générale de probabilité dans un contexte où on peut se restreindre à un sous-ensemble de $\mathcal{P}(\Omega)$ vérifiant certaines propriétés de stabilité par opérations ensemblistes.

Un premier objectif de ce chapitre sera de présenter ainsi la définition générale d'*espace probabilisé*, et de *variable aléatoire discrète* : c'est l'objet de la section **2**. Avant cela, il va nous falloir étudier les notions de *dénombrabilité* et de *sommabilité* (section **1**), qui interviennent dans ces définitions. Une fois tout le cadre théorique posé, nous pourrons aborder les notions de conditionnement et d'indépendance (section **3**) ainsi que celles d'espérance et de variance d'une variable aléatoire discrète (section **4**).

## 1 Prerequis : dénombrabilité et sommabilité

### 1.1 Ensembles dénombrables

**Définition 1.** Soient deux ensembles quelconques $E$ et $F$. On dit que $E$ est *équipotent* à $F$ lorsqu'il existe une application $\varphi : E \to F$ bijective.

**Remarques :**

- On pourra noter $E \simeq F$ lorsque $E$ et $F$ sont équipotents. Attention ! Cette notation $\simeq$ entre deux ensembles $E$ et $F$ signifie généralement qu'ils sont *isomorphes*, et il faut que le contexte indique clairement les structures algébriques concernées (groupe, anneaux, espace vectoriel, etc ..). Ici, pour des ensembles sans structure, il n'y a rien d'autre à vérifier que la bijectivité !
- $\simeq$ vérifie les propriétés d'une relation d'équivalence, mais affirmer qu'il en s'agit bien d'une pose un problème conceptuel qui sort du cadre du programme : on ne peut pas considérer l'ensemble de tous les ensembles ...
- Rappelons qu'un ensemble $E$ est fini lorsqu'il est vide ou lorsqu'il existe $n \in \mathbb{N}$ et une application $\varphi : [\![1, n]\!] \to E$ bijective, autrement dit lorsque $E \simeq [\![1, n]\!]$. $n$ est alors appelé *cardinal* de $E$ (on montre qu'un tel $n$ est unique).

**Définition 2.** On dit qu'un ensemble $E$ est *dénombrable* lorsqu'il existe une application $\varphi : \mathbb{N} \to E$ bijective.

**Remarques :**
- $E$ est donc dénombrable lorsque $E \simeq \mathbb{N}$.
- Ainsi un ensemble dénombrable est infini, mais un infini que l'on peut "compter", ou "enumérer".
- $E$ est dénombrable si, et seulement si, on peut le décrire en *extension* sous la forme

$$E = \{x_n, n \in \mathbb{N}\}, \qquad \text{avec les } x_n \text{ deux à deux distincts}$$

(si on autorise des éléments égaux, cela englobe alors aussi les ensembles finis).

**Exemples :**
- $2\mathbb{N} = \{2n, n \in \mathbb{N}\}$, l'ensemble des entiers naturels pairs, est dénombrable.
- $\{n^2, n \in \mathbb{N}\}$, ensemble des carrés naturels, est dénombrable (ne pas le noter $\mathbb{N}^2$ !)
- L'ensemble des nombres premiers étant infini et inclu dans $\mathbb{N}$, il est dénombrable, grâce à la proposition suivante.

**Proposition 1.**

***a)*** *Toute partie infinie de $\mathbb{N}$ est dénombrable.*

***b)*** *Un ensemble $E$ est fini ou dénombrable si, et seulement si, il est en bijection avec une partie de $\mathbb{N}$.*

**Remarque :** Un ensemble fini ou dénombrable est dit *au plus dénombrable*.

De nombreux ensembles apparemment "bien plus gros" que $\mathbb{N}$ sont en fait dénombrables, car la propriété d'être dénombrable reste stable par diverses opérations ensemblistes.

**Proposition 2.**

***a)*** *$\mathbb{N} \times \mathbb{N}$ est dénombrable.*

***b)*** *Pour $p \in \mathbb{N}^*$, si $E_1, \dots, E_p$ sont des ensembles dénombrables, alors $E_1 \times \cdots \times E_p$ est dénombrable.*

***c)*** *Si $(A_i)_{i\in I}$ est une famille dénombrable d'ensembles dénombrables, alors $\bigcup_{i\in I} A_i$ est dénombrable.*

**Remarque :** Les deux résultats **b)** et **c)** précédents sont encore vrais en remplaçant "dénombrable" par "au plus dénombrable".

**Corollaire 1.**
- *$\mathbb{Z}$ est dénombrable*
- *$\mathbb{Q}$ est dénombrable*
- *$\mathbb{Q}[X]$ est dénombrable*

**Théorème 1.** *$\mathbb{R}$ n'est pas dénombrable.*

**Remarques :**
- On peut montrer que $\mathbb{R}$ est équipotent à $\mathcal{P}(\mathbb{N})$, qui n'est pas équipotent à $N$ grâce au théorème de Cantor : pour tout ensemble $E$, il n'existe pas de bijection de $E$ sur $\mathcal{P}(E)$.
- L'hypothèse du continu affirme qu'il n'y a pas d'*intermédiaire* entre $\mathbb{N}$ et $\mathbb{R}$ : Si $E$ est une partie infinie de $\mathbb{R}$ alors $E \simeq \mathbb{N}$ ou $E \simeq \mathbb{R}$.

## 1.2 Sommation des familles positives

Nous allons maintenant présenter comment généraliser la notion de somme d'une famille finie $(u_i)_{1\leqslant i\leqslant n}$ de nombres réels ou complexes, à une famille quelconque $(u_i)_{i\in I}$. Vous vous dites que cela a déjà été fait via la notion de série $\sum u_n$. C'est vrai, mais cela ne concerne que les familles indexées par $\mathbb{N}$ (autrement dit les suites), et utilise explicitement l'ordre usuel de $\mathbb{N}$ pour définir une somme comme un passage à la limite. Ce que nous allons voir maintenant s'affranchit de ces contraintes. Nous commençons par le cas des familles positives, pour lesquelles aucune difficulté n'apparait, si on s'autorise à ajouter $+\infty$ comme valeur possible.

**Définition 3.**

**a)** Pour tout $a \in [0, +\infty] = \mathbb{R}_+ \cup \{+\infty\}$, on définit :
- $a + (+\infty) = +\infty + a = +\infty$
- $a \times (+\infty) = +\infty \times a = \begin{cases} +\infty \text{ si } a \neq 0 \\ 0 \text{ si } a = 0 \end{cases}$

**b)** On étend la relation d'ordre $\leqslant$ à $[0, +\infty]$ en imposant $a \leqslant +\infty$ pour tout $a \in [0, +\infty]$ ($a < +\infty$ si $a \neq +\infty$).

**c)** Pour toute partie $A$ non vide de $[0, +\infty]$, On note $\sup(A) = +\infty$ si $A$ n'a pas de majorant dans $\mathbb{R}_+$.

**Proposition 3.** *Avec les définitions précédentes :*
- $\leqslant$ *définit un ordre total sur* $[0,+\infty]$*, et* $+\infty$ *est le plus grand élément.*
- $+$ *définit une loi de composition interne associative, commutative, et compatibe avec l'ordre* $\leqslant$*.*
- *Toute partie non vide de* $[0,+\infty]$ *admet une borne supérieure dans* $[0,+\infty]$*.*

**Définition 4.** Soit $(u_i)_{i\in I}$ une famille de $[0,+\infty]$, avec $I$ un ensemble quelconque. On définit

$$\sum_{i\in I} u_i = \sup\left\{ \sum_{i\in F} u_i \ : \quad F\subset I,\ F \text{ fini} \right\}$$

Lorsque $\sum_{i\in I} u_i < +\infty$, on dit que la famille $(u_i)_{i\in I}$ est *sommable.*

Avec cette définition, et les propriétés de $[0,+\infty]$, toute famille positive admet une "somme", qui vaut éventuellement $+\infty$ lorsque la famille n'est pas sommable. Le résultat suivant généralise la propriété de commutativité.

**Proposition 4.** *Soit* $(u_i)_{i\in I}$ *une famille de* $[0,+\infty]$*. Pour toute bijection* $\sigma : I\to I$*, on a :*

$$\sum_{i\in I} u_i = \sum_{i\in I} u_{\sigma(i)}$$

Dans le cas $I=\mathbb{N}$, on peut voir que cette définition est bien compatible avec celle de la somme d'une série, avec la convention que la somme d'une série positive divergente est $+\infty$ :

**Proposition 5.** *Si* $\sum u_n$ *est une série positive, alors*

$$\sum_{n=0}^{+\infty} u_n = \sum_{n\in\mathbb{N}} u_n$$

Le propriétés classiques de linéarité et de croissance de la somme se généralisent parfaitement avec les familles de $[0,+\infty]$ :

**Proposition 6.** *Soient* $(u_i)_{i\in I}$ *et* $(v_i)_{i\in I}$ *deux familles de* $[0,+\infty]$*, et* $\lambda\in[0,+\infty]$*. Alors :*
- *Si* $u_i\leqslant v_i$ *pour tout* $i\in I$*, on a* $\sum_{i\in I} u_i \leqslant \sum_{i\in I} v_i$*.*
- $\sum_{i\in I} \lambda u_i = \lambda \sum_{i\in I} u_i$*.*
- $\sum_{i\in I}(u_i+v_i) = \sum_{i\in I} u_i + \sum_{i\in I} v_i$

Dans le cas des familles de $[0,+\infty]$, nous pouvons aussi écrire des propriétés de comparaison et de calcul avec des sous-familles, et il est utile d'introduire la notion de fonction indicatrice afin de faciliter les raisonnements.

**Définition 5.** Pour $A\subset I$, on note $\mathbb{1}_A : i\mapsto \begin{cases} 1 \text{ si } i\in A \\ 0 \text{ si } i\notin A \end{cases}$. On l'appelle *fonction indicatrice de* $A$ dans $I$.

**Proposition 7.** *Soit* $(u_i)_{i\in I}$ *une famille de* $[0,+\infty]$ *et* $A\subset I$*. Alors*

$$\sum_{i\in A} u_i = \sum_{i\in I} \mathbb{1}_A(i) u_i$$

**Proposition 8.** *Soient* $(u_i)_{i\in I}$ *une famille de* $[0,+\infty]$*, et* $A,B$ *deux parties non vides de* $I$*.*

***a)*** *Si* $A\subset B$*,* $\sum_{i\in A} u_i \leqslant \sum_{i\in B} u_i$*.*

***b)*** *Si* $A\cap B=\varnothing$*,* $\sum_{i\in A\sqcup B} u_i = \sum_{i\in A} u_i + \sum_{i\in B} u_i$

**Remarque :** On peut généraliser par récurrence le second point à toute famille finie $(A_j)_{1\leqslant j\leqslant n}$ de parties de $I$ deux à deux disjointes : si $A = \bigsqcup_{j=1}^{n} A_j$, on a $\sum_{i\in A} u_i = \sum_{j=1}^{n} \sum_{i\in A_j} u_i$. Mais on va voir maintenant un énoncé encore plus général.

**Théorème 2.** ***(de sommation par paquets, cas positif)***
*Soit $(u_i)_{i\in I}$ une famille de $[0,+\infty]$ et $(A_j)_{j\in J}$ une* partition *de $I$. Alors*

$$\sum_{i\in I} u_i = \sum_{j\in J} \sum_{i\in A_j} u_i$$

**Remarque :** On utilise typiquement ce résultat avec $I = \mathbb{N}^2$ et $J = \mathbb{N}$ dans les deux cas suivant :
- $\forall j \in \mathbb{N},\ A_j = \{j\} \times \mathbb{N}$ ou $\forall j \in \mathbb{N},\ A_j = \mathbb{N} \times \{j\}$. L'égalité des sommes suivant ces deux partitions donnent alors le théorème de Fubini (voir plus loin).
- $\forall j \in \mathbb{N},\ A_j = \{(n,p) \in \mathbb{N}^2 \,|\, n+p = j\}$.

Représentez graphiquement ces partitions de $\mathbb{N}^2$ pour bien comprendre de quoi il s'agit.

**Exercice 1.** La famille $\left(\dfrac{1}{nk(n+k)}\right)_{(n,k)\in(\mathbb{N}^*)^2}$ est-elle sommable ?

Comme dans l'exercice prédédent, une situation classique de sommation concerne une famille doublement indexée. Le théorème de sommation par paquets permet alors de se ramener à une somme double et il y a un théorème qui explicite le fait que l'ordre de ces deux sommes n'importe pas.

**Théorème 3.** ***(de Fubini, cas positif)***
*Soit deux ensembles $I$ et $J$, et soit $(u_{i,j})_{(i,j)\in I\times J}$ une famille de $[0,+\infty]$. ALors :*

$$\sum_{i\in I}\sum_{j\in J} u_{i,j} = \sum_{j\in J}\sum_{i\in I} u_{i,j}$$

**Remarque :** Ces deux sommes sont égales à $\sum_{(i,j)\in I\times J} u_{i,j}$.

**Exercice 2.** étudier la nature de la série $\sum R_n$, avec $R_n = \sum_{k=n+1}^{+\infty} \dfrac{1}{k!}$.

## 1.3 Sommabilité d'une famille réelle ou complexe

Vous savez peut-être que pour une série $\sum u_n$ semi-convergente, comme par exemple la série harmonique alternée $\sum (-1)^n \dfrac{1}{n+1}$, la valeur de la somme $\sum_{n=0}^{+\infty} u_n$ est intimement liée à l'ordre dans lequel les termes sont additionnés, via le passage à la limite des sommes partielles : en changeant l'ordre, on peut obtenir n'importe quel limite ! Cette difficulté disparait lorsqu'on ne considère que des séries absolument convergentes, et c'est seulement en généralisant cette situation qu'on va pouvoir définir proprement et de façon unique la somme d'une famille quelconque.

**Définition 6.** On dit qu'une famille $(u_i)_{i\in I}$ de $\mathbb{K}$ est *sommable* lorsque $\sum_{i\in I} |u_i| < +\infty$.
On note $\ell^1(I)$ ou $\ell(I)$ l'ensemble des familles sommables indexées par $I$.

**Remarques :**
- Dans le cas $I = \mathbb{N}$, on peut bien relier à la notion de série : $(u_n)_n$ est sommable si, et seulement si, la série $\sum u_n$ est absolument convergente.
- En général, l'ensemble $I$ est dénombrable (par exemple $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{N}\times\mathbb{N}$).

**Exercice 3.** Montrer que le *support* $J$ d'une famille sommable $(u_i)_{i\in I}$ de nombres complexes, c'est-à-dire l'ensemble $\{i \in I \,|\, u_i \neq 0\}$, est au plus dénombrable.

Pour une famille sommable $(u_i)_{i\in I}$ de $\mathbb{K}$, on va pouvoir attribuer une valeur dans $\mathbb{K}$ à la somme $\sum_{i\in I} u_i$ et retrouver les mêmes propriétés que dans le cas des familles positives. Pour le cas réel, l'idée est de séparer les termes positifs et négatifs. Pour le cas complexe, on separera alors les parties réelles et imaginaires de chaque terme. On peut en fait tout faire d'un coup, et ce que nous allons détailler maintenant.

**Définition 7.** Pour tout $x \in \mathbb{R}$, on note $x^+ = \max(x, 0)$ et $x^- = \max(-x, 0)$.
$x^+$ et $x^-$ s'appelle respectivement partie positive et négative de $x$.

**Proposition 9.** *Pour tout $x \in \mathbb{R}$, on a :*

$$x^+ \geqslant 0 \quad , \quad x^- \geqslant 0 \quad , \quad x = x^+ - x^- \quad , \quad |x| = x^+ + x^-.$$

**Définition 8.** Soit $(u_i)_{i\in I}$ une famille sommable réelle ou complexe. On définit la somme de cette famille par :

$$\sum_{i\in I} u_i = \left(\sum_{i\in I} \mathrm{Re}(u_i)^+ - \sum_{i\in I} \mathrm{Re}(u_i)^-\right) + i\left(\sum_{i\in I} \mathrm{Im}(u_i)^+ - \sum_{i\in I} \mathrm{Im}(u_i)^-\right)$$

**Remarque :** À moins qu'il ne s'agisse d'une famille réelle positive, on ne pourra parler de la somme d'une famille $(u_i)_{i\in I}$ réelle ou complexe que si cette famille est sommable.

Cette définition de la somme d'une famille sommable est bien cohérente avec celle de la somme d'une série absolument convergente :

**Proposition 10.** *Une suite $(u_n)_{n\in\mathbb{N}}$ réelle ou complexe est sommable si, et seulement si, $\sum u_n$ est une série absolument convergente et on a alors :*

$$\sum_{n\in\mathbb{N}} u_n = \sum_{n=0}^{+\infty} u_n$$

**Remarque :** $\ell^1(\mathbb{N})$ désigne donc l'ensemble des suites qui sont le terme général d'une série absolument convergente.

Lorsqu'une famille $(u_i)_{i\in I}$ est sommable, la valeur de la somme peut être arbitrairement approchée par la somme d'une sous-famille finie. C'est ce qui va permettre de généraliser tous les résultats connus pour les sommes finies, et déjà vus dans le cas des familles positives.

**Proposition 11.** *Soit $(u_i)_{i\in I}$ une famille sommable et $\varepsilon > 0$. Alors il existe une partie finie $F \subset I$ telle que :*

$$\left|\sum_{i\in I} u_i - \sum_{i\in F} u_i\right| \leqslant \varepsilon$$

Voyons déjà comment on retrouve une situation générale de commutativité. Ce qui suit ne peut pas fonctionner pour une série semi-convergente !

**Proposition 12.** *Si $(u_i)_{i\in I}$ est une famille sommable et $\sigma$ une permutation de $I$ alors $(u_{\sigma(i)})_{i\in I}$ est sommable et*

$$\sum_{i\in I} u_{\sigma(i)} = \sum_{i\in I} u_i$$

Poursuivons avec les propriétés de linéarité et la généralisation de l'inégalité triangulaire.

**Proposition 13.**

***a)*** *Si $(u_i)_{i\in I}$ et $(v_i)_{i\in I}$ sont deux familles sommables et $\lambda \in \mathbb{K}$, alors $(u_i + \lambda v_i)_{i\in I}$ est sommable et :*

$$\sum_{i\in I}(u_i + \lambda v_i) = \sum_{i\in I} u_i + \lambda \sum_{i\in I} v_i$$

***b)*** *Si $(u_i)_{i\in I}$ est une famille sommable, alors*

$$\left|\sum_{i\in I} u_i\right| \leqslant \sum_{i\in I} |u_i|$$

Le corollaire suivant reformule ce qui précède en termes savants ...

**Corollaire 2.** *$\ell^1(I)$ est un sous-espace vectoriel de $\mathbb{K}^I$, et l'application $u \mapsto \sum_{i\in I} u_i$ est une forme linéaire continue sur l'espace vectoriel normé $(\ell^1(I), \|\cdot\|_1)$.*

Nous allons enfin donner une version du théorème de sommation par paquets et du théorème de Fubini dans le cas de familles réelles ou complexes, en tenant compte de la propriété de sommabilité.

**Théorème 4.** ***(de sommation par paquets)***
*Soit $(u_i)_{i\in I}$ une famille réelle ou complexe et $(A_j)_{j\in J}$ une* partition *de $I$. Alors les deux assertions suivantes sont équivalentes :*

*(i) la famille $(u_i)_{i\in I}$ est sommable.*

*(ii) pour tout $j \in J$, $(u_i)_{i\in A_j}$ est sommable, et la famille $\left(\sum_{i\in A_j} |u_i|\right)_{j\in J}$ est sommable.*

*Ces conditions étant vérifiées, on a alors :*

$$\sum_{i\in I} u_i = \sum_{j\in J} \sum_{i\in A_j} u_i$$

**Théorème 5.** ***(de Fubini)***
*Soit deux ensembles $I$ et $J$, et soit $(u_{i,j})_{(i,j)\in I\times J}$ une famille réelle ou complexe. Les trois assertions suivantes sont équivalentes :*

*(i) $(u_{i,j})_{(i,j)\in I\times J}$ est sommable*

*(ii) Pour tout $i \in I$, la famille $(u_{i,j})_{j\in J}$ est sommable, et la famille $\left(\sum_{j\in J} |u_{i,j}|\right)_{i\in I}$ est sommable.*

*(iii) Pour tout $j \in J$, la famille $(u_{i,j})_{i\in I}$ est sommable, et la famille $\left(\sum_{i\in I} |u_{i,j}|\right)_{j\in J}$ est sommable.*

*Ces conditions étant vérifiées, on a :*

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

Avec deux familles $(u_i)_{i\in I}$ et $(v_j)_{j\in J}$ on peut constituer la famille double $(u_i v_j)_{(i,j)\in I\times J}$. En cas de sommabilité, le théorème de Fubini permet alors d'écrire la somme de cette famille comme le produit des deux sommes, et de généraliser ainsi la propriété de double distributivité, lorsqu'on développe le produit de deux sommes.

**Proposition 14.** *Soient $(u_i)_{i\in I}$ et $(v_j)_{j\in J}$ deux familles sommables. Alors $(u_i v_j)_{(i,j)\in I\times J}$ est sommable et*

$$\left(\sum_{i\in I} u_i\right)\left(\sum_{j\in J} v_j\right) = \sum_{(i,j)\in I\times J} u_i v_j$$

**Remarque :** Ce résultat se généralise à tout produit fini de sommes de familles sommables : si pour tout $k \in [\![1, p]\!]$ ; $(u_{k,i})_{i\in I_k}$ est une famille sommable, alors $(u_{1,i_1}\cdots u_{p,i_p})_{(i_1,\dots,i_p)\in I_1\times\cdots\times I_p}$ est une famille sommable et :

$$\prod_{k=1}^{p}\left(\sum_{i_k\in I_k} u_{k,i_k}\right) = \sum_{(i_1,\dots,i_k)\in I_1\times\cdots\times I_k} u_{1,i_1}\cdots u_{p,i_p}$$

Comme on l'a vu, une façon classique de partitionner $\mathbb{N}^2$ consiste à écrire $\mathbb{N}^2 = \bigsqcup_{n\in\mathbb{N}}\{(k, n-k), 0 \leqslant k \leqslant n\}$. Cette partition a l'avantage de n'être constituée que de parties finies, de sorte que la somme d'une famille sommable $(u_{n,p})_{(n,p)\in\mathbb{N}^2}$ peut s'écrire comme une seule somme indexée par $\mathbb{N}$ de sommes finies. Dans le cas particulier d'une famille $(u_n v_p)_{(n,p)\in\mathbb{N}^2}$, on retrouve la notion de produit de Cauchy de deux séries et on est en mesure de démontrer facilement le résultat suivant :

**Proposition 15.** *Soient $\sum u_n$ et $\sum v_n$ deux séries réelles ou complexes absolument convergentes. Alors leur produit de Cauchy $\sum w_n$ est une série absolument convergente, et*

$$\sum_{n=0}^{+\infty} w_n = \left(\sum_{n=0}^{+\infty} u_n\right)\left(\sum_{n=0}^{+\infty} v_n\right) \qquad \text{avec} \quad w_n = \sum_{k=0}^{n} u_k v_{n-k}$$

# 2 Cadre théorique

Nous pouvons maintenant aborder pleinement la construction du cadre théorique permettant de modéliser des situations aléatoires et de calculer des probabilités. La finalité est la définition de la notion de variable aléatoire discrète et de loi de probabilité.

## 2.1 Tribu sur un ensemble

**Définition 9.** Soit $\Omega$ un ensemble quelconque. On appelle *tribu* sur $\Omega$, une partie $\mathcal{A}$ de $\mathcal{P}(\Omega)$ telle que :

**a)** $\Omega \in \mathcal{A}$,

**b)** Pour tout $A \in \mathcal{A}$, $\overline{A} = \Omega \backslash A \in \mathcal{A}$.

**c)** Pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{A}$, la réunion $\bigcup_{n\in\mathbb{N}} A_n$ est également un élément de $\mathcal{A}$.

Si $\mathcal{A}$ est une tribu sur $\Omega$, on dit que $(\Omega, \mathcal{A})$ (ou simplement $\Omega$ lorsque le contexte indique clairement la tribu utilisée) est un *espace probabilisable*.

**Remarques :**

- On peut résumer tout cela en disant qu'une tribu sur $\Omega$ est un ensemble de parties de $\Omega$ contenant $\Omega$, stable par passage au complémentaire et par réunion dénombrable.
- Pour une telle tribu $\mathcal{A}$, on doit aussi avoir $\varnothing = \overline{\Omega} \in \mathcal{A}$.
- On peut voir alors que la propriété de stabilité par réunion dénombrable reste vraie pour une réunion *finie* $\bigcup_{n\in[\![1,N]\!]} A_n$. Il suffit en effet de poser $A_n = \varnothing$ pour $n \geqslant N+1$. On peut donc dire qu'une tribu est stable par réunion au plus dénombrable.

**Exemples :**

- La tribu triviale $\{\varnothing, \Omega\}$
- La tribu pleine $\mathcal{P}(\Omega)$ : c'est le plus souvent celle-ci qui est utilisée lorsque $\Omega$ est au plus dénombrable. Lorsque $\Omega$ n'est pas dénombrable, on ne peut pas l'utiliser en général.
- Pour $A \subset \Omega$, $\{\Omega, \varnothing, A, \overline{A}\}$ : c'est la plus *petite* tribu (au sens qu'elle est contenue dans toutes les autres) qui contient $A$, on l'appelle *tribu engendrée* par $A$.

De la définition, on peut aussi déduire la propriété suivante, qui dit qu'une tribu est également stable par intersection au plus dénombrable :

**Proposition 16.** *Soit un ensemble $\Omega$ et une tribu $\mathcal{A}$ sur $\Omega$. Pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{A}$, l'intersection $\bigcap_{n\in\mathbb{N}} A_n$ est également un élément de $\mathcal{A}$.*

**Remarques :**

- Dans le contexte des probabilités, un tel ensemble $\Omega$ est appelé *univers*, et les éléments de $\mathcal{A}$ sont appelés *événements*. Il s'agit des parties de $\Omega$ dont on pourra mesurer la *probabilité*, une fois qu'aura été introduite une (mesure de) probabilité sur l'espace probabilisable $(\Omega, \mathcal{A})$.
- Tout élément $\omega \in \Omega$ peut s'interpréter comme la réalisation d'une alternative unique parmi toutes les possibilités, qui déterminera la réalisation de tout événement $A \in \mathcal{A}$ tel que $\omega \in A$. Si $\{\omega\} \in \mathcal{A}$ (ce n'est pas nécessairement le cas !), on dit qu'il s'agit d'un *événement élémentaire*.
- En général, on ne précise pas l'univers $\Omega$, et l'expérience aléatoire est modélisée par des *variables aléatoires* (voir plus loin). Dans des cas simples, on peut tout de même se débrouiller sans variable aléatoire et modéliser explicitement l'expérience à travers $\Omega$ (par exemple $\Omega = \{1,2,3,4,5,6\}$ pour le lancer d'un dé ordinaire).
- On introduit un vocabulaire adapté aux probabilités pour désigner les opérations ensemblistes. On a ainsi la terminologie suivante, pour $A, B \in \mathcal{A}$ :
  — "*non* $A$" désigne l'événement $\overline{A} = \Omega \backslash A$, appelé *événement contraire* de $A$ ;
  — "$A$ *ou* $B$" désigne l'événement $A \cup B$ ;
  — "$A$ *et* $B$" désigne l'événement $A \cap B$ ;
  — "$A$ implique $B$" se traduit par la condition $A \subset B$ ;
  On peut généraliser les notions de *ou* et de *et*, pour une suite $(A_n)_n$ d'événements :
  — "Il existe $n \in \mathbb{N}$ tel que $A_n$" désigne l'événement $\bigcup_{n\in\mathbb{N}} A_n$ ;
  — "Pour tout $n \in \mathbb{N}$, $A_n$" désigne l'événement $\bigcap_{n\in\mathbb{N}} A_n$.
  Enfin, les événements particuler $\varnothing$ et $\Omega$ sont appelés respectivement *événement impossible* et *événement certain*, et deux événements dont l'intersection est l'événement impossible sont dits *incompatibles*.

**Exercice 4.** Soit $\Omega$ un ensemble infini et $(A_n)_{n\in\mathbb{N}}$ une famille de parties de $\Omega$ réalisant une *partition* de $\Omega$ :

$$n \neq m \Rightarrow A_n \cap A_m = \varnothing \text{ et } \bigcup_{n\in\mathbb{N}} A_n = \Omega$$

On pose

$$\mathcal{A} = \left\{ \bigcup_{n\in T} A_n \;\middle|\; T \in \mathcal{P}(\mathbb{N}) \right\}$$

**a)** Montrer que $\mathcal{A}$ est une tribu de $\Omega$.

**b)** On suppose l'ensemble $\Omega$ dénombrable. Montrer que toute tribu infinie sur $\Omega$ est de la forme ci-dessus pour une certaine famille $(A_n)_{n\in\mathbb{N}}$.

**c)** Existe-t-il des tribus dénombrables ?

## 2.2 Espace probabilisé

**Définition 10.** Si $\Omega$ est un ensemble et $\mathcal{A}$ une tribu sur $\Omega$, on appelle *probabilité* sur $(\Omega, \mathcal{A})$ une application $\mathbb{P} : \mathcal{A} \to [0,1]$ vérifiant :

**a)** $\mathbb{P}(\Omega) = 1$

**b)** Pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'événements deux à deux incompatibles,

$$\mathbb{P}\left(\bigsqcup_{n=0}^{+\infty} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n)$$

$(\Omega, \mathcal{A}, \mathbb{P})$ est alors appelé *espace probabilisé*.

**Remarque :** La seconde des deux propriétés exigées par la définition porte le nom d'*additivité dénombrable*. De ces deux propriétés, on peut en déduire bien d'autres, exposées dans la proposition suivante.

**Proposition 17.** *Soit $\mathbb{P}$ une probabilité sur un espace probabilisable $(\Omega, \mathcal{A})$. On a alors les propriétés suivantes :*

***a)*** $\mathbb{P}(\varnothing) = 0$.

***b)*** *Pour tout $A \in \mathcal{A}$, $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$*

***c)*** *Pour toute famille finie $(A_k)_{1\leqslant k\leqslant n}$ d'événements deux à deux incompatibles,*

$$\mathbb{P}\left(\bigsqcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mathbb{P}(A_k) \qquad \textit{(additivité finie)}$$

***d)*** *Pour $A, B \in \mathcal{A}$, on a $A \subset B \Rightarrow \mathbb{P}(A) \leqslant \mathbb{P}(B)$ (croissance)*

**Proposition 18.** ***(formule de Grassmann)*** *Pour $A, B \in \mathcal{A}$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.*

**Remarques :**
- Il existe une formule générale pour toute réunion finie, appelée *formule du crible*. Sauriez-vous la deviner ?
- On en déduit immédiatement $\mathbb{P}(A \cup B) \leqslant \mathbb{P}(A) + \mathbb{P}(B)$, ce qui se généralise par récurrence.

**Corollaire 3.** ***(sous-additivité finie)*** *Pour toute famille finie $(A_k)_{0\leqslant k\leqslant n}$ d'événements, on a :*

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

**Proposition 19.** ***(continuité croissante et décroissante)***

- *si $(A_n)_n$ est une suite croissante d'événements (*ie $A_n \subset A_{n+1}$ *pour tout $n \in \mathbb{N}$), alors :*

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to+\infty} \mathbb{P}(A_n)$$

- *si $(A_n)_n$ est une suite décroissante d'événements (*ie $A_{n+1} \subset A_n$ *pour tout $n \in \mathbb{N}$), alors :*

$$\mathbb{P}\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{n\to+\infty} \mathbb{P}(A_n)$$

**Remarque :** des propriétés de continuité croissante et décroissante, on en déduit plus généralement que pour une suite quelconque $(A_n)_{n\in\mathbb{N}}$ d'événements :

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to+\infty} \mathbb{P}\left(\bigcup_{k=0}^{n} A_k\right) \qquad \text{et} \qquad \mathbb{P}\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{n\to+\infty} \mathbb{P}\left(\bigcap_{k=0}^{n} A_k\right)$$

**Corollaire 4.** ***(sous-additivité dénombrable)***

*Pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'événements, on a :*

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

Terminons cette sous-section avec un complément de vocabulaire à propos des événements ayant une probabilité nulle ou égale à 1.

**Définition 11.** Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $A \in \mathcal{A}$.

- On dit que $A$ est un événement *négligeable* lorsque $\mathbb{P}(A) = 0$.
- On dit que $A$ est un événement *presque sûr* lorsque $\mathbb{P}(A) = 1$.

**Remarque :** Attention, on a forcément $\mathbb{P}(\varnothing) = 0$ (l'événement impossible est toujours négligeable) mais $\mathbb{P}(A) = 0$ n'implique pas $A = \varnothing$ : un événement peut être de probabilité nulle sans être impossible, et cela peut dépendre de la probabilité $\mathbb{P}$ définie sur l'espace probabilisable $(\Omega, \mathcal{A})$ ). De même, $\mathbb{P}(\Omega) = 1$ mais on peut avoir $\mathbb{P}(A) = 1$ pour $A \neq \Omega$ (un événement presque sûr n'est pas forcément l'événement certain !).

**Proposition 20.**

- *Toute réunion au plus dénombrable d'événements négligeables est un événement négligeable.*
- *Toute intersection au plus dénombrable d'événements presque sûrs est un événement presque sûr.*

**Définition 12.** Une famille $(A_i)_{i\in I}$ au plus dénombrable d'événements est un système :

- *complet* lorsqu'elle réalise une partition de $\Omega$.
- *quasi-complet* lorsque les $A_i$ sont deux à deux incompatibles et $\mathbb{P}\left(\bigcup_{i\in I} A_i\right) = 1$.

**Remarque :** Ainsi, pour un système quasi-complet d'événements, la réalisation d'aucun des événements $A_i$ n'est pas forcément impossible mais reste négligeable, *i.e.* de probabilité nulle.

## 2.3 Variable aléatoire discrète

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

**Définition 13.** On appelle *variable aléatoire discrète* sur $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans un ensemble $E$ toute application $X : \Omega \to E$ vérifiant :

- $X(\Omega)$ est au plus dénombrable
- pour tout $x \in X(\Omega)$, $X^{-1}(\{x\}) \in \mathcal{A}$.

**Remarques :**
- La probabilité $\mathbb{P}$ n'intervient pas dans la définition d'une variable aléatoire, et on on peut donc aussi parler de variable aléatoire discrète sur $(\Omega, \mathcal{A})$. La probabilité $\mathbb{P}$ ne va intervenir que dans la définition de la *loi* d'une telle variable (voir juste après).
- De nombreux phénomènes aléatoires peuvent être modélisés par un ensemble (parfois infini) de variables aléatoires suposées vérifier des *lois de probabilité* particulières sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Un tel espace n'est généralement pas explicité car on dispose de théorèmes assurant son existence, sous certaines conditions. Nous ne nous occuperons pas de ces points théoriques en MPI, et verrons simplement ultérieurement un théorème d'existence dans le cas d'une suite de variables aléatoires discrètes.
- Le terme *variable aléatoire* porte un peu à confusion puiqu'il s'agit d'une application sur $\Omega$. Il provient des origines historiques de la notion de probabilité, mais l'interprétation reste intuitivement claire : la fonction $X$ prend ses valeurs dans $X(\Omega)$, et on peut la voir comme une quantité qui "varie" de façon "aléatoire" parmi ces valeurs. On va voir d'ailleurs une série de notations qui font écho à cette interprétation.

le fait que l'image réciproque de tout élément de $X(\Omega)$ soit un élément de $\mathcal{A}$ garantit la possibilité d'évaluer la probabilité de toutes les réalisations possibles de la variable, comme l'indique la proposition suivante.

**Proposition 21.** *Soit $X$ une variable aléatoire discrète sur $(\Omega, \mathcal{A})$. Pour toute partie $U \subset X(\Omega)$, on a $X^{-1}(U) \in \mathcal{A}$.*

**Remarque :** Cela reste vrai plus généralement pour toute partie $U$ de l'ensemble d'arrivée $E$, puisqu'on a alors $U = \big(U \cap X(\Omega)\big) \bigsqcup \big(U \cap \overline{X(\Omega)}\big)$ (réunion disjointe) dont l'image réciproque par $X$ est donc égale à $X^{-1}\big(U \cap X(\Omega)\big) \bigsqcup \varnothing = X^{-1}\big(U \cap X(\Omega)\big)$.

**Notation :** Soit $X$ une variable aléatoire discrète sur $(\Omega, \mathcal{A})$ et à valeur dans un ensemble $E$.
- Pour $U \subset X(\Omega)$ (ou plus généralement $U \subset E$) on note $\{X \in U\}$ ou $(X \in U)$ l'événement $X^{-1}(U)$
- Pour $x \in X$, on note plus simplement $\{X = x\}$ ou $(X = x)$ l'événement $X^{-1}(\{x\}$ (au lieu de $\{X \in \{x\}\}$ ou $(X \in \{x\})$)
- Si $E = \mathbb{R}$ (on dit alors que $X$ est une variable aléatoire *réelle*), on peut encore utiliser d'autre symboles relationnels comme par exemple $\{X \leqslant x\}$ (pour l'événement $\{X \in ]-\infty, x]\}$) ou encore $\{X > x\}$, etc ... (et analogues avec parenthèses).

## 2.4 Loi d'une variable aléatoire discrète

**Définition 14.** Soit $X$ une variable aléatoire discrète sur $(\Omega, \mathcal{A}, \mathbb{P})$. On appelle *loi de probabilité* de $X$ la fonction $\mathbb{P}_X : \mathcal{P}\big(X(\Omega)\big) \to [0,1]$ définie par :

$$\forall U \subset X(\Omega)\,, \quad \mathbb{P}_X(U) = \mathbb{P}(X \in U) = \mathbb{P}\big(X^{-1}(U)\big).$$

**Remarque :** On peut plus généralement écrire que cette loi $\mathbb{P}_X$ est définie sur $\mathcal{P}(E)$, avec $E$ l'ensemble d'arrivée de $X$ ($E$ contient donc $X(\Omega)$), puisque pour tout $U \subset E$, on a alors $X^{-1}(U) = X^{-1}\big(U \cap X(\Omega)\big)$.

**Exemple :** Si $X$ est à valeurs dans $\mathbb{R}$ mais avec $X(\Omega) = \mathbb{Z}$, on peut évoquer $\mathbb{P}_X(U)$ pour toute partie $U$ de $\mathbb{R}$. On aura :

$$\mathbb{P}_X(U) = \mathbb{P}(X \in U) = \sum_{n \in U \cap \mathbb{Z}} \mathbb{P}(X = n)$$

Dans l'exemple ci-dessus, on a ramené la probabilité de l'événement $\{X \in U\}$ à celle de la réunion disjointe de tous les événements $\{X = n\}$ pour $n \in U \cap X(\Omega)$. Il s'agit en fait d'un point très général : la loi de probabilité d'une variable aléatoire discrète est entièrement déterminée par sa valeur sur les singletons. C'est ce qu'exprime le résultat suivant.

**Proposition 22.** *Pour toute variable aléatoire discrète $X$ sur $(\Omega, \mathcal{A}, \mathbb{P})$ :*

***a)*** *la loi de $X$ est uniquement déterminée par la famille $\big(\mathbb{P}(X = x)\big)_{x \in X(\Omega)}$.*

***b)*** *$\mathbb{P}_X$ définit une probabilité sur l'espace probabilisable $\big(X(\Omega), \mathcal{P}(X(\Omega))\big)$.*

**Remarques :**
- On dit que la famille $\big(\mathbb{P}(X = x)\big)_{x \in X(\Omega)}$ est une *distribution de probabilité discrète* sur $X(\Omega)$ et que $\big(X(\Omega), \mathcal{P}(X(\Omega)), \mathbb{P}_X\big)$ est un *espace probabilisé discret*.

- En pratique, on pourra donc définir $\mathbb{P}_X$ en donnant simplement $\mathbb{P}(X = x)$ pour tout $x \in X(\Omega)$.
- Deux variables aléatoires $X$ et $Y$ suivent une même loi si, et seulement si, on a $X(\Omega) = Y(\Omega)$ et $\mathbb{P}(X = x) = \mathbb{P}(Y = x)$ pour tout $x \in X(\Omega)$. Il n'est même pas besoin en fait d'imposer que $X$ et $Y$ soient définies sur le même espace probabilisé.

**Notation :** Si $X$ et $Y$ sont deux variables aléatoires suivant une même loi, on note $X \sim Y$.

**Définition 15. Lois usuelles**

- **Loi uniforme** : $E = X(\Omega)$ fini de cardinal $n \in \mathbb{N}^*$ et pour tout $x \in E$ :

$$\mathbb{P}(X = x) = \frac{1}{n} \quad ; \quad \text{on note } X \sim \mathcal{U}(E).$$

- **Loi de Bernoulli de paramètre** $p$ : $X(\Omega) = \{0, 1\}$ et $p \in [0, 1]$ avec :

$$\mathbb{P}(X = 1) = p \quad \text{et donc} \quad \mathbb{P}(X = 0) = 1 - p \quad ; \quad \text{on note } X \sim \mathcal{B}(p).$$

- **Loi Binomiale de paramètre** $n$ **et** $p$ : $X(\Omega) = [\![0, n]\!]$, $n \in \mathbb{N}^*$ et $p \in [0, 1]$, avec pour tout $k \in [\![0, n]\!]$ :

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad ; \quad \text{on note } X \sim \mathcal{B}(n, p).$$

  On peut avoir aussi $X(\Omega) = \mathbb{N}$ avec la convention $\binom{n}{k} = 0$ si $k > n$.

- **Loi géométrique de paramètre** $p$ : $X(\Omega) = \mathbb{N}^*$ et $p \in ]0, 1]$, avec pour tout $k \in \mathbb{N}^*$ :

$$\mathbb{P}(X = k) = (1-p)^{k-1} p \quad ; \quad \text{on note } X \sim \mathcal{G}(p).$$

- **Loi de Poisson de paramètre** $\lambda$ : $X(\Omega) = \mathbb{N}$ et $\lambda \in \mathbb{R}_+$ avec pour tout $k \in \mathbb{N}$ :

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} \quad ; \quad \text{on note } X \sim \mathcal{P}(\lambda).$$

**Remarques :**

- La loi de Bernoulli $\mathcal{B}(p)$ modélise une expérience ayant deux issues seulement, "succès" (probabilité $p$) ou "échec" ( probabilité $1 - p$).
- La loi binomiale $\mathcal{B}(n, p)$ s'interprète comme le nombre de succès lors de la répétition de $n$ expériences indépendantes, ayant chacun une probabilité $p$ de succès.
- La loi géométrique $\mathcal{G}(p)$ s'intérprète comme donnant la probabilité du rang du premier succès dans une suite illimité d'expériences indépendantes ayant chacune une probabilité $p$ de succès.
- On verra (proposition 63) que pour "$n$ grand devant $\lambda$", la loi de Poisson $\mathcal{P}(\lambda)$ approche la loi Binomiale $\mathcal{B}\left(n, \frac{\lambda}{n}\right)$ : c'est la *loi des événements rares.*

## 2.5 Fonction d'une variable aléatoire

**Notation :** Soient $E$ et $F$ deux ensemble quelconques non vides, $X : \Omega \to E$ une variable aléatoire discrète et $f : E \to F$ une application. On note alors $f(X)$ l'application $f \circ X$.

**Proposition 23.** *Si $X$ est une variable aléatoire discrète à valeurs dans $E$ et $f : E \to F$, alors $f(X)$ est une variable aléatoire discrète à valeurs dans $F$. De plus :*

***a)*** *La loi de $f(X)$ est uniquement déterminée par $f$ et par la famille $\big(\mathbb{P}(X = x)\big)_{x \in X(\Omega)}$.*

***b)*** *Si $X \sim Y$, alors $f(X) \sim f(Y)$.*

**Remarque :** On obtient la loi de $f(X)$ à partir de celle de $X$ grâce à la relation :

$$\mathbb{P}(f(X) = y) = \sum_{x \in f^{-1}(\{y\})} \mathbb{P}(X = x)$$

**Exemple :** Soit $X : \Omega \to \{-1, 0, 1\}$ une variable aléatoire uniforme. On peut considérer la variable aléatoire $X^2$. On a $\mathbb{P}(X^2 = 0) = \mathbb{P}(X = 0) = \frac{1}{3}$ et $\mathbb{P}(X^2 = 1) = \mathbb{P}(X \in \{-1, 1\}) = \mathbb{P}(X = 1) + \mathbb{P}(X = -1) = \frac{2}{3}$.

# 3 Conditionnement et indépendance

Dans toute cette section, $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace probabilisé.

## 3.1 Probabilité conditionnelle.

La probabilité d'un événement $A$ correspondant d'une certaine façon à la "mesure" de l'ensemble $A$, on est naturellement conduit à définir la probabilité de $A$ *conditionnellement* à un événement $B$ comme la mesure de $A \cap B$ "relativement" à la mesure de $B$ : on examine la mesure de $A$ dans un univers "réduit à" $B$.

**Définition 16.** Soient $A, B \in \mathcal{A}$ deux événements tels que $\mathbb{P}(B) > 0$. On appelle *probabilité conditionnelle* de $A$ sachant $B$ le réel

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

On la note habituellement $\mathbb{P}(A \mid B)$.

**Proposition 24.** *Un événement $B \in \mathcal{A}$ tel que $\mathbb{P}(B) > 0$ étant fixé, l'application $\mathbb{P}_B : A \mapsto \mathbb{P}(A \mid B)$ est une probabilité sur $(\Omega, \mathcal{A})$.*

la relation $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$ peut s'appliquer en cascade, et donner le résultat suivant, *via* une récurrence immédiate.

**Proposition 25.** ***(formule des probabilités composées)*** *Soit $(A_i)_{1 \leqslant i \leqslant n}$ une suite finie d'événements tels que $\mathbb{P}\left(\bigcap_{i=1}^{n-1} A_i\right) > 0$. On a alors : $\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1})$.*

**Exercice 5.** Dans une urne contenant $n$ boules noires et $n$ boules blanches, on pioche successivement sans remise. Quelle est la probabilité de vider l'urne en alternant parfaitement les couleurs.

**Définition 17.** Si $X$ est une variable aléatoire discrète et $A$ un événement tel que $\mathbb{P}(A) > 0$, l'application $U \mapsto \mathbb{P}(X \in U \mid A)$, définie sur $\mathcal{P}\big(X(\Omega)\big)$, s'appelle *loi de $X$ conditionnellement* à l'événement $A$.

**Remarque :** Cette loi est uniquement déterminée par la famille $\big(\mathbb{P}(X = x \mid A)\big)_{x \in X(\Omega)}$.

## 3.2 Formule des probabilités totales et de Bayes

Nous revisitons maintenant des résulats vus en MP2I pour des suites finis d'événements.

**Théorème 6.** ***(formule des probabilités totales)*** *Soit $(A_i)_{i \in I}$ un système quasi-complet d'événements, tels que $\mathbb{P}(A_i) > 0$ pour tout $i \in I$. On a alors, pour tout événement $B$ :*

$$\mathbb{P}(B) = \sum_{i \in I} \mathbb{P}(B \cap A_i) = \sum_{i \in I} \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

**Remarques :**

- La formule est encore valide si certains événements $A_i$ sont de probabilité nulle : on s'autorisera à écrire $\mathbb{P}(B|A_i)\mathbb{P}(A_i) = 0$, alors même que $\mathbb{P}(B|A_i)$ n'est pas défini.
- En pratique on a $I = [\![1, n]\!]$ pour un certain $n \in \mathbb{N}^*$, ou $I = \mathbb{N}$ (ou éventuellement $\mathbb{N}^*$, $\mathbb{Z}$, $\mathbb{N}^2$, etc ...).

**Proposition 26.** ***(formule de Bayes)*** *Soient deux événements $A$ et $B$ de probabilité non nulle. Alors*

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

En combinant la formule précédente et la formule des probabilités totales, on obtient :

**Théorème 7.** ***(théorème de Bayes)***
*Soit $(A_i)_{i\in I}$ un sytème quasi-complet d'événements. On a alors, pour tout événement $B \in \mathcal{A}$ de probabilité non nulle, et pour tout $k \in I$ :*

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\sum_{i\in I}\mathbb{P}(B|A_i)\mathbb{P}(A_i)}.$$

**Remarque :** Là encore, on s'autorisera l'écriture de $\mathbb{P}(B|A_i)\mathbb{P}(A_i) = \mathbb{P}(B \cap A_i) = 0$ lorsque $\mathbb{P}(A_i) = 0$.

**Exercice 6.** Un test de dépistage pour un certain type rare de cancer, touchant en moyenne $0,01\%$ de la population, a un taux de fiabilité de 99%, à la fois pour les personnes atteintes et non atteintes. En cas de résultat positif au test pour une personne prise au hasard, quelle est la probabilité qu'elle soit atteinte ?

## 3.3 Indépendance d'événements

**Définition 18.** Soient $A$ et $B$ deux événements. On dit qu'ils sont *indépendants* lorsque $\mathbb{P}(A\cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

**Remarque :** Contrairement à l'incompatibilité, propriété *ensembliste* ne dépendant pas du choix de la probabilité $\mathbb{P}$ définie sur l'espace probabilisable $(\Omega, \mathcal{A})$, La notion d'indépendance est *probabiliste* : deux événements peuvent être indépendants pour un certain choix de $\mathbb{P}$ et pas pour un autre. Ne confondez pas les deux notions !

**Proposition 27.** *On suppose $\mathbb{P}(B) > 0$. $A$ et $B$ sont indépendants si, et seulement si, $\mathbb{P}(A|B) = \mathbb{P}(A)$.*

On peut interpréter ce résultat en disant que l'indépendance de $A$ et $B$ signifie que la connaissance de la réalisation ou non-réalisation de $B$ n'influence pas la probilité qu'a $A$ de se réaliser. Il apparait notamment clairement intuitivement qu'on peut remplacer $A$ et/ou $B$ par leur événement contraire, ce que corrobore la proposition suivante.

**Proposition 28.** *Supposons que $A$ et $B$ sont indépendants. Alors :*
***a)*** *$\overline{A}$ et $B$ sont indépendants*
***b)*** *$A$ et $\overline{B}$ sont indépendants*
***c)*** *$\overline{A}$ et $\overline{B}$ sont indépendants.*

On peut étendre la notion d'indépendance à une famille quelconque d'événements. Attention, la définition est plus contraignante que ce à quoi on pourrait s'attendre intuitivement.

**Définition 19.** On dit que $(A_i)_{i\in I}$ est une famille d'événements *(mutuellement) indépendants* lorsque pour toute partie finie $F \subset I$ :

$$\mathbb{P}\left(\bigcap_{i\in F} A_i\right) = \prod_{i\in F}\mathbb{P}(A_i).$$

**Remarques :**
- l'indépendance deux à deux n'implique pas l'indépendance (mutuelle). Exemple : $\Omega = [\![1,4]\!]$, avec la probabilité uniforme, $A = \{1,2\}$, $B = \{1,3\}$, $C = \{1,4\}$.
- La seule condition $P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$ pour une famille finie $(A_i)_{1\leqslant i\leqslant n}$ ne suffit pas non plus. Exemple : $\Omega = [\![1,6]\!]^2$, avec la probabilité uniforme, $A = [\![1,6]\!] \times \{1,2,5\}$, $B = [\![1,6]\!] \times \{4,5,6\}$, et enfin $C = \{(i,j) \in \Omega \mid i+j = 9\}$.

## 3.4 Loi conjointe et lois marginales d'un couple.

**Proposition 29.** *Soit $X$ et $Y$ deux variables aléatoires discrètes sur $(\Omega, \mathcal{A})$. L'application $(X,Y)$ définie sur $\Omega$ par :*

$$\forall\, \omega \in \Omega\,,\ \ (X,Y)(\omega) = \big(X(\omega), Y(\omega)\big).$$

*est une variable aléatoire discrète sur $(\Omega, \mathcal{A})$, et à valeurs dans $X(\Omega) \times Y(\Omega)$.*

**Notation :** Pour tout $(x,y) \in X(\Omega) \times Y(\Omega)$, l'événement $(X,Y)^{-1}(\{(x,y)\}) = X^{-1}(\{x\}) \cap Y^{-1}(\{y\})$ est noté $\{X = x, Y = y\}$ et sa probabilité éventuelle $P(X = x, Y = y)$.

**Définition 20.** On appelle *loi conjointe* de $(X,Y)$ la loi de probabilité du couple $(X,Y)$ en tant que variable aléatoire discrète sur $\Omega$.
Les lois de $X$ et $Y$ sont appelées *lois marginales* de $(X,Y)$.

Comme pour une seule variable $X$, la loi conjointe d'un couple $(X,Y)$ est uniquement déterminée par les probabilités des événements $\{X = x, Y = y\}$, pour tout $(x,y) \in (X,Y)(\Omega)$.
On retrouve les lois marginales à partir de la loi conjointe :

**Proposition 30.** *Soit $X$ et $Y$ deux variables aléatoires discrètes sur $\Omega$, Les lois marginales de $(X,Y)$ sont déterminées à partir de la loi conjointe par :*

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

**Remarques :**
- Ce procédé de sommation permettant de retrouver une loi marginale s'appelle *marginalisation.*
- La loi conjointe détermine donc les lois marginales, mais la réciproque est fausse, sauf, on le verra, lorsque $X$ et $Y$ sont indépendantes.

**Exemple :** On considère le couple $(X,Y)$ de variables aléatoires, suivant une loi uniforme sur $\{0,1\} \times \{0,1\}$ et $(X',Y')$ de loi $P(X' = 0, Y' = 0) = P(X' = 1, Y' = 1) = \dfrac{1}{8}$ et $P(X' = 0, Y' = 1) = P(X' = 1, Y' = 0) = \dfrac{3}{8}$. On calcule facilement que les lois marginales de $(X,Y)$ et de $(X',Y')$ sont les mêmes.

**Remarque :** On peut généraliser la notion de loi conjointe et de lois marginale à un $n$-uplet $(X_1, \dots, X_n)$ de variables aléatoires, pour tout $n \geqslant 2$. On aura ainsi, par exemple $(n = 4)$ :

$$P(X_1 = x_1, X_2 = x_2) = \sum_{x_3 \in X_3(\Omega)} \sum_{x_4 \in X_4(\Omega)} \mathbb{P}(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4)$$

## 3.5 Indépendance de deux variables aléatoires discrètes

**Définition 21.** On dit que deux variables aléatoires discrètes $X$ et $Y$ sur $(\Omega, \mathcal{A}, P)$ sont *indépendantes* lorsque pour tout $U \subset X(\Omega)$ et tout $V \subset Y(\Omega)$, les événements $\{X \in U\}$ et $\{Y \in V\}$ sont indépendants. On note alors $X \perp\!\!\!\perp Y$.

Cette définition de l'indépendance de $X$ et $Y$ sur les tous les événements possibles associés aux valeurs que peuvent prendre $X$ et $Y$ peut en fait être ramenée au cas des valeurs individuelles.

**Proposition 31.** *$X$ et $Y$ sont indépendantes si, et seulement si, pour tout $(x,y) \in X(\Omega) \times Y(\Omega)$ :*

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)P(Y = y).$$

**Corollaire 5.** *Si $X \perp\!\!\!\perp Y$, la loi conjointe de $(X,Y)$ est uniquement déterminée par les lois marginales de $X$ et $Y$.*

Plus généralement, l'indépendance de $X$ et $Y$ implique l'indépendance de toute fonction de $X$ et de toute fonction de $Y$, comme l'indique la proposition suivante, pour laquelle on rappelle que $f(X)$ et $g(Y)$ sont des notations pour $f \circ X$ et $g \circ Y$.

**Proposition 32.** *Si $X$ et $Y$ sont deux variable aléatoire discrète indépendantes, alors pour toutes fonctions $f$ et $g$ définies sur $X(\Omega)$ et $Y(\Omega)$ respectivement, on a $f(X) \perp\!\!\!\perp g(Y)$.*

Comme pour les événements, l'indépendance de $Y$ et de $Y$ peut s'interpréter par l'idée que la connaissance d'une valeur prise par $Y$ n'affecte pas les probabilités des valeurs prises par $X$ (et vice versa). L'utilisation de la loi conditionnelle concrétise cette idée :

**Proposition 33.** *Soient deux variables aléatoires discrètes $X$ et $Y$. Les assertions suivantes sont équivalentes :*

***(i)*** *$X$ et $Y$ sont indépendantes*

***(ii)*** *pour tout $x \in X(\Omega)$ et tout $y \in Y(\Omega)$ tel que $\mathbb{P}(Y = y) > 0$, on a $\mathbb{P}(X = x|Y = y) = \mathbb{P}(X = x)$*

La définition et les résultats précédents se généralisent au cas de $n$ variables aléatoires discrètes $X_1, \dots, X_n$.

**Définition 22.** On dit que les variables aléatoires discrètes $X_1, \dots, X_n$ sont (mutuellement) indépendantes lorsque pour tout $(U_1, \dots, U_n) \in \mathcal{P}(X_1(\Omega)) \times \cdots \times \mathcal{P}(X_n(\Omega))$, $\big(\{X_i \in U_i\}\big)_{1 \leqslant i \leqslant n}$ est une famille d'événements (mutuellement) indépendants.

**Remarque :** Comme pour les familles d'événements, l'indépendance deux à deux des $X_i$ n'implique pas l'indépendance.

**Exercice 7.** Soient $X$ et $Y$ deux variables aléatoires indépendantes telles que $X \sim Y \sim \mathcal{B}(1/2)$, et soit $Z = X + Y \bmod 2$. Étudier l'indépendance de $X, Y, Z$.

Pour vérifier l'indépendance de $X_1, \dots, X_n$, il suffit en fait de se restreindre aux événements de la forme $\{X_i = x\}$, comme le précise le résultat suivant.

**Proposition 34.** *$X_1, \dots, X_n$ sont indépendantes si, et seulement si, :*

$$\forall (x_1, \dots, x_n) \in X_1(\Omega) \times \cdots X_n(\Omega), \quad \mathbb{P}(X_1 = x_1, \cdots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i)$$

On peut généraliser alors le résultat concernant l'indépendance de $f(X)$ et $g(Y)$ à partir de celle de $X$ et $Y$. On peut même faire intervenir des fonctions qui "regroupent" plusieurs variables aléatoires, ce qu'on appelle des *coalitions*.

**Proposition 35.** ***(lemme des coalitions)*** *Soient $n \geqslant 2$ variables aléatoires $X_1, \dots, X_n$ indépendantes, soit $m \in [\![1, n-1]\!]$, et soient deux applications :*

$$f : \prod_{k=1}^{m} X_k(\Omega) \to E, \qquad g : \prod_{k=m+1}^{n} X_k(\Omega) \to F$$

*Alors $f(X_1, \dots, X_m)$ et $g(X_{m+1}, \dots, X_n)$ sont indépendantes.*

**Remarque :** Ce résultat se généralise bien entendu au cas de plus de deux coalitions.

De nombreuses situations seront modélisées par une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires dont on souhaitera supposer qu'elles sont indépendantes. Il nous faut donc généraliser la définition de l'indépendance des variables à des familles quelconques.

**Définition 23.** On dit qu'une famille $(X_i)_{i \in I}$ de variables aléatoires discrètes est une famille de variables *indépendantes* lorsque pour toute partie finie $F \subset I$, $(X_i)_{i \in F}$ est une famille finie de variables indépendantes.

**Définition 24.** On dit que $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires discrètes *indépendantes identiquement distribuées*, ce qu'on note *i.i.d* lorsque les $X_n$ sont indépendantes et suivent toutes la même loi.

**Exemple :** L'exemple le plus commun est celui d'une suite $(X_n)_{n \in \mathbb{N}}$ *i.i.d.* avec $X_n \sim \mathcal{B}(p)$. Le cas $p = \dfrac{1}{2}$ permet de modéliser le jeu de pile ou face infini.

# 4 Espérance et variance, compléments sur les variables aléatoires

## 4.1 Espérance

**Définition 25.** Soit $X$ une variable aléatoire discrète à valeurs dans $[0,+\infty]$. On appelle *espérance* de $X$ l'élément de $[0,+\infty]$ défini par :

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

**Remarques :**

- On a $\mathbb{E}(X) < +\infty$ si, et seulement si, la famille $\big(x\mathbb{P}(X=x)\big)_{x\in X(\Omega)}$ est sommable.
- Lorsque $\mathbb{P}(X=+\infty) > 0$, on a nécessairement $\mathbb{E}(X) = +\infty$, mais on peut très bien avoir $\mathbb{E}(X) = +\infty$ avec $\mathbb{P}(X=+\infty) = 0$.

**Exemples :**

- Pour $X$ à valeurs dans $\mathbb{N}$, avec $\mathbb{P}(X=n) = \dfrac{2n+1}{n^2(n+1)^2}$, On a $\mathbb{E}(X) < +\infty$.
- Pour $X$ à valeurs dans $\mathbb{N}$, avec $\mathbb{P}(X=n) = \dfrac{1}{n(n+1)}$, on a $\mathbb{E}(X) = +\infty$.

Dans le cas d'une variable aléatoires à valeurs dans $\mathbb{N}$ (comme c'est le cas pour les lois classiques), la formule de la proposition suivante peut s'avérer très utile. Attention de ne pas chercher à l'utiliser pour une variable aléatoire discrète à valeurs non entières !

**Proposition 36.** *Si $X$ est à valeurs dans $\mathbb{N}\cup\{+\infty\}$, on a :*

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

Pour une variable aléatoire $X$ qui ne prend pas que des valeurs positives, on ne peut pas toujours définir l'espérance : il faut en effet que cela ait un sens d'attribuer une valeur à la somme des $x\mathbb{P}(X=x)$ pour toutes les valeurs prises par $X$.

**Définition 26.** Soit $X$ une variable aléatoire discrète réelle ou complexe. On dit que $X$ est d'*espérance finie* lorsque la famille $\big(x\mathbb{P}(X=x)\big)_{x\in X(\Omega)}$ est sommable. L'espérance est alors définie par :

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

**Remarques :**

- Dans le cas d'une variable aléatoire discrète non positive, on peut aussi dire que $X$ *admet* une espérance Lorsque $\big(x\mathbb{P}(X=x)\big)_{x\in X(\Omega)}$ est sommable (étant entendu que l'espérance n'est pas définie si cette famille n'est pas sommable).
- Si $X(\Omega)$ est fini, l'espérance est toujours finie.
- Si $X(\Omega)$ est dénombrable, on peut écrire $X(\Omega) = \{x_n, n\in\mathbb{N}\}$. L'espérance est alors finie si, et seulement si, la série $\sum x_n\mathbb{P}(X=x_n)$ est *absolument convergente*, et l'espérance est égale dans ce cas à la somme de cette série.

**Notation :** Lorsqu'une variable aléatoire discrète $X$ est d'espérance finie, on écrit $X\in L^1$.

## 4.2 Espérance des lois classiques, Propriétés de l'espérance

**Proposition 37.** ***(espérances classiques)***

- *Si $X\sim\mathcal{B}(n,p)$, $\mathbb{E}(X) = np$.*
- *Si $X\sim\mathcal{G}(p)$, $\mathbb{E}(X) = \dfrac{1}{p}$.*
- *Si $X\sim\mathcal{P}(\lambda)$, $\mathbb{E}(X) = \lambda$.*

**Théorème 8.** ***(formule de transfert)***
*Soit $X$ une variable aléatoire discrète sur $(\Omega, \mathcal{A}, \mathbb{P})$, et soit $f : X(\Omega) \to \mathbb{R}$. Alors $f(X)$ est d'espérance finie si, et seulement si, la famille $(f(x)\mathbb{P}(X = x))_{x\in X(\Omega)}$ est sommable. On a dans ce cas :*

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

Le théorème de transfert permet de montrer un ensemble de propriétés classiques à propos de l'espérance :

**Proposition 38.** *Soient $X$ et $Y$ deux variables aléatoires discrètes réelles ou complexes sur $(\Omega, \mathcal{A}, P)$ d'espérances finies. Alors :*
- *Linéarité : pour tout $\lambda \in \mathbb{C}$, $X + \lambda Y$ est d'espérance finie et $\mathbb{E}(X + \lambda Y) = \mathbb{E}(X) + \lambda\mathbb{E}(Y)$.*
- *Positivité : si $X \geqslant 0$ (*ie. *$X(\Omega) \subset \mathbb{R}^+$), alors $\mathbb{E}(X) \geqslant 0$*
- *Croissance : si $X$ et $Y$ sont réelles et $X \leqslant Y$ (*ie $X(\omega) \leqslant Y(\omega)$ *pour tout $\omega \in \Omega$), alors $\mathbb{E}(X) \leqslant \mathbb{E}(Y)$.*
- *Inégalité triangulaire : $\big|\mathbb{E}(X)\big| \leqslant \mathbb{E}\big(|X|\big)$*

**Remarque :** La propriété de la linéarité cache une difficulté : elle ne correspond pas simplement à la linéarité de l'opérateur $\sum$. En effet, l'opération qui permet de passer de la loi de $(X, Y)$ à la loi de $X + Y$ n'est pas une simple somme, mais une opération plus compliquée qu'on appelle *convolution*. Par exemple, pour $X$ et $Y$ à valeurs dans $\mathbb{N}$, on a pour $n \in \mathbb{N}$ :

$$\mathbb{P}(X + Y = n) = \sum_{k=0}^{n} \mathbb{P}(X = k, Y = n - k)$$

En utilisant le théorème de transfert avec la fonction $x \mapsto |x|$, et la croissance de l'espérance, on peut montrer les deux critères utiles suivants pour savoir si une variable aléatoire discrète $X$ est d'espérance finie :

**Proposition 39.** *Soit $X$ une variable aléatoire discrète réelle ou complexe. On a :*
***a)*** *$X \in L^1$ si, et seulement si, $\mathbb{E}(|X|) < +\infty$.*
***b)*** *Si $|X| \leqslant Y$ pour une variable aléatoire discrète positive $Y \in L^1$, alors $X \in L^1$.*

La notion suivante et les deux propositions qui suivent concernent le cas des variables aléatoires discrètes d'espérance nulle.

**Définition 27.** On dit qu'une variable aléatoire discrète d'espérance finie est *centrée* lorsque $\mathbb{E}(X) = 0$.

**Proposition 40.** *Si $X$ est d'espérance finie, la variable aléatoire discrète $X - \mathbb{E}(X)$ est centrée.*

**Proposition 41.** *Si $X$ est à valeurs dans $\mathbb{R}^+$, on a $\mathbb{E}(X) = 0$ si, et seulement si, l'événement $\mathbb{P}(X \neq 0)$ est négligeable.*

On a vu que l'espérance d'une somme est la somme de l'espérance. Qu'en est-il pour un produit ?

**Proposition 42. (Espérance d'un produit)** *Soient $X$ et $Y$ deux variables aléatoires discrètes réelles ou complexes. Si $X$ et $Y$ sont dans $L^1$ et indépendantes, alors $XY \in L^1$ et :*

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

**Remarques :**
- Si $X$ et $Y$ ne sont pas indépendantes, elles peuvent très bien être d'espérance finie sans que $XY$ le soit. Par exemple $X = Y$ avec la loi $P(X = n) = \dfrac{1}{Sn^3}$, avec $S = \displaystyle\sum_{n=1}^{+\infty} \frac{1}{n^3}$. (Utiliser le théorème de transfert, avec $f : x \mapsto x^2$).
- On peut avoir $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ sans que $X$ et $Y$ soient indépendantes comme le montre l'exemple du couple $(X, Y)$ dont la loi conjointe est représentée dans le tableau suivant :

| | $X = 1$ | $X = -1$ | Loi de $Y$ |
|---|---|---|---|
| $Y = -1$ | 1/8 | 1/8 | 1/4 |
| $Y = 0$ | 3/8 | 1/8 | 1/2 |
| $Y = 1$ | 1/8 | 1/8 | 1/4 |
| Loi de $X$ | 5/8 | 3/8 | 1 |

- Ce résultat se généralise au cas de $n$ variables indépendantes $X_1, \ldots, X_n$.

## 4.3 Variance, écart-type

On supposera ici que les variables aléatoires considérées sont réelles.

**Proposition 43.** *Si $X$ une variable aléatoire discrète réelle vérifiant $\mathbb{E}(X^2) < +\infty$, alors $X \in L^1$.*

**Notation :** Lorsque $\mathbb{E}(X^2) < +\infty$, autrement dit lorque $X^2 \in L^1$, on note $X \in L^2$.

**Remarque :** La proposition précédente dit donc que $X \in L^2 \Rightarrow X \in L^1$.

**Proposition 44.** ***(inégalité de Cauchy-Schwarz)***
*Si $X$ et $Y$ sont deux variables aléatoires discrètes réelles dans $L^2$ alors $XY \in L^1$ et*

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

*De plus, il y a égalité si, et seulement si, il existe $\lambda \in \mathbb{R}$ tel que $\mathbb{P}(Y = \lambda X) = 1$*

**Remarque :** Dans le cas d'égalité, on dit que $Y = \lambda X$ *presque sûrement.*

**Définition 28.** Pour $X \in L^2$, on appelle *variance* de $X$ le réel :

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

On appelle *écart-type* de $X$ le réel :

$$\sigma(X) = \sqrt{\mathbb{V}(X)}.$$

## 4.4 Propriétés, variances des lois classiques

**Proposition 45.** *Pour $X \in L^2$ on a $\mathbb{V}(X) = 0$ si, et seulement si, $\mathbb{P}(X = \mathbb{E}(X)) = 1$.*

**Remarque :** On dit dans ces conditions que $X$ est *presque sûrement* constante.

**Proposition 46.** ***(formule de König-Huygens)*** *Pour $X \in L^2$, on a :*

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

**Proposition 47.** *Si $X \in L^2$ et $a, b \in \mathbb{R}$, alors $aX + b \in L^2$ et $\mathbb{V}(aX + b) = a^2\mathbb{V}(X)$*

**Définition 29.** On dit qu'une variable aléatoire discrète réelle $X \in L^2$ est *réduite* lorsque $\sigma(X) = 1$.

**Proposition 48.** *Si $\sigma(X) > 0$, la variable aléatoire $\dfrac{X - \mathbb{E}(X)}{\sigma(X)}$ est centrée réduite.*

**Proposition 49.** ***(Variances classiques)***

- *Si $X \sim \mathcal{B}(n, p)$, $\mathbb{V}(X) = np(1-p)$.*
- *Si $X \sim \mathcal{G}(p)$, $\mathbb{V}(X) = \dfrac{1-p}{p^2}$*
- *Si $X \sim \mathcal{P}(\lambda)$, $\mathbb{V}(X) = \lambda$*

## 4.5 Covariance de deux variables

**Définition 30.** Soient $X$ et $Y$ deux variables aléatoires discrètes réelles telles que $X, Y \in L^2$. On appelle *covariance* de $X$ et $Y$ le réel :

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

Si $Cov(X, Y) = 0$, on dit que $X$ et $Y$ sont *décorrélées*

**Remarques :**
- $\mathrm{Cov}(X,Y) > 0$ signifie que $Y$ a tendance à augmenter quand $X$ augmente (corrélation positive)
- $\mathrm{Cov}(X,Y) < 0$ signie que $Y$ a tendance à diminuer quand $X$ augmente (corrélation négative)
- On peut également introduire le *coefficient de corrélation* de $X$ et $Y$ (si $X$ et $Y$ sont de variance non nulle) :
$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

**Proposition 50.** ***(formule de König-Huygens)*** *Pour $X, Y \in L^2$, on a :*
$$\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

**Corollaire 6.** *Si $X$ et $Y$ sont indépendantes, alors $X$ et $Y$ sont décorrélées.*

**Remarque :** Comme on l'a déjà vu, la réciproque est fausse. Voici un autre exemple : si $X$ et $Y$ sont indépendantes, avec
$$\mathbb{P}(X=0) = \mathbb{P}(X=1) = \mathbb{P}(Y=-1) = \mathbb{P}(Y=1) = \frac{1}{2},$$
$X$ et $Z = XY$ ne sont pas indépendantes mais $\mathbb{E}(XZ) = \mathbb{E}(X)\mathbb{E}(Z)$, donc $\mathrm{Cov}(X,Y) = 0$.

**Proposition 51.** *Si $X, Y \in L^2$, on a :*
$$\mathbb{V}(X+Y) = \mathbb{V}(X) + 2\mathrm{Cov}(X,Y) + \mathbb{V}(Y)$$

**Remarque :** Plus généralement, $\mathbb{V}(aX+bY) = a^2\mathbb{V}(X) + 2ab\mathrm{Cov}(X,Y) + b^2\mathbb{V}(Y)$.

**Corollaire 7.** *Si $X$ et $Y$ sont indépendantes, alors $\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y)$.*

**Remarques :**
- Réciproquement, $\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y)$ implique que $X$ et $Y$ sont décorrélées, mais, comme on l'a vu, pas nécessairement indépendantes.
- Lorque $X$ et $Y$ ne sont pas décorrélées, on a toujours la majoration :
$$\left|\mathrm{Cov}(X,Y)\right| \leqslant \sigma(X)\sigma(Y)$$

**Exercice 8.** Montrer cette majoration et donner une condition nécessaire et suffisante d'égalité.

Les résultats précédents se généralisent au cas de $n$ variables aléatoires $X_1, \dots, X_n$ décorrélées 2 à 2, ce qui permet, entre autre de retrouver facilement la variance d'une loi binomiale.

**Proposition 52.** *Soit $(X_i)_{1\leqslant i\leqslant n}$ une famille finie de variables aléatoires réelles discrètes dans $L^2$. Alors :*
$$\mathbb{V}\left(\sum_{i=1}^n X_i\right) = \sum_{(i,j)\in[\![1,n]\!]^2} \mathrm{Cov}(X_i,X_j) = \sum_{i=1}^n \mathbb{V}(X_i) + 2\sum_{\substack{(i,j)\in[\![1,n]\!]^2 \\ i<j}} \mathrm{Cov}(X_i,X_j).$$

**Corollaire 8.** *Soit $(X_i)_{1\leqslant i\leqslant n}$ une famille finie de variables aléatoires discrètes réelles dans $L^2$, et* deux à deux *décorrélées. Alors :*
$$\mathbb{V}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{V}(X_i).$$

## 4.6 Inégalités probabilistes et applications

**Proposition 53.** ***(inégalités de Markov)***
*Soit $X$ une variable aléatoire discrète réelle dans $L^2$. Pour tout $t > 0$ :*

$$P\big(|X| \geqslant t\big) \leqslant \frac{\mathbb{E}\big(|X|\big)}{t} \qquad et \qquad P\big(|X| \geqslant t\big) \leqslant \frac{\mathbb{E}\big(|X|^2\big)}{t^2}$$

**Proposition 54.** ***(inégalité de Bienaymé-Tchebychev)***
*Soit $X$ une variable aléatoire discrète réelle dans $L^2$. Alors*

$$\forall\, \varepsilon > 0\,, \;\; P\big(|X - \mathbb{E}(X)| \geqslant \varepsilon\big) \leqslant \frac{\mathbb{V}(X)}{\varepsilon^2}.$$

**Remarque :** L'inégalité de Bienaymé-Tchebychef confirme l'interprétation intuitive de ce qu'est la variance : une mesure de *dispersion*, qui évalue l'écart à la moyenne. La probabilité que l'écart de $X$ à $\mathbb{E}(X)$ soit supérieur à $\varepsilon$ est d'autant plus petite que $\mathbb{V}(X)$ est faible.

**Proposition 55.** ***(loi faible des grands nombres)***
*Soient $(X_n)_n$ une suite de variables aléatoires discrètes réelles i.i.d, dans $L^2$ et d'espérance $m$. Soient $S_n = \sum_{k=1}^{n} X_k$ pour tout $n \geqslant 1$. Alors on a pour tout $\varepsilon > 0$ :*

$$P\left(\left|\frac{1}{n}S_n - m\right| \geqslant \varepsilon\right) \underset{n\to+\infty}{\longrightarrow} 0$$

**Remarque :** le résultat précédent stipule que pour une expérience aléatoire fournissant un résultat réel, un très grand nombre de répétitions de l'expérience donne presque sûrement un résultat moyen correspondant à l'espérance de la variable aléatoire modélisant les résultats possible de l'expérience. Cela donne une justification *a posteriori* d'une approche fréquentiste : si on obtient $m$ comme résultat moyen, il est cohérent de modéliser l'expérience par une variable aléatoire dont l'espérance est $m$.

## 4.7 Fonctions génératrices

**Définition 31.** Soit $X$ une variable aléatoire discrète à valeurs dans $\mathbb{N}$. On appelle *série génératrice* de $X$ la série entière réelle $\sum a_n x^n$, avec $a_n = \mathbb{P}(X = n)$. On note $G_X$ la somme de cette série et on l'appelle *fonction génératrice* de $X$ :

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

**Proposition 56.** *La série génératrice de $X$ converge normalement sur $D_f(0,1)$, d'où :*
- *le rayon $R$ de convergence vérifie $R \geqslant 1$*
- *$G_X$ est définie continue sur $[-1,1]$ (au moins).*

**Remarque :** Si $X(\Omega)$ est fini, ou même presque fini, c'est-à-dire si $E = \{n \in \mathbb{N},\ \mathbb{P}(X = n) > 0\}$ est fini, $G_X$ est une fonction polynômiale de degré $\max E$.

En vertu de l'unicité d'un développement en série entière, la loi d'une variable aléatoire à valeurs dans $\mathbb{N}$ est caractérisée par sa fonction génératrice, *ie* deux variables aléatoire ayant même fonction génératrice suivent la même loi :

**Proposition 57.** *Soient $X$ et $Y$ deux variables aléatoires discrètes sur $(\Omega, \mathcal{A}, \mathbb{P})$, et à valeurs dans $\mathbb{N}$. On suppose qu'il existe $r > 0$, tel que pour tout $t \in ]-r, r[$, $G_X(t) = G_Y(t)$. On a alors :*

$$\forall n \in \mathbb{N}, \quad \mathbb{P}(X = n) = \mathbb{P}(Y = n)$$

Remarquons que la fonction génératrice $G_X$ d'une variable aléatoire est de classe $\mathcal{C}^\infty$ sur $]-1,1[$. Qu'en est-il en 1 ?

**Proposition 58.** *Soit $X$ une variable aléatoire discrète dans $\mathbb{N}$ et $G_X$ sa fonction génératrice. Alors $X \in L^1$ si, et seulement si, $G_X$ est dérivable à gauche en 1 et on a dans ce cas*

$$\mathbb{E}(X) = G'_X(1).$$

**Démonstration :**

- Supposons $X \in L^1$, donc que $\sum n\mathbb{P}(X=n)$ converge. Cela implique que la série entière $\sum n\mathbb{P}(X=n)t^{n-1}$ converge normalement, donc uniformément, sur $[-1,1]$. Comme il s'agit de la série dérivée de la série génératrice $\sum \mathbb{P}(X=n)t^n$, on peut appliquer le théorème de dérivation d'une somme de série de fonctions : $G_X$ est de classe $\mathcal{C}^1$ sur $[-1,1]$, avec

$$\forall t \in [-1,1], \quad G'_X(t) = \sum n\mathbb{P}(X=n)t^{n-1}$$

  En particulier, $G_X$ est dérivable à gauche en 1, et $G'_X(1) = \mathbb{E}(X)$.
- Supposons que $G_X$ est dérivable à gauche en 1. Pour $t \in [0,1[$, on a

$$\frac{G_X(t) - G_X(1)}{t-1} = \sum_{n=0}^{+\infty} \mathbb{P}(X=n)\left(\frac{t^n-1}{t-1}\right) = \sum_{n=0}^{+\infty} \mathbb{P}(X=n)(1+t+\cdots+t^{n-1})$$

  Ce taux d'accroissement définit une fonction croissante sur $[0,1[$. Comme par définition de la dérivabilité, il tend vers $G'_X(1)$ (dérivée à gauche), il est majoré par $G'_X(1)$. Fixons alors $N \in \mathbb{N}$. On a, pour tout $t \in [0,1[$ :

$$\sum_{n=0}^{N} \mathbb{P}(X=n)(1+t+\cdots+t^{n-1}) \leqslant \sum_{n=0}^{+\infty} \mathbb{P}(X=n)(1+t+\cdots+t^{n-1}) \leqslant G'_X(1)$$

  En faisant tendre $t$ vers 1 dans le membre de gauche, on obtient

$$\sum_{n=0}^{N} n\mathbb{P}(X=n) \leqslant G'_X(1)$$

  Cette majoration étant vérifiée pour tout $N \in \mathbb{N}$, on a montré ainsi que la suite des sommes partielles de la série positive $\sum n\mathbb{P}(X=n)$ est bornée, et cela prouve donc sa convergence, autrement dit que $X \in L^1$.

**Remarques :**

- $G'_X(1)$ désigne ici la dérivée à gauche en 1 dans le cas particulier où $G_X$ n'est définie que sur $[-1,1]$, c'est-à-dire lorsque le rayon de convergence $R$ de la série génératrice vérifie $R = 1$.
- Dans ces conditions, $G_X$ est en fait alors nécessairement $\mathcal{C}^1$ sur $[-1,1]$.
- Si le rayons $R$ vérifie $R > 1$, on a directement $G_X$ de classe $\mathcal{C}^\infty$ au voisinage de 1.

**Proposition 59.** *Soit $X$ une variable aléatoire discrète à valeurs dans $\mathbb{N}$ et $G_X$ sa fonction génératrice. Alors $X \in L^2$, si et seulement si $G_X$ est deux fois dérivable en 1 et on a dans ce cas :*

$$\mathbb{E}(X^2) = G''_X(1) + G'_X(1).$$

**Corollaire 9.** *Si $X^2$ est d'espérance finie, on a :*

$$\mathbb{V}(X) = G''_X(1) + G'_X(1)\big(1 - G'_X(1)\big).$$

**Remarque :** Cette formule n'est pas à apprendre par coeur mais doit pouvoir être retrouvée facilement !

On rappelle que si $X$ et $Y$ sont indépendantes, $f(X)$ et $f(Y)$ sont également indépendantes. Appliquée à la fonction $f_t : n \mapsto t^n$, cette propriété conduit au résultat suivant :

**Proposition 60.** *Soient $X$ et $Y$ deux variables aléatoires discrètes indépendantes à valeurs dans $\mathbb{N}$. Pour tout $t \in [-1,1]$ :*

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

**Remarque :** Ce résultat se généralise par récurrence à toute somme finie de variables aléatoires discrètes indépendantes $X_1, \dots, X_n$ à valeur dans $\mathbb{N}$ :

$$G_{X_1+\cdots+X_n} = \prod_{i=1}^{n} G_{X_i}$$

Terminons par les fonctions génératrices des lois usuelles, qui doivent pouvoir être retrouvées rapidement :

**Proposition 61.**
- *Pour* $X \sim \mathcal{B}(n,p)$ *et* $q = 1-p$, $G_X(t) = (q+pt)^n$
- *Pour* $X \sim \mathcal{G}(p)$ *et* $q = 1-p$, $G_X(t) = \dfrac{pt}{1-qt}$.
- *Pour* $X \sim \mathcal{P}(\lambda)$, $G_X(t) = \mathrm{e}^{\lambda(t-1)}$.

## 4.8 Compléments

**Proposition 62.** ***(la loi géométrique est sans mémoire)***
*Soit $X$ une variable aléatoire discrète à valeurs dans $\mathbb{N}^*$. Alors $X$ suit une loi géométrique si, et seulement si, pour tout $(n,k) \in \mathbb{N}^2$ :*

$$P(X = n+k \,|\, X > n) = P(X = k).$$

**Proposition 63.** ***(la loi de Poisson approche la loi Binomiale)***
*Soit $(X_n)_n$ une suite de variables aléatoires discrètes telle que $X_n \sim \mathcal{B}(n, p_n)$, avec $np_n \to \lambda$. Alors, pour tout $k \in \mathbb{N}$,*

$$\lim_{n\to+\infty} P(X_n = k) = \mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!}.$$

Ce résultat justifie qu'on appelle parfois la loi de Poisson la loi des événements rares. Supposons par exemple qu'on sache qu'en moyenne $\lambda$ noyaux d'atomes parmi $N$ (très grand) se désintègrent sur une période $[0,T]$. Un élément radioactif a donc une probabilité très faible $p = \dfrac{\lambda}{N}$ de se désintégrer. Le nombre de désintégrations sur $[0,T]$ peut être modélisé par une loi Binomiale $\mathcal{B}(n,p)$, ce qui est très bien approché par la loi de Poisson $\mathcal{P}(\lambda)$.