PROBABILITÉS

I. Préliminaires.

1. Ensembles dénombrables.

Ce paragraphe a déjà été traité dans le poly sur les suites. Je rappelle les définitions et résultats essentiels.

Un ensemble E est dit **dénombrable** s'il est en bijection avec $\mathbb{N}$, c'est-à-dire s'il peut être décrit en extension sous la forme $\{x_i \ ; \ i \in \mathbb{N}\}$ avec des x_i distincts.

Exemples: les ensembles $\mathbb{N}$, $\mathbb{N}^*$, $\mathbf{Z}$, $\mathbb{Q}$ et $\mathbb{N}^2$ sont dénombrables.

De façon plus générale, tout produit cartésien d'un nombre fini d'ensembles dénombrables est dénombrable.

Toute partie infinie de $\mathbb{N}$ est dénombrable, autrement dit: toute partie de $\mathbb{N}$ est, soit finie (si elle est majorée), soit dénombrable (sinon).

Un ensemble E est alors dit **au plus dénombrable** s'il est fini ou dénombrable, c'est-à-dire s'il est en bijection avec une partie de $\mathbb{N}$, soit encore s'il peut être décrit en extension sous la forme $E = \{x_i \ ; \ i \in I\}$, où I est une partie de $\mathbb{N}$ et les x_i sont distincts. Une telle écriture sera appelée une **énumération** de l'ensemble E.

Toute partie d'un ensemble dénombrable est au plus dénombrable.

Toute union au plus dénombrable d'ensembles dénombrables est dénombrable.

Les ensembles $\mathbb{R}$ et $\{0,1\}^{\mathbb{N}}$ sont infinis non dénombrables.

2. Familles sommables de réels positifs.

Les résultats de ce paragraphe seront admis et ils seront utilisés dans les démonstrations de cours de ce chapitre.

Leur usage doit être strictement réservé au contexte probabiliste.

Dans ce paragraphe, nous nous placerons dans l'ensemble $[0,+\infty] = \overline{\mathbb{R}_+} = \mathbb{R}_+ \cup \{+\infty\}$. Nous étendrons d'abord l'addition usuelle de $\mathbb{R}_+$ en convenant que

$$\forall x \in [0,+\infty] \qquad x + (+\infty) = (+\infty) + x = +\infty \ .$$

On conviendra aussi que $a \cdot (+\infty) = (+\infty) \cdot a = +\infty$ pour tout réel a **strictement** positif.

Nous étendrons aussi la relation d'ordre usuelle en convenant que

$$\forall x \in \mathbb{R}_+ \qquad x < +\infty \ .$$

Nous admettrons que, si $(x_i)_{i\in I}$ est une famille au plus dénombrable d'éléments de $[0,+\infty]$ (i.e. si l'ensemble I des indices est au plus dénombrable), il est possible de définir la somme de cette famille, notée $\displaystyle\sum_{i\in I} x_i$, qui est aussi un élément de $[0,+\infty]$.

Si l'ensemble d'indices I est fini, il n'y a là rien de bien nouveau, si ce n'est que la somme vaut $+\infty$ lorsque l'un des éléments x_i de la famille vaut lui-même $+\infty$.

Si l'ensemble d'indices I est infini dénombrable, on peut considérer une énumération $I = \{i_n \ ; \ n \in \mathbb{N}\}$ de cet ensemble (rappelons que les i_n doivent être distincts, autrement dit l'application $n \mapsto i_n$ doit être une bijection de $\mathbb{N}$ vers I) et se ramener à la notion de somme d'une série en posant

$$\sum_{i\in I} x_i = \begin{cases} \displaystyle\sum_{n=0}^{+\infty} x_{i_n} & \text{si cette série converge, les } x_{i_n} \text{ étant tous des réels positifs} \\ +\infty & \text{si la série diverge, ou si au moins un des } x_{i_n} \text{ vaut } +\infty \end{cases} .$$

Pour rendre rigoureuse une telle définition, il faut s'assurer que le résultat (i.e. la nature de la série $\sum_{n\geq 0} x_{i_n}$ et la valeur de sa somme en cas de convergence) ne dépend pas de l'énumération choisie de l'ensemble I. Cette propriété sera admise et a déjà été mentionnée à la fin du poly de cours sur les séries, puisqu'il y est affirmé (sans preuve) que toute série à termes positifs convergente est "commutativement convergente".

On admet aussi que, pour tout **découpage en paquets** $I = \bigsqcup_{n\in\mathbb{N}} I_n$, i.e. si on écrit I comme une réunion dénombrable disjointe de sous-ensembles I_n, on a alors l'égalité (dite **sommation par paquets**), dans $[0,+\infty]$:

$$\sum_{i\in I} x_i = \sum_{n=0}^{+\infty} \Big(\sum_{i\in I_n} x_i \Big) .$$

En particulier, si $I = \mathbb{N}^2$ (cas des **sommes doubles**), en découpant $\mathbb{N}^2 = \bigsqcup_{k\in\mathbb{N}} \big(\{k\}\times\mathbb{N}\big)$ ou bien $\mathbb{N}^2 = \bigsqcup_{l\in\mathbb{N}} \big(\mathbb{N}\times\{l\}\big)$, si $(a_{k,l})_{(k,l)\in\mathbb{N}^2}$ est une famille de **réels positifs**, on a, dans $[0,+\infty]$, l'égalité (**interversion de sommes**)

$$\sum_{(k,l)\in\mathbb{N}^2} a_{k,l} = \sum_{k=0}^{+\infty} \Big(\sum_{l=0}^{+\infty} a_{k,l} \Big) = \sum_{l=0}^{+\infty} \Big(\sum_{k=0}^{+\infty} a_{k,l} \Big) .$$

La famille $(x_i)_{i\in I}$ d'éléments de $[0,+\infty]$ est alors dite **sommable** si on a $\sum_{i\in I} x_i < +\infty$.

Mentionnons enfin la **croissance de la somme**:

si $\forall i \in I \quad 0 \leq x_i \leq y_i$, alors $\sum_{i\in I} x_i \leq \sum_{i\in I} y_i$ dans $[0,+\infty]$.

Dans la pratique, et **dans le cas de réels positifs**, *on se permettra de découper, calculer, majorer des sommes directement sans justification, et la finitude de la somme pourra être considérée comme une preuve de sommabilité de la famille.*

Exemples. • La famille $\left(\frac{1}{n}\right)_{n\in\mathbb{N}^*}$ n'est pas sommable, on a $\sum_{n\in\mathbb{N}^*} \frac{1}{n} = +\infty$.

• La famille $\big(\frac{1}{n^2}\big)_{n\in\mathbf{Z}^*}$ est sommable et on a $\sum_{n\in\mathbf{Z}^*} \frac{1}{n^2} = 2\sum_{n=1}^{+\infty} \frac{1}{n^2} = \frac{\pi^2}{3}$.

• La famille $\left(\frac{1}{2^p 3^q}\right)_{(p,q)\in\mathbb{N}^2}$ est sommable. En effet, on peut écrire (*cf.* "cas des sommes doubles" ci-dessus):

$$\sum_{(p,q)\in\mathbb{N}^2} \frac{1}{2^p 3^q} = \sum_{p=0}^{+\infty} \left(\sum_{q=0}^{+\infty} \frac{1}{2^p 3^q} \right) = \sum_{p=0}^{+\infty} \left(\frac{3}{2} \times \frac{1}{2^p} \right) = 3 < +\infty .$$

• La famille $\Big(\dfrac{1}{(k+l)^\alpha}\Big)_{(k,l)\in(\mathbb{N}^*)^2}$ est sommable si et seulement si $\alpha > 2$. On le voit en sommant par paquets, en posant $(\mathbb{N}^*)^2 = \bigsqcup_{n=2}^{+\infty} I_n$ avec $I_n = \big\{(k,l) \in (\mathbb{N}^*)^2 \mid k+l = n\big\}$. Comme $\mathrm{Card}(I_n) = n-1$, on a, dans $[0,+\infty]$

$$\sum_{(k,l)\in(\mathbb{N}^*)^2} \frac{1}{(k+l)^\alpha} = \sum_{n=2}^{+\infty} \Big(\sum_{(k,l)\in I_n} \frac{1}{(k+l)^\alpha} \Big) = \sum_{n=2}^{+\infty} \frac{n-1}{n^\alpha} \ ,$$

et il résulte du cours sur les séries que cette dernière somme est finie si et seulement si $\alpha > 2$.

II. Espaces probabilisés.

1. Notion d'espace probabilisable.

Soit Ω un ensemble que nous appellerons **univers**. Cet "univers" est censé rendre compte de toutes les issues possibles d'une expérience aléatoire. La plupart du temps, il ne sera pas explicité.

Définition. On appelle **tribu sur** Ω tout ensemble $\mathcal{A}$ de parties de Ω, i.e. toute partie $\mathcal{A}$ de l'ensemble $\mathcal{P}(\Omega)$, tel que:

(P1): $\Omega \in \mathcal{A}$;

(P2): $\forall A \in \mathcal{A} \quad \overline{A} \in \mathcal{A}$;

(P3): Pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{A}$, la réunion $\bigcup_{n=0}^{+\infty} A_n$ appartient à $\mathcal{A}$.

Les éléments de $\mathcal{A}$ sont appelés les **événements**. On dit alors que le couple $(\Omega, \mathcal{A})$ est un **espace probabilisable**.

Commentaires. Dans la propriété **(P2)**, la notation $\overline{A}$ représente le complémentaire $\Omega \setminus A$, parfois aussi noté A^c, et appelé **événement contraire** de A. La propriété **(P2)** est la **stabilité** de la tribu **par passage au complémentaire.** La propriété **(P3)** s'appelle "**stabilité par union dénombrable**".

Conséquences. De ces axiomes de définition découlent quelques propriétés évidentes, par exemple le fait que $\emptyset \in \mathcal{A}$, ou encore que **toute tribu est aussi stable par réunion finie, et par intersection finie ou dénombrable**.

Preuve. D'abord, $\emptyset = \overline{\Omega}$ *et* $\Omega \in \mathcal{A}$*, donc* $\emptyset \in \mathcal{A}$ *d'après* **(P2)**.

Si $(A_0, A_1, \cdots, A_n)$ *est une famille finie d'événements, en posant* $A_k = \emptyset$ *pour tout* $k > n$*, on a*

$$\bigcup_{k=0}^{n} A_k = \bigcup_{k=0}^{+\infty} A_k \in \mathcal{A} \qquad \text{d'après } \textbf{(P3)} \ .$$

Enfin, si $(A_k)_{k\in\mathbb{N}}$ *est une suite d'événements, alors* $\overline{A_k} \in \mathcal{A}$ *pour tout* k *d'après* **(P2)***, puis par les "lois de De Morgan",*

$$\bigcap_{k\in\mathbb{N}} A_k = \overline{\bigcup_{k\in\mathbb{N}} \overline{A_k}} \in \mathcal{A} \qquad \text{d'après } \textbf{(P3)} \ .$$

Idem pour une intersection finie.

Signalons aussi que, si A et B sont deux événements, i.e. $A \in \mathcal{A}$ et $B \in \mathcal{A}$, alors $A \setminus B \in \mathcal{A}$: *en effet,* $A \setminus B = A \cap \overline{B}$. La différence de deux événements en est encore un.

Une tribu est donc stable par toutes les opérations ensemblistes finies ou dénombrables. Donc, si $(A_i)_{i\in I}$ est une famille d'événements, l'ensemble d'indices I étant au plus dénombrable, alors les ensembles $\bigcup_{i\in I} A_i$ et $\bigcap_{i\in I} A_i$ sont des événements.

Exemples.

• Si Ω est un ensemble non vide, alors $\mathcal{A} = \{\emptyset, \Omega\}$ est la **tribu grossière** sur Ω.

• L'ensemble $\mathcal{P}(\Omega)$ est une tribu sur Ω.

• Si $\Omega = \mathbb{R}$ (ou, plus généralement, $\mathbb{R}^n$), il existe une "plus petite" tribu sur $\mathbb{R}$ contenant toutes les parties ouvertes de $\mathbb{R}$, on l'appelle **tribu borélienne** sur $\mathbb{R}$.

Rappels de vocabulaire. L'univers Ω est l'**événement certain**, alors que $\emptyset$ est l'**événement impossible**. Deux événements A et B sont dits **incompatibles** (vocabulaire probabiliste) s'ils sont **disjoints** (vocabulaire ensembliste), i.e. si $A \cap B = \emptyset$.

Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements, alors l'événement $\bigcup_{n=0}^{+\infty} A_n$ est réalisé si et seulement si **au moins un** des événements A_n est réalisé. Tandis que l'événement $\bigcap_{n=0}^{+\infty} A_n$ est réalisé si et seulement si **tous** les événements A_n sont réalisés.

Exercice. Que signifie la réalisation des événements $\bigcup_{n=0}^{+\infty}\Big(\bigcap_{k=n}^{+\infty} A_k\Big)$ et $\bigcap_{n=0}^{+\infty}\Big(\bigcup_{k=n}^{+\infty} A_k\Big)$?

Définition. Dans un espace probabilisable $(\Omega, \mathcal{A})$, on appelle **système complet d'événements** (**SCE**) toute famille $(A_i)_{i\in I}$ d'événements, indexée par un ensemble I au plus dénombrable, telle que

$$\bigcup_{i\in I} A_i = \Omega \qquad \text{et} \qquad \forall (i,j) \in I^2 \quad i \neq j \Longrightarrow A_i \cap A_j = \emptyset\,.$$

Cela signifie donc que l'univers Ω est la réunion **disjointe** des A_i, ce que l'on écrira aussi $\Omega = \bigsqcup_{i\in I} A_i$ (*Attention! Ce n'est pas une notation standard!*).

2. Notion de probabilité. Propriétés élémentaires.

C'est encore une définition axiomatique:

Définition. Soit $(\Omega, \mathcal{A})$ un espace probabilisable. On appelle **probabilité** sur $(\Omega, \mathcal{A})$ toute application $P : \mathcal{A} \to [0,1]$ telle que

(P1): $P(\Omega) = 1$;

(P2): Pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'événements incompatibles, $P\Big(\bigsqcup_{n=0}^{+\infty} A_n\Big) = \sum_{n=0}^{+\infty} P(A_n)$.

On dit alors que $(\Omega, \mathcal{A}, P)$ est un **espace probabilisé**.

Commentaire. La propriété **(P2)** affirme bien sûr la convergence de la série $\sum P(A_n)$. On la nomme σ-**additivité** ou encore **additivité dénombrable**.

Conséquences de cette définition.

• $\boxed{P(\emptyset) = 0.}$

En effet, en prenant $A_n = \emptyset$ pour tout n, d'après **(P2)**, *la série de terme général (constant) $P(\emptyset)$ doit converger, donc nécessairement $P(\emptyset) = 0$.*

• **Additivité finie**: Si $A_0, \cdots, A_n$ sont des événements **deux à deux incompatibles**, alors $P\Big(\bigsqcup_{k=0}^{n} A_k\Big) = \sum_{k=0}^{n} P(A_k)$.

En effet, il suffit d'appliquer **(P2)** *en posant $A_k = \emptyset$ pour tout $k > n$.*

Commentaire. On aura donc, pour toute famille $(A_i)_{i\in I}$ d'événements disjoints, indexée par un ensemble I **au plus dénombrable**, l'égalité $P\Big(\bigsqcup_{i\in I} A_i\Big) = \sum_{i\in I} P(A_i)$.

• **Probabilité de l'événement contraire.**

Si $A \in \mathcal{A}$, alors $\boxed{P(\overline{A}) = 1 - P(A)\ .}$

En effet, on applique l'additivité finie ci-dessus avec $\Omega = A \sqcup \overline{A}$.

• **Probabilité de la différence de deux événements.**

$$\forall (A, B) \in \mathcal{A}^2 \qquad P(A \setminus B) = P(A) - P(A \cap B)\ .$$

Cela résulte de l'égalité $A = (A \setminus B) \sqcup (A \cap B)$ et de l'additivité finie.

Dans le cas particulier où $B \subset A$, cette différence $A \setminus B$ peut être appelée "complémentaire de B dans A" et on a alors $P(A \setminus B) = P(A) - P(B)$.

• **Probabilité de la réunion de deux événements.**

$$\forall (A, B) \in \mathcal{A}^2 \qquad P(A \cup B) = P(A) + P(B) - P(A \cap B)\ .$$

En effet, $A \cup B = (A \cap B) \sqcup (A \setminus B) \sqcup (B \setminus A)$, cela résulte alors de la formule précédente.

• **Croissance.** Si $A \in \mathcal{A}$ et $B \in \mathcal{A}$, $\boxed{\text{si } A \subset B, \text{ alors } P(A) \leq P(B).}$

En effet, $B = A \sqcup (B \setminus A)$ (union disjointe), donc par la propriété d'additivité finie, $P(B) = P(A) + P(B \setminus A) \geq P(A)$.

Définitions. Un événement $A \in \mathcal{A}$ est dit **négligeable** (ou encore **quasi-impossible**) s'il est de probabilité nulle, i.e. si $P(A) = 0$. Il est dit **presque sûr** (ou encore **quasi-certain**) si $P(A) = 1$.

3. Propriétés de continuité monotone, et conséquences.

Dans tout ce paragraphe, $(\Omega, \mathcal{A}, P)$ est un espace probabilisé.

Théorème de continuité croissante. Soit $(A_n)_{n\in\mathbb{N}}$ une suite croissante d'événements, i.e. telle que, pour tout n, on ait $A_n \subset A_{n+1}$. On a alors

$$\lim_{n\to+\infty} P(A_n) = P\Big(\bigcup_{n=0}^{+\infty} A_n\Big)\ .$$

Preuve. Posons $B_0 = A_0$, puis $B_n = A_n \setminus A_{n-1}$ pour tout $n \in \mathbb{N}^*$.
Alors les B_n sont des événements, et ils sont deux à deux incompatibles. En effet, pour $n \geq 1$, on a $B_n = A_n \cap \overline{A_{n-1}}$ donc, si $p < q$, on a $B_p \subset A_p$ alors que $B_q \subset \overline{A_{q-1}} \subset \overline{A_p}$, donc $B_p \cap B_q = \emptyset$.
Ensuite, $\bigcup_{n=0}^{+\infty} A_n = \bigsqcup_{n=0}^{+\infty} B_n$: en effet, l'inclusion dans le sens indirect est immédiate puisque $B_n \subset A_n$ pour tout n. Puis, si $\omega \in \bigcup_{n=0}^{+\infty} A_n$, soit $n_0 = \min\{n \in \mathbb{N} \mid \omega \in A_n\}$, on a alors $\omega \in B_{n_0}$ donc $\omega \in \bigcup_{n=0}^{+\infty} B_n$, ce qui montre l'autre inclusion.
La suite $\big(P(A_n)\big)$ est croissante, et majorée par 1, elle converge donc, la série télescopique $\sum_{n\geq 1} \big(P(A_n) - P(A_{n-1})\big)$ est donc aussi convergente, de somme $\lim_{n\to+\infty} P(A_n) - P(A_0)$.
En utilisant la propriété de σ-additivité de la probabilité P, on obtient

$$\begin{aligned} P\Big(\bigcup_{n=0}^{+\infty} A_n\Big) &= P\Big(\bigsqcup_{n=0}^{+\infty} B_n\Big) = \sum_{n=0}^{+\infty} P(B_n) \\ &= P(A_0) + \sum_{n=1}^{+\infty} P(A_n \setminus A_{n-1}) \\ &= P(A_0) + \sum_{n=1}^{+\infty} \big(P(A_n) - P(A_{n-1})\big) \\ &= \lim_{n\to+\infty} P(A_n)\ . \end{aligned}$$

<u>Théorème de continuité décroissante</u>. Soit $(A_n)_{n\in\mathbb{N}}$ une suite décroissante d'événements, i.e. telle que, pour tout n, on ait $A_{n+1} \subset A_n$. On a alors

$$\lim_{n\to+\infty} P(A_n) = P\Big(\bigcap_{n=0}^{+\infty} A_n\Big)\ .$$

<u>Preuve.</u> Par passage au complémentaire. Posons $C_n = \overline{A_n}$ pour tout n, alors la suite (C_n) est croissante, donc $\lim_{n\to+\infty} P(C_n) = P\Big(\bigcup_{n=0}^{+\infty} C_n\Big)$ d'après le théorème de continuité croissante. Mais $P(C_n) = 1 - P(A_n)$ pour tout n et, par les lois de De Morgan,

$$P\Big(\bigcup_{n=0}^{+\infty} C_n\Big) = P\Big(\bigcup_{n=0}^{+\infty} \overline{A_n}\Big) = P\Big(\overline{\bigcap_{n=0}^{+\infty} A_n}\Big) = 1 - P\Big(\bigcap_{n=0}^{+\infty} A_n\Big)\ .$$

On conclut alors facilement.

<u>Conséquences</u>. Soit $(A_n)_{n\in\mathbb{N}}$ une suite quelconque d'événements, on a alors

$$\lim_{n\to+\infty} P\Big(\bigcup_{k=0}^{n} A_k\Big) = P\Big(\bigcup_{k=0}^{+\infty} A_k\Big) \qquad \text{et} \qquad \lim_{n\to+\infty} P\Big(\bigcap_{k=0}^{n} A_k\Big) = P\Big(\bigcap_{k=0}^{+\infty} A_k\Big)\ .$$

Preuve. Pour la première égalité, posons $B_n = \bigcup_{k=0}^{n} A_k$ *pour tout* n*, la suite d'événements* (B_n) *est alors croissante. Le théorème de continuité croissante montre donc que*

$$\lim_{n\to+\infty} P(B_n) = P\Big(\bigcup_{n=0}^{+\infty} B_n\Big) = P\Big(\bigcup_{n=0}^{+\infty} A_n\Big)$$

puisqu'il est immédiat que la réunion des B_n *est aussi la réunion des* A_n*. Preuve analogue pour la deuxième égalité.*

Théorème de sous-additivité. Si $(A_n)_{n\in\mathbb{N}}$ **est une suite quelconque d'événements, on a**

$$P\Big(\bigcup_{n=0}^{+\infty} A_n\Big) \leq \sum_{n=0}^{+\infty} P(A_n)\ .$$

Commentaire. Si la série à termes positifs $\sum P(A_n)$ est divergente, on conviendra que $\sum_{n=0}^{+\infty} P(A_n) = +\infty$, ainsi cette inégalité reste vraie dans $[0, +\infty]$.

Preuve. On le montre d'abord pour une réunion finie: si $A_0, \cdots, A_n$ *sont des événements, alors* $P\Big(\bigcup_{k=0}^{n} A_k\Big) \leq \sum_{k=0}^{n} P(A_k)$*, par récurrence sur* n*.*

- pour $n = 1$*,* $P(A_0 \cup A_1) = P(A_0) + P(A_1) - P(A_0 \cap A_1) \leq P(A_0) + P(A_1)$*.*

- si c'est vrai pour n *événements, alors*

$$\begin{aligned} P\Big(\bigcup_{k=0}^{n+1} A_k\Big) &= P\Big(\Big(\bigcup_{k=0}^{n} A_k\Big) \cup A_{n+1}\Big) \leq P\Big(\bigcup_{k=0}^{n} A_k\Big) + P(A_{n+1}) \\ &\leq \sum_{k=0}^{n} P(A_k) + P(A_{n+1}) = \sum_{k=0}^{n+1} P(A_k)\ . \end{aligned}$$

Dans le cas d'une réunion infinie dénombrable, il suffit de faire tendre n *vers l'infini dans l'inégalité que l'on vient d'obtenir, en utilisant la première des deux égalités obtenues comme "conséquences" ci-dessus.*

Commentaire. On peut reformuler de la façon suivante: si $(A_i)_{i\in I}$ est une famille d'événements, l'ensemble d'indices I étant au plus dénombrable, on a, dans $[0, +\infty]$, l'inégalité

$$P\Big(\bigcup_{i\in I} A_i\Big) \leq \sum_{i\in I} P(A_i)\ .$$

Conséquence. Toute réunion dénombrable d'événements négligeables est encore négligeable. Et, par passage au complémentaire, toute intersection dénombrable d'événements presque sûrs est un événement presque sûr.

Définition. Dans un espace probabilisé $(\Omega, \mathcal{A}, P)$, on appelle **système quasi-complet d'événements** toute famille $(A_i)_{i\in I}$ d'événements deux à deux incompatibles, indexée par un ensemble I au plus dénombrable, telle que $\sum_{i\in I} P(A_i) = 1$, i.e. $P\Big(\bigsqcup_{i\in I} A_i\Big) = 1$.

Autrement dit, les A_i "recouvrent l'univers à un ensemble négligeable près". Le complémentaire de la réunion des A_i est en effet de probabilité nulle.

4. <u>Conditionnement</u>.

<u>Définition.</u> Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Soient $A \in \mathcal{A}$ et $B \in \mathcal{A}$ deux événements avec $P(B) > 0$. On appelle **probabilité conditionnelle de A sachant B** le réel

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

On le note aussi $P(A|B)$.

<u>Proposition</u>. L'application P_B ainsi définie sur $\mathcal{A}$ est une probabilité sur l'espace probabilisable $(\Omega, \mathcal{A})$.

Preuve. Par croissance de la probabilité P, on a $P(A \cap B) \leq P(B)$, donc $P_B(A) \in [0,1]$. On a bien $P_B(\Omega) = \dfrac{P(\Omega \cap B)}{P(B)} = 1$ et, si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements incompatibles, les $A_n \cap B$ sont aussi incompatibles, donc

$$P_B\Big(\bigsqcup_{n=0}^{+\infty} A_n\Big) = \frac{P\Big(\big(\bigsqcup_{n=0}^{+\infty} A_n\big) \cap B\Big)}{P(B)} = \frac{P\Big(\bigsqcup_{n=0}^{+\infty}(A_n \cap B)\Big)}{P(B)}$$

$$= \frac{\displaystyle\sum_{n=0}^{+\infty} P(A_n \cap B)}{P(B)} = \sum_{n=0}^{+\infty} \frac{P(A_n \cap B)}{P(B)} = \sum_{n=0}^{+\infty} P_B(A_n) \ ,$$

ce qu'il fallait démontrer (la σ-additivité).

<u>Formule des probabilités composées.</u>

(a): Soient A_1 et A_2 deux événements avec $P(A_1) > 0$, alors

$$P(A_1 \cap A_2) = P_{A_1}(A_2) \cdot P(A_1) \ .$$

(b): Si $A_1, \cdots, A_n$ sont des événements tels que $P(A_1 \cap \cdots \cap A_{n-1}) > 0$, alors

$$P(A_1 \cap \cdots \cap A_n) = P_{A_1 \cap \cdots \cap A_{n-1}}(A_n) \cdot P_{A_1 \cap \cdots \cap A_{n-2}}(A_{n-1}) \cdots P_{A_1}(A_2) \cdot P(A_1) \ .$$

<u>Preuve.</u> La formule **(a)** *n'est qu'une réécriture de la définition d'une probabilité conditionnelle. Elle peut être considérée comme initialisation d'une récurrence pour prouver* **(b)**.

<u>Commentaire.</u> Cette formule des probabilités composées généralisée **(b)** correspond à la situation où une suite finie d'expériences aléatoires est réalisée, l'issue de chacune de ces expériences pouvant être influencée par les résultats des expériences précédentes. Elle se prête bien à l'illustration par un arbre pondéré mais, à votre niveau, la seule représentation d'un arbre de probabilité ne peut être considérée comme une preuve valide, un minimum de formalisation est nécessaire.

Formule des probabilités totales (FPT).

Si $(A_n)_{n\in\mathbb{N}}$ est un système quasi-complet d'événements, si B est un événement, alors la série $\sum P(B \cap A_n)$ converge, et on a

$$P(B) = \sum_{n=0}^{+\infty} P(B \cap A_n) = \sum_{n=0}^{+\infty} P(B|A_n)\, P(A_n)\ .$$

Commentaire 1. En toute rigueur, les probabilités conditionnelles $P(B|A_n)$ n'ont de sens que si les A_n sont de probabilité non nulle. Cependant, si pour certains entiers n, on a $P(A_n) = 0$, on adoptera la convention $P(B|A_n)\, P(A_n) = 0$ dans ce cas, et la FPT reste alors valable.

Commentaire 2. La formule est valable en particulier si (A_n) est un système complet d'événements.

Preuve. Posons $A = \bigsqcup_{n\in\mathbb{N}} A_n$, alors A est un événement presque sûr, i.e. $P(A) = 1$, donc $P(\overline{A}) = 0$. Comme $B = (B \cap A) \sqcup (B \cap \overline{A})$, avec $P(B \cap \overline{A}) \leq P(\overline{A})$ donc $P(B \cap \overline{A}) = 0$, on a donc

$$P(B) = P(B \cap A) = P\Big(B \cap \Big(\bigsqcup_{n\in\mathbb{N}} A_n\Big)\Big) = P\Big(\bigsqcup_{n\in\mathbb{N}} (B \cap A_n)\Big) = \sum_{n=0}^{+\infty} P(B \cap A_n)$$

en utilisant la σ-additivité, ce qui donne la première égalité. La deuxième égalité résulte de la formule des probabilités composées (ou de la convention adoptée dans le cas où $P(A_n) = 0$).

Commentaire 3. On peut reformuler ce théorème de la façon suivante: si I est un ensemble au plus dénombrable, si $(A_i)_{i\in I}$ est un système quasi-complet d'événements, alors la famille $\big(P(B \cap A_i)\big)_{i\in I}$ est sommable, et on a

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

toujours avec la convention $P(B|A_i)\, P(A_i) = 0$ si $P(A_i) = 0$.

Formule de Bayes.

Soit $B \in \mathcal{A}$ un événement de probabilité non nulle, soit $(A_n)_{n\in\mathbb{N}}$ un système quasi-complet d'événements. On a alors, pour tout $n \in \mathbb{N}$,

$$P(A_n|B) = \frac{P(B|A_n)\, P(A_n)}{P(B)} = \frac{P(B|A_n)\, P(A_n)}{\displaystyle\sum_{k=0}^{+\infty} P(B|A_k)\, P(A_k)}\ .$$

Cette formule reste vraie si certains des A_k sont de probabilité nulle, en convenant alors que le produit $P(B|A_k)\, P(A_k)$ est nul.

Preuve. La première égalité est une conséquence immédiate de la définition des probabilités conditionnelles si $P(A_n) > 0$, elle résulte de la convention adoptée si $P(A_n) = 0$. La deuxième égalité résulte alors de la FPT.

5. Événements indépendants.

Définition. Deux événements A et B d'un espace probabilisé $(\Omega, \mathcal{A}, P)$ sont dits **indépendants** si on a $P(A \cap B) = P(A) \cdot P(B)$.

Commentaire. Si $P(B) > 0$, l'indépendance de A et B équivaut à $P(A|B) = P(A)$. Autrement dit, la réalisation ou non de l'événement B n'influe pas sur la probabilité de réalisation de l'événement A. Il faut toutefois se garder d'interpréter cette indépendance comme une absence de lien de cause à effet, puisque deux événements d'un même espace probabilisable $(\Omega, \mathcal{A})$ peuvent être indépendants pour une certaine probabilité P sans l'être pour une autre probabilité P'.

Définition. Soit $(A_1, \cdots, A_n)$ une famille finie d'événements. On dit qu'ils sont **indépendants** si, pour toute partie I non vide de l'intervalle $[\![1, n]\!]$, on a

$$P\Big(\bigcap_{i\in I} A_i\Big) = \prod_{i\in I} P(A_i)\ .$$

On parle parfois d'événements "mutuellement" indépendants (pour ne pas confondre avec l'indépendance deux à deux).

Attention! Il ne suffit pas d'imposer la relation $P\Big(\bigcap_{i=1}^{n} A_i\Big) = \prod_{i=1}^{n} P(A_i)$: cette dernière égalité serait par exemple vraie dès que l'un des événements A_i, disons A_n, est impossible (les deux membres de l'égalité seraient alors nuls), cela ne garantit pourtant pas l'indépendance de la sous-famille $(A_1, \cdots, A_{n-1})$. On désire en effet avoir le résultat:

Si une famille d'événements est indépendante, alors toute sous-famille l'est encore. *Avec la définition choisie, cette propriété de transmission aux sous-familles est trivialement vérifiée.*

En particulier, **l'indépendance "mutuelle" des événements A_1, $\cdots$, A_n entraîne leur indépendance deux à deux, la réciproque étant fausse.** *Pour cette réciproque, un contre-exemple est donné dans le poly de révisions sur le cours de 1ère année.*

Propriété. Si A et B sont indépendants, il en est de même de A et $\overline{B}$, et bien sûr aussi de $\overline{A}$ et $\overline{B}$, ou encore de $\overline{A}$ et B. Plus généralement, **si A_1, $\cdots$, A_n sont des événements indépendants, si B_1, $\cdots$, B_n sont des événements tels que, pour tout i, on ait $B_i = A_i$ ou $B_i = \overline{A_i}$, alors les événements B_1, $\cdots$, B_n sont indépendants.**

Preuve. Soient A_1, $\cdots$, A_n des événements indépendants. Il suffit de montrer que l'indépendance est conservée si l'on remplace l'un des événements, disons A_1, par son contraire, il conviendra ensuite d'itérer ce raisonnement si plusieurs événements sont remplacés par leur contraire.

Posons donc $B_1 = \overline{A_1}$ et $B_k = A_k$ pour $k \in [\![2, n]\!]$. Soit I une partie non vide de $[\![1, n]\!]$.

- si $1 \notin I$, alors $P\Big(\bigcap_{i\in I} B_i\Big) = P\Big(\bigcap_{i\in I} A_i\Big) = \prod_{i\in I} P(A_i) = \prod_{i\in I} P(B_i)$ puisque les A_i sont indépendants.

- si $1 \in I$, posons $J = I \setminus \{1\}$ alors, comme $\Omega = A_1 \sqcup \overline{A_1}$, et $B_i = A_i$ pour $i \in J$,

$$P\Big(\bigcap_{i\in J} B_i\Big) \;=\; P\Big(\bigcap_{i\in J} A_i\Big) \;=\; P\Big(A_1 \cap \Big(\bigcap_{i\in J} A_i\Big)\Big) + P\Big(\overline{A_1} \cap \Big(\bigcap_{i\in J} A_i\Big)\Big)$$

$$= \quad P\Big(\bigcap_{i\in I} A_i\Big) + P\Big(\bigcap_{i\in I} B_i\Big)\ .$$

Donc

$$\begin{aligned}
P\Big(\bigcap_{i\in I} B_i\Big) &= P\Big(\bigcap_{i\in J} A_i\Big) - P\Big(\bigcap_{i\in I} A_i\Big) \\
&= \prod_{i\in J} P(A_i) - \prod_{i\in I} P(A_i) \qquad\qquad \textit{par indépendance des } A_i \\
&= \prod_{i\in J} P(A_i) - P(A_1)\ \prod_{i\in J} P(A_i) \qquad \textit{comme } I = J\sqcup\{1\} \\
&= \big(1-P(A_1)\big)\ \prod_{i\in J} P(A_i) \\
&= P(B_1)\ \prod_{i\in J} P(B_i) \\
&= \prod_{i\in I} P(B_i)\ ,
\end{aligned}$$

ce qui achève la démonstration (et le lecteur).

De façon plus générale, si $(A_i)_{i\in I}$ est une famille au plus dénombrable d'événements, on dit qu'ils sont (mutuellement) **indépendants** si, pour tout partie finie J de I, on a $P\Big(\bigcap_{i\in J} A_i\Big) = \prod_{i\in J} P(A_i)$.

6. Cas des univers finis ou dénombrables. Si Ω est un ensemble fini, on choisit en général pour tribu $\mathcal{A} = \mathcal{P}(\Omega)$, une probabilité P sur l'espace probabilisable $\big(\Omega, \mathcal{P}(\Omega)\big)$ est alors déterminée par la donnée d'une famille finie $(p_\omega)_{\omega\in\Omega}$ de réels positifs tels que $\sum_{\omega\in\Omega} p_\omega = 1$ (ce que l'on appelle une **distribution de probabilités** sur Ω), en posant $P(A) = \sum_{\omega\in A} p_\omega$ pour tout $A \in \mathcal{P}(\Omega)$, notamment $P\big(\{\omega\}\big) = p_\omega$ pour tout événement élémentaire $\{\omega\}$. On retrouve ici la situation du programme de première année.

Proposition. Si l'univers Ω est au plus dénombrable, une probabilité P sur l'espace probabilisable $\big(\Omega, \mathcal{P}(\Omega)\big)$ est déterminée par la donnée d'une "distribution de probabilités" sur Ω, i.e. d'une famille $(p_\omega)_{\omega\in\Omega}$ de réels positifs tels que $\sum_{\omega\in\Omega} p_\omega = 1$. On pose alors $P(A) = \sum_{\omega\in A} p_\omega$ pour toute partie A de Ω, et on a en particulier $P\big(\{\omega\}\big) = p_\omega$ pour tout $\omega \in \Omega$.

Preuve. • *Si P est une probabilité sur Ω, les nombres $p_\omega = P\big(\{\omega\}\big)$ forment une famille de réels positifs, dont la somme vaut 1 grâce aux propriétés* **(P1)** *et* **(P2)** *puisque*

$$1 = P(\Omega) = P\Big(\bigsqcup_{\omega\in\Omega} \{\omega\}\Big) = \sum_{\omega\in\Omega} P\big(\{\omega\}\big) = \sum_{\omega\in\Omega} p_\omega$$

par σ-additivité (il s'agit bien d'une réunion finie ou dénombrable disjointe).

• *Inversement, soit* $(p_\omega)_{\omega\in\Omega}$ *une distribution de probabilités sur* Ω*, il faut montrer que l'application* $P : \mathcal{P}(\Omega) \to \mathbb{R}$ *définie par* $P(A) = \sum_{\omega\in A} p_\omega$ *est bien une probabilité sur l'espace probabilisable* $\big(\Omega, \mathcal{P}(\Omega)\big)$*. On a déjà* $P(A) \in [0, +\infty]$ *pour tout* $A \in \mathcal{P}(\Omega)$ *comme somme de réels positifs. Puis* $P(\Omega) = \sum_{\omega\in\Omega} p_\omega = 1$*, ce qui prouve la propriété* **(P1)***. Par addition d'inégalités, on déduit au passage que, si* $A \in \mathcal{P}(\Omega)$*, alors*

$$P(A) = \sum_{\omega\in A} p_\omega = \sum_{\omega\in\Omega} p_\omega \mathbb{1}_A(\omega) \leq \sum_{\omega\in\Omega} p_\omega = 1$$

par "croissance de la somme", ce qui montre que P *est bien à valeurs dans* $[0, 1]$*. Enfin, si* $(A_n)_{n\in\mathbb{N}}$ *est une suite d'événements incompatibles, soit* $A = \bigsqcup_{n\in\mathbb{N}} A_n$*, par sommation par paquets, on obtient*

$$P\Big(\bigsqcup_{n=0}^{+\infty} A_n\Big) = P(A) = \sum_{\omega\in A} p_\omega = \sum_{n=0}^{+\infty} \Big(\sum_{\omega\in A_n} p_\omega\Big) = \sum_{n=0}^{+\infty} P(A_n)\ ,$$

soit la propriété **(P2)**.

En revanche, si l'univers Ω est infini non dénombrable, la situation est plus compliquée. Il est impossible alors de définir des probabilités "intéressantes" sur l'ensemble $\mathcal{P}(\Omega)$ tout entier, ce qui motive l'introduction de la notion de tribu, il y aura donc des parties de Ω que l'on ne pourra pas qualifier d'événements. De plus, les événements élémentaires $\{\omega\}$ (si ce sont des événements!) ont bien souvent une probabilité nulle, ce qui n'est pas incompatible avec la propriété **(P2)** puisque l'écriture $\Omega = \bigsqcup_{\omega\in\Omega} \{\omega\}$ n'est plus une réunion dénombrable.
Tout ceci est bien difficile à expliquer sans sortir très largement du cadre du programme.

III. Variables aléatoires discrètes.

1. Définition.

Définition. Soit $(\Omega, \mathcal{A})$ un espace probabilisable. Une **variable aléatoire discrète** sur $(\Omega, \mathcal{A})$ est une application X définie sur Ω, dont l'image $X(\Omega)$ est au plus dénombrable, et telle que l'image réciproque de tout singleton de $X(\Omega)$ est un événement, i.e. appartient à la tribu $\mathcal{A}$.

Commentaire. Cette dernière condition, assez formelle, s'écrit

$$\forall x \in X(\Omega) \qquad X^{-1}\big(\{x\}\big) \in \mathcal{A}\ .$$

Mais nous utiliserons d'autres notations. L'événement $X^{-1}\big(\{x\}\big) = \big\{\omega \in \Omega \mid X(\omega) = x\big\}$ sera couramment noté $\{X = x\}$ ou encore $(X = x)$.

Commentaire. Conformément à votre programme, nous ne considérerons dans ce cours que des variables aléatoires (en abrégé v.a.) "discrètes". Il existe d'autres types de variables aléatoires, notamment celles dites "à densité", vous avez sans doute entendu parler de variables aléatoires réelles "suivant une loi normale (gaussienne) ou une loi exponentielle", nous n'en parlerons pas ici.

Proposition. Si X **est une variable aléatoire discrète sur** $(\Omega, \mathcal{A})$**, et si** U **est une partie de l'ensemble-image** $X(\Omega)$**, alors** $X^{-1}(U)$ **est un événement.**

Preuve. *Comme $X(\Omega)$ est au plus dénombrable, il en est de même de U qui en est un sous-ensemble. L'ensemble $X^{-1}(U)$ est alors une réunion finie ou dénombrable d'événements puisque*

$$X^{-1}(U) = \bigcup_{u\in U} X^{-1}\big(\{u\}\big) = \bigcup_{u\in U} \{X = u\} \ ,$$

donc $X^{-1}(U) \in \mathcal{A}$ (stabilité d'une tribu par réunion au plus dénombrable).

Commentaire. Ici aussi, l'événement $X^{-1}(U) = \big\{\omega \in \Omega \mid X(\omega) \in U\big\}$ sera noté $\{X \in U\}$, ou encore $(X \in U)$.

Remarque. La plupart des variables aléatoires étudiées dans ce cours sont à valeurs réelles, i.e. $X(\Omega) \subset \mathbb{R}$, on rencontrera toutefois lors de l'étude des couples ou n-uplets de variables aléatoires quelques "variables aléatoires vectorielles", ici à valeurs dans $\mathbb{R}^2$ ou $\mathbb{R}^n$.

Dans le cas d'une variable aléatoire réelle X, si $x \in \mathbb{R}$ est un réel, on considérera souvent des événements tels que

$$(X \geq x) = \{X \geq x\} = \big\{\omega \in \Omega \mid X(\omega) \geq x\big\} = X^{-1}\big([x, +\infty[\big) \ ,$$

c'est bien un événement d'après ce qui précède, puisque cet ensemble est aussi l'image réciproque par X de la partie $X(\Omega) \cap [x, +\infty[$ de $X(\Omega)$. On rencontrera aussi des $\{X > x\}$, $\{X < x\}$, $\{X \leq x\}$, écrits indifféremment avec des accolades ou des parenthèses.

Notion de fonction d'une variable aléatoire. Si X est une v.a. discrète sur $(\Omega, \mathcal{A})$, et si f est une application définie sur $X(\Omega)$, on peut considérer l'application $Y = f \circ X$, alors définie sur Ω. Alors Y est une variable aléatoire discrète sur $(\Omega, \mathcal{A})$. *En effet, l'ensemble-image $Y(\Omega) = f\big(X(\Omega)\big)$ est au plus dénombrable et, si $y \in Y(\Omega)$, alors $Y^{-1}\big(\{y\}\big) = X^{-1}(U)$ où $U = f^{-1}\big(\{y\}\big)$ est une partie de $X(\Omega)$, donc $Y^{-1}\big(\{y\}\big) \in \mathcal{A}$ d'après la proposition ci-dessus.* Cette nouvelle variable aléatoire Y est couramment notée $Y = f(X)$, et on dit que c'est une **fonction de la variable aléatoire** X.

2. Loi d'une variable aléatoire discrète.

Proposition et définition. Soit X une variable aléatoire discrète sur un espace probabilisé $(\Omega, \mathcal{A}, P)$. Pour toute partie A de $X(\Omega)$, on pose $P_X(A) = P(X \in A)$. L'application P_X ainsi définie est alors une probabilité sur l'espace probabilisable $\Big(X(\Omega), \mathcal{P}\big(X(\Omega)\big)\Big)$, appelée loi de la variable X.

Preuve. *Comme l'ensemble $X(\Omega)$ est au plus dénombrable, il suffit de montrer que la famille $(p_x)_{x\in X(\Omega)}$, où l'on pose $p_x = P_X\big(\{x\}\big) = P(X = x)$, est une distribution de probabilités sur $X(\Omega)$. Ce sont clairement des réels positifs et on a bien, par σ-additivité de P et par le fait que $\big(\{X = x\}\big)_{x\in X(\Omega)}$ est un système complet d'événements,*

$$\sum_{x\in X(\Omega)} p_x = \sum_{x\in X(\Omega)} P\big(\{X = x\}\big) = P\Big(\bigsqcup_{x\in X(\Omega)} \{X = x\}\Big) = P(\Omega) = 1 \ .$$

Commentaire 1. En fait, une variable aléatoire discrète permet de construire un nouvel espace probabilisé, ou "univers-image" fini ou dénombrable, qui est $\Big(X(\Omega), \mathcal{P}\big(X(\Omega)\big), P_X\Big)$, et tous les calculs ultérieurs (espérance, variance) seront faits à partir de cet univers-image, et se ramèneront donc à des sommes finies ou "infinies dénombrables", i.e. à des sommes de séries.

Commentaire 2. Dans les exercices et problèmes de probabilités, l'univers $(\Omega, \mathcal{A}, P)$, rendant compte de toutes les issues possibles d'une expérience aléatoire, ne sera presque jamais explicité. On se contentera d'étudier des variables aléatoires qui seront données par une loi, et ce qui précède montre que "se donner une loi de variable aléatoire" revient précisément à se donner un ensemble-image au plus dénombrable $X(\Omega)$, et une distribution de probabilités sur cet ensemble $X(\Omega)$. Remarquons aussi que l'ensemble-image $X(\Omega)$ peut comporter des valeurs atteintes avec une probabilité nulle, ces valeurs pourront être omises sans rien modifier aux calculs "intéressants" (espérance, variance), comme nous le verrons bientôt par exemple dans le cas de la loi géométrique.

Commentaire 3. Enfin, une distribution de probabilités étant donnée sur un ensemble E au plus dénombrable, c'est-à-dire une famille $(p_x)_{x\in E}$ de réels positifs telle que $\displaystyle\sum_{x\in E} p_x = 1$, on peut se poser la question de l'existence d'une variable aléatoire sur un certain espace probabilisé suivant la loi associée. La réponse est affirmative, il suffit de considérer l'espace probabilisé $\big(E, \mathcal{P}(E), P\big)$, où la probabilité P est celle associée à cette distribution, i.e. telle que, pour tout $A \in \mathcal{P}(E)$, on ait $P(A) = \displaystyle\sum_{x\in A} p_x$ (*cf.* proposition dans le paragraphe **II.6.**: "cas des univers finis ou dénombrables"), et de considérer l'application $X = \mathrm{id}_E : E \to E$, $x \mapsto x$, ainsi X est bien une variable aléatoire sur $\big(E, \mathcal{P}(E), P\big)$ dont la loi est donnée par cette distribution de probabilités $(p_x)_{x\in E}$.

Notation. On notera $X \sim Y$ pour signifier que deux variables aléatoires X et Y (pas nécessairement définies sur le même espace probabilisé) suivent la même loi.

Attention! Deux variables aléatoires ayant la même loi ne sont pas nécessairement égales! Par exemple, si $\Omega = [\![1, 6]\!]$, muni de la tribu $\mathcal{A} = \mathcal{P}(\Omega)$ et de la probabilité uniforme P (*on modélise ici le lancer d'un dé équilibré*), les variables aléatoires $X : \omega \mapsto \omega$ et $Y = 7 - X : \omega \mapsto 7 - \omega$ ont la même loi (loi uniforme sur $[\![1, 6]\!]$), mais ne sont bien sûr pas égales.

Propriété. Si $X \sim X'$, alors $f(X) \sim f(X')$.

Preuve. Soient X et X' deux v.a., définies respectivement sur $(\Omega, \mathcal{A}, P)$ et sur $(\Omega', \mathcal{A}', P')$, telles que $X \sim X'$. On a alors $X(\Omega) = X'(\Omega')$ et on a $P(X \in A) = P'(X' \in A)$ pour toute partie A de $X(\Omega)$. Soit f une application définie sur $X(\Omega)$, notons $F = f\big(X(\Omega)\big)$ son ensemble-image. Si B est une partie de F, alors $\big\{f(X) \in B\big\} = \big\{X \in f^{-1}(B)\big\}$, donc

$$P\big(f(X) \in B\big) = P\big(X \in f^{-1}(B)\big) = P'\big(X' \in f^{-1}(B)\big) = P'\big(f(X') \in B\big)\ ,$$

donc les variables aléatoires $f(X)$ et $f(X')$, qui sont définies respectivement sur $(\Omega, \mathcal{A}, P)$ et sur $(\Omega', \mathcal{A}', P')$, ont la même loi.

3. Une loi usuelle: la loi géométrique.

Définition. Soit $p \in]0, 1[$. On dit qu'une variable aléatoire X sur $(\Omega, \mathcal{A}, P)$ suit la **loi géométrique de paramètre** p si $X(\Omega) = \mathbb{N}^*$ et si

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

On note alors $X \sim \mathcal{G}(p)$.

Remarque. Comme on le verra dans l'interprétation ci-dessous, on peut aussi considérer que $X(\Omega) = \mathbb{N}^* \cup \{+\infty\}$, en rajoutant la condition $P(X = +\infty) = 0$, ce qui ne change rien fondamentalement.

Commentaire. Pour simplifier, nous poserons $q = 1 - p$. Vérfions la cohérence de cette définition: les nombres $a_k = p\, q^{k-1}$, avec $k \in \mathbb{N}^*$, sont bien des réels positifs et on a

$$\sum_{k=1}^{+\infty} a_k = p \sum_{k=0}^{+\infty} q^k = p\, \frac{1}{1-q} = p\, \frac{1}{p} = 1\ .$$

D'après le "commentaire 4" du paragraphe précédent, si $p \in]0, 1[$, il existe bien des variables aléatoires X dont la loi est $\mathcal{G}(p)$.

Interprétation. La loi géométrique $\mathcal{G}(p)$ peut être interprétée comme rang du premier succès dans une suite illimitée d'épreuves de Bernoulli indépendantes de même paramètre p. En effet, considérons cette expérience aléatoire (suite infinie de pile ou face, jouer au loto chaque semaine, ...). Supposons donc qu'à chaque essai, la probabilité de succès est p et, sans construire explicitement un univers, introduisons les événements

S_k: "la k-ième tentative conduit à un succès" $\quad (k \in \mathbb{N}^*)$,

et notons X le nombre de tentatives nécessaires pour obtenir un premier succès. Admettons que X puisse être considérée comme une variable aléatoire sur un certain espace probabilisé. On a alors $X(\Omega) = \mathbb{N}^*$ et, pour tout $k \in \mathbb{N}^*$, on observe que

$$\{X = k\} = \overline{S_1} \cap \cdots \cap \overline{S_{k-1}} \cap S_k\ .$$

Par indépendance des événements S_1, $\cdots$, S_k (dont on peut remplacer certains par leur complémentaire), on a

$$P(X = k) = \Big(\prod_{i=1}^{k-1} P(\overline{S_i}) \Big) \cdot P(S_k) = q^{k-1}\, p\ .$$

Donc $X \sim \mathcal{G}(p)$. On dira que X est le **temps d'attente** du premier succès.

Notons que l'indépendance mutuelle des événements S_1, $\cdots$, S_k est une façon de traduire la notion intuitive d'"épreuves de Bernoulli indépendantes".

En toute rigueur, il faudrait considérer que $X(\Omega) = \mathbb{N}^* \cup \{+\infty\}$. En effet, l'événement $\{X = +\infty\}$, qui traduirait le fait que l'on n'obtient jamais aucun succès lors de cette répétition infinie d'épreuves, n'est pas a priori impossible. Il est seulement négligeable: en effet, son complémentaire est l'événement $\{X < +\infty\} = \bigsqcup_{k \in \mathbb{N}^*} \{X = k\}$, dont la probabilité est $P(X < +\infty) = \sum_{k=1}^{+\infty} P(X = k) = \sum_{k=1}^{+\infty} pq^{k-1} = 1$.

Un petit calcul. Si une variable aléatoire X suit la loi géométrque $\mathcal{G}(p)$, on peut être amené à calculer la probabilité de l'événement $\{X > n\}$ pour $n \in \mathbb{N}$ donné. Pour cela, on peut bien sûr écrire que $\{X > n\} = \bigsqcup_{k=n+1}^{+\infty} \{X = k\}$, et en déduire, par additivité dénombrable, que

$$P(X > n) = \sum_{k=n+1}^{+\infty} P(X = k) = \sum_{k=n+1}^{+\infty} pq^{k-1} = pq^n \sum_{j=0}^{+\infty} q^j = pq^n\, \frac{1}{1-q} = q^n$$

(*on reconnaît un reste de série géométrique, on factorise donc par q^n et on fait un décalage d'indice*). Mais il est plus rapide de se souvenir de l'interprétation en termes de temps d'attente d'un succès: l'événement $\{X > n\}$ signifie alors que les n premières tentatives se sont soldées par des échecs. Avec des notations déjà introduites ci-dessus, $\{X > n\} = \bigcap_{k=1}^{n} \overline{S_k}$, donc par indépendance des épreuves,

$$P(X > n) = \prod_{k=1}^{n} P(\overline{S_k}) = q^n = (1-p)^n \ .$$

4. <u>Une autre loi usuelle: la loi de Poisson</u>.

<u>Définition.</u> Soit λ un réel strictement positif. On dit qu'une variable aléatoire sur $(\Omega, \mathcal{A}, P)$ suit la **loi de Poisson de paramètre** λ, et on note $X \sim \mathcal{P}(\lambda)$, si $X(\Omega) = \mathbb{N}$ et

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

Les réels $a_n = e^{-\lambda} \dfrac{\lambda^n}{n!}$, pour $n \in \mathbb{N}$, sont bien positifs et de somme égale à 1, d'où la cohérence de cette définition.

<u>Commentaire.</u> Cette loi est souvent utilisée pour modéliser le nombre d'événements survenant pendant un certain intervalle de temps (nombre de désintégrations de noyaux radioactifs, nombre de clients entrant dans une boutique, nombre de voitures circulant sur l'autoroute, nombre d'anges qui passent, ...). On l'appelle **loi des événements rares**. Ceci est lié au résultat asymptotique suivant (hors programme), dit "**approximation de la loi binomiale par la loi de Poisson**". Commençons par l'aspect calculatoire, avec un

<u>Théorème d'approximation (HP).</u> Soit $\lambda > 0$, soit (X_n) une suite de variables aléatoires sur $(\Omega, \mathcal{A}, P)$. On suppose que, pour tout n, $X_n \sim \mathcal{B}(n, p_n)$, où (p_n) est une suite de réels de $]0,1[$ telle que $\lim\limits_{n\to+\infty} np_n = \lambda$. Alors, pour tout $k \in \mathbb{N}$, on a

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

<u>Commentaire.</u> On dit alors (mais ce vocabulaire est hors programme) que la suite de variables aléatoires (X_n) **converge en loi** vers une variable aléatoire X telle que $X \sim \mathcal{P}(\lambda)$.

<u>Preuve.</u> On a $X_n(\Omega) = [\![0, n]\!]$ pour tout n, et pour tout k entier naturel,

$$P(X_n = k) = \begin{cases} \binom{n}{k} p_n^k (1-p_n)^{n-k} & \text{si} \quad k \leq n \\ 0 & \text{si} \quad k > n \end{cases} .$$

Si k est fixé, dès que $n \geq k$, on a

$$P(X_n = k) = \frac{n(n-1)\cdots(n-k+1)}{k!} p_n^k (1-p_n)^n (1-p_n)^{-k} \ .$$

Or, $p_n \underset{n\to+\infty}{\sim} \dfrac{\lambda}{n}$, donc $\lim\limits_{n\to+\infty} p_n = 0$ et $\lim\limits_{n\to+\infty} (1-p_n)^{-k} = 1$.

Une fonction polynomiale étant équivalente en $+\infty$ à son terme dominant, on a

$$\frac{n(n-1)\cdots(n-k+1)}{k!} \underset{n\to+\infty}{\sim} \frac{n^k}{k!} .$$

Ensuite, comme $\lim\limits_{n\to+\infty} p_n = 0$,

$$\ln\big((1-p_n)^n\big) = n\,\ln(1-p_n) = n\,\big(-p_n + o(p_n)\big) \quad \underset{n\to+\infty}{\longrightarrow} \quad -\lambda\,,$$

donc $\lim\limits_{n\to+\infty}(1-p_n)^n = e^{-\lambda}$. *Au final, on obtient*

$$P(X_n = k) \underset{n\to+\infty}{\sim} \frac{n^k}{k!}\left(\frac{\lambda}{n}\right)^k e^{-\lambda} = e^{-\lambda}\,\frac{\lambda^k}{k!}\,,$$

ce qui est le résultat annoncé (cet équivalent lorsque n tend vers $+\infty$ ne dépend pas de n, c'est donc aussi la limite lorsque n tend vers $+\infty$).

Commentaire. Ce "résultat asymptotique" permet d'utiliser la loi de Poisson $\mathcal{P}(np)$, dont le calcul est facile, pour approcher une loi binomiale $\mathcal{B}(n,p)$ dont le calcul, utilisant des coefficients binomiaux, est plus délicat, notamment lorsque le paramètre n (nombre de répétitions d'une épreuve de Bernoulli) est grand, et que le paramètre p (probabilité de succès à chaque épreuve) est petit, autrement dit lorsque l'on s'intéresse, sur une longue durée, au nombre d'occurrences d'un "événement rare".

Poussons un peu plus loin cette modélisation: sur un intervalle $[0,T]$ avec $T>0$, notons X le nombre de fois qu'un événement "rare" se produit. Découpons l'intervalle de temps $[0,T]$ en n sous-intervalles de longueur $\dfrac{T}{n}$, à savoir $I_k = \left[\dfrac{kT}{n}, \dfrac{(k+1)T}{n}\right[$, avec $0 \le k \le n-1$. Si n est suffisamment grand, il est raisonnable de faire les hypothèses suivantes:

(H1): sur chaque intervalle de temps I_k, il se produit au plus un événement ;

(H2): il y a indépendance mutuelle de la survenue ou non d'événements dans les différents intervalles I_k.

(H3): La probabilité de présence d'un événement dans un (petit) intervalle de temps est proportionnelle à la durée de cet intervalle.

L'hypothèse **(H1)** correspond à la notion d'"événement rare", elle se traduit par le fait que le nombre U_k d'événements se produisant dans l'intervalle de temps I_k est une variable aléatoire qui suit une loi de Bernoulli.

L'hypothèse **(H2)** signifie l'indépendance des variables aléatoires U_k, $1 \le k \le n$.

L'hypothèse **(H3)** entraîne que, pour un n donné, les variables de Bernoulli U_k ont toutes le même paramètre p_n, et que si l'on fait varier n, ce paramètre est de la forme $p_n = \dfrac{\lambda}{n}$ avec $\lambda > 0$.

Le nombre d'événements se produisant dans l'intervalle $[0,T]$ est $X_n = \sum\limits_{k=1}^{n} U_k$, et on sait alors que $X_n \sim \mathcal{B}(n,p_n)$ (cours de 1ère année). Le théorème d'approximation ci-dessus montre que, si X est une variable aléatoire suivant la loi de Poisson de paramètre λ, alors pour tout k entier naturel, $P(X=k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$ donne "une bonne approximation" de la probabilité que k événements surviennent pendant l'intervalle de temps $[0,T]$.

5. Couples de variables aléatoires.

Soient X et Y deux variables aléatoires discrètes sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$. Pour tout $\omega \in \Omega$, posons $U(\omega) = \big(X(\omega), Y(\omega)\big) \in X(\Omega) \times Y(\Omega)$. On définit ainsi une nouvelle variable aléatoire U, dont l'ensemble-image $U(\Omega)$ est inclus dans le produit cartésien $X(\Omega) \times Y(\Omega)$. *En effet, l'ensemble $X(\Omega) \times Y(\Omega)$ est au plus dénombrable comme produit de deux ensembles ayant cette même propriété, et $U(\Omega)$ en est une partie, donc est aussi au plus dénombrable. Enfin, si $u = (x, y) \in U(\Omega)$, son image réciproque par U s'écrit*

$$U^{-1}\big(\{u\}\big) = \{U = u\} = \big\{(X, Y) = (x, y)\big\} = \{X = x\} \cap \{Y = y\} ,$$

et c'est bien un événement comme intersection de deux événements. On notera $U = (X, Y)$.

Remarque. Si X et Y sont des variables aléatoires réelles, i.e. $X(\Omega) \subset \mathbb{R}$ et $Y(\Omega) \subset \mathbb{R}$, on dira parfois que U est une "variable aléatoire vectorielle", ou encore un "vecteur aléatoire", puisqu'alors $U(\Omega) \subset \mathbb{R}^2$.

Définition. Avec les notations ci-dessus, on définit la **loi conjointe** du couple $U = (X, Y)$, c'est tout simplement la loi de cette nouvelle variable aléatoire U. Plutôt que sur l'espace probabilisable $\Big(U(\Omega), \mathcal{P}\big(U(\Omega)\big)\Big)$, parfois difficile à décrire, on considère cette loi P_U comme une probabilité sur l'espace probabilisable $\Big(X(\Omega) \times Y(\Omega), \mathcal{P}\big(X(\Omega) \times Y(\Omega)\big)\Big)$. Elle est déterminée par sa "distribution", i.e. par les valeurs qu'elle prend sur les singletons $\big\{(x, y)\big\}$:

$$\forall (x, y) \in X(\Omega) \times Y(\Omega) \qquad P_U(x, y) = P_U\Big(\big\{(x, y)\big\}\Big) = P\big(\{X = x\} \cap \{Y = y\}\big) .$$

Pour simplifier, on notera $P_U(x, y) = P(X = x\ ,\ Y = y)$.

Quant aux lois des variables X et Y, on les appelle **lois marginales** du couple $U = (X, Y)$.

Proposition. La loi conjointe détermine les lois marginales.

Preuve. Pour connaître la loi P_X de X, il suffit de connaître $P_X\big(\{x\}\big) = P(X = x)$ pour tout $x \in X(\Omega)$. Or, $\big(\{Y = y\}\big)_{y \in Y(\Omega)}$ est un système complet dénombrable d'événements. La formule des probabilités totales donne alors

$$\forall x \in X(\Omega) \qquad P(X = x) = \sum_{y \in Y(\Omega)} P\big(\{X = x\} \cap \{Y = y\}\big) = \sum_{y \in Y(\Omega)} P_U(x, y) .$$

De même,

$$\forall y \in Y(\Omega) \qquad P(Y = y) = \sum_{x \in X(\Omega)} P\big(\{X = x\} \cap \{Y = y\}\big) = \sum_{x \in X(\Omega)} P_U(x, y) .$$

Rappelons que la réciproque est fausse: **les lois marginales ne déterminent pas la loi conjointe**, cela a déjà été vu en première année dans le cadre particulier des univers finis.

Généralisation. Si $X_1, \cdots, X_n$ sont des variables aléatoires discrètes sur $(\Omega, \mathcal{A}, P)$, on peut considérer le n-uplet $V = (X_1, \cdots, X_n)$ comme une variable aléatoire discrète à valeurs dans le produit cartésien $X_1(\Omega) \times \cdots \times X_n(\Omega)$. Ici aussi, on peut considérer la **loi conjointe** qui est la loi de V, et les **lois marginales** qui sont les lois des différentes variables X_i, $1 \leq i \leq n$. La loi conjointe détermine alors les lois marginales puisque, par exemple si $x_1 \in X_1(\Omega)$, on a

$$P(X_1 = x_1) = \sum_{(x_2,\cdots,x_n)\in X_2(\Omega)\times\cdots\times X_n(\Omega)} P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) \ .$$

Enfin, si $A \in \mathcal{A}$ est un événement tel que $P(A) > 0$, la probabilité conditionnelle P_A est bien une probabilité sur l'espace probabilisable $(\Omega, \mathcal{A})$. Si Y est une variable aléatoire sur ce même espace probabilisable, la loi de Y, considérée comme variable aléatoire sur l'espace probabilisé $\big(\Omega, \mathcal{A}, P_A\big)$ est appelée **loi conditionnelle de Y sachant l'événement** A. On a ainsi, pour toute partie F de $Y(\Omega)$,

$$P_A(Y \in F) = P(Y \in F \mid A) = \frac{P\big(A \cap \{Y \in F\}\big)}{P(A)} \ .$$

Une application. Si X et Y sont deux variables aléatoires discrètes (v.a.d.) sur un même espace probabilisable $(\Omega, \mathcal{A})$, à valeurs réelles ou complexes, si α est un scalaire, alors $\alpha X+Y$ et XY sont des variables aléatoires discrètes sur $(\Omega, \mathcal{A})$. En effet, on a montré au début de ce paragraphe que le couple $U = (X, Y)$ est une v.a.d. sur $(\Omega, \mathcal{A})$. Enfin, $\alpha X + Y = f(U)$ et $XY = g(U)$, en considérant

$$f : \begin{cases} X(\Omega)\times Y(\Omega) & \to & \mathbb{C} \\ (x,y) & \mapsto & \alpha x + y \end{cases} \qquad \text{et} \qquad g : \begin{cases} X(\Omega)\times Y(\Omega) & \to & \mathbb{C} \\ (x,y) & \mapsto & xy \end{cases} .$$

Ce sont donc des v.a.d. en tant que "fonctions" de variables aléatoires discrètes, *cf.* fin du paragraphe **III.1.** On peut généraliser bien sûr à des sommes ou produits de n v.a.d., ou à bien d'autres expressions.

6. Variables aléatoires indépendantes.

Définition 1. Soient X et Y deux variables aléatoires discrètes définies sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$. On dit que X et Y sont **indépendantes** si, pour toute partie A de $X(\Omega)$ et toute partie B de $Y(\Omega)$, les événements $\{X \in A\}$ et $\{Y \in B\}$ sont indépendants, autrement dit si on a

$$\forall (A, B) \in \mathcal{P}\big(X(\Omega)\big) \times \mathcal{P}\big(Y(\Omega)\big) \qquad P(X \in A \ , \ Y \in B) = P(X \in A) \cdot P(Y \in B) \ .$$

On note alors $X \perp\!\!\!\perp Y$.

Proposition. Les deux variables X et Y sont indépendantes si et seulement si la distribution de probabilités du couple (X, Y) est donnée par

$$\forall (x, y) \in X(\Omega) \times Y(\Omega) \qquad P(X = x \ , \ Y = y) = P(X = x)\, P(Y = y) \ .$$

Preuve.

• *Le sens direct est immédiat, il suffit d'appliquer la définition avec $A = \{x\}$ et $B = \{y\}$.*

• *Réciproquement, supposons $P(X = x \ , \ Y = y) = P(X = x)\, P(Y = y)$ pour tout couple (x, y), soient A une partie de $X(\Omega)$ et B une partie de $Y(\Omega)$. On a alors*

$$\begin{aligned} P(X\in A\,,\,Y\in B) &= P\big((X,Y)\in A\times B\big) = \sum_{(x,y)\in A\times B} P(X=x\,,\,Y=y) \\ &= \sum_{(x,y)\in A\times B} P(X=x)\cdot P(Y=y) \qquad \textit{par hypothèse} \\ &= \Big(\sum_{x\in A} P(X=x)\Big)\,\Big(\sum_{y\in B} P(Y=y)\Big) \qquad \textit{(produit de deux sommes)} \\ &= P(X\in A)\cdot P(Y\in B)\,, \end{aligned}$$

donc X et Y sont indépendantes.

Un exemple de calcul. Soient X et Y deux variables indépendantes à valeurs dans $\mathbb{N}$, posons $Z=X+Y$. Alors Z est aussi à valeurs dans $\mathbb{N}$, et sa loi peut se déduire des lois de X et de Y. En effet, pour tout n entier naturel, on a

$$\{Z=n\}=\{X+Y=n\}=\bigsqcup_{p+q=n}\big(\{X=p\}\,\cap\,\{Y=q\}\big)=\bigsqcup_{k=0}^{n}\big(\{X=k\}\,\cap\,\{Y=n-k\}\big)\,,$$

donc par additivité (finie), puis par indépendance des variables X et Y, on déduit

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

Un cas particulier est à connaître, concernant les lois de Poisson:

Proposition. Soient $\lambda>0$ et $\mu>0$, soient X et Y deux variables aléatoires indépendantes telles que $X\sim\mathcal{P}(\lambda)$ et $Y\sim\mathcal{P}(\mu)$. Alors $X+Y\sim\mathcal{P}(\lambda+\mu)$.

Preuve. Posons $Z=X+Y$. Alors, pour tout n entier naturel, on a (cf. ci-dessus):

$$P(Z=n)=\sum_{k=0}^{n}P(X=k)\cdot P(Y=n-k)$$

Donc

$$\begin{aligned} P(Z=n) &= \sum_{k=0}^{n} e^{-\lambda}\,\frac{\lambda^k}{k!}\,e^{-\mu}\,\frac{\mu^{n-k}}{(n-k)!} \\ &= \frac{1}{n!}\,e^{-(\lambda+\mu)}\,\sum_{k=0}^{n}\binom{n}{k}\lambda^k\mu^{n-k} \\ &= e^{-(\lambda+\mu)}\,\frac{(\lambda+\mu)^n}{n!}\,, \end{aligned}$$

ce qui achève la démonstration.

Définition 2. On dit que n variables aléatoires $X_1,\cdots,X_n$ sur $(\Omega,\mathcal{A},P)$ sont **indépendantes** si, quelles que soient les parties $A_1,\,\cdots,\,A_n$ de $X_1(\Omega),\,\cdots,\,X_n(\Omega)$ respectivement, les événements $\{X_1\in A_1\},\,\cdots,\,\{X_n\in A_n\}$ sont indépendants.

On parle parfois de variables "mutuellement indépendantes", pour ne pas confondre avec l'indépendance deux à deux.

Proposition. Les variables $X_1, \cdots, X_n$ sont indépendantes si et seulement si, quelles que soient les parties $A_1, \cdots, A_n$ de $X_1(\Omega), \cdots, X_n(\Omega)$ respectivement, on a

$$P(X_1 \in A_1, \cdots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i) \ .$$

Preuve. Il est clair que cette condition est nécessaire.

Réciproquement, si cette condition est réalisée, soient $B_1, \cdots, B_n$ des parties de $X_1(\Omega), \cdots, X_n(\Omega)$ respectivement. Soit I une partie non vide de l'intervalle $[\![1,n]\!]$, soit $J = [\![1,n]\!] \setminus I$, il faut montrer que $P\Big(\bigcap_{i\in I}\{X_i \in B_i\}\Big) = \prod_{i\in I} P(X_i \in B_i)$. Posons alors $A_i = B_i$ pour $i \in I$ et $A_i = X_i(\Omega)$ pour $i \in J$. On a alors $\{X_i \in A_i\} = \Omega$ pour $i \in J$. L'égalité $P(X_1 \in A_1, \cdots, X_n \in A_n) = P\Big(\bigcap_{i=1}^{n}\{X_i \in A_i\}\Big) = \prod_{i=1}^{n} P(X_i \in A_i)$ se ramène donc à

$$P\Big(\bigcap_{i\in I}\{X_i \in A_i\}\Big) = \prod_{i\in I} P(X_i \in A_i) \ ,$$

et c'est bien ce qu'il fallait démontrer puisque $A_i = B_i$ pour tout $i \in I$.

Proposition. Les variables $X_1, \cdots, X_n$ sont indépendantes si et seulement si la distribution de probabilités du n-uplet $(X_1, \cdots, X_n)$ est donnée par

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

Preuve analogue à celle faite ci-dessus dans le cas de deux variables.

Il est clair que, si $X_1, \cdots, X_n$ sont des variables indépendantes, alors toute sous-famille est constituée aussi de variables indépendantes. En particulier, l'indépendance "mutuelle" entraîne l'indépendance deux à deux, **la réciproque étant fausse**.

Proposition. Si X et Y sont deux variables aléatoires indépendantes sur $(\Omega, \mathcal{A}, P)$, alors quelles que soient les fonctions f et g définies respectivement sur $X(\Omega)$ et $Y(\Omega)$, les variables aléatoires $U = f(X)$ et $V = g(Y)$ sont indépendantes.

Preuve. Soient $u \in U(\Omega) = f\big(X(\Omega)\big)$ et $v \in V(\Omega) = g\big(Y(\Omega)\big)$. On a alors

$$\{U = u\} = \Big\{X \in f^{-1}\big(\{u\}\big)\Big\} \quad \text{et} \quad \{V = v\} = \Big\{Y \in g^{-1}\big(\{v\}\big)\Big\} \ .$$

Alors

$$\begin{aligned} P(U = u, V = v) &= P\Big(X \in f^{-1}\big(\{u\}\big) \ , \ Y \in g^{-1}\big(\{v\}\big)\Big) \\ &= P\Big(X \in f^{-1}\big(\{u\}\big)\Big) \cdot P\Big(Y \in g^{-1}\big(\{v\}\big)\Big) \\ &= P(U = u) \cdot P(V = v) \end{aligned}$$

ce qui prouve l'indépendance des variables U et V.

Ainsi, $X \perp\!\!\!\perp Y \Longrightarrow f(X) \perp\!\!\!\perp g(Y)$.

Généralisation. Si les variables X_1, $\cdots$, X_n sont indépendantes, alors les variables $f_1(X_1)$, $\cdots$, $f_n(X_n)$ sont indépendantes.

Preuve analogue à celle faite ci-dessus dans le cas de deux variables.

Lemme des coalitions. Soient X_1, $\cdots$, X_n des variables aléatoires mutuellement indépendantes, soit $k \in [\![1, n-1]\!]$, soient U et V deux variables aléatoires que l'on peut écrire respectivement sous la forme $U = f(X_1, \cdots, X_k)$ et $V = g(X_{k+1}, \cdots, X_n)$ où f et g sont des applications. Alors les variables U et V sont indépendantes.

Preuve. On va d'abord montrer que les "vecteurs" aléatoires $Y = (X_1, \cdots, X_k)$ et $Z = (X_{k+1}, \cdots, X_n)$ sont indépendants. Pour cela, considérons un k-uplet $y = (x_1, \cdots, x_k)$ dans $X_1(\Omega) \times \cdots \times X_k(\Omega)$, un $(n-k)$-uplet $z = (x_{k+1}, \cdots, x_n)$ dans $X_{k+1}(\Omega) \times \cdots \times X_n(\Omega)$ et montrons que $P(Y = y, Z = z) = P(Y = y) \cdot P(Z = z)$. C'est facile:

$$\begin{aligned} P(Y = y, Z = z) &= P(X_1 = x_1, \cdots, X_k = x_k, X_{k+1} = x_{k+1}, \cdots, X_n = x_n) \\ &= \prod_{i=1}^{n} P(X_i = x_i) \qquad \textit{car } X_1, \cdots, X_n \textit{ sont indépendantes} \\ &= \left(\prod_{i=1}^{k} P(X_i = x_i) \right) \left(\prod_{i=k+1}^{n} P(X_i = x_i) \right) \\ &= P(Y = y)\, P(Z = z) \end{aligned}$$

car X_1, $\cdots$, X_k sont indépendantes, et il en est de même de X_{k+1}, $\cdots$, X_n.

De $Y \perp\!\!\!\perp Z$, on tire $f(Y) \perp\!\!\!\perp g(Z)$, soit $U \perp\!\!\!\perp V$.

On peut généraliser ce résultat au cas de plus de deux coalitions: Si on partitionne l'intervalle $[\![1, n]\!]$ en k parties non vides et deux à deux disjointes (c'est une partition!): $[\![1, n]\!] = \bigsqcup_{j=1}^{k} I_j$, si X_1, $\cdots$, X_n sont des variables aléatoires indépendantes sur $(\Omega, \mathcal{A}, P)$, si Y_1, $\cdots$, Y_k sont des v.a. telles que, pour tout $j \in [\![1, k]\!]$, la variable Y_j est fonction des X_i avec $i \in I_j$, alors les variables Y_1, $\cdots$, Y_k sont indépendantes. *On admet.*

Une remarque. Des événements A_1, $\cdots$, A_n sont indépendants si et seulements les variables aléatoires indicatrices $\mathbb{1}_{A_i}$ $(1 \leq i \leq n)$ sont indépendantes.

Preuve: Posons $X_i = \mathbb{1}_{A_i}$ pour tout i, on a alors $A_i = \{X_i = 1\}$. Il résulte alors immédiatement de la définition de variables aléatoires indépendantes que l'indépendance des X_i entraîne l'indépendance des A_i. Réciproquement, si les événements A_i sont indépendants, il faut montrer que

$$\forall (\varepsilon_1, \cdots, \varepsilon_n) \in \{0,1\}^n \qquad P(X_1 = \varepsilon_1, \cdots, X_n = \varepsilon_n) = \prod_{i=1}^{n} P(X_i = \varepsilon_i)\ .$$

Soit donc $(\varepsilon_1, \cdots, \varepsilon_n) \in \{0,1\}^n$, posons $I = \left\{ i \in [\![1, n]\!] \mid \varepsilon_i = 1 \right\}$ et $J = [\![1, n]\!] \setminus I$. Alors

$$\bigcap_{i=1}^{n} \{X_i = \varepsilon_i\} = \left(\bigcap_{i \in I} A_i \right) \bigcap \left(\bigcap_{i \in J} \overline{A_i} \right)\ .$$

Par indépendance des événements A_i (dont certains peuvent être remplacés par leur contraire), on obtient

$$P\Big(\bigcap_{i=1}^{n}\{X_i=\varepsilon_i\}\Big)=\Big(\prod_{i\in I}P(A_i)\Big)\times\Big(\prod_{i\in J}P(\overline{A_i})\Big)=\prod_{i=1}^{n}P(X_i=\varepsilon_i)\ ,$$

ce qu'il fallait démontrer.

Définition 3. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires sur un même espace probabilisé $(\Omega,\mathcal{A},P)$. On dira que ces variables sont **indépendantes** si, pour tout n entier naturel, les variables X_0, X_1, $\cdots$, X_n sont indépendantes.

On rencontrera notamment des "**suites i.i.d.**", abréviation pour "suites de variables aléatoires **indépendantes et identiquement distribuées**", c'est-à-dire suivant la même loi.

Dans le programme officiel, il est écrit: "on ne soulèvera aucune difficulté quant à l'existence d'un espace probabilisé portant une suite i.i.d.", ce qui me semble confus. Même si ce n'est pas tout à fait conforme au programme, je préfère énoncer (sans démonstration) le résultat (difficile) suivant:

Proposition. Si $(p_x)_{x\in E}$ est une distribution de probabilités sur un ensemble au plus dénombrable E, alors il existe un espace probabilisé $(\Omega,\mathcal{A},P)$ sur lequel on peut construire une suite (X_n) de variables aléatoires indépendantes à valeurs dans E et ayant pour loi commune celle donnée par $(p_x)_{x\in E}$, i.e. telles que

$$\forall x\in E\quad\forall n\in\mathbb{N}\qquad P(X_n=x)=p_x\ .$$

Commentaire. On utilisera cette notion notamment pour **modéliser le jeu de pile ou face infini**. Les résultats des différents lancers de la pièce (en codant par 0 ou 1 les deux issues possibles) seront considérés comme des variables de Bernoulli X_k, $k\in\mathbb{N}^*$, ayant le même paramètre, ces variables X_k étant indépendantes. On a donc une suite i.i.d. de variables de Bernoulli sur un certain espace probabilisé. Notons qu'ici, l'univers est naturellement $\Omega=\{0,1\}^{\mathbb{N}^*}$ (ensemble infini non dénombrable), mais il est plus difficile d'expliciter une tribu, ainsi qu'une probabilité sur cette tribu qui soit compatible avec les calculs intuitifs de probabilités de résultats élémentaires.

IV. Espérance et variance.

1. Espérance d'une variable aléatoire positive.

Définition. Soit X une variable aléatoire discrète sur $(\Omega, \mathcal{A}, P)$, à valeurs dans $[0, +\infty]$. On définit son **espérance** $\mathrm{E}(X)$ par la formule

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

On a alors $\mathrm{E}(X) \in [0, +\infty]$.

Remarque. Dans le cas où $x = +\infty$ et $P(X = +\infty) = 0$, on conviendra que le produit $x\, P(X = x) = (+\infty) \times 0$ est nul.

On dira alors que X est **d'espérance finie** si $\mathrm{E}(X) < +\infty$, autrement dit si la famille de nombres positifs $\big(x\, P(X = x)\big)_{x \in X(\Omega)}$ est sommable.

Exemple de la loi de Poisson.

Si $X \sim \mathcal{P}(\lambda)$ avec $\lambda > 0$, alors X est d'espérance finie, et $\mathrm{E}(X) = \lambda$.

En effet, $X(\Omega) = \mathbb{N}$ *et* $P(X = n) = e^{-\lambda}\, \dfrac{\lambda^n}{n!}$. *La convergence de la série* $\displaystyle\sum_{n \geq 0} n\, P(X = n)$ *est immédiate (par exemple règle de d'Alembert), et*

$$\mathrm{E}(X) = \sum_{n=1}^{+\infty} n\, e^{-\lambda}\, \frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{+\infty} \frac{\lambda^{n+1}}{n!} = \lambda\, e^{-\lambda}\, e^{\lambda} = \lambda\ .$$

Exemple de variable positive d'espérance infinie. Soit X une variable aléatoire telle que $X(\Omega) = \mathbb{N}^*$ et $P(X = n) = \dfrac{1}{n(n+1)}$ pour tout $n \in \mathbb{N}^*$. C'est bien cohérent puisque, par télescopage,

$$\sum_{k=1}^{n} \frac{1}{k(k+1)} = \sum_{k=1}^{n} \Big(\frac{1}{k} - \frac{1}{k+1}\Big) = 1 - \frac{1}{n+1} \quad \underset{n \to +\infty}{\longrightarrow} \quad 1\ , \quad \text{donc} \quad \sum_{n=1}^{+\infty} \frac{1}{n(n+1)} = 1$$

la famille $\left(\dfrac{1}{n(n+1)}\right)_{n \in \mathbb{N}^*}$ est donc une distribution de probabilités. La série de terme général $n\, P(X = n) = \dfrac{1}{n+1}$ diverge, donc $\mathrm{E}(X) = \displaystyle\sum_{n \in \mathbb{N}^*} \frac{1}{n+1} = +\infty$.

Dans le cas (fréquent) des variables aléatoires entières, on a le résultat suivant:

Proposition. Si X est une variable aléatoire à valeurs dans $\mathbb{N} \cup \{+\infty\}$, alors

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

Preuve. On fait des calculs dans $[0, +\infty]$*, ce qui autorise les interversions de sommes. Ainsi,*
- si $P(X = +\infty) > 0$*, alors*

$$\sum_{n=1}^{+\infty} P(X \geq n) = \sum_{n=1}^{+\infty} \left(P(X = +\infty) + \sum_{k=n}^{+\infty} P(X = k) \right) = +\infty$$

car la série à termes positifs $\sum P(X \geq n)$ *est alors grossièrement divergente. Dans ce cas, on a aussi* $\mathrm{E}(X) = +\infty$ *en partant de la définition de l'espérance, d'où l'égalité recherchée.*

- si $P(X = +\infty) = 0$*, alors*

$$\sum_{n=1}^{+\infty} P(X \geq n) = \sum_{n=1}^{+\infty} \left(\sum_{k=n}^{+\infty} P(X = k) \right) = \sum_{k=1}^{+\infty} \left(\sum_{n=1}^{k} P(X = k) \right) = \sum_{k=1}^{+\infty} k\, P(X = k) = \mathrm{E}(X)\ .$$

Une variable aléatoire à valeurs dans $\mathbb{N} \cup \{+\infty\}$ est donc d'espérance finie si et seulement si la série à termes positifs $\sum P(X \geq n)$ est convergente.

Exemple de la loi géométrique.

Si $X \sim \mathcal{G}(p)$ avec $0 < p < 1$, alors X est d'espérance finie et $\mathrm{E}(X) = \dfrac{1}{p}$.

Preuve. En effet, $X(\Omega) = \mathbb{N}^*$ *et* $P(X = n) = pq^{n-1}$ *avec* $q = 1 - p$*. On sait aussi que, pour tout* $n \in \mathbb{N}^*$*, on a* $P(X \geq n) = P(X > n-1) = q^{n-1}$*, cf. paragraphe sur la loi géométrique. La série* $\sum P(X \geq n)$ *est géométrique de raison* $q \in]0, 1[$*, elle est donc convergente, et*

$$\mathrm{E}(X) = \sum_{n=1}^{+\infty} P(X \geq n) = \sum_{n=1}^{+\infty} q^{n-1} = \sum_{n=0}^{+\infty} q^n = \frac{1}{1-q} = \frac{1}{p}\ .$$

Remarque. Dans le cas de la loi géométrique, on a vu que l'on pouvait aussi considérer que $X(\Omega) = \mathbb{N}^* \cup \{+\infty\}$ avec $P(X = +\infty) = 0$, mais cela ne change rien au calcul.

Le résultat obtenu est simple à interpréter: en effet, si l'on répète une épreuve où la probabilité de gagner est par exemple $p = \dfrac{1}{10}$, il faudra jouer en moyenne 10 fois pour gagner, i.e. le "temps d'attente moyen" d'un succès est $\dfrac{1}{p}$.

2. Familles sommables de nombres complexes.

Si I est un ensemble d'indices au plus dénombrable, une famille $(x_i)_{i \in I}$ de nombres complexes est dite **sommable** si la famille de réels positifs $\left(|x_i|\right)_{i \in I}$ est sommable, i.e. si $\sum_{i \in I} |x_i| < +\infty$ (*cf.* paragraphe précédent).

C'est toujours le cas bien évidemment si l'ensemble I est fini.

Dans le cas où $I = \mathbb{N}$, la sommabilité de la famille $(x_n)_{n \in \mathbb{N}}$, i.e. de la **suite** (x_n), équivaut à la **convergence absolue** de la série $\sum_{n \geq 0} x_n$.

Si I est un ensemble dénombrable, soit $I = \{i_n\ ;\ n \in \mathbb{N}\}$ une énumération de cet ensemble, la sommabilité de la famille $(x_i)_{i \in I}$ équivaut à la **convergence absolue** de la série $\sum_{n \geq 0} x_{i_n}$,

on admettra que ce résultat ne dépend pas du choix de l'énumération de l'ensemble I.

Théorème de comparaison. Si $|x_i| \leq y_i$ pour tout $i \in I$, la famille de réels positifs $(y_i)_{i\in I}$ étant sommable, alors la famille $(x_i)_{i\in I}$ est sommable.

Définition de la somme. Si $(x_i)_{i\in I}$ est une famille sommable, si $I = \{i_n \ ; \ n \in \mathbb{N}\}$ est une énumération de I, on admettra que la valeur de la somme $\displaystyle\sum_{n=0}^{+\infty} x_{i_n}$ ne dépend pas de l'énumération choisie de cet ensemble I. *Ce résultat a déjà été énoncé sans preuve à la fin du poly de cours sur les séries, en affirmant que toute série absolument convergente est "commutativement convergente".* On posera alors

$$\sum_{i\in I} x_i = \sum_{n=0}^{+\infty} x_{i_n} \ ,$$

où $I = \{i_n \ ; \ n \in \mathbb{N}\}$ est une quelconque énumération de l'ensemble I, on a ainsi défini la **somme** de la famille $(x_i)_{i\in I}$.

On admettra alors les propriétés suivantes:

Croissance. Si $(x_i)_{i\in I}$ et $(y_i)_{i\in I}$ sont deux familles sommables de nombres **réels**, si $x_i \leq y_i$ pour tout $i \in I$, alors $\displaystyle\sum_{i\in I} x_i \leq \sum_{i\in I} y_i$.

Inégalité triangulaire. Si $(x_i)_{i\in I}$ est une famille sommable, alors $\displaystyle\Big|\sum_{i\in I} x_i\Big| \leq \sum_{i\in I} |x_i|$.

Linéarité. Si $(x_i)_{i\in I}$ et $(y_i)_{i\in I}$ sont deux familles sommables de nombres complexes, si $\alpha \in \mathbb{C}$ et $\beta \in \mathbb{C}$, alors la famille $(\alpha x_i + \beta y_i)_{i\in I}$ est sommable et on a

$$\sum_{i\in I} (\alpha x_i + \beta y_i) = \alpha \sum_{i\in I} x_i + \beta \sum_{i\in I} y_i \ .$$

Sommation par paquets. Si $I = \displaystyle\bigsqcup_{n\in\mathbb{N}} I_n$, autrement si $(I_n)_{n\in\mathbb{N}}$ est une partition de l'ensemble I, si la famille $(x_i)_{i\in I}$ est sommable, alors les familles $(x_i)_{i\in I_n}$ sont sommables pour tout n entier naturel, la série $\displaystyle\sum_{n\geq 0} \Big(\sum_{i\in I_n} x_i\Big)$ converge, et on a l'égalité

$$\sum_{i\in I} x_i = \sum_{n=0}^{+\infty} \Big(\sum_{i\in I_n} x_i\Big) \ .$$

Théorème de Fubini. Si I et J sont deux ensembles au plus dénombrables, il en est alors de même du produit cartésien $I \times J$, si $(x_{i,j})_{(i,j)\in I\times J}$ est une famille sommable de nombres complexes, alors on peut écrire

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

Ce théorème autorise à faire des **interversions de sommes**, soit dans le cas de familles de rées positifs (les sommes peuvent alors éventuellement valoir $+\infty$, *cf.* paragraphe précédent), soit dans le cas de familles **sommables** de nombres complexes.

En particulier, si $(x_{k,l})_{(k,l)\in\mathbb{N}^2}$ est une "suite double" sommable, on peut écrire

$$\sum_{(k,l)\in\mathbb{N}^2} x_{k,l} = \sum_{k=0}^{+\infty}\Big(\sum_{l=0}^{+\infty} x_{k,l}\Big) = \sum_{l=0}^{+\infty}\Big(\sum_{k=0}^{+\infty} x_{k,l}\Big)\ .$$

Produit de deux sommes. Si I et J sont au plus dénombrables, si les familles $(x_i)_{i\in I}$ et $(y_j)_{j\in J}$ sont sommables, alors la famille $(x_i y_j)_{(i,j)\in I\times J}$ est sommable et

$$\sum_{(i,j)\in I\times J} x_i y_j = \Big(\sum_{i\in I} x_i\Big)\Big(\sum_{j\in J} y_j\Big)\ .$$

Exemples. • Montrer que la famille $\left(q^{|n|}\right)_{n\in\mathbf{Z}}$ est sommable si et seulement si $|q|<1$ et que, dans ce cas, on a $\displaystyle\sum_{n\in\mathbf{Z}} q^{|n|} = \frac{1+q}{1-q}$.

• Soit q un complexe tel que $|q|<1$, soit l'ensemble $I = \big\{(m,n)\in(\mathbb{N}^*)^2 \mid m\le n\big\}$. Montrer que la famille $\left(\dfrac{m\,q^m}{n(n+1)}\right)_{(m,n)\in I}$ est sommable, et que sa somme vaut $\dfrac{q}{1-q}$.

• On admet que $\displaystyle\sum_{n=1}^{+\infty}\frac{1}{n^2} = \frac{\pi^2}{6}$. En déduire les valeurs des sommes $\displaystyle\sum_{p=0}^{+\infty}\frac{1}{(2p+1)^2}$ et $\displaystyle\sum_{k=1}^{+\infty}\frac{(-1)^k}{k^2}$, puis montrer que la famille $\left(\dfrac{(-1)^{kl}}{k^2l^2}\right)_{(k,l)\in(\mathbb{N}^*)^2}$ est sommable, et calculer sa somme.

3. Espérance d'une variable réelle ou complexe.

On utilisera désormais l'abréviation "**v.a.d.**" pour désigner une variable aléatoire discrète à valeurs réelles ou complexes.

a. Définition de l'espérance.

Définition. Soit X une variable aléatoire discrète sur $(\Omega,\mathcal{A},P)$, à valeurs réelles ou complexes. On dit que X **est d'espérance finie** si la famille $\big(x\,P(X=x)\big)_{x\in X(\Omega)}$ est sommable. Dans ce cas, on appelle **espérance** de X, et on note $\mathrm{E}(X)$, la somme de cette famille, soit

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

Commentaire. Si l'ensemble $X(\Omega)$ est dénombrable, soit $X(\Omega)=\{x_n\ ;\ n\in\mathbb{N}\}$ une énumération de cet ensemble. Alors X est d'espérance finie si et seulement si la série $\displaystyle\sum_{n\ge0} x_n\,P(X=x_n)$ est **absolument** convergente. Dans ce cas, l'espérance est

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

Vocabulaire. L'espérance d'une v.a.d. X est aussi appelée son **moment d'ordre un**. Pour affirmer que X est d'espérance finie, on dit parfois aussi que X **admet un moment d'ordre un**.

b. Propriétés de l'espérance.

Formule de transfert. Soit X une variable aléatoire discrète (dont les valeurs ne sont pas nécessairement des réels ou des complexes), soit $f : X(\Omega) \to \mathbb{C}$ une application. Alors la variable aléatoire $Y = f(X)$ est d'espérance finie si et seulement si la famille $\big(f(x)\,P(X=x)\big)_{x\in X(\Omega)}$ est sommable, et dans ce cas,

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

Preuve. Dans $[0,+\infty]$, on peut toujours faire le calcul suivant, par sommation par paquets, en écrivant que $X(\Omega) = \bigsqcup_{y\in Y(\Omega)} f^{-1}\big(\{y\}\big)$:

$$\begin{aligned}
\sum_{x\in X(\Omega)} \big|f(x)\big|\,P(X=x) &= \sum_{y\in Y(\Omega)} \left(\sum_{x\in f^{-1}(\{y\})} \big|f(x)\big|\,P(X=x) \right) \\
&= \sum_{y\in Y(\Omega)} |y| \left(\sum_{x\in f^{-1}(\{y\})} P(X=x) \right) \\
&= \sum_{y\in Y(\Omega)} |y|\,P\Big(X\in f^{-1}\big(\{y\}\big)\Big) \\
&= \sum_{y\in Y(\Omega)} |y|\,P(Y=y)\,.
\end{aligned}$$

L'une des deux sommes est donc finie si et seulement si l'autre l'est, donc $Y = f(X)$ est d'espérance finie si et seulement si la famille $\big(f(x)\,P(X=x)\big)_{x\in X(\Omega)}$ est sommable. Dans ce cas, en refaisant le même calcul par sommation par paquets sans les valeurs absolues (ou modules), on obtient l'égalité annoncée.

Commentaire. Le calcul ci-dessus montre aussi que, **si f est à valeurs positives**, alors on a toujours, dans $[0,+\infty]$, l'égalité

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

Commentaire. Dans la formule de transfert ci-dessus, la variable X n'est pas nécessairement à valeurs numériques (réelles ou complexes), il peut donc s'agir d'un couple ou d'un n-uplet. On le verra par exemple dans la preuve de la linéarité de l'espérance, ci-dessous.

Un exemple de calcul. Si X suit la loi de Poisson de paramètre λ avec $\lambda > 0$, alors X^2 est d'espérance finie et

$$\begin{aligned}
\mathrm{E}(X^2) &= \sum_{n=1}^{+\infty} n^2\,e^{-\lambda}\,\frac{\lambda^n}{n!} \;=\; \sum_{n=1}^{+\infty} \big(n(n-1)+n\big)\,e^{-\lambda}\,\frac{\lambda^n}{n!} \\
&= \sum_{n=2}^{+\infty} n(n-1)\,e^{-\lambda}\,\frac{\lambda^n}{n!} + \sum_{n=1}^{+\infty} n\,e^{-\lambda}\,\frac{\lambda^n}{n!}
\end{aligned}$$

$$= \lambda^2\, e^{-\lambda} \sum_{n=2}^{+\infty} \frac{\lambda^{n-2}}{(n-2)!} + \lambda\, e^{-\lambda} \sum_{n=1}^{+\infty} \frac{\lambda^{n-1}}{(n-1)!}$$
$$= \lambda^2 + \lambda\,.$$

Conséquence: inégalité triangulaire. De la formule de transfert, on peut déduire aussi que, si X est une v.a.d., alors X est d'espérance finie si et seulement si $|X|$ est d'espérance finie, i.e. si et seulement si $\mathrm{E}\big(|X|\big) < +\infty$, cette écriture ayant un sens dans le cas d'une variable positive. Dans ce cas, on a l'**inégalité triangulaire** $\big|\mathrm{E}(X)\big| \leq \mathrm{E}\big(|X|\big)$.

En effet, par l'inégalité triangulaire sur une somme, on a

$$\big|\mathrm{E}(X)\big| = \Big| \sum_{x\in X(\Omega)} x\, P(X=x)\Big| \leq \sum_{x\in X(\Omega)} |x|\, P(X=x) = \mathrm{E}\big(|X|\big)\,,$$

la dernière égalité résultant de la formule de transfert.

Linéarité de l'espérance. Soient X et Y deux v.a.d. d'espérance finie sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$, soient α et β deux nombres complexes. Alors $Z = \alpha X + \beta Y$ est aussi une v.a.d. d'espérance finie sur $(\Omega, \mathcal{A}, P)$, et on a

$$\mathrm{E}(\alpha X + \beta Y) = \alpha\, \mathrm{E}(X) + \beta\, \mathrm{E}(Y)\,.$$

Preuve. On a vu à la fin du paragraphe **III.1.** *que Z est bien une v.a.d. sur $(\Omega, \mathcal{A}, P)$. On peut écrire $Z = f(U)$, en introduisant le couple $U = (X, Y)$ et l'application $f : \mathbb{C}^2 \to \mathbb{C}$, $u = (x, y) \mapsto \alpha x + \beta y$. D'après la formule de transfert, pour montrer que Z est d'espérance finie, il faut montrer la sommabilité de la famille $\Big(f(u)\, P(U=u)\Big)_{u\in U(\Omega)}$, soit encore de*

$$\Big((\alpha x + \beta y)\, P(X=x\,,\, Y=y)\Big)_{(x,y)\in X(\Omega)\times Y(\Omega)}\,.$$

Faisons donc encore des calculs dans $[0, +\infty]$, où tous les coups sont permis:

$$\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} |\alpha x + \beta y|\, P(X=x\,,\, Y=y)$$
$$\leq \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} \big(|\alpha|\,|x| + |\beta|\,|y|\big)\, P(X=x\,,\, Y=y)$$
$$= |\alpha| \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} |x|\, P(X=x\,,\, Y=y) + |\beta| \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} |y|\, P(X=x\,,\, Y=y)$$
$$= |\alpha| \sum_{x\in X(\Omega)} \Big(|x| \sum_{y\in Y(\Omega)} P(X=x\,,\, Y=y)\Big) + |\beta| \sum_{y\in Y(\Omega)} \Big(|y| \sum_{x\in X(\Omega)} P(X=x\,,\, Y=y)\Big)$$
$$= |\alpha| \sum_{x\in X(\Omega)} |x|\, P(X=x) + |\beta| \sum_{y\in Y(\Omega)} |y|\, P(Y=y)$$
$$= |\alpha|\, \mathrm{E}\big(|X|\big) + |\beta|\, \mathrm{E}\big(|Y|\big) \;<\; +\infty\,,$$

cela prouve la sommabilité. On réécrit ensuite le même calcul dans $\mathbb{C}$, sans les modules (ou valeurs absolues), l'inégalité au début de la deuxième ligne devient une égalité, et on obtient $\mathrm{E}(Z) = \alpha\, \mathrm{E}(X) + \beta\, \mathrm{E}(Y)$.

Conséquence. L'ensemble des variables aléatoires discrètes sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$, et d'espérance finie, est muni d'une structure de $\mathbb{K}$-espace vectoriel (c'est un s.e.v. de $\mathcal{F}(\Omega, \mathbb{K}) = \mathbb{K}^\Omega$), on le note parfois $\mathbb{L}^1(\Omega, \mathcal{A}, P, \mathbb{K})$, ou plus simplement $\mathbb{L}^1(\Omega, \mathbb{K})$ si aucune confusion n'est à craindre.

Conséquence. De la formule de transfert, on déduit que, si Z est une v.a.d. complexe, alors Z est d'espérance finie si et seulement si $\overline{Z}$ est d'espérance finie, et qu'on a alors $\mathrm{E}(\overline{Z}) = \overline{\mathrm{E}(Z)}$. En utilisant les relations $\mathrm{Re}(Z) = \dfrac{1}{2}(Z + \overline{Z})$ et $\mathrm{Im}(Z) = \dfrac{1}{2i}(Z - \overline{Z})$, de la linéarité de l'espérance, on déduit que Z est d'espérance finie si et seulement si $\mathrm{Re}(Z)$ et $\mathrm{Im}(Z)$ sont d'espérance finie, et qu'on a alors

$$\mathrm{E}\big(\mathrm{Re}(Z)\big) = \mathrm{Re}\,\big(\mathrm{E}(Z)\big) \qquad \text{et} \qquad \mathrm{E}\big(\mathrm{Im}(Z)\big) = \mathrm{Im}\,\big(\mathrm{E}(Z)\big)\ .$$

Définition. Une v.a.d. Y est dite **centrée** si elle est d'espérance finie avec $\mathrm{E}(Y) = 0$. Si X est une v.a.d. d'espérance finie, alors $Y = X - \mathrm{E}(X)$ est une variable centrée. *C'est une conséquence immédiate de la linéarité de l'espérance.*

Théorème de comparaison. Soient X et Y deux v.a.d. On suppose que Y est d'espérance finie et que $|X| \leq Y$. Alors X est d'espérance finie.

Preuve. Considérons le couple $U = (X, Y)$, alors $U(\Omega)$ est une partie du produit cartésien $X(\Omega) \times Y(\Omega)$ telle que $\forall (x, y) \in U(\Omega) \quad |x| \leq y$. Enfin, si un couple $(x, y) \in X(\Omega) \times Y(\Omega)$ n'appartient pas à $U(\Omega)$, on a $P(X = x, Y = y) = P\big(U = (x, y)\big) = 0$. Cela justifie le calcul suivant dans $[0, +\infty]$, en utilisant le théorème de Fubini:

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

On a donc prouvé que $|X|$ est d'espérance finie, donc X aussi.

Positivité. Si une v.a.d. X est positive, i.e. si $X(\Omega) \subset \mathbb{R}_+$, alors $E(X) \geq 0$. De plus, l'espérance de X est nulle si et seulement si X est presque sûrement nulle, i.e. ssi $P(X = 0) = 1$.

Preuve. La première affirmation est immédiate, puisque $\mathrm{E}(X) \in [0,+\infty]$ *d'après le paragraphe* **a.** *Dans* $[0,+\infty]$*, on a* $\mathrm{E}(X) = \sum_{x\in X(\Omega)} x\,P(X=x)$*. Cette somme de termes positifs est nulle si et seulement si chaque terme est nul, on a donc nécessairement* $P(X=x)=0$ *pour tout* x **non nul** *dans* $X(\Omega)$*. Comme, d'autre part,* $\sum_{x\in X(\Omega)} P(X=x) = P(\Omega) = 1$*, il reste* $P(X=0)=1$*. La réciproque est immédiate.*

Croissance. Soient $X \in \mathbb{L}^1(\Omega,\mathbb{R})$ **et** $Y \in \mathbb{L}^1(\Omega,\mathbb{R})$**. Si** $X \le Y$**, c'est-à-dire si** $\forall\omega\in\Omega \quad X(\omega)\le Y(\omega)$**, alors** $\mathrm{E}(X)\le\mathrm{E}(Y)$**.**

Preuve. Cela résulte de la linéarité et de la positivité: on a $Y-X\ge 0$ *et* $Y-X\in\mathbb{L}^1(\Omega,\mathbb{R})$*, donc* $\mathrm{E}(Y-X)\ge 0$*, i.e.* $\mathrm{E}(Y)-\mathrm{E}(X)\ge 0$*.*

Proposition. Si X **et** Y **sont d'espérance finie et indépendantes, alors** XY **est d'espérance finie et**

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

Preuve. Introduisons le couple $U=(X,Y)$*, alors* $XY = f(U)$ *avec* $f : \mathbb{C}^2 \to \mathbb{C}$*,* $u=(x,y)\mapsto xy$*. De la formule de transfert, on déduit qu'il faut montrer la sommabilité de la famille* $\big(f(u)\ P(U=u)\big)_{u\in U(\Omega)}$*, ou encore de* $\big(xy\,P(X=x\ ,\ Y=y)\big)_{(x,y)\in X(\Omega)\times Y(\Omega)}$*. Or, en calculant encore dans* $[0,+\infty]$*, l'indépendance des variables* X *et* Y *permet de se ramener à un produit de deux sommes:*

$$\begin{aligned}\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} |xy|\,P(X=x\ ,\ Y=y) &= \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} |xy|\,P(X=x)\,P(Y=y)\\ &= \Big(\sum_{x\in X(\Omega)} |x|\,P(X=x)\Big)\Big(\sum_{y\in Y(\Omega)} |y|\,P(Y=y)\Big)\\ &= \mathrm{E}\big(|X|\big)\,\mathrm{E}\big(|Y|\big) \ <\ +\infty\ ,\end{aligned}$$

ce qui prouve la sommabilité. En refaisant le même calcul sans les modules, on obtient la relation $\mathrm{E}(XY)=\mathrm{E}(X)\,\mathrm{E}(Y)$*.*

Extension. Si $X_1, \cdots, X_n$ sont des v.a.d. indépendantes, chacune d'espérance finie, alors le produit $P=\prod_{i=1}^n X_i$ est d'espérance finie et

$$\mathrm{E}\Big(\prod_{i=1}^n X_i\Big) = \prod_{i=1}^n \mathrm{E}(X_i)\ .$$

On procède par récurrence en partant de la proposition ci-dessus, en remarquant pour l'hérédité que le lemme des coalitions entraîne l'indépendance des variables $\prod_{i=1}^{n-1} X_i$ *et* X_n*.*

c. Inégalité de Markov.

Proposition. Soit X **une variable aléatoire discrète positive et d'espérance finie. On a alors**

$$\forall a\in\mathbb{R}_+^* \qquad P(X\ge a)\le\frac{\mathrm{E}(X)}{a}\ .$$

Preuve. *Soit l'événement* $A = \{X \geq a\}$, *notons* $\mathbb{1}_A$ *la variable aléatoire indicatrice de* A, *c'est une variable de Bernoulli et* $P(X \geq a) = P(A) = \mathrm{E}(\mathbb{1}_A)$.

On constate que $X \geq a \cdot \mathbb{1}_A$. *En effet, pour* $\omega \in \Omega$,

- *si* $X(\omega) < a$, *alors* $0 = a \cdot \mathbb{1}_A(\omega) \leq X(\omega)$ *puisque* X *est à valeurs positives ;*
- *si* $X(\omega) \geq a$, *alors* $a = a \cdot \mathbb{1}_A(\omega) \leq X(\omega)$.

Par croissance de l'espérance et linéarité, on a immédiatement

$$\mathrm{E}(X) \geq a\,\mathrm{E}(\mathbb{1}_A) = a\,P(X \geq a)\ .$$

4. Moment d'ordre deux et variance.

Dans tout ce paragraphe, on considérera des **v.a.r.d.** (variables aléatoires réelles discrètes).

a. L'espace vectoriel $L^2(\Omega, \mathbb{R})$.

Proposition. Soit X une v.a.r.d. sur $(\Omega, \mathcal{A}, P)$. Si X^2 est d'espérance finie, alors X est d'espérance finie.

Preuve. *On a en effet* $|X| \leq \dfrac{1}{2}(X^2 + 1)$. *Or,* X^2 *et* 1 *(variable aléatoire constante) sont d'espérance finie, il en est donc de même de* $\dfrac{1}{2}(X^2 + 1)$ *par linéarité, puis de* X *par le théorème de comparaison énoncé au paragraphe précédent.*

Vocabulaire. Le réel $\mathrm{E}(X^2)$, lorsqu'il existe, est appelé **moment d'ordre deux** de la variable aléatoire X. Pour dire que X^2 est d'espérance finie, on dit donc aussi que X **admet un moment d'ordre deux**, ou encore que X **est de variance finie**. Enfin, je noterai $\mathbb{L}^2(\Omega, \mathcal{A}, P, \mathbb{R})$, ou plus simplement $\mathbb{L}^2(\Omega, \mathbb{R})$, l'ensemble des v.a.r.d. sur $(\Omega, \mathcal{A}, P)$ qui admettent un moment d'ordre deux.

Proposition. L'ensemble $\mathbb{L}^2(\Omega, \mathbb{R})$ des v.a.r.d. admettant un moment d'ordre deux est un $\mathbb{R}$-espace vectoriel, et l'application

$$b : \begin{cases} \mathbb{L}^2(\Omega, \mathbb{R}) \times \mathbb{L}^2(\Omega, \mathbb{R}) & \to \mathbb{R} \\ (X, Y) & \mapsto \mathrm{E}(XY) \end{cases}$$

est une forme bilinéaire symétrique positive sur cet espace vectoriel.

Preuve. *On utilise encore l'inégalité usuelle* $|XY| \leq \dfrac{1}{2}(X^2 + Y^2)$.

Si X *et* Y *sont dans* $L^2(\Omega, \mathbb{R})$, *comme* X^2 *et* Y^2 *sont d'espérance finie, il en est de même de* $\dfrac{1}{2}(X^2 + Y^2)$, *et donc de* XY *par le théorème de comparaison du paragraphe précédent.*

On a prouvé que $\big(X \in \mathbb{L}^2(\Omega, \mathbb{R}) \text{ et } Y \in \mathbb{L}^2(\Omega, \mathbb{R})\big) \Longrightarrow XY \in \mathbb{L}^1(\Omega, \mathbb{R})$.

Montrons que $\mathbb{L}^2(\Omega, \mathbb{R})$ *est un s.e.v. de* $\mathcal{F}(\Omega, \mathbb{R}) = \mathbb{R}^\Omega$. *D'abord,* $0 \in \mathbb{L}^2(\Omega, \mathbb{R})$, *puis la stabilité par multiplication par un scalaire est immédiate. Enfin, si* X *et* Y *sont des v.a.r.d. admettant un moment d'ordre deux, alors* X^2, Y^2 *et* XY *sont d'espérance finie, donc appartiennent à l'espace vectoriel* $\mathbb{L}^1(\Omega, \mathbb{R})$, *cf. paragraphe précédent. Donc*

$$(X + Y)^2 = X^2 + 2XY + Y^2 \in \mathbb{L}^1(\Omega, \mathbb{R})\ ,$$

i.e. $X + Y \in \mathbb{L}^2(\Omega, \mathbb{R})$.

Au passage, on peut noter que $\mathbb{L}^2(\Omega,\mathbb{R}) \subset \mathbb{L}^1(\Omega,\mathbb{R})$.

La bilinéarité de b résulte de la linéarité de l'espérance, la symétrie est immédiate, enfin

$$\forall X \in \mathbb{L}^2(\Omega,\mathbb{R}) \qquad b(X,X) = \mathrm{E}(X^2) \geq 0$$

par positivité de l'espérance.

Notons qu'il manque à la forme b le caractère défini pour être un produit scalaire sur $\mathbb{L}^2(\Omega,\mathbb{R})$. *En effet, d'après le paragraphe précédent,* $b(X,X) = \mathrm{E}(X^2)$ *est nul pour toute variable X "presque sûrement nulle", i.e. telle que* $P(X=0)=1$, *ce qui n'entraîne pas (pas tout à fait!) que* $X=0$. *On écrit parfois "*$X=0$ *p.s."*

<u>Inégalité de Cauchy-Schwarz.</u> Soient X et Y des v.a.r.d. telles que X^2 et Y^2 sont d'espérance finie. Alors XY est aussi d'espérance finie, et on a

$$\mathrm{E}(XY)^2 \leq \mathrm{E}(X^2)\,\mathrm{E}(Y^2)\ ,$$

avec égalité si et seulement s'il existe une relation de colinéarité presque sûre entre X et Y.

<u>Commentaire.</u> Cette dernière condition signifie que, soit $P(X=0)=1$, soit il existe un réel λ tel que $P(Y=\lambda X)=1$.

<u>Preuve.</u> Reprenons la preuve classique du cours sur les espaces préhilbertiens, la seule différence étant qu'ici on n'a pas tout à fait un produit scalaire sur $\mathbb{L}^2(\Omega,\mathbb{R})$ *puisque la forme b n'est pas "définie", cela changera juste l'étude du cas d'égalité.*

Soient $X \in \mathbb{L}^2(\Omega,\mathbb{R})$ *et* $Y \in \mathbb{L}^2(\Omega,\mathbb{R})$. *Alors, pour tout réel* λ, *on a* $\mathrm{E}\big((\lambda X+Y)^2\big) \geq 0$. *En développant, avec la linéarité de l'espérance, on obtient*

$$\textbf{(*) :} \qquad \forall \lambda \in \mathbb{R} \qquad \lambda^2\,\mathrm{E}(X^2) + 2\lambda\,\mathrm{E}(XY) + \mathrm{E}(Y^2) \geq 0\ .$$

Deux cas se présentent:

- si $\mathrm{E}(X^2)=0$: *cette condition équivaut à* $X^2=0$ *p.s., donc à* $X=0$ *p.s. Dans ce cas, on a aussi* $XY=0$ *p.s., donc* $\mathrm{E}(XY)=0$. *L'inégalité de Cauchy-Schwarz est alors satisfaite, et il s'agit dans ce cas d'une égalité, les deux membres étant nuls.*

- si $\mathrm{E}(X^2)>0$: *la relation* **(*)** *nous montre qu'un certain trinôme est toujours positif sur* $\mathbb{R}$, *ce qui exclut l'éventualité de deux racines réelles distinctes, son discriminant* Δ *est alors négatif ou nul, ce qui fournit l'inégalité de Cauchy-Schwarz.*

L'égalité dans Cauchy-Schwarz a lieu:

- si $\mathrm{E}(X^2)=0$, *et dans ce cas on a* $X=0$ *presque sûrement ;*

- si $\mathrm{E}(X^2)>0$ *et* $\Delta=0$, *le trinôme* **(*)** *admet alors une racine réelle (double)* λ_0, *ce qui signifie que* $\mathrm{E}\big((\lambda_0 X+Y)^2\big)=0$, *donc* $\lambda_0 X+Y=0$ *p.s., soit* $Y=-\lambda_0 X$ *presque sûrement.*

La réciproque est immédiate: *si* $X=0$ *p.s. ou* $Y=\lambda X$ *p.s., alors* $\mathrm{E}(XY)^2=\mathrm{E}(X^2)\,\mathrm{E}(Y^2)$.

b. Variance.

Définition. Soit $X \in \mathbb{L}^2(\Omega, \mathbb{R})$. On définit sa **variance** par la relation

$$\mathrm{V}(X) = \mathrm{E}\Big(\big(X - \mathrm{E}(X)\big)^2\Big) .$$

Commentaire 1. Cela a bien un sens: en effet, on a vu que, si X^2 est d'espérance finie, alors X est d'espérance finie, c'est l'inclusion $\mathbb{L}^2(\Omega, \mathbb{R}) \subset \mathbb{L}^1(\Omega, \mathbb{R})$, notons $m = \mathrm{E}(X)$ pour simplifier. Ensuite,

$$(X - m)^2 = X^2 - 2mX + m^2$$

est une combinaison linéaire de v.a. d'espérance finie, donc $(X - m)^2$ est d'espérance finie.

Commentaire 2. Si l'on interprète l'espérance d'une v.a. X comme une "moyenne" (c'est la moyenne des valeurs prises par X, pondérée par les probabilités d'obtenir ces différentes valeurs), la variance est alors **la moyenne des carrés des écarts à la moyenne**, c'est un **indicateur de dispersion** de la variable X.

Commentaire 3. Pour dire d'une variable aléatoire qu'elle appartient à $\mathbb{L}^2(\Omega, \mathbb{R})$, on dit parfois qu'elle est **de variance finie**.

Formule de Koenig-Huygens. $\boxed{\textbf{Pour } X \in \mathbb{L}^2(\Omega, \mathbb{R}), \textbf{ on a } \mathrm{V}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2.}$

Preuve. Posons $m = \mathrm{E}(X)$, alors par linéarité de l'espérance,

$$\begin{aligned} \mathrm{V}(X) &= \mathrm{E}\big((X-m)^2\big) = \mathrm{E}(X^2 - 2mX + m^2) \\ &= \mathrm{E}(X^2) - 2m\,\mathrm{E}(X) + m^2 \\ &= \mathrm{E}(X^2) - 2\,\mathrm{E}(X)^2 + \mathrm{E}(X)^2 = \mathrm{E}(X^2) - \mathrm{E}(X)^2 . \end{aligned}$$

Exemple de la loi zéta.

Rappelons que, pour $x > 1$, on pose $\zeta(x) = \displaystyle\sum_{n=1}^{+\infty} \frac{1}{n^x}$ (**fonction zéta de Riemann**).

On dit qu'une variable aléatoire X sur $(\Omega, \mathcal{A}, P)$ **suit la loi zéta de paramètre** x, et on note $X \sim \mathcal{Z}(x)$, si on a $X(\Omega) = \mathbb{N}^*$ et, pour tout $n \in \mathbb{N}^*$, $P(X = n) = \dfrac{1}{n^x\,\zeta(x)}$. *Le lecteur vérifiera rapidement la cohérence de cette définition.* Comme $n\,P(X = n) = \dfrac{1}{n^{x-1}\,\zeta(x)}$, on voit alors que X est d'espérance finie si et seulement si $x > 2$ et que, dans ce cas, on a $\mathrm{E}(X) = \dfrac{\zeta(x-1)}{\zeta(x)}$. Enfin, X admet un moment d'ordre deux si et seulement si $x > 3$ et, dans ce cas, $\mathrm{E}(X^2) = \dfrac{\zeta(x-2)}{\zeta(x)}$, puis $\mathrm{V}(X) = \dfrac{\zeta(x-2)\,\zeta(x) - \zeta(x-1)^2}{\zeta(x)^2}$.

Définition. Toujours pour $X \in \mathbb{L}^2(\Omega, \mathbb{R})$, on définit l'**écart-type** de X par

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

On a en effet $\mathrm{V}(X) = \mathrm{E}\big((X-m)^2\big) \geq 0$ par positivité de l'espérance.

Proposition. Soit $X \in \mathbb{L}^2(\Omega, \mathbb{R})$, soient a et b deux réels. Alors la variable $Y = aX + b$ appartient à $\mathbb{L}^2(\Omega, \mathbb{R})$, et on a $\boxed{\mathrm{V}(aX+b) = a^2\,\mathrm{V}(X).}$

Preuve. On a $Y^2 = a^2X^2 + 2ab\,X + b^2$, donc Y^2 est une combinaison linéaire de v.a. d'espérance finie, donc $Y \in \mathbb{L}^2(\Omega, \mathbb{R})$ et, par linéarité de l'espérance, $\mathrm{E}(Y) = a\,\mathrm{E}(X) + b$, *puis*

$$\begin{aligned}\mathrm{V}(Y) &= \mathrm{E}(Y^2) - \mathrm{E}(Y)^2 \\ &= a^2\,\mathrm{E}(X^2) + 2ab\,\mathrm{E}(X) + b^2 - \big(a\,\mathrm{E}(X) + b\big)^2 \\ &= a^2\,\big(\mathrm{E}(X^2) - \mathrm{E}(X)^2\big) \;=\; a^2\,\mathrm{V}(X)\,.\end{aligned}$$

Commentaire. En prenant la racine carrée, on a $\sigma(aX+b) = |a|\,\sigma(X)$.

Cette dernière relation s'interprète facilement: l'écart-type, comme la variance, est un indicateur de dispersion, donc le fait de translater les valeurs prises par une v.a. (remplacer X par $X+b$) ne modifie pas la dispersion, donc ne modifie pas l'écart-type. En revanche, multiplier par un coefficient réel a les valeurs prises (remplacer X par aX) multiplie les écarts par $|a|$, donc multiplie de même par $|a|$ l'écart-type.

Remarque. Une v.a.r.d. X, admettant un moment d'ordre deux, a une variance nulle si et seulement si elle est presque sûrement constante. En effet, posons $m = \mathrm{E}(X)$, alors la variable $(X-m)^2$ est positive et admet un moment d'ordre un, donc

$$\mathrm{V}(X) = 0 \iff \mathrm{E}\big((X-m)^2\big) = 0 \iff (X-m)^2 = 0 \text{ p.s.} \iff P(X=m) = 1\,.$$

Définition. Une v.a. $Y \in \mathbb{L}^2(\Omega, \mathbb{R})$ est dite **réduite** si $\sigma(Y) = 1$.

Si $X \in \mathbb{L}^2(\Omega, \mathbb{R})$ est telle que $\sigma(X) > 0$, i.e. si elle n'est pas presque sûrement constante, alors la variable $Y = \dfrac{X - \mathrm{E}(X)}{\sigma(X)}$ est centrée réduite.

Cas des lois usuelles.

• Si X est une variable de Poisson de paramètre $\lambda > 0$, alors $\mathrm{V}(X) = \lambda$.

On a vu, au paragraphe **IV.3.b.** *que X^2 est d'espérance finie, avec* $\mathrm{E}(X) = \lambda$ *et* $\mathrm{E}(X^2) = \lambda^2 + \lambda$, *on déduit* $\mathrm{V}(X) = \lambda$.

• Si X est une variable géométrique de paramètre $p \in]0,1[$, alors $\mathrm{V}(X) = \dfrac{q}{p^2}$ avec $q = 1-p$.

On a vu que X est d'espérance finie avec $\mathrm{E}(X) = \dfrac{1}{p} = \displaystyle\sum_{k=1}^{+\infty} k\,pq^{k-1}$, *cf. paragraphe* **IV.1.**
Par ailleurs, par la formule de transfert, la convergence de la série étant immédiate,

$$\mathrm{E}(X^2) \;=\; \sum_{k=1}^{+\infty} k^2\,pq^{k-1} \;=\; \sum_{k=1}^{+\infty} k(k-1)\,pq^{k-1} + \sum_{k=1}^{+\infty} k\,pq^{k-1}$$

$$= \quad pq \sum_{k=2}^{+\infty} k(k-1)\, q^{k-2} + \mathrm{E}(X)$$
$$= \quad pq\, \frac{2}{(1-q)^3} + \frac{1}{p} \;=\; \frac{2q}{p^2} + \frac{1}{p}\;.$$

On a en effet reconnu $\displaystyle\sum_{k=2}^{+\infty} k(k-1)\, x^{k-2} \;=\; \frac{\mathrm{d}^2}{\mathrm{d}x^2}\Big(\sum_{k=0}^{+\infty} x^k\Big) \;=\; \frac{\mathrm{d}^2}{\mathrm{d}x^2}\Big(\frac{1}{1-x}\Big) \;=\; \frac{2}{(1-x)^3}$, *relation valable pour* $x \in \,]-1,1[$ *d'après le cours sur les séries entières. Enfin,*

$$\mathrm{V}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2} = \frac{q}{p^2}\;.$$

c. <u>Covariance</u>.

<u>Définition.</u> Soient X et Y deux v.a.r.d. sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$, chacune admettant un moment d'ordre deux. On définit alors la **covariance** de X et Y par la formule

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

<u>Commentaire.</u> On sait que X et Y sont d'espérance finie, posons $m = \mathrm{E}(X)$ et $m' = \mathrm{E}(Y)$. Alors la variable

$$(X-m)(Y-m') = XY - mY - m'X + mm'$$

est une combinaison linéaire de v.a.r.d. appartenant à $\mathbb{L}^1(\Omega, \mathbb{R})$, donc elle est aussi d'espérance finie et, par linéarité,

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{E}\big((X-m)(Y-m')\big)\\ &= \mathrm{E}(XY) - m\,\mathrm{E}(Y) - m'\,\mathrm{E}(X) + mm'\\ &= \mathrm{E}(XY) - mm'\;.\end{aligned}$$

On a donc une **formule de Koenig-Huygens généralisée**:

Pour $X \in \mathbb{L}^2(\Omega, \mathbb{R})$ et $Y \in \mathbb{L}^2(\Omega, \mathbb{R})$, **on a** $\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)$.

<u>Proposition.</u> La covariance est une forme bilinéaire symétrique positive sur le $\mathbb{R}$-espace vectoriel $\mathbb{L}^2(\Omega, \mathbb{R})$.

Preuve. La symétrie est évidente. La bilinéarité est facile, et résulte essentiellement de la linéarité de l'espérance. Enfin, si $X \in \mathbb{L}^2(\Omega, \mathbb{R})$, *on a* $\mathrm{Cov}(X,X) = \mathrm{V}(X) \geq 0$, *d'où la positivité.*

<u>Commentaire.</u> Il manque à la covariance le caractère "défini" pour être un produit scalaire, il manque donc aussi à l'écart-type l'axiome de séparation pour être une norme. En effet, on a vu dans le paragraphe précédent que $\mathrm{Cov}(X,X) = \mathrm{V}(X)$ est nul si et seulement si la variable X est presque sûrement constante (i.e. il existe un réel C tel que $P(X=C)=1$).

<u>Commentaire.</u> Il y a donc aussi pour la covariance, qui est "presque" un produit scalaire sur $L^2(\Omega, \mathbb{R})$, une inégalité de Cauchy-Schwarz, qui peut s'écrire

$$\forall (X,Y) \in \big(L^2(\Omega,\mathbb{R})\big)^2 \qquad \big|\mathrm{Cov}(X,Y)\big| \leq \sigma(X)\,\sigma(Y)\;,$$

nous n'étudierons pas le cas d'égalité.

Proposition. Soient X et Y dans $\mathbb{L}^2(\Omega,\mathbb{R})$. Si les variables X et Y sont indépendantes, alors $\mathrm{Cov}(X,Y)=0$.

C'est une conséquence immédiate de la relation $\mathrm{E}(XY)=\mathrm{E}(X)\,\mathrm{E}(Y)$ lorsque X et Y sont indépendantes.

Commentaire. Deux variables X et Y telles que $\mathrm{Cov}(X,Y)=0$ sont dites **décorrélées**. Deux variables indépendantes sont donc décorrélées, mais la réciproque est fausse.

Exemple. On lance deux dés équilibrés, on note X_1 et X_2 les résultats du premier et du deuxième dé, puis on pose $S=X_1+X_2$ et $D=X_1-X_2$. Les variables S et D ne sont pas indépendantes. En effet, on a $S(\Omega)=[\![2,12]\!]$, $D(\Omega)=[\![-5,5]\!]$. Comme l'événement $\{S=12\}$ ne peut se produire que si $X_1=X_2=6$, ce qui entraîne $D=0$, on a par exemple

$$P(S=12\ ,\ D=1)=0\ ,$$

alors que $P(S=12)$ et $P(D=1)$ sont tous les deux non nuls.

En revanche, les variables S et D sont décorrélées: en effet, en développant par bilinéarité de la covariance, et en utilisant la symétrie,

$$\mathrm{Cov}(S,D)=\mathrm{Cov}(X_1+X_2,X_1-X_2)=\mathrm{Cov}(X_1,X_1)-\mathrm{Cov}(X_2,X_2)=\mathrm{V}(X_1)-\mathrm{V}(X_2)=0$$

puisque X_1 et X_2 ont la même loi.

Proposition. Soient $X_1,\cdots,X_n$ des variables appartenant à $\mathbb{L}^2(\Omega,\mathbb{R})$. On a alors

$$\mathrm{V}\Big(\sum_{i=1}^n X_i\Big)=\sum_{i=1}^n \mathrm{V}(X_i)+2\sum_{i<j}\mathrm{Cov}(X_i,X_j)\ .$$

En particulier, si ces variables sont deux à deux indépendantes, alors

$$\mathrm{V}\Big(\sum_{i=1}^n X_i\Big)=\sum_{i=1}^n \mathrm{V}(X_i)\ .$$

Preuve. La première relation est juste une conséquence du caractère bilinéaire et symétrique de la covariance, c'est un développement par bilinéarité. La deuxième relation est valable car l'indépendance deux à deux des X_i entraîne leur décorrélation.

Commentaire. La covariance étant "presque" un produit scalaire sur $\mathbb{L}^2(\Omega,\mathbb{R})$, deux variables décorrélées pourraient être qualifiées d'"orthogonales". Le fait que, si les variables X_i sont deux à deux décorrélées, alors la variance de leur somme est égale à la somme de leurs variances, est analogue à la "relation de Pythagore".

d. Inégalités probabilistes.

On a déjà vu l'inégalité de Markov, *cf.* paragraphe **IV.3.c.**

Mentionnons une autre inégalité classique:

Inégalité de Bienaymé-Tchebychev. Soit X une variable aléatoire réelle discrète admettant un moment d'ordre deux. Alors

$$\forall\varepsilon>0\qquad P\Big(\big|X-\mathrm{E}(X)\big|\geq\varepsilon\Big)\leq\frac{\mathrm{V}(X)}{\varepsilon^2}\ .$$

Preuve. *Posons* $m = \mathrm{E}(X)$. *Soit la variable* $Y = (X-m)^2$. *Comme* X *et* m *(v.a. constante) appartiennent à* $\mathbb{L}^2(\Omega, \mathbb{R})$, *il en est de même de* $X - m$, *donc* $Y \in \mathbb{L}^1(\Omega, \mathbb{R})$, *i.e.* Y *est d'espérance finie. Par ailleurs* Y *est positive, on peut donc lui appliquer l'inégalité de Markov. Notons enfin que* $\mathrm{E}(Y) = \mathrm{E}\big((X-m)^2\big) = \mathrm{V}(X)$. *Donc, par Markov,*

$$P\Big(\big|X - \mathrm{E}(X)\big| \geq \varepsilon\Big) = P(Y \geq \varepsilon^2) \leq \frac{\mathrm{E}(Y)}{\varepsilon^2} = \frac{\mathrm{V}(X)}{\varepsilon^2} \ .$$

Commentaire. Cette inégalité donne une majoration de la probabilité qu'une variable aléatoire sorte d'un intervalle centré autour de sa valeur moyenne (i.e. son espérance). Par exemple, si $X \in \mathbb{L}^2(\Omega, \mathbb{R})$ est centrée réduite ($\mathrm{E}(X) = 0$ et $\mathrm{V}(X) = 1$), en choisissant $\varepsilon = 2$, la probabilité pour que $X(\omega)$ soit en dehors de l'intervalle $]-2, 2[$ est majorée par $\dfrac{1}{4}$. La probabilité d'être en dehors de l'intervalle $]-10, 10[$ est majorée par $\dfrac{1}{100}$.

Proposition (loi faible des grands nombres). Soit $(X_n)_{n\geq 1}$ **une suite i.i.d. de variables aléatoires de variance finie, alors si** $S_n = \displaystyle\sum_{k=1}^{n} X_k$**, si on pose** $m = \mathrm{E}(X_1)$ **et** $\sigma = \sigma(X_1)$**, on a pour tout** $\varepsilon > 0$, $\displaystyle\lim_{n\to+\infty} P\left(\Big|\frac{S_n}{n} - m\Big| \geq \varepsilon\right) = 0$**. On a, plus précisément, l'estimation suivante:**

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

Preuve. *Les variables* X_k *ayant toutes la même loi, on a* $\mathrm{E}(X_k) = m$ *et* $\sigma(X_k) = \sigma$ *pour tout* $k \in \mathbb{N}^*$. *Posons* $Y_n = \dfrac{S_n}{n} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$. *Par linéarité de l'espérance, on a* $\mathrm{E}(Y_n) = m$. *Comme* $\mathbb{L}^2(\Omega, \mathbb{R})$ *est un espace vectoriel, on a* $Y_n \in \mathbb{L}^2(\Omega, \mathbb{R})$, *ce qui justifie le calcul suivant et l'utilisation de l'inégalité de Bienaymé-Tchebychev. Les variables* X_k *étant deux à deux indépendantes, on a*

$$\mathrm{V}(Y_n) = \frac{1}{n^2}\,\mathrm{V}\Big(\sum_{k=1}^{n} X_k\Big) = \frac{1}{n^2}\,\sum_{k=1}^{n}\mathrm{V}(X_k) = \frac{1}{n^2}\,n\sigma^2 = \frac{\sigma^2}{n} \ .$$

Pour tout $\varepsilon > 0$, *on a donc*

$$P\Big(\big|Y_n - \mathrm{E}(Y_n)\big| \geq \varepsilon\Big) \leq \frac{\mathrm{V}(Y_n)}{\varepsilon^2} \ , \quad \textit{soit} \quad P\big(|Y_n - m| \geq \varepsilon\big) \leq \frac{\sigma^2}{n\varepsilon^2} \ .$$

On en déduit immédiatement que, pour tout $\varepsilon > 0$, *on a* $\displaystyle\lim_{n\to+\infty} P\big(|Y_n - m| \geq \varepsilon\big) = 0$.

Interprétation. Imaginons une répétition d'expériences aléatoires identiques et indépendantes, le résultat de la k-ième expérience étant modélisé par une variable aléatoire X_k (variable indicatrice de l'obtention d'un succès, temps d'attente d'un succès dans un jeu), avec $k \in \mathbb{N}^*$. Notons $m = \mathrm{E}(X_1) = \mathrm{E}(X_k)$ l'espérance de la variable X_k, qui est alors la **moyenne théorique** des résultats obtenus. Lors de cette répétition d'expériences, pour tout n, la variable $Y_n = \dfrac{S_n}{n} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$ représente la **moyenne observée** des résultats sur les n premiers essais. La loi faible des grands nombres donne une estimation de la

différence entre moyenne observée et moyenne théorique, et dit d'une certaine façon que la moyenne observée "tend vers" la moyenne théorique lorsque le nombre n de répétitions de l'expérience tend vers $+\infty$. En vocabulaire des probabilités, on dit que la suite de variables aléatoires (Y_n) **converge en probabilité** (*notion hors programme*) vers la variable aléatoire constante de valeur m.

V. Fonction génératrice d'une variable aléatoire à valeurs dans $\mathbb{N}$.

1. Définition et premières propriétés.

Définition. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. On définit sa **fonction génératrice**, notée G_X, par la relation

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

pour tout réel t tel que cette série converge absolument.

Commentaire. La deuxième égalité résulte du théorème du transfert.

Proposition. La fonction génératrice de X est la somme d'une série entière, dont le rayon de convergence est au moins égal à 1, et qui converge normalement sur $[-1,1]$. La fonction génératrice est définie et continue sur l'intervalle $[-1,1]$.

Preuve. Posons $u_n(t) = P(X=n)\, t^n$. Chaque fonction u_n est continue sur $\mathbb{R}$, donc sur le segment $S=[-1,1]$, et $\|u_n\|_{\infty,S} = P(X=n)$ est le terme général d'une série convergente. On a donc prouvé la convergence normale de la série de fonctions $\sum u_n$ sur S, ce qui entraîne la continuité de la fonction somme G_X sur ce segment. Le rayon de convergence R de la série entière vérifie donc $R \geq 1$.

Remarque. On a toujours $G_X(1) = \sum_{n=0}^{+\infty} P(X=n) = 1$.

Proposition. La loi d'une variable aléatoire X à valeurs dans $\mathbb{N}$ est caractérisée par sa fonction génératrice G_X.

Preuve. En effet, la série entière définissant G_X sur $[-1,1]$ est nécessairement la série de Taylor de la fonction G_X. On a donc

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

2. Cas des lois usuelles.

Notons d'abord que, dans le cas d'une variable aléatoire X ne prenant qu'un nombre fini de valeurs (entières), alors G_X est une fonction polynomiale, elle est donc définie sur $\mathbb{R}$ tout entier. C'est notamment le cas des lois étudiées en première année.

• **Loi de Bernoulli.** Si $X \sim \mathcal{B}(p)$ avec $0<p<1$, en posant $q=1-p$, on a

$$\forall t \in \mathbb{R} \qquad G_X(t) = q + pt\ .$$

• **Loi binomiale.** Si $X \sim \mathcal{B}(n,p)$ avec $0<p<1$ et $n \in \mathbb{N}^*$, toujours avec $q=1-p$, on a

$$\forall t \in \mathbb{R} \qquad G_X(t) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} t^k = (q+pt)^n \ .$$

• **Loi uniforme sur $[\![1,n]\!]$.** Si $X \sim \mathcal{U}_{[\![1,n]\!]}$ avec $n \in \mathbb{N}^*$, alors

$$\forall t \in \mathbb{R} \qquad G_X(t) = \frac{1}{n} \sum_{k=1}^{n} t^k = \frac{t}{n} \frac{1-t^n}{1-t}$$

(la deuxième expression étant valable si $t \neq 1$).

• **Loi de Poisson.** Si $X \sim \mathcal{P}(\lambda)$ avec $\lambda > 0$, alors

$$\forall t \in \mathbb{R} \qquad G_X(t) = \sum_{n=0}^{+\infty} e^{-\lambda} \frac{\lambda^n}{n!} t^n = e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}$$

(on reconnaît une série exponentielle, donc de rayon de convergence infini).

• **Loi géométrique.** Si $X \sim \mathcal{G}(p)$ avec $0 < p < 1$, en posant $q = 1-p$, on a

$$G_X(t) = \sum_{n=1}^{+\infty} pq^{n-1} t^n = pt \sum_{n=0}^{+\infty} (qt)^n \ .$$

On reconnaît une série géométrique de raison qt, qui converge si et seulement si $|qt| < 1$, on a donc un rayon de convergence qui vaut $\frac{1}{q}$, et

$$\forall t \in \left]-\frac{1}{q}, \frac{1}{q}\right[\qquad G_X(t) = \frac{pt}{1-qt} \ .$$

Remarque. Pour toutes les lois "usuelles" listées ci-dessus, le rayon de convergence est strictement plus grand que 1. Voici un exemple où le rayon de convergence vaut 1: reprenons la loi zéta de paramètre x avec $x > 1$, notée $\mathcal{Z}(x)$, mentionnée dans le paragraphe **IV.4.b.** Si $X \sim \mathcal{Z}(x)$, alors pour tout n entier naturel non nul, $P(X = n) = \frac{1}{n^x \, \zeta(x)}$. On a donc $G_X(t) = \frac{1}{\zeta(x)} \sum_{n=1}^{+\infty} \frac{t^n}{n^x}$, la règle de d'Alembert montre immédiatement que cette série entière a toujours pour rayon de convergence 1.

3. Lien avec l'espérance et la variance.

Dans tout ce paragraphe, X est une variable aléatoire sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, à valeurs dans $\mathbb{N}$.

Commençons par un cas particulier simple:

Proposition 1. Si la série génératrice G_X a un rayon de convergence R tel que $R > 1$, alors $X \in \mathbb{L}^2(\Omega, \mathbb{R})$ et on a les relations

$$\mathrm{E}(X) = G'_X(1) \qquad \text{et} \qquad \mathrm{V}(X) = G'_X(1) + G''_X(1) - G'_X(1)^2 \ .$$

Preuve. On sait que, dans ce cas, la fonction génératrice G_X est de classe $\mathcal{C}^\infty$ sur l'intervalle de convergence $I =]-R, R[$, et que ses dérivées successives sur cet intervalle peuvent se calculer par dérivation terme à terme. On a donc, pour tout $t \in I$,

$$G'_X(t) = \sum_{n=1}^{+\infty} n\, P(X=n)\, t^{n-1} \qquad \text{et} \qquad G''_X(t) = \sum_{n=2}^{+\infty} n(n-1)\, P(X=n)\, t^{n-2}\,,$$

les séries considérées étant absolument convergentes. Comme $1 \in I$*, on déduit que* $X \in \mathbb{L}^1(\Omega, \mathbb{R})$ *et, en évaluant pour* $t = 1$*, on a* $G'_X(1) = \sum_{n=1}^{+\infty} n\, P(X=n) = \mathrm{E}(X)$ *directement, et par le théorème du transfert,* $X^2 - X \in \mathbb{L}^1(\Omega, \mathbb{R})$*, et*

$$G''_X(1) = \sum_{n=2}^{+\infty} n(n-1)\, P(X=n) = \mathrm{E}(X^2 - X)\,.$$

Par linéarité de l'espérance, on déduit que X *admet un moment d'ordre deux, à savoir* $\mathrm{E}(X^2) = \mathrm{E}(X^2 - X) + \mathrm{E}(X) = G''_X(1) + G'_X(1)$*, et enfin par la formule de Koenig-Huygens, on obtient l'expression annoncée pour la variance.*

Remarque. Si la série génératrice G_X a un rayon de convergence R avec $R > 1$, alors la variable X admet des moments de tout ordre, i.e. pour tout k entier naturel, X^k est d'espérance finie. *En effet, en reprenant la démonstration ci-dessus, on a, pour tout* k *entier naturel,* $G_X^{(k)}(1) = \sum_{n=k}^{+\infty} n(n-1)\cdots(n-k+1)\, P(X=n) = \sum_{n=k}^{+\infty} \frac{n!}{(n-k)!}\, P(X=n)$*, cette série étant absolument convergente. Par le critère des équivalents sur les séries à termes positifs, on déduit la convergence absolue de la série* $\sum_{n\geq 0} n^k\, P(X=n)$*, soit l'existence du moment d'ordre* k *de la variable* X*.*

Le programme de PSI va un peu plus loin et propose l'énoncé suivant (*seule la démonstration du sens direct est exigible*):

Proposition 2. La variable aléatoire X (à valeurs dans $\mathbb{N}$) est d'espérance finie si et seulement si G_X est dérivable en 1 et, si tel est le cas, $\mathrm{E}(X) = G'_X(1)$**.**

Preuve. Posons $p_n = P(X=n)$ *pour simplifier.*

• *Le sens direct est facile: en effet, si* X *est d'espérance finie, alors la série* $\sum np_n$ *converge. Posons* $u_n(t) = p_n t^n$ *pour* $n \in \mathbb{N}$ *et* $t \in [-1,1]$*, on a alors* $\|u'_n\|_\infty = np_n$*, on a donc prouvé la convergence normale (donc uniforme) de la série de fonctions* $\sum u'_n$ *sur* $[-1,1]$*. Comme d'autre part les* u_n *sont de classe* $\mathcal{C}^1$ *et la série* $\sum u_n$ *converge simplement sur* $[-1,1]$*, le théorème de dérivation des séries de fonctions s'applique: la fonction* G_X *est de classe* $\mathcal{C}^1$ *sur* $[-1,1]$ *et, en particulier, elle est dérivable au point 1. Une remarque: cette démonstration considère le cas où* $R = 1$ *puisque, dans ce cas, la fonction génératrice* G_X *est définie sur* $[-1,1]$ *et* **seulement** *sur cet intervalle, la dérivabilité en 1 est donc seulement une dérivabilité à gauche. Lorsque* $R > 1$*, je renvoie le lecteur à la preuve de la proposition 1 ci-dessus.*

• *Pour la réciproque, supposons* G_X *dérivable (à gauche) en 1. Soit* $t \in [0,1[$*. Comme* G_X *est dérivable sur* $[t,1]$*, d'après l'égalité des accroissements finis, il existe* $c_t \in]t,1[$ *tel que*

$$\frac{G_X(t) - G_X(1)}{t-1} = G'_X(c_t) = \sum_{n=1}^{+\infty} n\, p_n\, c_t^{n-1}\,.$$

Pour tout N entier naturel, comme $t \leq c_t$ et que les coefficients de la série entière sont positifs, on a alors

$$\sum_{n=1}^{N} n\, p_n\, t^{n-1} \leq \sum_{n=1}^{N} n\, p_n\, c_t^{n-1} \leq \sum_{n=1}^{+\infty} n\, p_n\, c_t^{n-1} = \frac{G_X(t) - G_X(1)}{t-1} \ .$$

En passant à la limite (pour $t \to 1^-$) dans cette inégalité entre les membres extrêmes, on obtient

$$\forall N \in \mathbb{N} \qquad \sum_{n=1}^{N} n\, p_n \leq G'_X(1) \ .$$

La série $\sum n\, p_n$ est une série à termes positifs dont les sommes partielles sont majorées, elle est donc convergente, ce qui prouve que X est d'espérance finie.

Application aux lois usuelles. Ce qui précède va nous permettre de trouver (ou retrouver) l'expression de l'espérance et de la variance pour les lois usuelles. Pour toutes ces lois, on a $R > 1$, ce qui facilite les justifications.

• **Loi binomiale.** Si $X \sim \mathcal{B}(n,p)$ avec $0 < p < 1$ et $n \in \mathbb{N}^*$, en posant $q = 1-p$, on a vu que $G_X(t) = (pt+q)^n$ pour tout réel t (c'est une fonction polynomiale), on a donc $G'_X(t) = np(pt+q)^{n-1}$ et $G''_X(t) = n(n-1)p^2\,(pt+q)^{n-2}$, on retrouve alors, puisque $p+q=1$,

$$\mathrm{E}(X) = G'_X(1) = np \ , \quad \text{puis} \quad G''_X(1) = n(n-1)p^2 \ ,$$

et enfin $\mathrm{V}(X) = n(n-1)p^2 + np - (np)^2 = np(1-p) = npq$.

Si $X \sim \mathcal{B}(n,p)$, alors $G_X(t) = (pt+1-p)^n$, puis $\mathrm{E}(X) = np$ et $\mathrm{V}(X) = np(1-p)$.

• **Loi de Poisson.** Si $X \sim \mathcal{P}(\lambda)$ avec $\lambda > 0$, on a vu que $G_X(t) = e^{\lambda(t-1)}$ pour tout $t \in \mathbb{R}$. Donc $G'_X(t) = \lambda\, e^{\lambda(t-1)}$, $G''_X(t) = \lambda^2\, e^{\lambda(t-1)}$. Donc $\mathrm{E}(X) = G'_X(1) = \lambda$, et

$$\mathrm{V}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda \ .$$

Si $X \sim \mathcal{P}(\lambda)$ avec $\lambda > 0$, alors $G_X(t) = e^{\lambda(t-1)}$, puis $\mathrm{E}(X) = \lambda$ et $\mathrm{V}(X) = \lambda$.

• **Loi géométrique.** Si $X \sim \mathcal{G}(p)$ avec $0 < p < 1$, en posant $q = 1-p$, on a vu que, pour $t \in \left]-\dfrac{1}{q}, \dfrac{1}{q}\right[$, on a $G_X(t) = \dfrac{pt}{1-qt}$. On calcule $G'_X(t) = \dfrac{p}{(1-qt)^2}$, puis $G''_X(t) = \dfrac{2pq}{(1-qt)^3}$. On en déduit que $\mathrm{E}(X) = G'_X(1) = \dfrac{p}{(1-q)^2} = \dfrac{1}{p}$, et enfin que

$$\mathrm{V}(X) = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q+p-1}{p^2} = \frac{q}{p^2} \ .$$

Si $X \sim \mathcal{G}(p)$ alors, avec $q = 1-p$, on a $G_X(t) = \dfrac{pt}{1-qt}$, puis $\mathrm{E}(X) = \dfrac{1}{p}$ et $\mathrm{V}(X) = \dfrac{q}{p^2}$.

4. Somme de deux variables indépendantes.

Proposition. Soient X et Y deux variables aléatoires indépendantes, à valeurs dans $\mathbb{N}$. On a alors

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

Preuve 1. Soit $t \in [-1,1]$. Comme X et Y sont indépendantes, il en est de même de t^X et t^Y qui sont fonctions de X et Y respectivement. Comme $|t| \leq 1$, les variables t^X et t^Y sont d'espérance finie. Il en est alors de même de leur produit $t^X\, t^Y = t^{X+Y}$, et on a la relation:

$$\mathrm{E}(t^{X+Y}) = \mathrm{E}(t^X)\,\mathrm{E}(t^Y)\ , \quad \text{soit} \quad G_{X+Y}(t) = G_X(t)\, G_Y(t)\ .$$

Preuve 2. Posons $Z = X + Y$. Selon un calcul déjà proposé dans le paragraphe **III.6.**, *pour tout n entier naturel, on a $P(Z=n) = \sum_{k=0}^{n} P(X=k)\, P(Y=n-k)$. Pour $t \in [-1,1]$, on a alors*

$$G_{X+Y}(t) = G_Z(t) = \sum_{n=0}^{+\infty} \Big(\sum_{k=0}^{n} P(X=k)\, P(Y=n-k) \Big)\, t^n\ .$$

On reconnaît donc un produit de Cauchy. Les séries définissant $G_X(t)$ et $G_Y(t)$ étant toutes deux absolument convergentes, on a donc $G_Z(t) = G_X(t)\, G_Y(t)$.

Exemple des lois de Poisson. Si $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ avec $\lambda > 0$ et $\mu > 0$, si X et Y sont indépendantes, on a alors $G_X(t) = e^{\lambda(t-1)}$, $G_Y(t) = e^{\mu(t-1)}$, donc

$$G_{X+Y}(t) = G_X(t)\, G_Y(t) = e^{(\lambda+\mu)(t-1)}\ .$$

Comme la fonction génératrice caractérise la loi d'une v.a. à valeurs dans $\mathbb{N}$, on déduit que $X + Y \sim \mathcal{P}(\lambda + \mu)$, résultat déjà obtenu dans le paragraphe **III.6.**

Généralisation. Soient X_1, $\cdots$, X_n des variables indépendantes à valeurs dans $\mathbb{N}$, on a alors

$$\forall t \in [-1,1] \qquad G_{X_1+\cdots+X_n}(t) = \prod_{k=1}^{n} G_{X_k}(t)\ .$$

Preuve. Récurrence immédiate, en utilisant pour l'hérédité le lemme des coalitions qui nous garantit que les variables $X_1 + \cdots + X_{n-1}$ et X_n sont indépendantes.

Exemple. On retrouve ainsi, à l'aide des fonctions génératrices, le résultat suivant du cours de 1ère année: si n variables indépendantes X_1, $\cdots$, X_n suivent chacune la loi de Bernoulli $\mathcal{B}(p)$, alors leur somme $S = \sum_{k=1}^{n} X_k$ suit la loi binomiale $\mathcal{B}(n,p)$. *En effet, pour tout $k \in [\![1,n]\!]$, on a $G_{X_k}(t) = pt + q$ avec $q = 1-p$, donc $G_X(t) = \prod_{k=1}^{n} G_{X_k}(t) = (pt+q)^n$.*

Comme la fonction génératrice caractérise la loi, on conclut que $S \sim \mathcal{B}(n,p)$.