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 $\mathbb{N}^2 = \bigsqcup_{k\in\mathbb{N}} (\{k\}\times\mathbb{N})$ 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**)

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

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 $\left(\frac{1}{n^2}\right)_{n\in\mathbf{Z}^*}$ 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 $\left(\dfrac{1}{(k+l)^\alpha}\right)_{(k,l)\in(\mathbb{N}^*)^2}$ 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 $I_n = \{(k,l) \in (\mathbb{N}^*)^2 \mid k+l = n\}$. 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

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

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",*

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

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**.

<u>Conséquences de cette définition.</u>

• $\boxed{P(\emptyset) = 0.}$

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 $k > n$.*

<u>Commentaire.</u> 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, $P(B) = P(A) + P(B \setminus A) \geq P(A)$.

<u>Définitions.</u> 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. <u>Propriétés de continuité monotone, et conséquences</u>.

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

<u>Théorème de continuité croissante</u>. 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, $\displaystyle\bigcup_{n=0}^{+\infty} A_n = \bigsqcup_{n=0}^{+\infty} B_n$*: en effet, l'inclusion dans le sens indirect est immédiate puisque* $B_n \subset A_n$ *pour tout* n. *Puis, si* $\omega \in \displaystyle\bigcup_{n=0}^{+\infty} A_n$, *soit* $n_0 = \min\{n \in \mathbb{N} \mid \omega \in A_n\}$, *on a alors* $\omega \in B_{n_0}$ *donc* $\omega \in \displaystyle\bigcup_{n=0}^{+\infty} B_n$, *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* $\displaystyle\sum_{n\geq 1} \big(P(A_n) - P(A_{n-1})\big)$ *est donc aussi convergente, de somme* $\displaystyle\lim_{n\to+\infty} P(A_n) - P(A_0)$.
En utilisant la propriété de σ*-additivité de la probabilité* P, *on obtient*

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

Théorème de continuité décroissante. 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

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

Preuve. *Par passage au complémentaire. Posons* $C_n = \overline{A_n}$ *pour tout* n, *alors la suite* (C_n) *est croissante, donc* $\displaystyle\lim_{n\to+\infty} P(C_n) = P\Big(\bigcup_{n=0}^{+\infty} C_n\Big)$ *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.

Conséquences. 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. Conditionnement.

Définition. 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 $P(A|B)$.

Proposition. 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

$$\begin{aligned} P_B\Big(\bigsqcup_{n=0}^{+\infty} A_n\Big) &= \frac{P\Big(\Big(\bigsqcup_{n=0}^{+\infty} A_n\Big) \cap B\Big)}{P(B)} = \frac{P\Big(\bigsqcup_{n=0}^{+\infty} (A_n \cap B)\Big)}{P(B)} \\ &= \frac{\sum_{n=0}^{+\infty} P(A_n \cap B)}{P(B)} = \sum_{n=0}^{+\infty} \frac{P(A_n \cap B)}{P(B)} = \sum_{n=0}^{+\infty} P_B(A_n) , \end{aligned}$$

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

Formule des probabilités composées.

(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

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

Preuve. 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)**.

Commentaire. 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

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

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 $P\Bigg(\bigcap_{i\in J} A_i\Bigg) = \prod_{i\in J} P(A_i)$.

6. **<u>Cas des univers finis ou dénombrables.</u>** 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.

<u>Proposition.</u> 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 $\omega \in \Omega$.

<u>Preuve.</u> • *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*

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

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*

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

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*

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

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 <u>dénombrable</u>.

Tout ceci est bien difficile à expliquer sans sortir très largement du cadre du programme.

III. <u>Variables aléatoires discrètes</u>.

1. <u>Définition</u>.

<u>Définition.</u> 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 $\mathcal{A}$.

<u>Commentaire.</u> 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 $(X = x)$.

<u>Commentaire.</u> 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.

<u>Proposition.</u> 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.**

<u>Preuve.</u> 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).

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

<u>Remarque.</u> 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.

<u>Notion de fonction d'une variable aléatoire.</u> 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. <u>Loi d'une variable aléatoire discrète</u>.

<u>Proposition et définition.</u> 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 <u>loi de la variable X</u>.

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,

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

<u>Commentaire 1.</u> 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.

<u>Commentaire 2.</u> 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.

<u>Commentaire 3.</u> 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 $\displaystyle P(A) = \sum_{x\in A} p_x$ (*cf.* proposition à la fin du 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 $(p_x)_{x\in E}$.

<u>Notation.</u> 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.

<u>Attention!</u> 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.

<u>Propriété.</u> Si $X \sim X'$, alors $f(X) \sim f(X')$.

<u>Preuve.</u> 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. <u>Une loi usuelle: la loi géométrique</u>.

<u>Définition.</u> 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,

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