Chapitre XIV
Probabilités discrètes

Table des matières

Partie A : Espaces probabilisés **2**
1. Espaces probabilisables . . . 2
2. Espaces probabilisés . . . 4
3. Propriétés élémentaires des probabilités . . . 8
4. Loi/Distribution de probabilité discrète . . . 11

Partie B : Probabilités conditionnelles **14**
1. Conditionnement . . . 14
2. Formules conditionnelles . . . 15
3. Événements indépendants . . . 20

Partie C : Variables aléatoires discrètes **24**
1. Variables aléatoires discrètes . . . 24
2. Rappels des lois usuelles sur un univers fini . . . 29
3. Lois usuelles sur un univers dénombrable . . . 30
4. Couples de variables aléatoires . . . 34
5. Variables aléatoires indépendantes . . . 37

Partie D : Espérance, variance **42**
1. Espérance . . . 42
2. Variance . . . 55
3. Covariance . . . 63
4. Espérance, variance et indépendance . . . 64
5. Inégalités de concentration . . . 68

Partie E : Fonctions génératrices **71**
1. Définition et expression d'une fonction génératrice . . . 71
2. Propriétés des fonctions génératrices . . . 73
3. Fonctions génératrices des lois usuelles . . . 80
4. Fonction génératrice d'une somme de variables indépendantes . . . 81

Dans ce chapitre, on utilisera la notation $\overline{A}$ pour désigner le complémentaire d'une sous-partie A d'un ensemble.

Partie A

Espaces probabilisés

1. Espaces probabilisables

a. Définitions

Définition 1. *Tribu*

Soit Ω un ensemble non vide et $\mathcal{T} \subset \mathcal{P}(\Omega)$. On dit que $\mathcal{T}$ est une **tribu** sur Ω si :

- $\Omega \in \mathcal{T}$;
- $\mathcal{T}$ est stable par passage au complémentaire i.e. pour tout $A \in \mathcal{T}$, $\overline{A} \in \mathcal{T}$;
- $\mathcal{T}$ est stable par réunion dénombrable i.e. pour toute suite $(A_n)_{n\in\mathbb{N}}$ à valeurs dans $\mathcal{T}$, $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{T}$.

Remarque 1.

On trouve parfois la terminologie σ-algèbre pour désigner une tribu et la notation $\mathcal{A}$ au lieu de $\mathcal{T}$.

Définition 2. *Espace probabilisable*

Soit Ω un ensemble non vide et $\mathcal{T}$ une tribu sur Ω. Le couple $(\Omega, \mathcal{T})$ est appelé **espace probabilisable**.

Exemple 1.

Soit Ω un ensemble non vide.

- $\mathcal{T} = \{\emptyset, \Omega\}$ est une tribu appelée **tribu grossière** ;
- $\mathcal{T} = \mathcal{P}(\Omega)$ est une tribu ;
- Si $A \subset \Omega$, $\mathcal{T} = \{\emptyset, A, \overline{A}, \Omega\}$ est une tribu.

Exercice 1.

Soit Ω un ensemble non vide et $\mathcal{F} \subset \mathcal{P}(\Omega)$.

1. Montrer que l'intersection de toutes les tribus contenant $\mathcal{F}$ (en existe-t-il ?) est une tribu. On note cette tribu $\sigma(\mathcal{F})$ et on l'appelle **tribu engendrée par** $\mathcal{F}$.
2. Montrer que $\sigma(\mathcal{F})$ est la plus petite tribu contenant $\mathcal{F}$.

Remarque 2. *Vocabulaire*

Soit $(\Omega, \mathcal{T})$ un espace probabilisable. On utilise la terminologie suivante :
- Ω est l'**univers** ;
- les éléments A de $\mathcal{T}$ sont les **événements** ;

b. Les événements

Proposition 1.

Soit $(\Omega, \mathcal{T})$ un espace probabilisable. On a les propriétés suivantes :
- $\emptyset \in \mathcal{T}$;
- $\mathcal{T}$ est stable par réunion finie ;
- $\mathcal{T}$ est stable par intersection au plus dénombrable.
- pour tout $A, B \in \mathcal{T}$, $A \setminus B \in \mathcal{T}$.

Démonstration.

- On a $\Omega \in \mathcal{T}$ donc $\emptyset = \overline{\Omega} \in \mathcal{T}$.
- Soit $n \in \mathbb{N}^*$ et $A_1, ..., A_n \in \mathcal{T}$. On considère la suite $(A_k)_{k\in\mathbb{N}}$ telle que $A_k = A_n \in \mathcal{T}$ pour tout entier $k > n$. Par stabilité de $\mathcal{T}$ par union dénombrable, on a $\bigcup_{k\in\mathbb{N}} A_k \in \mathcal{T}$. Par suite, on a :
$$A_1 \cup ... \cup A_{n-1} \cup A_n = A_1 \cup ... A_{n-1} \cup \left(\bigcup_{\substack{k\in\mathbb{N} \\ k\geq n}} A_n \right) = \bigcup_{k\in\mathbb{N}} A_k \in \mathcal{T}.$$
- Pour toute famille $(A_i)_{i\in I}$ au plus d'éléments de $\mathcal{T}$, $(\overline{A_i})_{i\in I}$ est une famille au plus dénombrable d'éléments de $\mathcal{T}$ donc, par stabilité de $\mathcal{T}$ par réunion au plus dénombrable et par passage au complémentaire :
$$\bigcap_{i\in I} A_i = \overline{\bigcup_{i\in I} \overline{A_i}} \in \mathcal{T}.$$
- Soit $A, B \in \mathcal{T}$. Par stabilité de $\mathcal{T}$ par intersection finie et par passage au complémentaire, on a :
$$A \setminus B = A \cap \overline{B} \in \mathcal{T}.$$

□

Remarque 3. *Vocabulaire*

Soit $(\Omega, \mathcal{T})$ un espace probabilisable et $A, B \in \mathcal{T}$ des événements. On utilise la terminologie suivante concernant les événements :

- Ω est l'événement **certain** et $\emptyset$ l'événement **impossible** ;
- $\overline{A}$ est l'événement **contraire** de A ;
- si $A \cap B = \emptyset$, on dit que A et B sont des événements **incompatibles**.

Définition 3. *Système complet d'événements*

Soit $(\Omega, \mathcal{T})$ un espace probabilisable et $(A_i)_{i\in I}$ une famille au plus dénombrable d'événements. On dit que la famille $(A_i)_{i\in I}$ est un **système complet d'événements** si :

$$\forall\, i, j \in I \text{ avec } i \neq j,\ A_i \cap A_j = \emptyset \text{ et } \bigcup_{i\in I} A_i = \Omega.$$

2. Espaces probabilisés

a. Définition

Définition 4. *Probabilité*

Soit $(\Omega, \mathcal{T})$ un espace probabilisable et $P : \mathcal{T} \to [0, 1]$ une application. On dit que P est une **mesure de probabilité** ou plus simplement, une **probabilité** sur $(\Omega, \mathcal{T})$ si :

i) $P(\Omega) = 1$;

ii) pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'événements deux à deux disjoints, la série $\sum P(A_n)$ converge et :

$$P\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} P(A_n). \qquad \sigma\text{-additivité}$$

Remarque 4. *Vocabulaire*

Soit $(\Omega, \mathcal{T})$ un espace probabilisable et P une probabilité sur $(\Omega, \mathcal{T})$. Pour $A \in \mathcal{T}$, on dit que $P(A)$ est **la probabilité de l'événement** A.

Définition 5. *Espace probabilisé*

Soit $(\Omega, \mathcal{T})$ un espace probabilisable et P une probabilité sur $(\Omega, \mathcal{T})$. Le triplet $(\Omega, \mathcal{T}, P)$ est appelé **espace probabilisé**.

Définition 6. *Événement négligeable/presque sûr*

Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé et $A \in \mathcal{T}$. On dit que :

— l'événement A est **négligeable** si $P(A) = 0$.

— l'événement A est **presque sûr** si $P(A) = 1$.

Définition 7. *Système quasi-complet d'événements*

Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé et $(A_i)_{i\in I}$ une famille au plus dénombrable d'événements. On dit que la famille $(A_i)_{i\in I}$ est un **système quasi-complet d'événements** si :

$$\forall\, i, j \in I \text{ avec } i \neq j,\ A_i \cap A_j = \emptyset \text{ et } P\left(\bigcup_{i\in I} A_i\right) = 1.$$

Remarque 5.

Un système quasi-complet d'événements porte bien son nom : il s'agit d'un système complet d'événement à "un événement négligeable près". En effet, en complétant un système quasi-complet d'événements $(A_i)_{i\in I}$ par l'événement négligeable $\overline{\bigcup_{i\in I} A_i}$, on obtient un système complet d'événements comme on le verra dans l'exercice 2.

b. Propriétés

Proposition 2.

Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé et $A, B \in \mathcal{T}$. On a les propriétés suivantes :

i) $P(\emptyset) = 0$;

ii) si A et B sont incompatibles, $P(A \cup B) = P(A) + P(B)$;

iii) $P(\overline{A}) = 1 - P(A)$;

iv) si $A \subset B$, $P(B \setminus A) = P(B) - P(A)$;

v) si $A \subset B$, $P(A) \leq P(B)$;

vi) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Démonstration.

i) Notons $c = P(\emptyset)$. Pour tout $n \in \mathbb{N}$, on pose $A_n = \emptyset \in \mathcal{T}$. Alors, pour tout $n, m \in \mathbb{N}$ avec $n \neq m$, $A_n \cap A_m = \emptyset \cap \emptyset = \emptyset$ donc $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements deux à deux disjoints. Par suite, la série $\sum P(A_n) = \sum c$ converge. Or, une suite constante est terme général d'une série convergente si, et seulement s'il s'agit de la suite nulle, d'où $P(\emptyset) = c = 0$.

ii) On suppose A et B incompatibles. Alors, pour $A_0 = A$, $A_1 = B$ et $A_n = \emptyset$ pour tout entier $n \geq 2$, $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements deux à deux disjoints et donc, $\sum P(A_n)$

converge et on a, en utilisant i) :

$$\begin{aligned}
P(A \cup B) &= P\left(A_0 \cup A_1 \cup \emptyset\right) \\
&= P\left(A_0 \cup A_1 \cup \bigcup_{\substack{n\in\mathbb{N}\\ n\geq 2}} A_n\right) \\
&= P\left(\bigcup_{n\in\mathbb{N}} A_n\right) \\
&= \sum_{n=0}^{+\infty} P(A_n) \\
&= P(A_1) + P(A_2) + \sum_{n=2}^{+\infty} \underbrace{P(A_n)}_{P(\emptyset)=0} \\
P(A \cup B) &= P(A) + P(B).
\end{aligned}$$

iii) Comme $A \in \mathcal{T}$, $\overline{A} \in \mathcal{T}$ et comme A et $\overline{A}$ sont incompatibles, d'après ii), on a :

$$1 = P(\Omega) = P(A \cup \overline{A}) = P(A) + P(\overline{A}).$$

D'où $P(\overline{A}) = 1 - P(A)$.

iv) On suppose $A \subset B$. On a alors, par les diverses stabilités d'une tribu : $\overline{B} \in \mathcal{T}$, $B \setminus A = B \cap \overline{A} \in \mathcal{T}$ et $A \cup \overline{B} = \overline{B \setminus A} \in \mathcal{T}$; donc, $\overline{B}$ et A étant incompatibles, on a, en utilisant ii) et iii) :

$$\begin{aligned}
P(B \setminus A) &= 1 - P\left(A \cup \overline{B}\right) \\
&= 1 - (P(A) + P(\overline{B})) \\
&= 1 - (P(A) + 1 - P(B)) \\
P(B \setminus A) &= P(B) - P(A).
\end{aligned}$$

v) On suppose $A \subset B$. Comme P est à valeurs dans $[0,1]$ et $B \setminus A \in \mathcal{T}$, on a $P(B \setminus A) \geq 0$. Or, d'après iv), on a $P(B \setminus A) = P(B) - P(A)$, d'où $P(B) - P(A) \geq 0$ et donc $P(A) \leq P(B)$.

vi) Par les diverses stabiltés d'une tribu, $A \cup B$, $A \cap B$, $A \setminus (A \cap B)$ et $B \setminus (A \cap B)$ appartiennent à $\mathcal{T}$. De plus, $(A \setminus (A \cap B), B \setminus (A \cap B), A \cap B)$ forme une partition de $A \cup B$ donc, $A \cap B$ étant inclus dans A et B, d'après ii) et iv) :

$$\begin{aligned}
P(A \cup B) &= P\left((A \setminus (A \cap B)) \cup (B \setminus (A \cap B)) \cup (A \cap B)\right) \\
&= P(A \setminus (A \cap B)) + P(B \setminus (A \cap B)) + P(A \cap B) \\
&= P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) \\
P(A \cup B) &= P(A) + P(B) - P(A \cap B).
\end{aligned}$$

□

Exercice 2.

Vérifier l'affirmation de la remarque 5.

Correction.

Soit $(A_i)_{i\in I}$ un système quasi-complet d'événements. On pose $A_\bullet = \overline{\bigcup_{i\in I} A_i}$ et $J = I \cup \{\bullet\}$. Comme la tribu $\mathcal{T}$ est stable par réunion au plus dénombrable et par passage au complémentaire, que I est au plus dénombrable et que, pour tout $i \in I$, $A_i \in \mathcal{T}$, on a $A_\bullet \in \mathcal{T}$ et :

$$P(A_\bullet) = P\left(\overline{\bigcup_{i\in I} A_i}\right) = 1 - P\left(\bigcup_{i\in I} A_i\right) = 1 - 1 = 0.$$

De plus, on a :

— $\displaystyle\bigcup_{j\in J} A_j = \left(\bigcup_{i\in I} A_i\right) \cup \overline{\bigcup_{i\in I} A_i} = \Omega.$

— Pour tous $j, k \in J$ avec $j \neq k$, si $j, k \in I$, $A_i \cap A_j = \emptyset$ car $(A_i)_{i\in I}$ un système quasi-complet d'événements et si $j = \bullet$, alors $k \in I$ et donc :

$$A_j \cap A_k = \overline{\bigcup_{i\in I} A_i} \cap A_k = \bigcap_{i\in I} \overline{A_i} \cap A_k = \emptyset$$

car $\overline{A_k} \cap A_k = \emptyset$.

Par suite, $(A_j)_{j\in J}$ est un système complet d'événements.

Corollaire 1.

Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé, $n \in \mathbb{N}^*$ et $A_1, ..., A_n$ des événements deux à deux incompatibles. On a :

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k).$$

Démonstration.

On procède par récurrence en utilisant le point ii) de la proposition précédente. □

Proposition 3.

Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé et $(A_i)_{i\in I}$ une famille au plus dénombrable d'événements deux à deux dijoints. Alors, $(P(A_i))_{i\in I}$ est une famille sommable et on a :

$$P\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} P(A_i).$$

En particulier, si $(A_i)_{i\in I}$ est un système complet ou quasi-complet d'événements, on a :

$$\sum_{i\in I} P(A_i) = 1.$$

Démonstration.

— Si I est fini, $(P(A_i))_{i\in I}$ est une famille finie donc sommable et on a, en numérotant $i_1, ..., i_n$ les éléments de I :

$$P\left(\bigcup_{i\in I} A_i\right) = P(A_{i_1} \cup ... \cup A_{i_n}) = P(A_{i_1}) + ... + P(A_{i_n}) = \sum_{i\in I} P(A_i).$$

car les A_{i_j} sont deux à deux disjoints.

— Si I est infini et donc dénombrable, il est en bijection avec $\mathbb{N}$ via une certaine bijection $n \mapsto i_n$. Ainsi $(A_{i_n})_{n\in\mathbb{N}}$ est une suite d'événements deux à deux disjoints, d'où $\sum P(A_{i_n})$ converge (absolument) et donc $(P(A_{i_n}))_{n\in\mathbb{N}}$ est une famille sommable. Ainsi, par changement d'indice, $(P(A_i))_{i\in I}$ est une famille sommable et on a :

$$P\left(\bigcup_{i\in I} A_i\right) = P\left(\bigcup_{n\in\mathbb{N}} A_{i_n}\right) = \sum_{n=0}^{+\infty} P(A_{i_n}) = \sum_{i\in I} P(A_i).$$

De plus, si $(A_i)_{i\in I}$ est un système complet ou quasi-complet d'événements, alors $P(\bigcup_{i\in I} A_i) = 1$, d'où :

$$\sum_{i\in I} P(A_i) = P\left(\bigcup_{i\in I} A_i\right) = 1.$$

□

3. Propriétés élémentaires des probabilités

Théorème 1. *Continuité croissante/décroissante*

Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé et $(A_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $\mathcal{T}$.

- Si $(A_n)_{n\in\mathbb{N}}$ est croissante i.e. pour tout $n \in \mathbb{N}$, $A_n \subset A_{n+1}$, alors la suite $(P(A_n))_{n\in\mathbb{N}}$ converge et :
$$P(\bigcup_{n\in\mathbb{N}} A_n) = \lim_{n\to+\infty} P(A_n).$$
- Si $(A_n)_{n\in\mathbb{N}}$ est décroissante i.e. pour tout $n \in \mathbb{N}$, $A_{n+1} \subset A_n$, alors la suite $(P(A_n))_{n\in\mathbb{N}}$ converge et :
$$P(\bigcap_{n\in\mathbb{N}} A_n) = \lim_{n\to+\infty} P(A_n).$$

Démonstration.

- Soit $(A_n)_{n\in\mathbb{N}}$ une suite croissante d'événements. Alors, pour tout $n \in \mathbb{N}$, on a $P(A_n) \leq P(A_{n+1})$ et $P(A_n) \leq 1$. Ainsi $(P(A_n))_{n\in\mathbb{N}}$ est une suite croissante et majorée de réels donc elle converge.

  On note $A_{-1} = \emptyset \in \mathcal{T}$.
  La série télescopique $\sum(P(A_n)-P(A_{n-1}))$ étant de même nature que la suite $(P(A_n))_{n\in\mathbb{N}}$, cette série converge, et on a :

$$\sum_{n=0}^{+\infty}(P(A_n) - P(A_{n-1})) = \lim_{n\to+\infty} P(A_n) - P(A_{-1}) = \lim_{n\to+\infty} P(A_n).$$

  On pose, pour $n \in \mathbb{N}$, $B_n = A_n \setminus A_{n-1} \in \mathcal{T}$. Pour tout $n \in \mathbb{N}$, on a $P(B_n) = P(A_n) - P(A_{n-1})$ car, pour $n = 0$, $A_{-1} = \emptyset \subset A_0$ et pour $n \geq 1$, la suite $(A_n)_{n\in\mathbb{N}}$ est croissante.
  De plus, pour $n, m \in \mathbb{N}$ avec $n > m$, on a, toujours par croissance, $B_m \subset A_m \subset A_{n-1}$ car $n - 1 \geq m$, donc :

$$B_m \cap B_n = B_m \cap (A_n \cap \overline{A_{n-1}}) = A_n \cap \underbrace{(B_m \cap \overline{A_{n-1}})}_{=\emptyset} = \emptyset.$$

  Ainsi, les B_n sont des événements deux à deux disjoints. Par suite, $\sum P(B_n)$ converge et on a :

$$P\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \sum_{n=0}^{+\infty} \underbrace{P(B_n)}_{=P(A_n)-P(A_{n-1})} = \lim_{n\to+\infty} P(A_n).$$

  Or, on a, par un raisonnement par double inclusion :

$$\bigcup_{n\in\mathbb{N}} B_n = \bigcup_{n\in\mathbb{N}} A_n,$$

  donc :

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

- Soit $(A_n)_{n\in\mathbb{N}}$ une suite croissante d'événements. Alors, pour tout $n \in \mathbb{N}$, $\overline{A_n} \in \mathcal{T}$ et $\overline{A_{n+1}} \subset \overline{A_n}$ car $A_n \subset A_{n+1}$. Par suite, $(\overline{A_n})_{n\in\mathbb{N}}$ est une suite croissante d'événements et donc, d'après la point précédent, $(P(\overline{A_n}))_{n\in\mathbb{N}}$ converge et :

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

  Or, pour tout $n \in \mathbb{N}$, $P(A_n) = 1 - P(\overline{A_n})$ donc, par combinaison linéaire de suites convergentes, $(P(A_n))_{n\in\mathbb{N}}$ converge et $\lim_{n\to+\infty} P(A_n) = 1 - \lim_{n\to+\infty} P(\overline{A_n})$.

  De plus, on a $\bigcap_{n\in\mathbb{N}} \overline{A_n} = \overline{\bigcup_{n\in\mathbb{N}} A_n}$, donc :

$$P\left(\bigcap_{n\in\mathbb{N}} A_n\right) = 1 - P\left(\bigcup_{n\in\mathbb{N}} \overline{A_n}\right) = 1 - \lim_{n\to+\infty} P(\overline{A_n}) = \lim_{n\to+\infty} P(A_n).$$

□

Proposition 4.

Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé et $(A_n)_{n\in\mathbb{N}}$ une suite d'événements.

- pour $n \in \mathbb{N}$, on a :
$$P\left(\bigcup_{k=0}^{n} A_k\right) \leq \sum_{k=0}^{n} P(A_k).$$
- On a :
$$P\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n=0}^{+\infty} P(A_n).$$
où $\sum_{n=0}^{+\infty} P(A_n) \in \mathbb{R}_+ \cup \{+\infty\}$.

Démonstration.

Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements. Alors, par stabilité d'une tribu par réunion au plus dénombrable, pour $n \in \mathbb{N}$, $B_n = \bigcup_{k=0}^{n} A_k$ est un événement et $\bigcup_{n\in\mathbb{N}} A_n$ est un événement.

- On démontre ce résultat par récurrence sur $\mathbb{N}$.
  — Initialisation. On a
  $$P(B_0) = P(A_0) \leq P(A_0) = \sum_{k=0}^{0} P(A_k).$$
  donc la propriété est vraie au rang $n = 0$.
  — Hérédité. Soit $n \in \mathbb{N}$. On suppose $P(B_n) \leq \sum_{k=0}^{n} P(A_k)$.
  On a $B_{n+1} = B_n \cup A_{n+1}$, donc :
  $$P(B_{n+1}) = P(B_n) + P(A_{n+1}) - \underbrace{P(B_n \cap A_{n+1})}_{\geq 0} \leq P(B_n) + P(A_{n+1}).$$
  Ainsi, par hypothèse de récurrence,
  $$P(B_{n+1}) \geq \underbrace{P(B_n)}_{\leq \sum_{k=0}^{n} P(A_k)} + P(A_{n+1}) \leq \sum_{k=0}^{n+1} P(A_k).$$
  Ce qui achève le raisonnement par récurrence.
  Ainsi, pour tout $n \in \mathbb{N}$,
  $$P\left(\bigcup_{k=0}^{n} A_k\right) = P(B_n) \leq \sum_{k=0}^{n} P(A_k).$$
- Si $\sum P(A_n)$ diverge, l'inégalité est vérifiée. On suppose que $\sum P(A_n)$ converge. La suite $(B_n)_{n\in\mathbb{N}}$ est une suite croissante d'événements, donc par continuité croissante, $(P(B_n))_{n\in\mathbb{N}}$ converge et on a :
  $$P\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \lim_{n\to+\infty} P(B_n).$$
  Ainsi, d'après le point précédent et le théorème des gendarmes, on a :
  $$P\left(\bigcup_{n\in\mathbb{N}} B_n\right) \leq \lim_{n\to+\infty} \sum_{k=0}^{n} P(A_k) = \sum_{n=0}^{+\infty} P(A_n).$$

Puis on conclut en remarquant que $\bigcup_{n\in\mathbb{N}} B_n = \bigcup_{n\in\mathbb{N}} A_n$.

□

Corollaire 2.

Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé et $(A_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $\mathcal{T}$. Si, pour tout $n \in \mathbb{N}$, A_n est négligeable, alors $\bigcup_{n\in\mathbb{N}} A_n$ est négligeable.

Démonstration.

On applique le deuxième point de la proposition précédente à la suite $(A_n)_{n\in\mathbb{N}}$; on a :

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

D'où $P\left(\bigcup_{n\in\mathbb{N}} A_n\right) = 0$ et donc $\bigcup_{n\in\mathbb{N}} A_n$ est négligeable. □

4. Loi/Distribution de probabilité discrète

a. Caractérisation d'une mesure de probabilité sur un univers au plus dénombrable

Théorème 2.

Soit Ω un ensemble non vide au plus dénombrable.

i) Soit P une probabilité sur $(\Omega, \mathcal{P}(\Omega))$. On pose, pour $\omega \in \Omega$, $p_\omega = P(\{\omega\})$. Alors, $(p_\omega)_{\omega\in\Omega}$ est une famille sommable de réels positifs et :

$$\sum_{\omega\in\Omega} p_\omega = 1.$$

ii) Si $(p_\omega)_{\omega\in\Omega}$ est une famille sommable de réels positifs telle que $\sum_{\omega\in\Omega} p_\omega = 1$, alors il existe une unique probabilité P sur $(\Omega, \mathcal{P}(\Omega))$ telle que, pour tout $\omega \in \Omega$, $P(\{\omega\}) = p_\omega$. Elle est définie, pour $A \in \mathcal{P}(\Omega)$, par :

$$P(A) = \sum_{\omega\in A} p_\omega.$$

Démonstration.

Comme, pour tout $\omega, \omega' \in \Omega$, $\{\omega\} \in \mathcal{P}(\Omega)$, $\{\omega\} \cap \{\omega'\} = \emptyset$ et $\bigcup_{\omega\in\Omega}\{\omega\} = \Omega$, la famille au plus dénombrable $(\{\omega\})_{\omega\in\Omega}$ est un sytème complet d'événements.

i) La famille au plus dénombrable $(\{\omega\})_{\omega\in\Omega}$ étant un sytème complet d'événements, d'après la proposition 3, la famille $(P(\{\omega\}))_{\omega\in\Omega}$ est sommable et on a :

$$\sum_{\omega\in\Omega} P(\{\omega\}) = \sum_{\omega\in\Omega} p_\omega = 1.$$

ii) Soit $A \in \mathcal{P}(\Omega)$ un événement. Alors la famille $(p_\omega)_{\omega\in A}$ est sommable comme sous-famille de la famille sommable $(p_\omega)_{\omega\in\Omega}$.

On définit alors l'application $P : \mathcal{P}(\Omega) \to \mathbb{R}$, pour $A \in \mathcal{P}(\Omega)$, par :

$$P(A) = \sum_{\omega\in A} p_\omega.$$

Montrons que P est une probabilité sur $(\Omega, \mathcal{P}(\Omega))$.

— On a, pour tout $A \in \mathcal{P}(\Omega)$, $P(A) = \sum_{\omega\in A} p_\omega \geq 0$ car $p_\omega \geq 0$ pour tout $\omega \in \Omega$ et $P(A) = \sum_{\omega\in A} p_\omega \leq \sum_{\omega\in\Omega} p_\omega = 1$ car $(p_\omega)_{\omega\in A}$ est une sous-famille de la famille sommable $(p_\omega)_{\omega\in\Omega}$. Ainsi, P est à valeurs dans $[0, 1]$.

— on a $P(\Omega) = \sum_{\omega\in\Omega} p_\omega = 1$.

— Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements deux à deux disjoints. On pose $A = \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{P}(\Omega)$. Alors $(p_\omega)_{\omega\in A}$ est une famille sommable, donc, d'après le théorème de sommation par paquets appliqué à la partition $(A_n)_{n\in\mathbb{N}}$ de A, la famille $(P(A_n))_{n\in\mathbb{N}} = (\sum_{\omega\in A_n} p_\omega)_{n\in\mathbb{N}}$ est sommable et donc $\sum P(A_n)$ converge, et de plus, on a :

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

Il en résulte que P est une probabilité sur $(\Omega, \mathcal{P}(\Omega))$.
Finalement, on a, pour tout $\omega \in \Omega$, $\{\omega\} \in \mathcal{P}(\Omega)$ et $P(\{\omega\}) = \sum_{\omega'\in\{\omega\}} p_{\omega'} = p_\omega$.

Par construction, P est unique.

□

b. Loi/Distribution de probabilité discrète

Définition 8.

Soit Ω un ensemble non vide au plus dénombrable.

— On appelle **distribution de probabilité discrète sur** Ω ou plus simplement **distribution discrète sur** Ω toute famille indexée par Ω de réels positifs, sommable et de somme 1.

— On appelle **loi de probabilité discrète sur** Ω ou plus simplement **loi discrète sur** Ω toute mesure de probabilité sur $(\Omega, \mathcal{P}(\Omega))$.

Remarque 6.

— En vertu du théorème 2, une loi de probabilité discrète sur Ω est entièrement déterminée par la donnée d'une distribution discrète sur Ω.

— On utilisera souvent le symbole $\mathcal{L}$ pour désigner une loi de probabilité discrète quelconque.

Partie B

Probabilités conditionnelles

Dans cette partie, $(\Omega, \mathcal{T}, P)$ désigne un espace probabilisé.

1. Conditionnement

Définition 9. *Probabilité conditionnelle*

Soit $B \in \mathcal{T}$ tel que $P(B) > 0$. Pour $A \in \mathcal{T}$, on appelle **probabilité (conditionnelle) de A sachant B** et on note $P_B(A)$ ou encore $P(A|B)$ la quantité :

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

L'application $P_B : \mathcal{T} \to [0,1]$ définie par :

$$P_B : A \mapsto P_B(A)$$

est appelée **probabilité conditionnelle sachant B**.

Théorème 3.

Soit $B \in \mathcal{T}$ tel que $P(B) > 0$. La probabilité conditionnelle $P_B : \mathcal{T} \to \mathbb{R}$ sachant B est une probabilité sur l'espace probabilisable $(\Omega, \mathcal{T})$.

Démonstration.

Montrons que P_B une probabilité sur $(\Omega, \mathcal{T})$.

- Pour tout $A \in \mathcal{T}$, $P_B(A) = \frac{P(A \cap B)}{P(B)} \geq 0$ car P est à valeurs positives et comme $A \cap B \subset B$, on a $P(A \cap B) \leq P(B)$ d'où $P_B(A) = \frac{P(A \cap B)}{P(B)} \leq 1$. Par suite, P_B est à valeurs dans $[0,1]$.
- On a $P_B(\Omega) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$.
- Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements deux à deux disjoints. On a :

$$\left(\bigcup_{n \in \mathbb{N}} A_n \right) \cap B = \bigcup_{n \in \mathbb{N}} (A_n \cap B)$$

De plus, pour tout $n, m \in \mathbb{N}$ avec $n \neq m$, $(A_n \cap B) \cap (A_m \cap B) = \underbrace{(A_n \cap A_m)}_{=\emptyset} \cap B = \emptyset$;

donc $(A_n \cap B)_{n \in \mathbb{N}}$ est une suite d'événements deux à deux disjoints.

Par suite, $\sum P(A_n \cap B)$ converge, donc $\sum P_B(A_n) = \sum \frac{P(A_n \cap B)}{P(B)}$ également et on a :

$$P_B\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \frac{P\left(\bigcup_{n\in\mathbb{N}}(A_n \cap B)\right)}{P(B)} = \sum_{n=0}^{+\infty} \frac{P(A_n \cap B)}{P(B)} = \sum_{n=0}^{+\infty} P_B(A_n).$$

Il en résulte que P_B est une une probabilité sur $(\Omega, \mathcal{T})$. □

2. Formules conditionnelles

a. Formule des probabilités composées

Théorème 4. *Formule des probabilités composées*

i) Soit $A, B \in \mathcal{T}$ tel que $P(B) > 0$. On a :

$$P(A \cap B) = P(A|B)P(B).$$

ii) Soit $n \in \mathbb{N}$ avec $n \geq 2$ et $A_1, \ldots, A_n \in \mathcal{T}$ tels que $P\left(\bigcap_{i=1}^{n-1} A_i\right) > 0$. On a :

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \times P(A_2|A_1) \times P(A_3|A_1 \cap A_2) \times \ldots \times P\left(A_n \middle| \bigcap_{i=1}^{n-1} A_i\right).$$

Démonstration.

i) Soit $A, B \in \mathcal{T}$ tel que $P(B) > 0$. On a $P(A|B) = \frac{P(A\cap B)}{P(B)}$ donc :

$$P(A \cap B) = P(A|B)P(B).$$

ii) Montrons, par récurrence sur $\mathbb{N} \setminus \{0, 1\}$, que, pour tout entier $n \geq 2$, la propriété $\mathcal{P}_n$ ="pour tout $A_1, \ldots, A_n \in \mathcal{T}$ tels que $P\left(\bigcap_{i=1}^{n-1} A_i\right) > 0$, on a :

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \prod_{k=1}^{n-1} P\left(A_{k+1} \middle| \bigcap_{i=1}^{k-1} A_i\right).$$

— Initialisation. On a, d'après i) :

$$P\left(\bigcap_{i=1}^{2} A_i\right) = P(A_1 \cap A_2) = P(A_2|A_1)P(A_1) = P(A_1) \times \prod_{k=1}^{2-1} P\left(A_{k+1} \middle| \bigcap_{i=1}^{k-1} A_i\right).$$

Donc $\mathcal{P}_2$ est vraie.

— Hérédité. Soit un entier $n \geq 2$. On suppose $\mathcal{P}_n$ vraie. Soit $A_1, \ldots, A_{n+1} \in \mathcal{T}$ tels que

$P\left(\bigcap_{i=1}^{n} A_i\right) > 0$. Alors, comme $\bigcap_{i=1}^{n} A_i \subset \bigcap_{i=1}^{n-1} A_i$, on a :

$$0 < P\left(\bigcap_{i=1}^{n} A_i\right) \leq P\left(\bigcap_{i=1}^{n-1} A_i\right).$$

Ainsi, par hypothèse de récurrence,

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \prod_{k=1}^{n-1} P\left(A_{k+1} \middle| \bigcap_{i=1}^{k-1} A_i\right).$$

De plus, on a, d'après i) :

$$P\left(\bigcap_{i=1}^{n+1} A_i\right) = P\left(A_{n+1} \cap \left(\bigcap_{i=1}^{n} A_i\right)\right) = P\left(A_{n+1} \middle| \bigcap_{i=1}^{n} A_i\right) P(\left(\bigcap_{i=1}^{n} A_i\right).$$

Ainsi, il en résulte que :

$$\begin{aligned} P\left(\bigcap_{i=1}^{n+1} A_i\right) &= P(A_1) \prod_{k=1}^{n-1} P\left(A_{k+1} \middle| \bigcap_{i=1}^{k-1} A_i\right) \times P\left(A_{n+1} \middle| \bigcap_{i=1}^{n} A_i\right) \\ P\left(\bigcap_{i=1}^{n+1} A_i\right) &= P(A_1) \prod_{k=1}^{n} P\left(A_{k+1} \middle| \bigcap_{i=1}^{k-1} A_i\right) \end{aligned}$$

Ce qui achève le raisonnement par récurrence.

Il en résulte que la propriété $\mathcal{P}_n$ est vraie pour tout $n \in \mathbb{N} \setminus \{0, 1\}$.

□

Remarque 7.

Soit $A, B \in \mathcal{T}$. Dans le cas où $P(B) = 0$, on a $P(A \cap B) = 0$ car $A \cap B \subset B$. Ainsi, en vertu du théorème précédent, on adoptera la convention suivante dans toute la suite du chapitre :

Si $P(B) = 0$, on convient que $P(A|B)P(B) = 0$.

b. Formule des probabilités totales

Exercice 3.

Soit $B \in \mathcal{T}$. Montrer que, pour tout $A \in \mathcal{T}$, $P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B})$.

Correction.

Soit $A \in \mathcal{T}$. On a $B \cup \overline{B} = \Omega$ donc :

$$A = A \cap \Omega = A \cap (B \cup \overline{B}) = (A \cap B) \cup (A \cap \overline{B})$$

Et, de plus, on a $B \cap \overline{B} = \emptyset$, donc :

$$(A \cap B) \cap (A \cap \overline{B}) = A \cap B \cap A \cap \overline{B} = A \cap \underbrace{(B \cap \overline{B})}_{\emptyset} = \emptyset.$$

Par suite, on obtient :

$$P(A) = P((A \cap B) \cup (A \cap \overline{B})) = P(A \cap B) + P(A \cap \overline{B}) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B}).$$

On peut généraliser le résultat et la démonstration de l'exercice précédent à un système complet ou quasi-complet d'événements quelconque à la place du système complet $(B, \overline{B})$: il s'agit de la formule des probabilités totales.

Théorème 5. *Formule des probabilités totales*

Soit $(B_i)_{i \in I}$ un système complet ou quasi-complet d'événements. Pour tout événement $A \in \mathcal{T}$, on a :

$$P(A) = \sum_{i \in I} P(A \cap B_i) = \sum_{i \in I} P(A|B_i)P(B_i).$$

On rappelle la convention $P(A|B_i)P(B_i) = 0$ *si* $P(B_i) = 0$.

Démonstration.

Soit $(B_i)_{i \in I}$ est un système complet d'événements. On a :

$$\bigcup_{i \in I} B_i = \Omega \quad \text{et} \quad B_i \cap B_j = \emptyset \text{ si } i \neq j.$$

Par suite, on a :

$$A = \bigcup_{i \in I} (A \cap B_i) \text{ avec } (A \cap B_i) \cap (A \cap B_j) = \emptyset \text{ si } i \neq j.$$

En effet,

$$\bigcup_{i \in I} (A \cap B_i) = \bigcup_{i \in I} A \cap \bigcup_{i \in I} B_i = A \cap \Omega = A;$$

et

$$(A \cap B_i) \cap (A \cap B_j) = A \cap (B_i \cap B_j) = A \cap \emptyset = \emptyset.$$

Par suite, par les propriétés de la fonction de probabilité P, on a :

$$P(A) = P\left(\bigcup_{i \in I} (A \cap B_i)\right) = \sum_{i \in I} P(A \cap B_i) = \sum_{i \in I} P(A|B_i)P(B_i).$$

Si $(B_i)_{i \in I}$ est un système quasi-complet d'événements, on le complète en un système complet d'événements avec $B = \Omega \setminus \bigcup_{i \in I} B_i$ et on lui applique la formule des probabilités totales prouvée précédemment : on remarque alors que, comme B est négligéable, alors $P(A \cap B) = 0$ et donc $P(A) = \sum_{i \in I} P(A \cap B_i)$. □

c. Formule de Bayes

Théorème 6. *Formule de Bayes*

Soit $A \in \mathcal{T}$ tel que $P(A) > 0$.

i) Soit $B \in \mathcal{T}$. On a :

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

ii) Soit $(B_i)_{i\in I}$ un système complet d'événements. On a, pour tout $j \in I$:

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i\in I} P(A|B_i)P(B_i)}.$$

Démonstration.

i) Soit $B \in \mathcal{T}$. On a :

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

ii) D'après i) puis la formule des propbabilités totales, on a :

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i\in I} P(A|B_i)P(B_i)}.$$

□

d. Exercice d'application des formules

Exercice 4.

Dans une population, on donne la probabilité définie, pour $n \in \mathbb{N}$ par p_n qu'une famille ait n enfants où, α étant un réel strictement positif :

$$p_n = \alpha \frac{2^n}{n!}.$$

On suppose qu'il est équiprobable qu'un enfant soit une fille ou un garçon.

1. Déterminer le nombre α.
2. Déterminer la probabilité qu'une famille ait au moins un garçon.
3. Déterminer la probabilité qu'une famille ait exactement 2 garçons et 3 filles.
4. On cherche à déterminer la probabilité qu'une famille ait exactement deux enfants sachant qu'elle a au moins deux filles.
   a) Déterminer la probabilité qu'une famille ait exactement deux filles.
   b) Déterminer la probabilité qu'une famille ait au moins deux filles.
   c) Conclure.

Correction.

1. Soit Ω l'ensemble des compositions (d'enfants) des familles de la population. La suite d'événements $(A_n)_{n\in\mathbb{N}}$ définie par :

   $$A_n=\text{"La famille possède exactement } n \text{ enfants"}.$$

   étant un système complet d'événements, on a :

   $$1 = P(\Omega) = P\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} A_n = \sum_{n=0}^{+\infty} p_n.$$

   Par suite,

   $$1 = \sum_{n=0}^{+\infty} p_n = \alpha \sum_{n=0}^{+\infty} \frac{2^n}{n!} = \alpha e^2.$$

   D'où $\alpha = e^{-2}$.

2. On note B l'événement :

   $$B=\text{"La famille possède uniquement des filles (ou aucun enfant)"}.$$

   Alors $\overline{B}$="La famille possède au moins un garçon" est l'événement dont la probabilité est recherchée.

   On a, d'après la formule des probabilités totales :

   $$P(B) = \sum_{n=0}^{+\infty} P(B|A_n)P(A_n) = \sum_{n=0}^{+\infty} \frac{1}{2^n} p_n = e^{-2}.e = \frac{1}{e}.$$

3. On note C l'événement recherché, D l'événement "la famille possède 2 garçons et 3 filles (au moins)" et on reprend la suite (A_n) de la question précédente. Alors on a $C = D \cap A_2$ donc :

   $$P(C) = P(D \cap A_5) = P(D|A_5)P(A_5) = \left(\binom{5}{2}\frac{1}{2^2}\frac{1}{2^3}\right).e^{-2}\frac{2^5}{5!} = \frac{1}{12e^2}.$$

4. On note F l'événement "une famille a au moins deux filles" et on reprend la suite (A_n) de la question 1.

   a) Avec les notations précédentes, ce qu'on doit calculer est :

   $$P(F \cap A_2) = P(F|A_2)P(A_2) = \frac{1}{4}.e^{-2}\frac{4}{2} = \frac{1}{2e^2}.$$

   b) D'après la formule des probabilités totales, on a :

   $$P(F) = 1 - P(\overline{F}) = 1 - \sum_{n=0}^{+\infty} P(\overline{F}|A_n)P(A_n).$$

   et on a :

   $$P(\overline{F}|A_n) = \binom{n}{0}\frac{1}{2^0}\frac{1}{2^n} + \binom{n}{1}\frac{1}{2^1}\frac{1}{2^{n-1}} = (n+1)\frac{1}{2^n}.$$

   Ainsi, on a :

$$\begin{aligned} P(\overline{F}) &= \sum_{n=0}^{+\infty} P(\overline{F}|A_n)P(A_n) \\ &= \sum_{n=0}^{+\infty} (n+1)\frac{1}{2^n}p_n \\ &= e^{-2}\sum_{n=0}^{+\infty}\frac{n+1}{n!} \\ &= e^{-2}\left(\sum_{n=1}^{+\infty}\frac{n}{n!}+\sum_{n=0}^{+\infty}\frac{1}{n!}\right) \\ &= 2e^{-2}\sum_{n=0}^{+\infty}\frac{1}{n!} \\ &= \frac{2}{e}. \end{aligned}$$

D'où $P(F) = 1 - \frac{2}{e}$.

c) On a :

$$P(A_2|F) = \frac{P(A_2 \cap F)}{P(F)} = \frac{1}{2e^2}.\frac{1}{1-\frac{2}{e}} = \frac{1}{2e(e-2)}.$$

3. Événements indépendants

a. Définition et premières propriétés

Définition 10. *Événements indépendants*

Soit $A, B \in \mathcal{T}$. On dit que A et B sont **indépendants** si

$$P(A \cap B) = P(A).P(B)$$

Proposition 5.

Soit $A \in \mathcal{T}$. Si A est négligeable, pour tout $B \in \mathcal{T}$, A et B sont indépendants.

Démonstration.

Si A est négligeable, pour tout $B \in \mathcal{T}$, comme $A \cap B \subset A$, $0 \leq P(A \cap B) \leq P(A) = 0$ donc :

$$P(A \cap B) = 0 = 0.P(B) = P(A).P(B)$$

d'où l'indépendance de A et B. □

On justifie le choix de la terminologie d'"indépendance" par la proposition suivante :

Proposition 6.

Soit $A, B \in \mathcal{T}$ avec $P(B) > 0$. Les événements A et B sont indépendants si, et seulement si, $P(A|B) = P(A)$.

Démonstration.

Comme $P(B) > 0$, on a :
A et B sont indépendants
si, et seulement si,
$P(A \cap B) = P(A).P(B)$
si, et seulement si,
$P(A|B)P(B) = P(A).P(B)$
si, et seulement si,
$P(A|B) = P(A)$. □

Proposition 7.

Soit $A, B \in \mathcal{T}$. Si A et B sont indépendants, alors :
- $\overline{A}$ et $\overline{B}$ sont indépendants ;
- $\overline{A}$ et B sont indépendants.

Démonstration.

On suppose A et B indépendants.

- On a, en remarquant que $A \cup B = \overline{\overline{A} \cap \overline{B}}$:

$$\begin{aligned} P(\overline{A})P(\overline{B}) &= (1 - P(A))(1 - P(B)) \\ &= 1 - (P(A) + P(B) - P(A)P(B)) \\ &= 1 - P(A \cup B) \\ &= P\left(\overline{\overline{A} \cap \overline{B}}\right) \\ P(\overline{A})P(\overline{B}) &= P\left(\overline{A} \cap \overline{B}\right). \end{aligned}$$

  Donc $\overline{A}$ et $\overline{B}$ sont indépendants.

- Si B est négligeable, alors, d'après la proposition 5, $\overline{A}$ et B sont indépendants.
  Supposons $P(B) > 0$. Comme A et B sont indépendants, d'après la proposition 6, $P(A|B) = P(A)$ d'où, comme P_B est une probabilité sur $(\Omega, \mathcal{T})$:

$$P(\overline{A}|B) = P_B(\overline{A}) = 1 - P_B(A) = 1 - P(A|B) = 1 - P(A) = P(\overline{A}).$$

  Ainsi, d'après la proposition 6, $\overline{A}$ et B sont indépendants.

□

Corollaire 3.

Soit $A \in \mathcal{T}$. Si A est presque sûr, pour tout $B \in \mathcal{T}$, A et B sont indépendants.

Démonstration.

On suppose A est presque sûr. Soit $B \in \mathcal{T}$. L'événement $A' = \overline{A}$ est négligeable donc, d'après la proposition 5, A' et B sont indépendants ; par suite, d'après la proposition 7, $A = \overline{A'}$ et B sont indépendants. □

b. Indépendance d'une famille d'événement

Définition 11. *Indépendance deux à deux/Indépendance*

Soit $(A_i)_{i\in I}$ une famille à valeurs dans $\mathcal{T}$.

- On dit que les événements A_i, pour $i \in I$ sont **deux à deux indépendants** si, pour tous $i, j \in I$ avec $i \neq j$, on a :

$$P(A_i \cap A_j) = P(A_i)P(A_j).$$

- On dit que les événements A_i, pour $i \in I$ sont **indépendants** (ou *mutuellement indépendants*) si, pour toute partie finie J de I, on a :

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

Remarque 8.

Il est clair que l'indépendance d'une famille d'événements implique l'indépendance deux à deux de ces événements mais **la réciproque est fausse !** comme en atteste l'exercice suivant :

Exercice 5.

On lance deux fois un dé équilibré à 6 faces numérotées de 1 à 6 et on considère les événements suivants :

A="La somme des deux lancers est pair" ; B="Le premier lancer est pair" et C="Le second lancer est pair".

Montrer que les événements A, B, C sont deux à deux indépendants mais pas mutuellement indépendants.

Correction.

On formalise l'exercice de la façon suivante :

— L'univers Ω : ici, $\Omega = [\![1, 6]\!]^2$;

— La tribu $\mathcal{T}$: ici, $\mathcal{T} = \mathcal{P}(\Omega)$;

— La mesure de probabilité : ici, $P : A \mapsto P(A) = \dfrac{\#A}{\#\Omega} = \dfrac{\#A}{36}$ car le dé est équilibré.

On illustre les événements A, B, C de la manière suivante :

$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$	$(1,5)$	$(1,6)$
A ~~B~~ ~~C~~	~~A~~ ~~B~~ C	A ~~B~~ ~~C~~	~~A~~ ~~B~~ C	A ~~B~~ ~~C~~	~~A~~ ~~B~~ C
$(2,1)$	$(2,2)$	$(2,3)$	$(2,4)$	$(2,5)$	$(2,6)$
~~A~~ B ~~C~~	A B C	~~A~~ B ~~C~~	A B C	~~A~~ B ~~C~~	A B C
$(3,1)$	$(3,2)$	$(3,3)$	$(3,4)$	$(3,5)$	$(3,6)$
A ~~B~~ ~~C~~	~~A~~ ~~B~~ C	A ~~B~~ ~~C~~	~~A~~ ~~B~~ C	A ~~B~~ ~~C~~	~~A~~ ~~B~~ C
$(4,1)$	$(4,2)$	$(4,3)$	$(4,4)$	$(4,5)$	$(4,6)$
~~A~~ B ~~C~~	A B C	~~A~~ B ~~C~~	A B C	~~A~~ B ~~C~~	A B C
$(5,1)$	$(5,2)$	$(5,3)$	$(5,4)$	$(5,5)$	$(5,6)$
A ~~B~~ ~~C~~	~~A~~ ~~B~~ C	A ~~B~~ ~~C~~	~~A~~ ~~B~~ C	A ~~B~~ ~~C~~	~~A~~ ~~B~~ C
$(6,1)$	$(6,2)$	$(6,3)$	$(6,4)$	$(6,5)$	$(6,6)$
~~A~~ B ~~C~~	A B C	~~A~~ B ~~C~~	A B C	~~A~~ B ~~C~~	A B C

Ainsi, on a :

— $A = \{(1,1),(1,3),(1,5),(2,2),(2,4),(2,6),(3,1),(3,3),(3,5),(4,2),(4,4),(4,6),(5,1),(5,3),(5,5),(6,2),(6,4),(6,6)\}$, donc $P(A) = \frac{18}{36} = \frac{1}{2}$;

— $B = \{(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$, donc $P(B) = \frac{18}{36} = \frac{1}{2}$ et

— $C = \{(1,2),(2,2),(3,2),(4,2),(5,2),(6,2),(1,4),(2,4),(3,4),(4,4),(5,4),(6,4),(1,6),(2,6),(3,6),(4,6),(5,6),(6,6)\}$, donc $P(C) = \frac{18}{36} = \frac{1}{2}$;

puis :

$$A \cap B = B \cap C = C \cap A = A \cap B \cap C = \{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\},$$

d'où

$$P(A \cap B) = P(B \cap C) = P(C \cap A) = P(A \cap B \cap C) = \frac{9}{36} = \frac{1}{4}.$$

Ainsi, on a :

$$P(A \cap B) = \frac{1}{4} = P(A)P(B); \quad P(B \cap C) = \frac{1}{4} = P(B)P(C); \quad P(C \cap A) = \frac{1}{4} = P(C)P(A).$$

donc A, B et C sont deux à deux indépendants ; et

$$P(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C)$$

donc A, B et C ne sont pas indépendants.

Partie C

Variables aléatoires discrètes

1. Variables aléatoires discrètes

a. Définition

Définition 12. *Variable aléatoire discrète*

Soit $(\Omega, \mathcal{T})$ un espace probabilisable, E un ensemble et $X : \Omega \to E$ une fonction. On dit que la fonction X est une **variable aléatoire discrète** sur $(\Omega, \mathcal{T})$ si :

i) l'ensemble $X(\Omega)$ est au plus dénombrable ;

ii) pour tout $x \in X(\Omega)$, $X^{-1}(\{x\})$ est un événement de $\mathcal{T}$, i.e.

$$X^{-1}(\{x\}) = \{\omega \in \Omega \mid X(\omega) = x\} \in \mathcal{T}.$$

De plus, si $E \subset \mathbb{R}$, on dit que X est une variable aléatoire discrète **réelle**.

Proposition 8.

Soit $(\Omega, \mathcal{T})$ un espace probabilisable, E un ensemble et $X : \Omega \to E$ une variable aléatoire discrète. Pour tout $F \subset E$, $X^{-1}(F) \in \mathcal{T}$.

Démonstration.

Soit $F \subset E$. On a $F \cap X(\Omega) = \bigcup_{x \in F \cap X(\Omega)} \{x\}$ donc, par propriétés de l'image réciproque d'une fonction, on a :

$$X^{-1}(F) = X^{-1}(F \cap X(\Omega)) = \bigcup_{x \in F \cap X(\Omega)} X^{-1}(\{x\})$$

Comme X est une variable aléatoire discrète, $X(\Omega)$ est au plus dénombrable donc $F \cap X(\Omega)$ l'est aussi et, de plus, pour tout $x \in F \cap X(\Omega) \subset X(\Omega)$, $X^{-1}(\{x\}) \in \mathcal{T}$.
Or la tribu $\mathcal{T}$ est stable par union au plus dénombrable donc

$$X^{-1}(F) = \bigcup_{x \in F \cap X(\Omega)} X^{-1}(\{x\}) \in \mathcal{T}.$$

□

Pour des questions de lisibilité et de simplicité, on emploie les notations suivantes :

Notation 1.

Soit $(\Omega, \mathcal{T})$ un espace probabilisable, E un ensemble et $X : \Omega \to E$ une variable aléatoire discrète.

— Pour $F \subset E$, l'événement $X^{-1}(F)$ est noté $(X \in F)$.

— Pour $x \in E$, l'événement $X^{-1}(\{x\})$ est noté :

$$(X = x).$$

— Si $E \subset \mathbb{R}$, pour $x \in E$, les événements $X^{-1}(]-\infty, x[)$, $X^{-1}(]-\infty, x])$, $X^{-1}(]x, +\infty[)$ et $X^{-1}([x, +\infty[)$ sont notés respectivement :

$$(X < x), \qquad (X \leq x), \qquad (X > x), \qquad (X \geq x).$$

Proposition 9.

Soit X une variable aléatoire discrète sur un espace probabilisé $(\Omega, \mathcal{T}, P)$ et à valeurs dans $\mathbb{N}$. On a, pour tout $n \in \mathbb{N}$:

$$P(X = n) = P(X > n-1) - P(X > n) \text{ et } P(X > n) = \sum_{k=n+1}^{+\infty} P(X = k).$$

Démonstration.

— On a, pour tout $n \in \mathbb{N}$:

$$(X = n) = (X > n-1) \setminus (X > n) \text{ avec } (X > n) \subset (X > n-1),$$

donc :

$$P(X = n) = P(X > n-1) - P(X > n).$$

— On a, pour tout $n \in \mathbb{N}$:

$$(X > n) = \bigcup_{\substack{k \in \mathbb{N} \\ k \geq n+1}} (X = k)$$

et la famille $(X = k)_{k \in \mathbb{N}}$ étant un système complet d'événements la série $\sum_{k \geq n} P(X = k)$ converge et on a :

$$P(X > n) = P\left(\bigcup_{\substack{k \in \mathbb{N} \\ k \geq n+1}} (X = k) \right) = \sum_{k=n+1}^{+\infty} P(X = k)$$

et donc $P(X > n)$ est le reste d'ordre n de la série convergente $\sum P(X = n)$.

□

b. Loi de probabilité d'une variable aléatoire

Définition 13. *Loi de probabilité d'une variable aléatoire*

Soit X une variable aléatoire discrète sur un espace probabilisé $(\Omega, \mathcal{T}, P)$. On appelle **loi de probabilité de** X l'application $P_X : \mathcal{P}(X(\Omega)) \to [0,1]$ définie par :

$$P_X : A \mapsto P(X \in A).$$

Proposition 10.

Soit X une variable aléatoire discrète sur un espace probabilisé $(\Omega, \mathcal{T}, P)$. Le triplet $(X(\Omega), \mathcal{P}(X(\Omega)), P_X)$ est un espace probabilisé.

Démonstration.

D'après l'exemple 1, $\mathcal{P}(X(\Omega))$ est une tribu sur $X(\Omega)$ donc $(X(\Omega), \mathcal{P}(X(\Omega)))$ est un espace probabilisable. Il reste à montrer que P_X est une mesure de probabilité sur cet espace.
Comme X est une variable aléatoire discrète l'application image réciproque $X^{-1} : F \mapsto X^{-1}(F)$ définie sur $\mathcal{P}(X(\Omega))$ est à valeurs dans $\mathcal{T}$ d'après la proposition 8.

- On a alors $P_X = P \circ X^{-1} : \mathcal{P}(\Omega) \to [0,1]$.
- On a $P_X(X(\Omega)) = P(X^{-1}(X(\Omega))) = P(\Omega) = 1$.
- Soit $(F_n)_{n\in\mathbb{N}} \in \mathcal{P}(X(\Omega))^{\mathbb{N}}$ une suite d'ensembles deux à deux disjoints. Pour $n \in \mathbb{N}$, on note $A_n = X^{-1}(F_n) \in \mathcal{T}$ et on a alors $P(A_n) = P_X(F_n)$. Pour tous $n, m \in \mathbb{N}$ avec $n \neq m$:

  $$A_n \cap A_m = X^{-1}(F_n) \cap X^{-1}(F_m) = X^{-1}(F_n \cap F_m) = X^{-1}(\emptyset) = \emptyset.$$

  Ainsi, $(A_n)_{n\in\mathbb{N}}$ est une suite à valeurs dans $\mathcal{T}$ d'ensembles deux à deux disjoints, donc, P étant une mesure de probabilité, $\sum P_X(F_n) = \sum P(A_n)$ converge et, par σ-additivité de P, on a :

$$\begin{aligned} P_X\left(\bigcup_{n\in\mathbb{N}} F_n\right) &= P\left(X^{-1}\left(\bigcup_{n\in\mathbb{N}} F_n\right)\right) \\ &= P\left(\bigcup_{n\in\mathbb{N}} X^{-1}(F_n)\right) \\ &= P\left(\bigcup_{n\in\mathbb{N}} A_n\right) \\ &= \sum_{n=0}^{+\infty} P(A_n) \\ P_X\left(\bigcup_{n\in\mathbb{N}} F_n\right) &= \sum_{n=0}^{+\infty} P_X(F_n). \end{aligned}$$

Il en résulte que P_X est une mesure de probabilité sur $(X(\Omega), \mathcal{P}(X(\Omega)))$ et donc $(X(\Omega), \mathcal{P}(X(\Omega)), P_X)$ est un espace probabilisé. □

Proposition 11.

Soit X une variable aléatoire discrète sur un espace probabilisé $(\Omega, \mathcal{T}, P)$. La famille $((X = x))_{x \in X(\Omega)}$ est un système complet d'événements et en particulier, la famille $(P(X = x))_{x \in X(\Omega)}$ est sommable de somme :

$$\sum_{x \in X(\Omega)} P(X = x) = 1.$$

Démonstration.

Comme X est une variable aléatoire discrète, pour tout $x \in X(\Omega)$, $(X = x) = X^{-1}(\{x\}) \in \mathcal{T}$ d'après la propostion 8 et pour tous $x, x' \in X(\Omega)$ avec $x \neq x'$, on a :

$$(X = x) \cap (X = x') = X^{-1}(\{x\}) \cap X^{-1}(\{x'\}) = X^{-1}(\{x\} \cap \{x\}) = X^{-1}(\emptyset) = \emptyset.$$

Donc $((X = x))_{x \in X(\Omega)}$ est une famille d'événements deux à deux incompatibles.
De plus, on a :

$$\bigcup_{x \in X(\Omega)} (X = x) = \bigcup_{x \in X(\Omega)} X^{-1}(\{x\}) = X^{-1}\left(\bigcup_{x \in X(\Omega)} \{x\} \right) = X^{-1}(X(\Omega)) = \Omega.$$

Par suite, $((X = x))_{x \in X(\Omega)}$ est un système complet d'événements. □

Remarque 9.

— D'après ce qui précède, si X une variable aléatoire *discrète* sur un espace probabilisé $(\Omega, \mathcal{T}, P)$, la loi de probabilité P_X est entièrement déterminée par les événements $(X = x)$ pour $x \in X(\Omega)$.

Ainsi, pour déterminer la loi de probabilité P_X de la variable aléatoire X, (il faut et) il suffit de déterminer les valeurs de $P(X = x)$ pour chaque $x \in X(\Omega)$.

— Si X est une variable aléatoire discrète à valeurs dans un ensemble au plus dénombrable E, $X(\Omega)$ n'est pas forcément égal à E. Dans la suite, on étendra la loi P_X de X à $\mathcal{P}(E)$ tout entier, en posant, pour $x \in E \setminus X(\Omega)$, $P_X(\{x\}) := 0$ ou encore $P(X = x) = 0$.
Ainsi, P_X définit une loi de probabilité sur $(E, \mathcal{P}(E))$.

Notation 2.

Soit E un ensemble au plus dénombrable, $\mathcal{L}$ une loi discrète sur E définie par une famille $(p_x)_{x \in E}$ et X, Y des variables aléatoires discrètes à valeurs dans E.

— On dit que X **suit la loi discrète** $\mathcal{L}$ et on note $X \sim \mathcal{L}$ si $P_X = \mathcal{L}$ i.e. si pour tout $x \in E$,

$$P(X = x) = p_x.$$

— On dit que X et Y **suivent la même loi** ou encore que X et Y sont **équidistribuées**

et on note $X \sim Y$ si $P_X = P_Y$ i.e. si pour tout $x \in E$,

$$P(X = x) = P(Y = x).$$

c. Fonction d'une variable aléatoire discrète

Proposition-Notation 12.

Soit X une variable aléatoire discrète sur un espace probabilisable $(\Omega, \mathcal{T})$, F un ensemble et $f : X(\Omega) \to F$ une fonction. On note $f(X)$ la variable aléatoire discrète $f \circ X$ sur $(\Omega, \mathcal{T})$.

Démonstration.

On pose $Y = f(X) = f \circ X$.

- L'application $f : X(\Omega) \to f(X(\Omega)) = Y(\Omega)$ est surjective et $X(\Omega)$ est au plus dénombrable car X est une variable aléatoire discrète donc $Y(\Omega)$ est au plus dénombrable.
- Soit $y \in Y(\Omega) = f(X(\Omega))$. Alors, comme $f^{-1}(\{y\}) \subset X(\Omega)$ et X est une variable aléatoire discrète à valeurs dans $X(\Omega)$, d'après la propostion 8, $X^{-1}\left(f^{-1}(\{y\})\right) \in \mathcal{T}$. Ainsi :

$$(Y = y) = Y^{-1}(\{y\}) = (f \circ X)^{-1}(\{y\}) = X^{-1}\left(f^{-1}(\{y\})\right) \in \mathcal{T}.$$

Il en résulte que $f(X) = Y$ est une variable aléatoire discrète sur $(\Omega, \mathcal{T})$. □

Proposition 13.

Soit E, F des ensembles, X, Y des variables aléatoires discrètes à valeurs dans un ensemble E et $f : E \to F$ une fonction. Si $X \sim Y$, alors $f(X) \sim f(Y)$.

Démonstration.

On suppose que X et Y suivent la même loi. On note $X' = f(X)$ et $Y' = f(Y)$ qui sont des variables aléatoires discrètes d'après la proposition précédente. Soit $x' \in F$. On a :

$$(X' = x') = X^{-1}\left(f^{-1}(\{x'\})\right) = \bigcup_{x \in f^{-1}(\{x'\})} X^{-1}(\{x\}) = \bigcup_{x \in f^{-1}(\{x'\})} (X = x)$$

et de manière analogue :

$$(Y' = x') = \bigcup_{x \in f^{-1}(\{x'\})} (Y = x)$$

Or, les événements $(X = x)$ étant deux à deux disjoints d'après la proposition 11, et de même pour les $(Y = x)$, on a obtient, comme $X \sim Y$:

$$P(X' = x') = \sum_{x \in f^{-1}(\{x'\})} P(X = x) = \sum_{x \in f^{-1}(\{x'\})} P(Y = x) = P(Y' = x').$$

Il en résulte que $f(X) = X' \sim Y' = f(Y)$. □

2. Rappels des lois usuelles sur un univers fini

a. Loi uniforme

Définition-Proposition 14. *Loi uniforme*

Soit E un ensemble **fini**. La famille $(p_x)_{x\in E}$, où, pour tout $x \in E$, $p_x = \frac{1}{\#E}$, est une distribution de probabilité discrète sur E. On appelle **loi uniforme sur** E et on note $\mathcal{U}(E)$ la loi de probabilité discrète définie par cette distribution.

Ainsi une variable aléatoire X **suit une loi uniforme sur** E i.e. $X \sim \mathcal{U}(E)$ si :

- $X(\Omega) = E$;
- pour tout $x \in E$,
$$P(X = x) = \frac{1}{\#E}.$$

Si $E = [\![1, n]\!]$ pour un certain $n \in \mathbb{N}^*$, on notera simplement $\mathcal{U}(n)$ la loi uniforme sur $[\![1, n]\!]$.

Démonstration.

L'ensemble E est au plus dénombrable car fini. Pour tout $x \in E$, $p_x = \frac{1}{\#E} \in \mathbb{R}_+$; la famille $(p_x)_{x\in E}$ est sommable car E est fini et on a :

$$\sum_{x\in E} p_x = \sum_{x\in E} \frac{1}{\#E} = \#E.\frac{1}{\#E} = 1.$$

Par suite, la famille $(p_x)_{x\in E}$ est une distribution discrète sur E. □

b. Loi de Bernoulli

Définition-Proposition 15. *Loi de Bernoulli*

Soit $p \in]0, 1[$. La famille $(p_x)_{x\in\{0,1\}}$, où $p_0 = p$ et $p_1 = 1-p$, est une distribution de probabilité discrète sur $\{0, 1\}$. On appelle **loi de Bernoulli de paramètre** p et on note $\mathcal{B}(p)$ la loi de probabilité discrète définie par cette distribution.

Ainsi une variable aléatoire X **suit une loi de Bernoulli de paramètre** p i.e. $X \sim \mathcal{B}(p)$ si :

- $X(\Omega) = \{0, 1\}$;
- $P(X = 1) = p$ et $P(X = 0) = 1 - p$.

Démonstration.

On note $E = \{0, 1\}$ qui est au plus dénombrable car fini. La famille $(p_x)_{x\in E}$ est sommable car E est fini, $p_0, p_1 \in \mathbb{R}_+$ car $p \in [0, 1]$ et on a :

$$\sum_{x\in E} p_x = p_0 + p_1 = p + (1 - p) = 1.$$

Par suite, la famille $(p_x)_{x\in\{0,1\}}$ est une distribution discrète sur $\{0,1\}$. □

c. Loi binomiale

Définition-Proposition 16. *Loi binomiale*

Soit $n \in \mathbb{N}^*$ et $p \in]0,1[$. La famille $(p_k)_{k\in[\![0,n]\!]}$, où, pour tout $k \in [\![0,n]\!]$, $p_k = \binom{n}{k}p^k(1-p)^{n-k}$, est une distribution de probabilité discrète sur $[\![0,n]\!]$. On appelle **loi binomiale de paramètres** n **et** p et on note $\mathcal{B}(n,p)$ la loi de probabilité discrète définie par cette distribution.

Ainsi une variable aléatoire X **suit une loi binomiale de paramètres** n **et** p i.e. $X \sim \mathcal{B}(n,p)$ si :

- $X(\Omega) = [\![0,n]\!]$;
- pour tout $k \in [\![0,n]\!]$,

$$P(X=k) = \binom{n}{k}p^k(1-p)^{n-k}.$$

Démonstration.

On note $E = [\![0,n]\!]$ qui est au plus dénombrable car fini. La famille $(p_k)_{k\in E}$ est sommable car E est fini, à valeurs réelles positives car $p \in [0,1]$ et on a, d'après la formule du binôme de Newton :

$$\sum_{k\in E} p_k = \sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k} = (p+(1-p))^n = 1^n = 1.$$

Par suite, la famille $(p_k)_{k\in[\![0,n]\!]}$ une distribution discrète sur $[\![0,n]\!]$. □

3. Lois usuelles sur un univers dénombrable

a. Loi géométrique

Définition-Proposition 17. *Loi géométrique*

Soit $p \in]0,1[$. La famille $(p_k)_{k\in\mathbb{N}^*}$, où, pour tout $k \in \mathbb{N}^*$, $p_k = p(1-p)^{k-1}$, est une distribution de probabilité discrète sur $\mathbb{N}^*$. On appelle **loi géométrique de paramètre** p et on note $\mathcal{G}(p)$ la loi de probabilité discrète définie par cette distribution.

Ainsi une variable aléatoire X **suit une loi géométrique de paramètre** p et on note $X \sim \mathcal{G}(p)$ si :

- $X(\Omega) = \mathbb{N}^*$;
- pour tout $k \in \mathbb{N}^*$,

$$P(X=k) = p(1-p)^{k-1}.$$

Démonstration.

L'ensemble $\mathbb{N}^*$ est dénombrable et la famille $(p_k)_{k\in\mathbb{N}^*}$, qui est une suite à valeurs réelles positives car $p \in [0,1]$, est sommable car terme général d'une série absolument convergente : en effet, il s'agit d'une suite géométrique de raison $1-p \in]0,1[$. De plus, on a :

$$\sum_{k\in\mathbb{N}^*} p_k = p\sum_{k=1}^{+\infty}(1-p)^{k-1} = p\sum_{k=0}^{+\infty}(1-p)^k = \frac{p}{1-(1-p)} = 1.$$

Par suite, la famille $(p_k)_{k\in\mathbb{N}^*}$ une distribution discrète sur $\mathbb{N}^*$. □

Remarque 10.

Une variable aléatoire suivant une loi géométrique de paramètre p modélise le temps d'attente du premier succès d'une suite d'épreuves de Bernoulli indépendantes de paramètre p.

Exercice 6.

Soit $p \in]0,1[$ et X une variable aléatoire qui suit une loi géométrique de paramètre p. Déterminer la valeur de $P(X > k)$ pour $k \in \mathbb{N}^*$.

Correction.

On a $X \sim \mathcal{G}(p)$ donc $X(\Omega) = \mathbb{N}^*$ et pour tout $k \in \mathbb{N}^*$,

$$P(X = k) = p(1-p)^{k-1}.$$

On remarque que (on redémontre ici le résultat de la proposition 9) :

$$(X > k) = (X = k+1) \cup (X = k+2) \cup \ldots = \bigcup_{n=k+1}^{+\infty} (X = n)$$

Les événements $(X = n)$ étant incompatibles, on a alors :

$$P(X > k) = \sum_{n=k+1}^{+\infty} P(X = n) = \sum_{n=k+1}^{+\infty} p(1-p)^{n-1} = p(1-p)^k \underbrace{\sum_{n=0}^{+\infty}(1-p)^n}_{=\frac{1}{p}} = (1-p)^k$$

Proposition 14.

Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N}^*$ et que $P(X = n) \neq 0$ pour tout $n \in \mathbb{N}^*$. Alors X suit une loi géométrique si, et seulement si, pour tous $k, l \in \mathbb{N}^*$,

$$P(X > k + l | X > k) = P(X > l).$$

Démonstration.

On remarque tout d'abord que, pour tous $k, l \in \mathbb{N}^*$, $(X > k+l) \subset (X > k)$ d'où $(X > k+l) \cap (X > k) = (X > k+l)$, et ainsi :

$$P(X > k+l | X > k) = \frac{P((X > k+l) \cap (X > k))}{P(X > k)} = \frac{P(X > k+l)}{P(X > k)}.$$

($\Rightarrow$) On suppose qu'il existe $p \in]0,1[$ tel que $X \sim \mathcal{G}(p)$. Alors, d'après l'exercice précedent, pour tous $k, l \in \mathbb{N}^*$:

$$P(X > k+l | X > k) = \frac{P(X > k+l)}{P(X > k)} = \frac{(1-p)^{k+l}}{(1-p)^k} = (1-p)^l = P(X > l).$$

($\Leftarrow$) On suppose que pour tous $k, l \in \mathbb{N}^*$,

$$\left(\frac{P(X > k+l)}{P(X > k)} = \right) P(X > k+l | X > k) = P(X > l).$$

On pose $p = 1 - P(X > 1) \in]0,1[$ (car les $P(X = n)$ sont non nuls) et pour $n \in \mathbb{N}^*$, $u_n = P(X > n) > 0$. Alors on a :

$$\frac{u_{n+1}}{u_n} = \frac{P(X > n+1)}{P(X > n)} \underbrace{=}_{k=n \text{ et } l=1} P(X > 1) = 1 - p$$

Ainsi, $(u_n)_{n \in \mathbb{N}^*}$ est une suite géométrique de raison $1-p$ et donc, pour tout $n \in \mathbb{N}^*$:

$$P(X > n) = u_n = (1-p)^{n-1} u_1 = (1-p)^{n-1} P(X > 1) = (1-p)^n$$

On remarque que la formule précédente est toujours valable pour $n = 0$ car X étant à valeurs dans $\mathbb{N}^*$, $P(X > 0) = 1$.

Par suite, comme, pour $k \in \mathbb{N}^*$,

$$(X = k) = (X \leq k) \cap (X > k-1) = \overline{(X > k)} \cap (X > k-1) = (X > k-1) \setminus (X > k)$$

et $(X > k) \subset (X > k-1)$, on a :

$$\begin{aligned} P(X = k) &= P(X > k-1) - P(X > k) \\ &= (1-p)^{k-1} - (1-p)^k \\ &= (1-p)^{k-1}(1 - (1-p)) \\ P(X = k) &= p(1-p)^{k-1} \end{aligned}$$

Il en résulte que $X \sim \mathcal{G}(p)$.

□

b. Loi de Poisson

Définition-Proposition 18. *Loi de Poisson*

Soit $\lambda \in \mathbb{R}_+^*$. La famille $(p_k)_{k\in\mathbb{N}}$, où, pour tout $k \in \mathbb{N}$, $p_k = e^{-\lambda}\frac{\lambda^k}{k!}$, est une distribution de probabilité discrète sur $\mathbb{N}$. On appelle **loi de Poisson de paramètre** λ et on note $\mathcal{P}(\lambda)$ la loi de probabilité discrète définie par cette distribution.

Ainsi une variable aléatoire X **suit une loi de Poisson de paramètre** λ i.e. $X \sim \mathcal{P}(\lambda)$ si :

- $X(\Omega) = \mathbb{N}$;
- pour tout $k \in \mathbb{N}$,

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Démonstration.

L'ensemble $\mathbb{N}$ est dénombrable et la famille $(p_k)_{k\in\mathbb{N}}$, qui est une suite à valeurs réelles positives car $\lambda \in \mathbb{R}_+^*$, est sommable car terme général d'une série absolument convergente : en effet, pour tout $k \in \mathbb{N}$, $p_k > 0$ et on a :

$$\frac{p_{k+1}}{p_k} = \frac{\lambda}{k+1} \xrightarrow[k\to+\infty]{} 0 < 1$$

d'où la convergence absolue de $\sum p_k$ d'après la règle de D'Alembert. De plus, on a, en reconnaissant la somme de la série exponentielle :

$$\sum_{k\in\mathbb{N}} p_k = \sum_{k=0}^{+\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!}e^{-\lambda}e^{\lambda} = 1.$$

Par suite, la famille $(p_k)_{k\in\mathbb{N}}$ une distribution discrète sur $\mathbb{N}$. □

Exercice 7.

Soit $\lambda \in \mathbb{R}_+^*$, $(p_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $]0,1[$ et $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires telle que pour chaque $n \in \mathbb{N}$, $X_n \sim \mathcal{B}(n, p_n)$.

Montrer que si $np_n \xrightarrow[n\to+\infty]{} \lambda$, alors, pour tout $k \in \mathbb{N}$,

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

Correction.

Soit $k \in \mathbb{N}$. Pour tout $n \in \mathbb{N}$ avec $n \geq k$, on a $k \in [\![0, n]\!] = X_n(\Omega)$ d'où :

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

Or, comme $np_n \underset{n\to+\infty}{\sim} \lambda$, on a $p_n \underset{n\to+\infty}{\sim} \frac{\lambda}{n} \xrightarrow[n\to+\infty]{} 0$ et $\underset{n\to+\infty}{o}(np_n) \xrightarrow[n\to+\infty]{} 0$, d'où :

- $\star$ $(1-p_n)^{-k} \xrightarrow[n\to+\infty]{} 1$;
- $\star$ $n(n-1)...(n-k+1)p_n^k \underset{n\to+\infty}{\sim} n^k p_n^k = (np_n)^k \xrightarrow[n\to+\infty]{} \lambda^k$;

⋆ et :

$$(1-p_n)^n = e^{n\ln(1-p_n)} = e^{n(-p_n+\underset{n\to+\infty}{o}(p_n))} = e^{-np_n} \times e^{\underset{n\to+\infty}{o}(np_n)} \xrightarrow[n\to+\infty]{} e^{-\lambda}.$$

Il en résulte que :

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

Remarque 11.

En pratique, lorsqu'on se donne une variable aléatoire qui suit une loi binomiale de paramètres n et p pour $n \in \mathbb{N}$ et $p \in]0,1[$ tels que $n \geq 30$, $p \leq 0,1$ et $np \leq 15$, on estime qu'une loi de Poisson de paramètre $\lambda = np$ approxime de manière satisfaisante la loi de X.

4. Couples de variables aléatoires

Dans ce paragraphe, $(\Omega, \mathcal{T}, P)$ désigne un espace probabilisé. Les variables aléatoires considérées sont définies sur cet espace.

a. Généralités

Définition 19. *Couple de variable aléatoires discrètes*

Soit X, Y des variables aléatoires discrètes sur un espace probabilisable. On appelle **couple de variables aléatoires discrètes sur** $(\Omega, \mathcal{T}, P)$ et on note (X, Y) l'application $\omega \mapsto (X(\omega), Y(\omega))$.

Proposition 15.

Soit X, Y des variables aléatoires discrètes à valeurs dans E, F respectivement. Le couple de variables aléatoires (X, Y) est une variable aléatoire discrète sur $(\Omega, \mathcal{T})$ à valeurs dans $E \times F$ et, pour tous $(A, B) \in \mathcal{P}(E) \times \mathcal{P}(F)$, on a :

$$((X, Y) \in A \times B) = (X \in A) \cap (Y \in B),$$

et en particulier, pour tout $(x, y) \in X(\Omega) \times Y(\Omega)$:

$$((X, Y) = (x, y)) = (X = x) \cap (Y = y).$$

Démonstration.

On note $Z = (X, Y) : \omega \mapsto (X(\omega), Y(\omega))$. Alors Z est une fonction de Ω dans $E \times F$ et on a, comme X et Y sont des variables aléatoires discrètes sur $(\Omega, \mathcal{T})$:

— $Z(\Omega) = X(\Omega) \times Y(\Omega)$ qui est au plus dénombrable comme produit cartésien d'ensembles au plus dénombrables ;

— pour $A \subset E$ et $B \subset F$, on a :

$$\begin{aligned}(Z \in A \times B) &= \{\omega \in \Omega \mid Z(\omega) \in A \times B\} \\ &= \{\omega \in \Omega \mid (X(\omega), Y(\omega)) \in A \times B\} \\ &= \{\omega \in \Omega \mid X(\omega) \in A \text{ et } Y(\omega) \in B\} \\ &= \{\omega \in \Omega \mid X(\omega) \in A\} \cap \{\omega \in \Omega \mid Y(\omega) \in B\} \\ (Z \in A \times B) &= (X \in A) \cap (Y \in B)\end{aligned}$$

En particulier, pour $(x, y) \in Z(\Omega) = X(\Omega) \times Y(\Omega)$, on a, comme $(X = x), (Y = y) \in \mathcal{T}$ et $\mathcal{T}$ est stable par intersection :

$$(Z = (x, y)) = (Z \in \{x\} \times \{y\}) = (X \in \{x\}) \cap (Y \in \{x\}) = (X = x) \cap (Y = y) \in \mathcal{T}.$$

Il en résulte que $Z = (X, Y)$ est une variable aléatoire discrète sur $(\Omega, \mathcal{T})$. □

Dans la littérature, on pourra croiser la notation suivante, mais dans la suite de ce chapitre, elle ne sera pas systématiquement utilisée :

Notation 3.

Soit X, Y des variables aléatoires discrètes à valeurs dans E, F respectivement. Pour $A \subset E$ et $B \subset F$, on note :

$$(X \in A,\, Y \in B) = (X \in A) \cap (Y \in B).$$

En particulier, pour $(x, y) \in E \times F$, la notation devient :

$$(X = x, Y = y) = (X = x) \cap (Y = y).$$

Définition-Proposition 20.

Soit (X, Y) un couple de variables aléatoires discrètes. La famille d'événements :

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

est un système complet d'événements que l'on appelle **système complet d'événements associé au couple** (X, Y).

Démonstration.

D'après la proposition 15, (X, Y) est une variable aléatoire discrète, donc, d'après la proposition 11, $((X, Y) = (x, y))_{(x,y) \in (X,Y)(\Omega)}$ est un système complet d'événements. Or, toujours d'après la proposition 15, on a :

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

Ainsi, $((X = x) \cap (Y = y))_{(x,y) \in X(\Omega) \times Y(\Omega)}$ est un système complet d'événements. □

b. Loi conjointe

Définition 21. *Loi conjointe*

Soit (X, Y) un couple de variables aléatoires discrètes. On appelle **loi conjointe de** (X, Y) et on note $P_{(X,Y)}$ la loi de probabilité du couple (X, Y).

Remarque 12.

Pour (X, Y) un couple de variables aléatoires discrètes, comme (X, Y) est une variable aléatoire discrète, la loi conjointe du couple est entièrement déterminée par la donnée de la famille $(P(X = x, Y = y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$.

c. Lois marginales

Définition 22. *Loi marginales*

Soit (X, Y) un couple de variables aléatoires discrètes. On appelle **première loi marginale du couple** la loi de probabilité de X et **deuxième loi marginale du couple** la loi de probabilité de Y.

Le théorème suivant donne une expression des lois marginales en fonction de la loi conjointe.

Théorème 7.

Soit (X, Y) un couple de variables aléatoires discrètes. On a :

— pour tout $x \in X(\Omega)$, $P(X = x) = \sum_{y\in Y(\Omega)} P(X = x,\ Y = y)$.

— pour tout $y \in Y(\Omega)$, $P(Y = y) = \sum_{x\in X(\Omega)} P(X = x,\ Y = y)$.

Démonstration.

— Soit $x \in X(\Omega)$. Comme Y est une variable aléatoire discrète, $((Y = y))_{y\in Y(\Omega)}$ est un système complet d'événements et donc, d'après la formule des probabilités totales (Théorème 5), on a :

$$P(X = x) = \sum_{y\in Y(\omega)} P((X = x) \cap (Y = y)) = \sum_{y\in Y(\Omega)} P(X = x,\ Y = y).$$

— On raisonne de manière similaire ou bien on applique le point précédent au couple (Y, X) en remarquant que, pour $(y, x) \in Y(\Omega) \times X(\Omega)$, $P(Y = y,\ X = x) = P(X = x,\ Y = y)$. □

d. Lois conditionnelles

Définition 23. *loi conditionnelle*

Soit (X, Y) un couple de variables aléatoires discrètes et $x \in X(\Omega)$ tel que $P(X = x) \neq 0$. On appelle **loi conditionnelle de Y sachant $X = x$** la probabilité notée $P_{(X=x)}$ sur $(\Omega, \mathcal{T})$ déterminée, pour $y \in Y(\Omega)$, par :

$$P_{(X=x)}(Y = y) = P(Y = y|X = x).$$

Exercice 8.

Soit (X, Y) un couple de variables aléatoires discrètes tel que, pour tout $x \in X(\Omega)$, $P(X = x) \neq 0$.

1. Soit $(x, y) \in X(\Omega) \times Y(\Omega)$. Montrer que
$$P((X = x) \cap (Y = y)) = P(X = x)P(Y = y|X = x).$$
2. Soit $y \in Y(\Omega)$. Montrer que
$$P(Y = y) = \sum_{x \in X(\Omega)} P(X = x)P(Y = y|X = x).$$

Correction.

1. Il s'agit de la définition de la probabilité conditionnelle :
$$P(Y = y|X = x) = \frac{P((X = x) \cap (Y = y))}{P(X = x)}$$
2. On applique la formule des probabilités totales au système complet d'événements $((X = x))_{x \in X(\Omega)}$.

Remarque 13.

On généralise toutes les notions précédentes aux cas d'un n-uplet de variables aléatoires.

5. Variables aléatoires indépendantes

a. Couple de variables aléatoires indépendantes

Définition 24. *Couple de variables aléatoires indépendantes*

Soit (X, Y) un couple de variables aléatoires discrètes. On dit que X et Y sont **indépendantes**

et on note $X \perp\!\!\!\perp Y$ si, pour tout $(x, y) \in X(\Omega) \times Y(\Omega)$,

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

Proposition 16.

Soit (X, Y) un couple de variables aléatoires discrètes. Les variables X et Y sont indépendantes si, et seulement si, pour tous $(A, B) \in \mathcal{P}(X(\Omega)) \times \mathcal{P}(Y(\Omega))$,

$$P(X \in A,\ Y \in B) = P(X \in A)P(Y \in B).$$

Démonstration.

($\Rightarrow$) On suppose X et Y indépendantes. Soit $(A, B) \in \mathcal{P}(X(\Omega)) \times \mathcal{P}(Y(\Omega))$. On a :

$$(X \in A, Y \in B) = \bigcup_{(x,y)\in A\times B} ((X, Y) = (x, y)) = \bigcup_{(x,y)\in A\times B} (X = x,\ Y = y).$$

Comme (X, Y) est une variable aléatoire, la famille $((X = x,\ Y = y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$ est un système complet d'événements et ainsi, la famille la famille $(P(X = x,\ Y = y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$ est sommable (de somme 1) d'après la proposition 3. Par suite, sa sous-famille $(P(X = x,\ Y = y))_{(x,y)\in A\times B}$ est sommable et, d'après le théorème de sommation par paquets appliqué à la partition verticale de $A \times B = \bigcup_{x\in A}\{(x, y) \mid y \in B\}$, on a :

$$\sum_{(x,y)\in A\times B} P(X = x,\ Y = y) = \sum_{x\in A}\sum_{y\in B} P(X = x,\ Y = y).$$

Ainsi, comme $((X = x,\ Y = y))_{(x,y)\in A\times B}$ est une famille au plus dénombrable d'événements deux à deux disjoints et comme X et Y sont indépendantes :

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

($\Leftarrow$) Si, pour tous $(A,B) \in \mathcal{P}(X(\Omega)) \times \mathcal{P}(Y(\Omega))$, $P(X \in A,\ Y \in B) = P(X \in A)P(Y \in B)$, alors, pour tout $(x,y) \in X(\Omega) \times Y(\Omega)$, en appliquant l'hypothèse à $A = \{x\}$ et $B = \{y\}$, on obtient que X et Y sont indépendantes.

□

Proposition 17.

Soit (X,Y) un couple de variables aléatoires discrètes et f, g des fonctions définies sur $X(\Omega)$ et $Y(\Omega)$ respectivement. Si X et Y sont indépendantes, alors $f(X)$ et $g(Y)$ sont indépendantes.

Démonstration.

On suppose $X \perp\!\!\!\perp Y$. On note $Z = f(X)$ et $T = g(Y)$. Alors, Z et T sont des variables aléatoires discrètes d'après la proposition-notation 12 et on a, pour tout $(z,t) \in Z(\Omega) \times T(\Omega)$:

$$(Z = z) = \{\omega \in \Omega \mid f(X(\omega)) = z\} = \{\omega \in \Omega \mid X(\omega) \in f^{-1}(\{z\})\} = (X \in f^{-1}(\{z\}))$$

et de manière analogue, $(T = t) = (Y \in g^{-1}(\{t\}))$.
Par suite, d'après la proposition 16, on a :

$$\begin{aligned} P(Z = z,\ T = t) &= P((X \in f^{-1}(\{z\})),\ (Y \in g^{-1}(\{t\}))) \\ &= P((X \in f^{-1}(\{z\})))P(Y \in g^{-1}(\{t\}))) \\ P(Z = z,\ T = t) &= P(Z = z)P(T = t) \end{aligned}$$

D'où $f(X) = Z$ et $g(Y) = T$ sont indépendantes. □

Proposition 18. *Loi d'une somme de variables aléatoires discrètes*

Soit (X,Y) un couple de variables aléatoires discrètes à valeurs dans $\mathbb{N}$. Si X et Y sont indépendantes, alors, pour tout $n \in \mathbb{N}$,

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

Démonstration.

On pose $Z = X + Y$ et on suppose $X \perp\!\!\!\perp Y$. Alors Z est à valeurs dans $\mathbb{N}$ et on remarque que, pour $n, x, y \in \mathbb{N}$, $x + y = n$ si, et seulement si, $x \in [\![0,n]\!]$ et $y = n - x$, donc :

$$(Z = n) = \bigcup_{k \in [\![0,n]\!]} ((X = k) \cap (Y = n - k))$$

Ainsi, comme les événements de la famille $((X = k))_{k \in [\![0,n]\!]}$ sont incompatibles, ceux de la famille

$((X=k)\cap(Y=n-k))_{k\in[\![0,n]\!]}$ le sont aussi, d'où :

$$P(Z=n)=P\left(\bigcup_{k\in[\![0,n]\!]}((X=k)\cap(Y=n-k))\right)=\sum_{k=0}^{n}P((X=k)\cap(Y=n-k))$$

et ainsi, comme X et Y sont indépendantes :

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

□

Exercice 9.

Soit (X,Y) un couple de variables aléatoires discrètes tel que X et Y sont indépendantes.

1. Soit $\lambda,\mu\in\mathbb{R}_+^*$. On suppose $X\sim\mathcal{P}(\lambda)$ et $Y\sim\mathcal{P}(\mu)$. Montrer que $X+Y\sim\mathcal{P}(\lambda+\mu)$.
2. Soit $p,q\in\,]0,1[$. On suppose $X\sim\mathcal{G}(p)$ et $Y\sim\mathcal{G}(q)$. Déterminer la loi de $X+Y$. *On donnera une expression simple de la distribution associée et on distinguera bien deux cas !*

Correction.

1. Soit $\lambda,\mu\in\mathbb{R}_+^*$. On suppose $X\sim\mathcal{P}(\lambda)$ et $Y\sim\mathcal{P}(\mu)$. On note $Z=X+Y$. Comme $X(\Omega)=\mathbb{N}$ et $Y(\Omega)=\mathbb{N}$, on a $Z(\Omega)=\{x+y\mid x,y\in\mathbb{N}\}=\mathbb{N}$ et, d'après la proposition 18, pour tout $n\in\mathbb{N}$, on a :

$$\begin{aligned}P(Z=n) &= \sum_{k=0}^{n}P(X=k)P(Y=n-k)\\ &= \sum_{k=0}^{n}e^{-\lambda}\frac{\lambda^k}{k!}.e^{-\mu}\frac{\mu^{n-k}}{(n-k)!}\\ &= e^{-(\lambda+\mu)}\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}\lambda^k\mu^{n-k}\\ P(Z=n) &= e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^n}{n!}\text{ d'après la formule du binôme de Newton.}\end{aligned}$$

   Par suite, $X+Y=Z$ suit une loi de Poisson de paramètre $\lambda+\mu$.

2. Soit $p,q\in\,]0,1[$. On suppose $X\sim\mathcal{G}(p)$ et $Y\sim\mathcal{G}(q)$. On note $Z=X+Y$. Comme $X(\Omega)=\mathbb{N}^*$ et $Y(\Omega)=\mathbb{N}^*$, on a $Z(\Omega)=\{x+y\mid x,y\in\mathbb{N}^*\}=\{n\in\mathbb{N}\mid n\geq 2\}$ et, d'après

la proposition 18, pour tout entier $n \geq 2$, on a :

$$\begin{aligned} P(Z=n) &= \sum_{k=0}^{n} P(X=k)P(Y=n-k) \\ &= \sum_{k=1}^{n-1} p(1-p)^{k-1}.q(1-q)^{n-k-1} \\ P(Z=n) &= pq(1-q)^{n-2}\sum_{k=0}^{n-2}\left(\frac{1-p}{1-q}\right)^{k} \end{aligned}$$

On distingue alors deux cas :
— Cas $p=q$. Dans ce cas, on a alors :

$$P(Z=n) = (n-1)p^2(1-p)^{n-2}$$

— Cas $p \neq q$. Dans ce cas, on a alors :

$$P(Z=n) = pq(1-q)^{n-2}\frac{1-\left(\frac{1-p}{1-q}\right)^{n-1}}{1-\frac{1-p}{1-q}} = pq\frac{(1-q)^{n-1}-(1-p)^{n-1}}{p-q}.$$

b. Famille finie de variables aléatoires indépendantes

Définition 25. *Variables aléatoires indépendantes*

Soit $n \in \mathbb{N}^*$ et $(X_k)_{k\in[\![1,n]\!]}$ une famille de variables aléatoires discrètes. On dit que les variables $X_1, \ldots, X_n$ sont **indépendantes** (ou mutuellement indépendantes) si, pour tout $(x_1, \ldots, x_n) \in X_1(\Omega) \times \ldots \times X_n(\Omega)$:

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{k=1}^{n} P(X_k = x_k).$$

Proposition 19.

Soit $n \in \mathbb{N}^*$ et $(X_k)_{k\in[\![1,n]\!]}$ une famille de variables aléatoires discrètes. Les variables $X_1, \ldots, X_n$ sont indépendantes si, et seulement si, pour tout $(A_1, \ldots A_n) \subset X_1(\Omega) \times \ldots \times X_n(\Omega)$,

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{k=1}^{n} P(X_k \in A_k).$$

Proposition 20. *Lemme des coalitions*

Soit $n, m \in \mathbb{N}^*$ avec $m < n$, $(X_k)_{k\in[\![1,n]\!]}$ une famille de variables aléatoires discrètes et f, g des fonctions définies sur $X_1(\Omega) \times \ldots \times X_m(\Omega)$ et $X_{m+1}(\Omega) \times \ldots \times X_n(\Omega)$ respectivement. Si $X_1, \ldots, X_n$ sont indépendantes, alors $f(X_1, \ldots, X_m)$ et $g(X_{m+1}, \ldots, X_n)$ sont indépendantes.

Exercice 10.

Soit $m \in \mathbb{N}^*$, $n_1, ..., n_m \in \mathbb{N}$, $p \in [0, 1]$ et $X_1, ..., X_m$ des variables aléatoires telles que, pour $k \in [\![1, m]\!]$, $X_k \sim \mathcal{B}(n_k, p)$.
On suppose $X_1, ..., X_n$ indépendantes. Déterminer la loi de $Y = X_1 + \ldots + X_n$.

c. Famille infinie de variables aléatoires indépendantes

Définition 26.

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires discrètes. On dit que les variables X_n sont **indépendantes** si, pour toute partie finie $I = \{i_1, ..., i_k\}$ de $\mathbb{N}$, les variables $X_{i_1}, \ldots, X_{i_k}$ sont indépendantes.

Le théorème suivant (dont la démonstration est admise) va nous permettre de pouvoir "parler sans crainte" de suites de variables aléatoires indépendantes suivant chacune une loi prescrite :

Théorème 8. *Théorème d'existence de Kolmogorov*

Si $(\mathcal{L}_n)_{n\in\mathbb{N}}$ est une suite de lois discrètes, alors il existe un espace probabilisé et une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires discrètes indépendantes sur cet espace telle que, pour tout $n \in \mathbb{N}$, $X_n \sim \mathcal{L}_n$.

Exemple 2.

Soit $p \in [0, 1]$ et $(X_n)_{n\in\mathbb{N}^*}$ une suites de variables aléatoires indépendantes suivant chacune une loi de Bernouilli de paramètre p - *le théorème précédent (et il n'est là que pour cela) garantit que cette suite est bien définie !*
La variable $T = \min\{n \in \mathbb{N}^* \cup \{\infty\} \mid X_n = 1\}$ suit une loi géométrique de paramètre p.

Partie D

Espérance, variance

Sauf mention contraire, les variables aléatoires discrètes considérées dans cette partie sont définies sur un espace probabilisé $(\Omega, \mathcal{T}, P)$ et sont à valeurs réelles.

1. Espérance

a. Définitions et exemples

Définition 27. *Espérance*

Soit X une variable aléatoire discrète .

- Si X est presque sûrement à valeurs réelles positives, on appelle **espérance de** X et on note $E(X)$ la quantité (potentiellement infinie) :
$$E(X) = \sum_{x \in X(\Omega)} xP(X = x).$$
- Si X est à valeurs dans $\mathbb{R}$, on dit que X **admet une espérance finie**, si la famille $(xP(X = x))_{x \in X(\Omega)}$ est sommable. Dans ce cas, on appelle **espérance de** X et on note $E(X)$ la quantité :
$$E(X) = \sum_{x \in X(\Omega)} xP(X = x).$$
- On note $L^1(\Omega, P)$, ou plus simplement L^1, l'ensemble des variables aléatoires discrètes admettant une espérance finie.

Remarque 14.

— Si $X \geq 0$ presque sûrement i.e. $P(X \geq 0) = 1$, $(xP(X = x))_{x \in X(\Omega)}$ est une famille de réels positifs car, pour tout $x \in X(\Omega)$ avec $x < 0$, on a $(X = x) \subset (X < 0)$ d'où $P(X = x) \leq P(X < 0) = 1 - P(X \geq 0) = 0$ et donc $xP(X = x) = 0$.
Ainsi, la somme de la famille $(xP(X = x))_{x \in X(\Omega)}$ est bien définie dans $\mathbb{R}_+ \cup \{+\infty\}$ ce qui justifie la bonne définition de l'espérance $E(X)$.

— En toute rigueur, la définition exacte de $L^1(\Omega, P)$ est l'ensemble des classes d'équivalence des variables aléatoires admettant une espérance finie modulo la relation $X \sim Y \Leftrightarrow X = Y$ presque sûrement. Dans le cadre du programme, nous ne développerons pas ces considérations.

Exemple 3. *Espérance des lois usuelles*

Soit X une variable aléatoire discrète.

— Si $X \sim \mathcal{G}(p)$ alors $X \in L^1$ et $E(X) = \dfrac{1}{p}$,

Si $X \sim \mathcal{G}(p)$, alors pour tout $k \in \mathbb{N}^*$, $P(X = k) = p(1-p)^{k-1}$. D'après le critère de D'Alembert, $u_k = kp(1-p)^{k-1}$ est le terme général d'une série convergente car $0 < 1 - p < 1$. Par suite, $X \in L^1$ et

$$E(X) = \sum_{k=1}^{+\infty} kp(1-p)^{k-1}.$$

Calculons la valeur de cette somme. La série entière $\sum_{k\geq 0} x^k$ a pour rayon de convergence 1, donc sa somme $f : x \mapsto \frac{1}{1-x}$ est C^1 sur $]-1,1[$ et on a, pour tout $x \in]-1,1[$:

$$\begin{aligned}
\frac{1}{(1-x)^2} = f'(x) &= \frac{\mathrm{d}}{\mathrm{d}x} \sum_{k=0}^{+\infty} x^k \\
&= \sum_{k=0}^{+\infty} \frac{\mathrm{d}}{\mathrm{d}x} x^k \\
&= \sum_{k=0}^{+\infty} kx^{k-1}
\end{aligned}$$

Par suite, on a, comme $0 < 1 - p < 1$:

$$\begin{aligned}
E(X) &= \sum_{k=1}^{+\infty} kp(1-p)^{k-1} \\
&= p \sum_{k=0}^{+\infty} k(1-p)^{k-1} \\
&= pf'(1-p) = \frac{p}{(1-(1-p))^2} \\
E(X) &= \frac{1}{p}
\end{aligned}$$

— Si $X \sim \mathcal{P}(\lambda)$ alors $X \in L^1$ et $E(X) = \lambda$.

Si $X \sim \mathcal{P}(\lambda)$, alors pour tout $k \in \mathbb{N}$, $P(X = k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$. D'après le critère de D'Alembert, $u_k = k\dfrac{\lambda^k}{k!}$ est le terme général d'une série convergente. Par suite, $X \in L^1$ et

$$E(X) = \sum_{k=1}^{+\infty} ke^{-\lambda}\frac{\lambda^k}{k!}.$$

Calculons la valeur de cette somme. On a :

$$\begin{aligned} E(X) &= \sum_{k=0}^{+\infty} ke^{-\lambda}\frac{\lambda^k}{k!} \\ &= p\sum_{k=1}^{+\infty} ke^{-\lambda}\frac{\lambda^k}{k!} \\ &= \lambda e^{-\lambda}\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda e^{-\lambda}\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!} \\ &= \lambda e^{-\lambda}e^{\lambda} \\ E(X) &= \lambda \end{aligned}$$

Exercice 11.

Déterminer les espérances des lois uniforme sur $[\![1, n]\!]$, de Bernoulli et binomiale.

Exercice 12.

Soit X une variable aléatoire discrète.

1. Soit $c \in \mathbb{R}$. On suppose que $X = c$ presque sûrement. Montrer que $X \in L^1$ et calculer son espérance X.
2. Soit $M \in \mathbb{R}_+$. On suppose que X est bornée par M presque sûrement. Montrer que $X \in L^1$ et que $|E(X)| \leq M$.

Correction.

1. On suppose que $X = c$ presque sûrement. Alors, pour tout $x \in X(\Omega)$, on a :

$$P(X = x) = \begin{cases} 0 & \text{si } x \neq c \\ 1 & \text{si } x = c \end{cases}$$

Par suite, la famille $(xP(X = x))_{x \in X(\Omega)}$ est sommable car tous ses éléments sont nuls sauf un nombre fini d'entre eux.

Ainsi, $X \in L^1$ et on a :

$$E(X) = \sum_{x \in X(\Omega)} xP(X = x) = cP(X = c) = c.$$

2. On suppose que $|X| \leq M$ presque sûrement. Alors $P(|X| \leq M) = 1$ et donc, pour tout $x \in X(\Omega)$ tel que $|x| > M$, $(X = x) \subset (|X| > M)$ d'où :

$$P(X = x) \leq P(|X| > M) = 1 - P(|X| \leq M) = 0$$

Par suite, pour tout $x \in X(\Omega)$, on a, $|xP(X = x)| \leq MP(X = x)$ car

$$|xP(X = x)| = |x|P(X = x) \begin{cases} = |x| \times 0 = 0 \leq MP(X = x) & \text{si } x > M \\ \leq MP(X = x) & \text{si } x \leq M \end{cases}.$$

Or, la famille $(P(X = x))_{x \in X(\Omega)}$ est sommable car $((X = x))_{x \in X(\Omega)}$ est un système complet d'événements, donc, par linéarité, $(MP(X = x))_{x \in X(\Omega)}$ l'est aussi.
Ainsi, par comparaison, $(|xP(X = x)|)_{x \in X(\Omega)}$ est sommable et donc, par définition, $(xP(X = x))_{x \in X(\Omega)}$ est une famille sommable.
Il en résulte que X admet une espérance finie et on a, par inégalité triangulaire :

$$\begin{aligned} |E(X)| &= \left| \sum_{x \in X(\Omega)} xP(X = x) \right| \\ &\leq \sum_{x \in X(\Omega)} |xP(X = x)| \\ &\leq \sum_{x \in X(\Omega)} MP(X = x) = M \underbrace{\sum_{x \in X(\Omega)} P(X = x)}_{=1} \\ |E(X)| &\leq M. \end{aligned}$$

Exercice 13.

Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N}^*$ telle que, pour $n \in \mathbb{N}^*$, $P(X = n) = \dfrac{\alpha}{n^2}$ pour un certain $\alpha > 0$.

1. Déterminer α.
2. Quelle est l'espérance de X ?

b. Propriétés de l'espérance

Proposition 21.

Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N}$. Les assertions suivantes sont équivalentes :

i) $X \in L^1$;

ii) la série $\sum P(X > k)$ converge ;

iii) la série $\sum_{k\geq 1} P(X \geq k)$ converge.

Et si l'une de ces trois assertions est vérifiée, on a :

$$E(X) = \sum_{k=0}^{+\infty} P(X > k) = \sum_{k=1}^{+\infty} P(X \geq k).$$

Démonstration.

— i) $\Leftrightarrow$ ii). On suppose $X \in L^1$.

On propose deux preuves : l'une reposant sur le théorème de Fubini positif et l'autre en se basant sur le cas fini puis en passant à la limite.

- *Première preuve.* Comme on veut sommer les $P(X > n)$ et que $P(X > n) = \sum_{k=n+1}^{+\infty} P(X = k)$, on a l'idée de considérer la famille $(p_{k,n})_{(k,n)\in\mathbb{N}^2}$ définie, pour $(k, n) \in \mathbb{N}^2$, par :

$$p_{k,n} = \begin{cases} 0 & \text{si } k \leq n \\ P(X = k) & \text{si } k > n \end{cases}$$

La famille $(p_{k,n})_{(k,n)\in\mathbb{N}^2}$ est une famille de réels positifs, donc, d'après le théorème de Fubini positif, les trois termes de l'égalité suivante sont de même nature et on a, dans $[0, +\infty]$:

$$\sum_{k=0}^{+\infty}\sum_{n=0}^{+\infty} p_{k,n} = \sum_{(k,n)\in\mathbb{N}^2} p_{k,n} = \sum_{n=0}^{+\infty}\sum_{k=0}^{+\infty} p_{k,n}.$$

Or, d'après la proposition 9, on a, pour tout $n \in \mathbb{N}$:

$$P(X > n) = \sum_{k=n+1}^{+\infty} P(X = k) = \sum_{k=0}^{+\infty} p_{k,n}$$

et, on a, pour tout $k \in \mathbb{N}$:

$$kP(X = k) = \sum_{n=0}^{k-1} P(X = k) = \sum_{n=0}^{+\infty} p_{k,n}.$$

D'où :

$$E(X) = \sum_{k=0}^{+\infty} kP(X = k) = \sum_{(k,n)\in\mathbb{N}^2} p_{k,n} = \sum_{n=0}^{+\infty} P(X > n).$$

Ainsi X est d'espérance finie si, et seulement si, $\sum P(X > n)$ converge et dans ce cas, on a dans $\mathbb{R}_+$:

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

- *Seconde preuve.* Soit $n \in \mathbb{N}$. D'après la proposition 9, on a, pour tout $k \in \mathbb{N}$, $P(X = k) = P(X > k-1) - P(X > k)$ donc :

$$\begin{aligned}
\sum_{k=0}^{n} kP(X=k) &= \sum_{k=1}^{n} kP(X=k) \\
&= \sum_{k=1}^{n} k(P(X>k-1) - P(X>k)) \\
&= \sum_{k=0}^{n-1} (k+1)P(X>k) - \sum_{k=1}^{n} kP(X>k) \\
&= \sum_{k=1}^{n-1} (k+1-k)P(X>k) + P(X>0) - nP(X>n) \\
\sum_{k=0}^{n} kP(X=k) &= \sum_{k=0}^{n} P(X>k) - nP(X>n)
\end{aligned}$$

  Par suite, on a :

$$\sum_{k=0}^{n} P(X>k) = \sum_{k=0}^{n} kP(X=k) + nP(X>n) \qquad (*)$$

  ⋆ i) ⇒ ii). Si $X \in L^1$, alors la série $\sum kP(X=k)$ converge donc :
  — la suite $(\sum_{k=0}^{n} kP(X=k))_{n\in\mathbb{N}}$ des sommes partielles de cette série converge ;
  — la suite $(\sum_{k=n+1}^{+\infty} kP(X=k))_{n\in\mathbb{N}}$ des restes de cette série tend vers 0 et on a, pour tout $n \in \mathbb{N}$, comme $P(X>n) = \sum_{k=n+1}^{+\infty} P(X=k)$:

$$0 \leq nP(X>n) = \sum_{k=n+1}^{+\infty} nP(X=k) \leq \sum_{k=n+1}^{+\infty} kP(X=k) \xrightarrow[n\to+\infty]{} 0.$$

  Par suite, d'après $(*)$, la série $\sum P(X>k) = (\sum_{k=0}^{n} P(X>k))_{n\in\mathbb{N}}$ converge comme somme de deux suites convergentes.

  ⋆ ii) ⇒ i). On suppose que $\sum P(X>k)$ converge. Comme, pour tout $k \in \mathbb{N}$, $(X > k+1) \subset (X>k)$, la suite $(P(X>k))_{k\in\mathbb{N}}$ est décroissante.
  Or, si $(u_n)_{n\in\mathbb{N}}$ est une suite de réels positifs décroissante telle que $\sum u_n = (S_n)_{n\in\mathbb{N}}$ converge alors $nu_n \xrightarrow[n\to+\infty]{} 0$. En effet, on a :

$$S_{2n} - S_n = \sum_{k=n+1}^{2n} \underbrace{u_k}_{\geq u_{2n}} \geq nu_{2n} \geq 0$$

  et

$$S_{2n+1} - S_{n+1} = \sum_{k=n+2}^{2n+1} \underbrace{u_k}_{\geq u_{2n+1}} \geq nu_{2n+1} \geq 0$$

  Comme, par convergence de $\sum u_n$, la suite $(S_n)_{n\in\mathbb{N}}$ converge vers un certain S, toutes ses sous-suites également et donc les termes de gauche des équations précédentes tendent vers 0. Ainsi, d'après le théorème des gendarmes les suites $(nu_{2n})_{n\in\mathbb{N}}$

et $(nu_{2n+1})_{n\in\mathbb{N}}$. Par combinaison linéaire, il en est donc de même pour les deux sous-suites complémentaires $(2nu_{2n})_{n\in\mathbb{N}}$ et $((2n+1)u_{2n+1})_{n\in\mathbb{N}}$ de la suite $(nu_n)_{n\in\mathbb{N}}$. Il en résulte que $nu_n \xrightarrow[n\to+\infty]{} 0$.

Par suite, $nP(X>n) \xrightarrow[n\to+\infty]{} 0$. Ainsi, en utilisant $(*)$, la série $\sum kP(X=k) = (\sum_{k=0}^{n} P(X=k))_{n\in\mathbb{N}}$ converge comme somme de deux suites convergentes et donc $X \in L^1$.

De plus, si on suppose i) ou ii), d'après ce qui précède, $X \in L^1$, $\sum P(X>k)$ converge et $nP(X>n) \xrightarrow[n\to+\infty]{} 0$ et on a ainsi, d'après $(*)$:

$$\sum_{k=0}^{+\infty} P(X>k) = \lim_{n\to+\infty} \left(\sum_{k=0}^{n} P(X>k) + nP(X>n) \right) = E(X).$$

— ii) $\Leftrightarrow$ iii). On a, pour tout $k \in \mathbb{N}$, $(X \geq k) = (X>k) \cup (X=k)$ où $(X>k)$ et $(X=k)$ sont incompatibles donc :

$$P(X \geq k) = P(X>k) + P(X=k)$$

Or, comme $((X=k))_{k\in\mathbb{N}}$ est un système complet d'événements, $\sum P(X=k)$ est une série convergente de somme 1 donc les séries $\sum_{k\geq 1} P(X\geq k)$ et $\sum P(X>k)$ sont de même nature et en cas de convergence, on a, en remarquant que $P(X\geq 0) = 1$ car X est à valeurs dans $\mathbb{N}$:

$$\begin{aligned}
\sum_{n=0}^{+\infty} P(X>k) &= \sum_{n=0}^{+\infty} (P(X\geq k) - P(X=k)) \\
&= \sum_{n=0}^{+\infty} P(X\geq k) - \sum_{n=0}^{+\infty} P(X=k) \\
&= \sum_{n=1}^{+\infty} P(X\geq k) + P(X\geq 0) - 1 \\
\sum_{n=0}^{+\infty} P(X>k) &= \sum_{n=1}^{+\infty} P(X\geq k).
\end{aligned}$$

□

Proposition 22. *Positivité*

Soit X une variable aléatoire discrète. Si X est presque sûrement à valeurs positives, alors $E(X) \geq 0$ et $E(X) = 0$ implique $X = 0$ presque sûrement.

Démonstration.

On suppose X presque sûrement à valeurs positives. Par définition, $E(X)$ est la somme d'une famille de réels positifs, donc $E(X) \geq 0$.

On suppose de plus $X \in L^1$ et $E(X) = 0$. Comme X est presque sûrement à valeurs positives (rappelons que cela implique $x'P(X = x') = 0$ si $x' < 0$), la somme $\sum_{x' \in X(\Omega)} x'P(X = x')$ est à termes positifs et donc, pour tout $x \in X(\Omega)$, on a :

$$0 \leq xP(X = x) \leq \sum_{x' \in X(\Omega)} x'P(X = x') = E(X) = 0,$$

donc $xP(X = x) = 0$. Ainsi, pour tout $x \in X(\Omega) \setminus \{0\}$, $P(X = x) = 0$, d'où :

$$P(X = 0) = 1 - \sum_{x \in X(\Omega) \setminus \{0\}} P(X = x) = 1.$$

D'où $X = 0$ presque sûrement. □

Théorème 9. *Comparaison*

Soit X, Y des variables aléatoires discrètes. Si $|X| \leq Y$ presque sûrement et $Y \in L^1$, alors $X \in L^1$ et $|E(X)| \leq E(Y)$.

Démonstration.

On suppose que $|X| \leq Y$ presque sûrement i.e. $P(|X| \leq Y) = 1$ et $Y \in L^1$. On remarque tout d'abord que Y est presque sûrement à valeurs positives donc $(yP(Y = y))_{y \in Y(\Omega)}$ est une famille de réels positifs.

On applique le théorème de sommation par paquets à la famille $(yP(X = x, Y = y))_{(x,y) \in X(\Omega) \times Y(\Omega)}$ sur la partition sur la partition horizontale $(H_y)_{y \in Y(\Omega)}$ où $H_y = \{(x, y) \mid x \in X(\Omega)\}$ pour $y \in Y(\Omega)$:

— Somme sur chaque paquet. Soit $y \in Y(\Omega)$. Pour tout $x \in X(\Omega)$, $P(X = x, Y = y) \leq P(X = x)$ car $(X = x, Y = y) \subset (X = x)$, donc :

$$P(X = x, Y = y) \leq P(X = x).$$

Or, la famille $(P(X = x))_{x \in X(\Omega)}$ est sommable donc, par comparaison, $(P(X = x, Y = y))_{x \in X(\Omega)}$ est sommable.
Par suite, $(yP(X = x, Y = y))_{x \in X(\Omega)}$ est sommable et on a, en notant s_y la somme de cette famille, d'après le théorème 7 :

$$\begin{aligned} s_y &= \sum_{x \in X(\Omega)} yP(X = x, Y = y) \\ &= y \sum_{x \in X(\Omega)} P(X = x, Y = y) \\ s_y &= yP(Y = y) \end{aligned}$$

— Sommes des sommes des paquets. La famille $(s_y)_{y \in Y(\Omega)} = (yP(Y = y))_{y \in Y(\Omega)}$ est sommable car $Y \in L^1$.

Ainsi, d'après le théorème de sommation par paquets, la famille $(yP(X = x, Y = y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$ est sommable et on a :

$$\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} yP(X = x, Y = y) = \sum_{y\in Y(\Omega)} yP(Y = y) = E(Y).$$

De plus, pour tout $(x, y) \in X(\Omega) \times Y(\Omega)$, on a $|xP(X = x, Y = y)| \leq yP(X = x, Y = y)$; en effet :

— si $|x| \leq y$, $|xP(X = x, Y = y)| = |x|P(X = x, Y = y) \leq yP(X = x, Y = y)$;
— si $|x| > y$, on a $(X = x, Y = y) \subset \overline{(|X| \leq Y)}$ d'où $P(X = x, Y = y) \leq 1 - P(|X| \leq Y) = 0$ et donc $|xP(X = x, Y = y)| = 0 \leq 0 = yP(X = x, Y = y)$.

Ainsi, par comparaison, la famille $(|x|P(X = x, Y = y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$ est sommable et donc $(xP(X = x, Y = y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$ l'est aussi.
Par suite, d'après le théorème de sommation par paquets sur la partition verticale $(V_x)_{x\in X(\Omega)}$ où $V_x = \{(x, y) \mid y \in Y(\Omega)\}$ pour $x \in Y(\Omega)$, on a :

— Somme sur chaque paquet. Pour tout $x \in X(\Omega)$, la famille $(xP(X = x, Y = y))_{y\in Y(\Omega)}$ est sommable et on a, en notant s_x la somme de cette famille, d'après le théorème 7 :

$$\begin{aligned} s_x &= \sum_{y\in Y(\Omega)} xP(X = x, Y = y) \\ &= x \sum_{y\in Y(\Omega)} P(X = x, Y = y) \\ s_x &= xP(X = x) \end{aligned}$$

— Sommes des sommes des paquets. La famille $(s_x)_{x\in X(\Omega)} = (xP(X = x))_{x\in X(\Omega)}$ est sommable donc $X \in L^1$.

Et de plus, on a, toujours d'après le théorème de sommation par paquets et le théorème 7 :

$$\begin{aligned} |E(X)| &= \left|\sum_{x\in X(\Omega)} xP(X = x)\right| \\ &\leq \sum_{x\in X(\Omega)} |x|P(X = x) = \sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega)} |x|P(X = x, Y = y) \\ &\leq \sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega)} yP(X = x, Y = y) = \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} yP(X = x, Y = y) \\ |E(X)| &\leq E(Y) \end{aligned}$$

□

Théorème 10. *Théorème de transfert*

Soit X une variable aléatoire discrète et $f : X(\Omega) \to \mathbb{R}$ une fonction. La variable aléatoire $f(X)$ est d'espérance finie si, et seulement si, la famille $(f(x)P(X = x))_{x\in X(\Omega)}$ est sommable. Dans

ce cas, on a :

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

Démonstration.

On considère la variable Y la variable aléatoire discrète $f(X)$ et on pose, pour $y \in Y(\Omega)$:

$$I_y = f^{-1}(\{y\}) = \{x \in X(\Omega) \mid f(x) = y\}.$$

On remarque qu'on a, pour tout $y \in Y(\Omega)$, $(X \in I_y) = (Y = y)$ par définition de I_y.

La famille $(I_y)_{y \in Y(\Omega)}$ forme une partition de $X(\Omega)$ et d'après le théorème de sommation par paquets appliqué à la famille $(f(x)P(X = x))_{x \in X(\Omega)}$ et à la partition précédente, on a :

la famille $(f(x)P(X = x))_{x \in X(\Omega)}$ est sommable si, et seulement si,

- Somme sur chaque paquet. Pour tout $y \in Y(\Omega)$, $(f(x)P(X = x))_{x \in I_y} = (yP(X = x))_{x \in I_y}$ est sommable. Or c'est le cas ici car $(yP(X = x))_{x \in I_y}$ est une sous-famille de la famille $(yP(X = x))_{x \in X(\Omega)}$ qui est sommable car $(P(X = x))_{x \in X(\Omega)}$ l'est. De plus, on a, en notant s_y la somme de la famille $(f(x)P(X = x))_{x \in I_y}$:

$$s_y = \sum_{x \in I_y} f(x)P(X = x) = y \sum_{x \in I_y} P(X = x) = yP(X \in I_y) = yP(Y = y).$$

- Sommes des sommes des paquets. La famille $(s_y)_{y \in Y(\Omega)} = (yP(Y = y))_{y \in Y(\Omega)}$ est sommable. Or, c'est le cas si, et seulement si, Y est d'espérance finie.

Il en résulte que la famille $(f(x)P(X = x))_{x \in X(\Omega)}$ est sommable si, et seulement si, $Y = f(X)$ est d'espérance finie. De plus, dans ce cas, on a, toujours le théorème de sommation par paquets :

$$E(f(X)) = E(Y) = \sum_{y \in Y(\Omega)} \underbrace{P(Y = y)}_{= s_y} = \sum_{y \in Y(\Omega)} \sum_{x \in I_y} f(x)P(X = x) = \sum_{x \in X(\Omega)} f(x)P(X = x).$$

□

Proposition 23. *Intégrabilité / Inégalité triangulaire*

Soit X une variable aléatoire discrète.

Intégrabilité. $X \in L^1$ si, et seulement si, $|X| \in L^1$.

Inégalité triangulaire. Si $X \in L^1$, alors :

$$|E(X)| \leq E(|X|)$$

Démonstration.

- La famille $(xP(X = x))_{x \in X(\Omega)}$ est sommable si, et seulement si, $(|xP(X = x)|)_{x \in X(\Omega)} = (|x|P(X = x))_{x \in X(\Omega)}$ donc, d'après le théorème de transfert, $X \in L^1$ si, et seulement si, $|X| \in L^1$.
- On suppose $X \in L^1$. D'après le point précédent, $Y = |X| \in L^1$. De plus, comme $|X| \leq Y$,

d'après le théorème de comparaison, on a :

$$|E(X)| \leq E(Y) = E(|X|).$$

□

Théorème 11. *Linéarité de l'espérance*

Soit X, Y des variables aléatoires discrètes et $\lambda, \mu \in \mathbb{R}$. Si $X, Y \in L^1$, alors la variable aléatoire discrète $\lambda X + \mu Y \in L^1$ et on a :

$$E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y).$$

Démonstration.

On considère la fonction $f : (x, y) \mapsto \lambda x + \mu y$.
Montrons que la famille $(f(x,y)P((X,Y) = (x,y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$ est sommable.
On remarque que, pour tout $(x, y) \in X(\Omega) \times Y(\Omega)$, on a :

$$|f(x,y)P(X = x, Y = y)| \leq \lambda x P(X = x, Y = y) + \mu y P(X = x, Y = y)$$

On étudie alors séparemment les familles $(xP(X = x, Y = y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$ et $(yP(X = x, Y = y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$.

- On applique le théorème de sommation par paquets à la famille $(xP(X = x, Y = y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$ sur la partition verticale $(V_x)_{x\in X(\Omega)}$ où $V_x = \{(x, y) \mid y \in Y(\Omega)\}$ pour $x \in X(\Omega)$.
  - Somme sur chaque paquet. Soit $x \in X(\Omega)$. Pour tout $y \in Y(\Omega)$, $P(X = x, Y = y) \leq P(Y = y)$ car $(X = x, Y = y) \subset (Y = y)$, donc :

    $$P(X = x, Y = y) \leq P(Y = y).$$

    Or, la famille $(P(Y = y))_{y\in Y(\Omega)}$ est sommable donc, par comparaison, $(P(X = x, Y = y))_{y\in Y(\Omega)}$ est sommable.
    Par suite, $(xP(X = x, Y = y))_{y\in Y(\Omega)}$ est sommable et on a, en notant s_x la somme de cette famille, d'après le théorème 7 :

    $$\begin{aligned} s_x &= \sum_{y\in Y(\Omega)} xP(X = x, Y = y) \\ &= x \sum_{y\in Y(\Omega)} P(X = x, Y = y) \\ s_x &= xP(X = x) \end{aligned}$$

  - Sommes des sommes des paquets. La famille $(s_x)_{x\in X(\Omega)} = (xP(X = x))_{x\in X(\Omega)}$ est sommable car $X \in L^1$.

Par suite, $(xP(X=x,Y=y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$ est une famille sommable et on a :

$$\begin{aligned}\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} xP(X=x,Y=y) &= \sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega)} xP(X=x,Y=y)\\ &= \sum_{x\in X(\Omega)} xP(X=x)=E(X).\end{aligned}$$

- De manière analogue, on applique le théorème de sommation par paquets à la famille $(yP(X=x,Y=y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$ sur la partition horizontale $(H_y)_{y\in Y(\Omega)}$ où $H_y=\{(x,y)\mid x\in X(\Omega)\}$ pour $y\in Y(\Omega)$; on obtient que la famille $(yP(X=x,Y=y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$ est sommable et on a :

$$\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} yP(X=x,Y=y)=\sum_{y\in Y(\Omega)}\sum_{x\in X(\Omega)} yP(X=x,Y=y)=\sum_{y\in Y(\Omega)} yP(Y=y)=E(Y).$$

Il en résulte que $(f(x,y)P((X,Y)=(x,y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$ est une famille sommable comme combinaison linéaire de famille sommable. Ainsi, d'après le théorème de transfert, $\lambda X+\mu Y=f(X,Y)$ est d'espérance finie et, par linéarité de la somme, on obtient :

$$\begin{aligned}E(\lambda X+\mu Y) &= E(f(X,Y))\\ &= \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} f(x,y)P(X=x,Y=y)\\ &= \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} (\lambda x+\mu y)P(X=x,Y=y)\\ &= \lambda\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} xP(X=x,Y=y)+\mu\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} yP(X=x,Y=y)\\ E(\lambda X+\mu Y) &= \lambda E(X)+\mu E(Y).\end{aligned}$$

D'où la linéarité de l'espérance. □

Corollaire 4.

L'ensemble $L^1=L^1(\Omega,P)$ des variables aléatoires discrètes d'espérances finies est muni, avec ses opérations usuelles, d'une structure d'espace vectoriel et l'espérance est une forme linéaire sur L^1.

Correction.

On montre qu'il s'agit d'un sous-espace vectoriel de l'espace $\mathcal{F}(\Omega,\mathbb{R})$ des fonctions de Ω dans $\mathbb{R}$.

— La fonction nulle $\mathbf{0}$ admet une esparance finie car la famille finie $(0)_{0\in\{0\}}$ est sommable, d'où $\mathbf{0}\in L^1$.

— D'après le théorème précédent, L^1 est stable par combinaison linéaire.

Il en résulte que L^1 est une sous-espace vectoriel de $\mathcal{F}(\Omega,\mathbb{R})$ et donc lui-même un espace vectoriel.

De plus, le théorème précédent implique que l'application $E : L^1 \to \mathbb{R}$ est linéaire et donc que l'espérance est une application linéaire de L^1 dans $\mathbb{R}$.

Proposition 24. *Croissance*

Soit X, Y des variables aléatoires discrètes. Si $X, Y \in L^1$ et $X \leq Y$ presque sûrement, alors $E(X) \leq E(Y)$.

Démonstration.

On suppose $X \leq Y$ presque sûrement et $X, Y \in L^1$. Alors, $Z = Y - X \in L^1$ d'après le corollaire 4 et Z est positive presque sûrement, donc, par positivité (Proposition 22) et par linéarité de l'espérance :

$$E(Y) - E(X) = E(Z) \geq 0,$$

d'où $E(X) \leq E(Y)$. □

2. Variance

a. Moments

Définition 28. *Moments*

Soit X une variable aléatoire discrète et $p \in \mathbb{N}^*$.

- On dit que X **admet un moment d'ordre** p si la variable aléatoire discrète X^p admet une espérance finie. Dans ce cas, on appelle **moment d'ordre** p la quantité $E(X^p)$.
- On note $L^p(\Omega, P)$, ou plus simplement L^p, l'ensemble des variables aléatoires discrètes admettant un moment d'ordre p.

Proposition 25.

Soit X, Y des variables aléatoires discrètes.

- Si $X, Y \in L^2$, alors $XY \in L^1$.
- Si $X \in L^2$, alors $X \in L^1$.

Démonstration.

- Remarquons tout d'abord que pour tous $x, y \in \mathbb{R}$, comme $(x \pm y)^2 \geq 0$, on a $|xy| \leq \frac{x^2+y^2}{2}$. Supposons X^2 et Y^2 d'espérance finie. Par linéarité de l'espérance, on a $Z = \frac{X^2+Y^2}{2} \in L^1$. De plus, d'après la remarque initiale, on a $|XY| \leq Z$, donc, par comparaison (Théorème 9), XY est d'espérance finie.
- On suppose X^2 d'espérance finie et on pose $Y = 1$ (variable aléatoire constante en 1). Alors $Y \in L^2$, donc, d'après le point précédent, $X = XY \in L^1$.

□

Corollaire 5.

L'ensemble $L^2(\Omega, P)$ des variables aléatoires discrètes admettant un moment d'ordre 2 est muni, avec ses opérations usuelles, d'une structure d'espace vectoriel et on a $L^2 \subset L^1$.

Démonstration.

De même que pour L^1, il suffit de montrer que L^2 est un sous-espace vectoriel de $\mathcal{F}(\Omega, \mathbb{R})$. La fonction nulle appartient à L^2 car la famille finie $(0)_{0\in\{0\}}$ est sommable et, pour tous $X, Y \in L^2$ et tous $\lambda, \mu \in \mathbb{R}$, $(\lambda X + \mu Y)^2$ admet une espérance finie : en effet, on a $(\lambda X + \mu Y)^2 = \lambda^2 X^2 + \lambda\mu XY + \mu^2 Y^2$ donc, comme $XY \in L^1$ d'après la proposition précédente, par linéarité de l'espérance, $(\lambda X + \mu Y)^2 \in L^1$.
De plus, d'après la proposition précédente, si $X \in L^2$, alors $X \in L^1$ d'où $L^2 \subset L^1$. □

Plus généralement, on a le resultat suivant :

Proposition 26.

Soit $p, q \in \mathbb{N}^*$ avec $q \geq p$. Montrer qu'avec ses opérations usuelles, $L^p(\Omega, P)$ est muni d'une structure d'espace vectoriel et que $L^q \subset L^p$.

Correction.

Soit $p \in \mathbb{N}^*$. Montrons que L^p est un sous-espace vectoriel de $\mathcal{F}(\Omega, \mathbb{R})$.
La fonction nulle appartient à L^p car la famille finie $(0)_{0\in\{0\}}$ est sommable. Montrons la stabilité par combinaison linéaire.
On remarque que pour tous $x, y \in R_+$, on a, par convexité de la fonction $t \mapsto t^p$ sur $\mathbb{R}_+$, $\left(\frac{x}{2} + \frac{y}{2}\right)^p \leq \frac{x^p + y^p}{2}$ et donc :

$$(x + y)^p \leq 2^{p-1}(x^p + y^p).$$

Par suite, pour tous $X, Y \in L^p$ et $\lambda, \mu \in \mathbb{R}$:

$$|(\lambda X + \mu Y)^p| = |\lambda X + \mu Y|^p \leq 2^{p-1}|\lambda|^p|X^p| + 2^{p-1}|\mu|^p|Y^p|.$$

Or, comme $X, Y \in L^p$, $X^p, Y^p \in L^1$ et donc $|X^p|, |Y^p| \in L^1$ d'après la proposition 23. Par suite, par linéarité de l'espérance $2^{p-1}|\lambda|^p|X^p| + 2^{p-1}|\mu|^p|Y^p| \in L^1$ et donc, par comparaison, $(\lambda X + \mu Y)^p \in L^1$ i.e. $\lambda X + \mu Y \in L^p$.
Il en résulte que L^p est un sous-espace vectoriel de $\mathcal{F}(\Omega, \mathbb{R})$ et donc un espace vectoriel lui-même.

Soit $q \in \mathbb{N}^*$ avec $q \geq p$. On remarque que, pour tout $x \in \mathbb{R}_+$,

$$x^p \leq \begin{cases} 1 & \text{si } x \leq 1 \\ x^q & \text{si } x > 1 \end{cases} \leq 1 + x^q.$$

Ainsi, pour tout $X \in L^q$, on a $|X^p| = |X|^p \leq 1 + |X|^q \in L^1$ et donc $X \in L^p$. D'où l'inclusion $L^q \subset L^p$.

Exercice 14.

Soit X une variable aléatoire discrète. Pour $n \in \mathbb{N}$, on note :

$$X_n = \prod_{k=0}^{n-1}(X-k) = X(X-1)...(X-n+1).$$

1. Soit $n, m \in \mathbb{N}^*$ avec $n \leq m$. Montrer : si $X_m \in L^1$ alors $X_n \in L^1$.
2. Soit $p \in \mathbb{N}^*$. Montrer que $X \in L^p$ si, et seulement si, $X_p \in L^1$.

Correction.

1. Soit $n, m \in \mathbb{N}^*$ avec $n \leq m$. Soit $x \in \mathbb{R}$.
   — Si $|x| \leq m$, alors, pour tout $k \in [\![0, m-1]\!]$,
   $$|x-k| \leq |x| + k \leq m + k.$$
   Donc :
   $$\prod_{k=0}^{n-1}|x-k| \leq \frac{(m+n-1)!}{(m-1)!}.$$
   — Si $|x| > m$, alors, pour tout $k \in [\![0, m-1]\!]$,
   $$|x-k| \geq |x| - k > m - k \geq 1.$$
   Donc :
   $$\prod_{k=0}^{n-1}|x-k| \leq |x|.|x-1|...|x-n+1|.\underbrace{|x-n|}_{\geq 1}...\underbrace{|x-m+1|}_{\geq 1} = \left|\prod_{k=0}^{m-1}(x-k)\right|.$$
   Ainsi,
   $$|x(x-1)...(x-n+1)| = \prod_{k=0}^{n-1}|x-k| \leq \frac{(m+n-1)!}{(m-1)!} + |x(x-1)...(x-m+1)|.$$
   On suppose $X_m \in L^1$. Alors $|X_m| \in L^1$ et donc, d'après ce qui précède et par linéarité de l'espérance,
   $$|X_n| \leq \underbrace{\frac{(m+n-1)!}{(m-1)!}}_{\in L^1} + |X_m| \in L^1.$$
   Ainsi, par comparaison, $X_n \in L^1$.
2. Soit $p \in \mathbb{N}^*$.
   On considère $\mathbb{R}_p[Y]$ l'espace vectoriel des polynômes de l'indéterminée Y (pour ne pas confondre avec la notation X de variable aléatoire) de degré au plus p et, pour $k \in [\![0, p]\!]$, on note $P_k = \prod_{i=0}^{k-1}(Y-i) \in \mathbb{R}_p[Y]$. Les familles $(Y^k)_{k\in[\![0,p]\!]}$ et $(P_k)_{k\in[\![0,p]\!]}$ sont des bases de $\mathbb{R}_p[Y]$: ce sont des familles libres car de degrés étagés et de cardinal $p+1 = \dim(\mathbb{R}_p[Y])$. Ainsi, comme $P_p, Y^p \in \mathbb{R}_p[Y]$, il existe $a_0, ..., a_p \in \mathbb{R}$ et $b_0, ..., b_p \in \mathbb{R}$ tels que :
   $$P_p = \sum_{k=0}^{p} a_k Y^k \text{ et } Y^p = \sum_{k=0}^{p} b_k P_k.$$

Par suite, $(\mathcal{F}(\Omega,\mathbb{R}),+,\times)$ étant une algèbre, pour $X\in\mathcal{F}(\Omega,\mathbb{R})$, on a :

$$X_p=P_p(X)=\sum_{k=0}^{p}a_kX^k \text{ et } X^p=\sum_{k=0}^{p}b_kP_k(X)=\sum_{k=0}^{p}b_kX_k.$$

Montrons l'équivalence.

($\Rightarrow$) On suppose $X\in L^p$. Alors, pour tout $k\in[\![0,p]\!]$, $X^k\in L^1$ car, pour $k=0$, $1\in L^1$ et, pour $k\geq 1$, $L^p\subset L^k$ (Proposition 26).
Par suite, on a, par linéarité de l'espérance :

$$X_p=\sum_{k=0}^{p}a_k\underbrace{X^k}_{\in L^1}\in L^1.$$

($\Leftarrow$) On suppose $X_p\in L^1$. Alors, pour tout $k\in[\![0,p]\!]$, $X_k\in L^1$ d'après la question précédente pour $k\geq 1$ et, pour $k=0$, car $1\in L^1$.
Par suite, on a, par linéarité de l'espérance :

$$X^p=\sum_{k=0}^{p}b_k\underbrace{X_k}_{\in L^1}\in L^1.$$

Ainsi, $X\in L^p$.

b. Définition et propriétés de la variance

Définition 29. *Variance*

Soit X une variable aléatoire discrète telle que $X\in L^2$. On appelle **variance de** X et on note $V(X)$ la quantité :

$$V(X)=E\left((X-E(X))^2\right).$$

Théorème 12. *Formule de Kœnig-Huygens*

Soit X une variable aléatoire discrète. Si $X\in L^2$, alors on a :

$$V(X)=E(X^2)-E(X)^2.$$

Démonstration.

Soit $X \in L^2$. On a :

$$\begin{aligned} V(X) &= E\left((X - E(X))^2\right) \\ &= E\left(X^2 - 2E(X)X + E(X)^2\right) \\ &= E(X^2) - E(2E(X)X) + E\left(E(X)^2\right) \\ &= E(X^2) - 2E(X)^2 + E(X)^2 \\ V(X) &= E(X^2) - E(X)^2. \end{aligned}$$

□

Astuce usuelle pour le calcul de la variance. Très souvent, pour simplifier le calcul de la variance d'une variable aléatoire discrète $X \in L^2$:

- On calcule $E(X(X-1))$ (on note tout de même que $X(X-1)$ est bien d'espérance finie par linéarité de l'espérance car $X(X-1) = X^2 - X$ et $X \in L^1$ car $X \in L^2$).
- On obtient $V(X) = E(X^2) - E(X)^2 = E(X(X-1)) + E(X) - E(X)^2$ (Théorème de Kœnig-Huygens et linéarité de l'espérance).

Exemple 4. *Variances des lois usuelles*

Soit X une variable aléatoire discrète, $p \in]0,1[$ et $\lambda \in \mathbb{R}_+^*$.

- Si $X \sim \mathcal{G}(p)$ alors $X \in L^2$ et $V(X) = \dfrac{1-p}{p^2}$,

On suppose que $X \sim \mathcal{G}(p)$. Alors

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

$X \in L^2$ si X^2 est d'espérance finie, ce qui est équivalent, d'après le théorème de transfert, à : la famille $(k^2 P(X = k))_{k\in\mathbb{N}}$ est sommable. Comme cette famille est une suite à termes positifs, il suffit de montrer que c'est le terme général d'une série convergente.
Pour tout $k \geq 1$, on a, par croissances comparées :

$$k^2 \times k^2 P(X = k) = k^4 p(1-p)^{k-1} \xrightarrow[k\to+\infty]{} 0$$

d'où $k^2 P(X = k) = \underset{k\to+\infty}{o}(\frac{1}{k^2})$, et donc, par comparasion au terme général d'une série de Riemann convergente, $\sum k^2 P(X = k)$ converge.
Par suite, $(k^2 P(X = k))_{k\in\mathbb{N}^*}$ est sommable et donc $X \in L^2$.
Ainsi, X admet une variance finie et également une espérance finie. Calculons tout d'abord $E(X(X-1))$ (la variable $X(X-1)$ est bien d'espérance finie par linéarité car X^2 et X sont d'espérance finie et on a $X(X-1) = X^2 - X$). D'après le théorème de transfert, on a :

$$E(X(X-1)) = \sum_{k=1}^{+\infty} k(k-1)p(1-p)^{k-1} = p(1-p)\sum_{k=2}^{+\infty} k(k-1)(1-p)^{k-2}.$$

On considère alors la série entière $\sum_{k\geq 2} k(k-1)z^{k-2}$ et sa somme f. Le rayon de convergence de cette série entière est 1 (par exemple d'après la règle de D'Alembert pour les séries entières) et on a, pour tout $x \in]-1, 1[$:

$$\begin{aligned} f(x) &= \sum_{k=2}^{+\infty} k(k-1)x^{k-2} \\ &= \sum_{k=0}^{+\infty} \frac{\mathrm{d}^2}{\mathrm{d}x^2} x^k = \frac{\mathrm{d}^2}{\mathrm{d}x^2} \sum_{k=0}^{+\infty} x^k \\ f(x) &= \frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\frac{1}{1-x}\right) = \frac{2}{(1-x)^3}. \end{aligned}$$

Par suite, on a :

$$E(X(X-1)) = p(1-p)f(1-p) = p(1-p) \times \frac{2}{(1-(1-p))^3} = \frac{2(1-p)}{p^2}.$$

On a, par linéarité, $E(X(X-1)) = E(X^2) - E(X)$ et d'après le théorème de Kœnig-Huygens, $V(X) = E(X^2) - E(X)^2$; par suite :

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

— Si $X \sim \mathcal{P}(\lambda)$ alors $X \in L^2$ et $V(X) = \lambda$.

On suppose que $X \sim \mathcal{P}(\lambda)$. Alors

- $X(\Omega) = \mathbb{N}$
- pour tout $k \in \mathbb{N}$, $P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$.

$X \in L^2$ si X^2 est d'espérance finie, ce qui est équivalent, d'après le théorème de transfert, à : la famille $(k^2 P(X = k))_{k\in\mathbb{N}}$ est sommable. Comme cette famille est une suite à termes positifs, il suffit alors de montrer que c'est le terme général d'une série convergente.
Or, pour tout $k \geq 1$, $u_k = k^2 P(X = k) = k^2 e^{-\lambda}\frac{\lambda^k}{k!} > 0$ et on a :

$$\frac{u_{k+1}}{u_k} = \frac{\lambda(k+1)}{k^2} \xrightarrow[k\to+\infty]{} 0 < 1$$

donc, d'après la régle de D'Alembert, $\sum u_k$ converge.
Par suite, $(k^2 P(X = k))_{k\in\mathbb{N}}$ est sommable et donc $X \in L^2$.
Ainsi, X admet une variance finie et également une espérance finie. Calculons tout d'abord $E(X(X-1))$ (la variable $X(X-1)$ est bien d'espérance finie car X^2 et X sont d'espérance finie et on a $X(X-1) = X^2 - X$) :

$$\begin{aligned} E(X(X-1)) &= \sum_{k=0}^{+\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} \\ &= \lambda^2 e^{-\lambda}\sum_{k=2}^{+\infty}\frac{\lambda^{k-2}}{(k-2)!} \\ &= \lambda^2 e^{-\lambda}\underbrace{\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!}}_{=e^{\lambda}} \\ E(X(X-1)) &= \lambda^2. \end{aligned}$$

On a, par linéarité : $E(X(X-1)) = E(X^2) - E(X)$ et d'après le théorème de Kœnig-Huygens, $V(X) = E(X^2) - E(X)^2$; par suite :

$$V(X) = E(X(X-1)) + E(X) - E(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Exercice 15.

Déterminer les variances des lois uniforme sur $[\![1, n]\!]$, de Bernoulli et binomiale.

Proposition 27.

Soit $\lambda, \mu \in \mathbb{R}$ et X une variable aléatoire discrète. Si $X \in L^2$, alors on a :

$$V(\lambda X + \mu) = \lambda^2 V(X).$$

c. Écart-type

Définition 30. *Écart-type*

Soit X une variable aléatoire discrète telle que $X \in L^2$. On appelle **écart-type de** X et on note $\sigma(X)$ la quantité :

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

Définition 31. *Variable centrée réduite*

Soit X une variable aléatoire discrète telle que $X \in L^2$. On dit que la variable aléatoire X est **centrée** si $E(X) = 0$ et **réduite** si $\sigma(X) = 1$.

Proposition 28.

Soit X une variable aléatoire discrète telle que $X \in L^2$ et $V(X) \neq 0$. Alors la variable aléatoire $X^* = \dfrac{X - E(X)}{\sigma(X)}$ est une variable aléatoire centrée réduite.

Théorème 13. *Inégalité de Cauchy-Schwarz*

Soit X, Y des variables aléatoires discrètes. Si $X, Y \in L^2$ alors :

$$E(XY) \leq \sqrt{E(X^2)}.\sqrt{E(Y^2)}.$$

avec égalité si, et seulement si, X et Y sont presque sûrement proportionnelles.

Démonstration.

Remarque : La première idée pour démontrer cette inégalité serait de montrer que l'application $< \cdot, \cdot >: (X, Y) \mapsto E(XY)$ est un produit scalaire sur L^2. Par linéarité et positivité de l'espérance, on montre aisément qu'il s'agit d'une forme bilinéaire symétrique positive. Mais est-elle définie positive ? Et bien, dans le cadre du programme, la réponse est non en général !
En effet, pour $X \in L^2$, $< X, X >= 0$, alors $E(X^2) = 0$ et donc, d'après la proposition 22, $X^2 = 0$ presque sûrement et ainsi $X = 0$ **presque sûrement** ! Ce qui n'implique pas $X = 0$.
En fait, avec la "vraie" définition de L^2 - l'ensemble des classes d'équivalence des variables aléatoires admettant un moment d'ordre 2 modulo la relation $X \sim Y \Leftrightarrow X = Y$ presque sûrement (muni d'une structure d'espace vectoriel quotient), on gagne bien la définie positivité et $< \cdot, \cdot >$ est bien un produit scalaire sur L^2. Ainsi, l'inégalité de Cauchy-Schwarz est bien vérifiée sur L^2 avec sa définition "hors-programme". On doit donc reproduire la démonstration de l'inégalité de Cauchy-Schwarz, pour rester dans les clous :

Soit $X, Y \in L^2$ et $t \in \mathbb{R}$. Alors, par positivité et linéarité de l'espérance, on a :

$$t^2 E(X^2) + 2tE(XY) + E(Y^2) = E((tX+Y)^2) \geq 0.$$

Par suite, $t \mapsto E((tX+Y)^2)$ est une fonction polynomiale du second degré (au plus) qui est positive, donc son discriminant $\Delta = 4(E(XY)^2 - E(X^2)E(Y^2)$ est négatif. On en déduit donc l'inégalité de Cauchy-Schwarz.
De plus, on a égalité si, et seulement si, $\Delta = 0$ ce qui équivaut à l'existence d'un réel t_0 tel que $E((t_0X+Y)^2) = 0$ i.e., d'après la proposition 22, $t_0X + Y$ est nulle presque sûrement. □

Corollaire 6.

Soit X, Y des variables aléatoires discrètes. Si X, Y admettent un moment d'ordre 2 et sont centrées, alors :

$$E(XY) \leq \sigma(X).\sigma(Y).$$

3. Covariance

a. Définition et propriétés

Définition 32. *Covariance*

Soit (X, Y) un couple de variables aléatoires discrètes tel que $X, Y \in L^2$. On appelle **covariance du couple** (X, Y) et on note $\mathrm{Cov}(X, Y)$ la quantité :

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

Théorème 14. *Formule de Kœnig-Huygens*

Soit (X, Y) un couple de variables aléatoires discrètes. Si $X, Y \in L^2$, alors on a :

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

Proposition 29.

L'application Cov est une forme bilinéaire symétrique positive et, pour tout $X \in L^2$, $\mathrm{Cov}(X, X) = V(X)$.

b. Variance d'une somme

Théorème 15. *Formule de la variance d'une somme*

Soit (X, Y) deux variables aléatoires discrètes. Si $X, Y \in L^2$, alors :

$$V(X+Y) = V(X) + V(Y) + 2\text{Cov}(X, Y).$$

Soit $n \in \mathbb{N}^*$. On a, plus généralement, si $(X_1, ..., X_n)$ est un n-uplet de variables aléatoires discrètes, on a :

$$V(X_1 + \ldots + X_n) = \sum_{i=1}^{n} V(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j).$$

4. Espérance, variance et indépendance

Théorème 16. *Espérance d'un produit de variables indépendantes*

Soit X, Y des variables aléatoires discrètes. Si X et Y sont **indépendantes** et d'espérance finie, alors la variable aléatoire XY est d'espérance finie et on a :

$$E(XY) = E(X)E(Y).$$

Démonstration.

On suppose X et Y indépendantes et d'espérance finie. On considère la variable aléatoire discrète (X, Y) et la fonction $f : (x, y) \mapsto xy$.
Montrons que la famille $(f(x,y)P(X = x, Y = y))_{(x,y) \in X(\Omega) \times Y(\Omega)}$ est sommable. Comme X et Y sont indépendantes, pour tout $(x, y) \in X(\Omega) \times Y(\Omega)$, on a :

$$f(x,y)P(X = x, Y = y) = xyP(X = x)P(Y = y) = xP(X = x).yP(Y = y).$$

On applique le théorème de sommation par paquets à la famille sur la partition horizontale $(H_y)_{y \in Y(\Omega)}$ où $H_y = \{(x, y) \mid x \in X(\Omega)\}$ pour $y \in Y(\Omega)$.

— Somme sur chaque paquet. Soit $y \in Y(\Omega)$. Comme X est d'éspérance finie, la famille $(xP(X = x))_{x \in X(\Omega)}$ est sommable donc, par linéarité, $(yP(Y = y)xP(X = x))_{x \in X(\Omega)}$ est sommable.
Par suite, $(f(x,y)P(X = x, Y = y))_{x \in X(\Omega)}$ est sommable et on a, en notant s_y la somme de cette famille :

$$\begin{aligned} s_y &= \sum_{x \in X(\Omega)} f(x,y)P(X = x, Y = y) \\ &= yP(Y = y) \sum_{x \in X(\Omega)} xP(X = x) \\ s_y &= yP(Y = y)E(X) \end{aligned}$$

— Sommes des sommes des paquets. Comme Y est d'éspérance finie, la famille $(yP(Y = y))_{y \in Y(\Omega)}$ est sommable donc, par linéarité, $(s_y)_{y \in Y(\Omega)} = (E(X)yP(Y = y))_{x \in X(\Omega)}$ est sommable.

Par suite, $(f(x,y)P(X=x,Y=y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$ est une famille sommable et on a :

$$\begin{aligned}\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} f(x,y)P(X=x,Y=y) &= \sum_{y\in Y(\Omega)}\sum_{x\in X(\Omega)} xP(X=x).yP(Y=y)\\ &= \sum_{y\in Y(\Omega)} yP(Y=y)\sum_{x\in X(\Omega)} xP(X=x)\\ &= E(X)\sum_{y\in Y(\Omega)} yP(Y=y)\\ \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} f(x,y)P(X=x,Y=y) &= E(X)E(Y)\end{aligned}$$

Il en résulte, d'après le théorème de sommation par paquets, que $(f(x,y)P(X=x,Y=y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$ est une famille sommable. Ainsi, d'après le théorème de transfert, $XY=f(X,Y)$ est d'espérance finie et on a :

$$E(XY)=\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} f(x,y)P(X=x,Y=y)=E(X)E(Y).$$

□

Exercice 16.

Soit $p,q\in\,]0,1[$ et X,Y des variables aléatoires indépendantes qui suivent des lois géométriques de paramètres p et q respectivement.

On pose $Z=\dfrac{X}{Y}$.

1. Sans calculer $E(Z)$, montrer que si $p=q$ alors $E(Z)>1$.
2. Sans calculer la loi de Z, déterminer $E(Z)$ en fonction de p et q puis retrouver le résultat de la question 1.
3. Déterminer la loi de Z.

Correction.

1. Tout d'abord, on remarque que si $T\sim\mathcal{G}(p)$, alors, pour tout $\alpha\in\mathbb{R}$, $T^\alpha\in L^1$. En effet : supposons $T\sim\mathcal{G}(p)$ et fixons $\alpha\in\mathbb{R}$; alors $(n^\alpha P(T=n))_{n\in T(\Omega)}=(n^\alpha p(1-p)^{n-1})_{n\in\mathbb{N}^*}$ est une famille sommable car, en posant, pour $n\in\mathbb{N}^*$, $u_n=n^\alpha p(1-p)^{n-1}>0$, on a :

$$\frac{u_{n+1}}{u_n}=\left(\frac{n+1}{n}\right)^\alpha(1-p)\xrightarrow[n\to+\infty]{}1-p<1 \text{ car } p>0,$$

et donc, d'après la règle de D'Alembert, $\sum n^\alpha p(1-p)^{n-1}$ converge absolument.
Supposons $p=q$. Alors $X,Y\sim\mathcal{G}(p)$ donc $\sqrt{X}\in L^2$ et, d'après ce qui précède, $\frac{1}{\sqrt{X}}\in L^2$ et $\frac{1}{Y}\in L^1$. Ainsi, d'après l'inégalité de Cauchy-Schwarz, comme $\sqrt{X}$ et $\frac{1}{\sqrt{X}}$ ne sont pas

presque sûrement colinéaires (car X n'est pas presque sûrement constante) :

$$1 = E(1) = E\left(\sqrt{X}.\frac{1}{\sqrt{X}}\right) < \sqrt{E(X)}.\sqrt{E\left(\frac{1}{X}\right)} = \sqrt{E(X)}.\sqrt{E\left(\frac{1}{Y}\right)}.$$

De plus, Comme X et Y sont indépendantes, d'après la proposition 17, X et $\frac{1}{Y}$ le sont aussi et donc, d'après le théorème 16, on a :

$$E(Z) = E\left(\frac{X}{Y}\right) = E\left(X.\frac{1}{Y}\right) = E\left(X\right).E\left(\frac{1}{Y}\right) > 1$$

2. *Remarque : dans la suite, tout est de nouveau justifié comme si la question 1 n'avait pas eu lieu.*

   Comme X et Y sont indépendantes, d'après la proposition 17, X et $\frac{1}{Y}$ le sont aussi. Par suite, d'après le théorème 16, on a :

$$E(Z) = E\left(\frac{X}{Y}\right) = E\left(X.\frac{1}{Y}\right) = E\left(X\right).E\left(\frac{1}{Y}\right)$$

   De plus, on a :
   — $E(X) = \frac{1}{p}$ car $X \sim \mathcal{G}(p)$;
   — Considérons la famille $(\frac{1}{n}P(Y=n))_{n\in Y(\Omega)}$. Comme $Y \sim \mathcal{G}(q)$, $(\frac{1}{n}P(Y=n))_{n\in Y(\Omega)} = (\frac{1}{n}q(1-q)^{n-1})_{n\in\mathbb{N}^*}$ qui est sommable car terme général d'une série absolument convergente (par exemple car, pour tout $n \in \mathbb{N}^*$, $\frac{1}{n}q(1-q)^{n-1} \leq (1-q)^{n-1}$ qui est le terme général d'une série géométrique convergente).

     Par suite, d'après le théorème de transfert (Théorème 10), $\frac{1}{Y}$ est d'espérance finie et on a, en remarquant que, pour tout $x \in]-1,1[$, $\displaystyle\sum_{n=1}^{+\infty}\frac{x^n}{n} = -\ln(1-x)$:

$$\begin{aligned}
E\left(\frac{1}{Y}\right) &= \sum_{n\in\mathbb{N}^*}\frac{1}{n}P(Y=n)\\
&= \sum_{n=1}^{+\infty}\frac{1}{n}q(1-q)^{n-1}\\
&= \frac{q}{1-q}\sum_{n=1}^{+\infty}\frac{1}{n}(1-q)^n\\
&= \frac{q}{1-q}\left(-\ln(1-(1-q))\right)\\
E\left(\frac{1}{Y}\right) &= -\frac{q\ln(q)}{1-q}.
\end{aligned}$$

   Par suite,

$$E(Z) = -\frac{q\ln(q)}{p(1-q)}$$

   Supposons $p = q$. Alors $E(Z) = -\frac{\ln(p)}{1-p}$. La fonction ln est concave et donc sa courbe est en dessous de ses tangentes et en particulier, en dessous de sa tangente en 1 qui est d'équation

$y = x - 1$ d'où, pour tout $x \in \mathbb{R}_+^*$, $\ln(x) \leq x - 1$ avec égalité si, et seulement si $x = 1$. Par suite, comme $p \in]0, 1[$, on obtient :

$$E(Z) = -\frac{\ln(p)}{1-p} > -\frac{p-1}{1-p} = 1.$$

3. Comme X et Y sont à valeurs dans $\mathbb{N}^*$, on a $Z(\Omega) = \mathbb{Q}_+^*$ $(= \mathbb{Q} \cap \mathbb{R}_+^*)$. Maintenant, déterminons $P(Z = x)$ pour $x \in \mathbb{Q}_+^*$.
Soit $x \in \mathbb{Q}_+^*$. Alors il existe $n, m \in \mathbb{N}^*$ avec $n \wedge m = 1$ tels que $x = \frac{n}{m}$ et on remarque, d'après le théorème de Gauss, que $x = \frac{a}{b}$ avec $a, b \in \mathbb{N}^*$ si, et seulement si, il existe $k \in \mathbb{N}^*$ tel que $a = nk$ et $b = mk$. Par suite, on a :

$$(Z = x) = \bigcup_{k \in \mathbb{N}^*} ((X = nk) \cap (Y = nk))$$

Ainsi comme : cette union est disjointe, X et Y sont indépendantes et $(1-p)^n(1-q)^m \in]0, 1[$, on a

$$\begin{aligned}
P(Z = x) &= P\left(\bigcup_{k \in \mathbb{N}^*} ((X = nk) \cap (Y = nk))\right) \\
&= \sum_{k=1}^{+\infty} P((X = nk)).P(Y = mk) \\
&= \sum_{k=1}^{+\infty} p(1-p)^{nk-1} q(1-q)^{mk-1} \\
&= \frac{pq}{(1-p)(1-q)} \sum_{k=1}^{+\infty} ((1-p)^n(1-q)^m)^k \\
&= \frac{pq}{(1-p)(1-q)} \frac{(1-p)^n(1-q)^m}{1-(1-p)^n(1-q)^m} \\
P(Z = x) &= \frac{pq(1-p)^{n-1}(1-q)^{m-1}}{1-(1-p)^n(1-q)^m}
\end{aligned}$$

On peut s'arrêter là, mais si on veut vraiment faire apparaître seulement du "x" dans le résultat (au prix de l'apparition de puissances non entières), on obtient :

$$P(Z = x) = \frac{pq}{(1-p)(1-q)} \cdot \frac{(1-p)^x(1-q)^{\frac{1}{x}}}{1-(1-p)^x(1-q)^{\frac{1}{x}}}.$$

Proposition 30.

Soit (X, Y) un couple de variables aléatoires discrètes tel que $X, Y \in L^2$. Si X et Y sont indépendantes, alors $\mathrm{Cov}(X, Y) = 0$.

Proposition 31. *Variance d'une somme de variables indépendantes*

Soit $n \in \mathbb{N}^*$ et $X_1, ..., X_n$ des variables aléatoires discrètes admettant un moment d'ordre 2. Si $X_1, ..., X_n$ sont deux à deux indépendantes, alors :

$$V(X_1 + \ldots + X_n) = V(X_1) + \ldots + V(X_n).$$

5. Inégalités de concentration

a. Inégalités de Markov et de Bienaymé-Tchebychev

Théorème 17. *Inégalité de Markov*

Soit X une variable aléatoire discrète à valeurs positives. Si $X \in L^1$, alors, pour tout $a > 0$, on a :

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Démonstration.

On suppose $X \in L^1$. Soit $a > 0$. On pose $Y = \mathbb{1}_{(X \geq a)}$ i.e. pour $\omega \in \Omega$,

$$Y(\omega) = \begin{cases} 1 & \text{si } X(\omega) \geq a \\ 0 & \text{si } X(\omega) < a \end{cases}$$

Alors Y est une variable aléatoire discrète à valeurs dans $\{0, 1\}$ et donc d'espérance finie. Par suite, comme $(Y = 1) = \{\omega \in \Omega \mid Y(\omega) = 1\} = \{\omega \in \Omega \mid X(\omega) \geq a\} = (X \geq A)$, on a :

$$E(X) = 0 \times P(Y = 0) + 1 \times P(Y = 1) = P(Y = 1) = P(X \geq a).$$

De plus, on remarque que $X \geq aY$; en effet, pour tout $\omega \in \Omega$:

— si $X(\omega) \geq a$, alors $Y(\omega) = 1$, et donc $X(\omega) \geq aY(\omega)$;
— si $X(\omega) < a$, alors $Y(\omega) = 0$ et comme X est à valeurs positives, $X(\omega) \geq 0 = aY(\omega)$.

Par croissance de l'espérance, on a donc $E(aY) \leq E(X)$ et donc

$$aP(X \geq a) = aE(Y) = E(aY) \leq E(X).$$

D'où le résultat en divisant par $a > 0$. □

Théorème 18. *Inégalité de Bienaymé-Tchebychev*

Soit X une variable aléatoire discrète. Si $X \in L^2$, alors on a, pour tout $\varepsilon > 0$:

$$P\left(|X - E(X)| \geq \varepsilon\right) \leq \frac{V(X)}{\varepsilon^2}.$$

Démonstration.

Soit $\varepsilon > 0$. On pose $Y = (X - E(X))^2$. Alors $Y \in L^1$ et Y est positive, donc, d'après le théorème de Markov,

$$P\left((X - E(X))^2 \geq \varepsilon^2\right) = P(Y \geq \varepsilon^2) \leq \frac{E(Y)}{\varepsilon^2}.$$

Or on a :

— $((X - E(X))^2 \geq \varepsilon^2) = (|X - E(X)| \geq \varepsilon)$, et

— $E(Y) = E((X - E(X))^2) = V(X)$.

Par suite,

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

□

Exercice 17.

Soit X une variable aléatoire discrète admettant un moment d'ordre 2 et telle que $V(X) > 0$. Pour $a \in \mathbb{R}_+^*$, minorer la probabilité $P(E(X) - a\sigma(X) < X < E(X) + a\sigma(X))$ en fonction de a.

En déduire que pour un jeu d'argent à gains entiers, d'espérance nulle et d'écart type égal à 100€, la probabilité de d'obtenir un gain strictement compris entre -200€ et 200€ est d'au moins $\frac{3}{4}$.

Correction.

Soit $n \in \mathbb{N}^*$. On a :

$$(E(X) - a\sigma(X) < X < E(X) + a\sigma(X)) = (|X - E(X)| < a\sigma(X)) = \overline{(|X - E(X)| \geq a\sigma(X))}$$

Comme $X \in L^2$, d'après l'inégalité de Bienaymé-Tchebychev appliquée à X et $\varepsilon = a\sigma(X) > 0$, on a :

$$\begin{aligned} P(E(X) - a\sigma(X) < X < E(X) + a\sigma(X)) &= 1 - P(|X - E(X)| \geq a\sigma(X)) \\ &\geq 1 - \frac{V(X)}{a^2\sigma(X)^2} \\ P(E(X) - a\sigma(X) < X < E(X) + a\sigma(X)) &\geq 1 - \frac{1}{a^2}. \end{aligned}$$

Ce jeu d'argent peut être modélisé par une variable aléatoire aléatoire discrète X à valeurs dans $\mathbb{Z}$, où $X \in L^2$, $E(X) = 0$ et $\sigma(X) = 100$. D'après ce qui précède, on a, avec $a = 2 > 0$:

$$P(X \in]-200, 200[) = P(E(X) - a\sigma(X) < X < E(X) + a\sigma(X)) \geq 1 - \frac{1}{2^2} = \frac{3}{4}.$$

b. Loi faible des grands nombres

Théorème 19. *Loi faible des grands nombres*

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires discrètes, deux à deux indépendantes, de même loi et admettant un moment d'ordre 2. On pose $m = E(X_1)$ et, pour $n \in \mathbb{N}^*$, $S_n = X_1 + \ldots + X_n$. Alors on a, pour tout $\varepsilon > 0$:

$$P\left(\left|\frac{S_n}{n} - m\right| \geq \varepsilon\right) \xrightarrow[n\to+\infty]{} 0.$$

Démonstration.

Soit $n \in \mathbb{N}^*$. On pose $Y = \frac{S_n}{n}$. Comme L^2 est un espace vectoriel, $Y \in L^2$ comme combinaison linéaire d'éléments de L^2. Par suite, $Y \in L^1$ et on a, par linéarité de l'espérance :

$$E(Y) = \frac{1}{n}E(S_n) = \frac{1}{n}\sum_{k=1}^{n}\underbrace{E(X_k)}_{=E(X_1)} = E(X_1) = m.$$

On applique l'inégalité de Bienaymé-Tchebychev à la variable Y pour $\varepsilon > 0$:

$$P\left(\left|\frac{S_n}{n} - m\right| \geq \varepsilon\right) = P(|Y - E(Y)| \geq \varepsilon) \leq \frac{V(Y)}{\varepsilon^2}.$$

Or on a, par indépendance deux à deux des X_k :

$$V(Y) = \frac{1}{n^2}V\left(\sum_{k=1}^{n}X_k\right) = \frac{1}{n}\sum_{k=1}^{n}\underbrace{V(X_k)}_{=V(X_1)} = \frac{V(X_1)}{n},$$

donc :

$$P\left(\left|\frac{S_n}{n} - m\right| \geq \varepsilon\right) \leq \frac{V(X_1)}{n\varepsilon^2} \xrightarrow[n\to+\infty]{} 0.‘$$

□

Partie E

Fonctions génératrices

1. Définition et expression d'une fonction génératrice

Définition 33. *Fonction génératrice*

Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N}$. On appelle **fonction génératrice de** X et on note G_X la fonction :

$$G_X : t \mapsto E(t^X).$$

Théorème 20.

Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N}$. La fonction génératrice G_X de X est bien définie sur $[-1,1]$ et on a, pour tout $t \in [-1,1]$:

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

Démonstration.

Soit $t \in [-1,1]$. On considère la famille $(t^n P(X=n))_{n\in\mathbb{N}}$. Pour tout $n \in \mathbb{N}$, on a :

$$|t^n P(X=n)| = |t|^n P(X=n) \leq P(X=n).$$

Comme $P(X=n)$ est le terme général d'une série convergente (dont la somme vaut 1), par comparaison, $\sum t^n P(X=n)$ converge absolument.
Par suite, $(t^n P(X=n))_{n\in\mathbb{N}}$ est une famille sommable, et ainsi, d'après le théorème de transfert, la variable aléatoire t^X est d'espérance finie. Donc $G_X(t) = E(t^X)$ est bien défini et on a :

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

□

Exemple 5.

Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N}$.

— Si, pour tout entier $n \in \mathbb{N}$, $P(X=n) = \dfrac{2}{3^{n+1}}$, alors, pour tout $t \in [-1,1]$, $G_X(t) = \dfrac{2}{3-t}$.

On vérifie tout d'abord que la formule proposée pour la loi de X est bien une loi de probabilité ! La série $\sum \frac{2}{3^{n+1}}$ est à termes positifs et convergente car géométrique de rapport $\frac{1}{3} \in]-1,1[$.

$$\sum_{n=0}^{+\infty} \frac{2}{3^{n+1}} = \frac{2}{3} \sum_{n=0}^{+\infty} \frac{1}{3^n} = \frac{2}{3} \frac{1}{1-\frac{1}{3}} = 1.$$

Ainsi, la loi proposée en est bien une !

D'après le théorème précédent, G_X est définie sur $[-1,1]$, et on a, pour tout $t \in [-1,1]$,

$$G_X(t) = \sum_{n=0}^{+\infty} t^n \frac{2}{3^{n+1}} = \frac{2}{3} \sum_{n=0}^{+\infty} \left(\frac{t}{3}\right)^n ;$$

or, $|\frac{t}{3}| \leq \frac{1}{3} < 1$ d'où :

$$G_X(t) = \frac{2}{3} \frac{1}{1-\frac{t}{3}} = \frac{2}{3-t}.$$

— S'il existe $N \in \mathbb{N}$ tel que $X(\Omega) \subset [\![0, N]\!]$, G_X est une fonction polynomiale de degré au plus N.

Pour $n \in \mathbb{N}$, on pose $a_n = P(X = n)$. Comme, pour tout entier $n \geq N$, $P(X = n) = 0$, $Q = \sum_{n=0}^{+\infty} a_n Y^n = \sum_{n=0}^{N} a_n Y^n$ est un polynôme de $\mathbb{R}_N[Y]$. D'après le théorème précédent, G_X est définie sur $[-1,1]$, et on a, pour tout $t \in [-1,1]$:

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

Par suite, G_X est une fonction polynomiale de degré au plus N.

Exercice 18.

Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N}$ dont la loi est donnée, pour tout $n \in \mathbb{N}$, par :

$$P(X = n) = \begin{cases} 0 & \text{si } n = 0 \text{ ou } 1 \\ \frac{1}{n(n-1)} & \text{si } n \geq 2 \end{cases}.$$

1. Vérifier que la loi de X en est bien une.
2. Considérons la fonction génératrice G_X de X. Montrer que, pour tout $t \in [-1,1]$, $G_X(t) = (1-t)\ln(1-t) + t$ (prolongée par la valeur 1 en 1).

Correction.

1. La série $\sum_{n\geq 2}\frac{1}{n(n-1)}$ est bien absolument convergente car à termes positifs et telescopique (en effet, $\frac{1}{n(n-1)}=\frac{1}{n-1}-\frac{1}{n}$), donc de même nature que la suite convergente $(\frac{1}{n})_{n\in\mathbb{N}^*}$.
De plus, on a, par télescopage :
$$\sum_{n=2}^{+\infty}\frac{1}{n(n-1)}=\frac{1}{1}-\lim_{n\to+\infty}\frac{1}{n}=1.$$
Ainsi, la loi proposée en est bien une !

2. D'après le théorème précédent, G_X est définie sur $[-1,1]$, et on a, pour tout $t\in[-1,1]$,
$$G_X(t)=\sum_{n=2}^{+\infty}\frac{t^n}{n(n-1)}.$$
On considère alors la série entière $\sum_{n\geq 2}\frac{z^n}{n(n-1)}$ qui de rayon de convergence 1 et dont G_X est la somme sur $]-1,1[$. On a donc, pour tout $t\in]-1,1[$:
$$\begin{aligned}
G_X(t) &= \sum_{n=2}^{+\infty}\int_0^t\frac{x^{n-1}}{n-1}\,\mathrm{d}x\\
&= \int_0^t\left(\sum_{n=2}^{+\infty}\frac{x^{n-1}}{n-1}\right)\mathrm{d}x\\
&= \int_0^t\left(\sum_{n=1}^{+\infty}\frac{x^{n}}{n}\right)\mathrm{d}x\\
&= -\int_0^t\ln(1-x)\,\mathrm{d}x\\
G_X(t) &= (1-t)\ln(1-t)+t
\end{aligned}$$
De plus, comme $\sum_{n\geq 2}\frac{(\pm 1)^n}{n(n-1)}$ convergent, d'après le théorème d'Abel radial, la somme de $\sum_{n\geq 2}\frac{z^n}{n(n-1)}$ est continue sur $[-1,1]$, tout comme la fonction $t\mapsto(1-t)\ln(1-t)+t$ (prolongée par la valeur 1 en 1) ; d'où l'égalité anoncée sur $[-1,1]$.

2. Propriétés des fonctions génératrices

Proposition 32.

Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N}$. Il existe $R\geq 1$ tel que la fonction génératrice G_X de X est développable en série entière sur $]-R,R[$, et si $\sum a_n z^n$ est la série entière associée à ce développement, on a, pour tout $n\in\mathbb{N}$:
$$P(X=n)=a_n.$$

En particulier, G_X est de classe C^∞ sur $]-R,R[$ et pour tout $n \in \mathbb{N}$:

$$P(X=n) = \frac{G_X^{(k)}(0)}{n!}.$$

Démonstration.

Pour $n \in \mathbb{N}$, on note $a_n = P(X=n)$ et R le rayon de convergence de la série entière $\sum a_n z^n$. Comme X est une variable aléatoire à valeurs dans $\mathbb{N}$, $\sum a_n 1^n = \sum P(X=n)$ converge, donc $R \geq 1$.
Soit $t \in]-R,R[$. D'après les propriétés du rayon de convergence d'une série entière, $\sum a_n t^n$ converge absolument et donc $(t^n P(X=n))_{n\in\mathbb{N}}$ est une famille sommable. Par suite, d'après le théorème de transfert, la variable aléatoire t^X est d'espérance finie et on a :

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

Il en résulte que G_X est développable en série entière sur $]-R,R[$ de série entière associée $\sum P(X=n)z^n$. Ainsi, G_X est de classe C^∞ sur $]-R,R[$, et, comme une fonction développable en série entière coïncide avec la somme de sa série de Taylor-Mac Laurin (sur l'intervalle de dévoloppement), on a :

$$P(X=n) = \frac{G_X^{(n)}(0)}{n!}.$$

□

Au delà du théorème 20, on peut même caractériser les fonctions pouvant être vues comme des fonctions génératrices :

Proposition 33.

Soit $f : [-1,1] \to \mathbb{R}$ une fonction. Alors f est la (restriction sur $[-1,1]$ de la) fonction génératrice d'une certaine variable aléatoire discrète X à valeurs dans $\mathbb{N}$ si, et seulement si, les conditions suivantes sont satisfaites :

- $f(1) = 1$,
- f est développable en série entière sur $]-1,1[$ de série entière associée $\sum a_n z^n$,
- La série $\sum a_n$ est convergente et à termes positifs.

Démonstration.

($\Rightarrow$) Si $f = G_X$ où X est une variable aléatoire discrète à valeurs dans $\mathbb{N}$, alors, d'après le théorème 20 et la proposition 32, f est définie sur $[-1,1]$ et développable en série entière sur $]-1,1[$ de série entière associée $\sum a_n z^n$ où $a_n = P(X=n) \geq 0$ pour tout $n \in \mathbb{N}$.
De plus, comme $(X=n)_{n\in\mathbb{N}}$ est un système complet d'événements, $\sum a_n = \sum P(X=n)$

converge et on a :

$$f(1) = G_X(1) = \sum_{n=0}^{+\infty} P(X = n) = 1.$$

Ainsi, toutes les conditions sont satisfaites.

($\Leftarrow$) On suppose les conditions satisfaites. On note $\sum a_n z^n$ la série entière associée au développement en série entière de f sur $]-1,1[$ et on pose $f_n : t \mapsto a_n t^n$. Pour $n \in \mathbb{N}$, on a :

$$\|f_n\|_\infty = \sup_{t\in[-1,1]} |a_n t^n| = a_n$$

donc, comme $\sum a_n$ converge, $\sum f_n$ converge normalement et donc uniformément sur $[-1,1]$. Par suite, les f_n étant continues sur $[-1,1]$, f est continue sur $[-1,1]$ et on a :

$$\sum_{n=0}^{+\infty} a_n = \sum_{n=0}^{+\infty} \lim_{t\to 1^-} (a_n t^n) = \lim_{t\to 1^-} \sum_{n=0}^{+\infty} a_n t^n = \lim_{t\to 1^-} f(t) = f(1) = 1.$$

Ainsi, la suite $(a_n)_{n\in\mathbb{N}}$ est à termes positifs et terme général d'une série convergente de somme 1 et donc, d'après le théorème 2, permet de définir une unique loi de probabilité P sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, définie, pour $n \in \mathbb{N}$, par $P(\{n\}) = a_n$.

On considère, sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}), P)$, la variable aléatoire discrète $X = \mathrm{Id}_{\mathbb{N}} : \mathbb{N} \to \mathbb{N}$ et sa fonction génératrice G_X. D'après le théorème 20, G_X est définie sur $[-1,1]$, et on a, pour tout $t \in [-1,1]$, comme, pour tout $n \in \mathbb{N}$, $(X = n) = \{n\}$:

$$G_X(t) = \sum_{n=0}^{+\infty} t^n \underbrace{P(X = n)}_{=P(\{n\})} = \sum_{n=0}^{+\infty} a_n t^n = f(t).$$

Il en résulte que $f = G_X$ sur $[-1,1]$ et donc que f est bien la restriction sur $[-1,1]$ de la fonction génératrice d'une certaine variable aléatoire discrète à valeurs dans $\mathbb{N}$.

□

Exercice 19.

Les fonctions suivantes sont-elles des fonctions caractéristiques de variables aléatoires discrètes à valeurs dans $\mathbb{N}$?

1. $f : t \mapsto \dfrac{1}{1-t}$ 2. $g : t \mapsto \dfrac{1}{2-t}$ 3. $h : t \mapsto \dfrac{2}{\pi}\arcsin(t)$ 4. $i : t \mapsto \dfrac{2}{1+t^2}$

Si oui, déterminer la loi d'une variable aléatoire associée.

Correction.

1. La fonction f n'est pas définie en 1, elle ne peut donc pas être une fonction caractéristique de variable aléatoire.
2. La fonction g vérifie :
   — g est bien définie sur $[-1,1]$ car $\mathcal{D}_g = \mathbb{R} \setminus \{2\}$ et $g(1) = \frac{1}{2-1} = 1$,

— Pour tout $t \in]-2,2[$, $\frac{t}{2} \in]-1,1[$ d'où

$$g(t) = \frac{1}{2}\frac{1}{1-\frac{t}{2}} = \sum_{n=0}^{+\infty} \frac{1}{2^{n+1}} t^n.$$

donc g est développable en série entière sur $]-2,2[$ et donc sur $]-1,1[$.

— La série $\sum \frac{1}{2^{n+1}}$ est convergente (série géométrique de raison $\frac{1}{2}$) et à termes positifs.

Par suite, g (ou plutôt sa restriction à $]-2,2[$) est une fonction génératrice de variable aléatoire discrète à valeurs dans $\mathbb{N}$ et si X est une telle variable, on a, pour tout $n \in \mathbb{N}$

$$P(X=n) = \frac{1}{2^{n+1}} = \frac{1}{2}(1-\frac{1}{2})^n$$

On remarque alors que $Y = X+1$ suit une loi géométrique de paramètre $\frac{1}{2}$.

3. La fonction h vérifie :

— h est bien définie sur $\mathcal{D}_h = [-1,1]$ et $h(1) = \frac{2}{\pi}\arcsin(1) = 1$,

— arcsin est développable en série entière donc h l'est aussi, et on a, pour tout $t \in]-1,1[$,

$$h(t) = \frac{2}{\pi}\sum_{n=0}^{+\infty} \frac{\binom{2n}{n}}{(2n+1)4^n} t^{2n+1}.$$

donc h est développable en série entière sur $]-1,1[$.

On note $(a_n)_{n\in\mathbb{N}}$ la suite associée à la série entière du développement de h. Alors, pour tout $n \in \mathbb{N}$:

$$a_n = \begin{cases} 0 & \text{si } n \text{ est pair} \\ \frac{2}{\pi}\frac{\binom{2k}{k}}{(2k+1)4^k} & \text{si } n = 2k+1 \text{ est pair} \end{cases}$$

— La série $\sum a_n$ est convergente car de même nature que la série convergente $\sum a_{2n+1}$ (utiliser par exemple la formule de Stirling ou voire le problème 2 "Calcul de $\zeta(2)$" dans le chapitre XII sur les suites et séries d'intégrales, pour une démonstration plus "économique") et $a_n \geq 0$ pour tout $n \in \mathbb{N}$.

Par suite, h est une fonction génératrice de variable aléatoire discrète à valeurs dans $\mathbb{N}$ et si X est une telle variable, on a, pour tout $n \in \mathbb{N}$

$$P(X=2n) = 0 \quad \text{et } P(X=2n+1) = \frac{2}{\pi}\frac{\binom{2n}{n}}{(2n+1)4^n}.$$

4. La fonction i vérifie :

— i est bien définie sur $[-1,1]$ car $\mathcal{D}_i = \mathbb{R}$ et $i(1) = \frac{1}{1+1^2} = 1$,

— Pour tout $t \in]-1,1[$, $-t^2 \in]-1,0[\subset]-1,1[$ d'où

$$i(t) = 2\frac{1}{1-(-t^2)} = \sum_{n=0}^{+\infty} 2(-1)^n t^{2n}.$$

donc i est développable en série entière sur $]-1,1[$.

— $2(-1)^n$ n'est pas positif en général !

Donc i n'est pas fonction génératrice de variable aléatoire.

Proposition 34.

Soit X, Y des variables aléatoires discrètes à valeurs dans $\mathbb{N}$.

— La loi de probabilité P_X de X est entièrement déterminée par sa fonction génératrice G_X.

— Les variables X et Y ont la même loi si, et seulement si, $G_X = G_Y$.

Remarque 15.

La proposition précédente permet donc de parler de la fonction génératrice associée à une loi de probabilité discrète sur $\mathbb{N}$, du fait que toutes les variables aléatoires discrètes à valeurs dans $\mathbb{N}$ suivant une loi donnée partagent la même fonction génératrice.

Proposition 35.

Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N}$ et G_X sa fonction génératrice. On a :

— La variable X appartient à L^1 si, et seulement si, G_X est dérivable en 1. Dans ce cas, on a :
$$G'_X(1) = E(X).$$

— La variable X appartient à L^2 si, et seulement si, G_X est deux fois dérivable en 1. Dans ce cas, on a :
$$G''_X(1) = E(X(X-1)).$$

Démonstration.

— ($\Rightarrow$) On suppose $X \in L^1$. Alors $\sum nP(X=n)$ converge et donc la série entière $\sum nP(X=n)z^{n-1}$ converge normalement sur $[-1,1]$. Par suite, d'après le théorème d'interversion limite/somme, on a :
$$G'_X(t) = \sum_{n=0}^{+\infty} nP(X=n)t^{n-1} \xrightarrow[t\to 1^-]{} \sum_{n=0}^{+\infty} nP(X=n) = E(X).$$
Il en résulte, d'après le théorème de la limite de la dérivée, que G_X est dérivable en 1 et que $G'_X(1) = E(X)$.

($\Leftarrow$) Raisonnons par contraposée. On suppose que $X \notin L^1$. La fonction G'_X est la somme d'une série de fonctions croissantes sur $[0,1[$ donc G'_X est croissante sur $[0,1[$. Ainsi, d'après le théorème de la limite montone, G'_X admet une limite ℓ en 1^- (potentiellement $+\infty$). Supposons $\ell < +\infty$. Alors, pour tout $n \in \mathbb{N}$ et tout $t \in [0,1[$, on a, la somme étant à termes positifs :
$$\sum_{k=0}^{n} kP(X=k)t^{k-1} \leq \sum_{k=0}^{+\infty} kP(X=k)t^{k-1} = G'_X(t)$$
d'où, en passant à la limite quand $t \to 1^-$:
$$\sum_{k=0}^{n} kP(X=k) \leq \ell.$$

Or X est à valeurs positives et donc $E(X) = +\infty$ i.e. $\sum kP(X=k)$ diverge vers $+\infty$. Par suite, $\ell \geq \sum_{k=0}^{n} kP(X=k) \xrightarrow[n\to+\infty]{}$. Contradiction !
Ainsi, $\ell = +\infty$ et donc G_X n'est pas dérivable en 1 d'après le théorème de la limite de la dérivée.

— ($\Rightarrow$) On suppose $X \in L^2$. Alors $X(X-1) = X^2 - X \in L^1$ car $L^2 \subset L^1$ et L^1 est un espace vectoriel. Ainsi $\sum n(n-1)P(X=n)$ converge et donc la série entière $\sum n(n-1)P(X=n)z^{n-2}$ converge normalement sur $[-1,1]$. Par suite, d'après le théorème d'interversion limite/somme, on a :

$$G''_X(t) = \sum_{n=0}^{+\infty} n(n-1)P(X=n)t^{n-2} \xrightarrow[t\to 1^-]{} \sum_{n=0}^{+\infty} n(n-1)P(X=n) = E(X(X-1)).$$

Il en résulte, d'après le théorème de la limite de la dérivée, que G'_X est dérivable en 1 et que $G''_X(1) = E(X(X-1))$.

($\Leftarrow$) Raisonnons par contraposée. On suppose que $X \notin L^2$. La fonction G''_X est la somme d'une série de fonctions croissantes sur $[0,1[$ donc G''_X est croissante sur $[0,1[$. Ainsi, d'après le théorème de la limite montone, G''_X admet une limite ℓ en 1^- (potentiellement $+\infty$). Supposons $\ell < +\infty$. Alors, pour tout $n \in \mathbb{N}$ et tout $t \in [0,1[$, on a, la somme étant à termes positifs :

$$\sum_{k=0}^{n} k(k-1)P(X=k)t^{k-2} \leq \sum_{k=0}^{+\infty} k(k-1)P(X=k)t^{k-2} = G''_X(t)$$

d'où, en passant à la limite quand $t \to 1^-$:

$$\sum_{k=0}^{n} k(k-1)P(X=k) \leq \ell.$$

De plus, comme $X \notin L^2$, alors $X(X-1) \notin L^1$ (en effet, pour Y une variable aléatoire discrète, si $Y(Y-1) \in L^1$, alors $Y^2 = Y(Y-1) + Y \in L^1$).
Or $X(X-1)$ est à valeurs positives et donc $E(X(X-1)) = +\infty$ i.e. $\sum k(k-1)P(X=k)$ diverge vers $+\infty$. Par suite, $\ell \geq \sum_{k=0}^{n} k(k-1)P(X=k) \xrightarrow[n\to+\infty]{}$. Contradiction !
Ainsi, $\ell = +\infty$ et donc G'_X n'est pas dérivable en 1 d'après le théorème de la limite de la dérivée.

□

Exercice 20.

Soit $X \in L^2$ une variable aléatoire discrète à valeurs dans $\mathbb{N}$ et G_X sa fonction génératrice. Exprimer $E(X^2)$ et $V(X)$ en fonction des dérivées de G_X en 1.

On a la généralisation suivante de la proposition précédente :

Exercice 21.

Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N}$ et G_X sa fonction génératrice. Montrer que, pour tout $k \in \mathbb{N}^*$, la variable X appartient à L^k si, et seulement si, G est k-fois dérivable en 1 et que dans ce cas, on a :

$$G_X^{(k)}(1) = E(X(X-1)...(X-k+1)).$$

Correction.

D'après la proposition 32, G_X est développable en série entière sur $]-1,1[$ et donc de classe C^∞ sur $]-1,1[$, et on a, pour tout $t \in]-1,1[$ et tout $k \in \mathbb{N}$:

$$G_X^{(k)}(t) = \sum_{n=0}^{+\infty} n(n-1)...(n-k+1)P(X=n)t^{n-k}.$$

Soit $k \in \mathbb{N}^*$.

($\Rightarrow$) On suppose $X \in L^k$. Alors, d'après l'exercice 14, $X(X-1)...(X-k+1) \in L^1$. Par suite, $\sum n(n-1)...(n-k+1)P(X=n)$ converge et donc la série entière $\sum n(n-1)...(n-k+1)P(X=n)z^{n-k}$ converge normalement sur $[-1,1]$.
Par suite, d'après le théorème d'interversion limite/somme et le théorème de transfert, on a :

$$\begin{aligned} G_X^{(k)}(t) &= \sum_{n=0}^{+\infty} n(n-1)...(n-k+1)P(X=n)t^{n-k} \\ &\xrightarrow[t\to 1^-]{} \sum_{n=0}^{+\infty} n(n-1)...(n-k+1)P(X=n) = E(X(X-1)...(X-k+1)). \end{aligned}$$

Il en résulte, d'après le théorème de la limite de la dérivée, que $G_X^{(k-1)}$ est dérivable en 1 et que $G_X^{(k)}(1) = E(X(X-1)...(X-k+1))$.

($\Leftarrow$) Raisonnons par contraposée. On suppose que $X \notin L^k$. Alors, d'après l'exercice 14, $X(X-1)...(X-k+1) \notin L^1$.
La fonction $G_X^{(k)}$ est la somme d'une série de fonctions croissantes sur $[0,1[$ donc $G_X^{(k)}$ est croissante sur $[0,1[$. Ainsi, d'après le théorème de la limite montone, $G_X^{(k)}$ admet une limite ℓ en 1^- (potentiellement $+\infty$). Supposons $\ell < +\infty$. Alors, pour tout $N \in \mathbb{N}$ et tout $t \in [0,1[$, on a, la somme étant à termes positifs :

$$\sum_{n=0}^{N} n(n-1)...(n-k+1)P(X=n)t^{n-k} \leq \sum_{n=0}^{+\infty} n(n-1)...(n-k+1)P(X=n)t^{n-k} = G_X^{(k)}(t)$$

d'où, en passant à la limite quand $t \to 1^-$:

$$\sum_{n=0}^{N} n(n-1)...(n-k+1)P(X=n) \leq \ell$$

Or, comme X est à valeurs dans $\mathbb{N}$, $X(X-1)...(X-k+1)$ est à valeurs positives et donc $E(X(X-1)...(X-k+1)) = +\infty$ i.e. $\sum n(n-1)...(n-k+1)P(X=n)$ diverge vers $+\infty$. Par suite, $\ell \geq \sum_{n=0}^{N} n(n-1)...(n-k+1)P(X=n) \xrightarrow[N\to+\infty]{} +\infty$. Contradiction !
Ainsi, $\ell = +\infty$ et donc $G_X^{(k-1)}$ n'est pas dérivable en 1 d'après le théorème de la limite de la dérivée. D'où G n'est pas k-fois dérivable en 1.

3. Fonctions génératrices des lois usuelles

Exemple 6.

Soit X une variable aléatoire discrète.

— Si $X \sim \mathcal{G}(p)$ alors $G_X(t) = \dfrac{pt}{1-(1-p)t}$,

On suppose $X \sim \mathcal{G}(p)$ avec $p \in]0,1[$. Alors :

$$\begin{cases} X(\Omega) = \mathbb{N}^* \\ P(X=k) = p(1-p)^{k-1} \text{ pour tout } k \in \mathbb{N}^* \end{cases}$$

D'après le théorème 20, G_X est définie sur $[-1,1]$ mais on remarque, en posant $a_n = P(X=n) > 0$ pour $n \geq 1$, que $\sum_{n\geq 1} a_n t^n$ est de rayon de convergence $\frac{1}{1-p}$. En effet, comme $|\frac{a_{n+1}}{a_n}| = (1-p) \xrightarrow[n\to+\infty]{} (1-p)$, d'après la règle de D'Alembert pour les séries entières, $R = \frac{1}{1-p}$.
Ainsi, si $X \sim \mathcal{G}(p)$, G_X est définie sur $]-\frac{1}{1-p}, \frac{1}{1-p}[$ (intervalle plus gros que $[-1,1]$ car $1-p<1$). De plus, pour tout $t \in]-\frac{1}{1-p}, \frac{1}{1-p}[$, on a :

$$\begin{aligned} G_X(t) &= \sum_{n=1}^{+\infty} t^n P(X=n) \\ &= \sum_{n=1}^{+\infty} t^n p(1-p)^{n-1} \\ &= pt \sum_{n=1}^{+\infty} (t(1-p))^{n-1} \\ &= pt \sum_{n=0}^{+\infty} ((1-p)t)^n \text{ et } t(1-p) \in]-1,1[\\ &= pt \frac{1}{1-(1-p)t} \\ G_X(t) &= \frac{pt}{1-(1-p)t} \end{aligned}$$

— Si $X \sim \mathcal{P}(\lambda)$ alors $G_X(t) = e^{\lambda(t-1)}$.

On suppose $X \sim \mathcal{P}(\lambda)$ avec $\lambda \in \mathbb{R}_+^*$. Alors :

$$\begin{cases} X(\Omega) = \mathbb{N} \\ P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \text{ pour tout } k \in \mathbb{N} \end{cases}$$

D'après le théorème 20, G_X est définie sur $[-1, 1]$ mais on remarque, en posant $a_n = P(X = n) > 0$ pour $n \in \mathbb{N}$, que $\sum a_n t^n$ est de rayon de convergence $+\infty$. En effet, comme $|\frac{a_{n+1}}{a_n}| = \frac{\lambda}{n+1} \xrightarrow[n\to+\infty]{} 0$, d'après la règle de D'Alembert pour les séries entières, $R = +\infty$.
Ainsi, si $X \sim \mathcal{P}(\lambda)$, G_X est définie sur $\mathbb{R}$. De plus, pour tout $t \in \mathbb{R}$, on a :

$$\begin{aligned} G_X(t) &= \sum_{n=1}^{+\infty} t^n P(X = n) \\ &= \sum_{n=0}^{+\infty} t^n e^{-\lambda}\frac{\lambda^n}{n!} \\ &= e^{-\lambda}\sum_{n=0}^{+\infty}\frac{(t\lambda)^n}{n!} \\ &= e^{-\lambda}e^{t\lambda} \\ G_X(t) &= e^{\lambda(t-1)}. \end{aligned}$$

Exercice 22.

Déterminer les fonctions génératrices des lois uniforme sur $[\![1, n]\!]$, de Bernoulli et binomiale.

4. Fonction génératrice d'une somme de variables indépendantes

Théorème 21.

Soit X, Y des variables aléatoires discrètes à valeurs dans $\mathbb{N}$. Si X et Y sont indépendantes, alors, pour tout $t \in [-1, 1]$:

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

Démonstration.

On suppose X et Y indépendantes.

On propose deux preuves : l'une reposant sur la formule de la loi d'un couple de variable indépendantes et un produit de Cauchy puis l'autre utilisant la formule : l'espérance d'un produit vaut le produit des espérances pour des variables indépendantes.

- Première preuve. Soit $t \in [-1, 1]$. On note $Z = X + Y$; alors Z est une variable aléatoire

discrète à valeurs dans $\mathbb{N}$ et on a :

$$G_Z(t) = E(t^Z) = \sum_{n=0}^{+\infty} t^n P(Z=n) = \sum_{n=0}^{+\infty} t^n P(X+Y=n)$$

Or, X et Y étant indépendantes, on a, pour tout $n \in \mathbb{N}$:

$$P(X+Y=n) = P\left(\bigcup_{k=0}^{n} ((X=k) \cap (y=n-k))\right) = \sum_{k=0}^{n} P(X=k)P(Y=n-k).$$

Ainsi :

$$G_Z(t) = \sum_{n=0}^{+\infty} t^n \sum_{k=0}^{n} P(X=k)P(Y=n-k) = \sum_{n=0}^{+\infty}\sum_{k=0}^{n} \underbrace{t^k P(X=k)}_{=a_k} \underbrace{t^{n-k} P(Y=n-k)}_{=b_{n-k}}.$$

où $a_n = t^n P(X=n)$ et $b_n = t^n P(Y=n)$ pour $n \in \mathbb{N}$.
On remarque que $G_X(t) = \sum_{n=0}^{+\infty} a_n$ et $G_Y(t) = \sum_{n=0}^{+\infty} b_n$ et que $\sum a_n, \sum b_n$ convergent absolument (car pour tout $n \in \mathbb{N}$, $|a_n| \leq P(X=n)$). Ainsi, le produit de Cauchy $\sum c_n$ de ces séries converge où $c_n = \sum_{k=0}^{n} a_n b_n$ et on a l'égalité :

$$G_{X+Y}(t) = G_Z(t) = \sum_{n=0}^{+\infty} c_n = \left(\sum_{n=0}^{+\infty} a_n\right)\left(\sum_{n=0}^{+\infty} b_n\right) = G_X(t)G_Y(t).$$

- Deuxième preuve. Soit $t \in [-1,1]$. On note $f : \mathbb{N} \to \mathbb{R}$ la fonction $n \mapsto t^n$. Comme X et Y sont à valeurs dans $\mathbb{N}$ et indépendantes, $t^X = f(X)$ et $t^Y = f(Y)$ sont bien définies et indépendantes, donc, d'après le théorème 16, on a :

$$G_{X+Y}(t) = E\left(t^{X+Y}\right) = E\left(t^X t^Y\right) = E(t^X)E(t^Y) = G_X(t)G_Y(t).$$

En effet, on a bien $t^{X+Y} = t^X t^Y$ car, pour tout $\omega \in \Omega$:

$$\begin{aligned} t^{X+Y}(\omega) &= f \circ (X+Y)(\omega) \\ &= f(X(\omega)+Y(\omega)) \\ &= t^{X(\omega)+Y(\omega)} \\ &= t^{X(\omega)}t^{Y(\omega)} = f(X(\omega))f(Y(\omega)) = (f(X)f(Y))(\omega) = \left(t^X t^Y\right)(\omega). \end{aligned}$$

□

Remarque 16.

Plus généralement, si $X_1, ..., X_n$ sont des variables aléatoires discrètes indépendantes, alors $G_{X_1+...+X_n} = G_{X_1} \times \ldots \times G_{X_n}$.

Exercice 23.

Soit $k, m, n \in \mathbb{N}^*$ tels que $k = nm$ et X, Y des variables aléatoires à valeurs dans $\mathbb{N}$ indépendantes. On pose $Z = X + Y$ et on suppose que $X \sim \mathcal{U}(\llbracket 0, n-1 \rrbracket)$ et $Z \sim \mathcal{U}(\llbracket 0, k-1 \rrbracket)$.

1. Déterminer la fonction génératrice de Y.
2. En déduire la loi de Y.