INTRODUCTION À LA THÉORIE DES PROBABILITÉS

FRANÇOIS CHAPON

Université de Toulouse

2024–2025

Table des matières

Introduction

La théorie des probabilités est une branche des mathématiques permettant de modéliser les expériences qui sont soumises au hasard. Dans ce cours, on ne s'intéressera qu'aux expériences aléatoires prenant un nombre dénombrable de valeurs. Le cadre général de la théorie des probabilités sera étudié l'année prochaine, une fois le cours de théorie de la mesure suivi.

1. Espace de probabilité

1.1. **Définition.** Dans ce cours, un espace de probabilité sera la donnée de :

- Un ensemble Ω, appelé l'univers, représentant l'ensemble de tous les résultats possibles d'une expérience aléatoire. Il sera supposé **dénombrable**, c'est-à-dire qu'il peut s'injecter dans $\mathbb{N}$ (il est donc fini ou infini dénombrable) ;
- L'ensemble $\mathcal{P}(\Omega)$ des sous-ensembles de Ω, appelée la tribu des événements. Il représente l'ensemble des événements dont on veut mesurer la probabilité ;
- Un application $\mathbb{P}$ sur les événements qui à un événement A lui associe sa probabilité de réalisation $\mathbb{P}(A)$. Plus précisément :

Définition 1.1. *Une (mesure de) probabilité sur* $(\Omega, \mathcal{P}(\Omega))$ *est une application de* $\mathcal{P}(\Omega)$ *dans* $[0,1]$ *vérifiant :*

- $\mathbb{P}(\Omega) = 1$ *;*
- σ*-additivité : pour toute suite* $(A_n)_{n\geq 0}$ *d'éléments de* $\mathcal{P}(\Omega)$ *deux à deux disjoints, c'est-à-dire que* $A_i \cap A_j = \varnothing$ *pour tous* $i \neq j$*, on a*

$$\mathbb{P}\left(\bigcup_{n\geq 0} A_n\right) = \sum_{n\geq 0} \mathbb{P}(A_n).$$

Les éléments de $\mathcal{P}(\Omega)$ *sont appelés des événements, et pour un événement* A*,* $\mathbb{P}(A)$ *est appelée la probabilité de réalisation de l'événement* A*.*

Dans toute la suite, $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ désignera un espace de probabilité. On ne fera plus forcément référence à la tribu des événements $\mathcal{P}(\Omega)$, et donc on dira que $(\Omega, \mathbb{P})$ est un espace de probabilité. Notons que la propriété de σ-additivité est assez naturelle : on veut que la probabilité de réalisation de deux événements incompatibles, c'est-à-dire disjoints, soit la somme de leurs probabilités de réalisations. La σ-additivité exprime exactement la même chose pour une suite d'événements.

Remarque 1.1. On prendra garde qu'on utilise le même mot "probabilité" pour désigner deux choses différentes : une probabilité est une application sur $\mathcal{P}(\Omega)$ vérifiant les deux propriétés ci-dessus, mais on parle aussi de la probabilité d'un événement (qui désigne ici le sens courant de probabilité).

Exemple 1.1. On lance un dé à 6 faces non truqué. On peut modéliser cette expérience en prenant l'univers $\Omega = \{1,2,3,4,5,6\}$, et on s'intéresse à des événements du type « le résultat du dé est 6 », ou encore « le résultat du dé est impair », c'est-à-dire des éléments de $\mathcal{P}(\Omega)$. La probabilité modélisant l'expérience est la probabilité uniforme (ou équiprobabilité) : chacun des événements élémentaires $\{\omega\}$, avec $\omega \in \Omega$, a même probabilité de réalisation, à savoir :

$$\mathbb{P}(\{\omega\}) = \frac{1}{6}, \quad \text{pour tout } \omega \in \Omega.$$

On a donc que pour tout événement A,

$$\mathbb{P}(A) = \frac{|A|}{6},$$

où $|A|$ désigne le cardinal de A. Si A est l'événement « le résultat du dé est impair », alors $A = \{1, 3, 5\}$, et donc

$$\mathbb{P}(A) = \frac{3}{6} = \frac{1}{2}.$$

En général, on appelle probabilité uniforme sur un ensemble fini Ω, la probabilité définie par,

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

pour tout $A \in \mathcal{P}(\Omega)$. C'est exactement la probabilité modélisant l'équiprobabilité :

$$\frac{\text{nombre de cas possibles}}{\text{nombre de cas total}}$$

1.2. **Vocabulaire.** Avant de donner quelques propriétés vérifiées par une probabilité, traduisons en language probabiliste les opérations ensemblistes usuelles.

Notation	Vocabulaire ensembliste	Vocabulaire probabiliste
Ω	ensemble plein	univers ou événement certain
$\varnothing$	ensemble vide	événement impossible
$A \in \mathcal{F}$	sous-ensemble	événement
A^c	complémentaire de A	événement contraire de A
$A \subset B$	A inclus dans B	A entraîne B
$A \cup B$	réunion de A et B	A ou B
$A \cap B$	intersection de A et B	A et B
$A \cap B = \varnothing$	A et B disjoints	A et B incompatibles
$A \setminus B = A \cap B^c$	différence ensembliste de A et B	A privé de B

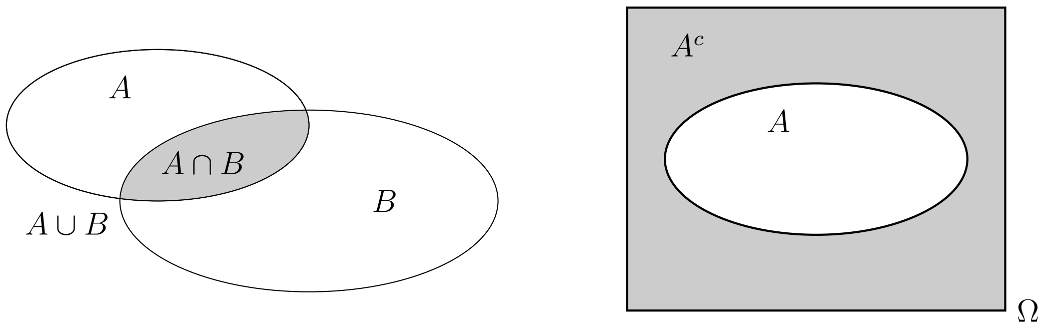

FIGURE 1.1. Représentation ensembliste de quelques événements.

1.3. **Propriétés.** Commençons par les propriétés élémentaires vérifiées par une probabilité :

Proposition 1.1. *Soient A et B des événements. Alors,*

(i) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ *;*

(ii) $\mathbb{P}(\varnothing) = 0$ *;*

(iii) $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$*. En particulier, si $A \subset B$, on a* $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$ *;*

(iv) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ *;*

(v) Si $A \subset B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Démonstration. On pourra se reporter à la figure 1.2. Comme $\Omega = A \cup A^c$, l'union étant bien entendu disjointe, on a

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c),$$

et les points (i) et (ii) en découlent.

Pour (iii) : il suffit d'écrire B comme l'union disjointe $B = (B \setminus A) \cup (A \cap B)$, ce qui donne

$$\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B).$$

En particulier, si $A \subset B$, alors $A \cap B = A$, et donc

$$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A).$$

Pour (iv) : on écrit $A \cup B$ comme l'union disjointe $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$, d'où

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B),$$

et on utilise le point (iii).

Pour (v) : par le point (iii), si $A \subset B$, alors

$$\mathbb{P}(B) - \mathbb{P}(A) = \mathbb{P}(B \setminus A) \geq 0,$$

d'où $\mathbb{P}(A) \leq \mathbb{P}(B)$. □

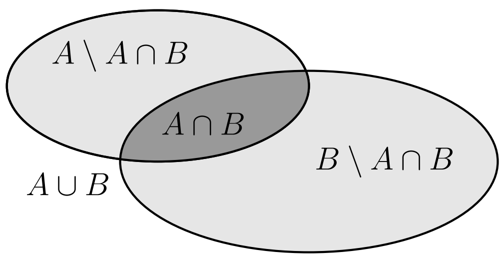

FIGURE 1.2.

Remarque 1.2. On pourra remarquer que si $\mathbb{P}(B) = 1$, alors pour tout événement A, $\mathbb{P}(A \cup B) = 1$ (car $B \subset A \cup B$ et $\mathbb{P}$ est croissante), et donc $\mathbb{P}(A) = \mathbb{P}(A \cap B)$.

De même, si B est de probabilité nulle, i.e. $\mathbb{P}(B) = 0$, alors pour tout événement A, $\mathbb{P}(A \cap B) = 0$ (car $A \cap B \subset B$ et $\mathbb{P}$ est croissante), et donc $\mathbb{P}(A) = \mathbb{P}(A \cup B)$.

Passons à quelques propriétés moins élémentaires.

Proposition 1.2 (Propriété de sous-additivité finie)**.** *Soit $A_1, \ldots, A_n$ des événements. Alors,*

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

Démonstration. Par récurrence sur n. C'est vrai à $n=2$ par la proposition précédente : comme
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B),$$
on a bien $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$. Supposons l'inégalité vraie à n, et soit $A_1, \dots, A_{n+1}$ des événements. Alors,
$$\mathbb{P}\left(\bigcup_{k=0}^{n+1} A_k\right) = \mathbb{P}\left(\bigcup_{k=0}^{n} A_k \cup A_{n+1}\right) \leq \mathbb{P}\left(\bigcup_{k=0}^{n} A_k\right) + \mathbb{P}(A_{n+1}),$$
en utilisant la propriété pour deux événements. Par hypothèse de récurrence, $\mathbb{P}\left(\bigcup_{k=0}^{n} A_k\right) \leq \sum_{k=0}^{n} \mathbb{P}(A_k)$, on obtient donc
$$\mathbb{P}\left(\bigcup_{k=0}^{n+1} A_k\right) \leq \sum_{k=0}^{n+1} \mathbb{P}(A_k).$$
□

Proposition 1.3. *Si $(A_n)_n$ est une suite croissante d'événements (i.e. $A_n \subset A_{n+1}$ pour tout n), alors*
$$\mathbb{P}\left(\bigcup_{n} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Démonstration. Soit $(A_n)_n$ est une suite croissante d'événements. Posons : $B_0 = A_0$, et pour tout $n \geq 1$, $B_n = A_n \setminus A_{n-1}$. Alors, les événements B_n sont deux à deux disjoints, on a $A_n = \bigcup_{k=0}^{n} B_k$, et
$$\bigcup_n B_n = \bigcup_n A_n.$$
On obtient donc, par σ-additivité,
$$\mathbb{P}\left(\bigcup_{n} A_n\right) = \mathbb{P}\left(\bigcup_{n} B_n\right) = \sum_n \mathbb{P}(B_n) = \lim_{n\to\infty} \sum_{k=0}^{n} \mathbb{P}(B_k) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=0}^{n} B_k\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$
□

Corollaire 1.1 (Propriété de sous-σ-additivité)**.** *Soit $(A_n)_n$ une suite d'événements. Alors,*
$$\mathbb{P}\left(\bigcup_{n} A_n\right) \leq \sum_n \mathbb{P}(A_n).$$

Démonstration. On pose $B_n = \bigcup_{k=0}^{n} A_k$. La suite $(B_n)_n$ est alors croissante, et $\bigcup_n B_n = \bigcup_n A_n$. On a donc, en utilisant la proposition précédente,
$$\mathbb{P}\left(\bigcup_{n} A_n\right) = \mathbb{P}\left(\bigcup_{n} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=0}^{n} A_k\right) \leq \lim_{n\to\infty} \sum_{k=0}^{n} \mathbb{P}(A_k) = \sum_n \mathbb{P}(A_n),$$
où l'inégalité provient de la sous-additivité finie. □

Proposition 1.4. *Si $(A_n)_n$ est une suite décroissante d'événements (i.e. $A_{n+1} \subset A_n$ pour tout n), alors*
$$\mathbb{P}\left(\bigcap_{n} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Démonstration. Il suffit d'appliquer la proposition précédente à la suite croissante des complémentaires $(A_n^c)_n$, et d'utiliser que $\mathbb{P}(A_n^c) = 1 - \mathbb{P}(A_n)$. □

1.4. **Caractérisation des probabilités sur un espaces dénombrable.** Dans notre cadre d'un espace Ω dénombrable, une probabilité sur Ω n'est rien d'autre qu'une suite de nombres dans $[0,1]$ de somme égale 1. En effet :

Proposition 1.5. *Soit Ω un ensemble dénombrable. Soit $(p(\omega))_{\omega\in\Omega}$ une suite de nombres telle que pour tout $\omega \in \Omega$, $p(\omega) \in [0,1]$ et telle que $\sum_{w\in\Omega} p(\omega) = 1$. Alors, l'application $\mathbb{P}$ définie par*

$$\mathbb{P}(A) = \sum_{\omega\in A} p(\omega), \quad \text{pour } A \in \mathcal{P}(\Omega),$$

est une probabilité sur $(\Omega, \mathcal{P}(\Omega))$. Réciproquement, toute probabilité sur $(\Omega, \mathcal{P}(\Omega))$ s'écrit de la sorte.

Démonstration. Vérifions les axiomes d'une probabilité. On a clairement $\mathbb{P}(\Omega) = \sum_{w\in\Omega} p(\omega) = 1$. Soit $(A_n)_n$ une suite d'éléments de $\mathcal{P}(\Omega)$, deux à deux disjoints. Alors,

$$\mathbb{P}\left(\bigcup_n A_n\right) = \sum_{\omega\in\bigcup_n A_n} p(\omega) = \sum_n \sum_{\omega\in A_n} p(\omega) = \sum_n \mathbb{P}(A_n),$$

et donc $\mathbb{P}$ satisfait la propriété de σ-additivité.

Réciproquement, si $\mathbb{P}$ est une probabilité, alors les nombres $p(\omega) = \mathbb{P}(\{\omega\})$ sont dans $[0,1]$ et de somme égale à 1. ☐

Exemple 1.2. On note δ_x la masse de Dirac au point x. C'est la probabilité définie par :

$$\delta_x(A) = \begin{cases} 1 & \text{si } x \in A, \\ 0 & \text{sinon.} \end{cases}$$

D'après la proposition précédente, une probabilité sur un ensemble dénombrable Ω peut toujours s'écrire sous la forme :

$$\mathbb{P} = \sum_{x\in\Omega} p_x \delta_x,$$

avec $p_x = \mathbb{P}(\{x\})$.

1.5. **Probabilités conditionnelles.** Lorsqu'on connaît une information supplémentaire, la probabilité de réalisation d'un événement peut s'en trouver modifiée. On introduit alors le concept de probabilité conditionnelle. Commençons par exemple dans un cadre équiprobable.

Exemple 1.3. On considère une population de 200 coccinelles, composée de 110 coccinelles femelles et de 90 coccinelles mâles. Les coccinelles sont de deux types différents : rouge et jaune, et parmi les 110 coccinelles femelles, il y en a 60 qui sont de couleurs rouges. La proportion de coccinelles rouges parmi les coccinelles femelles est donc évidemment

$$\frac{\text{nombre de coccinelles femelles et rouge}}{\text{nombre de coccinelles femelles}} = \frac{60}{110} = \frac{60/200}{110/200}$$

On définit alors naturellement la notion de probabilité conditionnelle :

Définition 1.2. *Soit B un événement tel que $\mathbb{P}(B) > 0$. Soit A un événement. La probabilité conditionnelle de A sachant B, notée $\mathbb{P}(A|B)$ est définie par :*

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

Proposition 1.6. *Soit B un événement tel que $\mathbb{P}(B) > 0$. Alors, l'application $\mathbb{P}(\,\cdot\,|B)\colon A \mapsto \mathbb{P}(A|B)$ est une probabilité.*

Démonstration. On a facilement $\mathbb{P}(\Omega) = 1$ (on a en fait $\mathbb{P}(B) = 1$). Montrons la σ-additivité. Soit $(A_n)_n$ une suite d'événements disjoints deux à deux. Comme,

$$\left(\bigcup_n A_n\right) \cap B = \bigcup_n (A_n \cap B),$$

l'union étant disjointe, on a alors,

$$\mathbb{P}\left(\bigcup_n A_n \,\Big|\, B\right) = \frac{\mathbb{P}\left(\bigcup_n A_n \cap B\right)}{\mathbb{P}(B)} = \sum_n \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_n \mathbb{P}(A_n|B). \qquad \square$$

Définition 1.3. *Soit $(A_n)_n$ une suite de sous-ensembles de Ω. On dit que $(A_n)_n$ est une partition de Ω si les sous-ensembles A_n sont deux à deux disjoints, et de réunion égale à Ω :*

$$i \neq j \Rightarrow A_i \cap A_i = \varnothing, \quad \text{et} \quad \Omega = \bigcup_n A_n.$$

En language probabiliste, on dit que $(A_n)_n$ forme un système complet d'événements.

On a alors :

Proposition 1.7 (Formule des probabilités totales)**.** *Soit $(A_n)_n$ une suite d'événements formant une partition de Ω, et B un événement. Alors,*

$$\mathbb{P}(B) = \sum_n \mathbb{P}(B|A_n)\,\mathbb{P}(A_n).$$

Démonstration. On a

$$\mathbb{P}(B) = \mathbb{P}\left(B \cap \bigcup_n A_n\right),$$

car $\mathbb{P}\left(\bigcup_n A_n\right) = \mathbb{P}(\Omega) = 1$, et comme les événements A_n sont deux à deux disjoints, on a par σ-additivité :

$$\mathbb{P}(B) = \mathbb{P}\left(\bigcup_n B \cap A_n\right) = \sum_n \mathbb{P}(B \cap A_n) = \sum_n \mathbb{P}(B|A_n)\,\mathbb{P}(A_n). \qquad \square$$

Exemple 1.4. On considère deux urnes numérotées 1 et 2. . L'urne n°1 contient 2 boules rouges et 1 boule noire, et l'urne n°2 contient 1 boule rouge et 4 boules noires. On tire à pile ou face pour choisir une urne, puis on tire une boule de l'urne choisie. Quelle est la probabilité de tomber sur une boule rouge ? On définit les événements :

U_1 ="choisir l'urne 1", U_2 ="choisir l'urne 2", et R ="tomber sur une boule rouge". La formule des probabilités totales nous donne

$$\begin{aligned}\mathbb{P}(R) &= \mathbb{P}(R|U_1)\,\mathbb{P}(U_1) + \mathbb{P}(R|U_2)\,\mathbb{P}(U_2)\\ &= \frac{2}{3} \times \frac{1}{2} + \frac{1}{5} \times \frac{1}{2} = \frac{13}{30}.\end{aligned}$$

La formule des probabilités totales correspond exactement à la notion bien connue d'arbre de probabilité. Dans notre exemple, cela correspond à l'arbre suivant :

Proposition 1.8 (Formule de Bayes)**.** *Soit A et B deux événements de probabilités non nulles. Alors,*

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

Démonstration. En effet,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\,\mathbb{P}(A)}{\mathbb{P}(B)}. \qquad \square$$

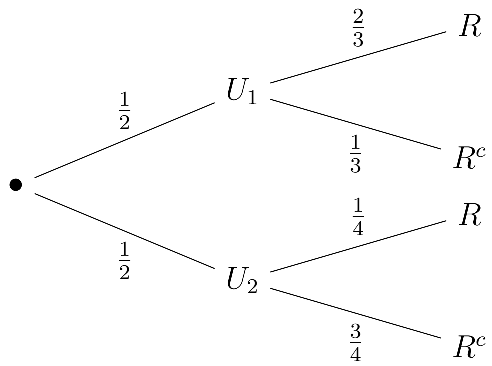

FIGURE 1.3. Arbre de probabilités et formule des probabilités totales.

Exemple 1.5. Exemple très classique d'un test d'une maladie. On considère une maladie rare touchant une personne sur 10000. On considère une test de dépistage : le test est positif à 99% sur les personnes malade et à 0.1% sur les personnes saines. Introduisons les événements :

$M =$"être malade", et $T =$"avoir un test positif". La formule de Bayes donne :

$$\mathbb{P}(M|T) = \frac{\mathbb{P}(M \cap T)}{\mathbb{P}(T)} = \frac{\mathbb{P}(T|M)\,\mathbb{P}(M)}{\mathbb{P}(T)}$$

Or, par la formule des probabilités totales,

$$\mathbb{P}(T) = \mathbb{P}(T|M)\,\mathbb{P}(M) + \mathbb{P}(T|M^c)\,\mathbb{P}(M^c) = 0.99 \times 0.0001 + 0.001 \times 0.9999$$

Donc

$$\mathbb{P}(M|T) = \frac{0.99 \times 10^{-4}}{10^{-4} \times 0.99 + 0.9999 \times 10^{-3}} \approx \frac{1}{11} \approx 0.09$$

Pas terrible...

1.6. **Indépendance.** Intuitivement, deux événements A et B sont indépendants si la connaissance de l'un n'influe pas sur la réalisation de l'autre, ce qui peut se traduire en termes de probabilités conditionnelles par $\mathbb{P}(A|B) = \mathbb{P}(A)$. Ceci suggère la définition suivante :

Définition 1.4. *Soit A et B deux événements. On dit que A et B sont indépendants si*

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

On note parfois $A \perp\!\!\!\perp B$ pour dire que A et B sont indépendants.

Exemple 1.6. On lance deux fois successivement une pièce de monnaie équilibrée. On modélise cette expérience aléatoire par $\Omega = \{0,1\} \times \{0,1\}$, et $\mathbb{P}$ la probabilité uniforme sur Ω. On vérifie facilement que les événements "tomber sur pile au premier lancer" et "tomber sur pile au second lancer" sont des événements indépendants.

Remarque 1.3. Si A et B sont indépendants, alors A et B^c sont indépendants (et bien sûr A^c et B, et A^c et B^c aussi) (exo !).

On généralise alors la définition d'indépendance à une suite quelconque d'événements.

Définition 1.5. *On dit que $(A_n)_n$ est une suite d'événements indépendants si pour tout $n \geq 1$, et pour tout $1 \leq i_1 < i_2 < \cdots < i_k \leq n$,*

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\,\mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_k}).$$

Remarque 1.4. L'indépendance d'événements deux à deux n'implique pas forcément l'indépendance mutuelle. Considérons l'exemple suivant : on lance deux fois une pièce de monnaie équilibrée. Soit les événements :

$$A = \text{« on obtient pile au premier lancer » ;}$$
$$B = \text{« on obtient face au second lancer » ;}$$
$$C = \text{« le résultat des deux lancers est identique ».}$$

On vérifie facilement que A, B, C sont deux à deux indépendants, mais ne sont pas mutuellement indépendants :

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\varnothing) = 0 \neq \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C).$$

2. Variables aléatoires

2.1. **Définition.** Une variable aléatoire représente le résultat d'une expérience aléatoire. On ne s'intéresse par la suite qu'aux variables aléatoires discrètes, c'est-à-dire ne prenant qu'un nombre dénombrable de valeurs. Dans toute la suite, $(\Omega, \mathbb{P})$ désigne un espace de probabilité, et E un sous-ensemble dénombrable de $\mathbb{R}$.

Définition 2.1. *On appelle variable aléatoire (discrète), en abrégé v.a., toute application $X : \Omega \to E$, où E est un ensemble dénombrable.*

Notation 2.1 (TRÈS IMPORTANTE)**.** On notera $\{X \in A\}$ l'image réciproque de A par X, c'est-à-dire

$$\{X \in A\} = X^{-1}(A) = \{\omega \in \Omega \,|\, X(\omega) \in A\}.$$

L'ensemble $\{X \in A\}$ est donc bien un sous-ensemble de Ω appartenant à $\mathcal{P}(\Omega)$, et on peut donc considérer sa probabilité de réalisation, qu'on écrira plus simplement $\mathbb{P}(X \in A)$.

En particulier, on écrira $\{X = a\}$ pour $\{\omega \in \Omega \,|\, X(\omega) = a\}$, $\{X \leq a\}$ pour $\{\omega \in \Omega \,|\, X(\omega) \leq a\}$, etc...

On utilisera une virgule pour désigner l'intersection d'événements portant sur des variables aléatoires : par exemple, $\{X \in A, Y \in B\}$ signifie $\{X \in A\} \cap \{Y \in B\}$.

Remarque 2.1. Ce qu'il faut comprendre, et c'est une des difficultés du formalisme de la théorie des probabilités, c'est que X représente le résultat d'une expérience aléatoire (par exemple le résultat du lancer d'un dé si $X : \Omega \to \{1, 2, \dots, 6\}$), et doit donc être vu comme un nombre aléatoire. L'espace Ω représente l'aléa, et ce à quoi on a accès est le résulat de l'expérience, donc ce qui se passe sur l'espace E. Une "réalisation" de cette expérience correspond alors à $X(\omega)$, pour un certain $\omega \in \Omega$, qui correspond cette fois à un nombre, et n'a plus rien d'aléatoire.

Définition 2.2. *Soit X une variable aléatoire discrète. La loi de X est définie par la suite $(p_k)_{k\in E}$ définie par*

$$p_k = \mathbb{P}(X = k), \quad \textit{pour tout } k \in E.$$

La suite $(p_k)_{k\in E}$ définit alors une probabilité sur $(E, \mathcal{P}(E))$.

Le fait que $(p_k)_{k\in E}$ soit une probabilité sur $(E, \mathcal{P}(E))$ vient du fait qu'on peut partitionner Ω selon les valeurs prises par X :

$$\Omega = \bigcup_{k\in E} \{\omega \in \Omega \,|\, X(\omega) = k\} = \bigcup_{k\in E} \{X = k\}.$$

La réunion étant disjointe, on a donc par σ-additivité,

$$\sum_{k\in E} p_k = \sum_{k\in E} \mathbb{P}(X = k) = \mathbb{P}(\Omega) = 1.$$

Notation 2.2. On notera $X \sim \mu$ pour dire que X suit la loi μ.

Exemple 2.1. Soit A un événement. Alors, l'application $\mathbb{1}_A : \Omega \to \{0, 1\}$, appelée fonction indicatrice de A, définie par

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \text{si } \omega \in A, \\ 0, & \text{sinon,} \end{cases}$$

est une variable aléatoire de loi (p_0, p_1) donnée par $p_1 = \mathbb{P}(\mathbb{1}_A = 1) = \mathbb{P}(A)$ et $p_0 = \mathbb{P}(1_A = 0) = \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

2.2. **Lois usuelles.** Dressons un catalogue (non exhaustif) des lois les plus courantes.

Exemple 2.2. Loi uniforme $\mathcal{U}\{1, \ldots, n\}$ sur $\{1, \ldots, n\}$. Soit $X : \Omega \to \{1, \ldots, n\}$ telle que

$$\mathbb{P}(X = k) = \frac{1}{n} \quad \text{pour tout } k \in \{1, \ldots, n\}$$

C'est l'équiprobabilité. Si on lance un dé à six faces non truqué, et si X désigne le résultat du dé, alors X suit la loi uniforme sur $\{1, 2, 3, 4, 5, 6\}$.

Exemple 2.3. Loi de Bernoulli $\mathcal{B}(p)$ de paramètre p. Soit $X : \Omega \to \{0, 1\}$ telle que

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p.$$

C'est donc la loi d'une expérience aléatoire à deux issues possibles. Si on lance une pièce de monnaie truquée dont la probabilité de tomber sur pile est p, en notant 1 pour pile et 0 pour face, et si X désigne le résultat de la pièce, alors X suit une loi de Benoulli de paramètre p.

Exemple 2.4. Loi binomiale $\mathcal{B}(n, p)$ de paramètres n et p. Soit $X : \Omega \to \{0, 1, \ldots, n\}$ telle que

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad \text{pour } k = 0, \ldots, n.$$

On verra que c'est la loi du nombre de succès dans une suite de n expériences indépendantes à deux issues possibles.

Exemple 2.5. Loi géométrique $\mathcal{G}_{\mathbb{N}^*}(p)$ sur $\mathbb{N}^*$ de paramètre p. Soit $X : \Omega \to \mathbb{N}^*$ telle que

$$\mathbb{P}(X = k) = (1 - p)^{k-1} p, \quad \text{pour } k \geq 1.$$

La série géométrique $\sum_{k \geq 0} x^n = \frac{1}{1-x}$, pour $x \in]0, 1[$ assure que ceci définit bien une loi de probabilité.

Exemple 2.6. Loi de Poisson $\mathcal{P}(\lambda)$ de paramètre $\lambda \in]0, +\infty[$. Soit $X : \Omega \to \mathbb{N}$ telle que

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \text{pour } k \geq 0.$$

Usuellement, la loi de Poisson modélise les événements rares. On rappelle que la série exponentielle est donnée par : $\sum_{k \geq 0} \frac{\lambda^k}{k!} = e^{\lambda}$.

2.3. **Fonction de répartition.**

Définition 2.3. *Soit X une variable aléatoire à valeurs dans $\mathbb{Z}$. La fonction de répartition de X (ou de la loi de X) est la fonction $F_X : \mathbb{R} \to [0, 1]$ définie par*

$$F_X(t) = \mathbb{P}(X \leq t), \quad \textit{pour } t \in \mathbb{R}.$$

On a donc :

$$F_X(t) = \mathbb{P}(X \leq n) = \sum_{k \leq n} \mathbb{P}(X = k), \quad \textit{si } t \in [n, n + 1[.$$

Au vu de la définition, on a que la fonction de répartition F_X d'une v.a. X est croissante, a pour limites 0 en $-\infty$ et 1 en $+\infty$, est constante par morceaux, continue à droite avec un saut pour chaque entier n tel que $\mathbb{P}(X = n) > 0$.

Remarque 2.2. On remarque que

$$\mathbb{P}(X = n) = \mathbb{P}(X \leq n) - \mathbb{P}(X \leq n - 1),$$

ainsi la fonction de répartition caractérise la loi de X. Donc X et Y sont de même loi si et seulement si elles ont la même fonction de répartition.

Exemple 2.7. La fonction de répartition n'est pas souvent facilement calculable dans le cas des lois discrètes. On peut néanmoins faire le calcul dans le cas d'une v.a. X de loi géométrique de paramètre p. On a en effet, pour tout $n \geq 1$, en posant $q = 1 - p$,

$$\mathbb{P}(X \leq n) = \sum_{k=1}^{n} \mathbb{P}(X = k) = \sum_{k=1}^{n} pq^{k-1} = p\frac{1 - q^n}{1 - q} = 1 - q^n.$$

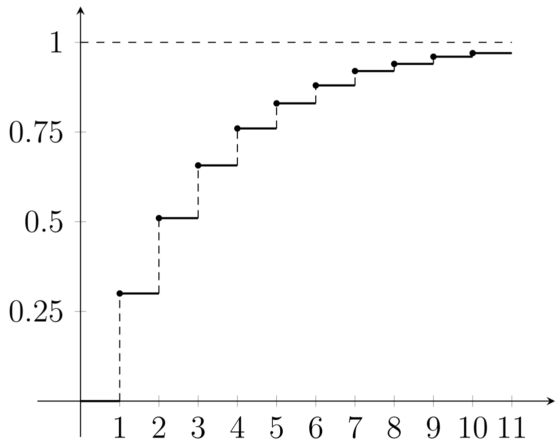

Figure 2.4. Fonction de répartition de la loi géométrique sur $\mathbb{N}^*$ de paramètre $p = 0.3$.

2.4. **Couple de variables aléatoires.** Soit $X : \Omega \to E$ et $Y : \Omega \to F$ deux variables aléatoires discrètes. Le couple (X, Y) peut être vu comme une variable aléatoire $(X, Y) : \Omega \to E \times F$ à valeurs dans l'espace produit $E \times F$, qui est dénombrable comme produit cartésien d'espaces dénombrables. On parle alors de couple aléatoire ou plus généralement de vecteur aléatoire. La loi du couple (X, Y) est alors définie comme la collection des nombres

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

pour tout $(x, y) \in E \times F$, où on rappelle que l'on a noté $\{X = x, Y = y\}$ l'intersection $\{X = x\} \cap \{Y = y\}$. On a donc

$$\sum_{(x,y) \in E \times F} \mathbb{P}(X = x, Y = y) = 1.$$

En général, si on connaît les lois de X et de Y, il n'y a aucune raison de connaître la loi du couple (X, Y). On a par contre :

Proposition 2.1 (Formule des lois marginales)**.** *Soit* (X,Y) *un couple de variables aléatoires à valeurs dans* $E \times F$*. Alors, on a*

$$\mathbb{P}(X=x) = \sum_{y\in F} \mathbb{P}(X=x, Y=y), \quad \textit{pour tout } x \in E$$

$$\mathbb{P}(Y=y) = \sum_{x\in E} \mathbb{P}(X=x, Y=y), \quad \textit{pour tout } y \in F.$$

La proposition nous dit que si on connaît la loi du couple (X,Y), on connaît les lois, dites marginales, de X et de Y.

Démonstration. C'est immédiat, en sommant sur les valeurs d'une des deux v.a. :

$$\begin{aligned}\sum_{y\in F} \mathbb{P}(X=x, Y=y) &= \mathbb{P}\left(\{X=x\} \cap \bigcup_{y\in F}\{Y=y\}\right)\\ &= \mathbb{P}(\{X=x\}\cap\{Y\in F\})\\ &= \mathbb{P}(X=x).\end{aligned}$$ □

2.5. **Indépendance des variables aléatoires.** Au vu de la définition de l'indépendance d'événements, on introduit naturellement la définition suivante.

Définition 2.4. *Soit* $X\colon \Omega \to E$ *et* $Y\colon \Omega \to F$ *deux variables aléatoires discrètes. On dit que* X *et* Y *sont indépendantes si pour tout* $x \in E$*, et tout* $y \in F$*, les événements* $\{X=x\}$ *et* $\{Y=y\}$ *sont indépendants, i.e.*

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

Le couple (X,Y) est une variable aléatoire à valeurs dans l'espace produit $E\times F$ (qui est dénombrable comme produit cartésien d'espaces dénombrables). La collection des nombres

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

pour $(x,y) \in E\times F$ est donc exactement la loi du couple (X,Y). Les v.a. X et Y sont donc indépendantes si la loi du couple (X,Y) s'écrit comme le produit des lois de X et de Y.

Remarque 2.3. Une définition équivalente de l'indépendance est : X et Y sont indépendantes si et seulement si pour tout $A \in \mathcal{P}(E)$, et tout $B \in \mathcal{P}(F)$,

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

Généralisons la notion d'indépendance à un nombre quelconque de variables aléatoires.

Définition 2.5. *Soit* $X_1,\dots,X_n$ *des variables aléatoires discrètes définies sur le même espace de probabilité, et à valeurs dans* $E_1,\dots,E_n$ *respectivement. On dit que* $X_1,\dots,X_n$ *sont indépendantes si pour tous* $x_1\in E_1,\dots,x_n\in E_n$*, les événements*

$$\{X_1=x_1\},\dots,\{X_n=x_n\}$$

sont indépendants

Par définition de l'indépendance (mutuelle) d'événements, il s'agit donc de vérifier que la probabilité de l'intersection de toute sous famille de ces événements est égale au produit des probabilités de ces événements. Mais en fait, l'indépendance de v.a. est beaucoup plus simple :

Proposition 2.2. *Les variables aléatoires* $X_1,\dots,X_n$ *sont indépendantes si et seulement si pour tous* $x_1\in E_1,\dots,x_n\in E_n$*,*

$$\mathbb{P}(X_1=x_1,\dots,X_n=x_n) = \mathbb{P}(X_1=x_1)\cdots\mathbb{P}(X_n=x_n).$$

Démonstration. La condition nécessaire est évidente. Réciproquement, supposons que

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n),$$

pour tous $x_1 \in E_1, \dots, x_n \in E_n$. Soit $I = \{i_1, \dots, i_k\} \subset \{1, \dots, n\}$. Par la formule des marginales, en sommant sur tous les x_j avec $j \notin I$, on a

$$\begin{aligned}
\mathbb{P}(X_{i_1} = x_{i_1}, \dots X_{i_k} = x_{i_k}) &= \sum_{\substack{x_j \in E_j \\ j \notin I}} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\
&= \sum_{\substack{x_j \in E_j \\ j \notin I}} \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n) \\
&= \mathbb{P}(X_{i_1} = x_{i_1}) \cdots \mathbb{P}(X_{i_k} = x_{i_k}),
\end{aligned}$$

car $\sum_{x_j \in E_j} \mathbb{P}(X_j = x_j) = 1$. Ainsi, on a bien que les événements $\{X_1 = x_1\}, \dots, \{X_n = x_n\}$ sont indépendants. $\square$

Enfin, on généralise la notion d'indépendance à une suite infinie de variables aléatoires. On admettra qu'une telle suite existe.

Définition 2.6. *Soit $(X_n)_n$ une suite de variables aléatoires définies sur le même espace de probabilité. On dit que les v.a. X_n sont indépendantes si pour tout $I \subset \mathbb{N}$ fini, les v.a. $(X_i)_{i \in I}$ sont indépendantes.*

Exemple 2.8 (Exemple fondamental)**.** On lance n fois une pièce de monnaie truquée, dont la probabilité de tomber sur pile est p, les lancers étant supposés indépendants. Soit S_n le nombre de fois où l'on a obtenu pile. Alors $S_n \sim \mathcal{B}(n, p)$. En effet, pour modéliser cette expérience, on introduit les v.a.

$$X_i = \begin{cases} 1 & \text{si on tombe sur pile au } i^{\text{e}} \text{ lancer,} \\ 0 & \text{sinon.} \end{cases}$$

Les v.a. $X_1, \dots, X_n$ sont alors indépendantes et de même loi de Bernoulli de paramètre p. Soit la v.a.

$$S_n = X_1 + \cdots + X_n,$$

qui représente alors le nombre de fois où l'on est tombé sur pile sur les n lancers. On cherche à déterminer la loi de S_n, donc

$$\mathbb{P}(S_n = k), \quad \text{pour tout } k \in \{0, 1, \dots, n\}.$$

Pour que $S_n = k$, il faut qu'il y ait exactement k variables X_i qui prennent la valeur 1 et $n - k$ la valeur 0. Il y a $\binom{n}{k}$ choix possibles de ces k variables, ainsi, par indépendance, on a

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

Exemple 2.9. On lance une pièce de monnaie truquée, dont la probabilité de tomber sur pile est p, et on s'arrête la première fois où l'on tombe sur pile. On modélise l'expérience par une suite $(X_i)_{i \geq 1}$ de v.a. indépendantes, de même loi de Bernoulli $\mathcal{B}(p)$, où

$$X_i = \begin{cases} 1 & \text{si on tombe sur pile au } i^{\text{e}} \text{ lancer,} \\ 0 & \text{sinon.} \end{cases}$$

On note T la variable aléatoire égale au nombre de lancers effectués. Alors T est à valeurs dans $\mathbb{N}^*$, et pour tout $n \geq 1$,

$$\begin{aligned}\mathbb{P}(T = n) &= \mathbb{P}(X_1 = 0, \ldots, X_{n-1} = 0, X_n = 1)\\ &= \mathbb{P}(X_1 = 0) \cdots \mathbb{P}(X_{n-1} = 0)\,\mathbb{P}(X_n = 1),\end{aligned}$$

par indépendance des v.a. X_i. Donc,

$$\mathbb{P}(T = n) = (1-p)^{n-1}p, \quad \text{pour } n \geq 1,$$

et T suit une loi géométrique de paramètre p.

On peut aussi utiliser la fonction de répartition. Calculons $\mathbb{P}(T > n)$. L'événement $\{T > n\}$ signifie que les n premiers lancers ont donné face, donc

$$\{T > n\} = \{X_1 = 0, \ldots, X_n = 0\}$$

et par indépendance,

$$\mathbb{P}(T > n) = \mathbb{P}(X_1 = 0) \cdots \mathbb{P}(X_n = 0) = (1-p)^n.$$

On a donc $\mathbb{P}(T \leq n) = 1 - (1-p)^n$, et donc la fonction de répartition de T est la fonction de répartition de la loi géométrique.

Remarque 2.4. Si X et Y sont indépendantes, alors $f(X)$ et $g(Y)$ sont indépendantes pour toutes fonctions f et g. De même, si $(X_n)_n$ est une suite de v.a. indépendantes, alors pour tous sous-ensembles finis I et J disjoints, on a indépendance de $(X_i)_{i\in I}$ et de $(X_j)_{j\in J}$ (indépendance par « paquet »).

3. Espérance, moments

3.1. Définition.

Définition 3.1. *Soit X une variables aléatoires discrète à valeurs dans un sous-ensemble dénombrable E de $\mathbb{R}$. On dit que X est intégrable (ou sommable) si*

$$\sum_{x\in E} |x|\,\mathbb{P}(X = x) < +\infty,$$

et dans ce cas, on définit l'espérance de X par

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

Remarque 3.1. L'espérance de X correspond donc à la moyenne de X : on somme les valeurs de X pondérées par leurs probabilités de réalisation.

Si X est à valeurs dans un ensemble fini, l'espérance est une somme finie et est donc toujours bien définie. Si X est à valeurs dans un ensemble infini dénombrable et est de plus *positive*, disons $X : \Omega \to \mathbb{N}$, alors la quantité $\sum_{x\in\mathbb{N}} x\,\mathbb{P}(X = x)$ est bien définie comme nombre de $[0, +\infty]$. En effet, la suite des sommes partielles

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

est une suite croissante, et sa limite est alors soit finie, soit égale à $+\infty$. On peut donc toujours définir $\mathbb{E}(X)$ quand X est une v.a. positive. On sera la plupart du temps dans ce cours dans le cas positif, et donc on ne se préoccupera pas de questions de convergence de série (voir le cours ana4 pour toutes ces questions).

L'espérance ne dépend que de la loi de X. Si X et Y ont même loi, alors $\mathbb{E}(X) = \mathbb{E}(Y)$. On parlera alors aussi de l'espérance d'une loi de probabilité.

3.2. **Espérances des lois usuelles.**

Exemple 3.1. Soit $X \sim \mathcal{U}\{1, \dots, n\}$. Alors,

$$\mathbb{E}(X) = \sum_{k=1}^{n} k\,\mathbb{P}(X = k) = \frac{1}{n}\sum_{k=1}^{n} k = \frac{1}{n}\frac{n(n+1)}{2} = \frac{(n+1)}{2}.$$

Exemple 3.2. Soit $X \sim \mathcal{B}(p)$. Alors,

$$\mathbb{E}(X) = \sum_{k\in\{0,1\}} k\,\mathbb{P}(X = k) = p.$$

Exemple 3.3. Soit $X \sim \mathcal{B}(n, p)$. Alors,

$$\mathbb{E}(X) = \sum_{k=0}^{n} k\,\mathbb{P}(X = k) = \sum_{k=1}^{n} k\binom{n}{k}p^k(1-p)^{n-k}.$$

On utilise alors la formule appelée parfois "formule du pion" :

$$k\binom{n}{k} = n\binom{n-1}{k-1}.$$

Pour démontrer cette formule, on peut utiliser une technique usuelle en combinatoire, dénombrer la même chose de deux façons différentes. En effet, considérons un groupe de n personnes. On cherche à former un comité de k personnes, avec un ou une présidente dans ce comité. Combien de possibilités a-t-on ? Soit on commence par choisir les k personnes parmi n qui formeront ce comité, puis le ou la présidente parmi les k personnes choisies, ce qui donne $k\binom{n}{k}$ possibilités, soit on commence par choisir le ou la présidente parmi les n personnes, puis on complète le comité, avec donc $k-1$ personnes à choisir parmi $n-1$, ce qui donne $n\binom{n-1}{k-1}$ possibilités. Ainsi,

$$k\binom{n}{k} = n\binom{n-1}{k-1}.$$

On peut bien sûr montrer cette formule directement en explicitant les coefficients binomiaux (exercice). On a donc,

$$\mathbb{E}(X) = n\sum_{k=1}^{n}\binom{n-1}{k-1}p^k(1-p)^{n-k} = np\sum_{j=0}^{n-1}\binom{n-1}{j}p^j(1-p)^{n-1-j} = np,$$

par la formule du binôme de Newton.

Exemple 3.4. Soit $X \sim \mathcal{G}_{\mathbb{N}^*}(p)$. Alors,

$$\mathbb{E}(X) = \sum_{k\geq 1} k\,\mathbb{P}(X = k) = \sum_{k\geq 1} kp(1-p)^{k-1}.$$

On rappelle que la série entière

$$\sum_{k\geq 0} x^k = \frac{1}{1-x}$$

a pour rayon de convergence 1, et est donc dérivable dans $]0, 1[$, de dérivée

$$\left(\sum_{k\geq 0} x^k\right)' = \sum_{k\geq 1} kx^{k-1} = \frac{1}{(1-x)^2}.$$

Ainsi,

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

Sans utiliser les séries entières, on peut procèder de la sorte :

$$\sum_{k\geq 1} kx^{k-1} = \sum_{j\geq 0}(j+1)x^j = \sum_{j\geq 0} jx^j + \sum_{j\geq 0} x^j = x\sum_{j\geq 1} jx^{j-1} + \sum_{j\geq 0} x^j,$$

d'où

$$(1-x)\sum_{k\geq 1} kx^{k-1} = \sum_{j\geq 0} x^j = \frac{1}{1-x},$$

et on obtient le résultat.

Exemple 3.5. Soit $X \sim \mathcal{P}(\lambda)$. Alors,

$$\mathbb{E}(X) = \sum_{n\geq 0} n\,\mathbb{P}(X=n) = \sum_{n\geq 1} n\frac{\lambda^n}{n!}e^{-\lambda} = \sum_{n\geq 1}\frac{\lambda^n}{(n-1)!}e^{-\lambda} = \lambda\sum_{k\geq 0}\frac{\lambda^k}{k!}e^{-\lambda} = \lambda.$$

3.3. **Propriétés.**

Proposition 3.1. *Soient X et Y deux v.a. discrètes à valeurs dans un sous-ensemble dénombrable de $\mathbb{R}$, et soit $\alpha \in \mathbb{R}$. L'espérance est linéaire, c'est à dire :*

$$\mathbb{E}\left(\alpha X + Y\right) = \alpha\,\mathbb{E}(X) + \mathbb{E}(Y),$$

et positive : si X est à valeurs dans dans un sous-ensemble de $\mathbb{R}_+$, alors $\mathbb{E}(X) \geq 0$. En particulier, l'espérance est croissante : si $X \leq Y$, alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

Démonstration. Notons $E = X(\Omega)$ et $F = Y(\Omega)$. Alors, $X+Y$ est une v.a. à valeurs dans $G = (X+Y)(\Omega) = \{x+y \,|\, x \in E, y \in F\}$ (qui est encore dénombrable). Supposons dans un premier temps que X et Y sont positives. Alors, en partitionnant Ω selon les valeurs de X, i.e. $\Omega = \bigcup_{x\in E}\{X = x\}$, on a

$$\begin{aligned}
\mathbb{E}(X+Y) &= \sum_{z\in G} z\,\mathbb{P}(X+Y=z) = \sum_{z\in G}\sum_{x\in E} z\,\mathbb{P}(X+Y=z, X=x)\\
&= \sum_{z\in G}\sum_{x\in E} z\,\mathbb{P}(Y=z-x, X=x) = \sum_{x\in E}\sum_{z\in G} z\,\mathbb{P}(Y=z-x, X=x)\\
&= \sum_{x\in E}\sum_{y\in F}(x+y)\,\mathbb{P}(Y=y, X=x)\\
&= \sum_{x\in E}\sum_{y\in F} x\,\mathbb{P}(Y=y, X=x) + \sum_{x\in E}\sum_{y\in F} y\,\mathbb{P}(Y=y, X=x)\\
&= \sum_{x\in E} x\,\mathbb{P}(X=x) + \sum_{y\in F} y\,\mathbb{P}(Y=y)\\
&= \mathbb{E}(X) + \mathbb{E}(Y).
\end{aligned}$$

Si X et Y sont maintenant intégrables, et de signe quelconque, alors, par l'inégalité triangulaire $|x+y| \leq |x| + |y|$, on a

$$\begin{aligned}
\sum_{x\in E}\sum_{y\in F}|x+y|\,\mathbb{P}(Y=y, X=x) &\leq \sum_{x\in E}\sum_{y\in F}|x|\,\mathbb{P}(Y=y, X=x) + \sum_{x\in E}\sum_{y\in F}|y|\,\mathbb{P}(Y=y, X=x)\\
&= \sum_{x\in E}|x|\,\mathbb{P}(X=x) + \sum_{y\in F}|y|\,\mathbb{P}(Y=y)\\
&< \infty
\end{aligned}$$

Ainsi, $X+Y$ est intégrable, et comme toutes les sommes sont finies, on peut faire le même calcul que dans le cas positif.

Soit $\alpha \in \mathbb{R}$. Alors, αX est à valeurs dans αE, et on a

$$\mathbb{E}(\alpha X) = \sum_{x\in E}\alpha x\,\mathbb{P}(\alpha X = \alpha x) = \sum_{x\in E}\alpha x\,\mathbb{P}(X=x) = \alpha\,\mathbb{E}(X).$$

Si X est positive, son espérance est évidemment positive, et donc en particulier, si $X \geq Y$, i.e. $X - Y \geq 0$, alors $\mathbb{E}(X - Y) \geq 0$, d'où $\mathbb{E}(X) \geq \mathbb{E}(Y)$. □

Théorème 3.1 (Théorème de transfert)**.** *Soit $X\colon \Omega \to E$ une variable aléatoire discrète et soit $f\colon E \to \mathbb{R}$ une fonction sur E, soit positive, soit de signe quelconque et vérifiant $\sum_{k\in E} |f(k)|\,\mathbb{P}(X = k) < \infty$. Alors,*

$$\mathbb{E}(f(X)) = \sum_{k\in E} f(k)\,\mathbb{P}(X = k).$$

Ce théorème est fondamental. Il dit donc que pour calculer l'espérance de $f(X)$, on a pas besoin de connaître la loi de $f(X)$, il suffit de connaître la loi de X.

Démonstration. Comme E est dénombrable, $f(E)$ est dénombrable. Supposons f positive. Par définition de l'espérance, on a donc

$$\mathbb{E}(f(X)) = \sum_{y\in f(E)} y\,\mathbb{P}(f(X) = y),$$

qui est bien définie dans $[0, +\infty]$ comme série à termes positifs. En partitionnant Ω selon les valeurs de X, on a

$$\{f(X) = y\} = \bigcup_{x\in E} \{f(X) = y, X = x\} = \bigcup_{x\in E} \{f(x) = y, X = x\} = \bigcup_{x\in f^{-1}(y)} \{X = x\},$$

l'union étant disjointe. On a donc,

$$\begin{aligned}\mathbb{E}(f(X)) &= \sum_{y\in f(E)} y\,\mathbb{P}(f(X) = y)\\ &= \sum_{y\in f(E)} \sum_{x\in f^{-1}(y)} y\,\mathbb{P}(X = x)\\ &= \sum_{y\in f(E)} \sum_{x\in f^{-1}(y)} f(x)\,\mathbb{P}(X = x)\\ &= \sum_{x\in E} f(x)\,\mathbb{P}(X = x),\end{aligned}$$

car $E = \bigcup_{y\in f(E)} f^{-1}(y)$, l'union étant clairement disjointe. Si f est de signe quelconque, la condition

$$\sum_{k\in E} |f(k)|\,\mathbb{P}(X = k) < \infty$$

assure que la série est bien définie, et on effectue le même raisonnement. □

Remarque 3.2. La condition X intégrable n'est donc rien d'autre que la condition $\mathbb{E}(|X|) < \infty$.

Exemple 3.6. Soit X une variable aléatoire à valeurs dans $\{-1, 1\}$, de loi $\mathbb{P}(X = \pm 1) = \frac{1}{2}$. Alors,

$$\mathbb{E}\left(e^{tX}\right) = \frac{1}{2}e^{-t} + \frac{1}{2}e^{t} = \cosh(t)$$

pour tout $t \in \mathbb{R}$.

Proposition 3.2. *Soit X et Y deux variables aléatoires. Alors, X et Y sont indépendantes si et seulement si pour toutes fonctions réelles f et g telles que les espérances ci-dessous ont un sens,*

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\,\mathbb{E}(g(Y)).$$

Démonstration. Par le théorème de transfert, appliqué à la v.a. (X, Y) et à la fonction $(x, y) \mapsto f(x)g(y)$, on a

$$\mathbb{E}(f(X)g(Y)) = \sum_{x \in E, y \in F} f(x)g(y)\,\mathbb{P}(X = x, Y = y).$$

Donc si X et Y sont indépendantes, $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y)$, et donc

$$\begin{aligned}\mathbb{E}(f(X)g(Y)) &= \sum_{x \in E, y \in F} f(x)g(y)\,\mathbb{P}(X = x)\,\mathbb{P}(Y = y) \\ &= \left(\sum_{x \in E} f(x)\,\mathbb{P}(X = x)\right)\left(\sum_{y \in F} g(y)\,\mathbb{P}(Y = y)\right) \\ &= \mathbb{E}(f(X))\,\mathbb{E}(g(Y)).\end{aligned}$$

Réciproquement, si on a

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\,\mathbb{E}(g(Y)),$$

pour tout f et g, alors l'égalité est vraie en particulier pour $f = \mathbb{1}_{\{x\}}$ et $g = \mathbb{1}_{\{y\}}$, ce qui donne que pour tout $x \in E$ et tout $y \in F$,

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

et donc X et Y sont indépendantes. □

Remarque 3.3. En particulier, si X et Y sont indépendantes, et si $\mathbb{E}(XY)$ est finie, alors $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$. La réciproque est fausse en général.

Finissons par la formule suivante dont la démonstration est une petite manipulation de séries à termes positifs.

Proposition 3.3. *Soit X une v.a. à valeurs dans $\mathbb{N}$. Alors,*

$$\mathbb{E}(X) = \sum_{k \geq 1} \mathbb{P}(X \geq k).$$

Démonstration. En partant du membre de droite, on a

$$\sum_{k \geq 1} \mathbb{P}(X \geq k) = \sum_{k \geq 1} \sum_{j \geq k} \mathbb{P}(X = j) = \sum_{j \geq 1} \sum_{k=1}^{j} \mathbb{P}(X = j) = \sum_{j \geq 1} j\,\mathbb{P}(X = j) = \mathbb{E}(X).$$

□

3.4. **Moments d'ordre supérieur.**

Définition 3.2. *Soit X une variable aléatoire, et $k \geq 1$. On dit que X admet un moment d'ordre k si*

$$\mathbb{E}(|X|^k) < +\infty.$$

Proposition 3.4. *Soit X une variable aléatoire discrète admettant un moment d'ordre n. Alors X admet un moment de tout ordre $k \leq n$.*

Démonstration. Soit $k \leq n$. On a

$$\begin{aligned}\mathbb{E}(|X|^k) &= \sum_{x \in E} |x|^k\,\mathbb{P}(X = k) \\ &= \sum_{x \in E} |x|^k \mathbb{1}_{\{|x| \leq 1\}}\,\mathbb{P}(X = k) + \sum_{x \in E} |x|^k \mathbb{1}_{\{|x| > 1\}}\,\mathbb{P}(X = k).\end{aligned}$$

Or $|x|^k$ est croissant en k pour $|x| > 1$, donc le deuxième terme ci-dessus est majoré par

$$\sum_{x \in E} |x|^n \mathbb{1}_{\{|x| > 1\}}\,\mathbb{P}(X = k) \leq \mathbb{E}(|X|^n),$$

qui est fini par hypothèse. Le premier terme est quand lui majoré trivialement par

$$\sum_{x\in E} |x|^k \mathbb{1}_{\{|x|\leq 1\}} \mathbb{P}(X=k) \leq \sum_{x\in E} \mathbb{P}(X=k) = 1.$$

Finalement, $\mathbb{E}(|X|^k) \leq 1 + \mathbb{E}(|X|^n)$, et X admet donc un moment d'ordre k, pour tout $k \leq n$. □

Définition 3.3. *Soit X une variable aléatoire admettant un moment d'ordre 2, i.e. $\mathbb{E}(|X|^2) < \infty$. La variance de X est la quantité*

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right).$$

L'écart-type de X, noté σ_X, est la racine de la variance :

$$\sigma_X = \sqrt{\mathbb{V}\mathrm{ar}(X)}.$$

La proposition précédente assure que la variance est bien définie pour X admettant un moment d'ordre 2. L'écart-type s'interprète comme une mesure de dispersion des valeurs de X par rapport à la moyenne de X.

Proposition 3.5. *On a :*

(i) $\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$ *;*

(ii) $\mathbb{V}\mathrm{ar}(aX+b) = a^2\, \mathbb{V}\mathrm{ar}(X)$.

Démonstration. Pour (i), il suffit de développer le carré dans la définition de la variance, et d'utiliser la linéarité de l'espérance :

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}\left((X-\mathbb{E}(X))^2\right) = \mathbb{E}\left(X^2 - 2X\,\mathbb{E}(X) + (\mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

La deuxième propriété est facile, et laissée en exercice. □

À titre d'exercice, on pourra calculer les variances des lois usuelles :

Exemple 3.7. Soit $X \sim \mathcal{B}(p)$. Alors,

$$\mathbb{V}\mathrm{ar}(X) = p(1-p).$$

Commençons par le moment d'ordre 2. On a

$$\mathbb{E}(X^2) = \sum_{k=0}^{1} k\,\mathbb{P}(X=k) = \mathbb{P}(X=1) = p.$$

Ainsi,

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

Exemple 3.8. Soit $X \sim \mathcal{B}(n,p)$. Alors,

$$\mathbb{V}\mathrm{ar}(X) = np(1-p).$$

Le calcul est similaire à celui de l'espérance. Plutôt que de calculer le moment d'ordre 2, on calcule $\mathbb{E}(X(X-1))$. En effet,

$$\mathbb{E}(X(X-1)) = \sum_{k=2}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k},$$

et on utilise la formule

$$k(k-1)\binom{n}{k} = n(n-1)\binom{n-2}{k-2}.$$

Ainsi,

$$\begin{aligned}\mathbb{E}(X(X-1)) &= n(n-1)\sum_{k=2}^{n} k(k-1)\binom{n-2}{k-2}p^k(1-p)^{n-k} \\ &= n(n-1)p^2\sum_{l=0}^{n-2}\binom{n}{l}p^l(1-p)^{n-l} \\ &= n(n-1)p^2.\end{aligned}$$

Finalement,

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - (\mathbb{E}\,X)^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p).$$

On verra un peu plus loin une autre façon de faire utilisant l'indépendance.

Exemple 3.9. Soit $X \sim \mathcal{U}\{1,\dots,n\}$. Alors,

$$\mathbb{V}\mathrm{ar}(X) = \frac{n^2-1}{12}.$$

On rappelle la formule de la somme des n premiers carrés : $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{12}$. Pour retrouver cette somme, on peut utiliser les cubes ! En effet, on a

$$k^3 - (k-1)^3 = 3k^2 - 3k + 1,$$

et par somme télescopique,

$$\sum_{k=1}^{n}(k^3 - (k-1)^3) = n^3.$$

D'autre part, la somme du terme de droite donne

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

Donc

$$6\sum_{k=1}^{n} k^2 = 2n^3 + 3n(n+1) - 2n = n(2n^2+3n+1) = n(n+1)(2n+1)$$

d'où le résultat.

Par cette astuce, on peut calculer n'importe quel somme de puissance d'un entier connaissant la somme des puissances inférieures.

Finalement, on a

$$\begin{aligned}\mathbb{V}\mathrm{ar}(X) &= \mathbb{E}(X^2) - (\mathbb{E}\,X)^2 \\ &= \frac{1}{n}\sum_{k=1}^{n} k^2 - (\mathbb{E}\,X)^2 \\ &= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \\ &= \frac{n^2-1}{12}.\end{aligned}$$

Exemple 3.10. Soit $X \sim \mathcal{G}_{\mathbb{N}^*}(p)$. Alors,

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

Là encore, plutôt que de calculer $\mathbb{E}(X^2)$, on calcule $\mathbb{E}(X(X-1))$, c'est plus simple. On a :

$$\mathbb{E}(X(X-1)) = \sum_{k\geq 2} k(k-1)pq^{k-1}.$$

On se débarrasse du $k(k-1)$ en écrivant $k(k-1) = 2\sum_{j=1}^{k-1} j$. Alors,

$$\mathbb{E}(X(X-1)) = 2\sum_{k\geq 2}\sum_{j=1}^{k-1} jpq^{k-1} = 2\sum_{j\geq 1} j \sum_{k\geq j+1} pq^{k-1} = 2pq\sum_{j\geq 1} jq^{j-1}\sum_{k\geq 0} q^k$$

Or

$$\sum_{k\geq 0} q^k = \frac{1}{p}$$

et

$$\sum_{j\geq 1} jpq^{j-1} = \mathbb{E}(X) = \frac{1}{p}$$

on obtient donc :

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

Finalement,

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

Si on connaît la théorie des séries entières, il suffit sinon de dériver deux fois la série géométrique pour avoir le résultat.

Exemple 3.11. Soit $X \sim \mathcal{P}(\lambda)$. Alors,

$$\mathbb{V}\mathrm{ar}(X) = \lambda.$$

(Exercice).

Définition 3.4. *Soit X et Y deux variables aléatoires à valeurs dans un sous ensemble dénombrable de $\mathbb{R}$, et admettant un moment d'ordre 2. La covariance de X et de Y, est la quantité*

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

La covariance est une quantité permettant de quantifier l'écarts joints de deux variables aléatoires par rapport à leurs espérances respectives.

Les propriétés suivantes sont faciles à voir :

Propriété.
- *On a :* $\mathbb{V}\mathrm{ar}(X) = \mathrm{Cov}(X,X)$.
- *Le fait que X et Y admettent un moment d'ordre 2 assure que le produit XY est intégrable (en utilisant par exemple l'inégalité $|xy| \leq \frac{1}{2}(x^2+y^2)$).*
- *En développant, on obtient la formule alternative :*

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

- *Si X et Y sont indépendantes, alors* $\mathrm{Cov}(X,Y) = 0$. *La réciproque est fausse en général.*
- *On montre facilement que* Cov *est une application bilinéaire symétrique et que* $\mathbb{V}\mathrm{ar}$ *est la forme quadratique associé à* Cov. *En particulier, on a :*

Proposition 3.6. *Soit X et Y deux variables aléatoires admettant un moment d'ordre 2. Alors,*

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

Si X et Y sont de plus indépendantes, on a

$$\mathbb{V}\mathrm{ar}(X+Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y).$$

Démonstration. Il suffit de développer :

$$\begin{aligned}\mathbb{V}\mathrm{ar}(X+Y) = \mathrm{Cov}(X+Y, X+Y) &= \mathrm{Cov}(X,X) + \mathrm{Cov}(Y,Y) + 2\,\mathrm{Cov}(X,Y)\\ &= \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y) + 2\,\mathrm{Cov}(X,Y).\end{aligned}$$

Si X et Y sont indépendantes, alors $\mathrm{Cov}(X,Y) = 0$, et on a $\mathbb{V}\mathrm{ar}(X+Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y)$. □

Remarque 3.4. Plus généralement, si $X_1, \ldots, X_n$ sont des v.a. admettant un moment d'ordre 2, alors

$$\mathbb{V}\mathrm{ar}(X_1 + \cdots X_n) = \sum_{i=1}^{n} \mathbb{V}\mathrm{ar}(X_i) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j),$$

et si de plus elles sont indépendantes, alors

$$\mathbb{V}\mathrm{ar}(X_1 + \cdots X_n) = \sum_{i=1}^{n} \mathbb{V}\mathrm{ar}(X_i).$$

Exemple 3.12. Calculer la variance d'une loi binomiale $\mathcal{B}(n,p)$ en utilisant la représentation de loi binomiale comme somme de n v.a. indépendantes de Bernoulli de paramètre p.

4. Fonction génératrice

En combinatoire, la série génératrice est une série dont les coefficients codent une suite de nombres. En probabilité, son utilisation est la suivante.

Définition 4.1. *Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. On appelle fonction génératrice de X, notée G_X, la série entière*

$$G_X(s) = \mathbb{E}(s^X) = \sum_{k \geq 0} s^k\, \mathbb{P}(X = k),$$

de rayon de convergence supérieur ou égal à 1.

Il est clair que G_X est bien convergente pour $|s| \leq 1$, et que dans ce cas, $|G_X(s)| \leq 1$. La rayon de convergence est donc au moins égal à 1.

La fonction génératrice ne dépend que de la loi de X, donc si X et Y ont même loi, elles ont même fonction génératrice. On va voir plus bas que la réciproque est également vraie.

4.1. **Fonctions génératrices des lois usuelles.**

Les calculs des fonctions génératrices des lois usuelles sont assez simples (exercice !).

Exemple 4.1. $X \sim \mathcal{U}\{1, \ldots, n\}$.

$$G_X(s) = \frac{s}{n} \frac{1 - s^n}{1 - s}.$$

Exemple 4.2. $X \sim \mathcal{B}(p)$

$$G_X(s) = 1 - p + ps.$$

Exemple 4.3. $X \sim \mathcal{B}(n,p)$

$$G_X(s) = (1 - p + ps)^n.$$

Exemple 4.4. $X \sim \mathcal{G}_{\mathbb{N}^*}(p)$

$$G_X(s) = \frac{ps}{1 - qs}, \quad \text{avec } q = 1 - p.$$

Exemple 4.5. $X \sim \mathcal{P}(\lambda)$

$$G_X(s) = e^{\lambda(s-1)}.$$

4.2. **Propriétés.** On admettra les résultats suivants. On renvoie à ana4 pour les justifications détaillées. On se contentera ici d'utiliser la fonction génératrice comme outil permettant de faire des calculs de lois.

Proposition 4.1. *Pour tout $k \geq 0$, on a*

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

On a donc que la fonction génératrice caractérise la loi : X et Y ont même loi si et seulement si X et Y ont même fonction génératrice.

Démonstration. Ceci provient de l'unicité du développement en série entière. Au voisinage de 0, la série de Taylor de G_X est donnée par

$$G_X(s) = \sum_{k \geq 0} \frac{G_X^{(k)}(0)}{k!} s^k,$$

et la formule suit par unicité. □

Remarque 4.1. En pratique, on utilise rarement cette formule, mais plutôt le fait que l'on connaît la fonction génératrice des lois usuelles.

Proposition 4.2. *Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. Alors X admet un moment d'ordre k si et seulement si G_X est k-fois dérivable à gauche de 1, et dans ce cas, on a*

$$\mathbb{E}\left(X(X-1)\cdots(X-k+1)\right) = \lim_{\substack{s \to 1 \\ s < 1}} G_X^{(k)}(s) = G_X^{(k)}(1^-).$$

Remarque 4.2. En particulier, X est intégrable si et seulement si G_X est dérivable à gauche de 1, et dans ce cas,

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

De même, X admet un moment d'ordre 2 si et seulement si G_X est 2 fois dérivable à gauche de 1, et on a

$$\mathbb{E}(X(X-1)) = G''_X(1^-),$$

ce qui permet de calculer la variance de X par la formule :

$$\mathbb{V}\mathrm{ar}(X) = G''_X(1^-) + G'_X(1^-) - \left(G'_X(1^-)\right)^2$$

Démonstration. Le rayon de convergence de G_X est supérieur ou égal à 1, donc par dérivation des séries entières, on a, pour tout $s \in]-1, 1[$,

$$G_X^{(k)}(s) = \sum_{n \geq k} n(n-1)\cdots(n-k+1) s^{n-k}\, \mathbb{P}(X = n).$$

Donc $G_X^{(k)}(1^-)$ existe signifie que

$$\sum_{n \geq k} n(n-1)\cdots(n-k+1)\, \mathbb{P}(X = n) < \infty,$$

i.e. $\mathbb{E}(X(X-1)\cdots(X-k+1)) < \infty$. On en déduit le résultat. □

Proposition 4.3. *Soit X et Y deux variables aléatoires indépendantes. Alors, pour tout $s \in [-1, 1]$,*

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

Démonstration. On a

$$G_{X+Y}(s) = \mathbb{E}(s^{X+Y}) = \mathbb{E}(s^X s^Y) = \mathbb{E}(s^X)\, \mathbb{E}(s^Y),$$

par indépendance de X et de Y. □

Remarque 4.3. Soit $X_1, \ldots, X_n$ n v.a. indépendantes de même loi de Bernoulli $\mathcal{B}(p)$. Alors $S_n = X_1 + \cdots + X_n$ suit la loi binomiale $\mathcal{B}(n,p)$, et on a

$$G_{X_1+\cdots+X_n}(s) = \prod_{i=1}^{n} G_{X_i}(s) = (G_{X_1}(s))^n = (1-p+ps)^n.$$

Exemple 4.6. Soient X et Y deux v.a. indépendantes, avec $X \sim \mathcal{P}(\lambda)$, et $Y \sim \mathcal{P}(\mu)$. Alors $X+Y \sim \mathcal{P}(\lambda+\mu)$. Utilisons les fonctions génératrices. On a, par indépendance,

$$G_{X+Y}(s) = G_X(s)G_Y(s),$$

et comme on connaît la fonction génératrice de la loi de Poisson, on obtient ici,

$$G_{X+Y}(s) = e^{\lambda(s-1)}e^{\mu(s-1)} = e^{(\lambda+\mu)(s-1)}.$$

On reconnaît ici la fonction génératrice de la loi de Poisson de paramètre $\lambda+\mu$. Comme la fonction génératrice caractérise la loi, on en déduit que $X+Y \sim \mathcal{P}(\lambda+\mu)$.