Mathématiques

Probabilité Semestre 4

Table des matières

Première partie

Espaces Probabilisés

1 Expériences aléatoires

En probabilité, le point de départ est ce qu'on appelle habituellement une expérience aléatoire (ou épreuve). Une telle expérience aléatoire $\mathcal{E}$ est caractérisée par :

— un espace des résultats possibles Ω

— un ensemble d'événements $\mathcal{F}$ (partie de Ω, sous ensemble de Ω) qui doit vérifier les propriétés d'une **tribu**

— la donnée d'une probabilité sur Ω (application de $\mathcal{F} \to [0,1]$ qui doit vérifier 2 axiomes)

3 exemples d'expériences aléatoires.

1. $\mathcal{E}_1$: on lance un dé équilibré
2. $\mathcal{E}_2$: on lance une pièce qui a une probabilité $p > 0$ de retomber sur face jusqu'à obtenir le premier face
3. $\mathcal{E}_3$: un générateur de nombre aléatoire renvoie un nombre réel compris entre 0 et 1 (exemple de ALEA() dans un tableur)

2 L'espace des résultats possibles

Définition. L'ensemble des résultats possibles d'une expérience aléatoire est appelé ensemble des **événements élémentaires** ou encore univers des possibles. On le note généralement Ω.

1. $\mathcal{E}_1 : \Omega_1 = \{1, 2, 3, 4, 5, 6\}$, Ω_1 est un ensemble fini, $card(\Omega_1) = \#\Omega_1 = 6$
2. $\mathcal{E}_2 : \Omega_2 = \{(F), (P, F), (P, P, F), ..., (P, ..., P, F), ...\} \cup \{(P, ..., P, ...)\}$
3. $\mathcal{E}_3 : \Omega_3 = [0, 1]$

Les ensembles Ω_2 et Ω_3 sont tous les 2 infinis mais Ω_2 est dénombrable alors que $[0, 1]$ ne l'est pas.

3 Ensemble dénombrables

Historiquement il existe deux définitions de la notion d'ensemble dénombrable.

Définition 1. On dit qu'un ensemble E est dénombrable s'il est en bijection avec $\mathbb{N}$ (on dit aussi que E est équipotent à $\mathbb{N}$)

définition originale de *Georg Cantor*

Définition 2. On dit que E est dénombrable s'il est en bijection avec une partie de $\mathbb{N}$. Avec cette définition les ensembles finis sont dénombrables. On parlera alors d'ensemble infini dénombrable pour le cas d'un ensemble infini.
Si on retient la définition 1, les ensembles dénombrables sont toujours infinis. Pour regrouper les notions d'ensemble fini et d'ensemble dénombrables, on parle d'ensemble **au plus dénombrables**, dans ce cours on utilisera la définition 1.

Remarque. En probabilité on dit aussi parfois qu'un ensemble fini ou dénombrables est un ensemble discret (probabilité **discrète**). Mais cela ne correspond pas à la "vraie" définition d'un ensemble discret. On dit qu'un sous-ensemble $D \subset \mathbb{R}$ est discret si pour tout $x \in D$, $\exists \varepsilon > 0$ tel que $]x-\varepsilon, x+\varepsilon[\cap D = \{x\}$

Propriété. On peut montrer qu'un sous ensemble de $\mathbb{R}$ discret est fini ou dénombrable, mais la réciproque est clairement fausse.
Soit $E = \left\{\frac{1}{n}, n \geq 1\right\} \cup \{0\}$. E est dénombrable (évident) mais E n'est pas discret au sens de la définition précédence. En effet, on ne peut trouver $\varepsilon > 0$ tel que $]-\varepsilon, \varepsilon[\cap E = \{0\}$. (0 est ce que l'on appelle un point d'accumulation).

Ensemble d'ensemble dénombrables

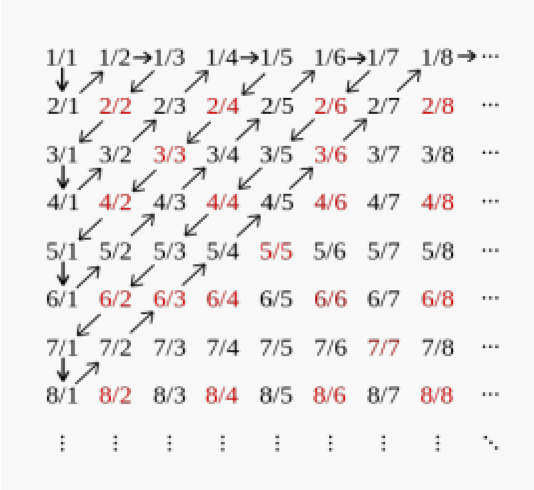
1. $\mathbb{N}^*$ est dénombrable.

   Démonstration. On considère $f : \begin{cases} \mathbb{N} \to \mathbb{N}^* \\ n \mapsto f(n) = n+1 \end{cases}$ □

2. $\mathbb{Z}$ est dénombrable.

   Démonstration. On peut écrire $\mathbb{Z} = \{0, -1, 1, -2, 2, ...\}$. On considère $f : \begin{cases} \mathbb{N} \to \mathbb{Z} \\ n \mapsto f(n) = \begin{cases} n/2 \text{ si } n \text{ pair} \\ -\dfrac{n+1}{2} \text{ si } n \text{ impair} \end{cases} \end{cases}$,
   on peut aussi écrire $(f(n) = (-1)^n E(\frac{n+1}{2}))$. □

3. $\mathbb{N}^2$ est dénombrable.



$\Pi(0,0) = 0, \Pi(1,0) = 1, \Pi(0,1) = 2, \Pi(0,2) = 3, \Pi(1,1) = 4...$ définit une bijection de $\mathbb{N}^2 \to \mathbb{N}$.

On peut aussi remarquer que la fonction $\phi(k,l) = 2^k(2l+1)$ définit une bijection de $\mathbb{N}^2 \to \mathbb{N}^*$

On voit bien qu'intuitivement, on peut dire qu'un ensemble E est dénombrable si on peut numéroter de façon exhaustive tous ses éléments autrement dit si on peut représenter les éléments de E sous la forme d'une suite ou encore si on peut "compter" E (countable set).

Définitions. Si E est un ensemble dénombrable et si $\phi : \mathbb{N} \to E$ est une bijection, on dit que la suite $\phi(0), \phi(1), ..., \phi(n), ...$ est une **énumération de E**.

3.1 Propriétés des ensembles dénombrables.

Théorème I.1. Toute partie infini de $\mathbb{N}$ est dénombrable.

Théorème I.2. Un ensemble est fini ou dénombrable si et seulement si il est en bijection avec une partie de $\mathbb{N}$

Théorème I.3. Toute partie d'un ensemble dénombrable est au plus dénombrable est au plus dénombrable

Corollaire 1. S'il existe une injection d'un ensemble E dans un ensemble dénombrable, alors E est au plus dénombrable.

3.2 Produit cartésien d'ensembles dénombrables.

Théorème I.4. Si E et F sont des ensembles dénombrables, alors $E \times F$ est dénombrable.

Corollaire 2. Si $E_1, ..., E_n$ sont des ensembles dénombrables alors $E_1 \times ... \times E_n$ est dénombrable.

Exemple. $\mathbb{Q}_+^*$ est dénombrable. Soit $\phi : \dfrac{p}{q} \to (p,q)$ où $\dfrac{p}{q}$ est irréductible. On voit bien que ϕ est injective de Q_+^* dans $\mathbb{N} \times \mathbb{N}$. (c'est à dire si s et s' sont deux rationnels positifs, $\phi(s) = \phi(s') \to s = s'$. Or $\mathbb{N} \times \mathbb{N}$ est dénombrable donc $\mathbb{Q}_+^*$ est dénombrable par le corollaire 3.1.

Théorème I.5. (Réunion) Soit $(E_i)_{i \in I}$ une famille d'ensemble dénombrables. Si $\forall i, E_i$ est au plus dénombrable et si I est également dénombrable alors $\bigcup_{i \in I} E_i$ est au plus dénombrable.

Un exemple d'ensemble non-dénombrable. Soit $E \subset \mathbb{R}_+$ l'ensemble des nombres s'écrivant sous la forme

$$x = \sum_{i=1}^{+\infty} \frac{\varepsilon_i}{10^i}, \text{ où } \varepsilon_i \in \{0,1\}$$

Exemple $x = 0,01001110...$ Supposons que E est dénombrable. Alors on peut trouver une bijection de $\mathbb{N}^* \to E$.
$f(1) = 0, \varepsilon_1^1 \varepsilon_2^1 ... \varepsilon_n^1 ...$
$f(2) = 0, \varepsilon_1^2 \varepsilon_2^2 ... \varepsilon_n^2 ...$
$\vdots$
$f(n) = 0, \varepsilon_1^n \varepsilon_2^n ... \varepsilon_n^n ...$
(énumère tout l'ensemble E).
Soit alors $x = 0, \varepsilon_1' \varepsilon_2' ... \varepsilon_n' ...$ avec $\forall i \geq 1, \varepsilon_i' = 1 - \varepsilon_i^i$ ($\forall n \geq 1, \varepsilon_n' = 1 - \varepsilon_n^n$). Alors on voit qu'on ne peut pas avoir $x = f(n)$ pour $n \in \mathbb{N}^*$. Donc E n'est pas dénombrable. Donc $\{0,1\}^{\mathbb{N}}$ ou encore $[0,1]$, ou bien $\mathbb{R}$, ne sont pas dénombrables.

4 Tribus

Définition. Soit un espace Ω. On dit qu'un ensemble $\mathcal{F}$ de $\begin{cases} \text{parties} \\ \text{sous-ensembles} \end{cases}$ de Ω est une tribu si :

1. $\Omega \in \mathcal{F}$
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ (stabilité par passage au complémentaire)
3. Pour toute suite $(A_n)_{n \geq 1}$ d'éléments de $\mathcal{F}$, $\bigcup_{n=1}^{+\infty} A_n \in \mathcal{F}$ (stabilité par réunion dénombrable).

Définitions. On appelle ensemble des parties de $\mathcal{F}$ l'ensemble de tous les sous-ensembles de Ω, y compris $\emptyset$ et Ω lui-même. On le note $\mathcal{P}(\Omega)$.

Remarque. $\mathcal{P}(\Omega)$ est une tribu sur Ω (c'est la plus grande), $\mathcal{F} = \{\emptyset \cap \Omega\}$ est une tribu (c'est la plus petite).

Définition. Soit un espace Ω muni d'une tribu $\mathcal{F}$ (espace "probabilité"). On appelle événement tout élément de $\mathcal{F}$ c'est à dire toute partie de Ω appartenant à $\mathcal{F}$.

Remarque. En pratique quand Ω est au plus dénombrable, on verra qu'on peut définir une probabilité pour toute partie de Ω, et on pourra donc prendre $\mathcal{F} = P(\Omega)$.

Exemples de tribus. Soit $\Omega = \{a, b, c\}$.
$\mathcal{F}_1 = \{\emptyset, \{a\}, \{b, c\}, \Omega\}, \mathcal{F}_2 = \{\emptyset, \{b\}, \{a, c\}, \Omega\}$.
Plus généralement , $\forall A \subset \Omega, \mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ est une tribu $\mathcal{G} = \{\emptyset, \{a\}, \{b\}, \Omega\}$ n'est pas une tribu.

Tribu borélienne. Soit $\Omega = \mathbb{R}$ et $\mathcal{M} = \{]-\infty, t], t \in \mathbb{R}\}$. La plus petite tribu contenant $\mathcal{M}$ est appelée tribu borélienne sur $\mathbb{R}$ et on la note $\mathcal{B}(\mathbb{R})$. Les éléments de $\mathcal{B}(\mathbb{R})$ sont appelés "boréliens".

Propriété élémentaires. Soit $\mathcal{F}$ une tribu.

1. $\emptyset \in \mathcal{F}$ (car $\emptyset = \Omega^c$)
2. Soit $(A_n)_{n\geq 1}$ une suite d'éléments de $\mathcal{F}$. Alors $\bigcap\limits_{n=1}^{\infty} A_n \in \mathcal{F}$ (stabilité par intersection dénombrable).

   Démonstration. $\bigcap\limits_{n=1}^{\infty} = (\bigcup\limits_{n=1}^{\infty} A_n^c)^c$ □

3. Soit A et B dans $\mathcal{F}$. Alors $A \setminus B = A \cap B^c \in \mathcal{F}$ (car $A \cap B^c = A \cap B^c \cap \Omega \cap \Omega \cap ... \cap \Omega ...$).

5 Probabilité

Définition. Soit Ω un espace muni d'une tribu d'événements $\mathcal{F}$. On appelle probabilité sur $(\Omega, \mathcal{F})$ toute application :

$$P : \begin{cases} \mathcal{F} \to [0,1] \\ A \in \mathcal{F} \mapsto P(A) \end{cases}$$

et vérifiant les axiomes suivants :

1. $P(\Omega) = 1$
2. Soit $(A_i)_{i\geq 1}$ une suite d'événements tels que $A_i \cap A_j = \emptyset, \forall i \neq j$ (on parle d'évenements 2 à 2 disjoints) alors $P(\bigcup\limits_{i=1}^{\infty} A_i) = \sum\limits_{i=1}^{\infty} P(A_i)$ (propriété de $\sigma - additive$).

Définition. Le triplet $(\Omega, \mathcal{F}, P)$ est appelée **espace probabilisé**.

Première conséquence. Probabilité sur un espace au plus dénombrable. Dans le cas $\Omega = \{\omega_1, ..., \omega_n, ...\}$ on peut définir une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ en définissant les valeurs de $P(\omega_i), \forall i$.
En effet, soit $A \in \mathcal{P}(\Omega)$, alors on peut toujours écrire

$$A = \bigcup_{\{i, \omega_i \in A\}} \{\omega_i\}$$

Or $\{\omega_i\} \cap \{\omega_j\} = \emptyset \forall i \neq j$, donc d'après l'axiome 2

$$P(A) = \sum_{i, \omega_i \in A} P(\omega_i)$$

Dans le cas où Ω est fini ou dénombrable, il est donc naturel de poser $\mathcal{F} = \mathcal{P}(\Omega)$.

Exemples de probabilités .

1. $\mathcal{E}_1 : P(1) = 1/6, ..., P(6) = 1/6$
2. $\mathcal{E}_2 : P_2(F) = p, P_2((P,F)) = (1-p) \times p, P_2((P, ..., P, F)) = (1-p)^{n-1} \times p, P_2((P, ..., P, ...)) = 0$
3. Pour l'expérience $\mathcal{E}_3$ (génération de nombres aléatoires), $\Omega_3 = [0,1]$, n'est pas dénombrable. On pose $\mathcal{F} = \mathcal{B}(\mathbb{R})_{|[0,1]} = \{A \cap [0,1], A \in \mathcal{B}(\mathbb{R})\}$. On peut montrer qu'il existe une (unique) probabilité P sur $(\Omega_3, \mathcal{F}_3)$ vérifiant

$$\forall (a,b) \in [0,1] \times [0,1], P([a,b]) = b - a \qquad \text{(loi uniforme)}$$

On voit bien dans ce cas que $\forall x \in [0,1], P(X=x)=0$ (voir preuve plus tard) et qu'on ne peut définir P par la donné de $P(X=x), x \in \Omega$.

Remarque. X est une variable aléatoire de la loi uniforme sur $[0,1]$. La densité est la fonction

$$f(x) = \mathbb{1}_{[0,1]}(x) = \begin{cases} 1 \text{ si } x \in [0,1] \\ 0 \text{ sinon} \end{cases}$$

Propriétés des probabilités

1. (a) $P(\emptyset) = 0$

   Démonstration. $\emptyset \cup \Omega = \Omega$ et $\emptyset \cap \Omega = \emptyset$ donc $P(\Omega) + P(\emptyset) = P(\Omega)$ donc $P(\emptyset) = 0$ □

   (b) $P(A^c) = 1 - P(A)$

   Démonstration. $A \cup A^c = \Omega$ et $A \cap A^c = \emptyset$ donc $P(A) + P(A^c) = P(\Omega) \Rightarrow P(A^c) = P(\Omega) - P(A) = 1 - P(A)$ □

2. Soit $A_1, ..., A_n$ n événements 2 à 2 disjoints. Alors $P(\bigcup\limits_{i=1}^{n} A_i) = \sum\limits_{i=1}^{n} P(A_i)$ (car $A_1 \cup ... \cup A_n = A_1 \cup ... \cup A_n \cup \emptyset \cup ... \cup \emptyset ...$)

3. $P(A \cup B) = P(A) + P(B \setminus A)$

   Démonstration. $A \cup (B \setminus A) = A \cup (B \cap A^c) = A \cup (B \cap A^c) = (A \cup B) \cap \overbrace{(A \cup A^c)}^{=\Omega} = A \cup B$ et $A \cap (B \setminus A) = A \cap B \cap A^c = \emptyset$ □

4. Si $A \setminus B$, alors $P(B \setminus A) = P(B) - P(A)$ (et donc $P(A) \leq P(B)$)

5. $P(B \setminus A) = P(B) - P(A \cap B)$

   Démonstration.

   $$\begin{aligned} B \setminus A &= B \cap A^c \\ &= (B \cap A^c) \cup (B \cap B^c) \\ &= B \cap (A^c \cup B^c) \\ &= B \cap (A \cap B)^c \\ &= B \setminus (A \cap B) \end{aligned}$$

   Or $A \cap B \subset B$, d'où $P(B \setminus A) = P(B \setminus (A \cap B)) \underbrace{=}_{\text{par } 4} P(B) - P(A \cap B)$ □

6. Donc $P(A \cup B) = P(A) + P(B \setminus A) = P(A) + P(B) - P(A \cap B)$

7. **Sous additivité.**

   $$P(\bigcup_{n=0}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$$

Démonstration. Si on pose $B_1 = A_1, \forall n \geq 2, B_n = A_n \setminus (A_1 \cup A_2 \cup ... \cup A_{n-1})$. On a $\bigcup\limits_{n=1}^{\infty} A_n = \bigcup\limits_{n=1}^{\infty} B_n$ et les événements $(B_n)_{n\geq 1}$ sont deux à deux distinct. D'autre part $\forall n, B_n \subset A_n$, donc

$$P(\bigcup_{n=1}^{\infty} A_n) = P(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} P(B_n) \leq \sum_{n=1}^{\infty} P(A_n)$$

□

8. **Monotonie.**

(a) Soit $(A_n)_{n\geq 1}$ avec $A_n \subset A_{n+1}$. Alors

$$P(\bigcup_{n=1}^{\infty} A_n) = \lim_{n\to+\infty} P(A_n)$$

Démonstration. Soit $B_n = A_n \setminus (A_1 \cup ... \cup A_{n-1})$ et $A_n = A_1 \cup ... \cup A_n = B_1 \cup ... \cup B_n, \forall n$ et $\bigcup\limits_{n=1}^{\infty} A_n = \bigcup\limits_{n=1}^{\infty} B_n$. Donc

$$\begin{aligned} P(\bigcup_{n=1}^{\infty} A_n) &= P(\bigcup_{n=1}^{\infty} B_n) \\ &= \sum_{n=1}^{\infty} P(B_n) \\ &= \lim_{n\to+\infty} \sum_{i=1}^{n} P(B_i) \\ &= \lim_{n\to+\infty} P(\bigcup_{i=1}^{n} B_i) \\ &= \lim_{n\to+\infty} P(B_1 \cup ... \cup B_n) \\ &= \lim_{n\to+\infty} P(A_n) \end{aligned}$$

□

(b) Soit $(A_n)_{n\geq 1}$ avec $A_{n+1} \subset A_n$. Alors

$$P(\bigcap_{n=1}^{\infty} A_n) = \lim_{n\to+\infty} P(A_n)$$

Démonstration. $A_n^c \subset A_{n+1}^c$ (suite croissante), $P(A_n) + P(A_n^c) = 1$. D'après 8.a.,

$$P(A_n^c) \to P(\bigcup_{n=1}^{\infty} A_n^c) \Leftrightarrow 1 - P(A_n) \to 1 - P(\bigcap_{n=1}^{\infty} A_n) \Leftrightarrow P(A_n) \to P(\bigcap_{n=1}^{\infty} A_n)$$

□

On peut utiliser 8.b. pour montrer que dans le cas de la probabilité sur $[0,1]$ donnée en exemple, on a bien $P(\{x\}) = P([x,x]) = 0, \forall x \in]0,1[$, on peut écrire $\{x\} = \bigcap\limits_{n=1} [x - \frac{1}{n}, x + \frac{1}{n}]$
$\Rightarrow P(\{x\}) = \lim\limits_{n\to+\infty} P([x - \frac{1}{n}, x + \frac{1}{n}]) = \lim\limits_{n\to+\infty} \frac{2}{n} = 0$

Autre application. Expérience $\mathcal{E}_2$ (on lance une pièce jusqu'à obtenir face).

$$P((P, ..., P)) = (\text{Probabilité de ne jamais obtenir face})$$

On a $(P, ..., P, ...) = \bigcap_{n\geq 1} A_n$ avec $A_n = (P, ..., P) =$ on obtient pile n fois de suite. On voit que $A_{n+1} \subset A_n$. D'où d'après 8.b.

$$\begin{aligned} P((P, ..., P, ...)) &= \lim_{n\to+\infty} P(A_n) \\ &= \lim_{n\to+\infty} (1-p)^n \\ &= 0 \ (p > 0) \end{aligned}$$

6 Indépendance et conditionnement

Intuitivement, 2 événements seront dit indépendants si la réalisation de l'un ne modifie par la probabilité de l'autre.

Définition. On dit que deux événements A et B sont indépendants pour la probabilité P si

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

Remarque. On voit que l'indépendance n'est pas une qualité intrinsèque des événements, elle dépend de la probabilité considérée.

Exemple. Soit $\Omega = \{1, 2, 3, 4, 5, 6\}$ muni de la tribu $\mathcal{P}(\Omega)$. Soit $A = \{1, 2\}$ et $B = \{2, 3\}$ 2 événements.
Soit P_1 et P_2 deux probabilités sur $(\Omega, \mathcal{P}(\Omega))$ défini par

i	1	2	3	4	5	6
$P_1(i)$	1/6	1/6	1/6	1/6	1/6	1/6
$P_2(i)$	1/6	1/6	1/3	1/9	1/9	1/9

Étudions l'indépendance de A et B pour P_1 et P_2. On a donc $P_1(A) = \frac{1}{3} = P_1(B)$ de plus $P_1(A \cap B) = \frac{1}{6} \neq \frac{1}{9}$ donc A et B ne sont pas indépendants pour P_1.
On a donc $P_2(A) = \frac{1}{3}$ et $P_2(B) = \frac{1}{2}$ de plus $P_2(A \cap B) = \frac{1}{6} = P_2(A)P_2(B)$ donc A et B sont indépendants pour P_2.

Proposition. Si A et B sont indépendants pour la probabilité P alors A et B^c, A^c et B, A^c et B^c sont indépendants pour P.

Démonstration. $P(A \cap B^c) + P(A \cap B) = P(A) \Leftrightarrow P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)\underbrace{(1 - P(B))}_{=P(B^c)}$ □

Indépendance d'une famille d'événements. Soit I un ensemble (ou une suite d'événements) d'indices fini ou dénombrable. On dit que les événements $(A_i)_{i\in I}$ sont indépendants (ou mutuellement indépendants) si $\forall p \geq 2, \forall i_1, ..., i_p \in I$ (avec $i_j \neq i_k \forall j \neq k$)

$$P(A_{i_1} \cap ... \cap A_{i_p}) = P(A_{i_1}) \times ... \times P(A_{i_p})$$

Exemple. Trois événements A, B, C sont (mutuellement) indépendants si $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, $P(B \cap C) = P(B)P(C)$

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

Définition. On voit que les événements $(A_i)_{i\in I}$ sont 2 à 2 indépendants si $\forall i \neq j, A_i$ et A_j sont indépendants.

Exemple fondamental : épreuves répétés. Soit $\mathcal{E}$ une expérience aléatoire caractérisée par l'espace probabilisée $(\Omega, \mathcal{F}, \mathcal{P})$.

On appelle une épreuve répétée n fois l'épreuve (l'expérience aléatoire) $\mathcal{E}^n$ caractérisé par l'espace probabilisé $(\Omega^n, \mathcal{F}^{\otimes n}, P^{\otimes n})$ où

- $\Omega^n = \{(\omega_1, ..., \omega_n), \omega_i \in \Omega, 1 \leq i \leq n\}$
- $\mathcal{F}^{\otimes n}$ est la plus petite tribu contenant les ensembles de la forme $A_1 \times ... \times A_n$ avec $A_i \in \mathcal{F}, \forall i$ (tribu produit).
- $P^{\otimes n}$ est l'unique probabilité sur $(\Omega^n, \mathcal{F}^{\otimes n})$ telle que $\forall A_1, ..., A_n \in \mathcal{F}$

$$P^{\otimes n}(A_1 \times ... \times A_n) = P(A_1) \times ... \times P(A_n) \qquad \text{(Probabilité produit)}$$

Propriété. Soit $\tilde{A}_1 = A_1 \times \Omega \times ... \times \Omega$ (j'obtiens un résultat donc A_n à la première occurrence)

$\tilde{A}_2 = \Omega \times A_2 \times \Omega \times ... \times \Omega$

$\vdots$

$\tilde{A}_n = \Omega \times \Omega \times ... \times \Omega \times A_n$

Alors on voit que $\tilde{A}_1 \cap ... \cap \tilde{A}_n = A_1 \times A_2 \times ... \times A_n$, d'où

$$P^{\otimes n}(\tilde{A}_1 \cap ... \cap \tilde{A}_n) = P(A_1) \times ... \times P(A_n) = P^{\otimes n}(\tilde{A}_1) \times ... \times P^{\otimes n}(\tilde{A}_n)$$

On voit donc que $\tilde{A}_1, ..., \tilde{A}_n$ sont indépendants pour $P^{\otimes n}$.

C'est pourquoi l'indépendance est souvent postulée dans ce cas de figure. Par exemple, si on lance un dé deux fois de suite, les événements relatifs au premier lancer son indépendnats de ceux relatifs au second lancer. Si $A = \{$le premier lancer est pair$\}$ et $B = \{$le premier lancer donne G$\}$, on admet que $P(A \cap B) = P(A).P(B)$

Définition. Soit $(\Omega, \mathcal{F}, P)$ unu espace probabilisé et $A \in \mathcal{F}$ avec $P(A) > 0$. On appelle **probabilité conditionnelle sachant** A l'application notée P_A et définie par $P_A : \begin{cases} \mathcal{F} \to [0,1] \\ B \mapsto P_A(B) = \dfrac{P(A \cap B)}{P(A)} \end{cases}$.

Remarque. On note aussi parfois $P_A(B) = P(B|A)$.

Remarque.

1. P_A est bien une probabilité
$$P_A(\Omega) = \frac{P(A \cap \Omega)}{P(A)} = \frac{P(A)}{P(A)} = 1$$

2. Soit $(B_n)_{n\geq 2}$ deux à deux disjoints

$$\begin{aligned}
P_A(\bigcup_{n=1}^{\infty} B_n) &= \frac{P((\bigcup\limits_{n=1}^{\infty} B_n) \cap A)}{P(A)} \\
&= \frac{P(\bigcup\limits_{n=1}^{\infty} (A \cap B_n))}{P(A)} \\
&= \frac{\sum\limits_{n=1}^{\infty} P(A \cap B_n)}{P(A)} \\
&= \sum_{n=1}^{\infty} \frac{P(A \cap B_n)}{P(A)} \\
&= \sum_{n=1}^{\infty} P_A(B_n)
\end{aligned}$$

Proposition. Si A et B deux événements avec $P(A) > 0$. Alors A et B indépendants si et seulement si $P_A(B) = P(B)$.

Démonstration. $P_A(B) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{P(A)P(B)}{P(A)} = P(B)$ (Si $P(B) > 0$, A et B indépendants $\Leftrightarrow P_B(A) = P(A)$) □

Formule des probabilités composées. Soit $A_1, ..., A_n \in \mathcal{F}$ tels que $P(A_1 \cap ... \cap A_{n-1}) > 0$. Alors

1. $P(\bigcap\limits_{i=1}^{n} A_i) = P(A_1 \cap ... \cap A_n) = P(A_1).P(A_2|A_1).P(A_3|A_1 \cap A_2) \times ... \times P(a_n|A_1 \cap ... \cap A_{n-1})$.
2. Si on a une suite d'événements $(A_i)_{i\geq 1}$, avec $P(\bigcap\limits_{i=1}^{n} A_i) > 0, \forall n$ alors

$$\begin{aligned}
P(\bigcap_{i=1}^{\infty} A_i) &= P(A_1) \times P(A_2|A_1) \times ... \times P(A_n|A_1 \cap ... \cap A_{n-1}) \times ... \\
&= \prod_{i=1}^{\infty} P(A_i|A_1 \cap ... \cap A_{i-1})
\end{aligned}$$

Démonstration. 1) est évident. Pour montrer 2) à partir de 1), on remarque que $P(\bigcap\limits_{i=1}^{\infty} A_i) = \lim\limits_{n\to+\infty} P(\bigcap\limits_{i=1}^{n} A_i)$ (propriété 8.b). □

Définition. Soit I fini ou dénombrable. On dit que $(A_i)_{i\in I}$ forme un système complet d'événements si et seulement si les 3 conditions suivantes sont vérifiées

— $\forall i \in I, A_i \neq \emptyset$

— $\forall i \neq j, A_i \cap A_j = \emptyset$

— $\bigcup_{i\in I} A_i = \Omega$

On dit que $(A_i)_{i\in I}$ forme une **partition** de Ω.

Théorème I.6. (formule des probabilités totales) Soit $(A_i)_{i\in I}$ un système complet d'événements d'un espace probabilisé $(\Omega, \mathcal{F}, P)$ tel que $\forall i \in I, P(A_i) > 0$, et soit $E \in \mathcal{F}$ un événement quelconque. Alors $P(E) = \sum_{i\in I} P(E|A_i)P(A_i) = \sum_{i\in I} P_{A_i}(E)P(A_i)$

Démonstration.

$$\begin{aligned}
P(E) &= P(\Omega \cap E)\\
&= P((\bigcup_i A_i) \cap E)\\
&= P(\bigcup_i (A_i \cap E))\\
&= \sum_i P(A_i \cap E)\\
&= \sum_i P(E|A_i).P(A_i)
\end{aligned}$$

□

Remarque. La formule reste vraie si on suppose que $P(\bigcup_{i\in I} A_i) = 1$ (sans nécessairement avec $\bigcup_{i\in I} A_i = \Omega$). En effet $P(E) = P((\bigcup A_i) \cap E) + P((\bigcup A_i)^c \cap E)$ or $P((\bigcup A_i)^c) = 1 - P(\bigcup A_i) = 0 \Rightarrow P((\bigcup A_i)^c \cap E) = 0$

Corollaire 5.1. (formule de Bayes) Soit $(A_i)_{i\in I}$ un système complet d'événements d'un espace probabilisé $(\Omega, \mathcal{F}, P)$ tel que $P(A_i) > 0, \forall i$. Soit $E \in \mathcal{F}$ avec $P(E) > 0$, alors $\forall i \in I, P_E(A_i) = P(A_i|E) = \dfrac{P(A_i \cap E)}{P(E)} = \dfrac{P(E|A_i).P(A_i)}{\sum_{j\in I} P(E|A_j)P(A_j)}$

Indépendance conditionnelle. Soit E un événement avec $P(E) > 0$. On dit que 2 événements A et B sont indépendants conditionnellement à E (par la probabilité p) si A et B sont indépendants pour la probabilité conditionnelle P_E, c'est à dire si $P_E(A \cap B) = P_E(A).P_E(B) \Leftrightarrow P(A \cap B|E) = P(A|E).P(B|E)$

Exemple. On dispose de deux pièces, une équilibrée et une telle que $P(\text{face}) = \frac{3}{4}$. On sélectionne une des 2 pièces au hasard et on la lance 2 fois. Soit F_i l'événement "on obtient face au lancer i" où $i = 1, 2$. Soit P_i lévénement on a sélectionné la pièce $i = 1, 2$. Soit P_i l'événement on a sélectionné la pièce i" où $i = 1, 2$. Alors F_1 et F_2 sont indépendants conditionnement à P_1 ou P_2.

$$P(F_1 \cap F_2|P_1) = P(F_1|P_1).P(F_2|P_1) = \frac{1}{4}$$

$$P(F_1 \cap F_2|P_2) = P(F_1|P_2).P(F_2|P_2) = \frac{9}{16}$$

Est ce que F_1 et F_2 sont indépendants (pour la probabilité P)

$$P(F_1 \cap F_2) \stackrel{?}{=} P(F_1)P(F_2)$$

$P(F_1) = P(F_1|P_1).P(P_1) + P(F_1|P_2)P(P_2) = \frac{1}{2}\frac{1}{2} + \frac{3}{4}\frac{1}{2} = \frac{5}{8} = P(F_2)$

$P(F_1 \cap F_2) = P(F_1 \cap F_2|P_1) \times P(P_1) + P(F_1 \cap F_2|P_2).P(P_2) = \frac{1}{4}.\frac{1}{2} + \frac{9}{16}.\frac{1}{2} = \frac{13}{52} \neq (\frac{5}{8})^2 = P(F_1) \times P(F_2)$

Donc F_1 et F_2 ne sont pas indépendants pour la probabilité P.

$$\begin{aligned}
P(P_2|F_1) &= \frac{P(F_1|P_2).P(P_2)}{P(F_1|P_1)P(P_1) + P(F_1|P_2)P(P_2)} \\
&= \frac{\frac{3}{4} \times \frac{1}{2}}{\frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{3}{4}} \\
&= \frac{\frac{3}{4}}{\frac{1}{2} + \frac{3}{4}}
\end{aligned}$$

$P(P_2|F_1 \cap F_2) = \dfrac{9/16 \times 1/2}{1/4 \times 1/2 + 9/16 \times 1/2} = 9/13$

Deuxième partie

Variables aléatoires discrètes

1 Définition, loi de probabilité a une variable aléatoire

Définition. On appelle variable aléatoire (réelle) sur un espace probabilité $(\Omega, \mathcal{F}, P)$ toute fonction X de Ω dans $\mathbb{R}$ telle que $\forall A \in \mathcal{B}(\mathbb{R})$ (tribu borélienne),

$$X^{-1}(A) = \{\omega \in \Omega / X(\omega) \in A\} = \{x \in A\} \in \mathcal{F}$$

Si de plus $X(\Omega)$ est **fini ou dénombrable**, alors on dit que X est une **variable aléatoire discrète**.

Définition. On appelle variable aléatoire discrète (réelle) sur un espace probabilisé $(\Omega, \mathcal{F}, P)$ toute fonction X de $\Omega \to \mathbb{R}$ telle que $X(\Omega)$ est fini ou dénombrable (Rappel : $X(\Omega) = \{x \in \mathbb{R} | \exists \omega \in \Omega\}$ et $\forall x \in X(\Omega)$

$$X^{-1} = (\{x\}) = \{\omega | X(\omega) = x\} = \{X = x\}$$

est un élément de la tribu $\mathcal{F}$.
On note $P(X = x)$ ou parfois $P_X(\{x\})$ la quantité

$$P(X^{-1}(\{x\})) = P(\{x \in \Omega | X(\omega) = x\})$$

Plus généralement, on écrit $P(X \in A) = P_X(A) = P(X^{-1}(A)) = P(\{\omega \in \Omega\} \, | X(\omega) \in A)$.

Définition. On appelle P_X **la loi de probabilité** de la variable aléatoire X (ou plus simplement la loi de X).

Théorème II.1. P_X est une probabilité sur $X(\Omega)$ muni de la tribu $\mathcal{P}(X(\Omega))$

Démonstration.

— $P_X(X(\Omega)) = P(X \in X(\Omega)) = P(X^{-1}(X(\Omega)) = P(\Omega) = 1$
— Soit $(A_n)_{n\geq 1}$ une suite de partie de $X(\Omega)$, 2 à 2 disjointes. On a $P_X(\bigcup\limits_{n=1}^{\infty} A_n) = P(X^{-1}(\bigcup_{n=1}^{\infty} A_n))$. Or $X^{-1}(\bigcup\limits_{n=1}^{\infty} A_n) = \bigcup\limits_{n=1}^{\infty} X^{-1}(A_n)$ car

$$\begin{aligned}
\omega \in X^{-1}(\bigcup_{n=1}^{\infty} A_n) &\Leftrightarrow X(\omega) \in \bigcup_{n=1}^{\infty} A_n \\
&\Leftrightarrow \exists n \geq 1, X(\omega) \in A_n \\
&\Leftrightarrow \exists n \geq 1, \omega \in X^{-1}(A_n) \\
&\Leftrightarrow \omega \in \bigcup_{n=1}^{\infty} X^{-1}(A_n)
\end{aligned}$$

□

Donc $P_X(\bigcup_{n=1}^{\infty} A_n) = P(\bigcup_{n=1}^{\infty} X^{-1}(A_n))$. Enfin si $i \neq j, X^{-1}(A_i) \cap X^{-1}(A_j) = \emptyset$ car on ne peut avoir $X(\omega) \in A_i$ et $X(\omega) \in A_j$ car $A_i \cap A_j = \emptyset$.
Finalement $P_X(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{+\infty} P(X^{-1}(A_n)) = \sum_{n=1}^{+\infty} P_X(A_n)$.

Réciproque. Soit X une application d'un espace probabilisable $(\Omega, \mathcal{F})$ à valeurs dans $\mathbb{R}$ telle que $X(\Omega)$ est fini ou dénombrable. Posons $X(\Omega) = \{x_1, ..., x_n, ...\}$ et sait d'autre part $(p_n)_{n \geq 1}$ une suite de nombres réels positifs avec $\sum_{n=1}^{+\infty} p_n = 1$. Alors il existe une probabilité. P sur $(\Omega, \mathcal{F})$ telle que, $\forall x_n$,

$$P(X = x_n) = p_n$$

2 Indépendance de variables aléatoires

Définition. Soit X et Y deux variables aléatoires discrètes. On dit que X et Y sont indépendantes si et seulement si $\forall (x, y) \in X(\Omega) \times Y(\Omega), P(X = x, Y = y) = P(X = x).P(Y = y)$.

Propriétés.

1. Si X et Y sont deux variables aléatoires discrètes indépendantes alors $\forall A \in \mathcal{P}(X(\Omega))$ et $\forall B \in \mathcal{P}(Y(\Omega)), P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$.

   Démonstration.

$$\begin{aligned} P(X \in A, Y \in B) &= \sum_{x \in A, y \in B} P(X = x, Y = y) \\ &= \sum_{x \in A, y \in B} P(X = x)P(Y = y) \\ &= \sum_{x \in A} P(X = a) \sum_{y \in B} P(Y = y) \\ &= P(X \in A)P(Y \in B) \end{aligned}$$

□

2. Soit X et Y deux variables aléatoires discrètes indépendantes et $f : X(\Omega) \to \mathbb{R}$ et $g : Y(\Omega) \to \mathbb{R}$ deux applications. Alors les variables aléatoires $f(X)$ et $g(Y)$ sont indépendantes (et elles sont discrètes).

   Démonstration. On a $P(f(X) = u, g(Y) = v) = P(X \in f^{-1}(\{u\}), Y \in g^{-1}(\{v\})) = P(X \in f^{-1}(\{u\})) \times P(Y \in g^{-1}(\{v\}))$ (avec la propriété 1)

$$P(f(X) = u) = P(g(Y) = v)$$

□

Définition. On dit que que n variable aléatoire discrète $X_1, ..., X_n$ sont (mutuellement) indépendants si et seulement $\forall (x_1, ..., x_n) \in X_1(\Omega) \times ... \times X_n(\Omega)$.

$$P(X_1 = x_1, ..., X_n = x_n) = P(X_1 = x_1) \times ... \times P(X_n = x_n)$$

Définition. Soit I un ensemble fini ou dénombrable. On dit que les variables aléatoires (discrètes) $(X_i)_{i \in I}$ sont (mutuellement) indépendantes si et seulement si $\forall p \geq 2, \forall i_1, ..., i_p, (i_j \neq i_k)$ $X_{i_1}, ..., X_{i_p}$ sont (mutuellement) indépendants.

Définition. On dit que n variable aléatoires $X_1, ..., X_n$ sont identiquement distribuées si et seulement si $P_{X_1} = ... = P_{X_n}$, c'est à dire si elles ont toutes la même loi de probabilité. De même, si I est fini ou dénombrable on dit que les variables aléatoires. $X_i, i \in I$ sont identiquement distribuées $\forall i, j \in J, P_{X_i} = P_{X_j}$

3 Exemples de lois de probabilité sur des ensemble dénombrables

3.1 Loi de Poisson, convergence de la loi binomiale vers la loi de Poisson

Définition. On dit que la loi de Poisson variable aléatoire discrète X à valeurs dans $\mathbb{N}$ suit une loi de Poisson de paramètre $\lambda > 0$ si $\forall k \in \mathbb{N}$

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

On note $X \sim \mathcal{P}(\lambda)$ La loi de Poisson est fréquemment (mais pas uniquement) utilisée pour modéliser un nombre de phénomènes se produisant pendant un intervalle de temps donné. Par exemple, on considère la situation suivante sur un intervalle de temps $[0, T]$ Des personnes se présentent l'une après l'autre à un guichet. Soit N_T la variable aléatoire représentant le nombre de personnes qui se sont présentées au guichet sur $[0, T]$. Plus généralement, $\forall t \in [0, T]$, on notera N_t le nombre de personnes qui se sont présentées au guichet pendant l'intervalle de temps $[0, t]$. On supposera que $N_0 = 0$ et on fera en outre les hypothèses suivantes

1. $\forall s_1, s_2, t_1, t_2$ avec $s_1 < s_2 \leq t_1 < t_2$, $N_{t_2} - N_{t_1}$ et $N_{s_2} - N_{s_1}$ sont des variables aléatoires indépendants (hypothèse d'accroissements indépendants)
2. $\forall s < t$, la loi de probabilité de $N_t - N_s$ ne dépend que de l'écart entre $t - s$ ("stationnarité")
3. (a) $P(N_{t+h} - N_t = 1) = \lambda h + o(h)$ quand $h \to 0$.
   (b) $P(N_{t+h} - N_t > 1) = o(h)$ quand $h \to 0$. Donc $P(N_{t+h} - N_t \geq 1) = \lambda h + o(h)$ quand $h \to 0$. $P(N_{t+h} - N_t = 0) = 1 - \lambda h + o(h)$ quand $h \to 0$.

Posons ici $h = \dfrac{T}{n}$ (on divise l'intervalle $[0, T]$ en un intervalles de longueur $\dfrac{T}{n}$). Alors $P(N_{\frac{iT}{n}} - N_{(i-1)\frac{T}{n}} \geq 1) = p_n$ avec $\lim\limits_{n \to +\infty} \frac{p_n}{T/n} = \lambda \Leftrightarrow \lim\limits_{n \to +\infty} np_n = \lambda T$.
D'autre part $P(N_{\frac{iT}{n}} - N_{(i-1)\frac{T}{n}} > 1) = o(\frac{T}{n})$

$$\lim_{n \to +\infty} nP(N_{\frac{iT}{n}} - N_{(i-1)\frac{T}{n}} > 1) = 0$$

Soit $\forall 1 \leq 1 \leq n$

$$U_i = \begin{cases} 1 \text{ si une personne ou plus se présente au guichet dans l'intervalle }](i-1)\frac{T}{n}, i\frac{T}{n}] \\ 0 \text{ sinon} \end{cases}$$

Les variables U_i sont indépendantes et identiquement distribuées de loi de Bernoulli $\mathcal{B}(p_n)$. Donc $X_n = \sum_{i=1}^{n} U_i$ suit une loi Binomiale $\mathcal{B}(n, p_n)$. Montrons que $P(N_T \neq X_n) \to 0$. En effet

$$P(N_T \neq X_n) = P(\left\{N_{\frac{T}{n}} - N_0 > 1\right\} \cup \left\{N_{\frac{2T}{n}} - N_{\frac{T}{n}} > 1\right\} \cup ... \cup \left\{N_T - N_{(n-1)\frac{T}{n}} > 1\right\}) \leq nP(N_{\frac{T}{n}} - N_0 > 1) \to 0$$

Donc

$$P(X_n = k) = \underbrace{P(X_n = k, N_T = X_n)}_{P(N_T = k, N_T = X_n)} + \underbrace{P(X_n = k, N_T \neq x_n)}_{\to 0}$$

$$\begin{aligned} \lim_{n \to +\infty} P(X_n = k) &= \lim_{n \to +\infty} P(N_T = k, N_T = X_n) \\ &= \lim_{n \to +\infty} [P(N_T = k) - P(N_T = k, N_T \neq X_n)] \\ &= P(N_T = k) \end{aligned}$$

Calculons $\lim_{n \to +\infty} P(X_n = k) = \lim_{n \to +\infty} C_n^k p_n^k (1-p_n)^{n-k} \approx \frac{n(n-1) \times ... \times (n-k+1)}{k!} \frac{(\lambda T)^k}{n^k} (1-p_n)^{n-k}$ (car $p_n \sim \frac{\lambda T}{n}$).

On voit que $\frac{n(n-1) \times ... \times (n-k+1)}{n^l k} \to 1$. De plus

$$\begin{aligned} (1-p_n)^{n-k} &= \exp((n-k)ln(1-p_n)) \\ &= \exp((n-k)(-p_n + o(p_n)) \end{aligned}$$

Comme $\frac{o(p_n)}{\lambda T/n} \to 0$.

On a $(n-k)o(p_n) \to 0$ et $\lim(n-k)p_n = \lim_{n \to +\infty} np_n = \lambda T$. Donc finalement $(1-p_n)^{n-k} \to e^{-\lambda T}$. D'où $\lim_{n \to +\infty} P(X_n = k) = e^{-\lambda T} \frac{(\lambda T)^k}{k!} = P(N_T = k)$ (loi de Poisson de paramètre $\lambda T \to N_t \sim \mathcal{P}(\lambda T)$)

Théorème II.2. Soit $\lambda > 0$ et $(X_n)_{n \geq 1}$ une suite de variable aléatoire de la binomiale $\mathcal{B}(n, p_n)$, où p_n est une suite tel que $\lim_{n \to +\infty} np_n = \lambda$. Alors , $\forall k \in \mathbb{N}$, $\lim_{n \to +\infty} P(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}$.

En pratique, on peut approcher la loi binomiale par la loi de Poisson dès que $\begin{cases} p < 0,1 \\ n > 30 \\ \lambda = np < 16 \end{cases}$ (où $np(1-p) < 10$).

Théorème II.3. Soit X et Y deux variables aléatoires indépendantes avec aux $X \sim P(\lambda)$ et $Y \sim P(\mu)$. Alors $X + Y \sim P(\lambda + \mu)$

Démonstration. $\forall n \in \mathbb{N}, P(X+Y=n) = \sum_{k=0}^{n} P(X=k, Y=n-k) = \sum_{k=0}^{n} P(X=k)P(Y=n-k) = \sum_{k=0}^{n} e^{-\lambda} \frac{\lambda^k}{k!} e^{-\mu} \frac{N^{n-k}}{(n-k)!} = e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \lambda^k \mu^{n-k} = e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^n}{n!}$ (binôme de Newton) $\square$

3.2 Loi géométrique

Considérons l'une des situations suivantes.

1. Une urne contient une proportion p de boules blanches (et une proportion $1-p$ de boules noires). On effectue des tirages avec remise jusqu'à ce qu'on obtienne une boule blanche.
2. On lance une pièce jusqu'à obtenir "face".
3. On lance un dé jusqu'à obtenir un "6".

On définit la variable aléatoire X à valeurs dans $\mathbb{N}^* \cup \{\infty\}$ par X le numéro du tirage et on tire la première boule blanche ($X = +\infty$ si on ne tire jamais de boule blanche). Posons $\forall i \geq 1$

$$X_i = \begin{cases} 1 \text{ si on tire une boule blanche au tirage } i \\ 0 \text{ sinon} \end{cases}$$

La suite $(X_i)_{i\geq 1}$ est une suite de variable aléatoire indépendante et identiquement distribuées de la loi de Bernoulli $\mathcal{B}(p)$. On a donc $P(X_i = 1) = p \Leftrightarrow P(X_i = 0) = 1 - p, \forall i \geq 1$. On peut écrire

$$\begin{aligned} P(X = 1) &= P(X_1 = 1) = p \\ P(X = 2) &= P(X_1 = 0).P(X_2 = 1) = (1-p)p \\ &\vdots \\ P(X = n) &= P(X_1 = 0, ..., X_{n-1} = 0, X_n = 1) = P(X_1 = 0) \times ... \times P(X_{n-1} = 0).P(X_n = 1) = (1-p)^{n-1}.p \end{aligned}$$

Propriété. $P(X > n) = P(X_1 = 0, ..., X_n = 0) = (1-p)^n \Rightarrow P(X \leq n) = 1 - (1-p)^n$.
On en déduit la valeur de $P(X = +\infty)$ car

$$\{X = +\infty\} = \bigcap_{n\geq 1} \{X > n\}$$

est donc $P(X = +\infty) = \lim\limits_{n\to+\infty} P(X > n) = \lim\limits_{n\to+\infty} (1-p)^n = \begin{cases} 0 \text{ si } p > 0 \\ 1 \text{ si } p = 0 \end{cases}$.
On dit que X suit une loi géométrique de paramètre p et on note $X \sim \mathcal{G}(p)$.

Remarque. On appelle parfois loi géométrique, la loi de la variable aléatoire $\tilde{X}$ égale au nombre de tirages (ou boule noires) **avant** l'apparition de la première boule blanche. Alors $\tilde{X}$ est une variable aléatoire à valeurs dans $\mathbb{N} \cup \{+\infty\}$. On voit que $\tilde{X} = X - 1$. Donc $P(\tilde{X} = n) = P(X = n+1) = (1-p)^n.p$. On note parfois $\tilde{X} \sim \mathcal{G}_{\mathbb{N}}(p)$.

3.3 Loi de Pascal (ou binomiale négative).

C'est une généralisation de la loi géométrique. Considérons une urne contenant une proportion p de boule blanche. On effectue des tirages avec remise jusqu'à l'apparition de la r-ième boule blanche. Soit X_r le n^c du tirage où on tire une boule blanche pour la r-ième fois. ($r = 1 \to$ loi géométrique). Déterminons la loi de probabilité de X_r.

$$X_r(\Omega) = \{r, r+1, ...\} \cup \{\infty\}$$

$\forall n \geq r, P_{X_r}(n) = P(X_r = n) = P(X_1 + ... + X_{n-1} = r-1, X_n = 1)P(X_1 + ... + X_{n-1} = r-1).P(X_n = 1)$.
$X_1 + ... + X_{n-1} \sim \mathcal{B}(n-1, p)$

$$\begin{aligned} \Leftrightarrow P(X_r = n) &= P(\mathcal{B}(n-1,p) = r-1) \\ &= C_{n-1}^{r-1} p^{r-1}(1-p)^{n-r}.p \\ &= C_{n-1}^{r-1} p^{r}(1-p)^{n-r} \end{aligned}$$

On dit que X_r suit une loi de Pascal (ou binomiale négative) de paramètre r et p.

$$X_r \sim Pa(r,p)$$

Posons $U_i = \begin{cases} 1 \text{ si on tire une boule blanche au tirage} \\ 0 \end{cases}$

Propriété.

1. $P(X_r > n) = P(U_1 + ... + U_n < r) = P(\mathcal{B}(n,p) < r)$
2. On pose $Y_1 = X_1$, $Y_2 = X_2 - X_1$, ... , $Y_r = X_r - X_{r-1}$. Alors les va $(Y_i)_{1 \leq i \leq r}$ sont indépendantes et identiquement distribuées de la loi géométrique de paramètre p. En résumé, la loi de Pascal de paramètre r et p sont obtenue. Comme la loi de probabilité de la somme de r variable aléatoire indépendantes et identiquement distribuées de la loi géométrique de paramètre p.

Remarque. Une conséquence de ce résultat est qu'étant données 2 variables aléatoires X et Y indépendantes, avec $X \sim Pa(r,p)$ et $Y \sim Pa(r',p)$, alors $X + Y \sim Pa(r + r', p)$

Remarque. On appelle parfois **la loi binomiale négative** la loi de la variable aléatoire Z_r égale au nombre de boules noires tirées avant l'apparition de la r-ième boule blanche. Ainsi $Z_r = X_r - r$ et donc $\forall n \geq 0$

$$P(Z_r = n) = P(X_r = n + r) = C_{n+r-1}^{r-1} p^r (1-p)^n$$

4 Fonctions de répartition

Définition. Soit X une variable aléatoire réelle. On appelle fonction de répartition de X et on note F_X la fonction :

$$F_X = \begin{cases} \mathbb{R} \to [0,1] \\ x \mapsto F_X(x) = P(X \leq x) \end{cases}$$

Propriétés.

1. F_X est croissante (au sens large)

   Démonstration. $x_1 \leq x_2 \Rightarrow \{X \leq x_1\} \subset \{X \leq x_2\} \Rightarrow P(X \leq x_1) \leq P(X \leq x_2)$ □

2. $\lim\limits_{x \to +\infty} F_X(x) = 1.$

   Démonstration. On a $\bigcup\limits_{n \geq 1} \{X \leq n\} = \{X \in \mathbb{R}\}$. D'où si X est à valeurs dans $\mathbb{R}$, on a $\lim\limits_{n \to +\infty} P(X \leq n) = P(X \in \mathbb{R}) = 1$ (propriété 8.a). □

   Remarque. Si X est à valeurs dans $\overline{\mathbb{R}} = [-\infty, +\infty]$ (fermeture de $\mathbb{R} = \mathbb{R} \cup \{-\infty, +\infty\}$), avec $P(x = -\infty) = \alpha$ et $P(X = +\infty) = \beta$. Alors $\bigcup\limits_{n \geq 1} \{X \leq\} \cup \{X = +\infty\} = \{X \in \overline{\mathbb{R}}\} = \Omega$. D'où $\lim\limits_{n \to +\infty} P(X \leq n) + P(X = +\infty) = 1 \Rightarrow \lim\limits_{n \to +\infty} P(X \leq n) = 1 - \beta$
   $\omega \in \bigcup\limits_{n=1}^{\infty} \{X \leq n\} \Leftrightarrow \exists n \geq 1$ tel que $X(\omega) \leq n$

3. $\lim\limits_{x \to -\infty} F_X(x) = 0$

   Démonstration. Si X à valeurs dans $\mathbb{R}$, $\bigcap\limits_{n \geq 1} \{X \leq n\} = \emptyset \Rightarrow \lim\limits_{n \to +\infty} P(X \leq -n) = P(\emptyset) = 0$ (d'après la propriété 8.b) □

   Remarque. si $P(X = -\infty) = \alpha$. Alors $\bigcap\limits_{n \geq 1} \{X \leq -n\} = \{X = -\infty\} \Rightarrow \lim\limits_{n \to +\infty} P(X \leq -n) = P(X = -\infty) = \alpha$

4. $\forall x \in \mathbb{R}$, F_X est continue à droite au point x.

   Démonstration. $\{X \leq x\} = \bigcap\limits_{n \geq 1} \left\{X \leq x + \dfrac{1}{n}\right\}$ donc $P(X \leq x) = \lim\limits_{n \to +\infty} P(X \leq x + \dfrac{1}{n} \Rightarrow F_X(x) = \lim\limits_{n \to +\infty} F_X(x + \dfrac{1}{n}) \forall x \in \mathbb{R}$, la fonction F_X admet une limite à gauche et $\lim\limits_{\epsilon \to 0^+} F_X(x - \epsilon) = P(X < x)$
   $\{X < x\} = \bigcup\limits_{n \geq 1} \left\{X \leq x - \dfrac{1}{n}\right\} \Rightarrow P(X < x) = \lim\limits_{n \to +\infty} P(X \leq x - \dfrac{1}{n}) = \lim\limits_{n \to +\infty} F_X(x - \dfrac{1}{n})$. On dit aussi F_X est une fonction continue à droite et limitée à gauche ("càdlàg") □

4.1 Lien entre la loi de probabilité et fonction de répartition (variable aléatoire discrète)

Etant donnée une variable discrète X avec $\forall x \in X(\Omega)$,

$$P_X(x) = P(X = x) \qquad \text{(loi de probabilité)}$$

On voit que $F_X(x) = P(X \leq x) = \sum\limits_{t \in]-\infty, x] \cap X(\Omega)} P(X = t)$.
Réciproquement, $\forall x \in X(\Omega)$, on peut écrire

$$P_X(x) = P(X = x) = P(X < x) - P(X < x) = F_X(x) - \lim_{x \to +\infty} F_X(x - \frac{1}{n})$$

Exemple. Soit X le résultat du lancer d'un dé équilibré (loi uniforme discrète sur $[1, ..., G]$), $P_X(i) = \frac{1}{G}, \forall i$. On voit que $\{F_X(x) = c, \forall x < 1\}$.

4.2 Rappel : fonction en escalier

Définition (subdivision d'un segment). Soit $a, b \in \mathbb{R}$, avec $a < b$. On appelle subdivision du segment $[a, b]$ toute famille **finie** de réels $\sigma = (a_0, a_1, ..., a_n)$ telle que $a = a_0 < a_1 < a_2 < ... < a_{n-1} < a_n = b$

Définition (fonction en escalier). Une fonction $f : [a, b] \to \mathbb{R}$ est dite en escalier s'il existe une subdivision $\sigma = (a_0, a_1, ..., a_n)$ de $[a, b]$ telle que f soit constante sur chaque intervalle $]a_{i-1}, a_i[, 1 \leq i \leq n$.

Proposition.

1. Soit X une variable aléatoire discrète à valeurs dans $X(\Omega)$ fini. Alors sa fonction de répartition est une fonction en escalier.
2. Si X est une variable aléatoire discrète à valeurs dans $X(\Omega)$ dénombrable (infini) et si de plus il existe. $\phi : \mathbb{N} \to X(\Omega)$ (ou plus générallement $\phi : \mathbb{Z} \to X(\Omega)$) telle que $\phi(n+1) > \phi(n)$ et $\lim_{n \to +\infty} \phi(n) = +\infty$ alors la fonction de répartition F_X est une fonction en escalier (sur tout intervalle $[a, b]$).

Remarque. Si $\lim_{n \to +\infty} \phi(n) = b \in \mathbb{R}$ alors $\forall m, F_X$, n'est pas en escalier sur $[b - \frac{1}{m}, b]$

Contre exemple remarquable. Soit X et Y deux variables aléatoires indépendants de loi géométrique de paramètre $p \in]0, 1[$ et posons $Z = \frac{X}{Y}$. Alors $\forall s \in \mathbb{Q}_+^*, P(Z = s) > 0$ et de plus $\forall 0 < a < b, \exists s \in \mathbb{Q}_+^* \cap]a, b[$. Donc F_Z ne peut être constante sur aucun intervalle $]a, b[$, et finalement F_Z n'est pas en escalier (sur aucun intervalle $]a, b[$).

Proposition. Si la fonction de répartition F_X d'une variable aléatoire X est continue alors X n'est pas discrète

Démonstration. supposons que $X(\Omega)$ est un ensemble fini ou dénombrable. Comme F_X est continue, alors $\forall x \in X(\Omega)$

$$P(X = x) = P(X \leq x) - P(X < x) = 0$$

$$P(X = x) = F_X(x) - \lim_{\epsilon \to 0^+} F_X(x - \epsilon) = F_X(x)$$

D'où une contradiction (en effet si X est discrète) on doit avoir

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

□

5 Vecteur aléatoires, couples de variables aléatoires

Définition. On appelle vecteur aléatoire de dimension n sur $(\Omega, \mathcal{F}, P)$ toute application $X = (X_1, ..., X_n)$ de $\Omega \to \mathbb{R}^n$ où $\forall 1 \leq i \leq n$, X_i est une variables réelle. Dans le cas où les variable aléatoires X_i sont discrets ($\forall i, X_i(\Omega)$ fini ou dénombrable), on parlera de vecteur aléatoire discret. Dans le cas où $n = 2$, étant données deux variables aléatoires discretes. Dans le cas où $n = 2$, étant données deux variables aléatoires X et Y, on dit que (X, Y) est un couple de variable aléatoire.

5.1 Loi de probabilité d'un vecteur aléatoire

Soit $(X_1, ..., X_n)$ un vecteur aléatoire discret. On appelle loi de probabilité du vecteur $(X_1, ..., X_n)$ la probabilité sur $X_1(\Omega) \times ... \times X_n(\Omega)$ muni de la tribu $\mathcal{P}(X_1(\Omega) \times ... \times X_n(\Omega))$ notée $P_{X_1,...,X_n}$ est définie par $\forall (x_1, ..., x_n) \in X_1(\Omega) \times ... \times X_n(\Omega)$

$$\begin{aligned} P_{X_1,...,X_n}(x_1, ..., x_n) &= P((X_1, ..., X_n) = (x_1, ..., x_n)) \\ &= P(X_1 = x_1, ..., X_n = x_n) \end{aligned}$$

Ainsi, étant données deux variables aléatoires discrètes X et Y, on appelle loi du couple (X, Y) la probabilité $P_{X,Y}$ définie par

$$P_{X,Y}(x, y) = P(X = x, Y = y) \ \ \forall x \in X(\Omega), y \in Y(\Omega)$$

5.2 Fonction de répartition d'un vecteur aléatoire

Définition. On appelle fonction de répartition du vecteur aléatoire $(X_1, ..., X_n)$ et on note $F_{X_1,...,X_n}$ la fonction

$$F_{X_1,...,X_n} : \mathbb{R}^n \to [0, 1]$$

$$F_{X_1,...,X_n} \begin{cases} \mathbb{R}^n \to [0, 1] \\ (x_1, ..., x_n) \mapsto F_{X_1,...,X_n}(x_1, ..., x_n) = P(X_1 \leq x_1, ..., X_n \leq x_n) \end{cases}$$

Dans le cas d'un couple de variables aléatoires (X, Y), on définit la fonction de répartition du couple (X, Y) par

$$F_{X,Y} : \begin{cases} \mathbb{R}^2 \to [0, 1] \\ (x, y) \mapsto F_{X,Y}(x, y) = P(X \leq x, Y \leq y) \end{cases}$$

5.3 Lien avec l'indépendance

Théorème II.4. Soit $X_1, ..., X_n$ n variables réelles alors : les variables aléatoires $X_1, ..., X_n$ sont (mutuellement) indépendants si et seulement si $\forall (x_1, ..., x_n) \in \mathbb{R}^n, F_{X_1,...,X_n}(x_1, ..., x_n) = F_{X_1}(x_1) \times ... \times F_{X_n}(x_n)$. En particulier, deux variables aléatoires X et Y sont indépendants si et seulement si $\forall (x, y) \in \mathbb{R}^2, F_{X,Y}(x, y) = F_X(x) F_Y(y)$.

5.4 Lois marginales

Définition. Soit $(X_1, ..., X_n)$ un vecteur aléatoire de la $P_{X_1,...,X_n}$. Alors la loi de probabilité P_{X_i} des variables aléatoires $X_i, 1 \leq i \leq n$ sont appelées "lois marginales".

Remarque. $P_{X_i}(A_i) = P_{X_1,...,X_i,...,X_n}(\mathbb{R} \times \mathbb{R} \times ... \times \mathbb{R} \times A_i \times \mathbb{R} \times ... \times \mathbb{R})$. Si les variables aléatoires X_i sont discrets

$$P(X_i = x_i) = P_{X_i}(x_i) = \sum_{x_1,...,x_{i-1},x_{i+1},...,x_n} P_{X_1,...,X_{i-1},X_l,X_{l+1},...,X_n}(x_1, ..., x_i, ..., x_n)$$

Par exemple si (X, Y) est un couple de variable aléatoires discrètes $\forall x \in X(\Omega)$

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

$\forall y \in Y(\Omega), P_Y(y) = P(Y = y) = \sum_{x \in X(\Omega)} P(X = x, Y = y)$

6 Espérance d'une variable aléatoire. Moment d'ordre supérieur

6.1 Espérance mathématiques

Définition. Soit X une variable aléatoire discrète sur un espace probabilisé $(\Omega, \mathcal{F}, P)$ avec $X(\Omega) = \{x_n, n \in \mathbb{N}\}$. On dit que X est d'espérance finie si

$$\sum_{n=0}^{+\infty} |x|_n P(X = x_n) < +\infty$$

(c'est à dire la série $x_n P(X = x_n)$ est absolument convergente et on appelle alors espérance de X le nombre

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

Remarque 1. On dit aussi parfois que X est intégrable (ie d'espérance finie)

Remarque 2. Si Ω est un espace fini ou dénombrable, alors X est d'espérance finie si et seulement si

$$\sum_{\omega \in \Omega} |X(\omega)| P(\omega) < +\infty$$

et on peut écrire

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

Remarque 3. Si par exemple, $\forall x_n \in X(\Omega), x_n \geq 0$ (X à valeurs positives), alors la valeur de $\sum_{n=0}^{+\infty} x_n P(X = x_n)$ est toujours bien définie (elle est réelle si X est d'espérance finie ou bien elle vaut $+\infty$). On peut généraliser la définition de la façon suivante.

Définition. Soit X une variable aléatoire discrète sur $(\Omega, \mathcal{F}, P)$ à valeurs dans $X(\Omega) = \{x_n, n \in \mathbb{N}\}$. Posons $X^+ = \max(X, 0)$ (partie positive) et $X^- = \max(-X, 0)$ (partie négative). On voit que $X^+ - X^- = X$ et $X^+ + X^- = |X|$.

On dire que la variable aléatoire X "admet une espérance" si et seulement si au moins une des deux contradiction suivantes est vérifiée.

1. $E(X^+) = \sum\limits_{x_n \in X(\Omega)}^{x_n > 0} x_n P(X = x_n) < +\infty$
2. $E(X^-) = \sum\limits_{x_n \in X(\Omega)}^{x_n < 0} -x_n P(X = x_n) < +\infty$

et on définit l'espérance de X par

$$E(X) = E(X^+) - E(X^-)$$

Remarque. Une variable aléatoire X est d'espérance finie si et seulement si

$$E(X^+) < +\infty \text{ et } E(X^-) < +\infty$$

Premier exemple. On rappelle que $\sum\limits_{n \geq 1} = \dfrac{\pi^2}{6}$.

Soit X une variable aléatoire à valeurs dans $\mathbb{N}$ avec $P(X = 0) = 0$ et $\forall n \geq 1, P(X = n) = \dfrac{6}{\pi^2 n^2}$. On voit que

$$\sum_{n \geq 1} nP(X = n) = \sum_{n \geq 1} \frac{6}{\pi^2 n} = +\infty$$

Donc X n'est pas d'espérance finie. Par contre, on peut dire que X admet une espérance ($E(X) = +\infty$) (on voit qu'ici $E(X^-) = 0$).

Considérons maintenant X à valeurs dans $\mathbb{Z}$ avec $P(X = 0) = 0$ et $\forall n \neq 0, P(X = n) = \dfrac{3}{\pi^2 n^2}$. Alors on se trouve dans le cas

$$E(X^+) = +\infty$$

$$E(X^-) = +\infty$$

Donc X n'admet pas d'espérance.

Théorème II.5. (théorème de transfert) Soit $X_1, ..., X_k$ k variable aléatoire discrètes. Soit f une fonction de $\mathbb{R}^k \to \mathbb{R}$. Alors la variable aléatoire $Z = f(X_1, ..., X_k)$ est d'espérance finie si et seulement

$$\sum |f(x_1, ..., x_k)| P(X_1 = x_1, ..., X_k = x_k) < +\infty$$

avec $x_1 \in X_1(\Omega), ..., x_k \in X_k(\Omega)$ et on a alors

$$E(Z) = \sum_{x_1 \in X_1(\Omega), \ldots, x_k \in X_k(\Omega)} f(x_1, ..., x_k) P(X_1 = x_1, ..., X_k = x_k)$$

Remarque. Dans le cas où $k = 1$ et $X(\Omega) = \{x_n, n \in \mathbb{N}\}$ on a donc

$$E(Z) = \sum_{n=0}^{+\infty} f(x_n) P(X = x_n)$$

Avantage. La définition de $E(Z)$ est $E(Z) = \sum_{z_n \in Z(\Omega)} z_n P(Z = z_n)$

Propriétés.

1. Si $Y \geq 0$ est d'espérance finie et si $|X| \leq Y$, alors X est d'espérance finie.
2. (a) Si X est d'espérance finie alors $\forall \lambda \in \mathbb{R}, \lambda X$ est d'espérance finie et $E(\lambda X) = \lambda E(X)$
   (b) si X et Y sont d'espérance finie, alors $X + Y$ est d'espérance finie et $E(X + Y) = E(X) + E(Y)$
   (c) Si $\exists b \in \mathbb{R}$ tel que $P(X = b) = 1$ alors $E(X) = b$
   (d) $X \geq 0 \Rightarrow E(X) \geq 0$
   (e) Si $X \geq 0$ et $E(X) = 0$ alors $P(X = 0) = 1$
   (f) $X \leq Y \Rightarrow E(X) \leq E(Y)$
   (g) Si X et Y sont 2 variables aléatoires indépendantes d'espérance finie, alors XY est d'espérance finie et
   $$E(XY) = E(X)E(Y)$$

Démonstration.

1. Admis
2. (a) Dans le théorème du transfert, on pose $k = 1$ et $f(x) = dx$
   (b) Dans le théorème du transfert on pose $k = 2$ et $f(x, y) = x + y$ e a et b, on déduit que $\forall \lambda, \mu \in \mathbb{R}$,
   $$E(\lambda x + \mu y) = \lambda E(X) + \mu E(y)$$
   (c) évident $E(X) = bP(X = b) = b$
   (d) évident
   (e) Soit $X \geq 0$ tel que $E(x) = 0$. On a donc $\sum_{x \in X(\Omega)} xP(X = x) = 0 \Rightarrow \sum_{x > 0} xP(X = x) = 0 \Rightarrow \forall x > 0, xP(X = x) = 0 \Leftrightarrow P(X = x) = 0, \forall x > 0 \Rightarrow P(X = 0) = 1$
   (f) $X \leq Y \Rightarrow Y - X \geq 0 \Rightarrow E(Y - X) \geq 0 \Leftrightarrow E(Y) + E(-X) \geq 0 \Leftrightarrow E(Y) - E(X) \geq 0$
   (g) On prend $k = 2$ et $f(x, y)xy$ dans le théorème du transfert.

□

Remarque. La condition d'indépendance est une condition suffisante mais plus necessaire. On peut avoir $E(XY) = E(X)E(Y)$ et X et Y non indépendants.

Proposition. Soit X avec une variable à valeurs dans $\mathbb{N}$, alors

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

Démonstration. $\sum_{k=1}^{n} P(X \geq k) = P(X \geq 1) + P(X \geq 2) + P(X \geq 3) + ... + P(X \geq n) = P(X = 1) + P(X = 2) + P(X = 3) + ... + P(X = n) + P(X \geq n+1) + P(X = 2) + P(X = 3) + ... + P(X = n) + P(X \geq n+1)... = P(X = 1) + 2P(X = 2) + 3P(X = 3) + ... + nP(X = n) + nP(X \geq n+1) = \sum_{k=1}^{n} kP(X)k) + nP(X \geq n+1)$

1. Si $E(X) < +\infty$ Alors on a $P(X \geq n+1) = n \sum_{k=n+1}^{+\infty} P(X = k) \leq \sum_{k=n+1}^{+\infty} kP(X = k)$ et on a $\sum_{k=n+1}^{\infty} kP(X = k) \to 0$ car $E(X) = \sum_{k=1}^{\infty} kP(X = k) < +\infty$. D'où l'égalité en passant à la limite.
2. Si $E(X) = +\infty$. Alors on a $\sum_{k=1}^{\infty} P(X \geq k) \geq \sum_{k=1}^{\infty} kP(X = k)$ d'où $\sum_{k=1}^{\infty} P(X \geq k) = +\infty$ et l'égalité reste vraie.

□

Remarque. Si X est à valeurs dans $\mathbb{N} \cup \{+\infty\}$ avec $P(X = +\infty) > 0$. Alors $E(X) = +\infty$ et $\forall n, \sum_{k=1}$

7 Moments d'ordre supérieur variance et covariance

Définition. Soit X une variable aléatoire discrète sur $(\Omega, \mathcal{F}, P)$ avec $X(\Omega) = \{x_n, n \in \mathbb{N}\}$. On dit que X admet un moment (non centré) à l'ordre k si

$$\sum_{n \geq 0} |x_n|^k P(X = x_n) < +\infty$$

et on appelle alors moment (non centré) d'ordre k le nombre

$$E(X^k) = \sum_{n \geq 0} x_n^k P(X = x_n)$$

Remarque. On sait aussi que X est de puissante k intégrable. Si $k = 2$, on parle de variable aléatoire de carré itégrable l'espérance est le moment (non centré) d'ordre 1

Propriété. Si X admet un moment (non centré d'ordre k, alors $\forall i \leq k$, X admet un moment (non centré) d'ordre i.

Démonstration. On peut écrire $(i \leq k)$

$$|X|^i = |X|^i \mathbb{1}_{\{|X| \leq 1\}} + |X|^i \mathbb{1}_{\{|X| > 1\}} \leq 1.\mathbb{1}_{\{|X| \leq 1\}} + |X|^k \mathbb{1}_{\{|X| \leq 1\}} \leq 1 + |X|^k$$

Et donc $E(|X|^i) \leq 1 + E(|X|^k)$ d'où le résultat. □

Corollaire. Si X admet un moment d'ordre $k \geq 1$ alors X est d'espérance finie.

Propriété Si X admet un moment d'ordre k, alors $\forall a \in \mathbb{R}, (X + a)$ admet un moment d'ordre k.

Démonstration. D'après la formule du binôme de Newton.

$$(X+a)^k = \sum_{i=0}^{k} C_n^i X^i a^{k-i}$$

d'où le résultat avec la propriété précédente. □

Définition. Soit X une variable aléatoire admettant un moment (non centré) d'ordre k. On appelle moment centré d'ordre k la quantité

$$E((X-E(X))^k) = \sum_{n\geq 0} (x_n - E(X))^k P(X = x_n)$$

Définition. On appelle variance de X le moment centré d'ordre $k = 2$ et on note cette quantité σ_x^2 ou $Var(X)$ (ou encore $V(X)$)

$$\sigma_X^2 = E[(X-E(X))^2]$$

Définition. On appelle écart type de X la racine carrée de la variance c'est à dire $\sigma = \sqrt{Var(X)}$

Propriétés de la variance.

1. $Var(X) = E(X^2) - E(X)^2$

   Démonstration. $Var(X) = E[(X-E(X))^2] = E[X^2 - 2 \times E(X) + E(X)^2] = E[X^2] - 2E(X)E(X) + E(X)^2 = E(X^2) - E(X)^2$ □

2. $\forall a, b \in \mathbb{R}, Var(aX+b) = a^2 Var(X)$

   Démonstration. $Var(aX+b) = E[(aX+b-E(aX+b)^2] = E[(a(X-E(X))^2] = a^2 E[(X-E(X))^2] = a^2 Var(X)$ □

3. $Var(X) = 0 \Leftrightarrow \exists b \in \mathbb{R}$ tel que $P(X = b) = 1$

   Démonstration. $Var(X) = 0 \Leftrightarrow E[-X-E(X))^2] = 0 \Leftrightarrow P(X-E(X)=0) = 1 \Leftrightarrow P(X = E(X))) = 1$ (d'après la propriété e de l'espérance) □

Remarque. On a donc $\forall a, b \in \mathbb{R}, \sigma_{aX+b} = |a|\sigma_X$

8 Covariance, variance d'une somme

Propriété. Soit X et Y deux variables aléatoires admettant un moment d'ordre 2. Alors XY est despérance finie.

Démonstration. On a $(|X|-|Y|)^2 \geq 0 \Leftrightarrow X^2 - 2|X||Y| + Y^2 \geq 0 \Leftrightarrow |X||Y| \leq \frac{1}{2}(X^2+Y^2)$ D'après les propriété 1 et 2bde l'espérance, XY est d'espérance finie. □

Définition. Soit X et Y deux variables aléatoires admettant un moment d'ordre 2. On appelle covariance de X et Y la quantité

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))$$

Propriété. $Cov(X,Y) = E(XY) - E(X)E(Y)$

Démonstration. $Cov(X,Y) = E[XY - XE(Y) - YE(X) + E(X)E(Y)] = E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) = E(XY) - E(X)E(Y)$ □

Définition. On dit que 2 variables aléatoires X et Y admettant un moment d'ordre 2 sont non corrélés si $Cov(X,Y) = 0$.

Remarque. On a que X et Y sont indépendants ce qui implique que X et Y sont non corrélées. La réciproque est fausse.

Variance de la somme. D'une manière générale, on peut écrire

$$\begin{aligned}
Var(X+Y) &= E[(X+Y-E(X+Y))^2] \\
&= E[(X-E(X)) + (Y-E(Y))]^2 \\
&= E[(X-E(X))^2 + (Y-E(Y))^2 + 2(X-E(X))(Y-E(Y))] \\
&= Var(X) + Var(Y) + 2Cov(X,Y)
\end{aligned}$$

Proportion. Donc variable aléatoire de carré intégrable sont non corrélées si et seulement si

$$Var(X+Y) = Var(X) + Var(Y)$$

Généralisation. Soit $X_1, ..., X_n$ n variables aléatoires intégrables. Alors

$$Var(X_1 + ... + X_n) = Var(X_1) + ... + Var(X_n) + 2\sum_{i<j} Cov(X_i, X_j)$$

Si de plus les variables aléatoires $X_1, ..., X_n$ sont 2 à 2 non corrélées, alors on a

$$Var(X_1 + ... + X_n) = Var(X_1) + ... + Var(X_n)$$

Propriété de la covariance

— $Cov(X,X) = Var(X)$

— $(X,Y) \to Cov(X,Y)$ est bilinéaire symétrique positive

Démonstration. $Cov(X,Y) = Cov(Y,X)$, $Cov(X,X) = Var(X) \geq 0$, $Cov(aX+bY,Z) = E((aX+bY)Z) - E(aX+bY)E(Z) = E(aXZ+bYZ) - aE(X)E(Z) - bE(Y)E(Z)$ (linéarité de l'espérance) $a(E(XZ) - E(X)E(Z)) + b(E(YZ) - E(Y)E(Z)) = aCov(X,Z) + bCov(Y,Z)$ □

— $|Cov(X,Y)| \leq \sqrt{Var(X)}\sqrt{Var(Y)} \Leftrightarrow |Cov(X,Y)| \leq \sigma_X\sigma_Y$

Démonstration. Résulte de l'inégalité de Cauchy-Schwartz. En effet ,$\forall a, P(a) = Cov(aX+Y, aX+Y) = Var(aX+Y) \geq 0 \Leftrightarrow P(a) = a^2Var(X)+2aCov(X,Y)+Var(Y) \geq 0$. D'où le delta de $P(a)$ est inférieur à 0 donc $YCov^2(X,Y) - 4Var(X)Var(Y) \geq 0 \Leftrightarrow Cov^2(X,Y) \leq Var(X)Var(Y) \Leftrightarrow |Cov(X,Y)| \leq \sigma_X\sigma_Y$ □

Remarque. On définit le coefficient de corrélation (linéaire) de 2 variables aléatoires par

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}$$

D'aprèes la propriété précédente, on a $|\rho(X,Y)| \leq 1$ On voit que X et Y non corrélées $\Leftrightarrow \rho(X,Y) = 0$

Propriété. $|\rho| = 1 \Leftrightarrow \exists a, b \in \mathbb{R}$ tel que $aX + Y = b$ (tel que $P(aX + Y = b) = 1$)

Démonstration. si $Y = b - aX$, $\sigma_Y = |a|\sigma_X$

$$Cov(X,Y) = Cov(X, b - aX) = Cov(X, b) - aVar(X) = -aVar(X)$$

$$|\rho| = \frac{|aVar(X)}{\sigma_X|a|\sigma_X} = 1$$

Réciproquement si $|\rho| = 1$ alors le delta de $P(a)$ est nul, et donc il existe une racine de $P(a) \Leftrightarrow \exists a$tel que$(aX + Y) = 0 \Leftrightarrow aX + Y = b$ □

9 Inégalités probabilistes

9.1 Loi des grands nombres

Inégalité de Markov. Soit X une variable aléatoire d'espérance finie, alors $\forall \lambda > 0$

$$P(|X| \geq) \leq \frac{E(|X|)}{\lambda}$$

Inégalité de Bunayme-Tchebytcheff. Soit X une variable aléatoire de carré intégrable, alors $\forall \epsilon > 0$,

$$P(|X - E(X)| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}$$
$$\Leftrightarrow P(|X - E(X)| < \epsilon) \geq 1 - \frac{Var(X)}{\epsilon^2}$$
$$\Leftrightarrow P(E(X) - \epsilon < X < E(X) + \epsilon) \geq 1 - \frac{Var(X)}{\epsilon^2}$$

Démonstration. Découle de l'inégalité de Markov. Posons $\lambda = \epsilon^2$ et $Y = (X - E(X))^2$ Markov implique $P(|Y| \geq \lambda) \leq \frac{E(|Y|)}{\lambda} \Rightarrow P((X - E(X))^2 \geq \epsilon^2) \leq \frac{Var(X)}{\epsilon^2} \Rightarrow P(|X - E(X)| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}$. □

Exemple. Soit X une variable aléatoire tel que $E(X) = 100$ et $Var(X) = 16$. Posons $\epsilon = 8$, alors

$$P(92 < X < 108) \geq 1 - \frac{1,6}{64}$$

On déduit de l'inégalité de Bunayme-Tchebytcheff une version simplifiée de la loi faible des grands nombres.

Théorème II.6. Soit $(X_n)_{n\geq 1}$ une suite de variable aléatoire i.i.d. avec $E(X_n) = \mu$ et $Var(X_n) = \sigma^2$. Posons $\forall n \geq 1, S_n = \sum_{k=1}^{n} X_k$ et $\overline{X_n} = \frac{S_n}{n} = \frac{1}{n}\sum_{k=1}^{n} X_k$ $\overline{X}_n$ s'appelle la moyenne d'échantillon.
Alors $\forall \epsilon > 0, \lim_{n\to+\infty} P(|\overline{X}_n - \mu| \geq \epsilon) = 0$ (on dit que $\overline{X}_n$ converge e probabilité vers μ).

Démonstration. D'après l'inégalité de B.T

$$P(|\overline{X}_n - E(\overline{X}_n)| \geq \epsilon) \leq \frac{Var(\overline{X}_n)}{\epsilon^2}$$

$E(\overline{X}_n) = \frac{1}{n}\sum_{k=1}^{n} E(X_k) = \frac{1}{n} n\mu = \mu$
$Var(\overline{X}_n) = \frac{1}{n^2} Var(\sum_{k=1}^{n} X_k) = \frac{1}{n^2}\sum_{k=1}^{n} Var(X_k) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$ □

10 Fonction génératrices

Définition. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$ et posons $\forall n \geq 0, p_n = P(X = n)$. On appelle **fonction génératrice** de X (ou fonction génératrice des probabilités ou ou encore fonction génératrice des moments) la fonction

$$G_x : \begin{cases} \mathbb{C} \to \mathbb{C} \\ z \mapsto G_X(z) = \sum_{n=0}^{+\infty} p_n z^n \end{cases}$$

Remarque.

— $G_X(z) = E(z^X)$ (théorème du transfert)

— Comme bien sur $\sum_{n=0}^{+\infty} p_n = 1$, on en déduit que le rayon de convergence de la série entière $\sum p_n z^n$ est supérieur à 1.

Théorème II.7. Soit X et Y deux variables aléatoires telles que $\forall z \in \mathbb{R}, G_X(z) = G_Y(z)$. Alors $\forall n \in \mathbb{N}, P(X = n) = P(Y = n)$ (X et Y ont même loi)

Démonstration. Puisque le rayon de convergence de G_X et G_Y, est supérieur à 1, $\forall r \in]0,1[$, on peut dériver G_X (et G_Y) indéfiniment sur $]-r,r[$, en particulier G_X et G_Y sont indéfiniment dérivables en O. Or $\forall z \in]-r,r[$

$$G_X^{(n)}(z) = n!p_n + \sum_{k=n+1}^{+\infty} k(k-1)\times ... \times(k-n+1)p_k z^{k-n} \Rightarrow G_X^{(n)}(0) = n!p_n \Leftrightarrow p_n = \frac{G_X^{(n)}(0)}{n!}$$

On a montré de même que $q_n = P(Y = n) = \frac{G_Y^{(n)}(0)}{n!}$ d'où si $G_X = G_Y$ on a forcément $p_n = q_n \forall n$ □

Théorème II.8. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$ de fonction génératrice G_X. Alors X admet un moment d'ordre m si et seulement si la dérivée à gauche d'ordre m au point $z = 1$ de la fonction G_X existe et est finie. Dans ce cas, on a de plus

$$E[(X(X-1)\times \ldots \times (X-m+1)] = G_X^{(m)}(1) = \sum_{k=m}^{+\infty} k(k-1)\times \ldots \times (k-m+1)p_k$$

En particulier

$$\begin{aligned}
E(X) &= G'_X(1) \\
E(X^2) &= E(X(X-1)) + E(X) \\
&= G''_X(1) + G'_X(1) \\
Var(X) &= G''_X(1) + G'_X(1) - (G'_X(1))^2
\end{aligned}$$

Démonstration. Dans le cas où le rayon de convergence de la série entière $\sum p_n z^n$ est $R > 1$, le résultat est évident car la fonction G_X est indéfiniment dérivable au point 1. Si $R = 1$ la démonstration est plus technique. □

Fonction génératrice d'une somme de variable aléatoire indépendantes. Soit X et Y deux variables aléatoires indépendantes (à valeurs dans $\mathbb{N}$) de fonction génératrice G_X et G_Y. Calculons la fonction génératrice de $X+Y$

$$\begin{aligned}
G_{X+Y}(z) &= \sum_{n=0}^{+\infty} P(X+Y=n)z^n \\
&= \sum_{n=0}^{+\infty}\sum_{k=0}^{n} P(X=k, Y=n-k)z^n \\
&= \sum_{n=0}^{+\infty}\sum_{k=0}^{n} P(X=k)P(Y=n-k)z^n \\
&= \sum_{n=0}^{+\infty}\sum_{k=0}^{n} P(X=k)z^k P(Y=n-k)z^{n-k} \\
&= (\sum_{n=0}^{+\infty} P(X=n)z^n) \times (\sum_{n=0}^{+\infty} P(Y=n)z^n) \\
&= G_X(z).G_Y(z)
\end{aligned}$$

Autre démonstration. On a vu que $G_X(z) = E(z^X)$ et $G_Y(z) = E(z^Y)$. Donc $G_{X+Y}(z) = E(z^{X+Y}) = E(z^X z^Y) = E(z^X)E(z^Y)$ (car z^X et z^Y sont indépendants). On peut généraliser le résultat à n variable aléatoire indépendantes.

Théorème II.9. Soit $X_1, ..., X_n$ n variables aléatoires à valeurs dans $\mathbb{N}$ (mutuellement) indépendants. Alors

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

Exemples de fonctions génératrices.

Loi de Bernoulli. $X \sim \mathcal{B}(p), G_X(z) = P(X=0)z^0 + P(X=1)z^1 = 1-p+pz$ de plus $G'_X(z) = p$ et $G''_X(z) = 0$ donc $E(X) = p$ et $Var(X) = 0 + p - p^2 = p(1-p)$

Loi binomiale. $X \sim \mathcal{B}(n,p)$, $G_X(z) = \sum_{k=0}^{n} C_k^n p^k (1-p)^{n-k} z^k = \sum_{k=0}^{n} C_k^n (pz)^k (1-p)^{n-k} = (pz+1-p)^n$ (Binôme de Newton)

Loi géométrique (sur $\mathcal{N}^*$). $X \sim \mathcal{G}(p)$, $G_X(z) = \sum_{n=1}^{\infty} P(X=n) z^n = \sum_{n=1}^{+\infty} p(1-p)^{n-1} - z^n = pz \sum_{n=1}^{+\infty} ((1-p)z)^{n-1} = pz \frac{1}{1-(1-p)z} = \frac{pz}{1-(1-p)z}$

$G'_X(z) = \frac{p(1-(1-p)z) + pz(1-p)}{(1-(1-p)z)^2} = \frac{p}{1-(1-p)z)^2}$

$G''_X(z) = p \frac{2(1-p)}{1-(1-p)z)^3}$

$E(X) = G'_X(1) = 1/p$ et $Var(X) = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$.

Loi de Pascal. $X \sim Pa(r,p), G_X(z) = (\frac{pz}{1-(1-p)z})^r$

Loi de Poisson. $X \sim \mathcal{P}(\lambda)$, $G_X(z) = \sum_{n=0}^{+\infty} e^{-\lambda} \frac{\lambda^n}{n!} z^n = \sum_{n=0}^{+\infty} e^{-\lambda} \frac{(\lambda z)^n}{n!} = e^{-\lambda} \sum_{n=0}^{+\infty} \frac{(\lambda z)^n}{n!} = e^{-\lambda} e^{\lambda z} = e^{-\lambda(z-1)}$.

$G'_X(z) = \lambda e^{\lambda(z-1)} \Rightarrow G'_X(1) = \lambda$

$G''_X = \lambda^2 e^{\lambda(z-1)} \Rightarrow Var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$

FIN.