

Définition : Ω est un ensemble quelconque.

Une tribu (ou σ -algèbre) de Ω est une partie \mathcal{T} de $\mathcal{P}(\Omega)$ telle que :

- $\Omega \in \mathcal{T}$
 - \mathcal{T} est stable par réunion dénombrable et par passage au complémentaire :
- $\forall A \in \mathcal{T}, \Omega \setminus A \in \mathcal{T}$ et $\forall (A_n) \in \mathcal{T}^{\mathbb{N}}, \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{T}$.

Remarque : • On a alors $\emptyset \in \mathcal{T}$ car c'est le complémentaire de Ω dans Ω .

- Cette définition implique que \mathcal{T} est stable par réunion finie.
- En passant au complémentaire, \mathcal{T} est stable par intersection dénombrable (ou finie).

Définition : un espace probabilisable est un couple (Ω, \mathcal{T}) où :

- \mathcal{T} est une tribu de Ω , dite tribu des évènements (les évènements étant les éléments de \mathcal{T}).
- Ω est l'univers, ses éléments sont les issues (ou aléas).

Remarque : Comme, le plus souvent, pour $\omega \in \Omega$, $\{\omega\} \in \mathcal{T}$, les singletons qui sont des évènements sont appelés évènements élémentaires.

Définition : pour (Ω, \mathcal{T}) espace probabilisable, une probabilité $P: \begin{cases} \mathcal{T} & \rightarrow \mathbb{R}_+ \\ A & \mapsto P(A) \end{cases}$ sur (Ω, \mathcal{T}) est une application telle que :

- $P(\Omega) = 1$
- $\forall (A_n) \in \mathcal{T}^{\mathbb{N}}, (\forall (i, j) \in \mathbb{N}, i \neq j \Rightarrow A_i \cap A_j = \emptyset) \Rightarrow P(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} P(A_n)$

Remarques : • $A \cap B = \emptyset$ se dit « A et B » sont incompatibles.

- On note traditionnellement $\bigsqcup_{i \in I} A_i$ pour $\bigcup_{i \in I} A_i$ et $\forall i \neq j, A_i \cap A_j = \emptyset$.
- Posant, $A_i = \emptyset$ pour $i \in \mathbb{N}$, $P(\bigsqcup_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} P(A_i) = \sum_{i \in \mathbb{N}} P(\emptyset) < \infty$ donc $P(\emptyset) = 0$.
- $A \in \mathcal{T} \Rightarrow A^c \in \mathcal{T}$ et $1 = P(A \sqcup A^c) = P(A) + P(A^c)$
- Pour $A \subset B \in \mathcal{T}^2$, $B = A \sqcup (B \setminus A)$ donc $P(A) + P(B \setminus A) = P(B)$ et donc $P(A) \leq P(B)$.

Théorème de continuité monotone :

- Soit $(A_n) \in \mathcal{T}^{\mathbb{N}}$ une suite croissante pour l'inclusion. Alors :

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} P(A_n)$$

- Soit $(A_n) \in \mathcal{T}^{\mathbb{N}}$ une suite décroissante pour l'inclusion. Alors :

$$P\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} P(A_n)$$

$$\left| \begin{array}{l} \text{Preuve : on pose } A_{-1} = \emptyset \text{ et } A = \bigcup_{k=0}^{\infty} A_k = \bigsqcup_{k \in \mathbb{N}} (A_k \setminus A_{k-1}) \\ P(A) = \sum_{k=0}^{\infty} P(A_k \setminus A_{k-1}) = \lim_{n \rightarrow \infty} \sum_{k=0}^n P(A_k \setminus A_{k-1}) = \lim_{n \rightarrow \infty} P\left(\bigsqcup_{k \in [0, n]} (A_k \setminus A_{k-1})\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k=0}^n A_k\right) \\ \qquad \qquad \qquad = \lim_{n \rightarrow \infty} P(A_n) \end{array} \right|$$

Par passage au complémentaire, on a le théorème pour les suites décroissantes.

Corollaire : soit $(A_n)_{n \in \mathbb{N}} \in \mathcal{T}^{\mathbb{N}}$ pas forcément croissante. On lit dans la preuve précédente :

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k=0}^n A_k\right)$$

$$P\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} P\left(\bigcap_{k=0}^n A_k\right)$$

Proposition : inégalité de Boole. Soit $(A_n) \in \mathcal{T}^{\mathbb{N}}$:

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

On démontre par récurrence sur n que $P(\bigcup_{k=0}^n A_k) \leq \sum_{k=0}^n P(A_k)$.

$n = 0$: évident.

$n \geq 0$: et supposons la propriété vraie au rang n .

$$\begin{aligned} P\left(\bigcup_{k=0}^{n+1} A_k\right) &= P\left(\bigcup_{k=0}^n A_k \cup A_{n+1}\right) = P\left(\bigcup_{k=0}^n A_k\right) + P(A_{n+1}) - P\left(\bigcup_{k=0}^n A_k \cap A_{n+1}\right) \\ &= P\left(\bigcup_{k=0}^n A_k\right) + P(A_{n+1}) \\ &\stackrel{HR}{\leq} \sum_{k=0}^n P(A_k) + P(A_{n+1}) = \sum_{k=0}^{n+1} P(A_k) \end{aligned}$$

En utilisant le corollaire précédent, on obtient le résultat par passage à la limite.

Remarque : $\bigcup_{k=0}^{\infty} A_k = \lim_{n \rightarrow \infty} \bigcup_{k=0}^n A_k$ n'a aucun sens car la limite d'un truc qui n'est pas dans un evn c'est bidon.

Proposition – définition :

- Evènement presque impossible : $P(A) = 0$
- Evènement presque certain : $P(A) = 1$

A presque certain $\iff A^c$ presque impossible.

- Si $P(A \cap B) = 0$, A et B sont presque incompatibles.

Pour $(A_i)_{i \in \mathbb{N}}$ presque incompatibles : $P(\biguplus_{i \in \mathbb{N}} A_i) = \sum_{i=0}^{\infty} P(A_i)$ (preuve plus bas)

- Système total : $\Omega = \bigsqcup_{i \in I} A_i$
- Système presque total : $P(\biguplus_{i \in I} A_i) = 1$
- Pour $(A_i)_I$ système total appréciable (dont les probabilités sont non nulles) :

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

(En effet, $B = B \cap \Omega = B \cap (\bigsqcup_{i \in I} A_i)$ et $P(B \mid A_i) := \frac{P(A_i \cap B)}{P(A_i)}$)

La preuve qui est plus bas : par récurrence :

$$n = 1 : P(A \uplus B) = P(A) + P(B)$$

$n \geq 1$:

$$P\left(\biguplus_{i=0}^{n+1} A_i\right) = P\left(\biguplus_{i=0}^n A_i \cup A_{n+1}\right)$$

$$\text{Et } P(A_{n+1} \cap \biguplus_{i=0}^n A_i) = P(\biguplus_{i=0}^n A_i \cap A_{n+1}) \stackrel{Boole}{\leq} \sum_{i=0}^n P(A_i \cap A_{n+1}) = 0$$

Donc la réunion est bien disjointe et on peut récuser.

$$P\left(\biguplus_{i=0}^n A_i\right) = \sum_{i=0}^n P(A_i) \text{ donc par le théorème de continuité monotone : } P\left(\biguplus_{i \in \mathbb{N}} A_i\right) = \sum_{i=0}^{\infty} P(A_i)$$

Proposition – définition : on considère un espace probabilisé (Ω, \mathcal{T}, P) .

- Deux évènements A et B sont dits indépendants si $P(A \cap B) = P(A)P(B)$.
 - Pour I fini, $(A_i)_{i \in I} \in \mathcal{T}^I$, les (A_i) sont deux à deux indépendants si $\forall i \neq j, P(A_i \cap A_j) = P(A_i)P(A_j)$.
- Les (A_i) sont mutuellement indépendants si $\forall J \in \mathcal{P}(I), P(\bigcap_{i \in J} A_i) = \prod_{i \in J} P(A_i)$.

Pour I infini, les (A_i) sont mutuellement indépendants si $\forall J \in \mathcal{P}_f(I), P(\bigcap_{i \in J} A_i) = \prod_{i \in J} P(A_i)$.

Cela équivaut à : tout sous famille finie des (A_i) est mutuellement indépendante.

Remarque : il est évident que l'indépendance mutuelle entraîne l'indépendance deux à deux.

Proposition : soit I fini.

- $(A_i)_{i \in I}$ mutuellement indépendants $\iff \forall (A_i^*)_{i \in I} \in \prod_{i \in I} \{A_i, A_i^c\}, (A_i^*)_{i \in I}$ mutuellement indépendants.
- $(A_i)_{i \in I}$ mutuellement indépendants $\iff \forall (A_i^*)_{i \in I} \in \prod_{i \in I} \{A_i, A_i^c\}, P(\bigcap_{i \in I} A_i^*) = \prod_{i \in I} P(A_i^*)$.

1^{ère} preuve par récurrence sur $n = \#\{i \in I \mid A_i^* = A_i^c\}$.

2^{nde} preuve : posons, pour $J \subset I, \mathcal{A}_J = (\bigcap_{i \in J} A_i) \cap (\bigcap_{i \in I \setminus J} A_i^c)$. Alors $(\mathcal{A}_J)_{J \in \mathcal{P}(I)}$ est un système total car :

Pour $\omega \in \Omega$ et $J \subset I, \omega \in \mathcal{A}_J \iff \forall i \in I, (\omega \in A_i \iff i \in J) \iff \{i \in I \mid \omega \in A_i\} = J$. Notant $J(\omega) = \{i \in I \mid \omega \in A_i\}$, chaque ω appartient donc à un \mathcal{A}_J et un seul, à savoir $\mathcal{A}_{J(\omega)}$, et $(\mathcal{A}_J)_{J \in \mathcal{P}(I)}$ est une partition de Ω .

Si $(A_i)_{i \in I}$ mutuellement indépendants, pour $(A_i^*)_{i \in I} \in \prod_{i \in I} \{A_i, A_i^c\}, (A_i^*)_{i \in I}$ mutuellement indépendants donc $P(\bigcap_{i \in J} A_i^*) = \prod_{i \in J} P(A_i^*)$.

Réciproquement : si $\forall (A_i^*)_{i \in I} \in \prod_{i \in I} \{A_i, A_i^c\}, P(\bigcap_{i \in J} A_i^*) = \prod_{i \in J} P(A_i^*)$.

Soit $K \subset I$ quelconque. Montrons que $\bigcap_{i \in K} A_i = \bigsqcup_{J \supset K} \mathcal{A}_J$. Plus rigoureusement, notant $H = \{J \in \mathcal{P}(I) \mid K \subset J\}$, que $\bigcap_{i \in K} A_i = \bigsqcup_{J \in H} \mathcal{A}_J$.

$$\begin{aligned} \omega \in \bigcap_{i \in K} A_i &\iff \forall i \in K, \omega \in A_i \\ &\iff J(\omega) = \{i \in I \mid \omega \in A_i\} \supset K \\ &\iff \exists J \subset I \mid K \subset J \text{ et } J(\omega) = J \\ &\iff \exists J \subset I \mid K \subset J \text{ et } \omega \in J \\ &\iff \omega \in \bigsqcup_{J \supset K} \mathcal{A}_J \end{aligned}$$

On a donc $P(\bigcap_{i \in K} A_i) = \sum_{J \supset K} P(\mathcal{A}_J)$.

Comme la famille définie par $A_i^* = A_i$ si $i \in J$ et $A_i^* = A_i^c$ si $i \notin J$ est indépendante donc :

$$P(\mathcal{A}_J) = \prod_{i \in J} P(A_i) \times \prod_{i \in I \setminus J} P(A_i^c).$$

$$P\left(\bigcap_{i \in K} A_i\right) = \sum_{J \in H} \left[\prod_{i \in J} P(A_i) \times \prod_{i \in I \setminus J} P(A_i^c) \right] = \sum_{J \in H} \left[\prod_{i \in K} P(A_i) \times \prod_{i \in J \setminus K} P(A_i) \times \prod_{i \in I \setminus J} P(A_i^c) \right]$$

$$P\left(\bigcap_{i \in K} A_i\right) = \prod_{i \in K} P(A_i) \times \sum_{J \in H} \left[\prod_{i \in J \setminus K} P(A_i) \times \prod_{i \in I \setminus J} P(A_i^c) \right]$$

Posons $I' = I \setminus K$ et pour $J \in H$, posons $J' = J \setminus K$.

$$\begin{cases} H & \longrightarrow \mathcal{P}(I') \\ J & \longmapsto J' \end{cases} \text{ est bijective car sa réciproque est } \begin{cases} \mathcal{P}(I') & \longrightarrow H \\ J' & \longmapsto J \sqcup K \end{cases}$$

On fait ce changement d'indice ($J = J' \sqcup K$) dans la somme (avec $I \setminus J = I' \setminus J'$) :

$$P\left(\bigcap_{i \in K} A_i\right) = \prod_{i \in K} P(A_i) \times \sum_{J' \in \mathcal{P}(I')} \left[\prod_{i \in J'} P(A_i) \times \prod_{i \in I' \setminus J'} P(A_i^c) \right]$$

Les $(A_i)_{i \in I'}$ étant indépendants, $\prod_{i \in J'} P(A_i) \times \prod_{i \in I' \setminus J'} P(A_i^c) = P\left(\bigcap_{i \in J'} A_i \cap \bigcap_{i \in I' \setminus J'} A_i^c\right) = P(\mathcal{A}'_{J'})$, et $(\mathcal{A}'_{J'})_{J' \in \mathcal{P}(I')}$ étant totale, $P\left(\bigsqcup_{J' \subset I'} \mathcal{A}'_{J'}\right) = \sum_{J' \subset I'} P(\mathcal{A}'_{J'}) = 1$ donc :
 $P\left(\bigcap_{i \in K} A_i\right) = \prod_{i \in K} P(A_i)$.

Remarques : • Dans la 2^{nde} preuve, on aurait pu factoriser notre expression en utilisant :

$$\prod_{i \in I} (a_i + b_i) = \sum_{J \in \mathcal{P}(I)} \prod_{i \in J} a_i \prod_{i \in I \setminus J} b_i$$

• La 1^{ère} caractérisation s'applique également aux familles I infinies.

Définition : limites inférieures et supérieures :

- $\mathcal{A}^{sup} = \bigcap_{n=0}^{\infty} \left(\bigcup_{k=n}^{\infty} A_k \right)$
- $\mathcal{A}^{inf} = \bigcup_{n=0}^{\infty} \left(\bigcap_{k=n}^{\infty} A_k \right)$

Remarques : • $\omega \in \mathcal{A}^{sup} \iff \{k \in \mathbb{N} \mid \omega \in A_k\}$ est infini. $\omega \in \mathcal{A}^{inf} \iff \{k \in \mathbb{N} \mid \omega \in A_k\}$ est fini.

- \mathcal{A}^{sup} est l'évènement : « une infinité des A_i est réalisée ».
- \mathcal{A}^{inf} est l'évènement : « presque tous les A_i sont réalisés ».

Lemme de Borel – Cantelli : (HP) soit $(A_k)_{k \in \mathbb{N}}$ telle que $\sum_{k=0}^{\infty} P(A_k) < \infty$. Alors $P(\mathcal{A}^{sup}) = 0$.

On pose $B_n = \left(\bigcup_{k=n}^{\infty} A_k \right)$. Par continuité décroissante, comme $B_{n+1} \subset B_n$, $P(\mathcal{A}^{sup}) = \lim_{n \rightarrow \infty} P(B_n)$.
Or $0 \leq P(B_n) \stackrel{\text{Boole}}{\leq} \sum_{k=n}^{\infty} P(A_k)$ et comme reste de série convergente : $\lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} P(A_k) = 0$.
Donc $\lim_{n \rightarrow \infty} P(B_n) = 0$ donc $P(\mathcal{A}^{sup}) = 0$.

Proposition : Soit $(A_k)_{k \in \mathbb{N}}$ indépendants. On a l'alternative :

- $\sum_{k=0}^{\infty} P(A_k) < \infty$ et alors $P(\mathcal{A}^{sup}) = 0$
- $\sum_{k=0}^{\infty} P(A_k) = \infty$ et alors $P(\mathcal{A}^{sup}) = 1$

Preuve : la première partie est le lemme de Borel-Cantelli.
Supposons que $\sum_{k=0}^{\infty} P(A_k) = \infty$. Posons $B_n = \bigcap_{k=n}^{\infty} A_k^c$.
 $\mathcal{A}^{supc} = \left(\bigcap_{n=0}^{\infty} \left(\bigcup_{k=n}^{\infty} A_k \right) \right)^c = \bigcup_{n=0}^{\infty} \bigcap_{k=n}^{\infty} A_k^c = \bigcup_{n=0}^{\infty} B_n$
Par continuité décroissante, comme $B_{n+1} \subset B_n$, $P(\mathcal{A}^{supc}) = \lim_{n \rightarrow \infty} P(B_n)$.
Posons $B_{n,l} = \bigcap_{k=n}^{n+l} A_k^c = A_{n+l}^c \cap B_{n,l-1}$.
Les (A_i) étant indépendants :
 $P(B_{n,l}) = \prod_{k=n}^{n+l} P(A_k^c) = \prod_{k=n}^{n+l} (1 - P(A_k))$
A n fixé, puisque $(B_{n,l})_l$ est décroissante, on a par continuité décroissante :
 $P(B_n) = \lim_{l \rightarrow \infty} P(B_{n,l}) = \lim_{l \rightarrow \infty} \prod_{k=n}^{n+l} (1 - P(A_k)) \leq \lim_{l \rightarrow \infty} \prod_{k=n}^{n+l} e^{-P(A_k)} = \lim_{l \rightarrow \infty} \exp\left(-\sum_{k=n}^{n+l} P(A_k)\right) = 0$
Ainsi, $P(\mathcal{A}^{supc}) = 0$ donc $P(\mathcal{A}^{sup}) = 1$.

Définition : probabilités conditionnelles :

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

Dans le cas où $P(A)$ est nulle, on fixe par convention $P_A(B) = 1$ en sorte que l'expression reste vraie.

▲ Ceci est une convention d'écriture, car elle sous-entend $P(A|B) = P(A)$, soit l'indépendance.

Remarque : si A et B sont appréciables : $P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$ donc on en déduit la formule de Bayes pour deux évènements :

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

Proposition : loi des probabilités totales : soit $(A_i)_{i \in I}$ un système quasi-complet avec I au plus dénombrable, c'est-à-dire tel que : $P(\bigcup_{i \in I} A_i) = 1$ et $\forall i \neq j, P(A_i \cap A_j) = 0$.

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

Preuve : posant $\mathcal{A} = \bigcup_{i \in I} A_i$, $P(\mathcal{A} \cap B) = P(B)$. En effet :
 $\mathcal{A} \subset \mathcal{A} \cup B$ donc $1 = P(\mathcal{A}) \leq P(\mathcal{A} \cup B) = 1$ donc $P(\mathcal{A} \cap B) = P(\mathcal{A}) + P(B) - P(\mathcal{A} \cup B) = P(B)$.
 $P(B) = P(\mathcal{A} \cap B) = P(B \cap \bigcup_{i \in I} A_i) = P(\bigcup_{i \in I} B \cap A_i)$ car les $B \cap A_i$ sont disjoints.
 $P(B) = \sum_{i \in I} P(A_i \cap B)$.

Remarque : si les (A_i) sont de plus appréciables :

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

Théorème : formule de Bayes générale : pour (A_i) système quasi-complet d'évènements appréciables :

$$\forall B \in \mathcal{T} : P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(A_i|B)P(A_i)}{\sum_{j \in I} P(B|A_j)P(A_j)}$$

Variables aléatoires

Définition : un espace probabilisé (E, \mathcal{T}, q) est discret lorsque :

- $\mathcal{T} = \mathcal{P}(E)$
- $\sum_{x \in E} q(x) = 1$

La famille $(q(x))_{x \in E}$ est le germe de la loi de probabilité q . Ce germe détermine entièrement q .

Remarques :

• On rappelle que si une famille de réels positifs est sommable, son support est au plus dénombrable. Par conséquent, $\Omega' = \{x \in E \mid q(x) \neq 0\}$ est dénombrable.

• Toute partie de E étant un évènement, Ω' et $E \setminus \Omega'$ sont des évènements.

Pour $A \subset E$, on notera $A' = A \cap \Omega'$. $q(A') + q(A \cup \Omega') = q(A) + q(\Omega')$

$\forall A \in \mathcal{P}(E) : \Omega'$ étant dénombrable, on peut appliquer : $q(\Omega') = q(\bigsqcup_{x \in \Omega'} \{x\}) = \sum_{x \in \Omega'} q(x) = 1$.

Il en découle que $\forall A \in \mathcal{P}(E), q(A) = q(A') + q(A \cup \Omega') - q(\Omega') = q(A') = \sum_{x \in A \cap \Omega'} q(x) = \sum_{x \in A} q(x)$.

→ On a défini une nouvelle probabilité, $q|_{\Omega'}$.

• Réciproquement, toute famille de réels positifs $(\gamma_x)_{x \in E}$ telle que $\sum_{x \in E} \gamma_x = 1$ est le germe d'une probabilité q sur $(E, \mathcal{P}(E))$ en définissant $\forall A \in \mathcal{P}(E), q(A) \stackrel{\text{def}}{=} \sum_{x \in A} \gamma_x$.

• A priori, si E n'est pas dénombrable, on peut donc toujours se ramener au cas dénombrable en prenant comme univers $\Omega' = \{x \in E \mid q(x) \neq 0\}$ et en considérant l'espace $(\Omega', \mathcal{P}(\Omega'), q|_{\Omega'})$ où q est la probabilité associée à $(E, \mathcal{P}(E))$.

Définition : variables aléatoires

Au sens général (HP), une variable aléatoire X est une application de $\Omega \rightarrow \Omega'$ où (Ω, \mathcal{T}) et (Ω', \mathcal{T}') sont des espaces probabilisables telle que $\forall A \in \mathcal{T}', X^{-1}(A) \in \mathcal{T}$.

Remarques :

- En probabilités, $X^{-1}(A)$ se note $\{X \in A\} = \{\omega \in \Omega \mid X(\omega) \in A\}$. On peut aussi parler de l'évènement « $X \in A$ », toujours abusivement.
- Le plus souvent, $\mathcal{T}' = \mathcal{P}(\Omega')$
- En sup, toute application de $\Omega \rightarrow \Omega'$ était une VA car on avait systématiquement $\mathcal{T} = \mathcal{P}(\Omega)$ et $\mathcal{T}' = \mathcal{P}(\Omega')$ et l'image réciproque d'une partie étant une partie, on vérifiait systématiquement $\forall A \in \mathcal{T}', X^{-1}(A) \in \mathcal{T}$.

Proposition : si X est une VA de (Ω, \mathcal{T}, p) vers (Ω', \mathcal{T}') , alors q est une probabilité sur (Ω', \mathcal{T}') , où q est définie par :

$$\forall A \in \mathcal{T}', q(A) = p(\{X \in A\}) = p(X \in A)$$

Preuve :

- $\{\omega \in \Omega \mid X(\omega) \in \Omega'\} = \Omega$ donc $q(\Omega') = p(\Omega) = 1$.
- Si I est au plus dénombrable, soit $(A_i) \in \mathcal{T}'$ telle que $\forall i \neq j, A_i \cap A_j = \emptyset$, alors :

$$\left\{X \in \bigsqcup_{i \in I} A_i\right\} = X^{-1}\left(\bigsqcup_{i \in I} A_i\right) = \bigsqcup_{i \in I} X^{-1}(A_i)$$
 Puisque $i \neq j$ donne $(X(\omega) \in A_i \implies X(\omega) \notin A_j)$, alors $\omega \in X^{-1}(A_i) \implies \omega \notin X^{-1}(A_j)$ donc $X^{-1}(A_i) \cap X^{-1}(A_j) = \emptyset$. Par conséquent $\left\{X \in \bigsqcup_{i \in I} A_i\right\} = \bigsqcup_{i \in I} X^{-1}(A_i)$. On a donc :

$$q\left(\bigsqcup_{i \in I} A_i\right) = p\left(X \in \bigsqcup_{i \in I} A_i\right) = \sum_{i \in I} P(X \in A_i) = \sum_{i \in I} q(A_i)$$
 q est donc bien une probabilité

Définition : Soit (Ω, \mathcal{T}, p) un espace probabilisé et E un ensemble quelconque.

$X \in E^\Omega$ est une variable aléatoire discrète si :

- $X(\Omega)$ est au plus dénombrable.
- $\forall A \in E, \{X = A\} = X^{-1}(A)$ est un évènement.

Proposition : une VAD est une VA.

Preuve : avec les notations précédentes, posant $\Omega' = X(\Omega)$:

$\Omega' = \bigsqcup_{x \in \Omega'} \{x\}$ donne $\{X \in \Omega'\} = \bigsqcup_{x \in \Omega'} \{X = x\} = \Omega$.

Pour $A, B \in \mathcal{P}(E)^2$: $X^{-1}(A \cup B) = X^{-1}(A) \cup X^{-1}(B)$

$$X^{-1}(A \cap B) = X^{-1}(A) \cap X^{-1}(B)$$

$$X^{-1}(E \setminus A) = \Omega \setminus X^{-1}(A)$$

$$A \subset B \implies X^{-1}(A) \subset X^{-1}(B)$$

$$A \cap B = \emptyset \implies X^{-1}(A) \cap X^{-1}(B) = \emptyset$$

Ainsi : $A = (\Omega' \cap A) \sqcup (A \setminus \Omega')$ (partition entre les éléments ayant des antécédents et ceux n'en ayant pas) donne $\{X \in A\} = \{X \in \Omega' \cap A\} \sqcup \{X \in A \setminus \Omega'\} = \{X \in \Omega' \cap A\} = \bigsqcup_{x \in A \cap \Omega'} \{X = x\}$ est une réunion dénombrable d'évènements, et est donc un évènement.

On a donc $\forall A \in \mathcal{P}(E), \{X \in A\} \in \mathcal{T}$.

Par conséquent, une VAD de $(\Omega, \mathcal{T}) \rightarrow E$ est bien une VA de $(\Omega, \mathcal{T}) \rightarrow (E, \mathcal{P}(E))$.

Proposition : soit P une probabilité sur (Ω, \mathcal{T}) et $\mathcal{L}_X: \begin{cases} \mathcal{P}(E) & \longrightarrow \mathbb{R}_+ \\ A & \longmapsto P(X \in A) \end{cases}$ la loi de X .

Alors $(E, \mathcal{P}(E), \mathcal{L}_X)$ est un espace probabilisé dont le germe est $(P(X = x))_{x \in E}$

Preuve:

$$\sum_{x \in E} P(X = x) = \sum_{x \in X(\Omega)} P(X = x) \stackrel{X(\Omega)}{\text{dénom}} P\left(\bigsqcup_{x \in X(\Omega)} \{x\}\right) = P(X(\Omega)) = P(X \in \Omega') = P(\Omega) = 1$$

Par conséquent, $(E, \mathcal{P}(E), \mathcal{L}_X)$ est un espace probabilisé dont le germe est $(P(X = x))_{x \in E}$.

Définition : Deux variables aléatoires X, Y sont identiquement distribuées si

$$\begin{aligned} \forall A \in \mathcal{P}(E), P(X \in A) &= P(Y \in A) \\ \iff \forall a \in E, P(X = a) &= P(Y = a) \end{aligned}$$

Elles ont donc la même loi de probabilité.

On note alors $X \sim Y$, $X \equiv Y$ ou encore X, Y id (identiquement distribuées).

Remarque : ⚠ Deux VA peuvent avoir la même loi de probabilité sans être identiques.

Proposition – définition : fonction indicatrice. Sur (Ω, \mathcal{T}, p) , pour $A \in \mathcal{T}$:

$$\mathbb{1}_A: \begin{cases} \Omega & \longrightarrow \{0,1\} \\ x & \longmapsto \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases} \end{cases} \text{ est une variable aléatoire discrète}$$

$$\text{Son germe est } \forall x \in \mathbb{R}, p(\mathbb{1}_A = x) = \begin{cases} 1 - P(A) & \text{si } x = 0 \\ P(A) & \text{si } x = 1 \\ 0 & \text{sinon} \end{cases}$$

Définition : VAD indépendantes. Soient X, Y deux VAD sur (Ω, \mathcal{T}, p) avec X à valeurs dans $X(\Omega) \subset E$ et Y à valeurs dans $Y(\Omega) \subset F$ avec E et F quelconques. X et Y sont indépendantes si :

$$\forall A \in \mathcal{P}(E), \forall B \in \mathcal{P}(F), p(X \in A \cap Y \in B) = p(X \in A) \times p(Y \in B)$$

Remarque : cela équivaut à $\forall (A, B) \in \mathcal{P}(E) \times \mathcal{P}(F), \{X \in A\} \cap \{Y \in B\}$ sont indépendants.

• Cela équivaut même à $\forall (a, b) \in E \times F, p(X = a \cap Y = b) = p(X = a) \times p(Y = b)$.

En effet, $\forall (A, B) \in \mathcal{P}(E) \times \mathcal{P}(F), \{X \in A \text{ et } Y \in B\} = \{\omega \in \Omega \mid (X(\omega), Y(\omega)) \in A \times B\} =$

$$\bigsqcup_{(a,b) \in A \times B} \{(X(\omega), Y(\omega)) = (a, b)\} = \bigsqcup_{(a,b) \in \underbrace{A \cap X(\Omega)}_{A'} \times \underbrace{B \cap Y(\Omega)}_{B'}} \{(X(\omega), Y(\omega)) = (a, b)\}$$

$$p(\{X \in A \text{ et } Y \in B\}) = \sum_{(a,b) \in A' \times B'} p(X = a \text{ et } Y = b) = \sum_{a \in A'} \sum_{b \in B'} p(X = a) p(Y = b)$$

$$p(\{X \in A \text{ et } Y \in B\}) = \sum_{a \in A'} p(X = a) \sum_{b \in B'} p(Y = b) = p(X \in A') p(Y \in B') = p(X \in A) p(Y \in B)$$

→ L'indépendance pour les événements élémentaires suffit !

Définition : soit, pour $i \in \llbracket 1, n \rrbracket$, X_i une VAD de (Ω, \mathcal{T}, p) vers E_i . Elles sont indépendantes si :

$$\forall (A_i)_{i \in \llbracket 1, n \rrbracket} \in \prod_{i \in \llbracket 1, n \rrbracket} \mathcal{P}(E_i), \{X_i \in A_i\}_{i \in \llbracket 1, n \rrbracket} \text{ mutuellement indépendants}$$

Théorème : caractérisation des précédents : $(X_i)_{i \in \llbracket 1, n \rrbracket}$ VAD indépendantes si et seulement si :

$$\forall (a_i)_{i \in \llbracket 1, n \rrbracket} \in \prod_{i \in \llbracket 1, n \rrbracket} E_i, p((X_i)_{i \in \llbracket 1, n \rrbracket} = (a_i)_{i \in \llbracket 1, n \rrbracket}) = \prod_{i \in \llbracket 1, n \rrbracket} p(X_i = a_i)$$

Proposition – définition : soit Z une VAD de (Ω, \mathcal{T}, p) vers un ensemble produit $E \times F$. Soit X, Y les composantes de $Z = (X, Y)$.

- La loi de la VA Z s'appelle loi conjointe de X et Y .
- Les lois marginales de Z sont les lois de X et Y , qui sont des VAD.
- Réciproquement, si X et Y sont des VAD à valeurs dans E et F , $Z = (X, Y)$ est bien une VAD.

Preuve : $\Pi_1: \begin{cases} E \times F & \longrightarrow & E \\ (x, y) & \longmapsto & x \end{cases}$ et $\Pi_2: \begin{cases} E \times F & \longrightarrow & F \\ (x, y) & \longmapsto & y \end{cases}$

- Si $Z = (X, Y)$ est une VAD.

$X(\Omega) = \Pi_1(Z(\Omega))$ et comme $Z(\Omega)$ est dénombrable, son image par Π_1 est dénombrable donc $X(\Omega)$ est dénombrable (idem pour Y).

$$\begin{aligned} x \in X(\Omega) &\iff \exists \omega \in \Omega \mid (X(\omega), Y(\omega)) = (x, Y(\omega)) \\ &\iff \exists y \in Y(\Omega) \mid (X(\omega), Y(\omega)) = (x, y) \\ &\iff \exists y \in Y(\Omega) \mid Z(\omega) = (x, y) \end{aligned}$$

Ainsi : $\{X = x\} = \bigsqcup_{y \in Y(\Omega)} \{Z = (x, y)\} \in \mathcal{T}$ comme réunion dénombrable d'éléments de \mathcal{T} .

- Réciproquement : si X et Y sont des VAD à valeurs dans E et F :

$$(x, y) \in Z \implies x \in X(\Omega) \text{ et } y \in Y(\Omega).$$

$Z(\Omega) \subset X(\Omega) \times Y(\Omega)$ et est donc dénombrable.

$\{Z = (x, y)\} = \{X = x\} \cap \{Y = y\}$ est un évènement comme intersection d'évènements.

Notation : on note $\mathcal{L}(\Omega, E)$ l'ensemble des VAD de (Ω, \mathcal{T}) vers E .

On a donc $(X, Y) \in \mathcal{L}(\Omega, E \times F) \iff X \in \mathcal{L}(\Omega, E)$ et $Y \in \mathcal{L}(\Omega, F)$.

Proposition : la loi conjointe de (X, Y) détermine entièrement les lois de X et Y par :

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

Preuve : pour $x \in E$:

$$\{X = x\} = \bigsqcup_{y \in F} \{X = x \cap Y = y\} = \bigsqcup_{y \in Y(\Omega)} \{X = x \cap Y = y\}$$

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

Par conséquent, la donnée de la loi marginale de Y (les $P(Y = y)$), et de, pour chaque $y \in F$, de $x \mapsto P(X = x \mid Y = y)$ (la loi de X conditionnée par Y) donne la loi de X .

Remarque : en cas d'indépendance de X et Y : $P((X, Y) = (x, y)) = P(X = x)P(Y = y)$ et la loi conjointe est entièrement déterminée par les lois marginales de X et de Y .

Trouve un nom à cette merde, tiens :

Soit, pour $n \in \mathbb{N}^*$, $k \in \llbracket 1, n \rrbracket$, $X_k \in \mathcal{L}(\Omega_k, E_k)$. Considérons la famille :

$$\begin{aligned} \left(\gamma_{(x_k)_{k \in \llbracket 1, n \rrbracket}} \right)_{(x_k) \in \prod_{i=1}^n X_i(\Omega_i)} &= \left(\prod_{k=1}^n P(X_k = x_k) \right)_{(x_k) \in \prod_{i=1}^n X_i(\Omega_i)} \\ \sum_{(x_k) \in \prod_{i=1}^n X_i(\Omega_i)} \prod_{k=1}^n P(X_k = x_k) &= \prod_{k=1}^n \sum_{x_i \in X_i(\Omega_i)} P(X_i = x_i) = \prod_{k=1}^n 1 = 1 \end{aligned}$$

Donc cette famille est le germe d'une VAD, qui a un modèle :

$X \in \mathcal{L}(\Omega, \prod_{k=1}^n E_k)$ dont on note les composantes X'_1, \dots, X'_n .

Bien sûr, $P((X'_1, \dots, X'_n) = (x_1, \dots, x_n)) = \prod_{k=1}^n P(X'_k = x_k) \rightarrow X'_1, \dots, X'_n$ sont indépendantes et les X_k et X'_k ont les mêmes lois.

On a donc : étant donné n VAD X_1, \dots, X_n , il y a n VAD indépendantes sur un espace probabilisé telles que pour $k \in \llbracket 1, n \rrbracket$ les X_k et X'_k ont les mêmes lois.

En particulier, pour $X_1 = \dots = X_n = X$, il y a n VAD indépendantes sur un espace probabilisé telles que X et chaque X'_k ont les mêmes lois.

Problème : si les $X_k(\Omega_k), k \in \mathbb{N}$ sont dénombrables, $\prod_{k \in \mathbb{N}} X_k(\Omega_k)$ ne l'est pas nécessairement.

Solution : le programme dit. : on admet que pour une VAD X quelconque sur $(\text{osef}, \text{osaf}, p)$, il y a un (Ω, \mathcal{T}, p) sur lequel il existe une famille $(X_n)_{n \in \mathbb{N}}$ de VAD indépendantes, de même loi que X .

Proposition : $(X_i)_{i \in I}$ famille de VAD sur (Ω, \mathcal{T}, p) mutuellement indépendantes.

On considère $J, K \in \mathcal{P}_f(I) \mid J \cap K = \emptyset$. Alors $X = (X_i)_{i \in J}, Y = (X_i)_{i \in K}$ sont indépendantes.

$$\left| \begin{array}{l} \text{Preuve : pour } (x_i)_{i \in J \cup K} \in \prod_{i \in J \cup K} E_i, \text{ par indépendance des } X_i : \\ P((X_i)_{i \in J \cup K} = (x_i)_{i \in J \cup K}) = \prod_{i \in J \cup K} p(X_i = x_i) = \prod_{i \in J} p(X_i = x_i) \times \prod_{i \in K} p(X_i = x_i) \\ P((X_i)_{i \in J \cup K} = (x_i)_{i \in J \cup K}) = P((X_i)_{i \in J} = (x_i)_{i \in J}) \times P((X_i)_{i \in K} = (x_i)_{i \in K}) \\ P((X_i)_{i \in J \cup K} = (x_i)_{i \in J \cup K}) = P(X = x)P(Y = y) \\ \rightarrow X \text{ et } Y \text{ sont indépendantes.} \end{array} \right.$$

Proposition : $(X_i)_{i \in I}$ famille de VAD sur (Ω, \mathcal{T}, p) mutuellement indépendantes et S un index fini.

On considère $(J_j)_{j \in S} \in \mathcal{P}_f(I)^{\#S} \mid \forall i \neq j \in S^2, J_j \cap J_i = \emptyset$.

Alors $((X_i)_{i \in J_j})_{j \in S}$ sont indépendantes.

Preuve identique à la proposition précédente, simplement un peu plus lourde.

Proposition : soit $X \in \mathcal{L}(\Omega, E)$ et $f \in F^E$. On note $f(X) = f \circ X$ en probabilités. On a alors :

$$f(X) \in \mathcal{L}(\Omega, F)$$

$$\left| \begin{array}{l} \text{Preuve : } (f \circ X)(\Omega) = f(X(\Omega)) \text{ est au plus dénombrable car l'image par une application d'un ensemble} \\ \text{(au plus) dénombrable est (au plus) dénombrable.} \\ \text{Pour un } y \in F, \quad \omega \in \{f(X) = y\} \iff f(X(\omega)) = y \iff \exists x \in X(\Omega) \mid f(x) = y \quad \text{et} \quad X(\omega) = x \\ \iff \omega \in \bigsqcup_{x \in X(\Omega) \cap f^{-1}(\{y\})} \{X = x\} \\ \text{Par conséquent, } \{f(X) = y\} = \bigsqcup_{x \in X(\Omega) \cap f^{-1}(\{y\})} \{X = x\} \text{ donc } \{f(X) = y\} \in \mathcal{T} \end{array} \right.$$

Remarque : pour A partie de F , on remarque qu'on a $\{f(X) \in A\} = \bigsqcup_{x \in X(\Omega) \cap f^{-1}(A)} \{X = x\}$, qui doit être un réflexe.

Exemple : $(X, Y, Z, T) \in \mathcal{L}(\Omega, \mathbb{K})^4$. Alors la VAD $(X, Y, Z, T) \in \mathcal{L}(\Omega, \mathbb{K}^4)$.

Prenant $f_1: \begin{cases} \mathbb{K}^4 & \longrightarrow \mathbb{K} \\ (x, y, z, t) & \longmapsto \frac{\sin x}{1+|yz|^2} \end{cases}$ (ou bien $x + y + zt$ ou n'importe quoi), $f_1(X, Y, Z, T)$ est une VAD.

Proposition : si $(X, Y) \in \mathcal{L}(\Omega, E) \times \mathcal{L}(\Omega, F)$ sont indépendantes, et $f \in E'^E, g \in F'^F$:

$$f(X) \text{ et } g(Y) \text{ sont indépendantes}$$

Les images par des fonctions quelconques de VAD indépendantes sont indépendantes.

$$\begin{aligned}
& \text{Preuve : soit } (x, y) \in E' \times F' \\
& p(f(X) = x \text{ et } g(Y) = y) = \sum_{(\alpha, \beta) \in (X(\Omega) \cap f^{-1}(\{x\})) \times (Y(\Omega) \cap g^{-1}(\{y\}))} p(X = \alpha \text{ et } Y = \beta) \\
& p(f(X) = x \text{ et } g(Y) = y) = \sum_{\alpha \in (X(\Omega) \cap f^{-1}(\{x\}))} \sum_{\beta \in (Y(\Omega) \cap g^{-1}(\{y\}))} p(X = \alpha \text{ et } Y = \beta) \\
& p(f(X) = x \text{ et } g(Y) = y) = \sum_{\alpha \in (X(\Omega) \cap f^{-1}(\{x\}))} \sum_{\beta \in (Y(\Omega) \cap g^{-1}(\{y\}))} p(X = \alpha) p(Y = \beta) \\
& p(f(X) = x \text{ et } g(Y) = y) = \sum_{\alpha \in (X(\Omega) \cap f^{-1}(\{x\}))} p(X = \alpha) \sum_{\beta \in (Y(\Omega) \cap g^{-1}(\{y\}))} p(Y = \beta) \\
& p(f(X) = x \text{ et } g(Y) = y) = p(f(X) = x) \times p(g(Y) = y)
\end{aligned}$$

Proposition : si X_1, \dots, X_n sont des VAD indépendantes à valeurs dans E_1, \dots, E_n , pour $f_k \in F_k^{E_k}$, alors :
 $(f_k(X_k))_{k \in \llbracket 1, n \rrbracket}$ sont indépendantes

Lemme des coalitions :

- Soit $(X_i)_{i \in I} \in \prod_{i \in I} (\mathcal{L}(\Omega, E_i))$ indépendantes.
- Soit $(J_j)_{j \in K}$ famille de parties finies de I 2 à 2 disjointes, pour $j \in K$, soit $F_j = \prod_{i \in J_j} E_i$.
- Soit $(G_j)_{j \in K}$ une famille d'ensembles quelconques.
- Soit $(f_j)_{j \in K} \in (G_j^{F_j})^K$.

Alors les VAD $(f_j(X_k)_{k \in J_j})_{j \in K}$ sont indépendantes.

La preuve est la compilation des propositions précédentes.

Définition : espérance. Pour $X \in \mathcal{L}(\Omega, \mathbb{K})$. On dit que X a une espérance si $(xP(X = x))_{x \in \mathbb{K}}$ est sommable.

Dans ce cas, l'espérance de X est :

$$E(X) = \sum_{x \in \mathbb{K}} xP(X = x) = \sum_{x \in X(\Omega)} xP(X = x)$$

On note $\mathcal{L}^1(\Omega, \mathbb{K})$ l'ensemble des VAD ayant une espérance finie.

Remarques : • Si $X(\Omega)$ est fini, X a une espérance.

- Si $X(\Omega)$ est borné (par M), X a une espérance car $\sum_{x \in X(\Omega)} |x|P(X = x) \leq M \sum_{x \in X(\Omega)} P(X = x) = M$.
- Lorsque X est à valeurs positives, même si $(xP(X = x))_{x \in \mathbb{K}}$ n'est pas sommable, $\sum_{x \in X(\Omega)} xP(X = x)$ a un sens et vaut $+\infty$. Ce n'est pas une VAD à espérance finie.

Théorème de transfert : (a.k.a TT) soit $X \in \mathcal{L}(\Omega, E)$, $f \in \mathbb{K}^E$ et $Y = f(X)$.

$$f(X) \in \mathcal{L}^1(\Omega, \mathbb{K}) \iff (f(x)P(X = x))_{x \in X(\Omega)} \text{ sommable}$$


En ce cas :

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

$$\begin{aligned}
& \text{Preuve : rappel : pour } y \in \mathbb{K}, \\
& \{f(X) = y\} = \bigsqcup_{x \in X(\Omega) \cap f^{-1}(y)} \{X = x\} = \bigsqcup_{\substack{x \in X(\Omega) \\ f(x) = y}} \{X = x\} \\
& \text{On remarque que : } X(\Omega) = \bigsqcup_{y \in \mathbb{K}} \{X(\Omega) \cap f^{-1}(y)\} \\
& \text{Sans souci de sommabilité pour une famille à termes positifs :}
\end{aligned}$$

$$\begin{aligned}
\sum_{y \in Y(\Omega)} |y| P(Y = y) &= \sum_{y \in Y(\Omega)} |y| P\left(\bigsqcup_{x \in X(\Omega) \cap f^{-1}(y)} \{X = x\}\right) = \sum_{y \in Y(\Omega)} |y| \sum_{x \in X(\Omega) \cap f^{-1}(y)} P(X = x) \\
\sum_{y \in Y(\Omega)} |y| P(Y = y) &= \sum_{x \in \bigsqcup_{y \in \mathbb{K}} \{X(\Omega) \cap f^{-1}(y)\}} |f(x)| P(X = x) = \sum_{x \in X(\Omega)} |f(x)| P(X = x) \\
\text{On a donc } f(X) &\in \mathcal{L}^1(\Omega, \mathbb{K}) \iff (f(x)P(X = x))_{x \in X(\Omega)} \text{ sommable. En ce cas, le même calcul donne :} \\
E(f(X)) &= \sum_{x \in X(\Omega)} f(x)P(X = x)
\end{aligned}$$

Remarques : • Il est essentiel de comprendre qu'ici, E est complètement quelconque.

•  Le théorème de transfert ne donne jamais l'existence de l'espérance, mais simplement son expression en cas d'existence.

Proposition : pour $X \in \mathcal{L}(\Omega, \mathbb{K})$, $E(|X|) < \infty \iff X$ a une espérance. En ce cas, $|E(X)| \leq E(|X|)$.

$$\begin{aligned}
\text{Soit } f: \begin{cases} \mathbb{K} & \longrightarrow \mathbb{R} \\ x & \longmapsto |x| \end{cases}. \text{ Par le théorème de transfert :} \\
E(|X|) = E(f(X)) < \infty &\iff \sum_{x \in \mathbb{R}} |f(x)| P(X = x) = \sum_{x \in \mathbb{R}} |x| P(X = x) < \infty \\
\text{En ce cas :} \\
|E(X)| = \left| \sum_{y \in \mathbb{K}} y P(X = y) \right| &\leq \sum_{y \in \mathbb{K}} |y| P(X = y) \stackrel{\text{TT}}{=} \sum_{y \in \mathbb{K}} |y| P(|X| = y) = E(|X|)
\end{aligned}$$

Théorème : $\mathcal{L}^1(\Omega, \mathbb{K})$ est un \mathbb{K} -ev, et, sur ce \mathbb{K} -ev, E est une forme linéaire.

$$\begin{aligned}
\text{Preuve : soit } (X, Y, \lambda) &\in \mathcal{L}^1(\Omega, \mathbb{K})^2 \times \mathbb{K}, Z = (X, Y) \text{ et } f: (x, y) \mapsto x + y. \\
\text{Existence de l'espérance de } f(Z) : \\
\sum_{z \in \mathbb{K}^2} |f(z)| P(Z = z) &= \sum_{(x, y) \in \mathbb{K}^2} |x + y| P((X, Y) = (x, y)) \stackrel{\text{fub}}{=} \sum_{x \in \mathbb{K}} \sum_{y \in \mathbb{K}} |x + y| P((X, Y) = (x, y)) \\
\sum_{z \in \mathbb{K}^2} |f(z)| P(Z = z) &\leq \sum_{x \in \mathbb{K}} \sum_{y \in \mathbb{K}} |x| P(X = x \text{ et } Y = y) + \sum_{x \in \mathbb{K}} \sum_{y \in \mathbb{K}} |y| P(X = x \text{ et } Y = y) \\
\sum_{z \in \mathbb{K}^2} |f(z)| P(Z = z) &\leq \sum_{x \in \mathbb{K}} |x| \left(\sum_{y \in \mathbb{K}} P(X = x \text{ et } Y = y) \right) + \sum_{y \in \mathbb{K}} |y| \left(\sum_{x \in \mathbb{K}} P(X = x \text{ et } Y = y) \right) \\
\sum_{z \in \mathbb{K}^2} |f(z)| P(Z = z) &\leq \sum_{x \in X(\Omega)} |x| P(X = x) + \sum_{x \in X(\Omega)} |y| P(Y = y) \\
\sum_{z \in \mathbb{K}^2} |f(z)| P(Z = z) &\leq E(|X|) + E(|Y|) \\
\text{On a donc } f(Z) &\in \mathcal{L}^1(\Omega, \mathbb{K}). \text{ Le même calcul donne alors } E(Z) = E(X) + E(Y). \\
\text{Pour la multiplication par un scalaire, on considère } g_\lambda: \begin{cases} \mathbb{K} & \longrightarrow \mathbb{K} \\ x & \longmapsto \lambda x \end{cases} \text{ et on applique le théorème de} \\
\text{transfert.} \\
\text{On en déduit également que } \mathcal{L}^1(\Omega, \mathbb{K}) &\text{ est un } \mathbb{K}\text{-ev.}
\end{aligned}$$

Proposition : monotonie de l'espérance. Soient $(X, Y) \in \mathcal{L}^1(\Omega, \mathbb{R})^2$ telles que $X < Y$. Alors :

$$E(X) \leq E(Y)$$

$$\begin{aligned}
\text{Preuve : si } Z &\in \mathcal{L}(\Omega, \mathbb{R}_+), E(Z) \geq 0. \\
\text{Soit } (X, Y) &\in \mathcal{L}^1(\Omega, \mathbb{R})^2 \text{ telles que } X < Y. \text{ Alors :} \\
0 \leq E(Y - X) &= E(Y) - E(X)
\end{aligned}$$

Remarques : • $E(|X|) = 0 \iff P(X = 0) = 1$.

• Pour $A \in \mathcal{T}$, $E(\mathbb{1}_A) = P(A)$. Cela montre que la connaissance de l'espérance mène à la connaissance de la loi de probabilité.

Proposition : pour $(X, Y) \in \mathcal{L}^1(\Omega, \mathbb{K})$ indépendantes, $XY \in \mathcal{L}^1(\Omega, \mathbb{K})$ et $E(XY) = E(X)E(Y)$.

$$\left| \begin{array}{l} \text{Preuve : le lemme des coalitions donne que } |X| \text{ et } |Y| \text{ sont indépendantes.} \\ E(|XY|) \stackrel{\text{TT}}{=} \sum_{(x,y) \in \mathbb{K}^2} |xy| P(X=x \text{ et } Y=y) \\ E(|XY|) = \sum_{x \in \mathbb{K}} \sum_{y \in \mathbb{K}} |x||y| P(X=x) P(Y=y) = E(X)E(Y) \\ \text{On a donc } XY \in \mathcal{L}^1(\Omega, \mathbb{K}) \text{ et } E(XY) = E(X)E(Y). \end{array} \right.$$

Remarque : il est absolument essentiel que les deux variables soient indépendantes. Par exemple : Prenons X telle que $\forall n \in \mathbb{N}^*, P(X=n) = \frac{1}{\zeta(3)n^3}$ ($\sum_{n \in \mathbb{N}^*} P(X=n) = \frac{\zeta(3)}{\zeta(3)} = 1$, on a bien une proba). $E(X) = \sum_{n \in \mathbb{N}^*} n P(X=n) = \frac{\zeta(2)}{\zeta(3)} \rightarrow$ Elle a une espérance finie. En appliquant la proposition précédente : $E(X^2) = \sum_{n \in \mathbb{N}^*} \frac{n^2}{\zeta(3)n^3} = +\infty \rightarrow$ Elle n'a pas d'espérance.

Formule du crible : (Ω, \mathcal{T}, P) et $(A_1, \dots, A_n) \in \mathcal{T}$.

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{J \in \mathcal{P}(\llbracket 1, n \rrbracket) \setminus \{\emptyset\}} (-1)^{\#J+1} P\left(\bigcap_{k \in J} A_k\right)$$

$$\left| \begin{array}{l} \text{Preuve : notant } B = \bigcup_{k=1}^n A_k. \\ P(B^c) = E(\mathbb{1}_{B^c}) \\ P(B^c) = E\left(\prod_{k=1}^n (1 - \mathbb{1}_{A_k})\right) \\ P(B^c) = E\left(\sum_{J \in \mathcal{P}(\llbracket 1, n \rrbracket)} \prod_{k \in J} (-\mathbb{1}_{A_k}) \prod_{k \notin J} 1\right) \\ P(B^c) = E\left(\sum_{J \in \mathcal{P}(\llbracket 1, n \rrbracket)} (-1)^{\#J} \mathbb{1}_{\bigcap_{k \in J} A_k}\right) \\ P(B^c) = \sum_{J \in \mathcal{P}(\llbracket 1, n \rrbracket)} (-1)^{\#J} E(\mathbb{1}_{\bigcap_{k \in J} A_k}) \\ P(B^c) = \sum_{J \in \mathcal{P}(\llbracket 1, n \rrbracket)} (-1)^{\#J} P\left(\bigcap_{k \in J} A_k\right) \\ P(B) = 1 - \sum_{J \in \mathcal{P}(\llbracket 1, n \rrbracket)} (-1)^{\#J} P\left(\bigcap_{k \in J} A_k\right) = \sum_{J \in \mathcal{P}(\llbracket 1, n \rrbracket) \setminus \{\emptyset\}} (-1)^{\#J+1} P\left(\bigcap_{k \in J} A_k\right) \end{array} \right.$$

Convention : on rencontre fréquemment des VA à valeurs dans $\overline{\mathbb{R}}$. En ce cas, lorsque $p(X = \pm\infty) = 0$, on convient que $\pm\infty p(X = \pm\infty) = 0$, en sorte que son espérance soit toujours définie.

Exemple : lors d'une suite de lancers de pile ou face, la probabilité de ne jamais obtenir d'affilée Pile, Face, Pile, Pile est nulle. Si $X = k$ signifie que l'on a Pile, Face, Pile, Pile au rang k et pas avant, $p(X = \infty) = 0$. Néanmoins, l'espérance (temps moyen pour obtenir le motif) n'est pas affectée pathologiquement.

Changement de convention : dorénavant, si $p(A) = 0$, nous conviendrons que $p(B|A) = p(B)$.

⚠ Ceci s'oppose à la convention habituelle définie page 4, qui disait $p(B|A) = 1$.

Inégalité de Markov : soit $X \in \mathcal{L}(\Omega, \mathbb{R}_+)$ et $a > 0$.

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

Preuve 1 :

$$E(X) = \sum_{x \in \mathbb{R}_+} xP(X=x) = \sum_{0 \leq x \leq a} xP(X=x) + \sum_{a \leq x} xP(X=x) \geq a \sum_{a \leq x} P(X=x) \geq aP(X \geq a)$$

Preuve 2 :

$$aP(X \geq a) = aE(\mathbb{1}_{X \geq a}) = E(a\mathbb{1}_{X \geq a})$$

Si $X(\omega) \geq a$, alors $a\mathbb{1}_{X \geq a} = a \leq X(\omega)$

Sinon, $a\mathbb{1}_{X \geq a} = 0 \leq X(\omega)$.

Ainsi, $a\mathbb{1}_{X \geq a} \leq X$ donc par croissance de l'espérance : $aP(X \geq a) \leq E(X)$

Remarques : • La seconde preuve est l'archétype d'une idée puissante : le fait qu'une inégalité proba-espérance est une inégalité espérance-espérance en passant par les fonctions indicatrices, et résulte souvent de la croissance de l'espérance.

Définition : on dit que $X \in \mathcal{L}(\Omega, \mathbb{K})$ a un moment d'ordre $k \in \mathbb{N}^*$ si $E(|X|^k) < \infty$.

Son moment est alors $m_k(X) \stackrel{\text{def}}{=} E(X^k)$

On note alors $\mathcal{L}^k(\Omega, \mathbb{K})$ l'ensemble des VAD ayant un moment d'ordre k .

Cette définition peut également fonctionner si $k \notin \mathbb{N}^*$, dans le cas où $X = e^{k \ln X}$ a un sens, soit pour $X > 0$.

Proposition : pour $\alpha > \beta \geq 1$, $\mathcal{L}^\alpha(\Omega, \mathbb{K}) \subset \mathcal{L}^\beta(\Omega, \mathbb{K})$.

Preuve :

Pour $\omega \in \Omega$, ou bien $|X(\omega)| \geq 1$ et alors $|X(\omega)| \leq |X(\omega)|^\alpha + 1$

ou bien $|X(\omega)| \leq 1$ et alors $|X(\omega)| \leq 1 + |X(\omega)|^\alpha$

Donc $|X| \leq 1 + |X|^\alpha$ et donc $E(|X|) \leq 1 + E(|X|^\alpha)$ et donc $\mathcal{L}^1(\Omega, \mathbb{K}) \subset \mathcal{L}^\alpha(\Omega, \mathbb{K})$.

Pour $\alpha > \beta \geq 1$, $\frac{\alpha}{\beta} \geq 1$ donc

$$E(|X|^\alpha)^{\frac{\beta}{\alpha}} < \infty \implies E(|X|^\alpha) < \infty$$

Proposition : $\mathcal{L}^2(\Omega, \mathbb{K})$ est un espace vectoriel.

Preuve : pour $X \in \mathcal{L}^2(\Omega, \mathbb{K})$, et $\lambda \in \mathbb{K}$, $m_2(\lambda X) = \lambda^2 m_2(X)$.

Pour $(X, Y) \in \mathcal{L}^2(\Omega, \mathbb{K})$. $XY \in \mathcal{L}^1(\Omega, \mathbb{K})$. En effet :

$$(|X| - |Y|)^2 \geq 0 \text{ donc } |XY| \leq \frac{|X|^2 + |Y|^2}{2} \text{ donc } E(|XY|) \leq E\left(\frac{|X|^2 + |Y|^2}{2}\right) < \infty.$$

$$E(|X| + |Y|) = E(|XY|) + E(|Y|^2) + E(|X|^2) < \infty \text{ et } X + Y \in \mathcal{L}^2(\Omega, \mathbb{K})$$

Remarque : Le résultat est en réalité vrai pour tous les $\mathcal{L}^k(\Omega, \mathbb{K})$.

Proposition : $\varphi: \begin{cases} \mathcal{L}^2(\Omega, \mathbb{R}) & \longrightarrow & \mathbb{R} \\ (X, Y) & \longmapsto & E(XY) \end{cases}$ est (presque) un produit scalaire.

Le « presque » est dû au fait que la séparation n'est pas totalement vérifiée. En effet :

$\varphi(X, X) = 0 \iff E(X^2) = 0 \iff P(X = 0) = 1$. La variable peut néanmoins prendre des valeurs non-nulles avec une probabilité nulle.

La preuve est triviale (symétrie, bilinéarité, positivité, séparation).

Définition : pour $(X, Y) \in \mathcal{L}^2(\Omega, \mathbb{K})$, la covariance de X et Y est :

$$\text{cov}(X, Y) := E\left((X - E(X))(Y - E(Y))\right) = E(XY) - E(X)E(Y)$$

On notera $\tilde{X} = X - E(X)$, en sorte que $\text{cov}(X, Y) = E(\tilde{X}\tilde{Y})$

Remarque : on a par exemple : $E(\tilde{X}) = 0$; $\widetilde{X + \lambda Y} = \tilde{X} + \lambda\tilde{Y}$. Cela montre que :

$\text{cov}(X, Y + \lambda Z) = \text{cov}(X, Y) + \lambda \text{cov}(X, Z) \rightarrow$ on a la bilinéarité. On a également la positivité.

Pour ce qui est de la séparation, $\text{cov}(X, X) = 0 \iff P(\tilde{X} = 0) = 1 \iff X$ est presque sûrement constante.

C'est toujours pas ça, mais on progresse. On associe dorénavant une norme : $\|X\|^2 = \text{cov}(X, X)$.

Définition : pour $(X, Y) \in \mathcal{L}^2(\Omega, \mathbb{K})$, la variance de X et Y est :

$$V(X) := \text{cov}(X, X) = E(X^2) - E(X)^2$$

Son écart-type est :

$$\sigma(X) := \sqrt{V(X)} = \|X\|$$

Remarque : deux variables indépendantes ont une covariance nulle mais la réciproque est fausse.

Théorème : inégalité de Cauchy-Schwartz. Soit $(X, Y) \in \mathcal{L}^2(\Omega, \mathbb{K})$.

$$|\text{cov}(X, Y)| \leq \sigma(X)\sigma(Y)$$

Preuve : Si la covariance était un vrai produit scalaire, on aurait :

si $X = 0$ ou $Y = 0$, on a toujours $\langle X|Y \rangle \leq \|X\|\|Y\|$. Sinon :

$$\left\|\frac{X}{\|X\|} - \frac{Y}{\|Y\|}\right\| = \left\|\frac{Y}{\|Y\|}\right\| = 1.$$

$$0 \leq \left\|\frac{X}{\|X\|} - \frac{Y}{\|Y\|}\right\|^2 = \left\|\frac{X}{\|X\|}\right\|^2 + \left\|\frac{Y}{\|Y\|}\right\|^2 - 2\frac{\langle X|Y \rangle}{\|X\|\|Y\|} \implies \langle X|Y \rangle \leq \|X\|\|Y\|.$$

Néanmoins, ici, il se peut que $Y \neq 0$ mais que $\|Y\| = 0$. Dans ce cas :

Pour $\lambda \in \mathbb{R}$:

$\|X + \lambda Y\|^2 \geq 0$ donne $\|X\|^2 + \lambda^2\|Y\|^2 + 2\lambda\langle X|Y \rangle \geq 0$ donc $\|X\|^2 + 2\lambda\langle X|Y \rangle \geq 0$, qui est affine sur \mathbb{R} donc négative à un moment (absurde), sauf si $\langle X|Y \rangle = 0$, ce qui est donc le cas.

Ainsi, $\langle X|Y \rangle \leq \|X\|\|Y\| \iff 0 \leq 0$ se vérifie encore.

Proposition – définition : pour $(X, Y) \in \mathcal{L}^2(\Omega, \mathbb{K})$ non presque sûrement constantes ($\sigma(X) > 0$ et $\sigma(Y) > 0$), le coefficient de corrélation de X et Y est :

$$\text{cor}(X, Y) := \text{cov}\left(\frac{X}{\sigma(X)}, \frac{Y}{\sigma(Y)}\right) = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)} \in [-1, 1]$$

Proposition : soit $X \in \mathcal{L}(\Omega, \mathbb{N})$:

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

Preuve :

$$\sum_{n=0}^{\infty} P(X > n) = \sum_{n=0}^{\infty} \sum_{k=n+1}^{\infty} P(X = k) = \sum_{k=1}^{\infty} P(X = k) \sum_{n=0}^{k-1} 1 = \sum_{k=1}^{\infty} kP(X = k) = E(X)$$

L'interversion de sommes est un Fubini en ajoutant des 0 pour que la somme triangulaire devienne rectangulaire.

Proposition : pour $(X_1, \dots, X_n) \in \mathcal{L}^2(\Omega, \mathbb{K})$:

$$V\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n V(X_k) + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j)$$

Proposition – définition : chaînes de Markov discrètes. Soit $(X_k)_{k \in \mathbb{N}}$ une suite de VAD à valeurs dans le même \mathcal{E} dénombrable, l'ensemble des états. On décrit une chaîne de Markov lorsque :

Fausse propriété de Markov :

$$\forall k \in \mathbb{N}, \forall i, j \in \mathcal{E}, P_{X_k=i}(X_{k+1} = j) = p_{i,j} \text{ ne dépend pas de } k.$$

Vraie propriété de Markov : soit $(H_k) = ((X_0, \dots, X_k))$

$$P(X_{k+1} = j | X_k = i \text{ et } H_{k-1} = h) = p_{i,j} \text{ ne dépend pas de } k$$

Ce résultat dénote une situation concrète : on a un ensemble d'état, et $p_{i,j}$ est la probabilité d'aller de l'état i à l'état j .

Méthode : Avec les notations précédentes, si $\mathcal{E} = \{1, \dots, N\}$ est fini :

Soit $A = [p_{i,j}]_{(i,j) \in \llbracket 1, n \rrbracket^2}$. Ses colonnes somment 1 et ses lignes aussi.

Notant $U_k = \begin{pmatrix} P(X_k = 1) \\ \vdots \\ P(X_k = n) \end{pmatrix}$ le vecteur d'état du système, on a $\forall p \in \mathbb{N} : U_p = ({}^t A)^p U_0$.

On le prouve en utilisant la loi des probabilités totales.

→ on peut donc décrire l'état du système à chaque instant grâce à la matrice A et ses puissances.

Il est donc bon de savoir la réduire.

Proposition – définition : soit $X \in \mathcal{L}(\Omega, \mathbb{R})$. Pour $f_t : x \mapsto t^x$, $t^X = f_t(X)$ est une VAD.

Sans présumer du domaine de définition, la fonction génératrice de X est :

$$G_X : \begin{cases} \mathbb{R} & \longrightarrow & \mathbb{R} \\ t & \longmapsto & E(t^X) \end{cases}$$

Par théorème de transfert, dans le cas où $E(t^X)$ est bien défini :

$$G_X : t \mapsto \sum_{x \in \mathbb{R}} t^x p(X = x)$$

Remarque : si (X, Y) sont indépendantes, pour $t \in \mathcal{D}_{G_X} \cap \mathcal{D}_{G_Y} : f_t(X)$ et $f_t(Y)$ sont indépendantes (coalitions) et ont une espérance donc leur produit aussi et $E(t^X t^Y) = E(t^X)E(t^Y)$.

Par conséquent, $\mathcal{D}_{G_X} \cap \mathcal{D}_{G_Y} \subset \mathcal{D}_{G_{X+Y}}$. Cela se généralise par récurrence :

Proposition : pour (X_1, \dots, X_n) mutuellement indépendantes :

$$\bigcap_{k=1}^n \mathcal{D}_{G_{X_k}} \subset \mathcal{D}_{G_{\sum_{k=1}^n X_k}}$$

Exemple : marche de l'ivrogne.

$$\forall k \geq 1, p(X_k = 1) = p \text{ et } p(X_k = -1) = 1 - p = q$$

Les X_k sont indépendantes et identiquement distribuées (de même loi que X).

$$\forall t \in \mathcal{D}_{G_X}, G_X(t) = \sum_{x \in \{-1, 1\}} t^x p(X = x) = pt + \frac{q}{t}$$

Par conséquent, S_n étant la somme des pas faits ($S_n = \sum_{k=0}^n X_k$) :

$$G_{S_n} = (G_X)^n = \begin{cases} \mathbb{R}^* & \rightarrow \mathbb{R} \\ t & \mapsto \left(pt + \frac{q}{t}\right)^n \end{cases}$$

$$\forall t \in \mathbb{R}_+^*, G_{S_n}(t) = \sum_{i=-n}^n t^i p(S_n = i) = \left(pt + \frac{q}{t}\right)^n = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} t^{2k-n}$$

On multiplie par t^n pour identifier les coefficients polynômiaux : $\sum_{i=-n}^n t^{i+n} p(S_n = i) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} t^{2k}$

Si $\neg(i \equiv n[2])$, $p(S_n = i) = 0$.

Si $i \equiv n[2]$, posant $k = \frac{i+n}{2}$: $p(S_n = i) = \binom{n}{\frac{i+n}{2}} p^{\frac{i+n}{2}} q^{\frac{n-i}{2}}$.

Proposition : soit $X \in \mathcal{L}(\Omega, \mathbb{R}) / X(\Omega) \subset \mathbb{N}$. $\forall t \in \mathcal{D}_{G_X}$:

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

\rightarrow La fonction génératrice est une série entière. Puisque $G_X(1) = 1$, le rayon de convergence de cette série est au moins 1.

Proposition : pour $(X, Y) \in \mathcal{L}(\Omega, \mathbb{N})$ telles que $G_X = G_Y$. Par unicité du DSE :

$$\forall n \in \mathbb{N}, P(X = n) = P(Y = n) = \frac{G^{(n)}(0)}{n!}$$

Proposition : soit $X \in \mathcal{L}(\Omega, \mathbb{N})$. $\forall t \in \mathcal{D}_{G_X}$:

$$G'_X(t) = \sum_{n=1}^{\infty} n P(X = n) t^{n-1}$$

Remarque : a priori, cette formule peut être en défaut sur le cercle d'incertitude ($t = 1$). En effet, on conserve le rayon de convergence en dérivant, mais on ne sait rien à sa frontière. Néanmoins :

Proposition : on a l'alternative :

- $G_X \in C^1([0,1])$ et alors $X \in \mathcal{L}^1(\Omega, \mathbb{N})$ et $E(X) = G'_X(1)$.
- $E(X) = +\infty = \lim_{t \rightarrow 1} G'_X(t)$.

Remarques : • $G'_X(t) = \sum_{n=1}^{\infty} n P(X = n) t^{n-1} \geq 0 \implies G_X$ est croissante.

- $G''_X(t) = \sum_{n=2}^{\infty} n(n-1) P(X = n) t^{n-2} \geq 0 \implies G_X$ est convexe.

- $X \in \mathcal{L}^2(\Omega, \mathbb{N}) \iff G''_X \in C^0([0,1])$ et en ce cas, $G''_X(1) = E(X^2) - E(X) = V(X) - G'_X(1) + (G'_X(1))^2$.

	Support	$P(X = k)$	Espérance Variance	Fonction génératrice
Uniforme	$\llbracket 1, n \rrbracket$	$\frac{1}{n}$	$\frac{(n+1)/2}{(n^2-1)/12}$	$\frac{t}{n} \frac{1-t^n}{1-t}$
Bernoulli \mathcal{B}_p	$\{0,1\}$	$P(X=1) = p$ $P(X=0) = q = 1-p$	$\frac{p}{pq}$	$q + pt$
Binomiale $\mathcal{B}_{n,p}$	$\llbracket 0, n \rrbracket$	$\binom{n}{k} p^k q^{n-k}$	$\frac{np}{npq}$	$(q + pt)^n$
Géométrique (p)	\mathbb{N}^*	pq^{k-1}	$\frac{1/p}{q/p^2}$	$\frac{pt}{1-qt} = \frac{pt}{pt-t+1}$
Poisson (λ)	\mathbb{N}	$e^{-\lambda} \frac{\lambda^k}{k!}$	$\frac{\lambda}{\lambda}$	$e^{\lambda(t-1)}$
(HP) Hypergéométrique $(m, n) \in \llbracket 1, N \rrbracket^2$	$\llbracket 0, \min(m, n) \rrbracket$	$\frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}}$	$\frac{mn/N}{\frac{mn(N-n)(N-m)}{N^2(N-1)}}$?
(HP) Binomiale négative (n, p)	\mathbb{N}	$\binom{k+n-1}{n-1} p^n (1-p)^k$	$\frac{n(1-p)/p}{n(1-p)/p^2}$	$\left(\frac{p}{1-qt}\right)^n$
(HP) Loi de Pascal (n, p)	$\llbracket n, +\infty \rrbracket$	$\binom{k-1}{n-1} p^n (1-p)^{k-n}$	$\frac{n/p}{n(1-p)/p^2}$	$\left(\frac{pt}{1-qt}\right)^n$

Loi géométrique : elle est décrite dans le tableau. Elle correspond à l'expérience :

On a une succession de tirages suivant une loi de Bernoulli de paramètre p .

$P(X = k)$ est la probabilité de gagner pour la première fois à l'instant k . Alors X suit une loi géométrique de paramètre p .

Loi de Pascal : la loi d'attente du n -ième succès dans une succession de tirages de Bernoulli de paramètre p .

Loi binomiale : X suit une loi binomiale de paramètres (n, p) si elle est la somme de n variables de Bernoulli indépendantes de paramètre p .

Proposition : si $(p_n) \in [0,1]^{\mathbb{N}}$ est telle que $p_n \xrightarrow{+\infty} \frac{\lambda}{n}$ et si X_n suit une loi de Bernoulli de paramètres n, p_n , alors, $\forall k \in \mathbb{N}, \lim_{n \rightarrow \infty} P(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}$.

Proposition : inégalité de Bienaymé-Tchebychev. Soit $X \in \mathcal{L}(\Omega, \mathbb{K})$ admettant une espérance. Soit $\alpha > 0$

$$P(|X - E(X)| > \alpha) \leq \frac{V(X)}{\alpha^2}$$

Proposition – définition : loi faible des grands nombres. Soit $(X_k)_{k \in \mathbb{N}} \in \mathcal{L}^2(\Omega, \mathbb{R})$ indépendantes 2 à 2 et identiquement distribuées. Soit $\mu = E(X) = E(X_k)$ et $S_n = \frac{1}{n} \sum_{k=1}^n X_k$. Alors :

$$\forall \varepsilon > 0 : p(|S_n - \mu| \geq \varepsilon) \leq \frac{V(X)}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

Preuve :

$$S_n = \frac{1}{n} \sum_{k=1}^n X_k \Rightarrow E(S_n) = \frac{1}{n} \sum_{k=1}^n E(X_k) = \frac{n\mu}{n} = \mu$$

Par conséquent, $\widetilde{S}_n = S_n - \mu$ est centrée. Par l'inégalité de Markov : (carré pour la positivité)

$$p(|\widetilde{S}_n| \geq \varepsilon) = p(\widetilde{S}_n^2 \geq \varepsilon^2) \leq \frac{E(\widetilde{S}_n^2)}{\varepsilon^2}$$

$E(\widetilde{S}_n)^2 = 0$, donc :

$$p(|S_n - \mu| \geq \varepsilon) \leq \frac{E(\widetilde{S}_n^2) - E(\widetilde{S}_n)^2}{\varepsilon^2} = \frac{V(\widetilde{S}_n)}{\varepsilon^2}$$

$$V(\widetilde{S}_n) = V(S_n - \mu) = V(S_n) = V\left(\frac{1}{n} \sum_{k=1}^n X_k\right) \stackrel{Pythagore}{=} \frac{1}{n^2} \sum_{k=1}^n V(X_k) = \frac{V(X)}{n}$$

Enfin : $p(|S_n - \mu| \geq \varepsilon) \leq \frac{V(X)}{n\varepsilon^2}$.

Proposition – définition : $X \in \mathcal{L}(\Omega, \mathbb{N}^*)$ est dite sans mémoire si :

$$\forall (m, n) \in \mathbb{N}^2, P(X > m + n \mid X > m) = P(X > n)$$

De manière équivalente : $\exists p \mid X$ suit la loi géométrique de paramètre p .