

Table des matières

12 Probabilités	3
12.1 Rappels et compléments	3
12.1.1 Espace probabilisé	3
12.1.1.1 Tribu	3
12.1.1.2 Tribu engendrée par une partie	3
12.1.1.3 Un exemple : la tribu borélienne de $\mathbb{R}$	4
12.1.1.4 Probabilités	5
12.1.1.5 Construction de probabilité dans le cas Ω fini	7
12.1.1.6 Construction de probabilité dans le cas Ω dénombrable	7
12.1.1.7 Continuité monotone séquentielle	7
12.1.1.8 Événement négligeable, événement quasi certain :	8
12.1.2 Indépendance, Probabilité conditionnelle	8
12.1.2.1 Événements indépendants	8
12.1.2.2 Probabilité conditionnelle	9
12.1.2.3 Formule des probabilités totales :	9
12.1.2.4 Formule des probabilités composées :	10
12.2 Variables aléatoires	11
12.2.1 Définitions, premières propriétés	11
12.2.1.1 Définitions	11
12.2.1.2 Quelques propriétés	11
12.2.2 Loi d'une variable aléatoire	12
12.2.3 Fonction de répartition	13
12.2.4 Indépendance des v.a.r, conditionnement	14
12.2.4.1 Indépendance	14
12.2.4.2 Conditionnement	14
12.2.5 Variables aléatoires réelles discrètes	14
12.2.5.1 Définitions	14
12.2.5.2 Variables aléatoires discrètes usuelles :	15
12.2.6 Variables aléatoires continues à densité	17
12.2.6.1 Définitions, propriétés	17
12.2.6.2 Variables continues à densité usuelles	18
12.2.7 Somme de deux variables aléatoires indépendantes	20
12.2.7.1 Deux variables discrètes	20
12.2.7.2 Une variable continue et une autre discrète	20
12.2.7.3 Deux variables continues	21
12.2.8 Moments, espérance, variance, écart type d'une v.a.r.	21
12.2.8.1 Espérance mathématique : Définition	21
12.2.8.2 Espérance des variables usuelles	22
12.2.8.3 Propriété de transfert	22
12.2.8.4 Transfert pour deux variables	22
12.2.8.5 Linéarité de l'espérance mathématique	22
12.2.8.6 Autres propriétés de l'espérance mathématique	23
12.2.8.7 Moments	24
12.2.8.8 Covariance	25
12.2.8.9 Variable aléatoire centrée, réduite	27
12.3 Fonction génératrice	27
12.3.1 Définition, propriétés	28
12.3.2 Fonction génératrice d'une somme de v.a.r. indépendantes	30
12.4 Inégalités, Convergence	30

12.4.1	Inégalité	30
12.4.1.1	Inégalité de Markov	30
12.4.1.2	Inégalité de Bienaymé-Chebychev	31
12.4.1.3	inégalité de Jensen	31
12.4.2	Convergence	32
12.4.2.1	Convergence en loi	32
12.4.2.2	Convergence en probabilité	33
12.4.3	Lois faible des grands nombres et théorème limite central	34
12.4.3.1	Loi faible des grands nombres	34
12.4.3.2	Théorème limite central	34
12.5	Annexe	36
12.5.1	Tableau récapitulatif des variables aléatoires réelles discrètes usuelles	36
12.5.2	Tableau récapitulatif des variables aléatoires réelles continues à densité usuelles :	37
12.5.3	Stabilité de certaines lois	38
12.5.3.1	Loi de Bernoulli, loi binomiale	38
12.5.3.2	Lois de Poisson	38
12.5.3.3	Loi Γ	38
12.5.3.4	Loi normale	39

Chapitre 12

Probabilités

12.1 Rappels et compléments

12.1.1 Espace probabilisé

12.1.1.1 Tribu

Definition 12.1.1

Soit Ω un ensemble. On appelle tribu sur Ω , toute partie $\mathcal{T}$ de $\mathcal{P}(\Omega)$ tel que :

- (i) $\Omega \in \mathcal{T}$
- (ii) $\forall X \in \mathcal{T}, X^c \in \mathcal{T}$
- (iii) Pour toute famille au plus dénombrable $(X_i)_{i \in I}$ d'éléments de $\mathcal{T}$, on a $\bigcup_{i \in I} X_i \in \mathcal{T}$.

Le couple $(\Omega, \mathcal{T})$ s'appelle *espace probabilisable*.

Remarques. Si $(\Omega, \mathcal{T})$ est un espace probabilisable, alors :

1. Si $\mathcal{T}$ est une tribu sur Ω , tout élément A de $\mathcal{T}$ s'appelle : *événement* de l'espace probabilisable $(\Omega, \mathcal{T})$
2. $\emptyset \in \mathcal{T}$
3. $\forall A, B \in \mathcal{T}$, on a : $A \cup B, A \cap B, A \setminus B, A \Delta B$ sont dans $\mathcal{T}$.
4. Pour tout $A_1, \dots, A_n \in \mathcal{T}$, on a : $\bigcup_{k=1}^n A_k$ et $\bigcap_{k=1}^n A_k$ sont dans $\mathcal{T}$.
5. Pour toute famille $(A_n)_{n \geq 0}$ d'éléments de $\mathcal{T}$, $\bigcap_{n \geq 0} A_n \in \mathcal{T}$.

12.1.1.2 Tribu engendrée par une partie

Soient Ω et I des ensembles.

Proposition 12.1.1. Si $(\mathcal{T}_i)_{i \in I}$ est une famille de tribus sur Ω alors $\mathcal{T} = \bigcap_{i \in I} \mathcal{T}_i$ est une tribu.

Preuve. Élémentaire, en utilisant la définition d'une tribu. ■

Considérons une partie $\mathcal{A}$ de $\mathcal{P}(\Omega)$ et soit $T_{\mathcal{A}}$ l'ensemble des tribus $\mathcal{T}$ tel que $\mathcal{A} \subset \mathcal{T}$. On note $\mathcal{T}(\mathcal{A})$ l'intersection de toutes les tribus éléments de $T_{\mathcal{A}}$, c'est-à-dire :

$$\mathcal{T}(\mathcal{A}) = \bigcap_{\mathcal{T} \in T_{\mathcal{A}}} \mathcal{T}.$$

La proposition 12.1.1 ci-dessus montre que $\mathcal{T}(\mathcal{A})$ est une tribu. On a :

Proposition-Définition 12.1.1. $\mathcal{T}(\mathcal{A})$ est la plus petite tribu sur Ω contenant $\mathcal{A}$. $\mathcal{T}(\mathcal{A})$ s'appelle la tribu engendrée par $\mathcal{A}$

Preuve. Soit $\mathcal{S}$ une tribu tel que $\mathcal{A} \subset \mathcal{S}$, alors $\mathcal{S} \in T_{\mathcal{A}}$, par suite $\bigcap_{\mathcal{T} \in T_{\mathcal{A}}} \mathcal{T} \subset \mathcal{S}$, donc $\mathcal{T}(\mathcal{A}) \subset \mathcal{S}$. ■

Remarques. On donne les remarques suivantes :

1. Si $\mathcal{A}_1$ et $\mathcal{A}_2$ sont deux parties de $\wp(\Omega)$ alors :

$$\mathcal{A}_1 \subset \mathcal{A}_2 \Rightarrow \mathcal{T}(\mathcal{A}_1) \subset \mathcal{T}(\mathcal{A}_2)$$

2. $\mathcal{T}(\{\emptyset\}) = \mathcal{T}_g = \{\emptyset, \Omega\}$

3. Si $\mathcal{A}$ est une tribu de Ω alors $\mathcal{T}(\mathcal{A}) = \mathcal{A}$.

Preuve. 1. La tribu $\mathcal{T}(\mathcal{A}_2)$ contient $\mathcal{A}_2$, et comme $\mathcal{A}_1 \subset \mathcal{A}_2$, elle contient $\mathcal{A}_1$ et comme $\mathcal{T}(\mathcal{A}_1)$ est la plus petite tribu contenant $\mathcal{A}_1$, on a $\mathcal{T}(\mathcal{A}_1) \subset \mathcal{T}(\mathcal{A}_2)$.
 2. La plus petite tribu contenant $\emptyset$ doit contenir $\Omega = \emptyset^c$, donc elle doit contenir $\mathcal{T}_g = \{\emptyset, \Omega\}$, or $\mathcal{T}_g$ est elle même une tribu, donc $\mathcal{T}(\{\emptyset\}) = \mathcal{T}$.
 3. Si $\mathcal{A}$ est une tribu, il est clair que $\mathcal{A}$ est la plus petite tribu contenant $\mathcal{A}$. ■

12.1.1.3 Un exemple : la tribu borélienne de $\mathbb{R}$

Definition 12.1.2

Soit $\mathcal{O}$ l'ensemble des ouverts de l'ensemble $\mathbb{R}$ des nombres réels muni de la valeur absolue. La tribu de $\mathbb{R}$, engendrée par $\mathcal{O}$ s'appelle la tribu borélienne de $\mathbb{R}$, notée $\mathcal{B}_{\mathbb{R}}$.

Ainsi la tribu borélienne $\mathcal{B}_{\mathbb{R}}$ de $\mathbb{R}$ est la tribu $\mathcal{T}(\mathcal{O})$ engendrée par les ouverts de $\mathbb{R}$. On donne le théorème suivants et des remarques qui lui sont attachées. Le théorème sera d'une utilité importante par la suite. Avant de donner ce théorème, on donne et on démontre un lemme :

Lemme 12.1.1. Tout ouvert non vide de $\mathbb{R}$ est une réunion au plus dénombrable d'intervalles ouverts bornés. Autrement dit pour tout $W \in \mathcal{O}$, il existe une partie I de $\mathbb{N}$ tel que $W = \bigcup_{n \in I}]a_n, b_n[$ avec pour tout $n \in I$, $a_n, b_n \in \mathbb{R}$ et $a_n < b_n$.

Preuve. Soit O un ouvert non vide de $\mathbb{R}$. Pour tout $x \in O$, il existe $u, v \in \mathbb{R}$ tel que $x \in]u, v[$ et $]u, v[\subset O$. Par densité de $\mathbb{Q}$ dans $\mathbb{R}$, il existe $r(x), r'(x)$ rationnels tel que $u < r(x) < x < r'(x) < v$. Il en découle que $O = \bigcup_{x \in O}]r(x), r'(x)[$. Si on note $I(x) =]r(x), r'(x)[$, alors l'ensemble $J = \{I(x)/x \in O\}$ est au plus dénombrable car l'application : $\phi : J \rightarrow \mathbb{Q}^2, I(x) \mapsto (r(x), r'(x))$ est injective et $\mathbb{Q}^2$ est dénombrable. ϕ est injective car si $I(x) =]r(x), r'(x)[$ et $I(y) =]r(y), r'(y)[$ sont deux éléments de J avec $x, y \in O$, si $\phi(I(x)) = \phi(I(y))$ alors $\begin{cases} r(x) = r(y) \\ r'(x) = r'(y) \end{cases}$, ce qui veut dire $I(x) = I(y)$. ■

Théorème 12.1.1. La tribu borélienne $\mathcal{B}_{\mathbb{R}}$ est égale à la tribu engendrée par l'ensemble des intervalles de la forme $] - \infty, a]$ avec $a \in \mathbb{R}$.

Preuve. Notons $\mathcal{A} = \{] - \infty, a] / a \in \mathbb{R} \}$. Nous allons démontrer que $\mathcal{B}_{\mathbb{R}} = \mathcal{T}(\mathcal{A})$.

• Soit $W \in \mathcal{O}$ un ouvert non vide de $\mathbb{R}$, alors en vertu de lemme 12.1.1, W est une union au plus dénombrable d'intervalles de la forme $]a, b[$ avec $a < b$. Or un tel intervalle s'écrit : $]a, b[=] - \infty, b[\setminus] - \infty, a]$ et comme $] - \infty, b[= \bigcup_{n \in \mathbb{N}^*}] - \infty, b - \frac{1}{n}]$, on a $]a, b[\in \mathcal{T}(\mathcal{A})$, donc $W \in \mathcal{T}(\mathcal{A})$ et $\mathcal{B}_{\mathbb{R}} \subset \mathcal{T}(\mathcal{A})$.

• Réciproquement si $a \in \mathbb{R}$ alors on a :

$$] - \infty, a] =] - \infty, a[\cup \{a\}$$

et :

$$\{a\} = \bigcap_{n \in \mathbb{N}^*} \left] a - \frac{1}{n}, a + \frac{1}{n} \right[$$

ce qui montre que $] - \infty, a]$ s'obtient à partir de parties ouvertes de $\mathbb{R}$ par intersection au plus dénombrable, différence et union finie, donc $] - \infty, a]$ est bien un élément de la tribu engendrée par $\mathcal{O}$ donc $\mathcal{A} \subset \mathcal{B}_{\mathbb{R}}$, et

par suite $\mathcal{T}(\mathcal{A}) \subset \mathcal{T}(\mathcal{B}_{\mathbb{R}}) = \mathcal{B}_{\mathbb{R}}$. ■

Outre le cas où $\mathcal{A}$ est l'ensemble des intervalles de la forme $] - \infty, a]$ avec $a \in \mathbb{R}$, on a aussi et on peut le démontrer : $\mathcal{B}_{\mathbb{R}} = \mathcal{T}(\mathcal{A})$ pour les cas suivants :

1. $\mathcal{A}$ l'ensemble des intervalle de $\mathbb{R}$.
2. $\mathcal{A}$ l'ensemble des intervalles ouverts bornés $]a, b[$ (resp. $[a, b], [a, b[,]a, b]$ avec a, b réels $a \leq b$.)
3. $\mathcal{A}$ l'ensemble des intervalles de la forme reps. $] - \infty, a[$ avec a réel.
4. $\mathcal{A}$ l'ensemble des intervalles de la forme $[a, +\infty[$ (reps. $]a, +\infty[$) avec a réel.

On s'intéressera plus au cas où $\mathcal{A} = \{] - \infty, a] / a \in \mathbb{R} \}$, qui sera à la base de la définition de la fonction de répartition d'une variable aléatoire réelle.

Preuve. Note sur la preuve(non nécessaire dans une première lecture) : Tenant compte du fait que tout ouvert de $\mathbb{R}$ est un réunion au plus dénombrable d'intervalles ouverts bornés, on procède comme suit :

- Pour montrer que la tribu borélienne est engendrée par les intervalles de la forme $]a, b]$, on peut remarquer que $]a, b[= \bigcup_{n \in \mathbb{N}^*}]a, b - \frac{1}{n}]$ où I est l'ensemble des entiers naturels non nuls tel que $a < b - \frac{1}{n}$, ainsi la tribu engendrée par les intervalles de la forme $]a, b]$ contient celle engendrée par les ouverts, or un intervalle de la forme $]a, b]$ peut s'écrire $]a, b] = \bigcap_{n \in \mathbb{N}^*}]a, b + \frac{1}{n}[$, d'où l'inclusion réciproque.
- Finalement pour les intervalles de la forme $] - \infty, a]$ on peut remarque que si (a_n) est une strictement décroissante de nombre réels tel que $\lim_{n \rightarrow +\infty} a_n = -\infty$ et $a_0 = a$ alors $] - \infty, a] = \bigcup_{n \in \mathbb{N}}]a_{n+1}, a_n]$ et on vient de voir que les intervalles de la forme $]x, y]$ avec $x, y \in \mathbb{R}$ et $x < y$ sont des éléments de la tribu borélienne. Réciproquement un intervalle borné ouvert $]a, b[$ peut s'écrire $]a, b[=] - \infty, b[\setminus] - \infty, a]$ et $\{b\} = \bigcap_{n \in \mathbb{N}^*}]b - \frac{1}{n}, b]$. ■

12.1.1.4 Probabilités

Definition 12.1.3

Soit $(\Omega, \mathcal{T})$ un espace probabilisable. on appelle probabilité sur cet espace toute application $\mathbb{P} : \mathcal{T} \rightarrow \mathbb{R}_+$ tel que :

- (1) $\mathbb{P}(\Omega) = 1$
 - (2) $\mathbb{P}(\bigcup_{n=0}^{+\infty} A_n) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n)$ pour toute famille (A_n) d'éléments deux à deux disjoints de $\mathcal{T}$.
- le triplet $(\Omega, \mathcal{T}, \mathbb{P})$ est appelé *espace probabilisé*.

Remarques. On considère un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$, alors :

1. Si (A_n) est une famille d'événements deux à deux disjoints alors $\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = 0$.
2. $\mathbb{P}(\emptyset) = 0$
3. $\forall A, B \in \mathcal{T}$ tel que $A \cap B = \emptyset$, on a $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
4. $\forall A, B \in \mathcal{T}$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
5. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ pour tout $A \in \mathcal{T}$.
6. Pour tout $A, B \in \mathcal{T}$, on a : $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$.
7. Pour tout $A \in \mathcal{T}$, on a $\mathbb{P}(A) \in [0, 1]$.
8. Pour toute famille $(A_k)_{0 \leq k \leq n}$ d'événements, on a $\mathbb{P}(\bigcup_{k=0}^n A_k) \leq \sum_{k=0}^n \mathbb{P}(A_k)$
9. Pour toute famille $(A_n)_{n \in \mathbb{N}}$ d'éléments de $\mathcal{T}$, on a : $\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n=0}^{+\infty} \mathbb{P}(A_n)$, le second membre de cette inégalité appartient à $\mathbb{R}_+ \cup \{+\infty\}$.
10. Si Ω est au plus dénombrable, le choix $\mathcal{T} = \mathcal{P}(\Omega)$ est le plus usuel.
11. Si $(\Omega, \mathcal{T}, \mathbb{P})$ est un espace probabilisé, un élément de la tribu $\mathcal{T}$ s'appelle un événement.

- Preuve.** 1. Comme la série $\sum \mathbb{P}(A_n)$ est convergente, le résultat en découle immédiatement.
2. Posons $A_n = \emptyset$ pour tout $n \in \mathbb{N}$, alors $(A_n)_{n \geq 0}$ est une famille d'événements deux à deux disjoints, donc, compte tenu du 1) ci-dessus, la suite constante $\mathbb{P}(\emptyset)$ converge vers 0, d'où $\mathbb{P}(\emptyset) = 0$.
3. Posons $A_0 = A, A_1 = B, A_n = \emptyset, \forall n \geq 2$, alors $(A_n)_{n \geq 0}$ est une suite d'événements deux à deux disjoints, et par l'axiome (2), on a $\mathbb{P}(A \cup B) = \mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n) = \mathbb{P}(A) + \mathbb{P}(B)$ puisque $\mathbb{P}(A_n) = \mathbb{P}(\emptyset) = 0, \forall n \geq 2$.
4. On a $A = (A \setminus B) \cup B$ et $(A \setminus B) \cap B = \emptyset$, donc $\mathbb{P}(A \cup B) = \mathbb{P}(A \setminus B) + \mathbb{P}(B)$. On a $\mathbb{P}(A) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B)$ (mêmes arguments), il en découle que $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$. Compte tenu de (1) et (2), on a : $\mathbb{P}(A \cup B) = \mathbb{P}(A) - \mathbb{P}(A \cap B) + \mathbb{P}(B)$.
5. Il suffit d'appliquer l'une des propriétés précédentes à A et A^c .
6. Si $A \subset B$, alors $B = A \cup (B \setminus A)$ et c'est une réunion disjointe, donc $\mathbb{P}(A) = \mathbb{P}(B) + \mathbb{P}(A \setminus B)$ donc $\mathbb{P}(B) \geq \mathbb{P}(A)$.
7. Comme $\emptyset \subset A \subset \Omega$, on déduit d'après la remarque ci-dessus que $0 = \mathbb{P}(\Omega) \leq \mathbb{P}(A) \leq \mathbb{P}(\Omega) = 1$
8. Par récurrence sur n : Pour $n = 0$ on a égalité, pour $n = 1$, on a

$$\mathbb{P}(A_0) + \mathbb{P}(A_1) = \mathbb{P}(A_0 \cup A_1) + \mathbb{P}(A_0 \cap A_1),$$

d'où le résultat. Si la propriété est vraie pour $n \in \mathbb{N}$, on a

$$\mathbb{P}\left(\bigcup_{k=0}^{n+1} A_k\right) = \mathbb{P}\left(\left(\bigcup_{k=0}^n A_k\right) \cup A_{n+1}\right).$$

Comme la propriété est vraie pour $n = 1$, on a

$$\mathbb{P}\left(\left(\bigcup_{k=0}^n A_k\right) \cup A_{n+1}\right) \leq \mathbb{P}\left(\bigcup_{k=0}^n A_k\right) + \mathbb{P}(A_{n+1}).$$

On conclut en appliquant l'hypothèse de récurrence.

9. Notons que la série $\sum \mathbb{P}(A_n)$ est une série à termes positifs donc deux cas sont possibles : soit $\sum_{n=0}^{+\infty} \mathbb{P}(A_n) = +\infty$ et l'inégalité à prouver est vérifiée puisque le membre de gauche est un réel (par convention $\forall x \in \mathbb{R}, x \leq +\infty$), soit la série est convergente de somme ℓ . Si on pose $B_0 = A_0$ et $B_n = A_n \setminus \bigcup_{0 \leq k \leq n-1} A_k$, alors :
- les B_n sont des événements deux à deux disjoints, en effet si $p, q \in \mathbb{N}$ tel que $p < q$, on a $B_q = A_q \setminus \bigcup_{k=0}^{q-1} A_k$, donc pour tout $x \in B_q$, on a $x \notin A_p$, donc $x \notin B_p$.
 - On a : $\bigcup_{k=0}^n B_k = \bigcup_{k=0}^n A_k$, en effet pour tout $n \in \mathbb{N}$, on a $B_n \subset A_n \subset \bigcup_{k \in \mathbb{N}} A_k$. Réciproquement, soit $\omega \in \bigcup_{k=1}^n A_k$, si $\omega \in A_0$ alors $\omega \in B_0$, donc $\omega \in \bigcup_{k=0}^n B_k$, sinon soit ℓ le plus petit entier naturel de $\llbracket 1, n \rrbracket$ tel que $\omega \in A_\ell$ alors $1 \leq \ell \leq n$ et $\omega \notin A_j, \forall j \in \llbracket 0, \ell - 1 \rrbracket$ et $\omega \in A_\ell$, donc $\omega \in B_\ell$ et finalement $\omega \in \bigcup_{k=0}^n B_k$. Il en résulte aussi que $A = \bigcup_{n=0}^{+\infty} A_n = \bigcup_{n=0}^{+\infty} B_n$, donc :

$$\begin{aligned} \mathbb{P}(A) &= \sum_{n=0}^{+\infty} \mathbb{P}(B_n) \\ &= \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{k=0}^n B_k\right) \\ &= \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{k=0}^n A_k\right) \\ &\leq \lim_{n \rightarrow +\infty} \sum_{k=0}^n \mathbb{P}(A_k) \\ &\leq \lim_{n \rightarrow +\infty} \sum_{k=0}^{+\infty} \mathbb{P}(A_k) \end{aligned}$$

■

12.1.1.5 Construction de probabilité dans le cas Ω fini

Si $n \in \mathbb{N}^*$ et $\Omega = \{\omega_1, \dots, \omega_n\}$ alors toute probabilité sur l'espace probabilisable $(\Omega, \mathcal{P}(\Omega))$ est parfaitement déterminée par un élément $p = (p_1, \dots, p_n) \in \mathbb{R}_+^n$ tel que $\sum_{k=1}^n p_k = 1$. La donnée de p permet de construire $\mathbb{P}$ comme suit :

- Pour tout $k \in \llbracket 1, n \rrbracket$, on pose $\mathbb{P}(\{\omega_k\}) = p_k$.
- Pour toute partie A de Ω , on pose $\mathbb{P}(A) = 0$ si $A = \emptyset$ et $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$ si $A \neq \emptyset$.
- Réciproquement si $\mathbb{P}$ est une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ alors en notant $p_k = \mathbb{P}(\{\omega_k\})$, pour tout $k \in \llbracket 1, n \rrbracket$, on a $p = (p_1, \dots, p_n) \in \mathbb{R}_+^n$ et $\sum_{k=1}^n p_k = 1$.

12.1.1.6 Construction de probabilité dans le cas Ω dénombrable

Si $\Omega = \{\omega_n / n \in \mathbb{N}\}$ alors toute probabilité sur l'espace probabilisable $(\Omega, \mathcal{P}(\Omega))$ est parfaitement déterminée par une famille $p = (p_n)_{n \geq 0}$ de nombres réels positifs tel que $\sum_{k=0}^{+\infty} p_k = 1$. La donnée de p permet de construire $\mathbb{P}$ comme suit :

- Pour tout $n \in \mathbb{N}$, on pose $\mathbb{P}(\{\omega_n\}) = p_n$.
- Pour toute partie A de Ω , on pose $\mathbb{P}(A) = 0$ si $A = \emptyset$ et $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$ si $A \neq \emptyset$.
- Réciproquement si $\mathbb{P}$ est une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ alors en notant $p_n = \mathbb{P}(\{\omega_n\})$, pour tout $n \in \mathbb{N}$, on a $p = (p_n)_{n \geq 0}$ est une famille de réels positifs et $\sum_{n=0}^{+\infty} p_n = 1$.

12.1.1.7 Continuité monotone séquentielle

Proposition 12.1.2. Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé. Alors :

- Pour toute suite croissante $(A_n)_{n \geq 1}$ d'événements, on a :

$$\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \sup_n \mathbb{P}(A_n) = \mathbb{P}(\bigcup_n A_n)$$

- Pour toute suite décroissante $(A_n)_{n \geq 1}$ d'événements, on a :

$$\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \inf_n \mathbb{P}(A_n) = \mathbb{P}(\bigcap_n A_n)$$

Preuve. Soit $(A_n)_{n \in \mathbb{N}}$ une suite croissante d'événements et $A = \bigcup_{n \in \mathbb{N}} A_n$. Posons $B_0 = A_0$ et pour tout

$B_n = A_n \setminus \bigcup_{k=0}^{n-1} A_k$. Alors :

- Les B_n sont deux à deux disjoints. En effet si $n, m \in \mathbb{N}$ tel que $n < m$ Soit $\omega \in B_m$ alors $\omega \in A_m$ et $\omega \notin A_k, \forall k \in \llbracket 0, m-1 \rrbracket$, en particulier $\omega \notin A_n$.

- On a : $\forall n \in \mathbb{N}, A_n = \bigcup_{k=0}^n B_k$. En effet, $B_k \subset A_k \subset A_n$ pour tout $k \in \mathbb{N}$ donc $\bigcup_{k=0}^n B_k \subset A_n$. Réciproquement, soit $\omega \in A_n$. Si $\omega \in A_0$ alors $\omega \in B_0$, sinon, on a forcément $n \geq 1$ et $\omega \notin A_0$ et $\exists j \in \llbracket 1, n \rrbracket, \omega \in A_j$. Choisissons j minimal, donc $\omega \in A_j$ et $\omega \notin A_k, \forall k \in \llbracket 0, j-1 \rrbracket$, c'est-à-dire : $\omega \in B_j$, donc $\omega \in \bigcup_{k=0}^n B_k$.

Notons que d'après ce qui précède on a aussi $A = \bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} B_n$ et les B_n deux à deux disjoints. n tel que $\omega \in \mathbb{N}$. prenons n minimal donc $\omega \in A_n$ et $\omega \notin A_k, \forall k \in \llbracket 0, n-1 \rrbracket$. d'où $\omega \in B_n$. La suite réelle $\mathbb{P}(A_n)$ est majorée (par $\mathbb{P}(A)$), croissante, donc elle est convergente. Comme $\mathbb{P}(A_n) = \sum_{k=0}^n \mathbb{P}(B_k)$, on a

$$\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \sum_{k=0}^{+\infty} \mathbb{P}(B_k) = \mathbb{P}(A).$$

Soit (A_n) une famille décroissante d'événements, alors la famille (A_n^c) des complémentaires est croissante. D'après le premier point ci-dessus, on a :

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n^c\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n^c) = \sup_{n \in \mathbb{N}} \mathbb{P}(A_n^c).$$

Comme

$$\bigcup_{n \in \mathbb{N}} A_n^c = \left(\bigcap_{n \in \mathbb{N}} A_n \right)^c,$$

on a :

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

Cette dernière égalité est justifiée par la convergence de la suite $\mathbb{P}(A_n)$ qui est décroissante positive. Finalement, et compte tenu de la décroissance de la suite $\mathbb{P}(A_n)$, on a :

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

■

12.1.1.8 Événement négligeable, événement quasi certain :

Definition 12.1.4

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé. Un événement $A \in \mathcal{T}$ tel que $\mathbb{P}(A) = 0$ est dit négligeable.

Proposition-Définition 12.1.2. Soit $(\Omega, \mathcal{T}, \mathbb{P})$ est un espace probabilisé et $A \in \mathcal{T}$ alors les assertions suivantes sont équivalentes :

1. $\mathbb{P}(A) = 1$
2. $\forall X \in \mathcal{T}, \mathbb{P}(A \cap X) = \mathbb{P}(X)$.

Si une d'elles est vraie, on dit que A est un événement quasi certain.

Preuve. Supposons qu'on a (1). Soit $X \in \mathcal{T}$, alors $X = (X \cap A) \cup (X \cap A^c)$, l'union étant disjointe, donc $\mathbb{P}(X) = \mathbb{P}(A \cap X) + \mathbb{P}(X \cap A^c)$. On a $X \cap A^c \subset A^c$ et $\mathbb{P}(A^c) = 1 - \mathbb{P}(A) = 0$, donc $\mathbb{P}(X \cap A^c) = 0$ et finalement $\mathbb{P}(A \cap X) = \mathbb{P}(X)$.

Réciproquement il suffit de prendre $X = \Omega$.

■

12.1.2 Indépendance, Probabilité conditionnelle

12.1.2.1 Événements indépendants

Definition 12.1.5

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé. et $(A_j)_{j \in J}$ une famille quelconque d'événements. On dit que les événements A_j sont indépendants si pour toute partie K finie non vide de J on a $\mathbb{P}\left(\bigcap_{k \in K} A_k\right) = \prod_{k \in K} \mathbb{P}(A_k)$.

On dit aussi que les événements $(A_j)_{j \in J}$ sont mutuellement indépendants.

Exemples. On lance deux fois de suite un dé non truqué et on marque le couple de nombres obtenus. On considère les événements suivants :

A : « Obtenir 1 au premier lancé ».

B : « Obtenir 2 au deuxième lancé ».

C : « Obtenir 1 ou 2 au premier lancer ».

Alors A et B sont indépendants tandis que A et C ne sont pas indépendants. En effet, on a : $\text{card}(A) = \text{card}(B) = 6$ et $\text{card}(A \cap B) = 1$ et $\text{card}(C) = 12$ et $\text{card}(A \cap C) = \text{card}(A) = 6$, alors :

$$\mathbb{P}(A \cap B) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(A)\mathbb{P}(B)$$

Tandis que :

$$\mathbb{P}(A \cap C) = \mathbb{P}(A) = \frac{1}{6}$$

et

$$\mathbb{P}(A)\mathbb{P}(C) = \frac{1}{6} \cdot \frac{1}{3} = \frac{1}{18}$$

12.1.2.2 Probabilité conditionnelle

Proposition-Définition 12.1.3. $(\Omega, \mathcal{T}, \mathbb{P})$ est un espace probabilisé et $B \in \mathcal{T}$ tel que $\mathbb{P}(B) \neq 0$. Pour tout $X \in \mathcal{T}$, on pose :

$$\mathbb{P}_B(X) = \frac{\mathbb{P}(X \cap B)}{\mathbb{P}(B)}$$

Alors $\mathbb{P}_B$ est une probabilité sur $(\Omega, \mathcal{T})$ appelée probabilité conditionnelle (conditionnée par l'événement B). Le réel $\mathbb{P}_B(X)$ noté aussi $\mathbb{P}(X|B)$

Preuve. • Il est clair que $\mathbb{P}_B(X) \in \mathbb{R}_+$, pour tout $X \in \mathcal{T}$.

• On a $\mathbb{P}_B(\Omega) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$.

• Si $(A_n)_{n \in \mathbb{N}}$ est une famille d'événements deux à deux incompatibles et $A = \bigcup_{n \in \mathbb{N}} A_n$, alors :

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

Or $(A_n \cap B)_{n \in \mathbb{N}}$ est une famille d'événements deux à deux disjoints, donc :

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

ce qui fournit :

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

Donc $\mathbb{P}_B$ est une probabilité sur $(\Omega, \mathcal{T})$. ■

Remarques. On peut faire les remarques suivantes :

1. Si $B \in \mathcal{T}$ alors $\mathcal{T}_B = \{X \cap B / X \in \mathcal{T}\}$ est une tribu sur B . On dispose de l'espace probabilisable $(B, \mathcal{T}_B)$. Si de plus $\mathbb{P}(B) \neq 0$ alors $\mathbb{P}_B$ définie sur $\mathcal{T}_B$ par : $\mathbb{P}_B(X) = \frac{\mathbb{P}(X)}{\mathbb{P}(B)}$ est une probabilité sur $(B, \mathcal{T}_B)$. On dispose ainsi de l'espace probabilisé : $(B, \mathcal{T}_B, \mathbb{P}_B)$
2. Si A et B sont deux événements tel que $\mathbb{P}(B) \neq 0$, alors $\mathbb{P}_B(A) = \mathbb{P}(A)$ si et seulement si A et B sont indépendants.

12.1.2.3 Formule des probabilités totales :

Definition 12.1.6

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé. On appelle système complet d'événements toute famille $(A_i)_{i \in I}$ d'événements tel que :

- (i) I est non vide au plus dénombrable.
- (ii) $\forall i \in I, \mathbb{P}(A_i) \neq 0$ et
- (iii) $(A_i)_{i \in I}$ est une partition de Ω .

Proposition 12.1.3. Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $(A_i)_{i \in I}$ un système complet d'événements de $(\Omega, \mathcal{T}, \mathbb{P})$ avec I fini. Alors, pour tout $X \in \mathcal{T}$, on a :

$$\mathbb{P}(X) = \sum_{i \in I} \mathbb{P}(A_i) \mathbb{P}_{A_i}(X).$$

Remarques. On peut faire les remarques suivantes :

1. si on considère, en général, une famille au plus dénombrable $(A_i)_{i \in I}$ d'événements tel que $\forall i \in I, \mathbb{P}(A_i) \neq 0$ et $(A_i)_{i \in I}$ est une partition de Ω alors on a : $\forall X \in \mathcal{T}, \mathbb{P}(X) = \sum_{i \in I} \mathbb{P}(A_i) \mathbb{P}_{A_i}(X)$.
2. Si A est un événement tel que $\mathbb{P}(A)(1 - \mathbb{P}(A)) \neq 0$ alors :

$$\forall X \in \Omega, \mathbb{P}(X) = \mathbb{P}(A) \mathbb{P}_A(X) + \mathbb{P}(A^c) \mathbb{P}_{A^c}(X).$$

12.1.2.4 Formule des probabilités composées :

Proposition 12.1.4. Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et A et B deux événements tel que $\mathbb{P}(A) \neq 0$ et $\mathbb{P}(B) \neq 0$ alors : $\mathbb{P}(B) \mathbb{P}_B(A) = \mathbb{P}(A) \mathbb{P}_A(B)$

Preuve. C'est une conséquence immédiate de la définition des probabilités conditionnelles : On a $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ et $\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$, donc :

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

■

Une classe est composée de 30 élèves répartis sur quatre rangées comme suit suivant les rangées et le nombre des garçons et de filles :

1. : Rangée I : 4 garçons et 4 filles.
2. : Rangée II : 3 garçons et 4 filles.
3. : Rangée III : 4 garçons et 3 filles.
4. : Rangée IV : 5 garçons et 3 filles.

On résume cette configuration par le tableau suivant :

	I	II	III	IV	
G	4	3	4	5	16
F	4	4	3	3	14
	8	7	7	8	30

On fait un tirage au sort d'un élève de cette classe.

1. Quelle est la probabilité que ce soit un garçon sachant qu'il a été tiré dans la rangée III ?
2. Sachant que l'élève tiré est une fille, quelle est la probabilité que ce soit dans la rangée III ?

On introduit les événements suivants :

R_3 : « L'élève tiré est dans la rangée 3 »

G : « L'élève tiré est un garçon » Puisqu'il y a quatre rangées et qu'elles ont la même chance que l'élève tiré provienne de chacune d'elles, on a $\mathbb{P}(R_3) = \frac{7}{30}$.

Le nombre totale des garçons dans cette classe à 30 élèves est 16, donc $\mathbb{P}(G) = \frac{16}{30} = \frac{8}{15}$.

1. $\mathbb{P}(G|R_3) = \frac{4}{7}$ car la rangée R_3 comprends 4 garçons et 3 filles.
2. D'après la formule des probabilités composées, on a :

$$\mathbb{P}(R_3|F) = \mathbb{P}(F|R_3) \frac{\mathbb{P}(R_3)}{\mathbb{P}(F)}$$

Or $\mathbb{P}(F|R_3) = \frac{3}{7}$ et $\mathbb{P}(F) = \frac{7}{15}$, il vient :

$$\mathbb{P}(R_3|F) = \frac{3}{7} \frac{7}{30} \frac{15}{7} = \frac{3}{14}$$

Remarque. On peut généraliser cette formule sous la forme suivante :

Soit $n \in \mathbb{N}, n \geq 2$ et $A_1, \dots, A_n$ des événements tel que $\mathbb{P}\left(\bigcap_{k=1}^{n-1} A_k\right) \neq 0$. Alors, on a la formule :

$$\mathbb{P}\left(\bigcap_{k=1}^n A_k\right) = \mathbb{P}(A_1) \prod_{k=2}^n \mathbb{P}\left(A_k \mid \bigcap_{j=1}^{k-1} A_j\right)$$

Preuve. Par récurrence sur n :

- Pour $n = 2$, la formule s'écrit : $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)$ avec $\mathbb{P}(A_1) \neq 0$. C'est la définition de la probabilité conditionnelles.
- Soit $n \in \mathbb{N}, n \geq 2$ tel que la propriété ci-dessus est vrai pour n . Soit $A_1, \dots, A_n, A_{n+1}$ des événements tel que $B_n = \bigcap_{k=1}^n A_k$ réalise $\mathbb{P}(B_n) \neq 0$ et notons $A = \bigcap_{k=1}^{n+1} A_k$. On a $A = B_n \cap A_{n+1}$, donc :

$$\mathbb{P}(A) = \mathbb{P}(B_n)\mathbb{P}(A_{n+1}|B_n)$$

Par hypothèse de récurrence et compte tenu que $\mathbb{P}(B_{n-1}) \neq 0$ puisque $B_n \subset B_{n-1}$ et $\mathbb{P}(B_n) \neq 0$ (ici $B_{n-1} = \bigcap_{k=1}^{n-1} A_k$), on a :

$$\mathbb{P}(B_n) = \mathbb{P}(A_1) \prod_{k=2}^n \mathbb{P}\left(A_k \mid \bigcap_{j=1}^{n-1} A_j\right)$$

Il en découle que :

$$\mathbb{P}(A) = \left[\mathbb{P}(A_1) \prod_{k=2}^n \mathbb{P}\left(A_k \mid \bigcap_{j=1}^{n-1} A_j\right) \right] \mathbb{P}(A_{n+1}|B_n)$$

Donc :

$$\mathbb{P}\left(\bigcap_{k=1}^{n+1} A_k\right) = \mathbb{P}(A_1) \prod_{k=2}^{n+1} \mathbb{P}\left(A_k \mid \bigcap_{j=1}^{k-1} A_j\right)$$

ce qui achève la démonstration par récurrence. ■

12.2 Variables aléatoires

Dans toute la suite, $(\Omega, \mathcal{T}, \mathbb{P})$ est un espace probabilisé.

12.2.1 Définitions, premières propriétés

12.2.1.1 Définitions

Definition 12.2.1

On appelle variable aléatoire réelle sur $(\Omega, \mathcal{T}, \mathbb{P})$, toute application $X : \Omega \rightarrow \mathbb{R}$ tel que : $\forall a \in \mathbb{R}, X^{-1}(]-\infty, a]) \in \mathcal{T}$.

Remarques. On peut faire les remarques suivantes :

1. Soit X une v.a.r sur $(\Omega, \mathcal{T}, \mathbb{P})$. Si $\mathcal{B}$ est l'ensemble des parties de $\mathbb{R}$, obtenues à partir des intervalles de la forme $I_a =]-\infty, a]$ et les opérations sur les ensembles : union au plus dénombrable et complémentaire par rapport à Ω (donc intersection au plus dénombrable), alors on a $\forall A \in \mathcal{B}, X^{-1}(A) \in \mathcal{T}$
2. On note $(X \in A)$ l'événement $X^{-1}(A)$, pour toute partie $A \in \mathcal{B}$.
3. On a en particulier, par exemple :
 - $(X \leq a) = \{\omega \in \Omega / X(\omega) \leq a\}$
 - $(X = a) = \{\omega \in \Omega / X(\omega) = a\}$
 - $(a \leq X \leq b) = \{\omega \in \Omega / a \leq X(\omega) \leq b\}$
 - $(X > a) = \{\omega \in \Omega / X(\omega) > a\}$

12.2.1.2 Quelques propriétés

Théorème 12.2.1. (admis) Si $(X_k)_{1 \leq k \leq m}$ une suite de v.a.r sur $(\Omega, \mathcal{T}, \mathbb{P})$ et $f : \mathbb{R}^m \rightarrow \mathbb{R}$ une application continue alors l'application $\omega \mapsto f(X_1(\omega), \dots, X_m(\omega))$ de Ω vers $\mathbb{R}$ est une v.a.r. notée $f(X_1, \dots, X_m)$.

Corollaire 12.2.1. Si $(X_1, \dots, X_m)$ est une famille de v.a.r sur $(\Omega, \mathcal{T}, \mathbb{P})$ alors $\sum_{k=1}^m X_k, \prod_{k=1}^m X_k, \sup_{1 \leq k \leq m} (X_k)$ et $\inf_{1 \leq k \leq m} (X_k)$ sont des v.a.r sur $(\Omega, \mathcal{T}, \mathbb{P})$.

Proposition 12.2.1. Si $f : \mathbb{R} \rightarrow \mathbb{R}$ est une application monotone alors pour toute v.a.r X sur $(\Omega, \mathcal{T}, \mathbb{P})$, on a $f \circ X$ est une v.a.r.

Preuve. Supposons, par exemple, que f est croissante. On sait que pour tout intervalle I de $\mathbb{R}$, on a $f^{-1}(I)$ est un intervalle de $\mathbb{R}$, en effet si $x, y \in f^{-1}(I)$ tel que $x < y$, Soit $z \in \mathbb{R}$ tel que $x < z < y$ alors par croissance de f , on a $f(x) \leq f(z) \leq f(y)$ et par suite $f(z) \in I$ car $f(x), f(y) \in I$ et I est un intervalle de $\mathbb{R}$. Il en découle que si I est un intervalle de $\mathbb{R}$, on a $(f \circ X)^{-1}(I) = X^{-1}(f^{-1}(I)) \in \mathcal{T}$, compte tenu de la remarque 2 ci-dessus juste après la définition d'une v.a.r. ■

Remarque. La proposition ci-dessus est valable si on remplace monotone par monotone par morceaux.

Proposition 12.2.2. Soit (X_n) une suite de v.a.r. sur $(\Omega, \mathcal{T}, \mathbb{P})$. Si X_n converge simplement vers X application de Ω vers $\mathbb{R}$ alors X est une v.a.r.

Preuve. Avant de donner la preuve on commence par donner deux remarques :

- Première : Si (u_n) est une suite réelle convergente de limite ℓ alors $\ell = \inf_{n \in \mathbb{N}} \sup_{k \geq n} u_k$.

En effet, d'abord $\ell \leq U_n$ pour tout $n \in \mathbb{N}$ où $U_n = \sup_{k \geq n} u_k$, car sinon, on aurait :

$$\exists N \in \mathbb{N}, \forall k \geq N, u_k < \ell$$

ce qui contredit la convergence de (u_n) vers ℓ . Donc ℓ est un minorant de $\{U_n/n \in \mathbb{N}\}$. Soit $\varepsilon > 0$. Par définition de la limite, il existe $N \in \mathbb{N}$ tel que : $\forall n \geq N, \ell - \varepsilon < u_n < \ell + \varepsilon$. en particulier : $U_N = \sup_{k \geq N} u_k \leq \ell + \varepsilon$. Ainsi $\ell = \inf\{U_n/n \in \mathbb{N}\}$.

- Deuxième : Soit A une partie non vide bornée de $\mathbb{R}$, alors pour tout $a \in \mathbb{R}$ on a :

$$\sup(A) < a \Leftrightarrow \exists \eta > 0, \forall x \in A, x < a - \eta.$$

$$\inf(A) < a \Leftrightarrow \exists x \in A, x < a$$

En effet supposons que η existe alors $\sup(A) \leq a - \eta < a$. Réciproquement si $\sup(A) < a$ alors en prenant $\eta = \frac{1}{2}(a - \sup(A))$, on a $a - \eta = \frac{1}{2}(a + \sup(A))$, donc $x < a - \eta$, pour tout $x \in A$.

Pour la seconde équivalence, supposons que $\exists x \in A$ tel que $x < a$, alors $\inf(A) \leq x < a$. Supposons que $\forall x \in A, x \geq a$ alors, on aurait $\inf(A) \geq a$, d'où l'autre sens de l'équivalence par contraposée. Soit alors

$\omega \in \Omega$, comme $X_n \rightarrow X$ simplement sur Ω , on a en particulier $X(\omega) = \lim_{n \rightarrow +\infty} X_n(\omega)$.

Donc $X(\omega) = \sup_{n \in \mathbb{N}} \inf_{k \geq n} (X_k(\omega))$. Compte tenu des remarques ci-dessus, on a : si $a \in \mathbb{R}$, alors :

$$X(\omega) < a \Leftrightarrow \exists \eta > 0, \forall n \in \mathbb{N}, \exists k \geq n, u_k < a - \eta$$

il en découle qu'il existe $\eta > 0$ tel que, si $a' = a - \eta$, alors :

$$(X < a) = \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} (X_k < a').$$

Par hypothèse les X_k sont des variables aléatoires, donc $(X < a) \in \mathcal{T}$, pour tout $a \in \mathbb{R}$, ce qui prouve le fait que X est une v.a.r. ■

12.2.2 Loi d'une variable aléatoire

Soit $\mathcal{I}(\mathbb{R})$ l'ensemble des intervalles de $\mathbb{R}$ et $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé.

Definition 12.2.2

Si X est une v.a.r. sur $(\Omega, \mathcal{T}, \mathbb{P})$, l'application $\varphi : \mathcal{I}(\mathbb{R}) \rightarrow \mathbb{R}$ tel que $\varphi(I) = \mathbb{P}(X \in I)$ pour tout $I \in \mathcal{I}(\mathbb{R})$ s'appelle la loi de la v.a.r. X .

Definition 12.2.3

Si $X = (X_1, \dots, X_m)$ est une famille de v.a.r. sur $(\Omega, \mathcal{T}, \mathbb{P})$ alors l'application : $\Phi : \mathcal{I}(\mathbb{R})^m \rightarrow \mathbb{R}$ tel que $\Phi(I_1, \dots, I_m) = \mathbb{P}(\bigcap_{1 \leq k \leq m} (X_k \in I_k))$, s'appelle la loi de la famille de v.a.r. X .

Remarque. On notera $(X_1 \in I_1, \dots, X_m \in I_m)$ l'événement : $\bigcap_{1 \leq k \leq m} (X_k \in I_k)$, ainsi :

$$\Phi(I_1, \dots, I_m) = \mathbb{P}(X_1 \in I_1, \dots, X_m \in I_m)$$

12.2.3 Fonction de répartition**Definition 12.2.4**

Soit X une variable aléatoire réelle sur un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$. L'application F_X de $\mathbb{R}$ vers $\mathbb{R}$ définie par : $F_X(t) = \mathbb{P}(X \leq t)$ pour tout $t \in \mathbb{R}$ est appelée la fonction de répartition de la v.a.r. X .

Exemple. On lance une pièce de monnaie une seule fois et on associe à cette expérience l'espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$ avec $\Omega = \{P, F\}$ et $\mathbb{P}(P) = p$ et $\mathbb{P}(F) = 1 - p$ avec p un réel tel que $p \in [0, 1]$. On considère la variable aléatoire X tel que $X(P) = 1$ et $X(F) = 0$. Alors, on a $X(\Omega) = \{0, 1\}$ et pour tout réel t , on a $F_X(t) = \mathbb{P}(X \leq t) = 0$ si $t < 0$ et $F_X(t) = 0$ si $0 \leq t < 1$ et $F_X(t) = 1$ si $t \geq 1$.

Théorème 12.2.2. L'application F_X possède les propriétés suivantes :

1. F_X est croissante sur $\mathbb{R}$
2. $\lim_{-\infty} F_X = 0$ et $\lim_{+\infty} F_X = 1$
3. F_X est continue en tout point $a \in \mathbb{R}$ à droite.

Preuve. 1) Soit $x, y \in \mathbb{R}$ tel que $x \leq y$, alors $] - \infty, x] \subset] - \infty, y]$ donc $\mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y)$, d'où $F_X(x) \leq F_X(y)$.

2) Comme F_X est croissante il suffit de prouver que $\lim_{n \rightarrow +\infty} F_X(-n) = 0$. Or c'est le cas puisque la famille d'intervalles (I_n) tel que $I_n =] - \infty, -n]$ est décroissante et $\bigcap_{n \in \mathbb{N}} I_n = \emptyset$, donc la famille $(X \in I_n)$ est une famille décroissante d'éléments de $\mathcal{T}$ et $\bigcap_{n \in \mathbb{N}} (X \in I_n) = \emptyset$. par le théorème de continuité séquentielle monotone, on a $0 = \mathbb{P}(\emptyset) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \in I_n) = \lim_{n \rightarrow +\infty} F_X(-n)$.

Pour $\lim_{+\infty} F_X = 1$, même raisonnement en remarquant que $\mathbb{R} = \bigcup_{n \in \mathbb{N}}] - \infty, n]$.

3) Soit $a \in \mathbb{R}$. Compte tenu du fait que F_X est croissante, pour démontrer que F_X est continue à droite en a , il suffit de prouver que $\lim_{n \rightarrow +\infty} F_X(a + \frac{1}{n+1}) = F_X(a)$, chose vraie car si $I_n =] - \infty, a + \frac{1}{n+1}]$ alors $] - \infty, a] = \bigcap_{n \in \mathbb{N}} I_n$, donc par le théorème de continuité séquentielle monotone, on conclut. ■

Proposition 12.2.3. Soit $t_0 \in \mathbb{R}$, alors $\mathbb{P}(X < t_0) = F_X(t_0-)$.

Preuve. Pour toute suite (x_n) strictement croissante qui converge vers t_0 , on a

$$]-\infty, t_0[= \bigcup_{n \in \mathbb{N}}]-\infty, x_n]$$

donc par la caractérisation séquentielle de la limite, et le théorème de continuité séquentielle, on a :

$$\mathbb{P}(X < t_0) = F_X(t_0-)$$

■

Corollaire 12.2.2. F_X est continue au point t_0 si et seulement si $\mathbb{P}(X = t_0) = 0$.

Conséquence : on peut à l'aide de F_X exprimer $\mathbb{P}(X \in I)$ pour tout intervalle I de $\mathbb{R}$. par exemple, on a :

1. $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$
2. $\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a-)$
3. $\mathbb{P}(a < X < b) = F_X(b-) - F_X(a)$
4. $\mathbb{P}(a \leq X < b) = F_X(b-) - F_X(a-)$

Preuve. Par exemple pour la dernière : on a : $\mathbb{R} =]-\infty, a[\cup [a, b[\cup [b, +\infty[$, union disjointe, d'où : $1 = F_X(a-) + \mathbb{P}(a \leq X < b) + 1 - F_X(b-)$, ce qui donne $\mathbb{P}(a \leq X < b) = F_X(b-) - F_X(a-)$. ■

Definition 12.2.5

Soit $X = (X_1, \dots, X_m)$ une famille de variables aléatoires réelles sur $(\Omega, \mathcal{T}, \mathbb{P})$.
L'application : $F_X : \mathbb{R}^m \rightarrow \mathbb{R}$ tel que $F_X(t_1, \dots, t_m) = \mathbb{P}(X_1 \leq t_1, \dots, X_m \leq t_m)$ s'appelle la fonction, de répartition de la famille X des v.a.r. $(X_j), 1 \leq j \leq m$.

12.2.4 Indépendance des v.a.r, conditionnement

12.2.4.1 Indépendance

Definition 12.2.6

Soit $(X_j)_{j \in J}$ une famille de variables aléatoires réelles d'un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$. On dit que les v.a.r. X_j sont indépendantes si pour toute famille $(I_j)_{j \in J}$ d'intervalles de $\mathbb{R}$, les événements $(X_j \in I_j)_{j \in J}$ sont indépendants.

Remarques. 1. Ne pas confondre avec deux à deux indépendantes.
2. La notion d'indépendance dépend de la probabilité $\mathbb{P}$ choisie.

Proposition 12.2.4. Si $X_1, \dots, X_n$ sont des variables aléatoires indépendantes et si $0 = n_0 < \dots < n_p = n$ et pour tout $j \in \{1, \dots, p\}$ on pose $Y_j = f_j(X_{n_{j-1}+1}, \dots, X_{n_j})$ et si $Y_1, \dots, Y_p$ sont des variables aléatoires alors $Y_1, \dots, Y_p$ sont indépendantes.

12.2.4.2 Conditionnement

Definition 12.2.7

Si X et Y sont deux v.a.r. sur $(\Omega, \mathcal{T}, \mathbb{P})$, on dispose des événements $(X \in A)$ et $(Y \in B)$ où $A, B \in \mathcal{B}_{\mathbb{R}}$. Si $\mathbb{P}(X \in B) \neq 0$, alors on sait que :

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

On parle de variable conditionnée : Z dont la loi est donnée par :

$$\mathbb{P}(Z \in A) = P(X \in A) / (Y \in B)$$

pour tout $A \in \mathcal{B}_{\mathbb{R}}$.

12.2.5 Variables aléatoires réelles discrètes

12.2.5.1 Définitions

Definition 12.2.8

Soit X une variable aléatoire réelle sur $(\Omega, \mathcal{T}, \mathbb{P})$. on dit que X est discrète s'il existe $\Omega' \in \mathcal{T}$ tel que $\mathbb{P}(\Omega') = 1$ et $X(\Omega')$ est au plus dénombrable.

Remarques. On a les remarques suivantes :

1. Si X est une v.a.r.d., la partie D de $\mathbb{R}$ tel que :

$$D = \{x \in X(\Omega') / \mathbb{P}(X = x) \neq 0\},$$

s'appelle l'ensemble des valeurs possibles de X . On a $D \subset D' = X(\Omega')$ et l'inclusion peut être stricte.

2. La loi de X est parfaitement déterminée par la donnée de D (ensemble des valeurs possibles) et $\mathbb{P}(X = t)$ pour tout $t \in D$.
3. Si $A \in \mathcal{I}(\mathbb{R})$ (ensemble des intervalles de $\mathbb{R}$) alors

$$(\star) \quad \mathbb{P}(X \in A) = \sum_{t \in A} \mathbb{P}(X = t)$$

Notons que la somme dans $(\star)$ a un sens car l'ensemble

$$A' = A \cap D = \{t \in A / \mathbb{P}(X = t) \neq 0\}$$

est au plus dénombrable, donc $(\star)$ s'écrit aussi :

$$(\star\star) \quad \mathbb{P}(X \in A) = \sum_{t \in A \cap D} \mathbb{P}(X = t)$$

4. On parle de variable aléatoire discrète usuelle s'il existe une bijection croissante $\varphi : J \rightarrow D$ d'un intervalle J de $\mathbb{Z}$ vers D .

Dans ce cas on peut noter $D = \{t_k / k \in J\}$ où $t_k = \varphi(k), \forall k \in J$.

Exemple. On considère $\Omega = \mathbb{R}$, on prends la tribu $\mathcal{T} = \mathcal{P}(\mathbb{R})$ et soit $\mathbb{P} : \mathcal{T} \rightarrow [0, 1]$ tel que pour tout $A \in \mathcal{T}$, on a : $\mathbb{P}(A) = 0$ si $A \cap \mathbb{Z} = \emptyset$ et si $A' = A \cap \mathbb{Z} \neq \emptyset$, alors $\mathbb{P}(A) = \sum_{k \in A'} \frac{1}{3 \cdot 2^{|k|}}$. Il est aisé de prouver que $(\Omega, \mathcal{T}, \mathbb{P})$ est un espace probabilisé.

Remarquons que l'on a :

$$\mathbb{P}(\mathbb{Z}) = \sum_{k \in \mathbb{Z}} \frac{1}{3 \cdot 2^{|k|}} = \frac{1}{3} \left(1 + 2 \sum_{n=1}^{+\infty} \frac{1}{2^n} \right) = 1$$

Soit $X : \mathbb{R} \rightarrow \mathbb{R}$ tel que $X(\omega) = \omega^2$. On remarque que $X(\Omega) = X(\mathbb{R}) = \mathbb{R}_+$ n'est pas dénombrable.

Soit $\Omega' = \{\pm \sqrt{n}/n \in \mathbb{N}\}$: il est clair que Ω' est dénombrable, et puisque $\mathbb{Z} \subset \Omega'$ et $\mathbb{P}(\mathbb{Z}) = 1$, on a $\mathbb{P}(\Omega') = 1$. Il en résulte que la v.a.r. X est discrète.

L'ensemble des valeurs possibles de X est $D = \{n^2/n \in \mathbb{N}\} = X(\mathbb{Z})$. On remarque que $D \subsetneq D'$, où $D' = X(\Omega') = \mathbb{N}$.

12.2.5.2 Variables aléatoires discrètes usuelles :

1- Loi uniforme discrète

Definition 12.2.9

On dit que la v.a.r. X sur $(\Omega, \mathcal{T}, \mathbb{P})$ suit une loi uniforme discrète et on note $X \hookrightarrow \mathcal{U}[1, n]$ si $X(\Omega) = [1, n]$ et pour tout $k \in X(\omega)$, $\mathbb{P}(X = k) = \frac{1}{n}$.

Généralement si $X(\Omega) = \{x_1, \dots, x_n\}$ avec $x_1, \dots, x_n$ des réels distincts et $\mathbb{P}(X = x_k) = \frac{1}{n}$, on dit aussi que X suit une loi uniforme discrète et on note $X \hookrightarrow \mathcal{U}(\{x_1, \dots, x_n\})$

Exemple. On lance un dé parfait à six faces numérotées de 1 à 6. Soit X la variable aléatoire réelle qui désigne le numéro de la face obtenue. Alors $\Omega = \{1, \dots, 6\}$ et $X(\Omega) = [1, 6]$ et pour tout $k \in [1, 6]$, on a $\mathbb{P}(X = k) = \frac{1}{6}$. Donc $X \hookrightarrow \mathcal{U}[1, 6]$.

Exemple. On prends $\Omega = \mathbb{R}$ et $\mathcal{T} = \mathcal{B}_{\mathbb{R}}$ la tribu borélienne. On admet qu'il existe une probabilité $\mathbb{P}$ sur l'espace probabilisable $(\Omega, \mathcal{T})$ tel que pour tout intervalle I non trivial de $\mathbb{R}$, on ait :

$$\mathbb{P}(I) = \frac{1}{\sqrt{\pi}} \int_I e^{-t^2} dt$$

Soit X l'application de $\mathbb{R}$ vers $\mathbb{R}$ tel que :

$$\forall \omega \in \mathbb{R} \quad X(\omega) = \begin{cases} 1 & \text{si } \omega < 0 \\ 2 & \text{si } \omega \geq 0 \end{cases}$$

Alors X est une variable aléatoire puisque pour tout intervalle J de $\mathbb{R}$, on a

$$X^{-1}(J) = \begin{cases}]-\infty, 0[& \text{si } 1 \in J \text{ et } 2 \notin J \\ [0, +\infty[& \text{si } 1 \notin J \text{ et } 2 \in J \\ \mathbb{R} & \text{si } 1 \in J \text{ et } 2 \in J \end{cases}$$

On a $X(\mathbb{R}) = \{1, 2\}$ et

$$\begin{cases} \mathbb{P}(X = 1) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^0 e^{-t^2} dt = \frac{1}{2} \\ \mathbb{P}(X = 2) = \frac{1}{\sqrt{\pi}} \int_0^{+\infty} e^{-t^2} dt = \frac{1}{2} \end{cases}$$

Il en découle que $X \hookrightarrow \mathcal{U}[1, 2]$.

2- Loi de Bernoulli

Definition 12.2.10

Soit p un nombre réel tel que $p \in [0, 1]$ et X une v.a.r sur $(\Omega, \mathcal{T}, \mathbb{P})$. On dit que suit la loi de Bernoulli de paramètre p et on note $X \hookrightarrow \mathcal{B}(p)$, si $X(\Omega) = \{0, 1\}$ et $\mathbb{P}(X = 1) = p$ et $\mathbb{P}(X = 0) = 1 - p$

Exemple. On lance une pièce de monnaie qui montre pile avec une probabilité $\frac{1}{3}$ et face avec la probabilité $\frac{2}{3}$. Soit X la variable aléatoire qui prend 1 si on obtient pile et 0 si on obtient face. Alors $X \hookrightarrow \mathcal{B}(\frac{1}{3})$

3-Loi binomiale

Definition 12.2.11

Soit $p \in [0, 1]$ et $n \in \mathbb{N}$ et X une v.a.r sur $(\Omega, \mathcal{T}, \mathbb{P})$. on dit que X suit la loi binomiale de paramètres n et p et on note : $X \hookrightarrow \mathcal{B}(n, p)$, si $X(\Omega) = \llbracket 0, n \rrbracket$ et pour tout $k \in \llbracket 0, n \rrbracket$, $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$, où $\binom{n}{k}$ dénote le coefficient binomial : $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Exemple. Sur 10 foyers d'un village, un seul est prêt d'acheter des livres. Un vendeur de livres fait le tour des 400 foyers de ce village de manière aléatoire. Quelles est la probabilité que 40 personnes exactement achètent des livres quand il finit le tour de tous les foyers du village? Soit X la variable aléatoire qui prends le nombre de foyers ayant acheté un livre à ce vendeur. On cherche $\mathbb{P}(X = 40)$. Il est facile de voir que $X \hookrightarrow \mathcal{B}(400, \frac{1}{10})$, $\Omega = \llbracket 1, 400 \rrbracket$, $X(\Omega) = \llbracket 0, 400 \rrbracket$ et $p = \frac{1}{10}$ et par suite :

$$\mathbb{P}(X = 40) = \binom{400}{40} \left(\frac{1}{10^{40}} \right) \left(\frac{9}{10} \right)^{360}$$

Grâce à maple, on trouve : $\mathbb{P}(X = 40) \simeq 0.066$

4- Loi géométrique $X \hookrightarrow \mathcal{G}(p)$ si $X(\Omega) = \mathbb{N}^*$ et $\mathbb{P}(X = n) = p(1 - p)^{n-1}$. On adopte la notation $q = 1 - p$, donc $\forall n \in \mathbb{N}^* \quad \mathbb{P}(X = n) = pq^{n-1}$.

Remarque. Il existe une autre définition qui adopte : $X \hookrightarrow \mathcal{G}(p)$ si $X(\Omega) = \mathbb{N}$ et pour tout $n \in \mathbb{N}$, on a : $\mathbb{P}(X = n) = p(1 - p)^n$

5- Loi de Poisson $X \hookrightarrow \mathcal{P}(\lambda)$ si $X(\Omega) = \mathbb{N}$ et $\mathbb{P}(X = n) = e^{-\lambda} \frac{\lambda^n}{n!}$.

12.2.6 Variables aléatoires continues à densité

12.2.6.1 Définitions, propriétés

La notion de densité de probabilité possède une définition plus générale que celle qu'on donnera ci-dessous. Les limitations imposées par le programme nous pousse à nous restreindre aux fonctions $f : \mathbb{R} \rightarrow \mathbb{R}$ qui sont des densités de probabilités dans le cas particulier où f est continue par morceaux sur $\mathbb{R}$.

Definition 12.2.12

Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une application continue par morceaux sur $\mathbb{R}$. On dira que f est une densité de probabilité si :

- L'ensemble A des points de discontinuité de f est une partie finie de $\mathbb{R}$.
- f est positive sur $\mathbb{R}$.
- f est intégrable sur $\mathbb{R}$ et $\int_{-\infty}^{+\infty} f(t)dt = 1$.

Si $f : \mathbb{R} \rightarrow \mathbb{R}$ est une application positive continue par morceaux sur $\mathbb{R}$ d'ensemble de points de discontinuité A fini et si f est intégrable et $\alpha = \int_{-\infty}^{+\infty} f(t)dt$ réalise $\alpha > 0$ alors $\tilde{f} = \frac{1}{\alpha}f$ est une densité de probabilité.

Definition 12.2.13

Soit X une variable aléatoire réelle sur un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$. On dit que X est continue à densité si sa fonction de répartition F_X vérifie ce qui suit :

- (i) F_X est continue sur $\mathbb{R}$.
- (ii) Il existe une partie finie A de $\mathbb{R}$ tel que F_X est de classe C^1 sur $\mathbb{R} \setminus A$ et la fonction f_X définie par $f_X(t) = \begin{cases} 0 & \text{si } t \in A \\ F'_X(t) & \text{si } t \in \mathbb{R} \setminus A \end{cases}$ est une densité de probabilité.

Si c'est le cas f_X est appelé densité de la variable aléatoire réelle X .

Proposition 12.2.5. Si X est une variable aléatoire réelle continue à densité sur $(\Omega, \mathcal{T}, \mathbb{P})$ alors il existe une partie finie A de $\mathbb{R}$ tel que :

- (i) $f_X \geq 0$ sur $\mathbb{R}$.
- (ii) f_X est continue sur $\mathbb{R} \setminus A$.
- (iii) $\int_{-\infty}^{+\infty} f_X(t)dt = 1$.

Remarque. Si une application $f : \mathbb{R} \rightarrow \mathbb{R}$ ne vérifie pas l'une des propriétés ci-dessus elle ne peut être la densité de probabilité pour une variable aléatoire continue à densité.

Proposition 12.2.6. Soit X une variable aléatoire réelle continue à densité. Si F_X et f_X sont respectivement la fonction de répartition et la densité de X alors on a :

$$\forall x \in \mathbb{R}, F_X(x) = \int_{-\infty}^x f_X(t)dt$$

Preuve. Comme X est continue à densité, il existe une partie finie A de $\mathbb{R}$ tel que F_X est de classe C^1 sur tout intervalle contenu dans $\mathbb{R} \setminus A$. Soit $x \in \mathbb{R}$. Trois cas sont possibles :

- Si $A \cap]-\infty, x[= \emptyset$ alors $F'_X(t) = f_X(t)$ pour tout $t \in \mathbb{R}$, en particulier, pour tout $x \in \mathbb{R}$, on a : $\int_{-\infty}^x f_X(t)dt = \int_{-\infty}^x F'_X(t)dt = [F_X(t)]_{-\infty}^x = F_X(x)$ car $\lim_{t \rightarrow -\infty} F_X(t) = 0$.

- Si $A \cap]-\infty, x[= \{a\}$ avec $a \in \mathbb{R}$, la formule ci-dessus reste valable si $x < a$. Si $x \geq a$ alors :

$$\int_{-\infty}^x f_x(t)dt = \int_{-\infty}^a f_X(t)dt + \int_a^x f_X(t)dt = F_X(a) + F_X(x) - F_X(a) = F_X(x)$$

car f_x est de classe C^1 sur chacun des intervalles $] -\infty, a[$ et $]a, x[$ et sa dérivée F_X est continue sur chacun des intervalles $] -\infty, a[$ et $[a, x]$.

- Si $A \cap]-\infty, x[= \{a_1, \dots, a_m\}$ avec $m \in \mathbb{N}, m \geq 2$ et $a_1 < \dots < a_m$ alors :

$$\int_{-\infty}^x f_X(t)dt = \int_{-\infty}^{a_1} f_X(t)dt + \sum_{k=1}^{m-1} \int_{a_k}^{a_{k+1}} f_X(t)dt + \int_{a_m}^x f_X(t)dt$$

Puisque f_X est de classe C_1 sur chacun des intervalles $] -\infty, a_1[,]a_k, a_{k+1}[(1 \leq k \leq m-1),]a_m, x[$ et que sa dérivée F_X est continue sur $\mathbb{R}$, on a alors :

$$\int_{-\infty}^x f_X(t)dt = F_X(a_1) + \sum_{k=1}^{m-1} (F_X(a_{k+1}) - F_X(a_k)) + F_X(x) - F_X(a_m) = F_X(x)$$

■

Proposition 12.2.7. Soit X une variable aléatoire réelle continue à densité, alors pour tout intervalle I de $\mathbb{R}$ de bornes a et b , on a :

$$\mathbb{P}(X \in I) = \int_a^b f_X(t)dt = F_X(b) - F_X(a).$$

En particulier, pour tout $t \in \mathbb{R}$, on a $\mathbb{P}(X = t) = 0$.

Preuve. On a :

$$\mathbb{P}(X \leq b) = \mathbb{P}(X \leq a) + \mathbb{P}(a < X \leq b) \mathbb{P}(a < X \leq b)$$

Donc :

$$p(X \in I) = \mathbb{P}(a < X \leq b) = F(b) - F(a)$$

D'après la proposition 12.2.6, on a alors :

$$\mathbb{P}(X \in I) = \int_{-\infty}^b f_X(t)dt - \int_{-\infty}^a f_X(t)dt$$

donc

$$\mathbb{P}(X \in I) = \int_a^b f_X(t)dt$$

■

12.2.6.2 Variables continues à densité usuelles

1) Loi uniforme continue

Definition 12.2.14

Soit X une variable continue à densité sur $(\Omega, \mathcal{T}, \mathbb{P})$ et $a, b \in \mathbb{R}$ tel que $a < b$. On dit que f suit la loi uniforme continue sur $[a, b]$ si la densité f_X de X est définie par :

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{sinon} \end{cases}$$

Notation : $X \hookrightarrow \mathcal{U}[a, b]$.

Proposition 12.2.8. Soit X une variable aléatoire sur $(\Omega, \mathcal{T}, \mathbb{P})$. Si $X \hookrightarrow \mathcal{U}[a, b]$ alors la fonction de répar-

tion F_X de X est définie par :

$$F_X(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } b \leq x \end{cases}$$

Preuve. Soit $x \in \mathbb{R}$ alors $F_X(x) = \int_{-\infty}^x f_X(t)dt$.

- Si $x < a$, comme f_X est nulle sur $] -\infty, x]$, on a $F_X(x) = 0$.
- Si $a \leq x \leq b$ alors :

$$F_X(x) = \int_{-\infty}^a 0dt + \int_a^x \frac{1}{b-a}dt = \frac{x-a}{b-a}$$

- Si $x > b$ alors :

$$F_X(x) = \int_{-\infty}^b f_X(t)dt + \int_b^{+\infty} 0dt = F_X(b) = 1$$

■

2) Loi exponentielle de paramètre λ

Definition 12.2.15

Soit X une variable aléatoire réelle continue à densité sur $(\Omega, \mathcal{T}, \mathbb{P})$ et Soit $\lambda \in \mathbb{R}_+^*$. On dit que X suit la loi exponentielle de paramètre λ si la fonction de densité de X est définie par :

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

Notation : $X \hookrightarrow \mathcal{E}(\lambda)$.

Proposition 12.2.9. Soit X une variable aléatoire réelle continue à densité sur $(\Omega, \mathcal{T}, \mathbb{P})$ et $\lambda \in \mathbb{R}_+^*$. Si $X \hookrightarrow \mathcal{E}(\lambda)$ alors la fonction de répartition F_X de X est définie par :

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

Preuve. Soit $x \in \mathbb{R}$.

- Si $x < 0$, comme f_X est nulle sur $] -\infty, x]$, on a $F_X(x) = 0$.
- Si $x > 0$, alors :

$$F_X(x) = \int_0^x \lambda e^{-\lambda t} dt = [-e^{-\lambda t}]_0^x = 1 - \lambda e^{-\lambda x}$$

■

3) Loi normale de Gauss

Definition 12.2.16

Soit X une variable aléatoire réelle continue à densité sur $(\Omega, \mathcal{T}, \mathbb{P})$ et soit $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$. On dit que X suit la loi normale de Gauss de paramètres μ et σ^2 si la densité de X est définie par :

$$\forall x \in \mathbb{R}, f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Notation : $X \hookrightarrow \mathcal{N}(\mu, \sigma^2)$.

4) Loi gamma

Definition 12.2.17

Soit X une variable aléatoire réelle continue à densité sur $(\Omega, \mathcal{T}, \mathbb{P})$ et soit $(a, \lambda) \in \mathbb{R}_+^2$. On dit que X suit la loi Gamma de paramètres a et λ si la densité f_X de X est définie par :

$$f_X(t) = \begin{cases} \frac{\lambda^a}{\Gamma(a)} t^{a-1} e^{-\lambda t} & \text{si } t \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Notation : $X \hookrightarrow \Gamma(a, \lambda)$.

Remarque. Pour tout réel strictement positif λ , on a $\Gamma(1, \lambda) = \mathcal{E}(\lambda)$.

12.2.7 Somme de deux variables aléatoires indépendantes

$(\Omega, \mathcal{T}, \mathbb{P})$ est un espace probabilisé et toutes les v.a.r. sont considérées sur cet espace.

12.2.7.1 Deux variables discrètes

Proposition 12.2.10. Si X et Y sont deux v.a.r. indépendantes discrètes tel que $X(\Omega) = D_1$ et $Y(\Omega) = D_2$ alors $X + Y$ est une v.a.r. discrète et $(X + Y)(\Omega) = D \subset D_1 + D_2$ et pour tout $s \in D$, on a :

$$\mathbb{P}(X = s) = \sum_{u \in D_1} \mathbb{P}(X = u) \mathbb{P}(Y = s - u) = \sum_{v \in D_2} \mathbb{P}(X = s - v) \mathbb{P}(Y = v)$$

Preuve. On peut toujours se ramener aux cas où $X(\Omega)$ et $Y(\Omega)$ sont au plus dénombrables. Pour tout $\omega \in \Omega$, on a $(X + Y)(\omega) = X(\omega) + Y(\omega) \in X(\Omega) + Y(\Omega)$, donc si $D = (X + Y)(\Omega)$ alors $D \subset D_1 + D_2$ où $D_1 = X(\Omega)$ et $D_2 = Y(\Omega)$. On peut sans nuire à toute généralité^a, supposer que $X(\omega) \neq 0$ pour tout $\omega \in \Omega$, de manière à avoir $(X = k)_{k \in D_1}$ est un système complet d'événement et appliquer la formule des probabilités totales : pour $s \in X + Y(\Omega)$, on a :

$$\mathbb{P}(X + Y = s) = \sum_{k \in D_1} \mathbb{P}(X = k) \mathbb{P}(X + Y = s | X = k)$$

Or :

$$\mathbb{P}(X + Y = s | X = k) = \frac{\mathbb{P}(X + Y = s, X = k)}{\mathbb{P}(X = k)} = \frac{\mathbb{P}(X = k, Y = s - k)}{\mathbb{P}(X = k)}$$

Par indépendance, il vient :

$$\mathbb{P}(X + Y = s | X = k) = \frac{\mathbb{P}(X = k) \mathbb{P}(Y = s - k)}{\mathbb{P}(X = k)} = \mathbb{P}(Y = s - k)$$

d'où :

$$\mathbb{P}(X + Y = s) = \sum_{k \in D_1} \mathbb{P}(X = k) \mathbb{P}(Y = s - k)$$

pour tout $s \in X + Y(\Omega)$. ■

^a. Pour ce faire on prends $\Omega_1 = X^{-1}(D'_1)$ où

$$D'_1 = \{x \in X(\Omega) / \mathbb{P}(X = x) \neq 0\}$$

c'est -à-dire l'ensemble des valeurs possibles de X et remplacer Ω par Ω_1 . On peut prouver que $\mathbb{P}(\Omega_1) = 1$, donc rien ne change au niveau du raisonnement.

12.2.7.2 Une variable continue et une autre discrète

Proposition 12.2.11. Soient X et Y deux variables aléatoires réelles indépendantes sur $(\Omega, \mathcal{T}, \mathbb{P})$ tel que X est discrète d'ensemble de valeurs possibles et Y continue de densité f_Y , alors $X + Y$ est une v.a.r.

continue et sa densité f_{X+Y} est définie par :

$$\forall s \in \mathbb{R}, f_{X+Y}(s) = \sum_{u \in D} P(X = u) f_Y(s - u)$$

sous réserve que f_{X+Y} est bien définie sur $\mathbb{R}$ et est continue sur $\mathbb{R} \setminus A$ où A est une partie finie de $\mathbb{R}$.

12.2.7.3 Deux variables continues

Proposition 12.2.12. Soient X et Y deux variables aléatoires réelles indépendantes sur $(\Omega, \mathcal{T}, \mathbb{P})$ tel que X et Y sont continues de densités respectives f_X et f_Y , alors $X + Y$ est une v.a.r. continue et sa densité f_{X+Y} est définie par :

$$\forall s \in \mathbb{R}, f_{X+Y}(s) = \int_{-\infty}^{+\infty} f_X(u) f_Y(s - u) du = \int_{-\infty}^{+\infty} f_X(s - v) f_Y(v) dv$$

sous réserve que f_{X+Y} ci-dessus est continue sur $\mathbb{R} \setminus A$ où A est une partie finie de $\mathbb{R}$.

Remarque. On appelle produit de convolution de deux fonctions continues par morceaux f et g de $\mathbb{R}$ vers $\mathbb{R}$, la fonction notée $f * g$ définie par :

$$f * g(x) = \int_{\mathbb{R}} f(t) g(x - t) dt = \int_{\mathbb{R}} f(x - t) g(t) dt$$

On remarque donc que $f_{X+Y} = f_X * f_Y$.

12.2.8 Moments, espérance, variance, écart type d'une v.a.r.

12.2.8.1 Espérance mathématique : Définition

Definition 12.2.18

Soit X une v.a.r sur $(\Omega, \mathcal{T}, \mathbb{P})$.

Si X est discrète d'ensemble de valeurs possibles D et si la famille $(x\mathbb{P}(X = x))_{x \in D}$ est sommable, sa somme :

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

est appelée l'espérance mathématique de X .

Si X est continue de densité f_X et si l'application $x \mapsto xf_X(x)$ est intégrable, son intégrale :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf_X(x) dx$$

est appelée l'espérance mathématique de X .

Remarques. On peut faire les remarques suivantes :

- la sommabilité dans le cas discret permet d'avoir la somme indépendamment de la permutation de ses termes. en particulier si $\varphi : \mathbb{N}^* \rightarrow D$ est une bijection et si on pose $x_n = \varphi(n)$ et $p_n = \mathbb{P}(X = x_n)$ alors

$$E(X) = \sum_{n=1}^{+\infty} x_n p_n.$$

- On parle souvent de variable aléatoire X discrète représentée par la famille (x_i, p_i) , ce qui signifie que $D = \{x_i\}_{i \in I}$ est l'ensemble des valeurs possibles et que pour tout $i \in I$, $p_i = \mathbb{P}(X = x_i)$. Dans ce cas, et sous réserve de sommabilité, l'espérance mathématique de X est :

$$\mathbb{E}(X) = \sum_{i \in I} p_i x_i$$

- L'espérance mathématique de la v.a.r. X peut se comprendre comme une moyenne de points pondérés :

$$(x_i, p_i), \text{ puisque } \mathbb{E}(X) = \frac{\sum_{i \in I} p_i x_i}{\sum_{i \in I} p_i}. \text{ N'oublions pas que } \sum_{i \in I} p_i = 1$$

12.2.8.2 Espérance des variables usuelles

Voir le tableau récapitulatif des variables aléatoires réelles usuelles discrètes et continues à densité. Toutes les 9 v.a.r. admettent une espérance.

12.2.8.3 Propriété de transfert

Note : On admet que si X est une variable aléatoire sur $(\Omega, \mathcal{T}, \mathbb{P})$ est $g : \mathbb{R} \rightarrow \mathbb{R}$ une application continue par morceaux, alors $g \circ X$ est une variable aléatoire sur $(\Omega, \mathcal{T}, \mathbb{P})$.

Théorème 12.2.3. Soit X un v.a.r. sur $(\Omega, \mathcal{T}, \mathbb{P})$, $g : \mathbb{R} \rightarrow \mathbb{R}$ une application et $Y = g \circ X$.

• Si X est discrète, représentée par la famille $(x_k, p_k)_{k \in I}$ avec I un ensemble au plus dénombrable, et Y est une v.a.r. alors Y admet une espérance si et seulement si la famille $(p_k g(x_k))_{k \in I}$ est sommable, auquel cas, on a : $\mathbb{E}(Y) = \sum_{k \in I} p_k g(x_k)$.

• Si X est continue de densité f_X et g est continue par morceaux, alors, Y admet une espérance si et seulement si l'application $t \mapsto g(t)f_X(t)$ est intégrable sur $\mathbb{R}$, auquel cas, on a : $\mathbb{E}(Y) = \int_{-\infty}^{+\infty} g(t)f_X(t)dt$.

12.2.8.4 Transfert pour deux variables

On considère deux variables aléatoires réelles discrètes X, Y sur $(\Omega, \mathcal{T}, \mathbb{P})$ de loi conjointe la famille $((x_i, y_j), p_{ij})_{(i,j) \in I}$. Soit $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ une application tel que $Z = g \circ (X, Y)$ est une v.a.r. sur $(\Omega, \mathcal{T}, \mathbb{P})$. Alors Z admet une espérance si et seulement si la famille $(p_{ij}g(x_i, y_j))_{(i,j) \in I}$ est sommable, auquel cas, on a

$$\mathbb{E}(Z) = \sum_{(i,j) \in I} p_{ij}g(x_i, y_j)$$

12.2.8.5 Linéarité de l'espérance mathématique

Soient X et Y sont deux variables aléatoires tel que chacune d'elle est soit discrète soit continue à densité et $\lambda \in \mathbb{R}$ et supposons que $X + \lambda Y$ est discrète ou continue à densité, alors si X et Y admettent des espérances mathématiques $\mathbb{E}(X)$ et $\mathbb{E}(Y)$, alors $X + \lambda Y$ en admet une et on a $\mathbb{E}(X + \lambda Y) = \mathbb{E}(X) + \lambda \mathbb{E}(Y)$. Ce résultat sera admis et peut être utilisé mais on va énoncé et démontrer ce résultat dans le cas particulier où X et Y sont toutes les deux discrètes :

Proposition 12.2.13. Soient $\lambda \in \mathbb{R}$ et X et Y deux variables aléatoires discrètes sur $(\Omega, \mathcal{T}, \mathbb{P})$. Si X et Y admettent des espérances mathématiques $\mathbb{E}(X)$ et $\mathbb{E}(Y)$ alors $X + \lambda Y$ admet une espérance mathématique et on a

$$\mathbb{E}(X + \lambda Y) = \mathbb{E}(X) + \lambda \mathbb{E}(Y)$$

Preuve. Soient X et Y deux variables aléatoires discrètes sur $(\Omega, \mathcal{T}, \mathbb{P})$ et notons $I = X(\Omega)$ et $J = Y(\Omega)$. Sans nuire à la généralité on peut supposer que X et Y vérifient la condition :

$$\forall (x, y) \in I \times J, \quad \mathbb{P}(X = x) \neq 0 \quad \text{et} \quad \mathbb{P}(Y = y) \neq 0.$$

On suppose que X et Y admettent des espérances mathématiques $\mathbb{E}(X)$ et $\mathbb{E}(Y)$.

(★) Soit $\lambda \in \mathbb{R}$. En appliquant le théorème de transfert, avec l'application :

$$g : \mathbb{R} \rightarrow \mathbb{R}; t \mapsto \lambda t,$$

et compte tenu du fait que la famille $(\lambda x \mathbb{P}(X = x))_{x \in \Omega(X)}$ est sommable, on a $\mathbb{E}(\lambda X)$ existe et $\mathbb{E}(\lambda X) = \lambda \mathbb{E}(X)$.

(★) On a $K = (X + Y)(\Omega) = I + J$. Démontrons que la famille $(|k| \mathbb{P}(X + Y = k))_{k \in K}$ est sommable :

• Pour tout $i \in I$, fixé, la famille $(|i| \mathbb{P}(X = i, Y = j))_{j \in J}$ est sommable de somme :

$$S_{1,i} = |i| \mathbb{P}(X = i).$$

• Pour tout $(i, j) \in I \times J$, on a $\mathbb{P}(X = i, Y = j) \leq \mathbb{P}(Y = j)$ et la famille $(|j| \mathbb{P}(Y = j))_{j \in J}$ est sommable

de somme $\mathbb{E}(|Y|)$, donc la famille $(|j|\mathbb{P}(X = i, Y = j))_{j \in J}$ est sommable de somme :

$$S_{2,i} = \sum_{j \in J} |j| \mathbb{P}(X = i, Y = j)$$

• Il en découle que, pour tout $i \in I$, la famille $((|i| + |j|)\mathbb{P}(X = i, Y = j))_{j \in J}$ est sommable de somme

$$S_i = S_{1,i} + S_{2,i} = |i| \mathbb{P}(X = i) + \sum_{j \in J} |j| \mathbb{P}(X = i, Y = j)$$

• La famille $(S_i)_{i \in I}$ est sommable car :
 • D'une part la famille $(|i| \mathbb{P}(X = i))_{i \in I}$ est sommable de somme $\mathbb{E}(|X|)$ qui existe par hypothèse puisque $\mathbb{E}(X)$ existe.
 • D'autre part, on a $\mathbb{E}(|Y|) = \sum_{j \in J} |j| \mathbb{P}(Y = j)$, donc par la formule des probabilités totales, compte tenu du fait que $(X = i)_{i \in I}$ est un système fondamental d'événements de $(\Omega, \mathcal{T}, \mathbb{P})$, on a

$$\mathbb{E}(|Y|) = \sum_{j \in J} |j| \sum_{i \in I} \mathbb{P}(X = i) \mathbb{P}(Y = j / X = i),$$

donc

$$\mathbb{E}(|Y|) = \sum_{(i,j) \in I \times J} |j| \mathbb{P}(X = i, Y = j).$$

Il en découle que la famille $(|j| \mathbb{P}(X = i, Y = j))_{(i,j) \in I \times J}$ est sommable et par suite la famille $(S_{2,i})_{i \in I}$ est sommable. Donc $(S_i)_{i \in I}$ est sommable et cela termine la preuve du fait que la famille $((|i| + |j|)\mathbb{P}(X = i, Y = j))_{(i,j) \in I \times J}$ est sommable.

• Il découle de ce qui précède que $\mathbb{E}(X + Y)$ existe et de plus, en notant $K = I + J = (X + Y)(\Omega) = X(\Omega) + Y(\Omega)$, on a :

$$\begin{aligned} \mathbb{E}(X + Y) &= \sum_{s \in K} s \mathbb{P}((X + Y = s)) = \sum_{s \in K} s \sum_{i \in I} \mathbb{P}(X = i) \mathbb{P}(X + Y = s / X = i) \\ &= \sum_{s \in K} \sum_{i \in I} s \mathbb{P}(X = i, Y = s - i) \\ &= \sum_{j \in J} \sum_{i \in I} (i + j) \mathbb{P}(X = i, Y = j) \quad (j = s - i). \\ &= \sum_{i \in I} \sum_{j \in J} i \mathbb{P}(X = i, Y = j) + \sum_{j \in J} \sum_{i \in I} j \mathbb{P}(X = i, Y = j) \\ &= \sum_{i \in I} i \sum_{j \in J} \mathbb{P}(X = i, Y = j) + \sum_{j \in J} j \sum_{i \in I} \mathbb{P}(X = i, Y = j) \\ &= \sum_{i \in I} i \mathbb{P}(X = i) + \sum_{j \in J} j \mathbb{P}(Y = j) = \mathbb{E}(X) + \mathbb{E}(Y). \end{aligned}$$

■

12.2.8.6 Autres propriétés de l'espérance mathématique

Proposition 12.2.14. Soient X, Y deux v.a.r. sur $(\Omega, \mathcal{T}, \mathbb{P})$ tel que $|X| \leq Y$. Si Y admet une espérance il en est de même pour X .

Preuve. Résultat admis. ■

Proposition 12.2.15. Soient X, Y deux v.a.r. sur $(\Omega, \mathcal{T}, \mathbb{P})$ et $\lambda \in \mathbb{R}$. Alors :

- (1) Si X admet une espérance mathématique et $X \geq 0$ alors $\mathbb{E}(X) \geq 0$.
- (2) Si X et Y admettent des espérances mathématiques $\mathbb{E}(X)$ et $\mathbb{E}(Y)$ et si $X \geq Y$ alors $\mathbb{E}(X) \geq \mathbb{E}(Y)$.
- (3) X admet une espérance mathématique si et seulement si $|X|$ admet une espérance mathématique et on a alors : $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$

12.2.8.7 Moments

Definition 12.2.19

Soit X une variable aléatoire réelle et m un entier naturel non nul. Si l'espérance de la variable aléatoire X^m existe, le nombre réel $\mathbb{E}(X^m)$ s'appelle le moment d'ordre m de X .

Proposition 12.2.16. Soit X une v.a.r. sur $(\Omega, \mathcal{T}, \mathbb{P})$ et $m, n \in \mathbb{N}^*$ tel que $m \leq n$. Si X admet un moment d'ordre n alors X admet un moment d'ordre m .

Preuve. On part de l'inégalité : $t^m \leq t^n + 1$ valable pour tout réel $t \geq 0$. Dans le cas discret et X représentée par (x_k, p_k) on applique l'inégalité à $t_k = |x_k|$ et on obtient :

$$p_k |x_k|^m \leq p_k |x_k|^n + p_k$$

et comme $\mathbb{E}(X^n)$ existe par hypothèse et que la famille (p_k) est sommable, on déduit que la famille $(x_k^m p_k)$ est sommable.

Dans le cas continu à densité on applique l'inégalité à $|f_X(t)|$ pour tout $t \in \mathbb{R}$ et du fait que $t \mapsto t^n f_X(t)$ et de $t \mapsto f_X(t)$ sont intégrables, on conclut. ■

Definition 12.2.20

Soit X une variable aléatoire réelle sur $(\Omega, \mathcal{T}, \mathbb{P})$ admettant un moment d'ordre 2, et soit $\mu = \mathbb{E}(X)$ alors la v.a.r. $(X - \mu)^2$ admet une espérance positive ou nulle, le nombre réel positif :

$$\mathbb{V}(X) = \mathbb{E}((X - \mu)^2)$$

s'appelle la variance de X . On a $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

Le nombre réel $\sigma(X) = \sqrt{\mathbb{V}(X)}$ s'appelle l'écart type de X .

Proposition 12.2.17. Soit X une variable aléatoire sur $(\Omega, \mathcal{T}, \mathbb{P})$ admettant un moment d'ordre 2 alors $\mathbb{V}(X) \geq 0$. De plus, on a :

- (1) Si X est continue à densité alors $\mathbb{V}(X) > 0$.
- (2) Si X est discrète alors, $\mathbb{V}(X) = 0$ si et seulement si $\exists \Omega_1 \in \mathcal{T}$ tel que $\mathbb{P}(\Omega_1) = 1$ et X est constante sur Ω_1 .

Preuve. On a $\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) \geq 0$.

1. Si X est continue de densité f alors $\mathbb{V}(X) = \int_{-\infty}^{+\infty} (t - m)^2 f(t) dt$. or on sait qu'il existe une partie A finie de $\mathbb{R}$ tel que f est continue sur $\mathbb{R} \setminus A$. On a $\int_{-\infty}^{+\infty} (t - m)^2 f(t) dt = \int_{\mathbb{R} \setminus A} (t - m)^2 f(t) dt$ et $\mathbb{R} \setminus A = \bigcup_{j=1}^p I_j$ où les I_j sont des intervalles deux à deux disjoints et f est continue sur chaque I_j . Alors si $\mathbb{V}(X) = 0$ on aurait f nulle sur chaque I_j , donc sur $\mathbb{R} \setminus A$ donc sur $\mathbb{R}$, ce qui contredit $\int_{-\infty}^{+\infty} f(t) dt = 1$.

2. Soit X une variable aléatoire discrète sur $(\Omega, \mathcal{T}, \mathbb{P})$.

• Si X admet une variance et on note $\mu = \mathbb{E}(X)$ et si $\mathbb{V}(X) = 0$ alors $\sum_{x \in X(\Omega)} (x - \mu)^2 \mathbb{P}(X = x) = 0$. Comme il

s'agit d'une somme à termes positifs, on a $(x - \mu)^2 \mathbb{P}(X = x) = 0$, pour tout $x \in X(\Omega)$. Il en découle que :

$$\forall x \in X(\Omega) \setminus \{\mu\}, \mathbb{P}(X = x) = 0$$

Posons

$$\Omega_1 = (X = \mu) = X^{-1}(\{\mu\})$$

On a $\mathbb{P}(\Omega_1) = 1$ car

$$\begin{aligned} 1 &= \mathbb{P}(\Omega) = \mathbb{P}(X \in X(\Omega)) \\ &= \sum_{x \in X(\Omega)} \mathbb{P}(X = x) \\ &= \mathbb{P}(X = \mu) \\ &= \mathbb{P}(\Omega_1) \end{aligned}$$

Par ailleurs et par définition de Ω_1 , on a $X(\Omega_1) = \{\mu\}$, donc X est constante sur Ω_1 .

• Réciproquement, supposons qu'il existe $\Omega_1 \in \mathcal{T}$ tel que $\mathbb{P}(\Omega_1) = 1$ et X est constante sur Ω_1 . Posons alors $X(\omega) = m$, pour tout $\omega \in \Omega_1$. Soit $x \in \mathbb{R}$ tel que $x \neq m$. Montrons que $\mathbb{P}(X = x) = 0$. On a $(X = x) \cap (X = m) = \emptyset$ car $\{m\} \cap \{x\} = \emptyset$, donc $X^{-1}(\{x\}) \cap X^{-1}(\{m\}) = \emptyset$ comme $\Omega_1 \subset (X = m)$, on a $\Omega_1 \cap (X = x) = \emptyset$ et comme $\mathbb{P}(\Omega_1) = 1$, on a : $\mathbb{P}(X = x) = 0$.

On a :

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} x \mathbb{P}(X = x) = m \mathbb{P}(X = m) + \sum_{\substack{x \in X(\Omega) \\ x \neq m}} x \mathbb{P}(X = x) = m$$

puisque $\mathbb{P}(X = m) = \mathbb{P}(\Omega_1) = 1$.

Il en découle que la variance de X est :

$$\mathbb{V}(X) = \sum_{x \in X(\Omega)} (x - m)^2 \mathbb{P}(X = x) = (m - m)^2 \mathbb{P}(X = m) = 0$$

■

Proposition 12.2.18. Soit X une v.a.r. admettant un moment d'ordre 2. Alors, on a :

- (1) Pour tout $\lambda \in \mathbb{R}$, la v.a.r. λX admet une variance et $\mathbb{V}(\lambda X) = \lambda^2 \mathbb{V}(X)$, en particulier $\sigma(\lambda X) = |\lambda| \sigma(X)$.
- (2) Pour tout $\lambda \in \mathbb{R}$, la v.a.r. $(X + \lambda)$ admet une variance et $\mathbb{V}(X + \lambda) = \mathbb{V}(X)$.

12.2.8.8 Covariance

Proposition 12.2.19. Soit X et Y deux v.a.r. sur $(\Omega, \mathcal{T}, \mathbb{P})$ admettant chacune un moment d'ordre 2.. Alors XY admet une variance et on a l'inégalité dite de Cauchy-Schwarz :

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

Preuve. Notons que $XY \leq \frac{1}{2}(X^2 + Y^2)$ permet de dire que $\mathbb{E}(XY)$ existe. Par ailleurs pour tout $t \in \mathbb{R}$ on a $\mathbb{E}((tX + Y)^2) \geq 0$, donc l'application polynômiale $t \mapsto (\mathbb{E}(X^2)t^2 + 2t\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(Y^2))$ est positive, donc son discriminant réduit $\Delta' = (\mathbb{E}(X)\mathbb{E}(Y))^2 - \mathbb{E}(X^2)\mathbb{E}(Y^2)$ vérifie $\Delta' \leq 0$, ce qui donne l'inégalité de Cauchy-Schwarz. ■

Remarque. L'inégalité de Cauchy-Schwarz s'écrit aussi :

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

Proposition 12.2.20. Si X et Y sont deux v.a.r. indépendante sur $(\Omega, \mathcal{T}, \mathbb{P})$ admettant des moments d'ordre 2, alors XY admet une espérance et $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Preuve. Soient X et Y deux variables aléatoires qui admettent chacune un moment d'ordre 2 et notons $I = X(\Omega)$, $J = Y(\Omega)$ et $K = (XY)(\Omega)$. D'après la proposition 12.2.19, la variable aléatoire XY admet une espérance mathématique et on a

$$\mathbb{E}(XY) = \sum_{k \in K} k \mathbb{P}(XY = k).$$

Pour tout $k \in K$, on pose $U_k = \{(i, j) \in I \times J / ij = k\}$, alors $(U_k)_{k \in K}$ est une partition de $I \times J$ et on a :

$$\forall k \in K, \quad (XY = k) = \bigcup_{(i,j) \in U_k} (X = i, Y = j)$$

et il s'agit d'une réunion disjointe, donc

$$\mathbb{E}(XY) = \sum_{k \in K} \sum_{(i,j) \in U_k} k \mathbb{P}(X = i, Y = j) = \sum_{k \in K} \sum_{(i,j) \in U_k} ij \mathbb{P}(X = i, Y = j),$$

et par indépendance, il vient :

$$\mathbb{E}(XY) = \sum_{k \in K} \sum_{(i,j) \in U_k} ij \mathbb{P}(X = i) \mathbb{P}(Y = j),$$

et comme $(U_k)_{k \in K}$ est une partition de $I \times J$, le théorème de sommation par paquets permet d'écrire

$$\mathbb{E}(XY) = \sum_{(i,j) \in I \times J} ij \mathbb{P}(X = i) \mathbb{P}(Y = j)$$

et en utilisant les produits des séries absolument convergentes et qu'une famille sommable peut se traduire en série absolument convergente il vient :

$$\mathbb{E}(X) = \left(\sum_{i \in I} i \mathbb{P}(X = i) \right) \times \left(\sum_{j \in J} j \mathbb{P}(Y = j) \right) = \mathbb{E}(X) \mathbb{E}(Y)$$

■

Proposition 12.2.21. Si X et Y sont deux v.a.r indépendantes sur $(\Omega, \mathcal{T}, \mathbb{P})$ admettant des moments respectifs d'ordre 2 alors $X + Y$ admet une variance et :

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

Definition 12.2.21

Soient X et Y deux variables aléatoires réelles sur $(\Omega, \mathcal{T}, \mathbb{P})$, toutes les deux discrètes ou toutes les deux continues à densité admettant chacune un moment d'ordre deux. Le nombre réel :

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

s'appelle la covariance des v.a.r X et Y .

Proposition 12.2.22. Soient X et Y deux variables aléatoires réelles sur $(\Omega, \mathcal{T}, \mathbb{P})$, toutes les deux discrètes ou toutes les deux continues à densité admettant chacune un moment d'ordre deux. Alors $X + Y$ admet une variance et on a :

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

Remarque. On peut généraliser à plusieurs variables aléatoires : Si $X_1, \dots, X_n$ sont des variables aléatoires admettant des moments d'ordre 2 alors : $\sum_{k=1}^n X_k$ admet un moment d'ordre 2 et on a :

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

Proposition 12.2.23. Soient X et Y deux variables aléatoires réelles sur $(\Omega, \mathcal{T}, \mathbb{P})$, toutes les deux discrètes ou toutes les deux continues à densité admettant chacune un moment d'ordre deux. Si X et Y sont indépendantes alors

$$\text{Cov}(X, Y) = 0$$

en particulier :

$$V(X + Y) = V(X) + V(Y).$$

Remarque. Attention ! la réciproque n'est pas vraie.

Definition 12.2.22

Soient X et Y deux variables aléatoires réelles sur $(\Omega, \mathcal{T}, \mathbb{P})$, toutes les deux discrètes ou toutes les deux continues à densité admettant chacune un moment d'ordre deux et tel que : $V(X)V(Y) \neq 0$. Si $V(X)V(Y) \neq 0$, le nombre réel :

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

est appelé le coefficient de corrélation de X et Y .

Proposition 12.2.24. Soient X et Y deux variables aléatoires réelles sur $(\Omega, \mathcal{T}, \mathbb{P})$, toutes les deux discrètes ou toutes les deux continues à densité admettant chacune un moment d'ordre deux et tel que : $V(X)V(Y) \neq 0$. Alors :

1. $\rho(X, Y) \in [-1, 1]$.
2. $\rho(X, Y) = 1 \Leftrightarrow \exists(\alpha, \beta) \in \mathbb{R}^2, \begin{cases} \alpha > 0 \\ Y = \alpha X + \beta \end{cases}$.
3. $\rho(X, Y) = -1 \Leftrightarrow \exists(\alpha, \beta) \in \mathbb{R}^2, \begin{cases} \alpha > 0 \\ Y = -\alpha X + \beta \end{cases}$.

Remarques. Soient X et Y deux variables aléatoires réelles toutes les deux discrètes ou toutes les deux continues à densité admettant chacune un moment d'ordre deux et tel que $V(X)V(Y) \neq 0$. Alors :

1. Si $\rho(X, Y) = 0$ on dit que X et Y sont non corrélées.
2. Si X et Y sont indépendantes, alors elles sont non corrélées, la réciproque étant fausse.

12.2.8.9 Variable aléatoire centrée, réduite

Definition 12.2.23

Soit X une v.a.r. sur $(\Omega, \mathcal{T}, \mathbb{P})$.

1. Si X admet un moment d'ordre 1, alors la variable aléatoire

$$\tilde{X} = X - \mathbb{E}(X),$$

s'appelle la variable aléatoire centrée associée à X .

2. Si X admet un moment d'ordre 2 et si $V(X) \neq 0$, la v.a.r :

$$X^* = \frac{1}{\sigma(X)}(X - \mathbb{E}(X)),$$

s'appelle variable aléatoire centrée réduite associée à X .

Remarque. On a $\mathbb{E}(X^*) = 0$ et $\sigma(X^*) = 1$

Exemple. Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et X une v.a.r. continue à densité tel que $X \hookrightarrow \mathcal{U}[1, 2]$. On sait que $\mathbb{E}(X) = \frac{3}{2}$ et $V(X) = \frac{1}{12}$, donc $\sigma(X) = \frac{1}{2\sqrt{3}}$, alors $X^* = 2\sqrt{3}(X - \frac{3}{2})$.

12.3 Fonction génératrice

Dans tout ce paragraphe, $(\Omega, \mathcal{T}, \mathbb{P})$ est un espace probabilisé et on ne considère que des variables aléatoires discrètes à valeurs dans $\mathbb{N}$.

12.3.1 Définition, propriétés

Proposition 12.3.1. Soit X un v.a.r. représentée par (n, p_n) . La série entière $\sum p_n t^n$ a un rayon de convergence R tel que $R \geq 1$ et la fonction G_X définie par

$$G_X(t) = \sum_{n=0}^{+\infty} p_n t^n$$

est continue sur $[-1, 1]$ et elle est de classe C^∞ sur $] -R, R[$. On a $G_X(1) = 1$

Definition 12.3.1

La fonction G_X s'appelle fonction génératrice de la v.a.r. discrete X .

Remarque. Si $t > 0$ alors $G_X(t) = \mathbb{E}(t^X) = \sum_{n=0}^{+\infty} t^n \mathbb{P}(X = n)$.

Exemple. $\Omega = \mathbb{R}, \mathcal{T} = \mathcal{P}(\mathbb{R})$ et $\mathbb{P} : \mathcal{T} \rightarrow \mathbb{R}$ tel que, pour $A \in \mathcal{T}$:

$$\mathbb{P}(A) = \begin{cases} \frac{1}{3} \sum_{k \in \mathbb{Z} \cap A} \frac{1}{2^{|k|}} & \text{si } A \cap \mathbb{Z} \neq \emptyset \\ 0 & \text{si } A \cap \mathbb{Z} = \emptyset \end{cases}$$

Soit $X : \mathbb{R} \rightarrow \mathbb{R}$ tel que $X(x) = | \lfloor x \rfloor |$ (partie entière de x .)

Il est clair que $X(\mathbb{R}) = \mathbb{N}$ et pour tout $n \in \mathbb{N}$, si $n > 0$ alors $(X = n) = \{x \in \mathbb{R} / | \lfloor x \rfloor | = n\} = [-n, -n+1[\cup [n, n+1[$, donc $\mathbb{P}(X = n) = \frac{1}{3 \cdot 2^{n-1}}$ et $\mathbb{P}(X = 0) = \frac{1}{3}$, donc la fonction génératrice de X est :

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

Donc pour tout $t \in] -2, 2[$, on a $G_X(t) = \frac{1}{3} + \frac{2t}{6} \frac{2}{2-t} = \frac{2+t}{3(2-t)}$.

Convention : Soit X une v.a.r. discrète à valeur dans $\mathbb{N}$ et G_X sa fonction génératrice de rayon de convergence $R \geq 1$. On conviendra que si $R = 1$, la dérivabilité k fois au point 1 signifie la dérivabilité k à gauche au point 1.

Théorème 12.3.1. Soit X un v.a.r. discrète à valeurs dans $\mathbb{N}$, et G_X sa fonction génératrice, alors :

1. X admet un espérance si et seulement si G_X est dérivable au point 1, auquel cas on a : $\mathbb{E}(X) = G'_X(1)$.
2. X admet un moment d'ordre 2 (donc une variance) si et seulement si G_X est deux fois dérivable au point 1, auquel cas, on a : $\mathbb{E}(X^2) - \mathbb{E}(X) = G''_X(1)$.

Preuve. Soit R le rayon de convergence de $G_X(t) = \sum p_n t^n$.

• **Premiers cas : $R > 1$**

Puisque 1 est un point du disque de convergence de la série entière $G_X(t)$, alors il est clair que G_X est infiniment dérivable au point 1 et :

$$G'_X(1) = \sum_{n=1}^{+\infty} n p_n = \sum_{n=0}^{+\infty} n p_n = \mathbb{E}(X)$$

et

$$G''_X(1) = \sum_{n=2}^{+\infty} n(n-1) p_n = \sum_{n=0}^{+\infty} n^2 p_n - \sum_{n=0}^{+\infty} n p_n = \mathbb{E}(X^2) - \mathbb{E}(X)$$

• **Deuxième cas : $R = 1$**

1) **Dérivée première :**

• Supposons que $\mathbb{E}(X)$ existe. Si on pose $f_n(t) = p_n t^n$, pour tout $n \in \mathbb{N}$ et tout $t \in [-1, 1]$ alors les f_n sont de classe C^1 sur $[-1, 1]$ et :

$$\forall n \in \mathbb{N}^*, \forall t \in [-1, 1], f'_n(t) = n p_n t^{n-1}$$

(Bien entendu $f'_0(t) = 0$ pour tout $t \in [-1, 1]$). Il en découle que $|f'_n(t)| \leq n p_n$ pour tout $n \in \mathbb{N}$ donc la série $\sum f'_n$ converge normalement donc uniformément sur $[-1, 1]$ donc on peut dériver terme à terme la série de fonctions $\sum f_n$ sur $[-1, 1]$, donc :

$$\forall t \in [-1, 1], G'_X(t) = \sum_{n=1}^{+\infty} n p_n t^{n-1}$$

en particulier :

$$G'_X(1) = \sum_{n=1}^{+\infty} n p_n = \sum_{n=0}^{+\infty} n p_n = \mathbb{E}(X).$$

• Réciproquement, supposons que G_X est dérivable au point 1. alors : $G'_x(1) = \lim_{x \rightarrow 1} \Delta_X(t)$ où $\Delta_X(t) = \frac{G_X(t) - G_X(1)}{t-1}$, pour tout $t \in [0, 1[$. Remarquons que l'on a :

$$\begin{aligned} G_X(t) - G_X(1) &= \sum_{n=0}^{+\infty} p_n t^n - \sum_{n=0}^{+\infty} p_n \\ &= \sum_{n=1}^{+\infty} p_n (t^n - 1) \\ &= \sum_{n=1}^{+\infty} (t-1) p_n \left(\sum_{k=0}^{n-1} t^k \right) \\ &= (t-1) \sum_{n=1}^{+\infty} p_n \sum_{k=0}^{n-1} t^k \end{aligned}$$

Donc :

$$\Delta_X(t) = \sum_{n=1}^{+\infty} p_n P_n(t), \quad \text{avec } P_n(t) = \sum_{k=0}^{n-1} t^k$$

On veut démontrer que $\mathbb{E}(X)$ existe, donc que la série à termes positifs $\sum n p_n, n \geq 1$ est convergente, pour cela il suffit de prouver que sa somme partielle d'ordre n , $S_n = \sum_{k=1}^n k p_k$ est majorée. Commençons par remarquer que si on pose

$$F_n(t) = \sum_{k=1}^n p_k P_k(t)$$

alors

$$\lim_{t \rightarrow 1} F_n(t) = \sum_{k=1}^n k p_k = S_n$$

Or :

$$\forall t \in [0, 1[, \quad F_n(t) \leq \Delta_n(t)$$

car les p_k sont positifs, donc par passage à la limite :

$$S_n(t) \leq G'_X(1)$$

ce qui prouve l'existence de $\mathbb{E}(X)$. D'après la première partie de cette démonstration, on a alors $\mathbb{E}(X) = G'_X(1)$.

Dérivée seconde :

• Supposons que X admet un moment d'ordre 2. Comme dans le cas de la dérivée première, il est aisé de voir que la série des dérivées secondes des applications $f_n : [-1, 1] \rightarrow \mathbb{R}; t \mapsto f_n(t) = p_n t^n$ est normalement

donc uniformément convergente sur $[-1, 1]$, ce qui permet en particulier de dire G_X est deux fois dérivable au point 1 et que $G'_X(1) = \mathbb{E}(X)$ et $G''_X(1) = \sum_{n=0}^{+\infty} n(n-1)p_n = \mathbb{E}(X^2) - \mathbb{E}(X)$.

• Réciproquement, si G_X est deux fois dérivable au point 1, alors elle est déjà dérivable une fois et d'après la première partie du théorème $\mathbb{E}(X)$ existe et $G'_X(1) = \mathbb{E}(X)$. Par ailleurs on a :

$$\forall t \in [0, 1[, G'_X(t) = \sum_{n=1}^{+\infty} np_n t^n \quad \text{et} \quad G'_X(1) = \sum_{n=1}^{+\infty} np_n$$

ce qui permet de dire que : $G''_X(1) = \lim_{t \rightarrow 1} \Delta'_X(t)$ avec

$$\Delta'_X(t) = \frac{G'_X(t) - G'_X(1)}{t - 1} = \sum_{n=2}^{+\infty} n(n-1)p_n Q_n(t)$$

avec $Q_n(t) = \sum_{k=0}^{n-2} t^k$. Comme dans la première partie, on va montrer que $T_n = \sum_{k=0}^n k(k-1)p_k$ est majoré,

on commence par remarquer que si on pose $H_n(t) = \sum_{k=2}^n k(k-1)p_k Q_k(t)$ alors $\lim_{t \rightarrow 1} H_n(t) = T_n$. Or $H_n(t) \leq \Delta'_X(t)$ et par passage à la limite, on obtient $T_n \leq G''_X(1)$ et par suite la série $\sum n(n-1)p_n$ converge et comme $\mathbb{E}(X)$ existe la série $\sum np_n$ converge donc la série $\sum n^2 p_n$ converge d'où l'existence du moment d'ordre 2 et d'après la preuve du premier sens on a $G''_X(2) = \mathbb{E}(X^2) - \mathbb{E}(X)$.

Ceci termine la preuve du théorème. ■

Remarque. Lorsque le rayon de convergence R de G_X réalise $R > 1$, il est clair que G_X est infiniment dérivable au point 1, en particulier X admet une espérance et une variance et on a les relations ci-dessus données par le théorème.

12.3.2 Fonction génératrice d'une somme de v.a.r. indépendantes

Théorème 12.3.2. Si $X_k, 1 \leq k \leq m$ sont des variables aléatoires réelles discrètes à valeurs dans $\mathbb{N}$ et si ces v.a.r. sont indépendantes alors :

$$G_{\sum_{k=1}^m X_k} = \prod_{k=1}^m G_{X_k}$$

Preuve. il suffit de le prouver pour deux variables, le reste s'en déduit par simple récurrence. Soit donc X et Y deux v.a.r. indépendantes tel que $X(\Omega) = Y(\Omega) = \mathbb{N}$. Alors :

$$\begin{aligned} G_{X+Y}(t) &= \sum_{n=0}^{+\infty} \mathbb{P}(X+Y=n) t^n \\ &= \sum_{n=0}^{+\infty} \sum_{k=0}^n \mathbb{P}(X=k) \mathbb{P}(Y=n-k) t^n \\ &= \sum_{n=0}^{+\infty} \sum_{k=0}^n \mathbb{P}(X=k) t^k \mathbb{P}(Y=n-k) t^{n-k} \end{aligned}$$

On remarque qu'il s'agit du produit de Cauchy des séries $\sum p_n t^n$ et $\sum q_n t^n$ où $p = \mathbb{P}(X = n)$ et $q_n = \mathbb{P}(Y = n)$, qui sont absolument convergente pour tout t tel que $|t| < 1$, donc :

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

où $G_X(t) = \sum p_n t^n$ et $G_Y(t) = \sum q_n t^n$ sont les fonctions génératrices respectives de X et Y . ■

12.4 Inégalités, Convergence

12.4.1 Inégalité

12.4.1.1 Inégalité de Markov

Théorème 12.4.1. Soit X une v.a.r positive sur $(\Omega, \mathcal{T}, \mathbb{P})$ tel que X admet une espérance, alors : pour tout réel $\alpha > 0$, on a :

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

(inégalité de Markov)

Preuve. Dans le cas d'une variable continue à densité comme $X \geq 0$, la fonction densité f_X est nulle sur $] -\infty, 0[$, puisque la fonction F_X de répartition est nulle sur cet intervalle (Si $t < 0$ alors $(X \leq t) = \emptyset$), donc :

$$\mathbb{E}(X) = \int_0^{+\infty} t f_X(t) dt = \int_0^{\alpha} t f_X(t) dt + \int_{\alpha}^{+\infty} t f_X(t) dt$$

Il en découle :

$$\mathbb{E}(X) \geq \int_{\alpha}^{+\infty} t f_X(t) dt \geq \int_{\alpha}^{+\infty} \alpha f_X(t) dt = \alpha \int_{\alpha}^{+\infty} f_X(t) dt = \alpha \mathbb{P}(X \geq \alpha)$$

Ce qui donne l'inégalité de Markov.

Si X est discrete d'ensemble de valeurs possibles D alors :

$$\mathbb{E}(X) = \sum_{x \in D} x \mathbb{P}(X = x) = \sum_{\substack{x \in D \\ x \geq \alpha}} x \mathbb{P}(X = x) + \sum_{\substack{x \in D \\ x < \alpha}} x \mathbb{P}(X = x)$$

Donc, compte tenu de $X \geq 0$ on a :

$$\mathbb{E}(X) \geq \sum_{\substack{x \in D \\ x \geq \alpha}} x \mathbb{P}(X = x) \geq \alpha \sum_{\substack{x \in D \\ x \geq \alpha}} \mathbb{P}(X = x) = \alpha \mathbb{P}(X \geq \alpha)$$

ce qu'il fallait démontrer. ■

12.4.1.2 Inégalité de Bienaymé-Chebychev

Théorème 12.4.2. Soit X une v.a.r sur $(\Omega, \mathcal{T}, \mathbb{P})$. Si X admet une variance alors on a :

$$\forall \beta > 0, \mathbb{P}(|X - E(X)| \geq \beta) \leq \frac{\mathbb{V}(X)}{\beta^2}$$

(inégalité de Bienaymé-Chebychev)

Preuve. On va appliquer Markov à $Y = (X - E(X))^2$. On a $\mathbb{E}(Y) = \mathbb{V}(X)$ et d'après Markov :

$$\mathbb{P}(Y \geq \beta^2) \leq \frac{\mathbb{E}(Y)}{\beta^2}$$

Or $Y \geq \beta^2 \Leftrightarrow |X - E(X)| \geq \beta$ et $\mathbb{E}(Y) = \mathbb{V}(X)$, d'où l'inégalité de Bienaymé-Chebychev. ■

12.4.1.3 inégalité de Jensen

Avant de donner l'énoncé de ce théorème on rappelle que toute application $f : \mathbb{R} \rightarrow \mathbb{R}$ convexe sur $\mathbb{R}$ est continue sur $\mathbb{R}$.

on admet que si X est une variable aléatoire sur $(\Omega, \mathcal{T}, \mathbb{P})$ est $f : \mathbb{R} \rightarrow \mathbb{R}$ continue (par morceaux) alors $f(X)$ est une variable aléatoire.

Théorème 12.4.3. Soit X une variable aléatoire réelle ayant une espérance et $f : \mathbb{R} \rightarrow \mathbb{R}$ convexe. Si $f(X)$ admet une espérance alors :

$$f(\mathbb{E}(X)) \leq \mathbb{E}(f(X))$$

Preuve. Dans la cas d'une variable aléatoire discrete. On suppose X représentée par $(x_k, p_k)_{k \in \mathbb{N}^*}$. Pour tout $n \in \mathbb{N}^*$, posons : $E_n = \sum_{k=1}^n p_k x_k$ alors $\mathbb{E}(X) = \lim_{n \rightarrow +\infty} (E_n)$. Si $\pi_n = \sum_{k=1}^n p_k$ alors par convexité de f , on a :

$$f\left(\frac{E_n}{\pi_n}\right) \leq \sum_{k=1}^n \frac{p_k}{\pi_n} f(x_k) = \frac{1}{\pi_n} \sum_{k=1}^n p_k f(x_k)$$

Par passage à la limite quand n tends vers $+\infty$ et comme $f(X)$ admet une espérance et par continuité de f et le fait que $\lim_{n \rightarrow +\infty} \pi_n = 1$, on obtient :

$$f(\mathbb{E}(X)) \leq \mathbb{E}(f(X))$$

■

12.4.2 Convergence

12.4.2.1 Convergence en loi

Definition 12.4.1

Soit (X_n) une suite de v.a.r et X une v.a.r. On dit que (X_n) converge en loi vers X si la suite (F_n) des fonctions de répartition des X_n converge simplement vers F la fonction de répartition de X sur $\mathbb{R} \setminus D$ où D est l'ensemble des points de discontinuité de F .

Proposition 12.4.1. Si X ainsi que les X_n sont discrètes à valeurs dans $\mathbb{N}$ alors $X_n \rightarrow X$ en loi si et seulement si

$$\forall k \in \mathbb{N}, \lim_{n \rightarrow +\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$$

Exemples. 1. Soit (X_n) sur $(\Omega, \mathcal{T}, \mathbb{P})$ tel que $X_n \hookrightarrow \mathcal{B}(p)$, alors $X_n \rightarrow X_1$ en loi.

2. Soit (X_n) une suite de v.a.r sur $(\Omega, \mathcal{T}, \mathbb{P})$ tel qu'il existe une suite $(p_n) \in [0, 1]^{\mathbb{N}}$ et $\lambda > 0$ tel que $\lim np_n = \lambda$. Si $X_n \hookrightarrow \mathcal{B}(n, p_n)$ pour tout n alors $X_n \rightarrow X$ en loi où $X \hookrightarrow \mathcal{P}(\lambda)$ (voir TD).

3. On considère ici $\Omega = \mathbb{R}$, $\mathcal{T} = \mathcal{P}(\mathbb{R})$ et $\mathbb{P}(A) = \begin{cases} \frac{1}{3} \sum_{k \in A \cap \mathbb{Z}} \frac{1}{2^{|k|}} & \text{si } A \cap \mathbb{Z} \neq \emptyset \\ 0 & \text{si } A \cap \mathbb{Z} = \emptyset \end{cases}$ et pour tout $n \in \mathbb{N}^*$.

Soit $x_0 \in \mathbb{R}$. On dit qu'une v.a.r. Y sur $(\Omega, \mathcal{T}, \mathbb{P})$ suit la loi de dirac de paramètre x_0 , et on note $Y \hookrightarrow \mathcal{D}(x_0)$ si $\mathbb{P}(Y = x_0) = 1$. Il en résulte que l'ensemble de valeurs possibles de Y est $D = \{x_0\}$. Un exemple simple de telle variables est l'application constante $\omega \mapsto x_0$ de $\mathbb{R}$ vers $\mathbb{R}$.

1 Soit X_n une variable aléatoire réelle tel que $X_n \hookrightarrow \mathcal{B}(\frac{1}{n})$. Alors $X_n \rightarrow X$ en loi où $X \hookrightarrow \mathcal{D}(0)$. En effet, nous avons, pour tout $n \in \mathbb{N}^*$:

$$F_n(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 - \frac{1}{n} & \text{si } 0 \leq t < 1 \\ 1 & \text{si } 1 \leq t \end{cases}$$

et par suite, (F_n) converge simplement sur $\mathbb{R}$ vers F définie par

$$F(t) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } 0 \leq x \end{cases}$$

On voit que F est la fonction de répartition d'une variable aléatoire X suivant $\mathcal{D}(0)$.

12.4.2.2 Convergence en probabilité

Definition 12.4.2

On dit qu'une suite (X_n) de v.a.r sur $(\Omega, \mathcal{T}, \mathbb{P})$ converge en probabilité vers une v.a.r X sur $(\Omega, \mathcal{T}, \mathbb{P})$ si :

$$\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} (\mathbb{P}(|X_n - X| \geq \varepsilon)) = 0$$

Proposition 12.4.2. Soit $(f_n)_{n \in \mathbb{N}}$ une suite d'applications de $\mathbb{R}$ vers $\mathbb{R}$ qui converge simplement sur $\mathbb{R}$ vers une application $f : \mathbb{R} \rightarrow \mathbb{R}$. Si X est une v.a.r. sur $(\Omega, \mathcal{T}, \mathbb{P})$ tel que $Y_n = f_n(X), n \in \mathbb{N}$ et $Y = f(X)$ alors (Y_n) tends vers Y en probabilité.

Preuve. Soit

$$\Omega_1 = \{\omega \in \Omega / \lim_{n \rightarrow +\infty} (Y_n(\omega)) = Y(\omega)\}$$

On a $\Omega_1 = \Omega$ car $\Omega_1 \subset \Omega$ par définition de Ω_1 ; réciproquement, si $\omega \in \Omega$ alors $X(\omega) \in \mathbb{R}$ et comme $(f_n) \rightarrow f$ simplement, on a $f_n(X(\omega)) \rightarrow f(X(\omega))$, d'où $Y_n(\omega) \rightarrow Y(\omega)$.

Notons que pour tout $\omega \in \Omega$, on a :

$$\omega \in \Omega_1 \Leftrightarrow \forall \alpha > 0, \exists N \in \mathbb{N}, \forall n \geq N, |Y_n(\omega) - Y(\omega)| < \alpha$$

Par suite, on peut dire que

$$\Omega = \Omega_1 = \bigcap_{\alpha > 0} \bigcup_{N \in \mathbb{N}} \bigcap_{n \geq N} A_{n, \alpha}$$

où

$$A_{n, \alpha} = (|Y_n - Y| < \alpha)$$

Soit $\varepsilon > 0$, alors :

$$\Omega \subset \bigcup_{N \in \mathbb{N}} \bigcap_{n \geq N} A_{n, \varepsilon} \subset \Omega$$

par suite :

$$\Omega = \bigcup_{N \in \mathbb{N}} \bigcap_{n \geq N} A_{n, \varepsilon}$$

Pour tout $N \in \mathbb{N}$, notons : $B_N = \bigcap_{n \geq N} A_{n, \varepsilon}$. Il est aisé de voir que $B_N \in \mathcal{T}$. Il découle de ce qui précède que :

$$\Omega = \bigcup_{N \in \mathbb{N}} B_N$$

La famille $(B_N)_{N \geq 0}$ est une famille croissante d'éléments de $\mathcal{T}$ puisque, $B_N = A_{N, \varepsilon} \cap B_{N+1} \subset B_{N+1}$.

Par le théorème de la continuité séquentielle monotone, on a : $1 = \mathbb{P}(\Omega) = \sup_{N \in \mathbb{N}} \mathbb{P}(B_N)$.

Soit $\delta \in \mathbb{R}_+^*$, par définition de la borne supérieure, il existe $N \in \mathbb{N}$ tel que $1 - \delta < \mathbb{P}(B_N)$. Remarquons que $\forall n \geq N, B_N \subset A_{n, \varepsilon}$, donc $\mathbb{P}(A_{n, \varepsilon}) \geq \mathbb{P}(B_N)$, et par suite :

$$\forall n \geq N, \mathbb{P}(A_{n, \varepsilon}) > 1 - \delta$$

Or : $A_{n, \varepsilon} = (|Y_n - Y| > \varepsilon)$, donc $\mathbb{P}(A_{n, \varepsilon}) = 1 - \mathbb{P}(|Y_n - Y| \geq \varepsilon)$, ce qui donne :

$$\forall n \geq N, 1 - \mathbb{P}(|Y_n - Y| \geq \varepsilon) > 1 - \delta$$

donc :

$$\forall \varepsilon > 0, \forall \delta > 0, \exists N \in \mathbb{N}, \forall n \geq N, \mathbb{P}(|Y_n - Y| \geq \varepsilon) < \delta$$

Ceci démontre que :

$$\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}(|Y_n - Y| \geq \varepsilon) = 0$$

ce qui signifie que Y_n converge en probabilité vers Y . ■

Théorème 12.4.4. Soit (X_n) une suite de v.a.r. et X une v.a.r. sur $(\Omega, \mathcal{T}, \mathbb{P})$. Si (X_n) converge en probabilité vers X alors (X_n) converge en loi vers X .

Preuve. Ce théorème est admis. ■

12.4.3 Lois faible des grands nombres et théorème limite central

12.4.3.1 Loi faible des grands nombres

Théorème 12.4.5. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a.r indépendantes et identiquement distribuées admettant un moment d'ordre 2. Soit μ l'espérance mathématique commune de ces variables aléatoires $X_n, n \geq 1$. Alors la suite $\left(M_n = \frac{1}{n} \sum_{k=1}^n X_k\right)_{n \geq 1}$ converge en probabilité vers la variable aléatoire constante μ .

Preuve. Soit $M_n = \frac{1}{n} \sum_{k=1}^n X_k$. Par linéarité de l'espérance, on a $E(M_n) = \mu$ et par indépendance on a $V(M_n) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$ où σ est l'écart-type commun des v.a.r. X_n . Par l'inégalité de Bienaymé-Chebychev, on a :

$$\forall \varepsilon > 0, \mathbb{P}(|M_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

Pa conséquent :

$$\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}(|M_n - \mu| \geq \varepsilon) = 0$$

■

12.4.3.2 Théorème limite central

Théorème 12.4.6. Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. indépendantes, identiquement distribuées admettant un moment d'ordre 2 et soit μ et σ l'espérance mathématique et l'écart type communs de ces variables aléatoires. Soit (M_n) la suite des v.a.r. tel que $M_n = \frac{1}{n} \sum_{k=1}^n X_k$. Alors la suite (M_n^*) des variables aléatoires réduites centrées des M_n converge en loi vers une v.a.r. qui suit la loi de Gauss standard : $\mathcal{N}(0, 1)$

Remarque. Dans la pratique, quand n est suffisamment grand, on assimile la moyenne M_n avec une variable aléatoire Y tel que $Y \hookrightarrow \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$

En effet, la variable aléatoire centrée M_n^* est liée à M_n par

$$M_n^* = \frac{1}{\sigma(M_n)}(M_n - \mathbb{E}(M_n)).$$

Or $M_n = \frac{1}{n}S_n$, où $S_n = \sum_{k=1}^n X_k$. Donc $E(M_n) = \frac{1}{n}(n\mu) = \mu$ et par indépendance, $V(M_n) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$, donc

$$\sigma(M_n) = \frac{\sigma}{\sqrt{n}},$$

donc

$$M_n^* = \frac{\sqrt{n}}{\sigma}(M_n - \mu),$$

ce qui donne aussi

$$M_n = \frac{\sigma}{\sqrt{n}}M_n^* + \mu.$$

On peut démontrer facilement que si X est une variable aléatoire et X^* la variable aléatoire centrée réduite associée alors $X \hookrightarrow \mathcal{N}(\mu, \sigma) \Leftrightarrow X^* \hookrightarrow \mathcal{N}(0, 1)$.

Ainsi le fait d'identifier M_n^* à une variable aléatoire qui suit la loi $\mathcal{N}(0, 1)$ est équivalent à identifier M_n à X tel que $X \hookrightarrow \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$

Exemple. 200 personnes contribuent à un jeu qui consiste à faire tourner une roue au hasard une seule fois pour chaque personne. On note X_k la variable aléatoire associée à l'angle modulo 2π duquel la roue tourne après la contribution de la k ème personne. On suppose que toutes les précautions sont prises pour que les X_k sont mutuellement indépendantes. Calculer la probabilité que la moyenne des angles associés aux 200 personnes ne dépasse pas un demi tour.

Réponse : Il s'agit d'une famille de variables aléatoires $(X_k), 1 \leq k \leq 200$, qui suivent la loi uniforme $\mathcal{U}[0, 2\pi]$. Les v.a.r X_k sont identiquement distribuées d'espérance $\mu = \pi$ et d'écart-type $\sigma = \frac{\pi}{\sqrt{3}}$. Si on applique le théorème central limite, on identifie la moyenne M_{200} à une v.a.r X tel que $X \hookrightarrow \mathcal{N}(\pi, \frac{\pi}{10\sqrt{6}})$. On veut :

$$\begin{aligned} \mathbb{P}(M_n \leq \pi) &= \mathbb{P}(X \leq \pi) \\ &= \frac{10\sqrt{6}}{\pi\sqrt{2\pi}} \int_{-\infty}^{\pi} e^{-\frac{1}{2}(\frac{10\sqrt{6}(t-\pi)}{\pi})^2} dt \\ &\simeq 0.5 \end{aligned}$$

12.5 Annexe

12.5.1 Tableau récapitulatif des variables aléatoires réelles discrètes usuelles

Nom	Notation	$X(\Omega)$	Loi de X	$\mathbb{E}(X)$	$\mathbb{V}(X)$	$G_X(t)$
Uniforme discrète	$X \hookrightarrow \mathcal{U}(A)$ $A = \{1, \dots, n\}$	A	$k \in \{1 \cdots n\}$ $\mathbb{P}(X = k) = \frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\begin{cases} \frac{t}{n} \frac{1-t^n}{1-t} & \text{si } t \neq 1 \\ 1 & \text{si } t = 1 \end{cases}$
Bernoulli de paramètre $p \in [0, 1]$	$X \hookrightarrow \text{Ber}(p)$	$\{0, 1\}$	$\begin{cases} \mathbb{P}(X = 0) = 1 - p \\ \mathbb{P}(X = 1) = p \end{cases}$	p	$p(1-p)$	$1 - p + pt$
Binomiale de paramètres n et p	$X \hookrightarrow B(n, p)$	$\llbracket 0, n \rrbracket$	$\mathbb{P}(X = k) = C_n^k p^k q^{n-k}$ avec $q = 1 - p$	np	$np(1-p)$	$(1 - p + pt)^n$
Poisson de paramètre λ , avec $\lambda \in \mathbb{R}_+^*$	$X \hookrightarrow \mathcal{P}(\lambda)$	$\mathbb{N}$	$\mathbb{P}(X = n) = e^{-\lambda} \frac{\lambda^n}{n!}$	λ	λ	$e^{\lambda(t-1)}$
Loi géométrique de paramètre p avec $p \in]0, 1[$	$X \hookrightarrow \mathcal{G}(p)$	$\mathbb{N}^*$	$\mathbb{P}(X = n) = p(1-p)^{n-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pt}{1 - (1-p)t}$

12.5.2 Tableau récapitulatif des variables aléatoires réelles continues à densité usuelles :

Nom	Notation	Loi de X	$\mathbb{E}(X)$	$\mathbb{V}(X)$
Uniforme continue sur le segment $[a, b]$	$X \hookrightarrow \mathcal{U}([a, b])$	X a pour fonction de répartition : $F_X(t) = \begin{cases} 0 & \text{si } t < 0 \\ \frac{t-a}{b-a} & \text{si } a \leq t \leq b \\ 1 & \text{si } t > b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Loi exponentielle de paramètre λ avec $\lambda \in \mathbb{R}_+^*$	$X \hookrightarrow \mathcal{E}(\lambda)$	X a pour fonction de répartition : $F_X(t) = \begin{cases} 1 - e^{-\lambda t} & \text{si } t \geq 0 \\ 0 & \text{si } t < 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Loi normale de Gauss de paramètres μ et σ avec $\mu, \sigma \in \mathbb{R}$ et $\sigma > 0$	$X \hookrightarrow \mathcal{N}(\mu, \sigma^2)$	X a pour densité f_X tel que : $f_X(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$	μ	σ^2
Loi Gamma de paramètres a et λ avec $a, \lambda \in \mathbb{R}_+^*$	$X \hookrightarrow \Gamma(a, \lambda)$	X a pour densité f_X tel que : $f_X(t) = \begin{cases} \frac{\lambda^a}{\Gamma(a)} e^{-\lambda t} t^{a-1} & \text{si } t > 0 \\ 0 & \text{si } t \leq 0 \end{cases}$	$\frac{a}{\lambda}$	$\frac{a}{\lambda^2}$

12.5.3 Stabilité de certaines lois

12.5.3.1 Loi de Bernoulli, loi binomiale

Proposition 12.5.1. X et Y sont deux v.a.r indépendantes sur $(\Omega, \mathcal{T}, \mathbb{P})$. Si $X \hookrightarrow \mathcal{B}er(p)$ et $Y \hookrightarrow \mathcal{B}er(q)$ et X et Y alors $XY \hookrightarrow \mathcal{B}er(pq)$

Remarque. Généralement si $X_1, \dots, X_m$ sont des v.a.r indépendantes tel que

$$\forall k \in \llbracket 1, m \rrbracket, X_k \hookrightarrow \mathcal{B}er(p_k)$$

alors

$$X = \prod_{k=1}^m X_k \hookrightarrow \mathcal{B}er(p) \text{ où } p = \prod_{k=1}^m p_k$$

Proposition 12.5.2. Si X, Y sont deux v.a.r indépendantes sur $(\Omega, \mathcal{T}, \mathbb{P})$ et si $X \hookrightarrow \mathcal{B}(n, p)$ et $Y \hookrightarrow \mathcal{B}(m, p)$ alors $X + Y \hookrightarrow \mathcal{B}(n + m, p)$

Remarques. 1. Plus généralement si $X_1, \dots, X_n$ indépendantes et $X_k \hookrightarrow \mathcal{B}(n_k, p)$ alors

$$X = \sum_{k=1}^n X_k \hookrightarrow \mathcal{B}\left(\sum_{k=1}^n n_k, p\right)$$

2. Si $(X_k)_{1 \leq k \leq n}$ indépendantes tel que $\forall k \in \llbracket 1, n \rrbracket, X_k \hookrightarrow \mathcal{B}er(p)$ alors

$$\sum_{k=1}^n X_k \hookrightarrow \mathcal{B}(n, p)$$

12.5.3.2 Lois de Poisson

Proposition 12.5.3. Soit $m \in \mathbb{N}^*$ et $(X_k)_{1 \leq k \leq m}$ une famille de v.a.r indépendantes. On note $X = \sum_{k=1}^m X_k$ et $\lambda = \sum_{k=1}^m \lambda_k$. Si pour tout $k \in \llbracket 1, m \rrbracket, X_k \hookrightarrow \mathcal{P}(\lambda_k)$ alors $X \hookrightarrow \mathcal{P}(\lambda)$

Preuve. Par récurrence : Si $n = 2$, alors $X_1 + X_2(\Omega) = \mathbb{N}$ et pour tout $n \in \mathbb{N}$, on a

$$\begin{aligned} \mathbb{P}(X_1 + X_2 = n) &= \sum_{j=0}^n \mathbb{P}(X_1 = j) \mathbb{P}(X_2 = n - j) \\ &= \sum_{j=0}^n e^{-\lambda_1} \frac{\lambda_1^j}{j!} e^{-\lambda_2} \frac{\lambda_2^{n-j}}{(n-j)!} \\ &= e^{-(\lambda_1 + \lambda_2)} \frac{1}{n!} \sum_{j=0}^n \binom{n}{j} \lambda_1^j \lambda_2^{n-j} \\ &= e^{-\lambda} \frac{\lambda^n}{n!} \end{aligned}$$

où $\lambda = \lambda_1 + \lambda_2$.

Supposons la propriété vraie pour $m \in \mathbb{N}^*$ et soit $X_k, k \in \llbracket 1, m+1 \rrbracket$ des v.a.r indépendantes tel que $X_k \hookrightarrow \mathcal{P}(\lambda_k)$ pour tout $k \in \llbracket 1, m+1 \rrbracket$. Par hypothèse de récurrence, $\sum_{k=1}^m X_k \hookrightarrow \mathcal{P}(\sum_{k=1}^m \lambda_k)$ et d'après la propriété vraie pour $m = 2$ on en déduit que $\sum_{k=1}^{m+1} X_k \hookrightarrow \mathcal{P}(\sum_{k=1}^{m+1} \lambda_k)$ ■

12.5.3.3 Loi Γ

Proposition 12.5.4. Si X et Y sont deux v.a.r indépendantes sur $(\Omega, \mathcal{T}, \mathbb{P})$ tel que $X \hookrightarrow \Gamma(\lambda, a)$ et $Y \hookrightarrow \Gamma(\lambda, b)$ alors $X + Y \hookrightarrow \Gamma(\lambda, a + b)$

On donne les remarques suivantes :

1. Généralement si $X_1, \dots, X_m$ sont des v.a.r indépendantes sur $(\Omega, \mathcal{T}, \mathbb{P})$ et si $X_k \hookrightarrow \Gamma(\lambda, a_k)$ alors $\sum_{k=1}^m X_k \hookrightarrow \Gamma(\lambda, \sum_{k=1}^m a_k)$
2. Remarquons que $\Gamma(\lambda, 1) = \mathcal{E}(\lambda)$ donc si $X_1, \dots, X_m$ indépendantes suivent $\mathcal{E}(\lambda)$ leur somme suit $\Gamma(\lambda, m)$.

12.5.3.4 Loi normale

Proposition 12.5.5. Si X et Y sont deux v.a.r indépendantes sur $(\Omega, \mathcal{T}, \mathbb{P})$ tel que $X \hookrightarrow \mathcal{N}(\mu_1, \sigma_1)$ et $Y \hookrightarrow \mathcal{N}(\mu_2, \sigma_2)$ alors $X + Y \hookrightarrow \mathcal{N}(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$

Généralement si $X_1, \dots, X_m$ indépendantes tel que $X_k \hookrightarrow \mathcal{N}(\mu_k, \sigma_k)$ alors

$$X = \sum_{k=1}^m X_k \hookrightarrow \mathcal{N}(\mu, \sigma)$$

où

$$\mu = \sum_{k=1}^m \mu_k \quad \text{et} \quad \sigma = \sqrt{\sum_{k=1}^m \sigma_k^2}$$