

Table des matières

1	Dénombrer et sommer	5
1.1	Rappels ensemblistes	5
1.1.1	Opérations ensemblistes	5
1.1.2	Bijections	7
1.2	Ensembles finis et dénombrement	8
1.3	Dénombrabilité	16
1.4	Rappels sur les séries	24
1.4.1	Généralités	24
1.4.2	Séries à termes positifs	25
1.4.3	Séries à termes de signe non constant	27
1.4.4	Opérations sur les séries	28
1.5	Familles sommables	34
1.6	Séries doubles	46
2	Événements et Probabilités	51
2.1	Notion de mesure	51
2.2	Modéliser l'aléatoire	57
2.2.1	Notion d'expérience aléatoire	57
2.2.2	Événements	58
2.2.3	Une question de dés	59
2.3	La probabilité comme mesure	62
2.4	Exemples	68
2.5	Remarques sur le choix d'un modèle	75
2.6	Probabilités conditionnelles	77
2.6.1	Introduction	77
2.6.2	Propriétés	79
2.6.3	Quelques exemples	82
2.7	Indépendance	84
2.7.1	Indépendance de deux événements	84
2.7.2	Indépendance mutuelle	86
2.7.3	Épreuves répétées	88

3	Variables aléatoires	91
3.1	Introduction	91
3.2	Généralités	94
3.2.1	Variables aléatoires réelles	94
3.2.2	Loi d'une variable aléatoire	96
3.2.3	Fonction de répartition	99
3.2.4	Lois à densité	101
3.3	Lois discrètes classiques	106
3.3.1	Lois de Bernoulli	107
3.3.2	Loi uniforme sur un ensemble fini de réels	107
3.3.3	Lois binomiales	107
3.3.4	Lois hypergéométriques	108
3.3.5	Lois géométriques	110
3.3.6	Lois de Poisson	111
3.3.7	Sur le caractère universel de la loi de Poisson	116
3.4	Lois à densité classiques	119
3.4.1	Lois uniformes	119
3.4.2	Lois exponentielles	121
3.4.3	Lois gaussiennes	124
3.4.4	Lois de Cauchy	125
4	Espérance	127
4.1	Introduction	127
4.2	Espérance d'une variable aléatoire positive	130
4.3	Espérance d'une variable aléatoire réelle	146
4.4	Moments	152
5	Vecteurs aléatoires et indépendance	161
5.1	Vecteurs aléatoires	162
5.1.1	Généralités	162
5.1.2	Covariance	171
5.2	Indépendance de variables et vecteurs aléatoires	173
5.2.1	Suites indépendantes	173
5.2.2	Indépendance des composantes	178
5.2.3	Indépendance et espérance de produits	186
6	Théorèmes limites	191
6.1	Convergences de suites de v.a.	191
6.1.1	Convergence presque sûre et en probabilité	191
6.1.2	Convergence en moyenne d'ordre p	198
6.1.3	Bilan sur les convergences de v.a.	203
6.2	Loi des grands nombres	204
6.2.1	Loi faible des grands nombres	204
6.2.2	Loi forte des grands nombres	206

6.2.3	L'aiguille de Buffon	208
A	Intégrale de Riemann sur $[a, b]$	213
A.1	Construction	213
A.2	Riemann intégrabilité	219
A.3	Propriétés de l'intégrale de Riemann	226
A.3.1	Propriétés de l'ensemble $\mathcal{R}[a, b]$	226
A.3.2	Propriétés relatives à l'intervalle d'intégration	232
A.4	Interversion limite intégrale	238
B	Intégrale généralisée	241
B.1	Construction	241
B.2	Critère de Cauchy pour intégrales généralisées	251
B.3	Intégrales généralisées de fonctions positives	257
B.4	Divers	262
B.4.1	Changements de variable	262
B.4.2	Intégration par parties	265
B.4.3	Comparaison des intégrales ordinaires et généralisées	267
	Tables de la loi normale standard	269

Chapitre 1

Dénombrer et sommer

Compter des objets et faire des additions, voilà bien les deux activités les plus élémentaires à la base des mathématiques. Et pourtant à y regarder de plus près, ce n'est pas si facile. Déjà pour un ensemble fini, la méthode qui consiste à regarder ses éléments l'un après l'autre et à les compter (donc à les numérotter) n'est applicable que pour de « petits » ensembles. Le plus souvent on s'en sort en faisant une représentation de l'ensemble à dénombrer à l'aide d'un autre ensemble plus familier. Cette représentation est ce que l'on appelle une bijection. Elle est d'ailleurs à la base du processus de comptage qui consiste simplement à mettre en bijection un ensemble avec un ensemble de nombres entiers. Cette notion de bijection permet d'étendre en un certain sens le dénombrement aux ensembles infinis.

L'extension de la notion de somme d'une suite finie de nombres à une suite infinie conduit naturellement à la notion de *série* que nous réviserons dans ce chapitre. La théorie des probabilités utilise implicitement une notion plus générale, celle de *famille sommable*. Il s'agit de définir la somme, si elle existe, d'une famille de nombres indexée par un ensemble infini qui n'est pas forcément \mathbb{N} ou \mathbb{N}^* . Nous présentons cette théorie dans la dernière partie du chapitre.

Dans tout ce qui suit, la notation $\{1, \dots, n\}$ pour $n \in \mathbb{N}^*$ désigne l'ensemble de tous les entiers compris au sens large entre 1 et n . L'écriture un peu abusive « $\forall i = 1, \dots, n$ » signifie « $\forall i \in \{1, \dots, n\}$ ».

1.1 Rappels ensemblistes

1.1.1 Opérations ensemblistes

Soit Ω un ensemble; A est un *sous-ensemble* (ou une *partie*) de Ω si tout élément de A est aussi un élément de Ω ($\forall \omega \in A, \omega \in \Omega$). On note $A \subset \Omega$. On appelle $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω , ce que l'on peut noter ¹

$$\mathcal{P}(\Omega) = \{A; A \subset \Omega\}.$$

1. Dans toutes les écritures d'ensembles entre accolades, nous utilisons le point virgule au sens de « tel que ».

Ainsi les écritures $A \subset \Omega$ et $A \in \mathcal{P}(\Omega)$ sont deux façons de dire la même chose².

Si A et B sont deux parties du même ensemble Ω , on dit que A est *incluse* dans B (notation $A \subset B$) si tout élément de A est aussi élément de B ($\forall \omega \in A, \omega \in B$), autrement dit, si l'appartenance à A *implique* l'appartenance à B :

$$A \subset B \quad \text{signifie} \quad \forall \omega \in \Omega, \quad (\omega \in A) \Rightarrow (\omega \in B).$$

Soit I un ensemble quelconque d'indices (fini ou infini) et $(A_i)_{i \in I}$ une famille de parties de Ω . On définit son intersection $\bigcap_{i \in I} A_i$ et sa réunion, $\bigcup_{i \in I} A_i$ par :

$$\bigcap_{i \in I} A_i := \{\omega \in \Omega; \forall i \in I, \omega \in A_i\} \quad \text{et} \quad \bigcup_{i \in I} A_i = \{\omega \in \Omega; \exists i \in I, \omega \in A_i\}. \quad (1.1)$$

Remarque 1.1. La réunion et l'intersection d'une famille de parties de Ω sont définies de façon globale, elles s'obtiennent d'un coup, *sans passage à la limite* quand I est infini et sans qu'un ordre éventuel sur l'ensemble d'indices I n'ait d'importance.

Réunion et intersection sont très utiles pour la traduction automatique des quantificateurs. Si I est un ensemble quelconque d'indices, (π_i) une propriété dépendant de l'indice i et A_i l'ensemble des $\omega \in \Omega$ vérifiant (π_i) , on a :

$$\begin{aligned} \{\omega \in \Omega; \forall i \in I, \omega \text{ vérifie } (\pi_i)\} &= \bigcap_{i \in I} A_i, \\ \{\omega \in \Omega; \exists i = i(\omega) \in I, \omega \text{ vérifie } (\pi_i)\} &= \bigcup_{i \in I} A_i. \end{aligned}$$

Ainsi le quantificateur \forall peut se traduire par une intersection et le quantificateur \exists par une réunion.

L'intersection et l'union sont distributives l'une par rapport à l'autre, c'est à dire

$$B \cup \left(\bigcap_{i \in I} A_i \right) = \bigcap_{i \in I} (A_i \cup B) \quad B \cap \left(\bigcup_{i \in I} A_i \right) = \bigcup_{i \in I} (A_i \cap B).$$

Le *complémentaire* de A (dans Ω) est l'ensemble $A^c := \{\omega \in \Omega; \omega \notin A\}$. L'opération passage au complémentaire (qui est une bijection de $\mathcal{P}(\Omega)$ dans lui-même) vérifie $(A^c)^c = A$, $\Omega^c = \emptyset$, $\emptyset^c = \Omega$ et échange réunions et intersections grâce aux très utiles formules :

$$\left(\bigcap_{i \in I} A_i \right)^c = \bigcup_{i \in I} A_i^c \quad \left(\bigcup_{i \in I} A_i \right)^c = \bigcap_{i \in I} A_i^c.$$

On définit le *produit cartésien* de deux ensembles E et F , noté $E \times F$ par :

$$E \times F := \{(x, y); x \in E, y \in F\}.$$

Attention, dans cette écriture (x, y) ne désigne en aucune façon un ensemble mais un couple d'éléments (l'ordre d'écriture a une importance). Pour éviter toute confusion,

2. Noter cependant la différence de statut de A : dans la première écriture, A est considéré comme un ensemble, dans la deuxième comme un élément d'un ensemble d'un type un peu particulier.

on utilise des accolades pour la description des ensembles et des parenthèses pour les couples d'éléments.

On définit de manière analogue le produit cartésien d'une suite finie d'ensembles E_1, \dots, E_n par³

$$E_1 \times \dots \times E_n := \{(x_1, \dots, x_n); \forall i = 1, \dots, n, x_i \in E_i\}.$$

L'ensemble $E^2 := E \times E = \{(x_1, x_2); x_1 \in E, x_2 \in E\}$ peut être utilisé pour représenter l'ensemble de toutes les applications de $\{1, 2\}$ dans E , le couple (x_1, x_2) correspondant à l'application $f : \{1, 2\} \rightarrow E$ définie par $f(1) = x_1$ et $f(2) = x_2$. Il pourrait de la même façon, représenter les applications d'un ensemble à deux éléments dans E (remplacer les chiffres 1 et 2 par n'importe quelle paire de symboles distincts : 0 et 1, a et b , etc.).

Plus généralement, pour $n \geq 2$, E^n est l'ensemble des n -uplets ou listes de longueur n d'éléments de E . Dans un n -uplet (x_1, \dots, x_n) , il peut y avoir des répétitions. On peut aussi utiliser E^n pour représenter toutes les applications de l'ensemble $\{1, \dots, n\}$ (ou de n'importe quel ensemble à n éléments) dans E .

Soit I un ensemble quelconque, fini ou infini. Par analogie avec ce qui précède, l'ensemble de toutes les applications $f : I \rightarrow E$ sera noté E^I . Par exemple avec $E = \{0, 1\}$ et $I = \mathbb{N}$, on obtient l'ensemble $\{0, 1\}^{\mathbb{N}}$ de toutes les suites de chiffres binaires indexées par \mathbb{N} : $\{0, 1\}^{\mathbb{N}} = \{u = (u_i)_{i \in \mathbb{N}}; u_i = 0 \text{ ou } 1\}$. Avec $E = \mathbb{R}$ et $I = [0, 1]$, on obtient l'ensemble $\mathbb{R}^{[0,1]}$ des fonctions définies sur l'intervalle $[0, 1]$ et à valeurs dans \mathbb{R} .

1.1.2 Bijections

Définition 1.2 (injection). Une application $f : E \rightarrow F$ est dite injective si deux éléments distincts de E ont toujours des images distinctes dans F :

$$\forall x \in E, \forall x' \in E, \quad (x \neq x') \Rightarrow (f(x) \neq f(x')).$$

Une formulation équivalente est :

$$\forall x \in E, \forall x' \in E, \quad (f(x) = f(x')) \Rightarrow (x = x').$$

Une application injective $f : E \rightarrow F$ est appelée injection de E dans F .

Définition 1.3 (surjection). Une application $f : E \rightarrow F$ est dite surjective si tout élément de l'ensemble d'arrivée a au moins un antécédent par f :

$$\forall y \in F, \exists x \in E, \quad f(x) = y.$$

Une application surjective $f : E \rightarrow F$ est appelée surjection de E sur F .

3. Noter qu'ici le quantificateur « $\forall i$ » ne se traduit pas par une intersection. Ne confondez pas « $\forall i = 1, \dots, n, x \in E_i$ » qui traduit l'appartenance de x à l'intersection des E_i avec « $\forall i = 1, \dots, n, x_i \in E_i$ ».

Définition 1.4 (bijection). Une application $f : E \rightarrow F$ est dite bijective si elle est à la fois injective et surjective, autrement dit si tout élément de l'ensemble d'arrivée F a un unique antécédent par f dans l'ensemble de départ E :

$$\forall y \in F, \exists! x \in E, \quad f(x) = y.$$

Une application bijective $f : E \rightarrow F$ est appelée bijection de E sur F .

Remarque 1.5. Si $f : E \rightarrow F$ est une injection, en restreignant son ensemble d'arrivée à $f(E) := \{y \in F; \exists x \in E; f(x) = y\}$ la nouvelle application $f : E \rightarrow f(E)$ est une bijection. En effet cette opération préserve clairement l'injectivité et rend f surjective.

Définition 1.6 (application réciproque). Soit $f : E \rightarrow F$ une bijection. Tout $y \in F$ admet un unique antécédent x par f dans E . En posant $f^{-1}(y) := x$, on définit une application $f^{-1} : F \rightarrow E$ appelée application réciproque de f ou inverse de f . Cette application f^{-1} est bijective.

Justification. Pour vérifier l'injectivité de f^{-1} , soient y et y' deux éléments de F tels que $f^{-1}(y) = f^{-1}(y')$. Cela signifie qu'ils ont le même antécédent x par f , donc que $y = f(x)$ et $y' = f(x)$, d'où $y = y'$.

Pour la surjectivité, soit $x \in E$ quelconque. Posons $y = f(x)$. Alors x est antécédent de y par f , donc $f^{-1}(y) = x$ et ainsi y est antécédent de x par f^{-1} . Tout élément de E a donc un antécédent dans F par f^{-1} . Autrement dit, f^{-1} est surjective. \square

Remarque 1.7. Ainsi l'existence d'une bijection $E \rightarrow F$ équivaut à celle d'une bijection $F \rightarrow E$. On dira que E « est en bijection avec » F s'il existe une bijection $E \rightarrow F$ (ou $F \rightarrow E$).

Proposition 1.8. Soient $f : E \rightarrow F$ et $g : F \rightarrow G$ deux bijections. Alors $g \circ f$ est une bijection de E sur G . De plus $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Preuve. Rappelons que $(g \circ f)(x) := g(f(x))$ pour tout $x \in E$. Pour vérifier l'injectivité, soient x et x' dans E tels que $(g \circ f)(x) = (g \circ f)(x')$. Cette égalité réécrite $g(f(x)) = g(f(x'))$ implique par injectivité de g l'égalité $f(x) = f(x')$, laquelle implique $x = x'$ par injectivité de f . Pour la surjectivité de $g \circ f$, soit $z \in G$ quelconque. Par surjectivité de g , z a au moins un antécédent y dans F avec $g(y) = z$. À son tour $y \in F$ a un antécédent $x \in E$ par la surjection f . Finalement $y = f(x)$ et $z = g(y)$, d'où $z = g(f(x)) = (g \circ f)(x)$, ce qui montre que z a pour antécédent x par $g \circ f$. Comme z était quelconque, la surjectivité de $g \circ f$ est établie. Ainsi $g \circ f$ est une bijection de E sur G . En conservant les notations on a $(g \circ f)^{-1}(z) = x$. D'autre part $x = f^{-1}(y)$ et $y = g^{-1}(z)$, d'où $x = f^{-1}(g^{-1}(z)) = (f^{-1} \circ g^{-1})(z)$. On a donc pour z quelconque dans G l'égalité $(g \circ f)^{-1}(z) = x = (f^{-1} \circ g^{-1})(z)$, d'où $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. \square

1.2 Ensembles finis et dénombrement

Définition 1.9. Un ensemble E est dit fini s'il est vide ou s'il est en bijection avec un ensemble $\{1, \dots, n\}$ pour un certain entier $n \geq 1$. Un tel n est alors unique et est appelé cardinal de E (notation $\text{card } E$). Par convention le cardinal de l'ensemble vide est 0.

La cohérence de la définition 1.9 repose sur le lemme suivant (pourquoi?).

Lemme 1.10. *Si n et m sont deux entiers distincts, il n'existe pas de bijection entre $\{1, \dots, n\}$ et $\{1, \dots, m\}$.*

Preuve. On peut toujours supposer sans perte de généralité que $0 \leq n < m$. Dans le cas où $n = 0$, l'ensemble $\{1, \dots, n\}$ est l'ensemble des entiers j tels que $1 \leq j \leq 0$, c'est donc l'ensemble vide. On prouve le résultat par récurrence sur n en adoptant comme hypothèse de récurrence :

$$(\mathcal{H}_n) \quad \forall m > n, \quad \text{il n'existe pas de bijection } \{1, \dots, n\} \rightarrow \{1, \dots, m\}.$$

Initialisation. (\mathcal{H}_0) est clairement vraie, car on ne peut définir aucune application sur l'ensemble vide donc *a fortiori* aucune bijection.

Induction. Montrons que si (\mathcal{H}_n) est vérifiée pour un certain n , alors (\mathcal{H}_{n+1}) l'est aussi. Pour cela on raisonne par l'absurde : si (\mathcal{H}_{n+1}) n'était pas vérifiée, il existerait un entier $m > n+1$ et une bijection $f : \{1, \dots, n+1\} \rightarrow \{1, \dots, m\}$. Nous allons construire à partir de f une bijection $g : \{1, \dots, n+1\} \rightarrow \{1, \dots, m\}$ telle que $g(n+1) = m$. Notons $j = f(n+1)$. Si $j = m$, il suffit de prendre $g = f$. Si $j \neq m$, considérons la transposition $\tau : \{1, \dots, m\} \rightarrow \{1, \dots, m\}$ qui échange j et m et laisse les autres éléments inchangés. C'est une bijection et l'application composée $g = \tau \circ f$ est une bijection $\{1, \dots, n+1\} \rightarrow \{1, \dots, m\}$ vérifiant $g(n+1) = m$.

La restriction \tilde{g} de g à $\{1, \dots, n\}$ est une bijection de $\{1, \dots, n\}$ sur $\{1, \dots, m-1\}$ et comme $m > n+1$, on a bien $m-1 > n$, ce qui contredit (\mathcal{H}_n) . Nous venons d'établir l'implication $(\mathcal{H}_n) \Rightarrow (\mathcal{H}_{n+1})$, ce qui achève la récurrence. \square

Remarque 1.11. Si l'ensemble F est en bijection avec un ensemble fini E , alors F est fini et a même cardinal que E . En effet en notant $n = \text{card } E$, il existe une bijection $f : \{1, \dots, n\} \rightarrow E$ et une bijection $g : E \rightarrow F$. La composée $g \circ f$ réalise alors une bijection de $\{1, \dots, n\}$ sur F .

Proposition 1.12. *Soient E et F deux ensembles finis.*

$$\text{Si } E \cap F = \emptyset, \quad \text{card}(E \cup F) = \text{card } E + \text{card } F. \quad (1.2)$$

Preuve. Dans le cas où l'un des deux ensembles est vide, (1.2) est triviale. On suppose désormais que $\text{card } E = n \geq 1$ et $\text{card } F = m \geq 1$. Il existe alors des bijections

$$f : \{1, \dots, n\} \rightarrow E, \quad g : \{1, \dots, m\} \rightarrow F.$$

On prouve (1.2) en construisant une bijection h de $\{1, \dots, n+m\}$ sur $E \cup F$. La translation

$$t : \{n+1, \dots, n+m\} \rightarrow \{1, \dots, m\}, \quad i \mapsto i - n$$

est une bijection et l'application $g \circ t$ réalise une bijection de $\{n+1, \dots, n+m\}$ sur F . Définissons alors h par

$$h(i) := \begin{cases} f(i) & \text{si } i \in \{1, \dots, n\}, \\ (g \circ t)(i) & \text{si } i \in \{n+1, \dots, n+m\}. \end{cases}$$

Pour vérifier la surjectivité de h , soit z un élément quelconque de $E \cup F$. Si $z \in E$, alors il a un antécédent i dans $\{1, \dots, n\}$ par la surjection f et comme $h(i) = f(i) = z$, i est aussi antécédent de z par h . Si $z \in F$, il a un antécédent j dans $\{1, \dots, m\}$ par la surjection g et j a un antécédent i dans $\{n+1, \dots, n+m\}$ par la surjection t . Alors $h(i) = g(t(i)) = g(j) = z$ donc i est antécédent de z par h .

Pour vérifier l'injectivité de h , notons i et i' deux éléments distincts de $\{1, \dots, n+m\}$. S'ils sont l'un dans $\{1, \dots, n\}$ et l'autre dans $\{n+1, \dots, n+m\}$, leurs images $h(i)$ et $h(i')$ sont l'une dans E et l'autre dans F qui sont *disjoints* ($E \cap F = \emptyset$) donc $h(i) \neq h(i')$. sinon i et i' sont tous deux dans $\{1, \dots, n\}$ (resp. dans $\{n+1, \dots, n+m\}$) et $h(i) \neq h(i')$ en raison de l'injectivité de f (resp. de $g \circ t$). \square

Corollaire 1.13 (de la proposition 1.12).

- a) Si E_1, \dots, E_d sont d ensembles finis deux à deux disjoints, $E_1 \cup \dots \cup E_d$ est un ensemble fini et

$$\text{card} \left(\bigcup_{i=1}^d E_i \right) = \sum_{i=1}^d \text{card } E_i.$$

- b) Si E et F sont des ensembles finis quelconques (pas forcément disjoints), $E \cup F$ est fini et

$$\text{card}(E \cup F) = \text{card } E + \text{card } F - \text{card}(E \cap F).$$

Preuve. Laissée en exercice. \square

Proposition 1.14. Soient E et F deux ensembles finis. Leur produit cartésien a pour cardinal

$$\text{card}(E \times F) = \text{card } E \times \text{card } F. \quad (1.3)$$

Preuve. Le cas où l'un des deux ensembles E ou F est vide étant trivial, on suppose désormais qu'aucun des deux n'est vide. On fait une récurrence sur $n = \text{card } E$, en prenant pour hypothèse de récurrence :

$$(\mathcal{H}_n) \quad \text{si } \text{card } E = n, \text{ alors } \text{card}(E \times F) = n \text{ card } F \text{ pour tout } F \text{ fini non vide.}$$

Initialisation. Si $\text{card } E = 1$, E n'a qu'un élément x_1 et l'application $h : \{x_1\} \times F \rightarrow F$, $(x_1, y) \mapsto y$ est clairement une bijection donc $\text{card}(\{x_1\} \times F) = \text{card } F$ et (\mathcal{H}_1) est vérifiée.

Induction. Supposons (\mathcal{H}_n) vraie pour un certain n et soit E un ensemble de cardinal $n+1$. Il existe une bijection $f : \{1, \dots, n+1\} \rightarrow E$ permettant de numérotter les éléments de E en posant pour tout $i \in \{1, \dots, n+1\}$, $x_i := f(i)$. La restriction de f à $\{1, \dots, n\}$ est une bijection de $\{1, \dots, n\}$ sur son image $E' = \{x_1, \dots, x_n\}$. Ainsi E' est de cardinal n et E est l'union de ses deux sous-ensembles disjoints E' et $\{x_{n+1}\}$. On en déduit immédiatement que $E \times F$ est l'union des deux produits cartésiens disjoints $E' \times F$ et $\{x_{n+1}\} \times F$. Par la proposition 1.12 on a alors

$$\text{card}(E \times F) = \text{card}(E' \times F) + \text{card}(\{x_{n+1}\} \times F).$$

En utilisant (\mathcal{H}_n) et (\mathcal{H}_1) , on obtient alors

$$\text{card}(E \times F) = n \text{card } F + \text{card } F = (n + 1) \text{card } F,$$

donc (\mathcal{H}_{n+1}) est vérifiée. \square

Remarque 1.15. On aurait pu aussi prouver (1.3) en construisant explicitement une bijection $E \times F \rightarrow \{1, \dots, nm\}$ (avec $\text{card } E = n$ et $\text{card } F = m$). Voici une façon de la construire. On note $f : \{1, \dots, n\} \rightarrow E$, $i \mapsto x_i := f(i)$ et $g : \{1, \dots, m\} \rightarrow F$, $j \mapsto y_j := g(j)$ des numérotations bijectives de E et F . On définit alors $h : E \times F \rightarrow \{1, \dots, nm\}$ en posant $h(x_i, y_j) := m(i - 1) + j$, ou de manière plus formelle $h(x, y) := m(f^{-1}(x) - 1) + g^{-1}(y)$ pour tout $(x, y) \in E \times F$. On laisse en exercice la vérification de la bijectivité de h . L'idée de sa construction est simplement de ranger les couples éléments de $E \times F$ sous la forme d'un tableau où le couple (x_i, y_j) se trouve à l'intersection de la ligne i et de la colonne j et de numéroté les éléments de ce tableau en balayant chaque ligne de gauche à droite de la ligne 1 jusqu'à la ligne n .

Corollaire 1.16 (de la proposition 1.14). *Si E_1, \dots, E_d sont d ensembles finis,*

$$\text{card}(E_1 \times \dots \times E_d) = \prod_{i=1}^d \text{card } E_i.$$

Preuve. Une récurrence immédiate sur d fournit le résultat. \square

Proposition 1.17 (nombre d'applications $E \rightarrow F$ et cardinal de $\mathcal{P}(E)$).

(a) *Si $\text{card } E = n$ et $\text{card } F = p$, l'ensemble F^E des applications de E dans F est fini et a pour cardinal p^n , autrement dit :*

$$\text{card}(F^E) = (\text{card } F)^{\text{card } E}.$$

(b) *Comme $\mathcal{P}(E)$ est en bijection avec l'ensemble $\{0, 1\}^E$ des applications de E dans $\{0, 1\}$,*

$$\text{card } \mathcal{P}(E) = 2^n = 2^{\text{card } E}.$$

Preuve. Le (a) se démontre facilement par récurrence sur le cardinal de E , en notant que si on ajoute un élément x_{n+1} à E , il y a p façons différentes de prolonger $f : E \rightarrow F$ en attribuant comme image à x_{n+1} l'un des éléments de F . La rédaction détaillée est laissée en exercice.

Une bijection naturelle entre $\mathcal{P}(E)$ et $\{0, 1\}^E$ est l'application φ qui à toute partie A de E associe sa fonction *indicatrice* :

$$\varphi : \mathcal{P}(E) \rightarrow \{0, 1\}^E \quad A \mapsto \varphi(A) := \mathbf{1}_A.$$

Rappelons que l'indicatrice d'une partie A de E est l'application

$$\mathbf{1}_A : E \rightarrow \{0, 1\} \quad \omega \mapsto \mathbf{1}_A(\omega) := \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{si } \omega \notin A. \end{cases}$$

La vérification de la bijectivité de φ est laissée en exercice⁴. \square

4. Ni la définition de φ , ni la preuve de sa bijectivité n'utilisent la finitude de E . Ainsi $\mathcal{P}(E)$ et $\{0, 1\}^E$ sont en bijection quel que soit l'ensemble $E \neq \emptyset$, fini ou infini.

Définition 1.18 (arrangement). Si E est un ensemble de cardinal n et k un entier tel que $1 \leq k \leq n$, on appelle arrangement de k éléments de E tout k -uplet (x_1, x_2, \dots, x_k) d'éléments tous distincts de E . Un tel arrangement représente une injection de $\{1, \dots, k\}$ dans E .

Proposition 1.19 (dénombrement des arrangements). Le nombre d'arrangements de k éléments de E ($1 \leq k \leq n = \text{card } E$) est

$$A_n^k = n(n-1)(n-2) \cdots (n-k+1) = \frac{n!}{(n-k)!}. \quad (1.4)$$

A_n^k est aussi le nombre d'injections d'un ensemble I de cardinal k , par exemple $\{1, \dots, k\}$, dans E . En particulier pour $I = E$ (et donc $k = n$), on obtient le nombre de bijections de E dans lui même (appelées aussi permutations de E) :

$$\text{nombre de permutations de } E = A_n^n = n!$$

Preuve. On prouve (1.4) par récurrence finie sur k , le cas $k = 1$ étant évident. Supposons donc (1.4) vraie pour un $k < n$ et montrons qu'alors elle est aussi vraie pour $k + 1$. Une application $f : \{1, \dots, k+1\} \rightarrow E$ est déterminée de manière unique par la donnée de sa restriction \tilde{f} à $\{1, \dots, k\}$ et de $f(k+1)$. L'application f est injective si et seulement si sa restriction \tilde{f} est injective et $f(k+1) \notin \tilde{f}(\{1, \dots, k\})$. Comme le cardinal de $\tilde{f}(\{1, \dots, k\})$ est k , cela laisse $n - k$ choix possibles pour $f(k+1)$. On en déduit que

$$A_n^{k+1} = A_n^k(n-k) = n(n-1)(n-2) \cdots (n-k+1)(n-k),$$

ce qui montre que (1.4) est vérifiée au rang $k + 1$. □

Définition 1.20 (combinaison). On appelle combinaison de k éléments de E ($1 \leq k \leq n = \text{card } E$) toute partie de cardinal k de E . Une combinaison a tous ses éléments distincts comme un arrangement, mais l'ordre d'écriture n'a pas d'importance.

Proposition 1.21 (dénombrement des combinaisons). Le nombre de combinaisons de k éléments de E ($1 \leq k \leq n = \text{card } E$) est

$$C_n^k = \frac{n(n-1)(n-2) \cdots (n-k+1)}{k(k-1) \cdots 1} = \frac{n!}{k!(n-k)!}. \quad (1.5)$$

Preuve. Notons $\mathcal{A}_k(E)$ l'ensemble de tous les arrangements de k éléments de E . Il est alors clair que l'on a la décomposition en réunion d'ensembles disjoints

$$\mathcal{A}_k(E) = \bigcup_{B \subset E, \text{card } B=k} \mathcal{A}_k(B). \quad (1.6)$$

Autrement dit on a partitionné $\mathcal{A}_k(E)$ en regroupant dans une même classe $\mathcal{A}_k(B)$ tous les arrangements formés à partir des éléments d'une même partie B de cardinal k . Il

y a donc autant de classes distinctes dans cette décomposition que de parties B de cardinal k dans E , c'est-à-dire C_n^k classes. D'autre part chaque classe $\mathcal{A}_k(B)$ contient autant d'arrangements que de bijections $B \rightarrow B$ (ou permutations sur B), c'est-à-dire $k!$. Compte-tenu du corollaire 1.13 a), on déduit alors de (1.6) que :

$$\text{card } \mathcal{A}_k(E) = \sum_{B \subset E, \text{ card } B=k} \text{card } \mathcal{A}_k(B) = C_n^k k!.$$

Or $\text{card } \mathcal{A}_k(E) = A_n^k$ par la définition des A_n^k , d'où $A_n^k = C_n^k k!$, ce qui donne (1.5). \square

Les nombres C_n^k sont appelés aussi *coefficients binomiaux* en raison de leur rôle dans la formule du binôme de Newton, que nous retrouverons ci-dessous (corollaire 1.24) comme cas particulier de la formule du multinôme. Avant de nous y attaquer, il n'est peut-être pas inutile de faire un rappel sur le développement d'un produit de sommes finies.

Proposition 1.22. *Soient J_1, \dots, J_n des ensembles finis non vides d'indices et pour $i = 1, \dots, n$, $j \in J_i$, des nombres réels ou complexes $x_{i,j}$. En notant $K_n := J_1 \times \dots \times J_n$, on a*

$$\prod_{i=1}^n \left(\sum_{j \in J_i} x_{i,j} \right) = \sum_{(j_1, \dots, j_n) \in K_n} \left(\prod_{i=1}^n x_{i,j_i} \right) = \sum_{j_1 \in J_1} \dots \sum_{j_n \in J_n} \left(\prod_{i=1}^n x_{i,j_i} \right). \quad (1.7)$$

Voici une traduction sans formules de cet énoncé : « le produit de n sommes finies est égal à la somme de tous les produits de n facteurs que l'on peut former en sélectionnant un facteur parmi les termes de chacune des n sommes ». La figure 1.1 illustre l'application de cette règle au développement de $(a + b + c)(d + e)(s + t + u + v)$. Ici $J_1 = \{1, 2, 3\}$, $x_{1,1} = a, \dots, x_{1,3} = c$, $J_2 = \{1, 2\}$, $x_{2,1} = d, x_{2,2} = e$, $J_3 = \{1, \dots, 4\}$, $x_{3,1} = s, \dots, x_{3,4} = v$. La proposition 1.22 se démontre facilement par récurrence sur n .

Proposition 1.23 (formule du multinôme). *Pour tous entiers $d \geq 2$, $n \geq 2$ et tous nombres réels ou complexes a_1, \dots, a_d ,*

$$(a_1 + \dots + a_d)^n = \sum_{\substack{(k_1, \dots, k_d) \in \mathbb{N}^d \\ k_1 + \dots + k_d = n}} \frac{n!}{k_1! \dots k_d!} a_1^{k_1} \dots a_d^{k_d}. \quad (1.8)$$

Preuve. On commence par appliquer la formule (1.7) avec $J_1 = \dots = J_n = \{1, \dots, d\}$ pour obtenir :

$$(a_1 + \dots + a_d)^n = \underbrace{(a_1 + \dots + a_d) \times \dots \times (a_1 + \dots + a_d)}_{n \text{ parenthèses}} = \sum_{(i_1, \dots, i_n) \in \{1, \dots, d\}^n} a_{i_1} \dots a_{i_n}. \quad (1.9)$$

Parmi les n facteurs du produit $a_{i_1} \dots a_{i_n}$, il peut y avoir des répétitions et il peut aussi manquer certains des a_i . En regroupant les facteurs identiques, on peut toujours écrire ce produit sous la forme $a_1^{k_1} \dots a_d^{k_d}$, où pour $j = 1, \dots, d$, on a noté k_j le nombre de facteurs égaux à a_j dans ce produit (éventuellement $k_j = 0$ si a_j ne figure pas dans le

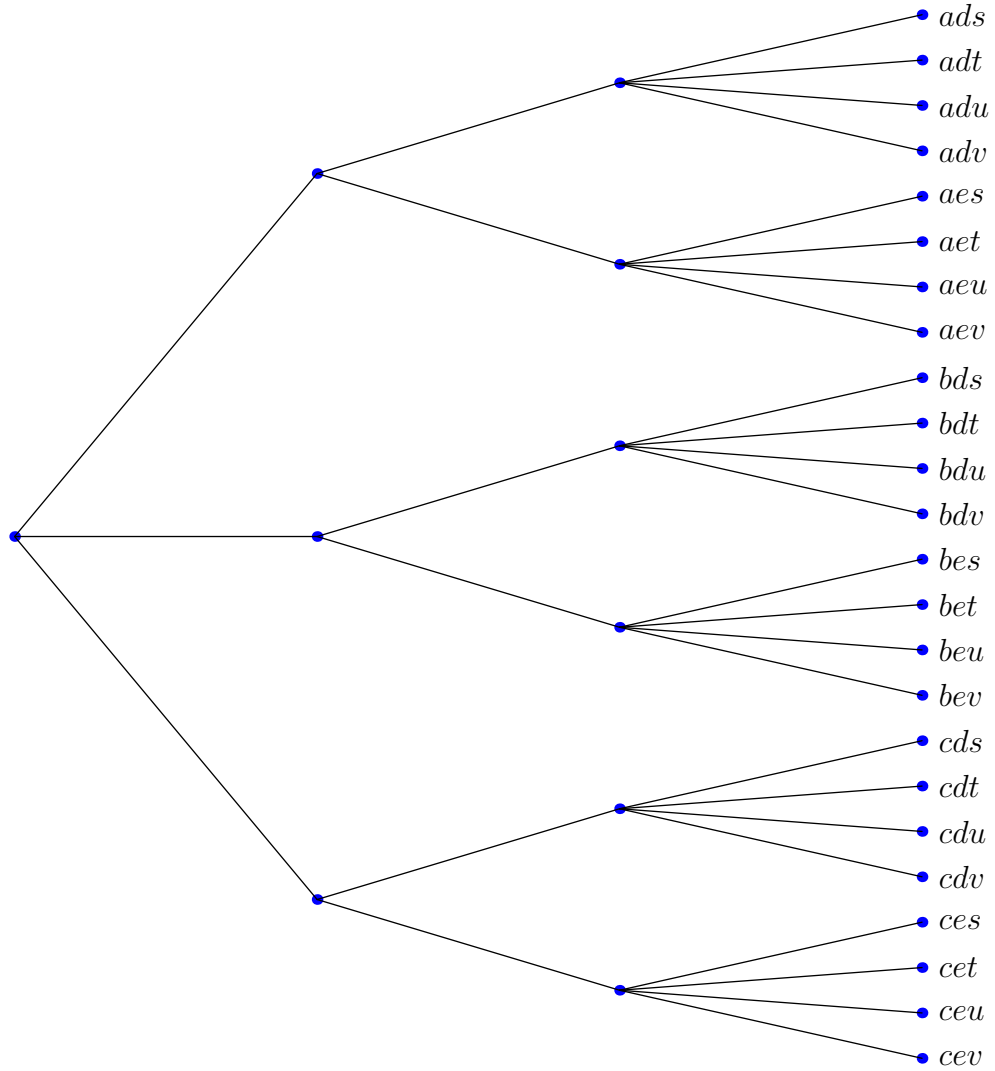


FIG. 1.1 – Arbre de développement de $(a + b + c)(d + e)(s + t + u + v)$

produit). Comme il y a n facteurs, les k_j vérifient $k_1 + \dots + k_d = n$. Notons maintenant $M(n, k_1, \dots, k_d)$, le nombre d'apparitions du produit $a_1^{k_1} \dots a_d^{k_d}$ dans le développement de $(a_1 + \dots + a_d)^n$. Avec cette notation, on peut réécrire (1.9) sous la forme

$$(a_1 + \dots + a_d)^n = \sum_{\substack{(k_1, \dots, k_d) \in \mathbb{N}^d \\ k_1 + \dots + k_d = n}} M(n, k_1, \dots, k_d) a_1^{k_1} \dots a_d^{k_d}. \quad (1.10)$$

Il ne nous reste plus qu'à faire un peu de dénombrement pour expliciter le coefficient $M(n, k_1, \dots, k_d)$ pour k_1, \dots, k_d entiers fixés tels que $k_1 + \dots + k_d = n$. Cela revient à compter de combien de façons on peut choisir les k_1 parenthèses fournissant a_1 , les k_2 parenthèses fournissant a_2 , ..., les k_d parenthèses fournissant a_d , pour former le produit $a_1^{k_1} \dots a_d^{k_d}$. Commençons par choisir les k_1 parenthèses dont la contribution est a_1 , ceci peut se faire de $C_n^{k_1}$ façons. Parmi les $n - k_1$ parenthèses restantes, on choisit ensuite les

k_2 parenthèses qui fournissent a_2 , ce qui nous laisse $C_{n-k_1}^{k_2}$ possibilités. Et ainsi de suite, jusqu'au choix de k_d parenthèses parmi les $n - k_1 - \dots - k_{d-1}$ dernières⁵ pour obtenir a_d . On en déduit que

$$\begin{aligned} M(n, k_1, \dots, k_d) &= C_n^{k_1} \times C_{n-k_1}^{k_2} \times C_{n-k_1-k_2}^{k_3} \times \dots \times C_{n-k_1-\dots-k_{d-1}}^{k_d} \\ &= \frac{n!}{k_1!(n-k_1)!} \times \frac{(n-k_1)!}{k_2!(n-k_1-k_2)!} \times \frac{(n-k_1-k_2)!}{k_3!(n-k_1-k_2-k_3)!} \times \dots \\ &\quad \dots \times \frac{(n-k_1-\dots-k_{d-1})!}{k_d!(n-k_1-\dots-k_d)!} \\ &= \frac{n!}{k_1! \dots k_d!}, \end{aligned}$$

après simplifications et en notant que $(n - k_1 - \dots - k_d)! = 0! = 1$. \square

Les nombres $M(n, k_1, \dots, k_d)$ sont appelés *coefficients multinomiaux*. On peut aussi les interpréter comme

- a) le nombre de façons de répartir les éléments d'un ensemble de cardinal n en d sous-ensembles de cardinal respectif k_1, \dots, k_d ;
- b) le nombre d'applications f de $\{1, \dots, n\}$ dans $\{1, \dots, d\}$ telles que chaque i dans $\{1, \dots, d\}$ ait exactement k_i antécédents dans $\{1, \dots, n\}$ par f .

Notons en passant que d'après l'interprétation b), le nombre total d'applications de $\{1, \dots, n\}$ dans $\{1, \dots, d\}$, soit d^n d'après la proposition 1.17 a), doit être égal à la somme de tous les $M(n, k_1, \dots, k_d)$, ce qui s'écrit :

$$d^n = \sum_{\substack{(k_1, \dots, k_d) \in \mathbb{N}^d \\ k_1 + \dots + k_d = n}} M(n, k_1, \dots, k_d).$$

C'est bien ce que donne la formule du multinôme en prenant tous les a_i égaux à 1 dans (1.8).

Corollaire 1.24 (formule du binôme). *Pour tous réels ou complexes a et b et tout entier $n \geq 2$,*

$$(a + b)^n = \sum_{\substack{(k,l) \in \mathbb{N}^2 \\ k+l=n}} \frac{n!}{k!l!} a^k b^l = \sum_{k=0}^n C_n^k a^k b^{n-k}. \quad (1.11)$$

Preuve. Il suffit d'appliquer la formule du multinôme avec $d = 2$, $a_1 = a$, $a_2 = b$. \square

Remarque 1.25. Dans le calcul des $M(n, k_1, \dots, k_d)$ ci-dessus, il peut sembler que l'on ait avantagé les a_1 en commençant par les choisir en premier et pénalisé les a_d en les gardant pour la fin. Vous pourrez vous convaincre qu'il n'en est rien en refaisant ce calcul en plaçant les a_i dans l'ordre qui vous convient, par exemple en commençant par a_d , ou par a_3, \dots

5. Comme $n - k_1 - \dots - k_{d-1} = k_d$, il n'y a qu'un seul choix possible à ce stade : prendre toutes les parenthèses restantes.

Voici d'ailleurs une suggestion pour une preuve alternative qui pourrait être développée en exercice. On considère l'ensemble \mathcal{S}_n de toutes les permutations de $\{1, \dots, n\}$. Son cardinal est $n!$. Fixons k_1, \dots, k_d entiers de somme n . On pose $E_i = \{1 + k_{i-1}, \dots, k_i\}$, avec $k_0 := 0$ et $i = 1, \dots, d$. Si (F_1, \dots, F_d) est un d -uplet de parties de $\{1, \dots, n\}$, deux à deux disjointes et de réunion $\{1, \dots, n\}$, combien y a-t-il de bijections $f \in \mathcal{S}_n$ vérifiant $f(E_i) = F_i$ pour tout $i = 1, \dots, d$? En déduire que $n! = M(n, k_1, \dots, k_d)k_1! \dots k_d!$.

1.3 Dénombrabilité

On peut comparer les ensembles finis par leur nombre d'éléments. Cette notion n'a plus de sens pour des ensembles infinis. Néanmoins on peut généraliser cette comparaison en disant que deux ensembles ont même cardinal s'ils sont en bijection. Ceci permet de comparer les ensembles infinis. Il convient de se méfier de l'intuition courante basée sur les ensembles finis. Par exemple si A et B sont finis et A est inclus strictement dans B , alors $\text{card } A < \text{card } B$ et il n'existe pas de bijection entre A et B . Ceci n'est plus vrai pour les ensembles infinis. Par exemple \mathbb{N}^* est strictement inclus dans \mathbb{N} mais est en bijection avec \mathbb{N} par l'application $f : \mathbb{N} \rightarrow \mathbb{N}^*$, $n \mapsto n + 1$, donc \mathbb{N}^* et \mathbb{N} ont même cardinal. Nous nous intéressons maintenant aux ensembles ayant même cardinal que \mathbb{N} .

Définition 1.26. *Un ensemble est dit dénombrable s'il est en bijection avec \mathbb{N} . Il est dit au plus dénombrable s'il est fini ou dénombrable.*

Exemple 1.27. L'ensemble $2\mathbb{N}$ des entiers pairs est dénombrable. Pour le voir, il suffit de considérer l'application $f : \mathbb{N} \rightarrow 2\mathbb{N}$, $n \mapsto 2n$ qui est clairement une bijection. On vérifie de même que l'ensemble des entiers impairs est dénombrable.

Exemple 1.28. L'ensemble \mathbb{Z} est dénombrable. On peut en effet « numéroté » les entiers relatifs par les entiers naturels en s'inspirant du tableau suivant.

\mathbb{Z}	...	-4	-3	-2	-1	0	+1	+2	+3	+4	...
\mathbb{N}	...	8	6	4	2	0	1	3	5	7	...

Plus formellement, définissons l'application $f : \mathbb{N} \rightarrow \mathbb{Z}$ par

$$\forall n \in \mathbb{N}, \quad f(n) = \begin{cases} \frac{n+1}{2} & \text{si } n \text{ est impair,} \\ \frac{-n}{2} & \text{si } n \text{ est pair.} \end{cases}$$

Pour vérifier que f est une bijection, montrons que pour tout $k \in \mathbb{Z}$ donné, l'équation $f(n) = k$ a une solution unique. Si $k > 0$, il ne peut avoir d'antécédent pair par f puisque f envoie l'ensemble des entiers pairs dans \mathbb{Z}^- . L'équation $f(n) = k$ se réduit donc dans ce cas à $\frac{n+1}{2} = k$ qui a pour unique solution $n = 2k + 1$. Si $k \leq 0$, il ne peut avoir d'antécédent impair par f et l'équation $f(n) = k$ se réduit à $\frac{-n}{2} = k$ qui a pour unique solution $n = -2k$. Ainsi f est bien une bijection puisque tout élément k de \mathbb{Z} a un antécédent n unique par f . La bijection inverse est donnée par

$$\forall k \in \mathbb{Z}, \quad f^{-1}(k) = \begin{cases} 2k + 1 & \text{si } k > 0, \\ -2k & \text{si } k \leq 0. \end{cases}$$

Exemple 1.29. L'ensemble \mathbb{N}^2 est dénombrable. Cet exemple relève de la proposition 1.34 ci-dessous, mais la vérification directe est instructive. Voici une façon de construire une bijection $f : \mathbb{N}^2 \rightarrow \mathbb{N}$. L'idée est de fabriquer une numérotation des couples de \mathbb{N}^2 par les entiers en s'inspirant du schéma de la figure 1.29. Un peu de

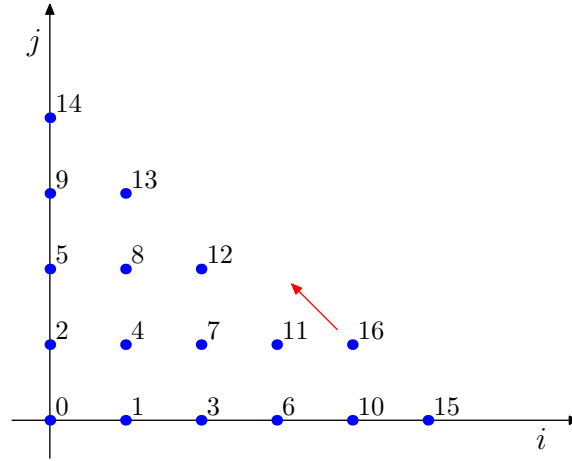


FIG. 1.2 – Une numérotation des couples (i, j) de \mathbb{N}^2

dénombrement nous conduit à proposer la définition⁶

$$\forall (i, j) \in \mathbb{N}^2, \quad f(i, j) := \frac{(i+j)(i+j+1)}{2} + j.$$

Preuve. La vérification de la bijectivité de f repose sur la remarque suivante. Définissons la suite $(u_k)_{k \in \mathbb{N}}$ par

$$\forall k \in \mathbb{N}, \quad u_k := \frac{k(k+1)}{2}.$$

Il s'agit clairement d'une suite *strictement croissante d'entiers*, de premier terme $u_0 = 0$. On a donc

$$\forall l \in \mathbb{N}, \quad \exists ! k = k_l \in \mathbb{N}, \quad u_k \leq l < u_{k+1}. \quad (1.12)$$

De plus,

$$\text{si } l = f(i, j), \quad k_l = i + j \text{ et } l = u_{k_l} + j. \quad (1.13)$$

Surjectivité de f . Soit l quelconque dans \mathbb{N} et $k = k_l$ défini par (1.12). Posons $j = l - u_k$ et $i = k - j$. Alors

$$f(i, j) = \frac{(i+j)(i+j+1)}{2} + j = \frac{k(k+1)}{2} + j = u_k + l - u_k = l.$$

Le couple (i, j) ainsi défini à partir de l est donc antécédent de l par f et comme l était quelconque, f est surjective.

6. Justification laissée en exercice.

Injectivité de f . Soient (i, j) et (i', j') tels que $f(i, j) = f(i', j')$ et notons l cette valeur commune. D'après (1.13), on a $k_l = i + j = i' + j'$ et $l = u_{k_l} + j = u_{k_l} + j'$. On en déduit immédiatement que $j = j'$, puis que $i = i'$, d'où $(i, j) = (i', j')$, ce qui établit l'injectivité de f . \square

Lemme 1.30. *Toute partie infinie de \mathbb{N} est dénombrable.*

Preuve. Soit E une partie infinie de \mathbb{N} . On construit une bijection f de \mathbb{N} sur E par récurrence en utilisant le fait que toute partie non vide de \mathbb{N} admet un plus petit élément. On initialise la récurrence en posant :

$$E_0 := E, \quad f(0) := \min E_0.$$

Ensuite, pour $n \geq 1$, si on a défini $f(k)$ et E_k pour $k = 0, \dots, n-1$, on pose

$$E_n := E \setminus f(\{0, \dots, n-1\}), \quad f(n) := \min E_n.$$

L'ensemble $f(\{0, \dots, n-1\}) = \{f(0), \dots, f(n-1)\}$ est fini donc E_n est non vide puisque E est infini. On peut donc bien construire ainsi de proche en proche tous les $f(n)$ pour $n \in \mathbb{N}$. De plus il est clair par construction que pour tout $n \geq 1$, $f(n-1) < f(n)$. L'application f est donc strictement croissante, ce qui entraîne son injectivité. Pour voir qu'elle est surjective, soit m un élément quelconque de E . Comme m est un entier, il n'y a qu'un nombre fini d'entiers strictement inférieurs à m , donc *a fortiori* qu'un nombre fini n d'éléments de E inférieurs strictement à m (éventuellement aucun). Ainsi m est le $(n+1)$ -ième plus petit élément de E , d'où $f(n) = m$ (comme on commence à 0, n est le $(n+1)$ -ième plus petit entier de \mathbb{N}). Nous venons de montrer qu'un élément quelconque de E a au moins un antécédent par f , autrement dit que f est surjective. \square

Proposition 1.31. *Toute partie infinie d'un ensemble dénombrable est elle-même dénombrable.*

Preuve. Soit A une partie infinie d'un ensemble dénombrable B . Il existe alors une bijection $g : B \rightarrow \mathbb{N}$. Sa restriction \tilde{g} à A est une bijection de A sur $g(A)$. L'ensemble $g(A)$ est une partie infinie de \mathbb{N} , car si elle était finie, il en serait de même pour A . Par le lemme 1.30, il existe une bijection f de $g(A)$ sur \mathbb{N} . L'application $f \circ \tilde{g} : A \rightarrow \mathbb{N}$ est une bijection comme composée de deux bijections. L'ensemble A est donc dénombrable. \square

Remarque 1.32. La proposition 1.31 nous permet de caractériser les ensembles *au plus* dénombrables comme ceux qui *sont en bijection avec une partie de \mathbb{N}* , ou encore comme ceux qui *s'injectent dans \mathbb{N}* . De même, les ensembles dénombrables sont les ensembles *infinis* qui s'injectent dans \mathbb{N} .

Remarque 1.33. Il résulte immédiatement de la proposition 1.31 que si l'ensemble B contient une partie infinie A non dénombrable, B est lui même infini non dénombrable.

Proposition 1.34. *Le produit cartésien d'une suite finie d'ensembles dénombrables est dénombrable.*

Preuve. Notons E_1, \dots, E_n la suite finie d'ensembles dénombrables considérée. Pour $i = 1, \dots, n$, nous disposons d'une bijection $E_i \rightarrow \mathbb{N}$, qui par composition avec la bijection $\mathbb{N} \rightarrow \mathbb{N}^*$, $n \mapsto n + 1$ donne une bijection $f_i : E_i \rightarrow \mathbb{N}^*$. Comme $E = E_1 \times \dots \times E_n$ est clairement un ensemble infini, il nous suffit de construire une injection de E dans \mathbb{N} . Pour cela il est commode d'utiliser les nombres premiers dont nous notons $(p_j)_{j \geq 1}$ la suite ordonnée : $p_1 = 2, p_2 = 3, p_3 = 5, p_4 = 7, p_5 = 11, \dots$

Définissons $f : E \rightarrow \mathbb{N}$, par

$$\forall x = (x_1, \dots, x_n) \in E, \quad f(x) := p_1^{f_1(x_1)} \dots p_n^{f_n(x_n)} = \prod_{j=1}^n p_j^{f_j(x_j)}.$$

Remarquons que pour tout $x \in E$, $f(x) \geq 2^{f_1(x_1)} \geq 2$.

Pour vérifier l'injectivité de f , soient x et y dans E tels que $f(x) = f(y)$. En vertu de l'unicité de la décomposition en facteurs premiers d'un entier supérieur ou égal à 2, cette égalité équivaut à :

$$\forall i = 1, \dots, n, \quad f_i(x_i) = f_i(y_i).$$

Comme chaque f_i est injective, ceci entraîne l'égalité $x_i = y_i$ pour tout i , d'où $x = y$. \square

Corollaire 1.35. *Pour tout entier $d \geq 1$, $\mathbb{N}^d, \mathbb{Z}^d$ sont dénombrables. L'ensemble \mathbb{Q} des nombres rationnels est dénombrable (de même que \mathbb{Q}^d , $d \geq 1$).*

Vérification. La dénombrabilité de \mathbb{Q} s'obtient facilement en l'injectant dans le produit cartésien d'ensembles dénombrables $\mathbb{Z} \times \mathbb{N}^*$ via l'unicité de l'écriture en fraction irréductible (avec dénominateur positif) d'un rationnel. Les autres affirmations du corollaire découlent immédiatement de la Proposition 1.34. \square

Voici un premier exemple d'ensemble infini non dénombrable.

Proposition 1.36. *L'ensemble $\{0, 1\}^{\mathbb{N}}$ des suites infinies de 0 ou de 1 est infini mais n'est pas dénombrable.*

Il est clair que $\{0, 1\}^{\mathbb{N}}$ est un ensemble infini, puisque qu'il contient un sous-ensemble en bijection avec \mathbb{N} , par exemple l'ensemble de suites $\{(\mathbf{1}_{\{n\}}(k))_{k \in \mathbb{N}}, n \in \mathbb{N}\}$.

Preuve. Supposons que $\{0, 1\}^{\mathbb{N}}$ soit dénombrable, on peut alors numérotter ses éléments par les entiers, de sorte que $\{0, 1\}^{\mathbb{N}} = \{x_n; n \in \mathbb{N}\}$, chaque x_n étant une suite $(x_{n,k})_{k \in \mathbb{N}}$ de chiffres binaires. Construisons alors la suite $y = (y_k)_{k \in \mathbb{N}}$ de chiffres binaires en posant :

$$\forall k \in \mathbb{N}, \quad y_k = 1 - x_{k,k} = \begin{cases} 1 & \text{si } x_{k,k} = 0 \\ 0 & \text{si } x_{k,k} = 1. \end{cases}$$

Alors par construction, la suite y diffère de chaque suite x_n (au moins par son n^{e} terme). Or y est un élément de $\{0, 1\}^{\mathbb{N}}$, donc la numérotation considérée ne peut être surjective. \square

Corollaire 1.37. $\mathcal{P}(\mathbb{N})$ n'est pas dénombrable. Le segment $[0, 1]$ de \mathbb{R} n'est pas dénombrable. \mathbb{R} n'est pas dénombrable, \mathbb{C} n'est pas dénombrable. Un intervalle de \mathbb{R} est soit infini non dénombrable, soit réduit à un singleton, soit vide.

Preuve. Comme $\mathcal{P}(\mathbb{N})$ est en bijection avec $\{0, 1\}^{\mathbb{N}}$ par l'application $A \mapsto \mathbf{1}_A$, $\mathcal{P}(\mathbb{N})$ est infini non dénombrable. D'après la remarque 1.33, la non dénombrabilité de \mathbb{R} ou de \mathbb{C} résulte immédiatement de celle de $[0, 1]$. Pour vérifier cette dernière, nous utilisons à nouveau la remarque 1.33 en construisant une partie de $[0, 1]$ en bijection avec $\{0, 1\}^{\mathbb{N}}$. La première idée qui vient à l'esprit pour une telle construction est d'utiliser le développement des nombres réels en base 2. Mais pour éviter les difficultés techniques liées à l'existence de développements propre et impropre pour les nombres de la forme $k2^{-n}$, nous utiliserons plutôt la base 3 en ne conservant que les chiffres binaires 0 et 1. Définissons donc ⁷

$$f : \{0, 1\}^{\mathbb{N}} \rightarrow [0, 1], \quad u = (u_k)_{k \in \mathbb{N}} \mapsto f(u) := \sum_{k=0}^{+\infty} \frac{u_k}{3^{k+1}}.$$

Comme les u_k ne peuvent prendre que les valeurs 0 ou 1, la série à termes positifs définissant $f(u)$ converge puisque son terme général vérifie l'encadrement $0 \leq u_k 3^{-k-1} \leq 3^{-k-1}$. Sa somme $f(u)$ vérifie donc

$$0 \leq f(u) \leq \sum_{k=0}^{+\infty} \frac{1}{3^{k+1}} = \frac{1}{3} \frac{1}{1 - \frac{1}{3}} = \frac{1}{2}.$$

Ainsi f est bien une application de $\{0, 1\}^{\mathbb{N}}$ dans $[0, 1]$ et d'après la remarque 1.5, il suffit de vérifier son injectivité pour qu'elle réalise une bijection de $\{0, 1\}^{\mathbb{N}}$ sur son image $A := \{f(u); u \in \{0, 1\}^{\mathbb{N}}\}$. Par la proposition 1.36, on en déduira la non dénombrabilité de A .

Pour montrer l'injectivité de f , supposons qu'il existe deux suites $u \neq u'$ éléments de $\{0, 1\}^{\mathbb{N}}$ telles que $f(u) = f(u')$. Comme $u \neq u'$, l'ensemble des entiers k tels que $u_k \neq u'_k$ est non vide et a donc un plus petit élément que nous notons j . On a ainsi $u_k = u'_k$ pour tout $k < j$ et $u_j \neq u'_j$. Quitte à permuter u et u' , on ne perd pas de généralité en supposant que $u_j = 1$ et $u'_j = 0$. L'égalité $f(u) = f(u')$ implique alors :

$$\frac{1}{3^{j+1}} = \sum_{k=j+1}^{+\infty} \frac{u'_k - u_k}{3^{k+1}}.$$

Comme $u'_k - u_k$ ne peut prendre que les valeurs $-1, 0$ ou 1 , il est majoré par 1, d'où :

$$\frac{1}{3^{j+1}} \leq \sum_{k=j+1}^{+\infty} \frac{1}{3^{k+1}} = \frac{1}{3^{j+2}} \frac{1}{1 - \frac{1}{3}} = \frac{1}{2 \times 3^{j+1}},$$

7. La lecture de ce qui suit requiert la connaissance des séries dont les principales propriétés sont rappelées section 1.4 ci-après, voir notamment les séries géométriques (exemple 1.49).

ce qui est impossible. On en déduit que si $f(u) = f(u')$, nécessairement $u = u'$, ce qui établit l'injectivité de f .

Pour vérifier la non dénombrabilité d'un intervalle non vide et non réduit à un singleton de \mathbb{R} , il suffit de remarquer que cet intervalle a au moins deux éléments a et b et qu'il contient alors $[a, b]$. Il suffit maintenant de construire une bijection $[0, 1] \rightarrow [a, b]$. L'application

$$f : [0, 1] \rightarrow [a, b], \quad t \mapsto ta + (1 - t)b$$

fait l'affaire. \square

Remarque 1.38. Il est facile de construire une bijection entre $] - 1, 1[$ et \mathbb{R} , par exemple $x \mapsto \tan(\pi x/2)$ et d'en déduire une bijection entre \mathbb{R} et n'importe quel intervalle ouvert non vide. En fait on peut montrer que les ensembles suivants ont tous même cardinal : $\{0, 1\}^{\mathbb{N}}$, \mathbb{R} , \mathbb{C} , tout intervalle non vide et non réduit à un singleton de \mathbb{R} . On dit qu'ils ont la *puissance du continu*. Sans aller jusqu'à démontrer complètement cette affirmation, nous nous contenterons de présenter ci-dessous une construction explicite d'une bijection entre $\{0, 1\}^{\mathbb{N}^*}$ et $[0, 1[$. Il est clair que $\{0, 1\}^{\mathbb{N}^*}$ est lui même en bijection avec $\{0, 1\}^{\mathbb{N}}$ (pourquoi?).

Exemple 1.39 (une bijection entre $\{0, 1\}^{\mathbb{N}^*}$ et $[0, 1[$). Commençons par un rappel sur le développement en base 2 des réels de $[0, 1[$, *i.e.* l'existence pour un $x \in [0, 1[$ d'une suite $(a_k)_{k \in \mathbb{N}^*} \in \{0, 1\}^{\mathbb{N}^*}$ telle que

$$x = \sum_{k=1}^{+\infty} \frac{a_k}{2^k}. \quad (1.14)$$

On appelle *nombre dyadique* de $[0, 1[$, tout $x \in [0, 1[$ de la forme $k2^{-n}$ avec $k \in \mathbb{N}$ et $n \in \mathbb{N}^*$. Un tel dyadique admet une écriture irréductible unique de la forme $l2^{-n}$ avec l impair. Notons Δ l'ensemble des dyadiques de $[0, 1[$.

Si $x \in [0, 1[\setminus \Delta$, il existe une unique suite $(a_k)_{k \in \mathbb{N}^*} \in \{0, 1\}^{\mathbb{N}^*}$ vérifiant (1.14). De plus cette suite $(a_k)_{k \in \mathbb{N}^*}$ comporte à la fois une infinité de 0 et une infinité de 1.

Si $x \in \Delta$, il existe deux suites $(a_k)_{k \in \mathbb{N}^*} \in \{0, 1\}^{\mathbb{N}^*}$ et $(a'_k)_{k \in \mathbb{N}^*} \in \{0, 1\}^{\mathbb{N}^*}$ vérifiant (1.14). L'une $(a_k)_{k \in \mathbb{N}^*}$ a tous ses termes nuls à partir d'un certain rang. C'est le *développement propre* du dyadique x . L'autre a tous ses termes égaux à 1 à partir d'un certain rang, c'est le *développement impropre* de x . Nous noterons $p(x)$ le développement propre de x et $i(x)$ son développement impropre. Par exemple le dyadique $\frac{3}{8}$ a pour développement propre $(0, 1, 1, 0, 0, 0, \dots)$ car $\frac{3}{8} = \frac{1}{4} + \frac{1}{8}$. Son développement impropre est $(0, 1, 0, 1, 1, 1, \dots)$ puisque $\frac{1}{8} = \sum_{k=4}^{+\infty} 2^{-k}$.

Définissons maintenant $f : [0, 1[\rightarrow \{0, 1\}^{\mathbb{N}^*}$ comme suit.

- Si $x \in [0, 1[\setminus \Delta$, on prend pour $f(x)$ l'unique suite $(a_k)_{k \in \mathbb{N}^*} \in \{0, 1\}^{\mathbb{N}^*}$ vérifiant (1.14).
- Si $x \in \Delta$ et $x \neq 0$, $x \neq 1/2$, il existe une écriture unique $x = l2^{-n}$ avec l impair.

On pose alors

$$f\left(\frac{l}{2^n}\right) = \begin{cases} i\left(\frac{l+1}{2^n}\right) & \text{si } l \equiv 1 \pmod{4}, \\ p\left(\frac{l-1}{2^n}\right) & \text{si } l \equiv 3 \pmod{4}. \end{cases}$$

- Pour $f(0)$ on prend la suite ne comportant que des 0 et pour $f(1/2)$ la suite ne comportant que des 1.

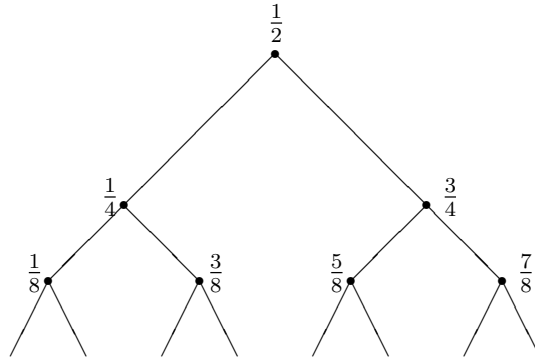


FIG. 1.3 – Arbre binaire des dyadiques de $]0, 1[$

Le mécanisme de construction de f sur les dyadiques autres que 0 et $1/2$ peut être décrit de manière informelle à l'aide de l'arbre binaire de la figure 1.3. Chaque individu de cet arbre hérite un développement binaire de son ascendant direct et engendre lui-même deux enfants et deux développements binaires. L'enfant de gauche hérite du développement impropre et celui de droite du développement propre.

La compréhension de la définition de f demandant plus d'effort que la vérification de sa bijectivité, ce dernier point est laissé au lecteur.

Proposition 1.40. *Soit J un ensemble au plus dénombrable d'indices et pour tout $j \in J$, soit A_j un ensemble au plus dénombrable. Alors $A := \bigcup_{j \in J} A_j$ est au plus dénombrable.*

Preuve. Le cas $J = \emptyset$ est trivial puisqu'alors $A = \emptyset$. Si tous les A_j sont vides, A l'est aussi (quel que soit J). On suppose désormais que J n'est pas vide et qu'au moins un des A_j est non vide. On va montrer que A est au plus dénombrable en construisant une injection h de A dans \mathbb{N} .

On peut traduire les hypothèses en écrivant qu'il existe une injection⁸ $f : J \rightarrow \mathbb{N}^*$ et que pour tout j tel que $A_j \neq \emptyset$, il existe une injection $g_j : A_j \rightarrow \mathbb{N}^*$. Admettons pour un instant que l'on peut construire une famille $(A'_j)_{j \in J}$ d'ensembles deux à deux disjoints tels que pour tout $j \in J$, $A'_j \subset A_j$ et que

$$A = \bigcup_{j \in J} A_j = \bigcup_{j \in J} A'_j.$$

On définit alors l'application $h : A \rightarrow \mathbb{N}$ comme suit. Si $x \in A$, il existe un *unique* $j \in J$ tel que $x \in A'_j$. On pose alors :

$$h(x) := p_{f(j)}^{g_j(x)},$$

8. Toute injection d'un ensemble dans \mathbb{N} peut se transformer en injection du même ensemble dans \mathbb{N}^* par composition avec la bijection $\mathbb{N} \rightarrow \mathbb{N}^*$, $n \mapsto n + 1$.

où p_k désigne le k -ième nombre premier. Remarquons que $h(x) \geq 2$ puisque $f(j) \geq 1$, la suite (p_k) est croissante de premier terme $p_1 = 2$ et $g_j(x) \geq 1$.

Pour vérifier l'injectivité de h , soient x et y dans A tels que $h(x) = h(y)$. En notant l l'unique indice tel que $y \in A'_l$, cette égalité s'écrit

$$p_{f(j)}^{g_j(x)} = p_{f(l)}^{g_l(y)}.$$

En raison de l'unicité de la décomposition d'un entier $n \geq 2$ en facteurs premiers, ceci entraîne $f(j) = f(l)$ et $g_j(x) = g_l(y)$, puis par injectivité de f , $j = l$ et $g_j(x) = g_j(y)$. Comme g_j est injective, on en déduit que $x = y$. L'injectivité de h est ainsi établie.

Il reste à justifier la construction de la famille $(A'_j)_{j \in J}$. On commence par transporter l'ordre de \mathbb{N} sur J via l'application f en écrivant pour $j, l \in J$, que $j \leq l$ signifie que $f(j) \leq f(l)$. Ceci nous permet de noter les éléments de J en suivant cet ordre $J = \{j_0, j_1, \dots\}$, où $j_0 \leq j_1 \leq \dots$. Ensuite on construit A'_j par récurrence en posant

$$\begin{array}{ll} A'_{j_0} := A_{j_0}, & B_{j_0} := A_{j_0} \\ A'_{j_1} := A_{j_1} \cap (A \setminus B_{j_0}), & B_{j_1} := B_{j_0} \cup A_{j_1} \\ A'_{j_2} := A_{j_2} \cap (A \setminus B_{j_1}), & B_{j_2} := B_{j_1} \cup A_{j_2} \\ \dots\dots & \dots\dots \end{array}$$

Les détails de la vérification sont laissés au lecteur. □

Proposition 1.41 (dénombrabilité par image surjective). *Soient A et B deux ensembles tels qu'il existe une surjection f de A sur B . Alors si A est dénombrable, B est au plus dénombrable. Si A est fini, B est fini et $\text{card } B \leq \text{card } A$.*

Il est possible que A soit infini et B fini, un exemple évident étant fourni par l'application nulle $f : x \mapsto 0$ de $A = \mathbb{N}$ dans $B = \{0\}$ qui est clairement une surjection.

Preuve. Supposons A dénombrable. Définissons sur A la relation d'équivalence $x \sim x'$ si $f(x) = f(x')$. Les classes d'équivalences pour cette relation réalisent une partition de A . Dans chaque classe d'équivalence, choisissons un représentant particulier⁹. Soit A_0 la partie de A formée de tous les représentants ainsi choisis. Si x et x' sont deux éléments distincts de A_0 , ils sont dans deux classes c et c' disjointes, donc $f(x) \neq f(x')$. La restriction de f à A_0 est donc injective. Par ailleurs, puisque f est surjective, tout $y \in B$ a au moins un antécédent x dans A . Si c est la classe de x , il y a dans cette classe un (unique) élément de A_0 qui a lui aussi y pour image par f . Donc la restriction de f à A_0 reste surjective et c'est finalement une bijection de A_0 sur B . L'ensemble B est en bijection avec une partie A_0 de l'ensemble dénombrable A , il est donc au plus dénombrable.

Le cas A fini est une adaptation facile de ce qui précède. □

9. Par exemple en fixant une numérotation de A par les entiers ($k \mapsto x_k$) et en décidant de prendre dans la classe c l'élément x_k d'indice k minimal. On évite ainsi l'invocation de l'axiome du choix...

1.4 Rappels sur les séries

Dans cette section nous rappelons sans démonstrations les points essentiels de la théorie des séries numériques vue en deuxième année. Nous détaillerons seulement la question de la convergence commutative en raison de son rôle dans la théorie des familles sommables.

1.4.1 Généralités

Définition 1.42. Soit $(u_k)_{k \in \mathbb{N}}$ une suite de nombres réels ou complexes. Pour tout $n \in \mathbb{N}$, on appelle somme partielle de rang n le nombre

$$S_n := \sum_{k=0}^n u_k.$$

- Si S_n tend vers une limite finie S quand n tend vers $+\infty$, on dit que la série de terme général u_k converge et a pour somme S , ce que l'on écrit

$$S = \sum_{k=0}^{+\infty} u_k := \lim_{n \rightarrow +\infty} S_n.$$

Dans ce cas, $R_n := S - S_n$ est appelé reste de rang n de la série.

- Si S_n n'a pas de limite finie (i.e. tend vers $-\infty$ ou $+\infty$ ou n'a pas de limite du tout), on dit que la série diverge.

Remarque 1.43. Par un abus d'écriture courant, on désigne la série (convergente ou divergente) par la notation « $\sum_{k=0}^{+\infty} u_k$ », mais on ne peut faire intervenir cette expression dans des calculs que si elle représente vraiment un nombre, i.e. si la série converge.

Remarque 1.44. Si les u_k sont complexes, posons $x_k := \operatorname{Re} u_k$ et $y_k := \operatorname{Im} u_k$, $S'_n := \operatorname{Re} S_n$ et $S''_n := \operatorname{Im} S_n$. La suite S_n converge dans \mathbb{C} vers S si et seulement si S'_n et S''_n convergent dans \mathbb{R} vers respectivement $S' := \operatorname{Re} S$ et $S'' := \operatorname{Im} S$. On en déduit immédiatement que la série de terme général complexe $u_k = x_k + iy_k$ converge si et seulement si les séries de terme général x_k et y_k convergent dans \mathbb{R} et que dans ce cas

$$\sum_{k=0}^{+\infty} (x_k + iy_k) = \sum_{k=0}^{+\infty} x_k + i \sum_{k=0}^{+\infty} y_k.$$

Remarque 1.45. La définition 1.42 se généralise immédiatement au cas où les u_k sont éléments d'un espace vectoriel normé E , la convergence dans E de S_n vers le vecteur S signifiant $\lim_{n \rightarrow +\infty} \|S - S_n\| = 0$.

Proposition 1.46. Si une série converge, son terme général u_n tend vers 0 quand n tend vers l'infini.

Remarque 1.47. La réciproque de la proposition 1.46 est fausse. Un contre exemple bien connu est la série harmonique de terme général $u_k = 1/k$ pour $k \geq 1$ (et $u_0 = 0$).

Remarque 1.48. La contraposée de la proposition 1.46 s'énonce : si u_n ne tend pas vers 0 quand n tend vers l'infini, alors la série diverge. On parle dans ce cas de divergence grossière de la série.

Exemple 1.49 (séries géométriques). Soit q un nombre réel ou complexe. La série géométrique standard $\sum_{k=0}^{+\infty} q^k$ converge si et seulement si $|q| < 1$. Dans ce cas, sa somme est donnée par

$$\sum_{k=0}^{+\infty} q^k = \frac{1}{1-q} \quad (|q| < 1).$$

Cette formule permet de calculer la somme de n'importe quelle série géométrique convergente (donc de raison q vérifiant $|q| < 1$). Il suffit de mettre en facteur le premier terme. Si $(u_k)_{k \in \mathbb{N}}$ est une suite géométrique de raison q (i.e. $u_{k+1} = qu_k$ pour tout k avec q ne dépendant pas de k), vérifiant $|q| < 1$, on a pour tout $j \in \mathbb{N}$,

$$\sum_{k=j}^{+\infty} u_k = \sum_{k=j}^{+\infty} u_j q^{k-j} = u_j \sum_{l=0}^{+\infty} q^l = \frac{u_j}{1-q}.$$

Théorème 1.50 (critère de Cauchy). *La série à termes réels ou complexes $\sum_{k=0}^{+\infty} u_k$ converge si et seulement si*

$$\forall \varepsilon > 0, \exists N \in \mathbb{N}, \forall n, m \geq N, |S_n - S_m| < \varepsilon.$$

Ce théorème n'est que l'application du critère de Cauchy à la suite des sommes partielles $S_n = \sum_{k=0}^n u_k$. Il se généralise immédiatement au cas où les termes u_k sont des éléments d'un espace vectoriel normé *complet*¹⁰, en remplaçant $|S_n - S_m| < \varepsilon$ par $\|S_n - S_m\| < \varepsilon$.

Corollaire 1.51 (convergence absolue). *Si la série $\sum_{k=0}^{+\infty} |u_k|$ converge, alors la série $\sum_{k=0}^{+\infty} u_k$ converge aussi. On dit qu'elle est absolument convergente.*

Le résultat s'étend immédiatement aux séries à termes dans un espace vectoriel normé *complet* en remplaçant $|u_k|$ par $\|u_k\|$. On parle alors de convergence normale.

Remarque 1.52. La réciproque du corollaire 1.51 est fausse. Un contre exemple bien connu est la série alternée de terme général $u_k = (-1)^k/k$ ($k \geq 1$) qui converge alors que la série des valeurs absolues est la série harmonique qui diverge.

1.4.2 Séries à termes positifs

Si tous les u_k sont positifs, $(S_n)_{n \in \mathbb{N}}$ est une suite *croissante* car pour tout $n \geq 1$, $S_n - S_{n-1} = u_n \geq 0$. Il n'y a alors que deux cas possibles :

- ou bien S_n a une limite finie S dans \mathbb{R}_+ , la série converge et a pour somme S ;

10. C'est le cas en particulier pour les séries à valeurs dans \mathbb{R}^d . Un espace vectoriel normé complet est appelé espace de Banach.

– ou bien S_n tend vers $+\infty$ et la série diverge.

La divergence d'une série à termes positifs équivaut donc à la convergence vers $+\infty$ de la suite de ses sommes partielles. Il est commode dans ce cas de considérer que la série « converge dans $\overline{\mathbb{R}}_+$ » et que sa somme vaut $+\infty$. Ainsi lorsque les u_k sont tous positifs, l'écriture $\sum_{k=0}^{+\infty} u_k$ a toujours un sens comme représentant un élément S de $\overline{\mathbb{R}}_+$: S étant un réel positif si la série converge au sens de la définition 1.42, $S = +\infty$ sinon.

Il importe de bien comprendre que ceci est particulier aux séries à termes positifs. Une série à termes de signe quelconque¹¹ peut très bien diverger sans que la suite des sommes partielles tende vers $+\infty$ ou $-\infty$. Un exemple évident est la série de terme général $u_k = (-1)^k$.

Théorème 1.53 (comparaison). *On suppose qu'il existe un $k_0 \in \mathbb{N}$ tel que pour tout $k \geq k_0$, $0 \leq u_k \leq v_k$. Alors*

- a) la convergence¹² de $\sum_{k=0}^{+\infty} v_k$ implique celle de $\sum_{k=0}^{+\infty} u_k$;
- b) la divergence de $\sum_{k=0}^{+\infty} u_k$ implique celle de $\sum_{k=0}^{+\infty} v_k$.

Parmi les applications du théorème de comparaison figure la comparaison à une série géométrique qui a donné naissance aux règles dites de D'Alembert et de Cauchy (basées respectivement sur l'étude du comportement asymptotique de u_{n+1}/u_n et de $u_n^{1/n}$). Ces règles n'ont d'autre utilité que de résoudre des exercices *ad hoc* et on peut facilement s'en passer.

Théorème 1.54 (règle des équivalents). *On suppose que u_k et v_k sont équivalents¹³ quand n tend vers l'infini et qu'à partir d'un rang k_0 , $v_k \geq 0$. Alors les séries $\sum_{k=0}^{+\infty} u_k$ et $\sum_{k=0}^{+\infty} v_k$ sont de même nature (toutes deux convergentes ou toutes deux divergentes).*

Attention, l'hypothèse d'équivalence de u_k et v_k n'implique pas à elle seule que les deux séries soient de même nature si le signe de v_k n'est pas constant à partir d'un certain rang. Comme contre exemple on peut proposer $u_k = (-1)^k k^{-1/2}$ et $v_k = u_k + 1/k$ (exercice).

Théorème 1.55 (comparaison série-intégrale). *Soit f continue sur $[k_0, +\infty[$, décroissante et positive sur cet intervalle. La série $\sum_{k=k_0}^{+\infty} f(k)$ converge si et seulement si $\int_{k_0}^n f(t) dt$ a une limite finie quand n tend vers $+\infty$.*

La démonstration repose sur l'encadrement

$$\forall n > k_0, \quad \sum_{k=k_0}^{n-1} f(k+1) \leq \int_{k_0}^n f(t) dt \leq \sum_{k=k_0}^{n-1} f(k)$$

illustré par la figure 1.4. Cet encadrement a son intérêt propre pour contrôler le reste d'une série à l'aide d'une intégrale généralisée (ou *vice versa*).

En appliquant le théorème 1.55 avec $f(t) = t^{-\alpha}$, $\alpha > 0$ et $k_0 = 1$, on obtient la caractérisation de la convergence des séries de Riemann.

11. Plus précisément une série dont la suite des termes présente une infinité de changements de signe.

12. La convergence et la divergence dans cet énoncé sont entendues au sens de la définition 1.42.

13. Ce qui signifie que l'on peut écrire $u_k = c_k v_k$ avec $\lim_{k \rightarrow +\infty} c_k = 1$.

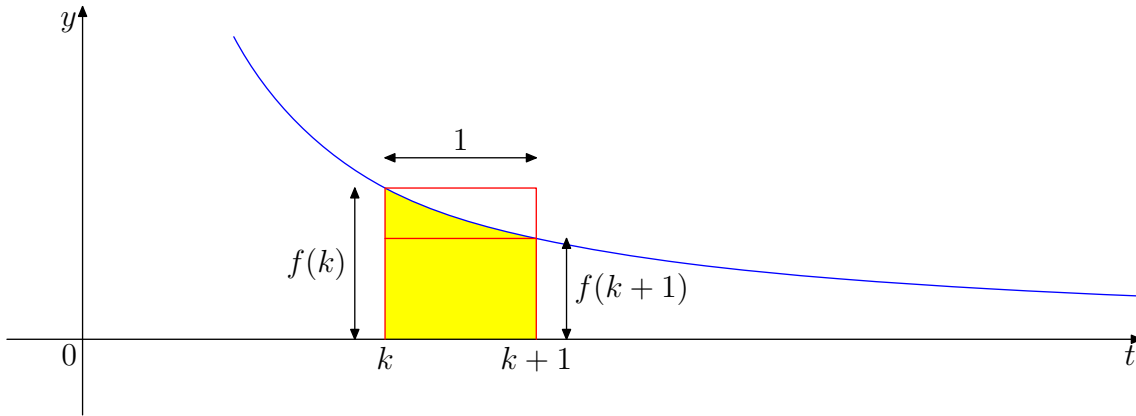


FIG. 1.4 – Comparaison série-intégrale : $f(k+1) \leq \int_k^{k+1} f(t) dt \leq f(k)$

Corollaire 1.56 (séries de Riemann).

$$\text{La série } \sum_{k=1}^{+\infty} \frac{1}{k^\alpha} \text{ est } \begin{cases} \text{convergente pour } \alpha > 1, \\ \text{divergente pour } 0 < \alpha \leq 1. \end{cases}$$

La série est aussi divergente pour $\alpha \leq 0$, puisqu'alors son terme général ne tend pas vers 0.

Corollaire 1.57 (séries de Bertrand).

$$\text{La série } \sum_{k=2}^{+\infty} \frac{1}{k(\ln k)^\beta} \text{ est } \begin{cases} \text{convergente pour } \beta > 1, \\ \text{divergente pour } 0 < \beta \leq 1. \end{cases}$$

Remarque 1.58. Pour $\alpha \neq 1$ (ou pour $\alpha = 1$ et $\beta \leq 0$), la nature de la série de Bertrand $\sum_{k=2}^{+\infty} k^{-\alpha} (\ln k)^{-\beta}$ s'obtient directement par comparaison avec une série de Riemann.

1.4.3 Séries à termes de signe non constant

Après avoir vu que la convergence de $\sum_{k=0}^{+\infty} |u_k|$ implique celle de $\sum_{k=0}^{+\infty} u_k$, on peut se demander s'il existe des séries qui convergent sans converger absolument.

Remarquons d'abord que si le signe de u_k est constant à partir d'un certain rang k_0 , l'étude de la convergence de $\sum_{k=0}^{+\infty} u_k$ se ramène à celle d'une série à termes positifs. En effet, $\sum_{k=0}^{+\infty} u_k$ et $\sum_{k=k_0}^{+\infty} u_k$ sont de même nature et si $u_k \leq 0$ pour $k \geq k_0$, il suffit de considérer $\sum_{k=k_0}^{+\infty} (-u_k)$. Ainsi pour une série à termes de signe constant à partir d'un certain rang, la convergence équivaut à la convergence absolue et la divergence ne peut se produire que si S_n tend vers $+\infty$ ou vers $-\infty$. Les seules séries susceptibles d'être convergentes sans l'être absolument sont donc celles dont le terme général *change de signe une infinité de fois*.

Par exemple, la série $\sum_{k=1}^{+\infty} (-1)^k / k$ converge mais pas absolument. C'est une application du théorème des séries alternées.

Théorème 1.59 (des séries alternées). Soit $(u_k)_{k \in \mathbb{N}}$ une suite alternée (i.e. pour tout k , u_k et u_{k+1} sont de signes contraires) telle que $|u_k|$ décroisse et tende vers 0. Alors

- a) la série $\sum_{k=0}^{+\infty} u_k$ converge,
- b) pour tout $n \in \mathbb{N}$, les sommes partielles consécutives S_n et S_{n+1} encadrent¹⁴ la somme S et le reste $R_n = \sum_{k=n+1}^{+\infty} u_k$ vérifie

$$|R_n| \leq |u_{n+1}|.$$

Exemple 1.60. Pour $0 < \alpha \leq 1$, la série $\sum_{k=1}^{+\infty} (-1)^k k^{-\alpha}$ converge mais pas absolument. Pour $\alpha > 1$ elle est absolument convergente.

Exemple 1.61. La série $\sum_{k=1}^{+\infty} (-1)^k / \ln k$ converge mais pas absolument.

Remarque 1.62. Même si une série alternée est absolument convergente, le b) du théorème reste intéressant pour le calcul numérique de la somme S .

1.4.4 Opérations sur les séries

La somme des séries de terme général u_k et v_k est la série de terme général $(u_k + v_k)$. Le produit de la série de terme général u_k par la constante a est la série de terme général au_k . Le produit de deux séries sera étudié plus tard à l'aide des familles sommables.

Proposition 1.63. Les affirmations suivantes sont vérifiées pour des séries à termes réels ou complexes ou dans un espace vectoriel normé.

- a) La somme de deux séries convergentes $\sum_{k=0}^{+\infty} u_k$ et $\sum_{k=0}^{+\infty} v_k$ est une série convergente et

$$\sum_{k=0}^{+\infty} (u_k + v_k) = \sum_{k=0}^{+\infty} u_k + \sum_{k=0}^{+\infty} v_k.$$

- b) Le produit de la série convergente $\sum_{k=0}^{+\infty} u_k$ par la constante a est une série convergente et

$$\sum_{k=0}^{+\infty} au_k = a \sum_{k=0}^{+\infty} u_k.$$

- c) La somme d'une série convergente et d'une série divergente est une série divergente.
- d) La somme de deux séries divergentes peut être convergente ou divergente.

Dans une somme d'un nombre fini de termes, les propriétés de commutativité et d'associativité de l'addition permettent

- de changer l'ordre des termes sans changer la valeur de la somme,
- de faire des groupements de termes sans changer la valeur de la somme.

14. Attention, une fois sur deux on aura $S_{n+1} \leq S_n$.

Que deviennent ces propriétés pour les séries considérées comme sommes d'une *infinité de termes*? En gros la réponse est : si la série est absolument convergente, tout se passe bien, sinon on peut avoir des situations très pathologiques.

Le seul cas où l'on n'ait pas besoin de convergence absolue pour retrouver une situation conforme aux propriétés des sommes d'un nombre fini de termes est celui de la sommation par paquets de taille finie fixe.

Proposition 1.64 (sommation par paquets de taille finie fixe). *Soit $(u_n)_{n \in \mathbb{N}}$ une suite tendant vers 0 à l'infini. Fixons un entier $p \geq 2$ et définissons les « paquets de p termes consécutifs »*

$$v_j = \sum_{i=0}^{p-1} u_{jp+i}, \quad j \in \mathbb{N}.$$

Alors les séries $\sum_{k=0}^{+\infty} u_k$ et $\sum_{j=0}^{+\infty} v_j$ sont de même nature et ont même somme lorsqu'elles convergent.

Remarque 1.65. Sans l'hypothèse de convergence vers 0 de la suite (u_n) , le résultat devient grossièrement faux, contre exemple $u_n = (-1)^n$.

Remarque 1.66. Si $\sum_{k=0}^{+\infty} u_k$ est absolument convergente, la série des paquets $\sum_{j=0}^{+\infty} v_j$ l'est aussi. La réciproque est fausse, comme on peut le voir en groupant deux par deux les termes de la série $\sum_{k=0}^{+\infty} (-1)^k / (k+1)$, obtenant ainsi la série de terme général positif $v_j = (2j+1)^{-1}(2j+2)^{-2}$ équivalent à $4j^{-2}$.

Définition 1.67 (convergence commutative). *La série $\sum_{k=0}^{+\infty} u_k$ est dite commutativement convergente et de somme S si pour toute bijection $f : \mathbb{N} \rightarrow \mathbb{N}$ la série $\sum_{k=0}^{+\infty} u_{f(k)}$ converge et a pour somme S . Dans ce cas nous pouvons utiliser la notation*

$$S = \sum_{k \in \mathbb{N}} u_k \quad \text{au lieu de} \quad S = \sum_{k=0}^{+\infty} u_k,$$

puisque la somme de la série ne dépend pas de l'ordre dans lequel on effectue la sommation. L'écriture avec indexation par « $k \in \mathbb{N}$ » ne présuppose aucun ordre d'écriture des termes¹⁵.

Théorème 1.68. *La convergence absolue d'une série à termes réels ou complexes, implique sa convergence commutative.*

Nous allons démontrer le théorème en examinant successivement les cas des séries à termes positifs, réels et complexes.

15. Par analogie avec l'écriture $\sum_{i \in I} u_i$, où I est un ensemble fini d'indices. Dans ce cas le résultat de l'addition des u_i pour $i \in I$ ne dépend pas de l'ordre des termes et d'ailleurs l'ensemble I n'a nul besoin ici d'être ordonné.

Preuve du th. 1.68, cas d'une série à termes positifs. Nous traitons d'abord le cas d'une série à termes positifs pour laquelle la convergence absolue se réduit à la convergence et signifie que la suite croissante des sommes partielles S_n a une limite finie S dans \mathbb{R}_+ . Soit f une bijection $\mathbb{N} \rightarrow \mathbb{N}$. Définissons les deux suites d'entiers $(p_n)_{n \in \mathbb{N}}$ et $(q_n)_{n \in \mathbb{N}}$ comme suit. Pour p_n nous prenons le plus petit entier p tel que $\{f(k); k \leq p\}$ recouvre $\{0, \dots, n\}$. L'existence de tels p et donc de p_n découle clairement de la surjectivité de f . Pour q_n on prend $\max\{f(k); 0 \leq k \leq p_n\}$. Autrement dit p_n sert à recouvrir $\{0, \dots, n\}$ de la manière la plus économique par les premiers termes de la suite des $f(k)$ et q_n sert à « boucher les trous » de la suite d'entiers ainsi formée. On a ainsi

$$\forall n \in \mathbb{N}, \quad \{0, \dots, n\} \subset \{f(k); k \leq p_n\} \subset \{0, \dots, q_n\}. \quad (1.15)$$

On a clairement $n \leq p_n$ et $n \leq q_n$ donc p_n et q_n tendent vers l'infini avec n . De plus ces deux suites sont croissantes de par leur construction. Posons

$$S_n := \sum_{k=0}^n u_k, \quad T_n := \sum_{k=0}^n u_{f(k)}.$$

En raison de la positivité des u_k et de (1.15), nous disposons de l'encadrement

$$\forall n \in \mathbb{N}, \quad S_n \leq T_{p_n} \leq S_{q_n}. \quad (1.16)$$

Si S_n converge vers S , la sous-suite (S_{q_n}) converge aussi vers S et (1.16) implique alors la convergence de (T_{p_n}) vers S . En raison de la positivité des u_k , la suite (T_n) est croissante. Nous venons de voir qu'elle a une sous-suite qui converge vers S , c'est donc toute la suite (T_n) qui converge vers S . Nous venons d'établir que $\sum_{k=0}^{+\infty} u_{f(k)}$ converge et a pour somme $S = \sum_{k=0}^{+\infty} u_k$. Comme la bijection f est quelconque, ceci achève la preuve du théorème 1.68 dans le cas d'une série à termes positifs. \square

Remarque 1.69. Pour établir (1.16), nous n'avons pas utilisé l'hypothèse de convergence de la série de terme général u_k , mais seulement la positivité des u_k . Par conséquent (1.16) reste valable si la série à termes positifs $\sum_{k=0}^{+\infty} u_k$ diverge. Dans ce cas S_n tend vers $+\infty$, donc aussi T_{p_n} et la suite croissante (T_n) ayant une sous-suite tendant vers $+\infty$, c'est toute la suite (T_n) qui tend vers l'infini. Ainsi on a $\sum_{k=0}^{+\infty} u_{f(k)} = +\infty = \sum_{k=0}^{+\infty} u_k$.

Pour la commodité de référence et en raison du rôle essentiel joué par les séries à termes positifs dans ce cours, nous rassemblons dans l'énoncé suivant les résultats obtenus dans la preuve du théorème 1.68 (cas $u_k \geq 0$) et dans la remarque 1.69.

Proposition 1.70. Si $(u_k)_{k \in \mathbb{N}}$ est une suite de réels positifs, on a pour toute bijection $f : \mathbb{N} \rightarrow \mathbb{N}$,

$$\sum_{k=0}^{+\infty} u_k = \sum_{k=0}^{+\infty} u_{f(k)},$$

cette égalité ayant lieu dans $\overline{\mathbb{R}}_+$. Ainsi si les u_k sont positifs, la notation $\sum_{k \in \mathbb{N}} u_k$ est toujours légitime et représente la somme $S \in \overline{\mathbb{R}}_+$ de la série.

Dans la preuve du théorème 1.68, le cas d'une série à termes réels de signe quelconque va se ramener à celui d'une série à termes positifs en utilisant la décomposition d'un réel en partie positive et partie négative. Ceci requiert une digression préalable (de la définition 1.71 à la remarque 1.74).

Définition 1.71 (parties positive et négative). *Pour tout x réel, sa partie positive notée x^+ et sa partie négative notée x^- sont les réels positifs définis par*

$$x^+ := \max(0, x), \quad x^- := \max(0, -x).$$

On a alors pour tout x réel

$$x = x^+ - x^-, \quad (1.17)$$

$$|x| = x^+ + x^-. \quad (1.18)$$

Par exemple si $x = 3,12$, $x^+ = 3,12$ et $x^- = 0$, si $x = -2,3$, $x^+ = 0$ et $x^- = 2,3$. Il importe de bien noter que la partie négative d'un réel est un nombre positif ou nul. Voyons d'abord comment intervient cette décomposition pour une série absolument convergente.

Lemme 1.72. *Si $\sum_{k=0}^{+\infty} u_k$ est une série à termes réels absolument convergente et de somme S , les séries à termes positifs $\sum_{k=0}^{+\infty} u_k^+$ et $\sum_{k=0}^{+\infty} u_k^-$ sont convergentes et on a*

$$\sum_{k=0}^{+\infty} u_k = \sum_{k=0}^{+\infty} (u_k^+ - u_k^-) = \sum_{k=0}^{+\infty} u_k^+ - \sum_{k=0}^{+\infty} u_k^-. \quad (1.19)$$

Preuve. Par (1.18), on a $0 \leq u_k^+ \leq |u_k|$ et $0 \leq u_k^- \leq |u_k|$. Les séries de terme général u_k^+ et u_k^- sont donc convergentes par le théorème de comparaison. La première égalité dans (1.19) résulte alors de (1.17) appliquée à $x = u_k$. La deuxième égalité s'obtient par la proposition 1.63 a) et b). \square

Remarque 1.73. La réciproque est vraie : si les séries à termes positifs $\sum_{k=0}^{+\infty} u_k^+$ et $\sum_{k=0}^{+\infty} u_k^-$ sont convergentes, la série $\sum_{k=0}^{+\infty} u_k$ est absolument convergente et on a (1.19) grâce à la proposition 1.63. Ceci montre que l'on ne peut pas se passer de l'hypothèse de convergence absolue dans le lemme 1.72. Il est d'autre part facile d'exhiber un contre exemple : la série de terme général $u_k = (-1)^k/k$.

Remarque 1.74. L'égalité « $\sum_{k=0}^{+\infty} u_k^+ = (\sum_{k=0}^{+\infty} u_k)^+$ » est grossièrement fausse ! Elle n'est même pas vraie pour des sommes finies. Comparez $(a+b)^+$ et $a^+ + b^+$ pour $a = -1$ et $b = 2$ pour vous en convaincre.

Preuve du th. 1.68 dans le cas d'une série à termes réels. Soit $\sum_{k=0}^{+\infty} u_k$ une série absolument convergente à termes réels et f une bijection quelconque $\mathbb{N} \rightarrow \mathbb{N}$. On utilise la décomposition $u_{f(k)} = u_{f(k)}^+ - u_{f(k)}^-$. Par le lemme 1.72, la convergence absolue de la série de terme général u_k implique la convergence des séries à termes généraux positifs u_k^+ et u_k^- . La preuve du théorème 1.68 dans le cas des termes positifs nous fournit alors la

convergence des séries $\sum_{k=0}^{+\infty} u_{f(k)}^+$ et $\sum_{k=0}^{+\infty} u_{f(k)}^-$ et l'égalité de leurs sommes avec $\sum_{k=0}^{+\infty} u_k^+$ et $\sum_{k=0}^{+\infty} u_k^-$ respectivement. Par la remarque 1.73, on en déduit la convergence absolue de $\sum_{k=0}^{+\infty} u_{f(k)}$. Compte-tenu du lemme 1.72, on a de plus

$$\sum_{k=0}^{+\infty} u_{f(k)} = \sum_{k=0}^{+\infty} u_{f(k)}^+ - \sum_{k=0}^{+\infty} u_{f(k)}^- = \sum_{k=0}^{+\infty} u_k^+ - \sum_{k=0}^{+\infty} u_k^- = \sum_{k=0}^{+\infty} (u_k^+ - u_k^-) = \sum_{k=0}^{+\infty} u_k.$$

□

Preuve du th. 1.68 dans le cas d'une série à termes complexes. Pour des u_k complexes, notons $x_k := \operatorname{Re} u_k$ et $y_k := \operatorname{Im} u_k$. Grâce aux inégalités $0 \leq |x_k| \leq |u_k|$ et $0 \leq |y_k| \leq |u_k|$, les séries de terme général réel x_k et y_k héritent de la convergence absolue de la série de terme général complexe u_k . On conclut alors en combinant la preuve du th. 1.68 dans le cas réel et la remarque 1.44. □

Remarque 1.75. L'argument utilisé ci-dessus dans le cas complexe se généralise aux séries à valeurs dans un espace vectoriel normé E de *dimension finie*, disons pour simplifier $E = \mathbb{R}^d$. En effet sur cet espace toutes les normes sont équivalentes et on peut donc choisir la norme donnée par $\|x\| = \max_{1 \leq i \leq d} |x_i|$ pour $x = (x_1, \dots, x_d)$. La convergence d'une suite dans $(E, \|\cdot\|)$ équivaut à la convergence composante par composante et la convergence de $\sum_{k=0}^{+\infty} \|u_k\|$ implique la convergence absolue des d séries de terme général $u_{k,i}$, en posant $u_k = (u_{k,1}, \dots, u_{k,d})$. L'adaptation de la preuve ci-dessus est alors immédiate. Par contre si E est de dimension infinie, ce raisonnement n'est plus valable.

Théorème 1.76. *La convergence commutative d'une série à termes réels ou complexes, implique sa convergence absolue.*

Preuve. Il est clair qu'il suffit de traiter le cas des séries à termes réels. Nous allons démontrer la *contraposée* de l'énoncé, c'est-à-dire que la négation de la conclusion implique la négation de l'hypothèse. Dans ce but nous supposons que la série de terme général u_k converge vers le réel S mais pas absolument, autrement dit que :

$$\sum_{k=0}^{+\infty} u_k = S \in \mathbb{R} \quad \text{et} \quad \sum_{k=0}^{+\infty} |u_k| = +\infty. \quad (1.20)$$

Nous allons construire une bijection $f : \mathbb{N} \rightarrow \mathbb{N}$ telle que $\sum_{k=0}^{+\infty} u_{f(k)}$ ne converge pas vers S .

On commence par noter que (1.20) implique que

$$\sum_{k=0}^{+\infty} u_k^+ = +\infty \quad \text{et} \quad \sum_{k=0}^{+\infty} u_k^- = +\infty. \quad (1.21)$$

En effet si ces deux sommes sont finies, la série $\sum_{k=0}^{+\infty} u_k$ est absolument convergente par la remarque 1.73. Si l'une est finie et l'autre infinie, la série $\sum_{k=0}^{+\infty} u_k$ diverge par la proposition 1.63 b) et c).

Nous allons montrer que l'on peut construire une bijection f pour laquelle les sommes partielles $\sum_{k=0}^n u_{f(k)}$ oscillent indéfiniment entre -1 et 1 et donc n'ont pas de limite. L'écriture explicite de la définition de f étant assez lourde, nous nous contenterons de donner l'idée de sa construction. Le point clé est le lemme suivant.

Lemme 1.77. *Soit $v = (v_k)_{k \in \mathbb{N}}$ une suite de réels positifs telle que $\sum_{k=0}^{+\infty} v_k = +\infty$. Pour tout intervalle borné I de \mathbb{R}_+ , définissons*

$$T(I, v) := \sum_{k \in \mathbb{N} \cap I} v_k,$$

avec la convention $T(\emptyset, v) := 0$. On a alors

$$\forall i \in \mathbb{N}, \forall x \in \mathbb{R}_+, \exists ! j = j(i, x, v) \in \mathbb{N}; \quad T([i, j[, v) \leq x < T([i, j], v). \quad (1.22)$$

Preuve du lemme. Notons $t_{i,j} := T([i, j[, v)$. On remarque que pour tout $i \in \mathbb{N}$ fixé, la suite $(t_{i,j})_{j \geq i}$ est croissante à cause de la positivité des v_k et tend vers $+\infty$ quand j tend vers $+\infty$. Ce dernier point résulte de l'hypothèse $\sum_{k=0}^{+\infty} v_k = +\infty$ et du fait que l'on ne change pas la divergence de cette série en supprimant ses i premiers termes. De plus on a $t_{i,i} = 0$ car $[i, i[= \emptyset$. Il est alors clair que l'on obtient (1.22) avec $j := \max\{l \geq i; t_{i,l} \leq x\}$. \square

Partageons maintenant \mathbb{N} en les deux sous-ensembles complémentaires :

$$A^+ := \{l \in \mathbb{N}; u_l \geq 0\}, \quad A^- := \{l \in \mathbb{N}; u_l < 0\}.$$

Il résulte immédiatement de (1.21) que ces deux ensembles sont infinis (donc dénombrables comme parties infinies de \mathbb{N}). Définissons alors les suites $v = (v_k)_{k \in \mathbb{N}}$ et $w = (w_k)_{k \in \mathbb{N}}$ en prenant pour v_k (resp. w_k) le terme de rang k dans la numérotation croissante des u_l indexés par A^+ (resp. A^-).

En appliquant le lemme 1.77 alternativement avec les suites v et $-w$, on peut construire f par récurrence de façon à ce que les sommes partielles $T_n = \sum_{k=0}^n u_{f(k)}$ se trouvent à gauche de -1 pour une infinité de n et à droite de $+1$ pour une infinité de n . Voici le début de la construction. On applique d'abord le lemme avec $i = 0$, $x = 1$ et v . On obtient alors $T_{n_1} \geq 1$ pour $n_1 = j(0, 1, v)$ donné par (1.22). Ensuite on revient en arrière en appliquant le lemme avec $-w$, $i = 0$ et $x = 1 + T_{n_1}$. Pour $n_2 = n_1 + j(0, x, -w)$, on a alors $T_{n_2} < -1$. Pour la troisième étape, on utilise le lemme avec v , $i = j(0, 1, v) + 1$, $x = -T_{n_2} + 1$ et pour $n_3 = n_2 + j(i, x, v)$ on obtient $T_{n_3} \geq 1$. Et ainsi de suite... \square

On peut adapter la démonstration ci-dessus pour construire d'autres bijections f de façon à faire osciller les sommes partielles entre a et b fixés, ou pour les faire converger vers n'importe quel réel fixé¹⁶, pour les faire tendre vers $+\infty$, vers $-\infty$, etc.

Pour conclure cette section, retenons que pour les séries à termes réels ou complexes, *convergence commutative et convergence absolue sont équivalentes*.

16. Il faudra utiliser alors la convergence vers 0 de u_n , ce dont nous n'avons pas eu besoin ci-dessus.

1.5 Familles sommables

Dans cette section, nous allons généraliser la notion de série en essayant de donner un sens à une expression de la forme $\sum_{i \in I} u_i$, où I est un ensemble infini¹⁷. Dans le cadre de ce cours, nous n'aurons besoin que du cas où les u_i sont réels ou complexes. Néanmoins, pour faciliter l'utilisation dans d'autres branches des mathématiques (par exemple analyse complexe, séries de Fourier, analyse fonctionnelle...), nous nous situerons d'emblée dans le cas où les u_i sont des éléments d'un espace vectoriel normé E . *Le lecteur que cela gênerait pourra facilement adapter les énoncés au cas des u_i réels ou complexes en remplaçant les normes par des valeurs absolues ou des modules.* Cette simplification des énoncés n'apporte d'ailleurs pas de simplification notable des démonstrations.

Partons de la définition de la convergence d'une série de vecteurs de terme général u_k dans un espace vectoriel normé E . Ici l'ensemble d'indexation est \mathbb{N} . On dit que cette série converge et a pour somme le vecteur S de E si la suite des sommes partielles $S_n := u_0 + u_1 + \dots + u_n$ converge vers S , autrement dit si $\|S_n - S\|$ tend vers 0 quand n tend vers $+\infty$. Ceci s'écrit encore :

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}, \forall n \geq n_0, \quad \|S_n - S\| < \varepsilon. \quad (1.23)$$

Si on veut généraliser ceci à une famille de vecteurs $(u_i)_{i \in I}$ où I est un ensemble infini quelconque, on se heurte immédiatement à une difficulté, c'est qu'une écriture comme « $\forall i \geq i_0$ », n'a en général pas de sens dans I qui n'a aucune raison d'être muni d'une relation d'ordre total. Essayons alors de traduire l'idée exprimée par (1.23) sans faire appel à la structure d'ordre de \mathbb{N} . On remarque pour cela que S_n réalise une approximation de S par la somme d'un *nombre fini* de termes de la famille $(u_k)_{k \in \mathbb{N}}$ avec une erreur $\|S_n - S\|$ inférieure à ε . Cette approximation peut être réalisée par une somme finie indexée par n'importe quel ensemble $K_n := \{0, 1, 2, \dots, n\}$ d'entiers consécutifs entre 0 et n pourvu que $n \geq n_0$. Pour se débarrasser de la relation d'ordre intervenant dans cette dernière écriture on la reformule en $K_n \supset K_{n_0}$. Réécrivons maintenant (1.23) à l'aide des ensembles emboîtés K_n .

$$\forall \varepsilon > 0, \exists K_{n_0} = \{0, \dots, n_0\}, \forall K_n \supset K_{n_0}, \quad \left\| \sum_{k \in K_n} u_k - S \right\| < \varepsilon. \quad (1.24)$$

Au risque d'insister lourdement, notons que dans l'écriture, « $\forall K_n \supset K_{n_0}$ », K_n désigne non pas n'importe quelle partie finie de \mathbb{N} , mais un ensemble fini de la forme $\{0, 1, 2, \dots, n\}$. Avons nous réussi ainsi à expurger (1.23) de la relation d'ordre sur \mathbb{N} ? En fait non, nous avons seulement réussi à la cacher dans la définition des ensembles K_n d'entiers *consécutifs* entre 0 et n . Si on veut vraiment généraliser à un ensemble d'indexation I quelconque, on est donc condamné à renoncer à cette structure particulière des K_n et à ne retenir que leur finitude. On est ainsi amené à introduire une *nouvelle*

17. L'exposé présenté ici est essentiellement une adaptation du chapitre XIV *Séries dans les espaces vectoriels normés* de l'ouvrage de L. SCHWARTZ, *Topologie générale et analyse fonctionnelle*, Hermann, Paris 1970.

notion de convergence pour la série de terme général u_k :

$$\forall \varepsilon > 0, \exists J \text{ fini } \subset \mathbb{N}, \forall K \text{ fini tel que } J \subset K \subset \mathbb{N}, \quad \left\| \sum_{k \in K} u_k - S \right\| < \varepsilon. \quad (1.25)$$

Pourquoi avons nous pris la précaution de qualifier cette convergence de *nouvelle* ? Parce que s'il est clair que (1.25) implique (1.24) et donc (1.23), rien ne nous permet d'affirmer que la réciproque soit vraie. Nous verrons d'ailleurs ci-dessous que cette réciproque est fausse pour les séries qui ne sont pas commutativement convergentes.

Après ce *briefing*, lançons nous résolument dans l'aventure de la sommation d'une famille $(u_i)_{i \in I}$ indexée par un ensemble quelconque, les u_i étant des éléments d'un espace vectoriel normé E . Commençons par une notation. Si K est un sous-ensemble *fini* de I , on pose :

$$S_K := \sum_{i \in K} u_i.$$

Cette définition est cohérente puisque dans un espace vectoriel E la somme d'un nombre fini d'éléments de E est encore un élément de E et ne dépend pas de l'ordre d'écriture des termes. Cette notation n'est pas contradictoire avec celle utilisée pour les sommes partielles S_n d'une série de terme général u_i , $i \in \mathbb{N}$, en considérant que S_n est une abréviation de $S_{\{0, \dots, n\}}$ qu'il ne faut pas confondre avec $S_{\{n\}} = u_n$. Dans le cas particulier où $K = \emptyset$, on fait la convention $S_\emptyset := 0$ (vecteur nul de E). Dans tout ce qui suit, les lettres majuscules J, K désigneront toujours des parties finies de I et nous omettrons parfois d'écrire $K \subset I$ si le contexte lève toute ambiguïté.

Définition 1.78 (famille sommable). *Soit E un espace vectoriel normé, I un ensemble d'indices et $(u_i)_{i \in I}$ une famille d'éléments de E . On dit que cette famille est sommable de somme $S \in E$ si*

$$\forall \varepsilon > 0, \exists J \text{ fini } \subset I, \forall K \text{ fini } \supset J, \quad \|S_K - S\| < \varepsilon. \quad (1.26)$$

On écrira alors $S = \sum_{i \in I} u_i$.

Insistons sur le fait que cette définition ne suppose aucune structure d'ordre sur l'ensemble d'indices I et donc aucun ordre privilégié d'écriture des termes. Si I est fini, toute famille $(u_i)_{i \in I}$ est trivialement sommable, il suffit de prendre $S = S_I$ et $J = I$ dans (1.26) pour avoir $\|S_K - S\| = 0$. La définition 1.78 n'a donc d'intérêt que pour I infini. Nous verrons bientôt qu'en fait, on peut se ramener à un I au plus dénombrable (prop. 1.85).

Remarque 1.79. Si $\{u_i, i \in I\}$ est sommable, sa somme S est unique, c'est-à-dire que si les vecteurs S et S' vérifient tous les deux (1.26), alors $S = S'$. Bien sûr, l'ensemble fini « J » associé à S' et à ε par (1.26) n'a aucune raison d'être le même que pour S , nous le noterons donc plutôt J' . Pour tout $\varepsilon > 0$, l'ensemble fini $K = J \cup J'$ contient à la fois J et J' , donc $\|S - S_K\| \leq \varepsilon$ et $\|S' - S_K\| \leq \varepsilon$. Par inégalité triangulaire, $\|S - S'\| \leq 2\varepsilon$. Dans cette dernière inégalité, K qui dépendait de ε a disparu et le premier membre ne dépend pas de ε . L'inégalité étant vraie pour tout $\varepsilon > 0$, $\|S - S'\| = 0$ et $S = S'$.

La sommabilité est préservée par combinaison linéaire. L'énoncé précis est le suivant.

Proposition 1.80 (sommabilité et combinaison linéaire). *Soient $(u_i)_{i \in I}$ et $(u'_i)_{i \in I}$ deux familles sommables dans le même espace vectoriel normé E , ayant même ensemble d'indexation I . Notons S et S' les sommes respectives. Alors pour tous scalaires a et b , la famille $(au_i + bu'_i)_{i \in I}$ est sommable dans E , de somme $aS + bS'$.*

Preuve. Il est clair qu'il suffit de traiter les deux cas particuliers de $(u_i + u'_i)_{i \in I}$ et de $(au_i)_{i \in I}$. Nous détaillons seulement le premier, laissant l'autre au lecteur. Par hypothèse de sommabilité des familles $(u_i)_{i \in I}$ et $(u'_i)_{i \in I}$, on a pour tout $\varepsilon > 0$, deux parties finies J et J' de I telles que pour tous K, K' finis tels que $J \subset K \subset I$ et $J' \subset K' \subset I$, $\|S - S_K\| < \varepsilon/2$ et $\|S' - S'_{K'}\| < \varepsilon/2$. Ces deux inégalités sont vraies en particulier pour $K = K' = L$ où L est n'importe quelle partie finie de I contenant $J \cup J'$. Par inégalité triangulaire, on a alors $\|(S + S') - (S_L + S'_L)\| < \varepsilon$. Par les propriétés de l'addition dans E et la finitude de L , on a

$$S_L + S'_L = \sum_{i \in L} u_i + \sum_{i \in L} u'_i = \sum_{i \in L} (u_i + u'_i).$$

On a donc bien vérifié que $(u_i + u'_i)_{i \in I}$ est sommable et a pour somme $S + S'$. \square

La sommabilité est préservée par permutation sur les indices.

Proposition 1.81 (invariance par permutation). *Soit $(u_i)_{i \in I}$ une famille sommable de somme S . Alors pour toute bijection $\varphi : I \rightarrow I$, la famille $(u_{\varphi(i)})_{i \in I}$ est sommable de somme S .*

Preuve. Il s'agit de montrer que $(v_i)_{i \in I}$ est sommable de même somme S que $(u_i)_{i \in I}$, les v_i étant définis par $v_i := u_{\varphi(i)}$. L'hypothèse de sommabilité de $(u_i)_{i \in I}$ s'écrit

$$\forall \varepsilon > 0, \exists J \text{ fini } \subset I, \forall K \text{ fini } \supset J, \quad \|S_K - S\| < \varepsilon \quad (1.27)$$

Posons $J' := \varphi^{-1}(J) = \{\varphi^{-1}(i); i \in J\}$. L'ensemble J' est fini car en bijection avec l'ensemble fini J par φ^{-1} . Pour tout K' fini contenant J' , l'ensemble fini $\varphi(K')$ contient $\varphi(J')$ et ce dernier ensemble est égal à J . On a donc en appliquant (1.27) avec $\varphi(K')$ au lieu de K , $\|S_{\varphi(K')} - S\| < \varepsilon$ et ceci est vrai pour tout K' fini contenant J' . D'autre part

$$S_{\varphi(K')} = \sum_{\ell \in \varphi(K')} u_\ell = \sum_{i \in K'} u_{\varphi(i)} = \sum_{i \in K'} v_i.$$

Nous avons donc montré que

$$\forall \varepsilon > 0, \exists J' \text{ fini } \subset I, \forall K' \text{ fini } \supset J', \quad \left\| \sum_{i \in K'} v_i - S \right\| < \varepsilon,$$

autrement dit que $(v_i)_{i \in I}$ est sommable de somme S . \square

Remarque 1.82 (sommabilité d'une série). Plaçons nous un instant dans le cas particulier $I = \mathbb{N}$. Nous savions déjà, cf. l'implication (1.25) \Rightarrow (1.23) vue en introduction, que la sommabilité de $(u_k)_{k \in \mathbb{N}}$ implique la convergence de la série $\sum_{k=0}^{+\infty} u_k$. La proposition 1.81 nous apprend en plus que cette convergence est nécessairement commutative. Dans le cas où les u_k sont réels ou complexes, on peut donc dire que la sommabilité de $(u_k)_{k \in \mathbb{N}}$ implique la convergence commutative et absolue de la série $\sum_{k=0}^{+\infty} u_k$, cf. théorème 1.76. Ainsi comme nous l'avions annoncé en introduction, l'implication « (1.23) \Rightarrow (1.25) » est fausse. Par exemple la série de terme général $u_k = (-1)^k / (k+1)$ est convergente, mais la famille $(u_k)_{k \in \mathbb{N}}$ n'est pas sommable.

Il est facile de donner une caractérisation de la sommabilité pour les familles de réels positifs.

Proposition 1.83. Soit $(u_i)_{i \in I}$ une famille d'éléments de \mathbb{R}_+ .

a) Si $(u_i)_{i \in I}$ est telle que

$$M := \sup_{K \text{ fini } \subset I} S_K < +\infty, \quad (1.28)$$

alors elle est sommable de somme M .

b) Réciproquement, si $(u_i)_{i \in I}$ est sommable de somme S , on a

$$S = \sup_{K \text{ fini } \subset I} S_K, \quad (1.29)$$

donc ce sup est fini.

Preuve du a). Soit $\varepsilon > 0$. Comme M est fini, $M - \varepsilon$ est strictement inférieur à M . Par minimalité du supremum M parmi tous les majorants de l'ensemble $\{S_K; K \text{ fini } \subset I\}$, il existe une partie finie J de I telle que

$$M - \varepsilon < S_J \leq M.$$

De plus pour toute partie finie K de I , contenant J ,

$$M - \varepsilon < S_J \leq S_K \leq M. \quad (1.30)$$

En effet, $S_K - S_J = S_{K \setminus J}$ est une somme de réels positifs, donc positive, ce qui justifie la deuxième inégalité dans (1.30). La troisième découle de la définition de M . Comme (1.30) implique $|M - S_K| < \varepsilon$, nous avons ainsi établi que $\{u_i, i \in I\}$ est sommable et de somme M . \square

Preuve du b). Supposons maintenant que $(u_i)_{i \in I}$ est sommable dans \mathbb{R}_+ , de somme S . Notons M le supremum défini par (1.29), qu'il soit fini ou infini. On commence par remarquer que pour toute partie finie J de I , on a l'égalité suivante dans \mathbb{R}_+ :

$$M := \sup_{L \text{ fini } \subset I} S_L = \sup_{K \text{ fini}, J \subset K \subset I} S_K. \quad (1.31)$$

En effet la positivité des u_i et l'inclusion de L fini dans l'ensemble fini $L \cup J$ contenant J , nous donnent pour tout L fini l'inégalité $S_L \leq S_{L \cup J}$. On en déduit l'inégalité des suprema

$$M \leq \sup_{K \text{ fini}, J \subset K \subset I} S_K.$$

L'inégalité inverse étant évidente, (1.31) est vérifiée.

Soit $\varepsilon > 0$ quelconque. Il existe une partie finie $J = J(\varepsilon)$ de I telle que pour tout K fini vérifiant $J \subset K \subset I$, $S_K \in]S - \varepsilon, S + \varepsilon[$. Ceci implique

$$\sup_{K \text{ fini}, J \subset K \subset I} S_K \in [S - \varepsilon, S + \varepsilon]$$

et compte-tenu de (1.31) appliqué avec $J = J(\varepsilon)$, $M \in [S - \varepsilon, S + \varepsilon]$. Ceci nous montre déjà que M est fini. On a de plus $|S - M| \leq \varepsilon$ pour tout $\varepsilon > 0$, d'où $M = S$. \square

Remarque 1.84. Il résulte de la proposition 1.83 qu'une famille de réels positifs u_i est non sommable si et seulement si le supremum M des sommes S_K pour K partie finie de I vaut $+\infty$. La situation est analogue à celle des séries à termes positifs qui ne peuvent diverger que si la suite des sommes partielles tend vers $+\infty$. On pourra donc toujours donner un sens à l'expression $\sum_{i \in I} u_i$, considérée comme élément de $\overline{\mathbb{R}}_+$, en posant

$$\sum_{i \in I} u_i := \sup_{K \text{ fini}, K \subset I} S_K = \begin{cases} S \in \mathbb{R}_+, & \text{si } (u_i)_{i \in I} \text{ est sommable de somme } S, \\ +\infty, & \text{sinon.} \end{cases} \quad (1.32)$$

Cette situation est particulière aux familles à termes positifs et c'est *le seul cas* où nous donnerons un sens à $\sum_{i \in I} u_i$ sans que la famille soit forcément sommable.

Proposition 1.85 (sommabilité et dénombrabilité). *Si $(u_i)_{i \in I}$ est sommable, l'ensemble d'indices $I' := \{i \in I; u_i \neq 0\}$ est au plus dénombrable.*

Preuve. Pour chaque $n \in \mathbb{N}^*$, appliquons (1.26) avec $\varepsilon = 1/n$ et choisissons l'un¹⁸ des ensembles finis J donnés par (1.26), que nous noterons J_n . Posons

$$H := \bigcup_{n \in \mathbb{N}^*} J_n.$$

L'ensemble H est au plus dénombrable comme réunion dénombrable d'ensembles au plus dénombrables (prop. 1.40). Si $H = I$, alors I' qui est inclus dans H est donc au plus dénombrable. En dehors de ce cas trivial, $I \setminus H$ n'est pas vide. Soit i_0 un élément quelconque de $I \setminus H$. En appliquant (1.26) avec $\varepsilon = 1/n$, $J = J_n$ et chacun des deux ensembles finis $K = J_n$, $K' = J_n \cup \{i_0\}$, on obtient les deux inégalités :

$$\|S_{J_n} - S\| \leq \frac{1}{n} \quad \text{et} \quad \|u_{i_0} + S_{J_n} - S\| \leq \frac{1}{n},$$

18. La propriété (1.26) nous assure qu'il existe au moins un J pour ε donné, mais ne dit rien sur l'unicité.

vraies pour tout $n \in \mathbb{N}^*$. En écrivant $u_{i_0} = (u_{i_0} + S_{J_n} - S) + (S - S_{J_n})$, l'inégalité triangulaire nous donne

$$\forall n \geq 1, \quad \|u_{i_0}\| \leq \frac{2}{n}.$$

Faisant tendre n vers l'infini (noter que i_0 est fixé et ne dépend pas de n), on en déduit $\|u_{i_0}\| = 0$, d'où $u_{i_0} = 0$. Ce raisonnement étant valable pour n'importe quel $i_0 \in I \setminus H$, on en déduit que I' est inclus dans H . Par conséquent, I' est au plus dénombrable. \square

Nous allons voir maintenant que pour I dénombrable, sommabilité de $(u_i)_{i \in I}$ et convergence commutative de la série associée *via* une numérotation de I sont équivalentes.

Théorème 1.86. *Si I est dénombrable, les propriétés suivantes sont équivalentes :*

- a) $(u_i)_{i \in I}$ est sommable de somme S ;
- b) pour toute bijection $f : \mathbb{N} \rightarrow I$, la série $\sum_{k=0}^{+\infty} u_{f(k)}$ converge et a pour somme S .

Avant de démontrer ce théorème, remarquons que la propriété b) implique que la série de terme général $v_k := u_{f(k)}$ est *commutativement* convergente et ceci pour tout choix d'une bijection $f : \mathbb{N} \rightarrow I$. En effet, fixons une telle bijection et donc la suite $(v_k)_{k \in \mathbb{N}}$ correspondante. Soit σ une bijection quelconque de \mathbb{N} sur \mathbb{N} . On a pour tout $k \in \mathbb{N}$, $v_{\sigma(k)} = u_{f(\sigma(k))} = u_{g(k)}$, où $g := f \circ \sigma$ est une bijection de \mathbb{N} sur I . Par b), la série $\sum_{k=0}^{+\infty} u_{g(k)}$ converge et a pour somme S . Or cette série est exactement $\sum_{k=0}^{+\infty} v_{\sigma(k)}$. Par arbitrarité de la bijection σ on en déduit que $\sum_{k=0}^{+\infty} v_k$ est commutativement convergente.

Réciproquement, supposons qu'il existe au moins une bijection $h : \mathbb{N} \rightarrow I$ telle que la série $\sum_{k=0}^{+\infty} u_{h(k)}$ converge commutativement et ait pour somme S . Alors $(u_i)_{i \in I}$ vérifie la propriété b). En effet toute bijection $f : \mathbb{N} \rightarrow I$ peut s'écrire $f = (h \circ h^{-1}) \circ f = h \circ (h^{-1} \circ f) = h \circ \sigma$, où $\sigma := h^{-1} \circ f$ est une bijection de \mathbb{N} sur \mathbb{N} . Alors pour tout $k \in \mathbb{N}$, $u_{f(k)} = u_{h(\sigma(k))} = v_{\sigma(k)}$, où l'on a noté $v_k := u_{h(k)}$ le terme général de la série $\sum_{k=0}^{+\infty} u_{h(k)}$. La série $\sum_{k=0}^{+\infty} v_k$ étant commutativement convergente de somme S , on en déduit que la série de terme général $u_{f(k)} = v_{\sigma(k)}$ converge et a pour somme S .

Preuve de a) \Rightarrow b). Avec les J_n choisis comme dans la preuve de la proposition 1.85, on a pour toute partie finie K de I contenant J_n , $\|S_K - S\| \leq 1/n$. Puisque f est une bijection de \mathbb{N} sur I et J_n une partie finie de I , il existe pour chaque $n \in \mathbb{N}^*$ un entier $k_0(n)$ tel que

$$f(\{0, 1, 2, \dots, k_0(n)\}) \supset J_n.$$

Il suffit de prendre pour cela, $k_0(n) := \max\{f^{-1}(i); i \in J_n\}$. Alors pour tout entier $l \geq k_0(n)$, $K := f(\{0, 1, 2, \dots, l\})$ est une partie finie de I , contenant J_n et donc $\|S - S_K\| \leq 1/n$. Comme $S_K = \sum_{k=0}^l u_{f(k)}$, nous venons ainsi d'établir que :

$$\forall n \in \mathbb{N}^*, \exists k_0(n), \quad \forall l \geq k_0(n), \quad \left\| \sum_{k=0}^l u_{f(k)} - S \right\| \leq \frac{1}{n},$$

ce qui équivaut à la convergence vers S de la série de terme général $u_{f(k)}$ (c'est juste ce que l'on obtient en discrétisant le « $\forall \varepsilon > 0$ » en le remplaçant par $\varepsilon = 1/n$ dans la définition de cette convergence). \square

Preuve de $b) \Rightarrow a)$. Nous allons prouver cette implication en démontrant sa contraposée : non $a) \Rightarrow$ non $b)$. La négation de $a)$ s'écrit

$$\exists \varepsilon > 0, \forall J \text{ partie finie de } I, \exists K \text{ fini, } J \subset K \subset I \text{ et } \|S - S_K\| > \varepsilon. \quad (1.33)$$

Fixons une première bijection $\varphi : \mathbb{N} \rightarrow I$. Posons $k_0 := 0$ et $J_0 := \{\varphi(0)\}$. Par (1.33) il existe une partie finie K_0 de I , contenant J_0 et telle que $\|S - S_{K_0}\| > \varepsilon$. On peut maintenant choisir un entier $k_1 > k_0$ tel que $J_1 := \varphi(\{0, 1, \dots, k_1\})$ contienne *strictement* K_0 , il suffit pour cela de prendre $k_1 = 1 + \max\{\varphi^{-1}(i); i \in K_0\}$. Une nouvelle invocation de (1.33) nous fournit une partie finie K_1 de I , contenant J_1 et telle que $\|S - S_{K_1}\| > \varepsilon$. Nous venons ainsi d'amorcer une récurrence qui basée sur l'utilisation itérée de (1.33), nous permet de construire une suite strictement croissante d'entiers (k_n) , deux suites (J_n) et (K_n) de parties finies de I , vérifiant :

$$\forall n \in \mathbb{N}, \quad \varphi(\{0, 1, \dots, k_n\}) = J_n \subset K_n \subsetneq J_{n+1} \text{ et } \|S - S_{K_n}\| > \varepsilon. \quad (1.34)$$

Par cette construction, la suite (K_n) est strictement croissante pour l'inclusion. La réunion de cette suite est exactement I . En effet elle est évidemment incluse dans I puisque chaque K_n est une partie de I . Dans l'autre sens, si i est un élément quelconque de I , $\varphi^{-1}(i)$ est un entier qui est majoré par k_n pour n assez grand (la suite *strictement* croissante d'entiers (k_n) tend vers l'infini). Par construction de J_n , l'inégalité $\varphi^{-1}(i) \leq k_n$ implique l'appartenance de i à J_n , donc aussi à K_n . Ainsi un élément quelconque i de I appartient toujours à au moins un K_n , donc $I \subset \bigcup_{n \in \mathbb{N}} K_n$ et finalement $I = \bigcup_{n \in \mathbb{N}} K_n$.

Posant maintenant

$$I_0 := K_0 \quad \text{et} \quad \forall n \in \mathbb{N}^*, \quad I_n := K_n \setminus K_{n-1},$$

il est clair qu'aucun des I_n n'est vide (stricte croissance de (K_n)), qu'ils sont deux à deux disjoints (pour la même raison) et que leur réunion est I . La famille $\{I_n; n \in \mathbb{N}\}$ constitue donc une partition de I en sous-ensembles finis. Notons maintenant

$$\forall n \in \mathbb{N}, \quad m_n := \text{card } K_n - 1.$$

Comme (K_n) est strictement croissante pour l'inclusion, la suite d'entiers (m_n) est strictement croissante. On obtient alors une partition $\{A_n; n \in \mathbb{N}\}$ de \mathbb{N} en posant :

$$A_0 := \{0, \dots, m_0\}, \quad \text{et} \quad \forall n \in \mathbb{N}^*, \quad A_n := \{m_{n-1} + 1, \dots, m_n\}.$$

Les A_n sont finis, $\text{card } A_0 = m_0 + 1 = \text{card } K_0 = \text{card } I_0$ et pour $n \geq 1$,

$$\text{card } A_n = m_n - m_{n-1} = \text{card } K_n - \text{card } K_{n-1} = \text{card } I_n.$$

Dire qu' A_n et I_n ont même cardinal c'est dire précisément qu'il existe une bijection $f_n : A_n \rightarrow I_n$. En « recollant » ces bijections f_n entre ensembles finis, on construit une application $f : \mathbb{N} \rightarrow I$: tout $k \in \mathbb{N}$ appartient à un unique A_n (puisque les A_n partitionnent \mathbb{N}) et on pose alors

$$f(k) := f_n(k), \quad (k \in A_n).$$

Vérifions que f ainsi définie est une bijection. Tout $i \in I$ appartient à un I_n (puisque les I_n partitionnent I). Il a alors pour antécédent par f l'entier $f_n^{-1}(i) \in A_n$. Ceci montre que chaque $i \in I$ a au moins un antécédent par f , donc que f est surjective. Pour vérifier l'injectivité, soient k et l deux entiers distincts. Ou bien ils sont dans le même A_n et alors $f(k) = f_n(k) \neq f_n(l) = f(l)$ par injectivité de f_n . Ou bien $k \in A_n$ et $l \in A_{n'}$ avec $n \neq n'$. Alors $f(k) \in I_n$ et $f(l) \in I_{n'}$ et comme I_n et $I_{n'}$ sont disjoints, $f(k)$ et $f(l)$ sont forcément distincts. Nous avons finalement construit une suite strictement croissante d'entiers m_n (donc tendant vers l'infini) et une bijection $f : \mathbb{N} \rightarrow I$ telles que

$$\forall n \in \mathbb{N}, \quad f(\{0, \dots, m_n\}) = K_n.$$

Au vu de (1.34), nous avons ainsi

$$\forall n \in \mathbb{N}, \quad \left\| \sum_{k=0}^{m_n} u_{f(k)} - S \right\| > \varepsilon,$$

ce qui interdit la convergence vers S de la série de terme général $u_{f(k)}$. L'existence d'une bijection f pour laquelle la série de terme général $u_{f(k)}$ ne converge pas vers S est précisément la propriété « non b) ». \square

Le prochain théorème est l'analogue du critère de Cauchy pour les familles sommables. Il est surtout important par ses corollaires. Il s'applique avec E espace vectoriel normé *complet* donc en particulier avec $E = \mathbb{R}$ ou \mathbb{C} .

Théorème 1.87 (critère de Cauchy pour la sommabilité). *Soit E un espace vectoriel normé complet. La famille $(u_i)_{i \in I}$ d'éléments de E est sommable si et seulement si*

$$\forall \varepsilon > 0, \exists J \text{ fini} \subset I, \forall H \text{ fini} \subset I \setminus J, \quad \|S_H\| < \varepsilon. \quad (1.35)$$

Avant d'attaquer la preuve du théorème, faisons un peu de recherche en paternité en comparant (1.35) avec le critère de Cauchy classique pour une série. Supposons donc pour un instant que $I = \mathbb{N}$. Le critère de Cauchy classique pour la série de terme général u_k s'écrit :

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}, \forall p, q \geq n_0, \quad \|S_p - S_q\| < \varepsilon. \quad (1.36)$$

On peut toujours par confort d'écriture supposer $p \leq q$. Définissons les ensembles d'entiers consécutifs $H_{p,q} := \{p, \dots, q\} = [p, q] \cap \mathbb{N}$. Avec cette notation, on peut réécrire (1.36) sous la forme équivalente

$$\forall \varepsilon > 0, \exists H_{0,n_0} \subset \mathbb{N}, \forall H_{p,q} \subset \mathbb{N} \setminus H_{0,n_0}, \quad \|S_{H_{p,q}}\| < \varepsilon. \quad (1.37)$$

Cette écriture établit la filiation de (1.35) à partir de (1.36) et nous permet de voir que (1.35) implique (1.36). La réciproque est fautive en raison de la remarque 1.82 et du théorème 1.87. Par exemple la série de terme général $u_k = (-1)^k/(k+1)$ est convergente donc vérifie le critère de Cauchy (1.36). Pourtant elle n'est pas sommable, donc ne vérifie pas (1.35).

Preuve de « sommabilité \Rightarrow (1.35) ». Supposons d'abord $(u_i)_{i \in I}$ sommable. Pour tout $\varepsilon > 0$, il existe une partie finie J de I telle que pour toute partie finie K de I contenant J , $\|S_K - S\| < \varepsilon/2$. En appliquant cette inégalité avec $K = J$, puis $K = J \cup H$, pour H partie finie quelconque de $I \setminus J$, on obtient :

$$\|S_J - S\| < \frac{\varepsilon}{2} \quad \text{et} \quad \|S_{J \cup H} - S\| < \frac{\varepsilon}{2}.$$

Les ensembles d'indices J et H étant disjoints et finis, $S_{J \cup H} = S_J + S_H$. En écrivant $S_H = (S_{J \cup H} - S) - (S_J - S)$, l'inégalité triangulaire nous donne $\|S_H\| < \varepsilon$. Ceci établit (1.35). Notons que nous n'avons pas utilisé la complétude de E dans cette partie. \square

Preuve de « (1.35) \Rightarrow sommabilité ». Pour établir la réciproque, supposons que la famille $(u_i)_{i \in I}$ d'éléments de E vérifie (1.35).

Vérifions d'abord que l'ensemble d'indices $I' := \{i \in I; u_i \neq 0\}$ est au plus dénombrable¹⁹. En appliquant (1.35) avec $\varepsilon = 1/n$, pour tout $n \geq 1$, on obtient une suite de parties finies J_n de I telles que si $H \subset I \setminus J_n$ est fini, $\|S_H\| < 1/n$. Ceci est vrai en particulier avec $H = \{i\}$, pour $i \notin J_n$. Donc si $i \notin J_n$, $\|u_i\| < 1/n$. Ainsi si $i \notin \bigcup_{n \geq 1} J_n$, autrement dit si $i \in \bigcap_{n \geq 1} (I \setminus J_n)$, on a $\|u_i\| < 1/n$ pour tout $n \geq 1$, d'où $\|u_i\| = 0$, puis $u_i = 0$. On en déduit que I' est inclus dans $\bigcup_{n \geq 1} J_n$. Ce dernier ensemble est au plus dénombrable comme union dénombrable d'ensembles finis, donc I' est lui-même au plus dénombrable.

Le cas I' fini est immédiat : on pose $J = I'$, $S := S_{I'}$ (bien défini comme somme d'un nombre fini de vecteurs de E). Pour tout K fini contenant I' , $S_K = S_{I'} + S_{K \setminus I'} = S_{I'}$ puisque pour les $i \in K \setminus I'$, $u_i = 0$. On a donc $S_K = S$ pour tout $K \supset I'$, donc *a fortiori* $\|S_K - S\| < \varepsilon$ pour tout $\varepsilon > 0$ (ici on a même J indépendant de ε). Ainsi $(u_i)_{i \in I}$ est sommable de somme S .

Passons au cas I' infini (donc ici dénombrable). Fixons une bijection $f : \mathbb{N} \rightarrow I'$ et définissons pour $k \in \mathbb{N}$, $v_k := u_{f(k)}$. Notons $T_n = v_0 + \dots + v_n$ la n -ième somme partielle de la série de vecteurs $\sum_{k=0}^{+\infty} v_k$. Vérifions que la suite $(T_n)_{n \in \mathbb{N}}$ satisfait au critère de Cauchy classique dans E . Pour $\varepsilon > 0$, choisissons l'un des ensembles J associés à ε par (1.35). Il est clair que l'on peut remplacer J par $J' := J \cap I'$ puisque pour les $i \notin I'$, $u_i = 0$. J' est fini donc par surjectivité de f , il existe un entier m_ε tel que $f(\{0, \dots, m_\varepsilon\})$ recouvre J' . Alors pour $m_\varepsilon < m \leq n$, $H := f(\{m, \dots, n\})$ est fini et disjoint de J' , donc par (1.35), $\|S_H\| < \varepsilon$, ce qui peut encore s'écrire $\|T_n - T_m\| < \varepsilon$. Ainsi la suite T_n est de Cauchy dans l'espace *complet* E . Elle est donc convergente. Notons S sa limite.

Nous allons montrer pour finir que $(u_i)_{i \in I}$ est sommable de somme S . En conservant les notations qui ont servi à vérifier que (T_n) est de Cauchy et en choisissant pour tout $n \geq m_\varepsilon$, $H_n := f(\{m_\varepsilon, \dots, n\})$, on a $\|T_n - T_{m_\varepsilon}\| < \varepsilon$. Compte-tenu de la convergence de T_n vers S , on en déduit en faisant tendre n vers l'infini que $\|S - T_{m_\varepsilon}\| \leq \varepsilon$. Notons $J'' := f(\{0, \dots, m_\varepsilon\})$ et rappelons que $J'' \supset J' = I' \cap J$, où J est associé à ε par (1.35).

19. Attention, ceci ressemble à la proposition 1.85, mais ici on suppose (1.35) vraie au lieu de la sommabilité de $(u_i)_{i \in I}$.

D'autre part $S_{J''} = T_{m_\varepsilon}$, d'où $\|S_{J''} - S\| \leq \varepsilon$. Soit K une partie finie de I , contenant J'' . En écrivant que

$$S_K = S_{J''} + S_{K \setminus J''} = S + (S_{J''} - S) + S_{K \setminus J''},$$

on obtient par inégalité triangulaire

$$\|S_K - S\| \leq \|S_{J''} - S\| + \|S_{K \setminus J''}\| \leq \varepsilon + \|S_{K \setminus J''}\|. \quad (1.38)$$

Puisque $K \setminus J''$ est fini et disjoint de J' (car $J' \subset J''$), on a $\|S_{K \setminus J''}\| < \varepsilon$. En reportant ceci dans (1.38), on voit que pour tout K fini contenant J'' , $\|S_K - S\| < 2\varepsilon$. Ce raisonnement étant valable pour tout $\varepsilon > 0$ (avec J'' dépendant de ε), on conclut que $(u_i)_{i \in I}$ est sommable de somme S . \square

Corollaire 1.88. *Soit $(u_i)_{i \in I}$ une famille d'éléments de l'espace vectoriel normé complet E . Si $(\|u_i\|)_{i \in I}$ est sommable dans \mathbb{R}_+ , alors $(u_i)_{i \in I}$ est sommable dans E . On dit dans ce cas qu'elle est normalement sommable ou si $E = \mathbb{R}$ ou \mathbb{C} , absolument sommable.*

Preuve. En appliquant la partie « sommabilité \Rightarrow (1.35) » du théorème 1.87 à la famille $(\|u_i\|)_{i \in I}$, ici l'espace vectoriel est \mathbb{R} , on obtient :

$$\forall \varepsilon > 0, \exists J \text{ fini } \subset I, \forall H \text{ fini } \subset I \setminus J, \sum_{i \in H} \|u_i\| < \varepsilon.$$

Par inégalité triangulaire on en déduit :

$$\|S_H\| \leq \sum_{i \in H} \|u_i\| < \varepsilon.$$

La partie « (1.35) \Rightarrow sommabilité » du théorème 1.87 appliquée cette fois à la famille $(u_i)_{i \in I}$ dans l'espace E nous fournit la sommabilité de cette dernière. \square

Nous sommes maintenant en mesure de faire un point définitif sur les différents modes de sommation étudiés pour une famille dans \mathbb{R} ou \mathbb{C} .

Proposition 1.89. *Pour une famille $(u_i)_{i \in I}$ avec I dénombrable et les u_i dans \mathbb{R} ou \mathbb{C} , les trois propriétés sommabilité, convergence commutative et sommabilité absolue sont équivalentes.*

Preuve. Cette équivalence résulte du théorème 1.86, du corollaire 1.88 et des théorèmes 1.68 et 1.76. \square

Le prochain corollaire nous sera utile pour établir le théorème de sommation par paquets.

Corollaire 1.90 (du th. 1.87). *Soit $(u_i)_{i \in I}$ une famille sommable d'éléments de l'espace vectoriel normé complet E .*

- a) *Pour toute partie L de I , finie ou non, $(u_i)_{i \in L}$ est sommable. On notera S_L sa somme.*

- b) Pour tout $\varepsilon > 0$, soit J une partie finie de I telle que pour toute partie finie $K \supset J$ de I , $\|S - S_K\| < \varepsilon$. Alors on a $\|S - S_L\| \leq \varepsilon$ pour toute partie $L \supset J$ de I , finie ou non.
- c) Si L_1, \dots, L_n sont des parties deux à deux disjointes de I , de réunion L , alors $S_L = S_{L_1} + \dots + S_{L_n}$.

Preuve du a). Utilisons le critère de Cauchy (1.35).

$$\forall \varepsilon > 0, \exists J \text{ fini } \subset I, \forall K \text{ fini } \subset I \setminus J, \quad \|S_K\| < \varepsilon.$$

Ceci est vrai en particulier pour toute partie finie K de L disjointe de J , autrement dit pour toute partie finie K de L disjointe de $J' := J \cap L$. Nous avons donc

$$\forall \varepsilon > 0, \exists J' \text{ fini } \subset L, \forall K \text{ fini } \subset L \setminus J', \quad \|S_K\| < \varepsilon.$$

Ainsi $(u_i)_{i \in L}$ est sommable par le critère de Cauchy (1.35). \square

Preuve du b). Par le a), la famille $(u_i)_{i \in L}$ est sommable. Il existe donc pour tout $n \in \mathbb{N}^*$ une partie finie J_n de L telle que pour tout ensemble fini K_n vérifiant $J_n \subset K_n \subset L$, $\|S_L - S_{K_n}\| < 1/n$. D'autre part on a pour tout K fini vérifiant $J \subset K \subset L \subset I$, $\|S - S_K\| < \varepsilon$. Pour l'ensemble fini $K'_n := K \cup K_n$ qui contient J_n et J , on a donc à la fois $\|S - S_{K'_n}\| < \varepsilon$ et $\|S_L - S_{K'_n}\| < 1/n$. Par inégalité triangulaire, on en déduit $\|S - S_L\| < \varepsilon + 1/n$. Cette inégalité étant vraie pour tout n et son premier membre ne dépendant pas de n , on en déduit en faisant tendre n vers $+\infty$ avec ε fixé, que $\|S - S_L\| \leq \varepsilon$. \square

Preuve du c). Par le a), on sait que les n sous-familles $(u_i)_{i \in L_l}$, $l = 1, \dots, n$ sont sommables. Ainsi pour tout $\varepsilon > 0$, tout $l = 1, \dots, n$, il existe J_l partie finie de L_l telle que pour tout K_l fini vérifiant $J_l \subset K_l \subset L_l$, $\|S_{K_l} - S_{L_l}\| < \varepsilon/n$. Posons $J = J_1 \cup \dots \cup J_n$ et soit K une partie finie quelconque de $L = L_1 \cup \dots \cup L_n$, contenant J . En prenant $K_l := K \cap L_l$, les K_l des ensembles finis deux à deux disjoints et de réunion K , d'où $S_K = S_{K_1} + \dots + S_{K_n}$. On a alors par inégalité triangulaire :

$$\left\| S_K - \sum_{l=1}^n S_{L_l} \right\| = \left\| \sum_{l=1}^n (S_{K_l} - S_{L_l}) \right\| \leq \sum_{l=1}^n \|S_{K_l} - S_{L_l}\| < n \frac{\varepsilon}{n} = \varepsilon.$$

On a ainsi vérifié que $(u_i)_{i \in L}$ est sommable de somme $\sum_{l=1}^n S_{L_l}$. \square

Théorème 1.91 (de sommation par paquets). Soit $(u_i)_{i \in I}$ une famille d'éléments de l'espace vectoriel normé complet E . On suppose que $I = \bigcup_{\alpha \in A} I_\alpha$, les I_α étant non vides et deux à deux disjoints. Si $(u_i)_{i \in I}$ est sommable de somme S , alors chacune des familles $(u_i)_{i \in I_\alpha}$ est sommable et en notant S_{I_α} sa somme, la famille $(S_{I_\alpha})_{\alpha \in A}$ est sommable de somme S . Autrement dit on a la formule de « sommation par paquets » :

$$\sum_{i \in I} u_i = \sum_{\alpha \in A} \left(\sum_{i \in I_\alpha} u_i \right). \quad (1.39)$$

Preuve. Pour tout $\alpha \in A$, la sous-famille $(u_i)_{i \in I_\alpha}$ est sommable par le corollaire 1.90 a). Pour toute partie finie C de A , nous pouvons alors définir $T_C := \sum_{\alpha \in C} S_{I_\alpha}$, somme d'un nombre fini de vecteurs de E . De même, $I_C := \bigcup_{\alpha \in C} I_\alpha$ étant une partie de I , la famille $(u_i)_{i \in I_C}$ est sommable et nous notons sa somme S_{I_C} . Par le c) du corollaire 1.90, nous avons $T_C = S_{I_C}$.

Puisque $(u_i)_{i \in I}$ est sommable de somme S , on dispose pour tout $\varepsilon > 0$ d'un ensemble fini $J \subset I$ tel que pour tout K fini vérifiant $J \subset K \subset I$, $\|S - S_K\| < \varepsilon$. Comme J est fini et $I = \bigcup_{\alpha \in A} I_\alpha$, il existe une partie finie B de A telle que I_B contienne J . Pour tout C fini inclus dans A et contenant B , $I_C \supset I_B \supset J$, donc par le corollaire 1.90 b) (même avec I_C infini) on a $\|S - S_{I_C}\| \leq \varepsilon$. Compte-tenu de l'égalité $S_{I_C} = T_C$, on a ainsi montré que

$$\forall \varepsilon > 0, \exists B \text{ fini } \subset A, \forall C \text{ fini tel que } B \subset C \subset A, \quad \|S - T_C\| \leq \varepsilon.$$

Autrement dit, $(S_{I_\alpha})_{\alpha \in A}$ est sommable de somme S . \square

Remarque 1.92. Il importe de noter que la sommabilité de la famille $(S_{I_\alpha})_{\alpha \in A}$ n'implique pas celle de la famille $(u_i)_{i \in I}$. Voici un contre exemple avec $E = \mathbb{R}$: on prend $I = \mathbb{Z}$, $u_i = i$, $A = \mathbb{N}$ et pour tout $k \in A$, $I_k = \{-k, k\}$. Alors pour tout $k \in A$, $S_{I_k} = 0$ et la famille $(S_{I_k})_{k \in \mathbb{N}}$ est sommable de somme 0. Par contre la famille $(u_i)_{i \in \mathbb{Z}}$ n'est pas sommable puisque $\sum_{i \in \mathbb{Z}} |i| = +\infty$.

Les familles de réels positifs constituent à nouveau un cas particulier pour la sommation par paquets (et une exception à la remarque 1.92). Il est naturel ici d'élargir un peu le problème en considérant les familles d'éléments de $\overline{\mathbb{R}}_+$. On prolonge l'addition de \mathbb{R}_+ à $\overline{\mathbb{R}}_+$ en posant

$$\forall x \in \mathbb{R}_+, x + (+\infty) = (+\infty) + x = +\infty \quad \text{et} \quad (+\infty) + (+\infty) = +\infty.$$

Avec cette convention, on peut toujours définir la somme de n'importe quelle famille d'éléments de $\overline{\mathbb{R}}_+$ en utilisant (1.32).

Théorème 1.93 (de sommation par paquets dans $\overline{\mathbb{R}}_+$). *Toute famille $(u_i)_{i \in I}$ d'éléments $\overline{\mathbb{R}}_+$ vérifie la formule de sommation par paquets (1.39) interprétée comme égalité dans $\overline{\mathbb{R}}_+$.*

Preuve. Si $M := \sum_{i \in I} u_i < +\infty$, nécessairement tous les u_i sont finis et la famille $(u_i)_{i \in I}$ est sommable (prop 1.83). Il n'y a alors rien à démontrer puisqu'on est dans le champ d'application du théorème 1.91. La démonstration générale par les suprema que l'on présente maintenant englobe ce cas et a l'avantage d'éviter de discuter suivant la finitude ou non des u_i ou des S_{I_α} . Notons

$$M := \sup_{K \text{ fini } \subset I} S_K \quad \text{et} \quad M' := \sup_{B \text{ fini } \subset A} S_{I_B}.$$

Avec ces notations, l'égalité (1.39) s'écrit $M = M'$. Nous allons montrer que $M' \leq M$ et que $M \leq M'$. Pour la première inégalité, il suffit de remarquer que pour tout B fini inclus dans A ,

$$S_{I_B} = \sup_{L \text{ fini } \subset I_B} S_L, \tag{1.40}$$

d'où en majorant S_L par M , $S_{I_B} \leq M$, puis en prenant le supremum pour tout B fini inclus dans A , $M' \leq M$. Dans l'autre sens, soit K une partie finie quelconque de I . Il existe une partie finie B de A telle que $I_B = \cup_{\alpha \in B} I_\alpha$ recouvre K . D'après (1.40) on a alors $S_K \leq S_{I_B}$. En majorant S_{I_B} par M' , on voit que pour tout K fini inclus dans I , $S_K \leq M'$. En prenant le supremum sur K , on obtient $M \leq M'$. \square

Théorème 1.94. *Soit $(u_i)_{i \in I}$ une famille d'éléments de l'espace vectoriel normé complet E . On suppose que $I = \cup_{\alpha \in A} I_\alpha$, les I_α étant non vides et deux à deux disjoints. On pose $T_{I_\alpha} := \sum_{i \in I_\alpha} \|u_i\|$. La famille $(u_i)_{i \in I}$ est normalement sommable dans E si et seulement si $\sum_{\alpha \in A} T_{I_\alpha} < +\infty$. Elle vérifie alors la formule de sommation par paquets (1.39).*

Preuve. La sommabilité dans \mathbb{R}_+ de $(\|u_i\|)_{i \in I}$ équivaut à $\sum_{\alpha \in A} T_{I_\alpha} < +\infty$ d'après le théorème 1.93 et la proposition 1.83. Si $(u_i)_{i \in I}$ est normalement sommable dans E (cf. corollaire 1.88) elle est sommable. Elle vérifie donc la formule de sommation par paquets par le théorème 1.91. \square

1.6 Séries doubles

Nous examinons maintenant le cas particulier des familles $(u_i)_{i \in I}$ indexées par un produit cartésien $I = I' \times I''$. On parle dans ce cas de « série double ». Nous nous limiterons au cas où $I = \mathbb{N}^2$, mais il est facile d'adapter les énoncés ci-dessous à $I = I' \times I''$ avec I' et I'' dénombrables, une fois fixées des bijections $\mathbb{N} \rightarrow I'$ et $\mathbb{N} \rightarrow I''$. L'indice i est désormais un couple d'entiers, $i = (k, l) \in \mathbb{N}^2$. Pour des raisons typographiques, nous écrirons $\{u_{k,l}; (k, l) \in \mathbb{N}^2\}$ de préférence à $(u_{(k,l)})_{(k,l) \in \mathbb{N}^2}$.

Définition 1.95 (série double). *Soit $(u_{k,l})_{k,l \in \mathbb{N}}$ une suite double d'éléments de l'espace vectoriel normé E . On dit que la série double de terme général $u_{k,l}$ est convergente (resp. normalement convergente) si la famille $\{u_{k,l}; (k, l) \in \mathbb{N}^2\}$ est sommable (resp. normalement sommable). La somme S de cette famille est alors appelée somme de la série double et notée*

$$S = \sum_{(k,l) \in \mathbb{N}^2} u_{k,l} = \sum_{k,l \in \mathbb{N}} u_{k,l}.$$

Dans tout ce qui suit, nous donnerons les énoncés relatifs aux séries doubles à termes réels ou complexes en nous contentant d'indiquer sous forme de remarques ou de commentaires les modifications à apporter pour la généralisation aux espaces vectoriels normés. Pour $E = \mathbb{R}$ ou \mathbb{C} , la sommabilité normale s'appelle sommabilité absolue et équivaut à la sommabilité. Il importe de bien comprendre que contrairement au cas des séries simples, la définition 1.95 n'autorise pas l'existence d'une série double à termes dans \mathbb{R} ou \mathbb{C} qui soit convergente sans l'être absolument. Il peut arriver que pour une certaine bijection $f : \mathbb{N} \rightarrow \mathbb{N}^2$, la série simple $\sum_{j=0}^{+\infty} u_{f(j)}$ soit convergente sans l'être absolument, mais dans ce cas la série double $\sum_{k,l} u_{k,l}$ est divergente (cf. th. 1.86 et th. 1.76).

Voici un premier critère de convergence pour les séries doubles.

Proposition 1.96. *Pour que la série double de terme général réel ou complexe $u_{k,l}$ ($k, l \in \mathbb{N}$) soit convergente, il faut et il suffit que la suite des sommes finies*

$$T_n := \sum_{k+l \leq n} |u_{k,l}|, \quad n \in \mathbb{N},$$

soit bornée. On a le même énoncé avec

$$T'_n := \sum_{\max(k,l) \leq n} |u_{k,l}|, \quad n \in \mathbb{N}.$$

Attention à bien noter l'indexation de la somme T_n par la condition $k + l \leq n$ et non pas $k + l = n$. Cet énoncé se généralise aux séries à termes dans E espace vectoriel normé *complet* en remplaçant « convergente » par « normalement convergente » et $|u_{k,l}|$ par $\|u_{k,l}\|$.

Preuve. Vérifions d'abord que la condition est nécessaire en supposant que la série double de terme général $u_{k,l}$ est convergente. Alors la famille $\{u_{k,l}; (k,l) \in \mathbb{N}^2\}$ est absolument sommable. On a alors d'après la proposition 1.83 b) :

$$M := \sup_{K \text{ fini } \subset \mathbb{N}^2} \sum_{(k,l) \in K} |u_{k,l}| < +\infty. \quad (1.41)$$

Comme T_n est la somme des termes indexés par le sous-ensemble fini de \mathbb{N}^2 , $D_n := \{(k,l) \in \mathbb{N}^2; k+l \leq n\}$, on a $T_n \leq M$ pour tout n donc la suite $(T_n)_{n \in \mathbb{N}}$ est bornée. Le même argument vaut pour (T'_n) en remplaçant D_n par $C_n := \{0, \dots, n\}^2$.

Réciproquement, supposons que la suite $(T_n)_{n \in \mathbb{N}}$ soit bornée par un réel positif M_1 . Soit K une partie finie quelconque de \mathbb{N}^2 . Il existe alors un entier n tel que D_n recouvre K : il suffit de choisir pour cela $n = \max\{k+l; (k,l) \in K\}$. Alors

$$\sum_{(k,l) \in K} |u_{k,l}| \leq \sum_{(k,l) \in D_n} |u_{k,l}| = T_n \leq M_1$$

et comme l'ensemble fini K est quelconque, ceci montre que (1.41) est vérifiée avec $M \leq M_1$. La sommabilité absolue de $\{u_{k,l}; (k,l) \in \mathbb{N}^2\}$ en découle par la proposition 1.83 a). L'adaptation à T'_n est immédiate. \square

À titre d'exercice, on pourra montrer qu'une condition nécessaire et suffisante de convergence de la série double de terme général réel ou complexe $u_{k,l}$ ($k, l \in \mathbb{N}$) est l'existence d'une suite $(K_n)_{n \in \mathbb{N}}$ croissante pour l'inclusion²⁰ de parties finies de \mathbb{N}^2 , telle que $\bigcup_{n \in \mathbb{N}} K_n = \mathbb{N}^2$ et que la suite de terme général

$$T_n := \sum_{(k,l) \in K_n} |u_{k,l}|$$

soit bornée.

Le théorème suivant fournit à la fois un critère de convergence des série doubles et une méthode de calcul de la somme.

20. i.e. telle que pour tout $n \in \mathbb{N}$, $K_n \subset K_{n+1}$.

Théorème 1.97 (d'interversion des sommations).

1. Si les $v_{k,l}$ sont des éléments de $\overline{\mathbb{R}}_+$, les égalités

$$\sum_{(k,l) \in \mathbb{N}^2} v_{k,l} = \sum_{k=0}^{+\infty} \left\{ \sum_{l=0}^{+\infty} v_{k,l} \right\} = \sum_{l=0}^{+\infty} \left\{ \sum_{k=0}^{+\infty} v_{k,l} \right\} \quad (1.42)$$

sont toujours vérifiées dans $\overline{\mathbb{R}}_+$.

2. Si les $u_{k,l}$ sont réels ou complexes, la série double de terme général $u_{k,l}$ converge si et seulement si l'une des deux conditions suivantes est réalisée :

$$a) \sum_{k=0}^{+\infty} \left\{ \sum_{l=0}^{+\infty} |u_{k,l}| \right\} < +\infty \quad \text{ou} \quad b) \sum_{l=0}^{+\infty} \left\{ \sum_{k=0}^{+\infty} |u_{k,l}| \right\} < +\infty. \quad (1.43)$$

Dans ce cas, pour tout $k \in \mathbb{N}$, la série simple $\sum_{l=0}^{+\infty} u_{k,l}$ est absolument convergente et il en va de même en échangeant les rôles de k et l . On a de plus la formule d'interversion des sommations :

$$\sum_{(k,l) \in \mathbb{N}^2} u_{k,l} = \sum_{k=0}^{+\infty} \left\{ \sum_{l=0}^{+\infty} u_{k,l} \right\} = \sum_{l=0}^{+\infty} \left\{ \sum_{k=0}^{+\infty} u_{k,l} \right\}. \quad (1.44)$$

Ce théorème s'étend au cas où les $u_{k,l}$ sont dans E espace vectoriel normé complet en remplaçant $|u_{k,l}|$ par $\|u_{k,l}\|$, la convergence de la série double par sa convergence normale et la convergence absolue des séries simples par leur convergence normale.

Preuve. Le cas des termes positifs est une application immédiate du théorème 1.93 en considérant les deux décompositions suivantes de \mathbb{N}^2 , cf. figure 1.5 :

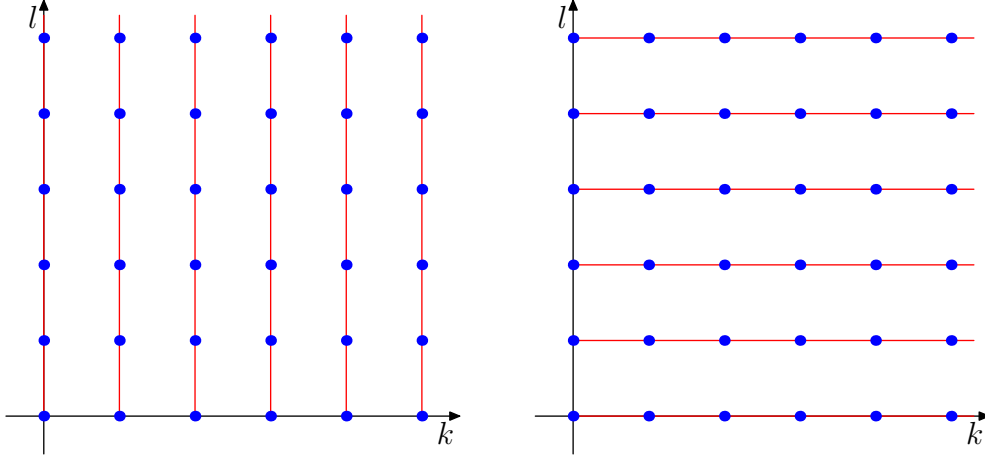
$$a) \mathbb{N}^2 = \bigcup_{k \in \mathbb{N}} (\{k\} \times \mathbb{N}) \quad b) \mathbb{N}^2 = \bigcup_{l \in \mathbb{N}} (\mathbb{N} \times \{l\}). \quad (1.45)$$

Dans le cas des termes réels ou complexes, la condition (1.43 a) utilisée avec la décomposition (1.45 a) équivaut à la sommabilité absolue de $\{u_{k,l}; (k,l) \in \mathbb{N}^2\}$ par le théorème 1.94. Il en va de même pour (1.43 b) utilisée avec la décomposition (1.45 b).

La sommabilité absolue de $\{u_{k,l}; (k,l) \in \mathbb{N}^2\}$ implique *chacune* des conditions (1.43 a) et (1.43 b) par le théorème 1.94. La première des deux implique la convergence absolue de toutes les séries $\sum_{l=0}^{+\infty} u_{k,l}$, la deuxième celle des séries $\sum_{k=0}^{+\infty} u_{k,l}$.

Enfin lorsque $\{u_{k,l}; (k,l) \in \mathbb{N}^2\}$ est sommable, on obtient (1.44) en appliquant la formule de sommation par paquets (th. 1.91) aux décompositions (1.45 a) et (1.45 b). \square

Une application classique de la théorie des séries doubles est le produit de deux séries *absolument* convergentes.

FIG. 1.5 – Découpages a) et b) de \mathbb{N}^2

Théorème 1.98 (produit de deux séries). *Soient $\sum_{k=0}^{+\infty} u_k$ et $\sum_{l=0}^{+\infty} v_l$ deux séries à termes réels ou complexes, absolument convergentes. Alors la série double de terme général $u_k v_l$ est convergente et on a*

$$\sum_{(k,l) \in \mathbb{N}^2} u_k v_l = \left\{ \sum_{k=0}^{+\infty} u_k \right\} \left\{ \sum_{l=0}^{+\infty} v_l \right\} = \sum_{n=0}^{+\infty} \left\{ \sum_{k+l=n} u_k v_l \right\}. \quad (1.46)$$

La série simple de terme général $w_n = \sum_{k+l=n} u_k v_l$ est absolument convergente. On l'appelle série produit de $\sum_{k=0}^{+\infty} u_k$ et $\sum_{l=0}^{+\infty} v_l$.

Preuve. Commençons par vérifier la convergence de la série double. Posons

$$U_n := \sum_{k=0}^n u_k, \quad U := \sum_{k=0}^{+\infty} u_k, \quad U_n^{\text{abs}} := \sum_{k=0}^n |u_k|, \quad U^{\text{abs}} := \sum_{k=0}^{+\infty} |u_k|$$

et définissons de même V_n , V , V_n^{abs} et V^{abs} en remplaçant u_k par v_l . Le produit

$$U_n^{\text{abs}} V_n^{\text{abs}} = \sum_{(k,l) \in \{0, \dots, n\}^2} |u_k| |v_l|$$

est exactement la somme T'_n de la proposition 1.96. Comme il est borné par la constante $U^{\text{abs}} V^{\text{abs}}$ indépendante de n , la proposition 1.96 nous assure de la convergence de la série double de terme général $u_k v_l$.

En appliquant la formule de sommation par paquets à cette série avec le découpage $\mathbb{N}^2 = \cup_{k \in \mathbb{N}} (\{k\} \times \mathbb{N})$, on obtient :

$$\sum_{(k,l) \in \mathbb{N}^2} u_k v_l = \sum_{k=0}^{+\infty} \left\{ \sum_{l=0}^{+\infty} u_k v_l \right\},$$

où chacune des séries simples $\sum_{l=0}^{+\infty} u_k v_l$ est absolument convergente. Dans une telle série, on peut mettre en facteur la « constante » u_k , d'où

$$\sum_{(k,l) \in \mathbb{N}^2} u_k v_l = \sum_{k=0}^{+\infty} u_k \left\{ \sum_{l=0}^{+\infty} v_l \right\}.$$

Dans la série indexée par k , on peut mettre en facteur la « constante » $\left\{ \sum_{l=0}^{+\infty} v_l \right\}$, d'où

$$\sum_{(k,l) \in \mathbb{N}^2} u_k v_l = \left\{ \sum_{l=0}^{+\infty} v_l \right\} \sum_{k=0}^{+\infty} u_k,$$

ce qui achève la vérification de la première égalité dans (1.46).

La deuxième égalité dans (1.46) résulte de l'application du théorème de sommation par paquets à la famille sommable $\{u_k v_l; (k, l) \in \mathbb{N}^2\}$ avec le découpage $\mathbb{N}^2 = \cup_{n \in \mathbb{N}} \{(k, l) \in \mathbb{N}^2; k + l = n\}$. Le même découpage utilisé avec le théorème 1.94 nous

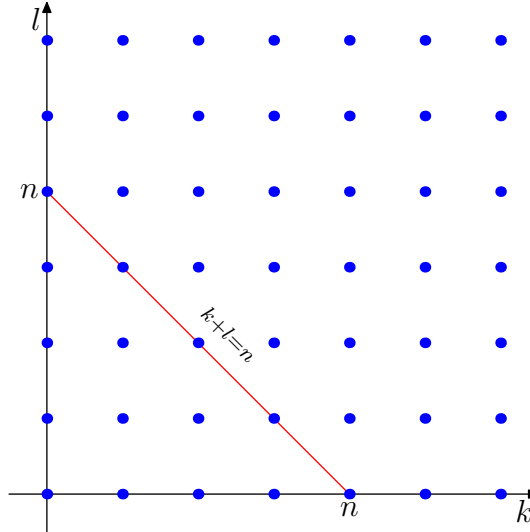


FIG. 1.6 – Découpage $\mathbb{N}^2 = \cup_{n \in \mathbb{N}} \{(k, l) \in \mathbb{N}^2; k + l = n\}$

fournit la convergence de la série à termes positifs

$$w_n^{\text{abs}} := \sum_{k+l=n} |u_k v_l|.$$

Cette dernière convergence entraîne par comparaison, la convergence absolue de la série de terme général w_n , puisque $|w_n| \leq w_n^{\text{abs}}$ par inégalité triangulaire.

□

Chapitre 2

Événements et Probabilités

Ce chapitre reprend essentiellement les deux premiers chapitres du cours de deuxième année¹ : *Espaces Probabilisés* et *Conditionnement et Indépendance*. La principale innovation est la mise en place de rudiments de théorie de la mesure (essentiellement du vocabulaire) en vue de pouvoir disposer d'emblée d'exemples de probabilités sortant du cadre des probabilités discrètes. Nous pourrons ainsi modéliser l'expérience aléatoire consistant à choisir un point *au hasard* sur le segment $[0, 1]$. Le langage de la mesure est présenté dans la première section. La deuxième section contient une approche informelle de la modélisation de l'aléatoire avant de présenter à la section 3 les axiomes de la théorie des probabilités et les propriétés générales des probabilités. On examine ensuite les notions de conditionnement et d'indépendance.

2.1 Notion de mesure

Soit Ω un ensemble, on note $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω . Une mesure sur Ω est une *fonction d'ensembles* m qui à certaines parties A de Ω , associe un réel positif $m(A)$ appelé mesure de A . La théorie que nous allons présenter fournit un cadre mathématique commun pour des notions comme le dénombrement, la mesure des grandeurs géométriques (longueur, aire, volume), physiques (masse) et les probabilités. À partir du concept de mesure, on peut bâtir une intégrale² sur l'ensemble Ω permettant d'unifier les notions d'intégrale simple ou multiple au sens classique, de série absolument convergente et d'espérance mathématique d'une variable aléatoire. Il est commode d'élargir d'emblée l'ensemble d'arrivée de m à $\overline{\mathbb{R}}_+$ au lieu de \mathbb{R}_+ , par exemple pour pouvoir dire que l'aire d'un quart de plan est $+\infty$.

Intuitivement, en pensant par exemple à la mesure des aires, les propriétés minimales que l'on puisse exiger de m sont

- a) la *croissance* : si $A \subset B$, $m(A) \leq m(B)$,
- b) l'*additivité* : si $A \cap B = \emptyset$, $m(A \cup B) = m(A) + m(B)$, sous réserve que $m(A)$, $m(B)$ et $m(A \cup B)$ soient définies.

1. *Introduction au Calcul des Probabilités*, <http://math.univ-lille1.fr/~suquet/index.html>

2. Appelée « intégrale de Lebesgue ». Sa construction n'est pas au programme de ce cours.

Il est facile de voir que la croissance est une conséquence de l'additivité en écrivant si $A \subset B$, $B = A \cup (B \setminus A)$ et $m(B) = m(A) + m(B \setminus A) \geq m(A)$, avec les mêmes réserves d'existence. L'additivité s'étend par une récurrence immédiate, aux suites finies A_1, \dots, A_n d'ensembles, donnant l'*additivité finie* : $m(A_1 \cup \dots \cup A_n) = m(A_1) + \dots + m(A_n)$ si les A_i sont deux à deux disjoints. Par contre elle ne s'étend pas automatiquement aux suites infinies d'ensembles deux à deux disjoints.

Pour avoir une théorie assez riche, on doit pouvoir effectuer certains passages à la limite, par exemple pour pouvoir mesurer l'aire d'un disque dans le plan en le « pavant » par des carreaux à côtés parallèles aux axes. La propriété correspondante est appelée *σ -additivité* :

$$\text{si les } A_n \text{ sont deux à deux disjoints, } m\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} m(A_n),$$

pourvu que toutes ces quantités soient définies.

Pourquoi ces clauses restrictives sur l'existence des $m(A)$? Il se trouve que dans la plupart des cas intéressants (sauf lorsque Ω est au plus dénombrable), on ne sait pas définir $m(A)$ pour tout $A \in \mathcal{P}(\Omega)$. Souvent pour construire une mesure m , on commence par attribuer une valeur à $m(A)$ pour chaque A dans une famille \mathcal{C} bien particulière de parties de Ω . Par exemple si $\Omega = \mathbb{R}$, on peut prendre pour \mathcal{C} la famille des intervalles $]a, b]$ et « décider » que $m(]a, b]) := b - a$, choisissant ainsi de mesurer $]a, b]$ par sa longueur³. De même si $\Omega = \mathbb{R}^2$, on peut partir de la famille \mathcal{C} des « rectangles » $R =]a, b] \times]c, d]$ et décider que $m(R) := (b - a)(d - c)$. Dans un deuxième temps, on essaie de prolonger m à une famille \mathcal{F} de parties de Ω plus grande que \mathcal{C} , tout en préservant la σ -additivité. Il se trouve qu'il n'est pas toujours possible de réaliser ce prolongement jusqu'à prendre $\mathcal{F} = \mathcal{P}(\Omega)$. Certaines parties de \mathbb{R} sont d'une trop grande complexité pour qu'on puisse leur attribuer une « longueur ». Ainsi la fonction d'ensembles m a un *ensemble de définition* qui est une sous-famille de $\mathcal{P}(\Omega)$. Cet ensemble de définition de m est ce que l'on appelle une *tribu* de parties de Ω .

Il est temps maintenant de formaliser les définitions suivantes.

Définition 2.1 (tribu). *Une famille \mathcal{F} de parties de Ω est appelée tribu (ou σ -algèbre) sur Ω si elle*

- a) *possède l'ensemble vide : $\emptyset \in \mathcal{F}$;*
- b) *est stable par passage au complémentaire : $\forall A \in \mathcal{F}, A^c \in \mathcal{F}$;*
- c) *est stable par union dénombrable : $(\forall i \in \mathbb{N}^*, A_i \in \mathcal{F}) \Rightarrow \bigcup_{i \in \mathbb{N}^*} A_i \in \mathcal{F}$.*

On vérifie facilement à partir de cette définition qu'une tribu est stable par unions ou intersections finies et par intersections dénombrables.

Définition 2.2 (mesure). *Soit \mathcal{F} une tribu sur Ω . On appelle mesure positive sur (Ω, \mathcal{F}) une application*

$$m : \mathcal{F} \longrightarrow [0, +\infty],$$

vérifiant

3. Il y a bien d'autres choix possibles, on pourrait poser plus généralement $m(]a, b]) := F(b) - F(a)$ où F est une fonction croissante $\mathbb{R} \rightarrow \mathbb{R}$, continue à droite.

a) $m(\emptyset) = 0$;

b) m est σ -additive : pour toute suite $(A_i)_{i \in \mathbb{N}^*}$ d'éléments de \mathcal{F} deux à deux disjoints,

$$m\left(\bigcup_{i \in \mathbb{N}^*} A_i\right) = \sum_{i=1}^{+\infty} m(A_i). \quad (2.1)$$

Dans la définition de la σ -additivité, ou de la stabilité par union dénombrable, on aurait évidemment pu tout aussi bien indexer la suite (A_i) par \mathbb{N} au lieu de \mathbb{N}^* .

Remarque 2.3. La réunion des A_i est invariante par permutation sur les indices et si chaque A_i est à son tour union dénombrable d'ensembles $B_{i,j} \in \mathcal{F}$ ($j \in \mathbb{N}^*$) deux à deux disjoints, on a clairement

$$\bigcup_{i \in \mathbb{N}^*} A_i = \bigcup_{i \in \mathbb{N}^*} \bigcup_{j \in \mathbb{N}^*} B_{i,j} = \bigcup_{(i,j) \in \mathbb{N}^* \times \mathbb{N}^*} B_{i,j}.$$

On voit ici que l'on a un besoin crucial des propriétés de convergence commutative et de sommation par paquets dans $\overline{\mathbb{R}}_+$. Sans elles la définition de la σ -additivité serait incohérente.

Voyons maintenant des exemples de tribus qui nous seront utiles. Les trois exemples les plus simples sont les suivants.

- La tribu triviale sur Ω est $\mathcal{F} = \{\Omega, \emptyset\}$.
- $\mathcal{P}(\Omega)$ est une tribu.
- Si A est une partie de Ω , alors $\mathcal{F} := \{\Omega, \emptyset, A, A^c\}$ est une tribu. C'est la plus petite tribu possédant A comme élément, *i.e.* toute tribu \mathcal{G} telle que $A \in \mathcal{G}$ contient \mathcal{F} . On dit que \mathcal{F} est la tribu engendrée par A .

Cette notion de tribu engendrée se généralise en remarquant que si $(\mathcal{G}_i)_{i \in I}$ est une famille quelconque de tribus sur Ω , $\mathcal{G} := \bigcap_{i \in I} \mathcal{G}_i$ est une tribu sur Ω (vérification immédiate à partir de la définition 2.1).

Définition 2.4 (tribu engendrée). Soit \mathcal{C} une famille de parties d'un ensemble Ω . On appelle tribu engendrée par \mathcal{C} , et on note $\sigma(\mathcal{C})$, la plus petite tribu contenant \mathcal{C} (c'est l'intersection de toutes les tribus sur Ω contenant \mathcal{C}).

Définition 2.5 (tribu borélienne). On appelle tribu borélienne sur \mathbb{R}^d la tribu engendrée par la famille \mathcal{O} des ensembles ouverts⁴ de \mathbb{R}^d . On la notera $\text{Bor}(\mathbb{R}^d)$. Ainsi $\text{Bor}(\mathbb{R}^d) = \sigma(\mathcal{O})$. Les sous-ensembles de \mathbb{R}^d qui sont éléments de sa tribu borélienne sont appelés boréliens de \mathbb{R}^d ou boréliens tout court quand il n'y a pas d'ambiguïté.

Remarque 2.6. On peut démontrer que $\text{Bor}(\mathbb{R})$ est aussi la tribu engendrée par les fermés de \mathbb{R} , ou par les intervalles ouverts, ou les intervalles fermés, ou les semi-ouverts, ou les intervalles (ouverts ou fermés) à extrémités rationnelles, ou les intervalles $]-\infty, a]$, ou les intervalles $[a, +\infty[$. De même, $\text{Bor}(\mathbb{R}^d)$ est engendrée par les pavés ouverts ou par les pavés de la forme $\prod_{k=1}^d]a_k, b_k]$.

4. Un ensemble ouvert de \mathbb{R}^d est une réunion (quelconque) de pavés ouverts $\prod_{k=1}^d]a_k, b_k[$. Un fermé est le complémentaire d'un ouvert.

Voyons maintenant des exemples importants de mesures.

Exemple 2.7 (masse de Dirac). Soit x_0 un élément fixé de Ω . On appelle *masse de Dirac* au point x_0 ou *mesure de Dirac* au point x_0 , la mesure δ_{x_0} sur $(\Omega, \mathcal{P}(\Omega))$ définie par

$$\forall A \in \mathcal{P}(\Omega), \quad \delta_{x_0}(A) := \begin{cases} 1 & \text{si } x_0 \in A, \\ 0 & \text{si } x_0 \notin A. \end{cases}$$

Par restriction, δ_{x_0} est aussi une mesure sur (Ω, \mathcal{F}) pour toute tribu \mathcal{F} sur Ω .

Vérification. Il est clair que $\delta_{x_0}(\emptyset) = 0$. Pour montrer la σ -additivité, soit $(A_n)_{n \in \mathbb{N}}$ une suite d'éléments de $\mathcal{P}(\Omega)$, deux à deux disjoints. Nous distinguons deux cas.

- a) $x_0 \notin \bigcup_{n \in \mathbb{N}} A_n$. Alors $\delta_{x_0}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = 0$. D'autre part, x_0 ne peut appartenir à aucun des A_n , donc pour tout $n \in \mathbb{N}$, $\delta_{x_0}(A_n) = 0$ et $\sum_{n \in \mathbb{N}} \delta_{x_0}(A_n) = 0$.
- b) $x_0 \in \bigcup_{n \in \mathbb{N}} A_n$. Alors $\delta_{x_0}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = 1$. D'autre part, comme les A_n sont deux à deux disjoints, x_0 doit appartenir à *un seul* des A_n , disons A_{n_0} . Ainsi, $\delta_{x_0}(A_{n_0}) = 1$ et pour tout $n \neq n_0$, $\delta_{x_0}(A_n) = 0$, d'où $\sum_{n \in \mathbb{N}} \delta_{x_0}(A_n) = 1$.

Dans les deux cas, on a

$$\delta_{x_0}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \delta_{x_0}(A_n),$$

δ_{x_0} est donc bien σ -additive. C'est une mesure sur $(\Omega, \mathcal{P}(\Omega))$. □

Exemple 2.8 (série de mesures finies). Si les $(\mu_k)_{k \in \mathbb{N}}$ sont des mesures finies⁵ sur (Ω, \mathcal{F}) et si $(a_k)_{k \in \mathbb{N}}$ est une suite de réels positifs, la fonction d'ensembles $\mu = \sum_{k \in \mathbb{N}} a_k \mu_k$ définie sur \mathcal{F} par

$$\mu : \mathcal{F} \rightarrow \overline{\mathbb{R}}_+, \quad A \mapsto \mu(A) := \sum_{k \in \mathbb{N}} a_k \mu_k(A)$$

est une mesure sur (Ω, \mathcal{F}) . Le résultat se généralise à $\mu = \sum_{i \in I} a_i \mu_i$, avec I au plus dénombrable.

Vérification. Pour tout k , $0 \leq a_k < +\infty$, donc $a_k \mu_k(\emptyset) = 0$ et $\mu(\emptyset) = \sum_{k \in \mathbb{N}} a_k \mu_k(\emptyset) = 0$. Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'éléments de \mathcal{F} , deux à deux disjoints. En utilisant la σ -additivité de chaque μ_k et l'interversion des sommations pour les séries doubles à termes positifs, on peut écrire :

$$\begin{aligned} \mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) &= \sum_{k \in \mathbb{N}} a_k \mu_k\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{k \in \mathbb{N}} a_k \left\{ \sum_{n \in \mathbb{N}} \mu_k(A_n) \right\} = \sum_{k \in \mathbb{N}} \left\{ \sum_{n \in \mathbb{N}} a_k \mu_k(A_n) \right\} \\ &= \sum_{n \in \mathbb{N}} \left\{ \sum_{k \in \mathbb{N}} a_k \mu_k(A_n) \right\} \\ &= \sum_{n \in \mathbb{N}} \mu(A_n), \end{aligned}$$

5. La mesure μ_k est *finie* si $\mu_k(A) < +\infty$ pour tout $A \in \mathcal{F}$. Elle est donc aussi *bornée* puisque $\mu_k(A) \leq \mu_k(\Omega) < +\infty$. L'hypothèse « μ_k finie » nous évite la gestion du conflit $a_k = 0$ et $\mu_k(A) = +\infty$.

ce qui établit la σ -additivité de μ . \square

Exemple 2.9 (mesure ponctuelle). Soient I un ensemble au plus dénombrable, $(x_i)_{i \in I}$ une famille dans Ω et $(a_i)_{i \in I}$ une famille de réels positifs. Alors $\mu := \sum_{i \in I} a_i \delta_{x_i}$ est une mesure sur $(\Omega, \mathcal{P}(\Omega))$, donc aussi par restriction sur tout espace mesurable (Ω, \mathcal{F}) . C'est un cas particulier de l'exemple précédent. Les mesures de ce type sont appelées *mesures ponctuelles*.

Remarque 2.10. Si $\Omega = \{\omega_i; i \in I\}$ est au plus dénombrable, toute mesure sur $(\Omega, \mathcal{P}(\Omega))$ qui est *finie sur les singletons* est une mesure ponctuelle.

Vérification. La tribu $\mathcal{P}(\Omega)$ possède les singletons (quand Ω est fini ou dénombrable, c'est même la seule tribu ayant cette propriété). Soit μ une mesure sur $(\Omega, \mathcal{P}(\Omega))$. On peut définir

$$\forall i \in I, \quad a_i := \mu(\{\omega_i\}),$$

en notant que puisque μ est finie sur les singletons, $0 \leq a_i < +\infty$. Considérons alors la mesure ponctuelle

$$\nu := \sum_{i \in I} a_i \delta_{\omega_i}.$$

Soit $A \in \mathcal{P}(\Omega)$, alors A est fini ou dénombrable et on peut l'écrire comme union disjointe finie ou dénombrable de singletons :

$$A = \bigcup_{\omega_i \in A} \{\omega_i\}.$$

Par σ -additivité (ou par additivité quand A est fini, cf. prop. 2.16 ci-dessous), on a donc

$$\mu(A) = \sum_{\omega_i \in A} \mu(\{\omega_i\}) = \sum_{\omega_i \in A} a_i = \sum_{i \in I} a_i \delta_{\omega_i}(A) = \nu(A).$$

Ainsi pour tout $A \in \mathcal{P}(\Omega)$, $\mu(A) = \nu(A)$, donc $\mu = \nu$ et μ est une mesure ponctuelle. \square

Exemple 2.11 (mesure de comptage). Si $\Omega = \{\omega_i; i \in I\}$ est au plus dénombrable, l'application $\mu : \mathcal{P}(\Omega) \rightarrow \overline{\mathbb{R}}_+$ définie par

$$\forall A \in \mathcal{P}(\Omega), \quad \mu(A) = \begin{cases} \text{card } A & \text{si } A \text{ est fini,} \\ +\infty & \text{sinon,} \end{cases}$$

est une mesure, appelée *mesure de comptage*. Pour le voir, il suffit de remarquer que pour tout A , $\mu(A) = \nu(A)$, où ν est définie comme la mesure ponctuelle $\nu = \sum_{i \in I} \delta_{\omega_i}$.

Exemple 2.12. Soit μ une mesure sur (Ω, \mathcal{F}) et $B \in \mathcal{F}$. La fonction d'ensembles $\nu = \mu(\cdot \cap B)$ définie sur \mathcal{F} par

$$\forall A \in \mathcal{F}, \quad \nu(A) := \mu(A \cap B)$$

est une mesure sur (Ω, \mathcal{F}) . Si de plus $0 < \mu(B) < +\infty$, la fonction d'ensembles μ_B définie sur \mathcal{F} par

$$\forall A \in \mathcal{F}, \quad \mu_B(A) := \frac{\mu(A \cap B)}{\mu(B)}$$

est une mesure sur (Ω, \mathcal{F}) , vérifiant $\mu_B(\Omega) = 1$, autrement dit une *probabilité*⁶. Quand $\mu = P$ est déjà une probabilité sur (Ω, \mathcal{F}) , la mesure P_B est appelée *probabilité conditionnelle*; notation : $P_B(A) =: P(A \mid B)$.

Vérification. On se contente de vérifier que ν est σ -additive, tout le reste étant évident. Soit $(A_i)_{i \in \mathbb{N}}$ une suite d'éléments de \mathcal{F} , deux à deux disjoints. Pour $i \neq j$, $A_i \cap A_j = \emptyset$ et comme $(A_i \cap B) \cap (A_j \cap B) \subset A_i \cap A_j$, les $A_i \cap B$ sont aussi deux à deux disjoints. De plus on a

$$\left(\bigcup_{i \in \mathbb{N}} A_i \right) \cap B = \bigcup_{i \in \mathbb{N}} (A_i \cap B).$$

La σ -additivité de ν découle alors clairement de celle de μ . □

Exemple 2.13 (mesure de Lebesgue λ_d). Nous admettons qu'il existe une unique mesure μ sur $(\mathbb{R}^d, \text{Bor}(\mathbb{R}^d))$ telle que pour tout pavé $\prod_{i=1}^d]a_i, b_i]$ non vide⁷,

$$\mu \left(\prod_{i=1}^d]a_i, b_i] \right) = \prod_{i=1}^d (b_i - a_i).$$

On l'appelle *mesure de Lebesgue* sur \mathbb{R}^d et on la note λ_d . Pour $d = 1$, on étend ainsi la notion de longueur des intervalles à tous les ensembles membres de $\text{Bor}(\mathbb{R})$, pour $d = 2$ on étend de même la notion d'aire des rectangles à tous les ensembles boréliens du plan \mathbb{R}^2 , pour $d = 3$, on étend la notion de volume. Pour $d > 3$, on continuera à parler de volume ou d'hypervolume. Nous admettrons que les formules classiques de calcul d'aire ou de volume se réécrivent à l'aide de λ_d . Par exemple si D est un disque de rayon r de \mathbb{R}^2 , $\lambda_2(D) = \pi r^2$. Plus généralement, voici les principales propriétés de la mesure de Lebesgue.

Proposition 2.14. *La mesure de Lebesgue λ_d sur $(\mathbb{R}^d, \text{Bor}(\mathbb{R}^d))$ a les propriétés suivantes.*

- i) λ_d est invariante par translations : si $h : x \mapsto x + c$ est une translation de \mathbb{R}^d , alors pour tout $B \in \text{Bor}(\mathbb{R}^d)$, $h(B) \in \text{Bor}(\mathbb{R}^d)$ et $\lambda_d(B) = \lambda_d(h(B))$.
- ii) λ_d est invariante par toute isométrie euclidienne de \mathbb{R}^d : symétrie, rotation, etc.
- iii) Si h est l'homothétie $x \mapsto cx$ dans \mathbb{R}^d , pour tout borélien B , $\lambda_d(h(B)) = |c|^d \lambda_d(B)$.
- iv) λ_d ne charge pas les points : $\forall x \in \mathbb{R}^d$, $\lambda_d(\{x\}) = 0$. Si $A \subset \mathbb{R}^d$ est fini ou dénombrable, $\lambda_d(A) = 0$.

6. En anticipant un peu sur la suite du chapitre.

7. Ce qui suppose implicitement que $a_i < b_i$ pour chaque $i = 1, \dots, d$.

v) $\lambda_d \left(\prod_{i=1}^d [a_i, b_i] \right) = \lambda_d \left(\prod_{i=1}^d]a_i, b_i[\right)$ et cette égalité implique bien sûr, l'égalité des mesures des 4^d pavés obtenus en jouant sur l'ouverture ou la fermeture des extrémités a_i, b_i des intervalles.

vi) Si E est un sous-espace affine de \mathbb{R}^d et $E \neq \mathbb{R}^d$, $\lambda_d(E) = 0$.

Un bon exercice consiste à établir la formule $\lambda_2(D) = \pi r^2$ en utilisant le calcul de l'aire de l'hypographe de la fonction $f : [-1, 1] \rightarrow \mathbb{R}, x \mapsto \sqrt{1 - x^2}$ par une intégrale de Riemann, cf. proposition A.5 p. 218 et certaines des propriétés de λ_2 énoncées ci-dessus.

2.2 Modéliser l'aléatoire

2.2.1 Notion d'expérience aléatoire

La théorie des probabilités fournit des modèles mathématiques permettant l'étude d'expériences dont le résultat ne peut être prévu avec une totale certitude. Le tableau 2.1 en donne quelques exemples.

Expérience	Résultat observable
Lancer d'un dé	Un entier $k \in \{1, \dots, 6\}$
Prélèvement de n objets en sortie d'une chaîne de production	Nombre d'objets défectueux dans l'échantillon
Questionnaire à 100 questions binaires	Suite ω de 100 réponses $\omega \in \{\text{oùi}, \text{non}\}^{100}$
Lancer d'une pièce jusqu'à la première obtention de pile	Un entier $k \in \mathbb{N}$: le temps d'attente du premier succès
Mise en service d'une ampoule	Durée de vie $T \in \mathbb{R}$
Lancer d'une fléchette sur une cible	Point d'impact
Mouvement d'un grain de pollen dans un liquide	Une fonction continue : la trajectoire
Mélange de deux gaz	Répartition spatiale de deux types de molécules

TAB. 2.1 – Quelques expériences aléatoires typiques

Bien que le résultat précis de chacune de ces expériences soit imprévisible, l'observation et l'intuition nous amènent à penser que ces phénomènes obéissent à certaines lois. Par exemple si on jette 6 000 fois un dé, on s'attend à ce que le nombre d'apparitions de la face « 3 » soit *voisin* de 1 000. Si on met en service 100 ampoules, leurs durées de vie observées seront *concentrées* autour d'une certaine valeur moyenne.

La théorie des probabilités permet de donner un sens précis à ces considérations un peu vagues. La *statistique* permet de confronter les modèles probabilistes avec la réalité

observée afin de les valider ou de les invalider. Par exemple si quelqu'un a 60 bonnes réponses sur 100 au questionnaire, est-il légitime de considérer qu'il a « mieux fait » que le hasard ? Sur les n objets prélevés en sortie de chaîne, k sont défectueux. Peut-on en déduire quelque chose sur la qualité de la production globale ?

2.2.2 Événements

La théorie moderne des probabilités utilise le langage des ensembles pour modéliser une expérience aléatoire. Nous noterons Ω un ensemble dont les éléments représentent tous les résultats possibles ou *événements élémentaires* d'une expérience aléatoire donnée. Les *événements* (ou événements composés) seront représentés par des parties (sous-ensembles) de Ω .

Il n'est pas toujours facile de trouver un ensemble Ω permettant de modéliser l'expérience aléatoire. Voici une règle pratique pour y arriver : les événements élémentaires sont ceux qui contiennent *l'information maximale* qu'il est possible d'obtenir de l'expérience. Par exemple si on jette un dé, l'événement A : « obtention d'un chiffre pair » n'est pas élémentaire. Il est composé des trois événements élémentaires 2, 4, 6 : $A = \{2, 4, 6\}$. Ici $\Omega = \{1, 2, 3, 4, 5, 6\}$. De même si on lance trois fois une pièce de monnaie, les événements élémentaires sont des triplets comme (p, f, p) indiquant le résultat précis de chacun des trois lancers. Ici $\Omega = \{f, p\}^3$. L'événement B « obtention de pile au deuxième des trois lancers » est composé : $B = \{(f, p, f); (f, p, p); (p, p, f); (p, p, p)\}$.

Notations	Vocabulaire ensembliste	Vocabulaire probabiliste
\emptyset	ensemble vide	événement impossible
Ω	ensemble plein	événement certain
ω	élément de Ω	événement élémentaire
A	sous-ensemble de Ω	événement
$\omega \in A$	ω appartient à A	Le résultat ω est une des réalisations possibles de A
$A \subset B$	A inclus dans B	A implique B
$A \cup B$	réunion de A et B	A ou B
$A \cap B$	intersection de A et B	A et B
A^c	complémentaire de A dans Ω	événement contraire de A
$A \cap B = \emptyset$	A et B sont disjoints	A et B sont incompatibles

TAB. 2.2 – Langage ensembliste - langage probabiliste

Avec ce mode de représentation, les opérations logiques sur les événements : « et », « ou », « négation » se traduisent par des opérations ensemblistes : intersection, réunion,

passage au complémentaire. Le tableau 2.2 présente la correspondance entre les deux langages.

Les opérations logiques sur les événements peuvent bien sûr faire intervenir plus de deux événements. Ainsi, si A_1, \dots, A_n sont des événements,

$$\bigcup_{i=1}^n A_i = A_1 \cup A_2 \cdots \cup A_n$$

est l'ensemble des ω qui sont dans l'un au moins des A_i . C'est donc l'événement « réalisation de l'un au moins des A_i ($1 \leq i \leq n$) ». De même :

$$\bigcap_{i=1}^n A_i = A_1 \cap A_2 \cdots \cap A_n$$

est l'ensemble des ω qui sont dans tous les A_i . C'est donc l'événement « réalisation de chacun des A_i ($1 \leq i \leq n$) ». Ceci s'étend aux réunions et intersections d'une suite infinie d'événements :

$$\begin{aligned} \bigcup_{i \in \mathbb{N}^*} A_i &= \{\text{réalisation de l'un au moins des } A_i, i \in \mathbb{N}^*\}, \\ \bigcap_{i \in \mathbb{N}^*} A_i &= \{\text{réalisation de tous les } A_i, i \in \mathbb{N}^*\}. \end{aligned}$$

Ces opérations logiques sur des suites d'événements sont très utiles pour analyser des événements complexes à l'aide d'événements plus simples et, comme nous le verrons plus tard, calculer ainsi des probabilités.

2.2.3 Une question de dés

Pour finir cette introduction informelle, nous allons discuter un problème d'énoncé très simple pour voir comment les notions de tribu et de mesure s'imposent naturellement dès que l'on veut évaluer une probabilité dans une expérience où apparaît l'infini. Voici la question. On effectue des lancers répétés d'une paire de dés et on observe pour chaque lancer, la *somme* des points indiqués par les deux dés. On se propose d'attribuer une probabilité à l'événement E défini ainsi : *dans la suite des résultats observés, la première obtention d'un 9 a lieu avant la première obtention d'un 7*. On suppose ici déjà connue la définition d'une probabilité sur un espace Ω fini : à savoir une application *additive*⁸ $P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}_+$ telle que $P(\Omega) = 1$.

Commençons par modéliser un lancer. Disons que l'on a un dé bleu et un dé rouge. L'information maximale que l'on puisse envisager ici est de savoir le résultat du dé bleu et celui du rouge. On prendra donc comme espace Ω_1 , le carré cartésien $\{1, \dots, 6\}^2$ en convenant de représenter par le couple $(i, j) \in \Omega_1$ le résultat du dé bleu (première composante) et celui du rouge (deuxième composante). Notons pour $i \in \mathbb{N}^*$,

$$\begin{aligned} F_i &:= \{\text{obtention de la somme 9 au } i^{\text{e}} \text{ lancer}\}, \\ G_i &:= \{\text{obtention de la somme 7 au } i^{\text{e}} \text{ lancer}\}, \\ H_i &:= \{\text{obtention d'une autre somme que 7 ou 9 au } i^{\text{e}} \text{ lancer}\}. \end{aligned}$$

8. Si Ω est fini, la σ -additivité se réduit à l'additivité car dans toute suite (A_n) de parties de Ω deux à deux disjointes, seul un nombre fini d'entre elles ne sont pas vides.

Pour l'instant, nous n'envisageons qu'un lancer et nous attribuerons par symétrie la même probabilité à tous les événements élémentaires en prenant $P_1(\{(i, j)\}) = 1/36$ pour tout événement élémentaire (i, j) . En remarquant que F_1 est constitué de 4 événements élémentaires : $(3, 6)$, $(4, 5)$, $(5, 4)$ et $(6, 3)$, on en déduit que $P_1(F_1) = 4/36 = 1/9$. De même, G_1 étant constitué de 6 événements élémentaires, $P_1(G_1) = 6/36 = 1/6$, cf. figure 2.1. On en déduit que $P_1(H_1) = 1 - 10/36 = 13/18$.

	1	2	3	4	5	6
1						7
2					7	
3				7		9
4			7		9	
5		7		9		
6	7		9			

FIG. 2.1 – Événements F_1 et G_1

Pour modéliser les n premiers lancers, on voit que l'information maximale que l'on puisse envisager est la connaissance de la suite finie des n couples d'entiers représentant les résultats des n lancers. Ceci conduit à prendre comme espace $\Omega_n = \Omega_1 \times \cdots \times \Omega_1 = \Omega_1^n$. Cet Ω_n est un ensemble fini de cardinal 36^n . Si nous convenons là encore d'attribuer la même probabilité à tous les événements élémentaires (donc maintenant 36^{-n}), on obtient la probabilité P_n définie sur $(\Omega_n, \mathcal{P}(\Omega_n))$ par

$$\forall A \in \mathcal{P}(\Omega_n), \quad P_n(A) = \frac{\text{card } A}{\text{card } \Omega_n}.$$

Regardons le cas particulier où A est de la forme $A_1 \times A_2 \times \cdots \times A_n$, chaque A_i étant une partie de $\{1, \dots, 6\}^2$. Cela signifie que la réalisation de A_i ne dépend que du résultat du i^{e} lancer. La formule sur le cardinal d'un produit cartésien nous donne alors :

$$\begin{aligned} P_n(A_1 \times \cdots \times A_n) &= \frac{\text{card}(A_1 \times \cdots \times A_n)}{\text{card } \Omega_n} = \frac{(\text{card } A_1) \times \cdots \times (\text{card } A_n)}{(\text{card } \Omega_1)^n} \\ &= \frac{\text{card } A_1}{36} \times \cdots \times \frac{\text{card } A_n}{36} \\ &= P_1(A_1) \times \cdots \times P_1(A_n). \end{aligned} \quad (2.2)$$

Appliquons ceci à l'événement E_n défini par

$$E_n := \{\text{les } n-1 \text{ premiers lancers ne donnent ni 7 ni 9 et le } n^{\text{e}} \text{ donne 9}\}. \quad (2.3)$$

On peut représenter⁹ E_n dans Ω_n en écrivant $E_n = H_1 \times H_2 \times \cdots \times H_{n-1} \times F_n$, donc

$$P_n(E_n) = \frac{1}{9} \left(\frac{13}{18} \right)^{n-1},$$

9. L'événement E_n est défini par la phrase entre les accolades dans (2.3). Son écriture ensembliste dépend de l'espace Ω considéré. D'habitude, on ne fait pas cette distinction parce qu'on travaille avec un seul Ω , mais ici elle s'impose.

formule valable pour tout $n \geq 1$, le cas $n = 1$ se réduisant à $E_1 = F_1$ et $P_1(F_1) = 1/9$.

Pour $1 \leq k < n$, on peut représenter E_k dans Ω_n en écrivant $E_k = H_1 \times H_2 \times \cdots \times H_{k-1} \times F_k \times \Omega_1^{n-k}$. La formule (2.2) nous donne alors

$$\forall n \geq k, \quad P_n(E_k) = P_1(H_1) \times \cdots \times P_1(H_{k-1}) \times P_1(F_k) \times 1^{n-m} = P_k(E_k), \quad (2.4)$$

où dans l'écriture « $P_k(E_k)$ », on interprète E_k comme une partie de Ω_k , explicitement $H_1 \times H_2 \times \cdots \times H_{k-1} \times F_k$. Ainsi (2.4) exprime une sorte de *compatibilité ascendante* des modèles $(\Omega_n, \mathcal{P}(\Omega_n), P_n)$. Si on note E'_n l'évènement *au cours des n premiers lancers, la première obtention de la somme 9 a lieu avant la première obtention de la somme 7*, on a grâce à cette compatibilité

$$\forall j \geq n, \quad P_j(E'_n) = P_j\left(\bigcup_{k=1}^n E_k\right) = \sum_{k=1}^n P_j(E_k), \quad (2.5)$$

en notant que les E_k sont deux à deux disjoints et en utilisant l'additivité de P_j .

Pour attribuer une probabilité à E , on ne peut malheureusement pas se contenter de (2.5), même avec n « grand ». En effet on ne peut pas exclure que la question de la priorité entre le 9 et le 7 ne soit tranchée qu'au delà du n^{e} lancer, voire jamais. Ceci nous conduit à prendre pour Ω l'ensemble des suites infinies de couples $(i, j) \in \{1, \dots, 6\}^2$, autrement dit, $\Omega = \Omega_1^{\mathbb{N}^*}$. Notons $\mathbb{N}_k := \{j \in \mathbb{N}; j > k\}$. On peut alors représenter E_k dans Ω par $E_k = H_1 \times H_2 \times \cdots \times H_{k-1} \times F_k \times \Omega_1^{\mathbb{N}_k}$. Au vu de la formule de compatibilité (2.4), il est naturel d'attribuer une probabilité à E_k considéré comme évènement de Ω en définissant

$$P(E_k) := P_k(E_k) = \frac{1}{9} \left(\frac{13}{18}\right)^{k-1}.$$

Finalement, pour pouvoir attribuer une probabilité à E , on remarque que E est l'union des E_k pour $k \in \mathbb{N}^*$, ces ensembles étant deux à deux disjoints. On a donc bien envie d'écrire

$$P(E) = P\left(\bigcup_{k \in \mathbb{N}^*} E_k\right) = \sum_{k \in \mathbb{N}^*} P(E_k) = \sum_{k=1}^{+\infty} \frac{1}{9} \left(\frac{13}{18}\right)^{k-1} = \frac{1}{9} \frac{1}{1 - \frac{13}{18}} = \frac{2}{5}.$$

La propriété qui nous manque pour justifier cette écriture est précisément la σ -additivité.

Une autre façon d'obtenir le même résultat est de passer à la limite quand n tend vers l'infini dans (2.5), après l'avoir réécrit sous la forme

$$P(E'_n) = \sum_{k=1}^n P(E_k).$$

Remarquons que la suite d'ensembles E'_n est croissante pour l'inclusion ($E'_n \subset E'_{n+1}$) et que $\bigcup_{n \geq 1} E'_n = E$. La propriété qui nous permettrait de faire cela s'appelle continuité séquentielle croissante de P (voir prop. 2.16-6.a).

On voit ainsi qu'il est souhaitable de définir la probabilité P sur Ω comme une *mesure*. Ceci pose la question de la tribu \mathcal{F} . En fait tout ce que nous savons faire

ici, c'est définir la probabilité d'événements comme les E_k , de la forme $A \times \Omega_1^{\mathbb{N}_k}$, avec $A \subset \Omega_k$. Ces événements sont ceux dont la réalisation ne dépend que du résultat des k premières épreuves, pour un certain k . On peut dire aussi qu'il s'agit des événements dont la réalisation ne dépend que d'un nombre fini d'épreuves. Notons \mathcal{C} la famille de ces événements. Un événement comme E n'appartient pas à \mathcal{C} , mais à la tribu engendrée par \mathcal{C} , $\mathcal{F} = \sigma(\mathcal{C})$. Cette tribu est plus petite¹⁰ que $\mathcal{P}(\Omega)$.

Enfin, soit H l'événement *aucun lancer ne produit la somme 7 ou la somme 9*. Avec

$$K_n := H_1 \times \cdots \times H_n \times \Omega_1^{\mathbb{N}_n},$$

on a clairement

$$H = H_1^{\mathbb{N}^*} = \bigcap_{n \geq 1} K_n.$$

L'ensemble H apparaît ainsi comme intersection *dénombrable* d'ensembles $K_n \in \mathcal{C}$, ce qui nous assure de l'appartenance de H à la tribu \mathcal{F} . On a de plus $H \subset K_n$ pour tout $n \geq 1$, donc par croissance de P pour l'inclusion (c'est une conséquence immédiate de l'additivité, cf. prop. 2.16-4.),

$$\forall n \geq 1, \quad P(H) \leq \left(\frac{13}{18}\right)^n.$$

Cette inégalité étant vérifiée pour tout n , on peut faire tendre n vers l'infini pour obtenir $P(H) = 0$. On a ainsi un exemple d'événement non vide et de probabilité nulle (H est l'ensemble des suites infinies de couples éléments de H_1 , il est infini non dénombrable).

2.3 La probabilité comme mesure

La *probabilité* P , telle que nous allons la définir ci-dessous, est une fonction qui à un événement, associe un nombre compris entre 0 et 1 et censé mesurer les chances de réalisation de cet événement. Pour des raisons sortant du cadre de ce cours, il n'est pas toujours possible d'attribuer ainsi de manière cohérente une probabilité à chaque partie de Ω . En d'autres termes, P ne peut pas être considérée comme une *application* de l'ensemble $\mathcal{P}(\Omega)$ de toutes les parties de Ω dans $[0, 1]$ mais comme une *fonction* ayant pour domaine de définition une tribu \mathcal{F} généralement plus petite que $\mathcal{P}(\Omega)$. La tribu \mathcal{F} est aussi appelée famille des événements observables¹¹.

Définition 2.15. Soit Ω un ensemble et \mathcal{F} une tribu sur Ω . On appelle *probabilité* sur (Ω, \mathcal{F}) toute application P de \mathcal{F} dans $[0, 1]$ vérifiant :

- (i) $P(\Omega) = 1$.

10. Nous l'admettons, mais si vous n'êtes pas convaincu, essayez de montrer que $\mathcal{F} = \mathcal{P}(\Omega)$...

11. La définition générale d'une tribu \mathcal{F} ne suppose pas que tous les singletons $\{\omega\}$ soient des éléments de \mathcal{F} . Donc un « événement élémentaire » n'est pas toujours un événement observable. Néanmoins dans la plupart des exemples que nous étudierons, la tribu possèdera les singletons.

(ii) Pour toute suite $(A_j)_{j \geq 1}$ d'événements de \mathcal{F} deux à deux disjoints (incompatibles) :

$$P\left(\bigcup_{j \in \mathbb{N}^*} A_j\right) = \sum_{j=1}^{+\infty} P(A_j).$$

Le triplet (Ω, \mathcal{F}, P) s'appelle *espace probabilisé*.

Définir une probabilité sur (Ω, \mathcal{F}) c'est en quelque sorte attribuer une « masse » à chaque événement observable, avec par convention une masse totale égale à 1 pour l'événement certain Ω . Une formulation équivalente à la définition 2.15 est : *P est une mesure sur (Ω, \mathcal{F}) telle que $P(\Omega) = 1$.*

Proposition 2.16 (propriétés générales d'une probabilité).

Toute probabilité P sur (Ω, \mathcal{F}) vérifie les propriétés suivantes :

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

2. *Additivité*

a) Si $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$.

b) Si les A_i ($1 \leq i \leq n$) sont deux à deux disjoints :

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

3. $\forall A \in \mathcal{F}$, $P(A^c) = 1 - P(A)$.

4. $\forall A \in \mathcal{F}$, $\forall B \in \mathcal{F}$, $A \subset B \Rightarrow P(A) \leq P(B)$.

5. $\forall A \in \mathcal{F}$, $\forall B \in \mathcal{F}$, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

6. *Continuité monotone séquentielle*

a) Si $(B_n)_{n \geq 0}$ est une suite croissante d'événements de \mathcal{F} convergente¹² vers $B \in \mathcal{F}$, alors $P(B) = \lim_{n \rightarrow +\infty} P(B_n)$. Notation :

$$B_n \uparrow B \Rightarrow P(B_n) \uparrow P(B) \quad (n \rightarrow +\infty).$$

b) Si $(C_n)_{n \geq 0}$ est une suite décroissante d'événements de \mathcal{F} convergente¹³ vers $C \in \mathcal{F}$, alors $P(C) = \lim_{n \rightarrow +\infty} P(C_n)$. Notation :

$$C_n \downarrow C \Rightarrow P(C_n) \downarrow P(C) \quad (n \rightarrow +\infty).$$

7. a) $\forall A \in \mathcal{F}, \forall B \in \mathcal{F} \quad P(A \cup B) \leq P(A) + P(B)$.

b) $\forall A_1, \dots, A_n \in \mathcal{F}, \quad P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$.

c) $\forall A_1, \dots, A_n, \dots \in \mathcal{F}, \quad P\left(\bigcup_{i \in \mathbb{N}^*} A_i\right) \leq \sum_{i=1}^{+\infty} P(A_i)$.

12. Ce qui signifie : $\forall n \geq 0, B_n \subset B_{n+1}$ et $B = \bigcup_{n \geq 0} B_n$.

13. Ce qui signifie : $\forall n \geq 0, C_{n+1} \subset C_n$ et $C = \bigcap_{n \geq 0} C_n$.

Preuve. Soit P une fonction d'ensembles $\mathcal{F} \rightarrow [0, 1]$ satisfaisant aux conditions (i) et (ii) de la définition 2.15, il s'agit de démontrer que P vérifie les propriétés 1 à 7.

Preuve de 1. Comme $P(A_j) \geq 0$ pour tout $A_j \in \mathcal{F}$, on a toujours

$$\sum_{j \in \mathbb{N}^*} P(A_j) \geq P(A_1) + P(A_2),$$

le premier membre pouvant être égal à $+\infty$. En choisissant $A_j = \emptyset$ pour tout $j \in \mathbb{N}^*$ et en utilisant la σ -additivité (ii), on en déduit :

$$P(\emptyset) = P\left(\bigcup_{j \in \mathbb{N}^*} A_j\right) = \sum_{j=1}^{+\infty} P(A_j) \geq P(\emptyset) + P(\emptyset).$$

Par conséquent, $P(\emptyset) \geq 2P(\emptyset)$ et comme $P(\emptyset) \geq 0$, ceci entraîne $P(\emptyset) = 0$. \square

Preuve de 2. Soient A_1, \dots, A_n , n événements de \mathcal{F} deux à deux disjoints. Pour $j > n$, posons $A_j = \emptyset$. On a ainsi une suite infinie $(A_j)_{j \geq 1}$ d'événements deux à deux disjoints. En utilisant la σ -additivité, on obtient alors :

$$P\left(\bigcup_{j=1}^n A_j\right) = P\left(\bigcup_{j \in \mathbb{N}^*} A_j\right) = \sum_{j=1}^n P(A_j) + \sum_{j=n+1}^{+\infty} P(A_j).$$

D'après 1, la somme pour $j \geq n+1$ vaut 0, ceci prouve 2 b). Bien sûr 2 a) n'est que le cas particulier $n = 2$. \square

Preuve de 3. Prendre $B = A^c$ dans 2 a) et utiliser (i). \square

Preuve de 4. Si $A \subset B$, alors $B = A \cup (B \cap A^c)$ et cette réunion est disjointe. D'après 2 a) on a $P(B) = P(A) + P(B \cap A^c)$ et comme $P(B \cap A^c) \geq 0$, on en déduit $P(B) \geq P(A)$. \square

Preuve de 5. On a les décompositions suivantes en unions disjointes :

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

En utilisant l'additivité on en déduit :

$$\begin{aligned} P(A \cup B) &= P(A \cap B^c) + P(A \cap B) + P(A^c \cap B) \\ &= [P(A \cap B^c) + P(A \cap B)] + [P(A \cap B) + P(A^c \cap B)] \\ &\quad - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

\square

Preuve de 6. Il suffit de prouver 6 a), la propriété 6 b) s'en déduit en appliquant 6 a) à la suite d'événements $B_n = C_n^c$. Admettons, pour l'instant, que pour tout $n \geq 1$, B_n vérifie la décomposition suivante en union disjointe (cf. figure 2.2)

$$B_n = B_0 \cup \left(\bigcup_{i=1}^n (B_i \setminus B_{i-1}) \right).$$

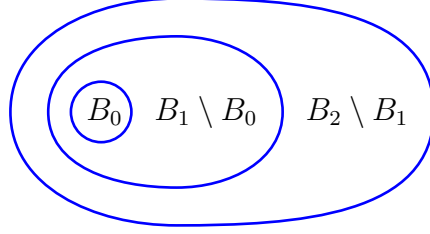


FIG. 2.2 – Décomposition de $B_0 \cup B_1 \cup B_2$ en union disjointe

En écrivant la réunion infinie des B_n à l'aide de cette décomposition et en « effaçant » toutes les répétitions des $B_i \setminus B_{i-1}$, on en déduit immédiatement que B vérifie la décomposition en union disjointe :

$$B = B_0 \cup \left(\bigcup_{i \in \mathbb{N}^*} (B_i \setminus B_{i-1}) \right).$$

Passant aux probabilités, ces deux décompositions nous donnent :

$$\begin{aligned} P(B_n) &= P(B_0) + \sum_{i=1}^n P(B_i \setminus B_{i-1}), \\ P(B) &= P(B_0) + \sum_{i=1}^{+\infty} P(B_i \setminus B_{i-1}). \end{aligned}$$

Comme cette série converge, sa somme est la limite de la suite de ses sommes partielles de rang n , ce qui s'écrit :

$$P(B) = \lim_{n \rightarrow +\infty} \left\{ P(B_0) + \sum_{i=1}^n P(B_i \setminus B_{i-1}) \right\} = \lim_{n \rightarrow +\infty} P(B_n).$$

Ainsi pour compléter la preuve, il ne reste plus qu'à justifier la décomposition de B_n . Posons :

$$D_n = B_0 \cup \left(\bigcup_{i=1}^n (B_i \setminus B_{i-1}) \right).$$

Pour montrer que $B_n = D_n$, il suffit de montrer que $D_n \subset B_n$ et $B_n \subset D_n$. La première inclusion est évidente puisque pour tout $i \leq n$, $B_i \setminus B_{i-1} \subset B_i \subset B_n$. Pour prouver l'inclusion inverse, on note ω un élément *quelconque* de B_n et on montre que ω appartient à D_n . Soit $i_0 = i_0(\omega)$ le plus petit des indices i tels que $\omega \in B_i$. Comme cet ensemble

d'indices contient au moins n , on a $0 \leq i_0 \leq n$. Si $i_0 = 0$, $\omega \in B_0$ et comme $B_0 \subset D_n$, $\omega \in D_n$. Si $i_0 \geq 1$, par la définition même de i_0 , on a $\omega \in B_{i_0}$ et $\omega \notin B_{i_0-1}$, donc $\omega \in B_{i_0} \setminus B_{i_0-1}$ et comme $i_0 \leq n$, $B_{i_0} \setminus B_{i_0-1} \subset D_n$ donc $\omega \in D_n$. Le raisonnement précédent étant valable pour tout ω de B_n , on en déduit $B_n \subset D_n$. \square

Preuve de 7 a). D'après 5 :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B),$$

car $P(A \cap B) \geq 0$. \square

Preuve de 7 b). On remarque que pour tout $n \geq 1$ on a :

$$\bigcup_{i=1}^n A_i = \bigcup_{i=1}^n B_i,$$

où les B_i sont des événements *deux à deux disjoints* définis comme suit :

$$B_1 = A_1, B_2 = A_2 \cap B_1^c, B_3 = A_3 \cap (B_1 \cup B_2)^c, \dots, B_n = A_n \cap (B_1 \cup \dots \cup B_{n-1})^c.$$

Par additivité :

$$P\left(\bigcup_{i=1}^n A_i\right) = P\left(\bigcup_{i=1}^n B_i\right) = \sum_{i=1}^n P(B_i).$$

Par construction pour tout i , $B_i \subset A_i$, d'où $P(B_i) \leq P(A_i)$ et finalement,

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(B_i) \leq \sum_{i=1}^n P(A_i).$$

\square

Preuve de 7 c). Posons pour tout $n \geq 1$,

$$D_n = \bigcup_{i=1}^n A_i, \quad D = \bigcup_{n \geq 1} D_n = \bigcup_{i \in \mathbb{N}^*} A_i.$$

La suite $(D_n)_{n \geq 1}$ est *croissante* et a pour limite D . Donc d'après 6 a), $P(D_n) \uparrow P(D)$ ($n \rightarrow +\infty$). D'après 7 b) on a :

$$\forall n \geq 1, \quad P(D_n) \leq \sum_{i=1}^n P(A_i).$$

Les deux membres de cette inégalité étant les termes généraux de deux suites croissantes de réels positifs, on obtient en passant à la limite quand n tend vers $+\infty$:

$$P\left(\bigcup_{i \in \mathbb{N}^*} A_i\right) = P(D) \leq \sum_{i=1}^{+\infty} P(A_i).$$

Ce qui prouve 7 c). Remarquons que les sommes partielles de la série convergent dans $\overline{\mathbb{R}}_+$. Bien sûr, l'inégalité obtenue n'a d'intérêt que lorsque la série de terme général $P(A_i)$ converge et a une somme strictement inférieure à 1. \square

La vérification de la proposition 2.16 est maintenant complète. \square

Remarque 2.17 (propriétés d'une mesure). Les propriétés d'une probabilité énoncées par la proposition 2.16 s'étendent à une mesure positive quelconque, avec les exceptions suivantes. La propriété 3 est valable seulement pour $\mu(\Omega)$ fini en remplaçant 1 par $\mu(\Omega)$. La propriété 5 est vraie à condition que $\mu(A \cap B)$ soit fini. Pour la continuité séquentielle décroissante 6 b), il faut rajouter l'hypothèse $\mu(C_0) < +\infty$.

Le calcul de probabilités de réunions ou d'intersections est une question cruciale. La propriété 5 montre qu'en général on ne peut pas calculer $P(A \cup B)$ à partir de la *seule connaissance* de $P(A)$ et $P(B)$ et qu'on se heurte à la même difficulté pour $P(A \cap B)$. Le calcul des probabilités d'intersections sera discuté plus tard, à propos du conditionnement. Pour les probabilités de réunions, on peut se demander comment se généralise la propriété 5 lorsqu'on réunit plus de deux événements. Il est facile de vérifier (faites-le!) que :

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

Le cas général est donné par la formule de Poincaré qui exprime $P(A_1 \cup \dots \cup A_n)$ à l'aide des probabilités de *toutes les intersections* des A_i : 2 à 2, 3 à 3, etc.

Proposition 2.18 (formule de Poincaré).

Pour tout entier $n \geq 2$ et tous événements A_1, \dots, A_n :

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) + \sum_{k=2}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}). \quad (2.6)$$

Preuve. On raisonne par récurrence¹⁴. La formule est vraie pour $n = 2$, car dans ce cas elle se réduit à

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (2.7)$$

Supposons la formule de Poincaré vraie au rang n (plus précisément on suppose que pour toute suite de n événements A_1, \dots, A_n , l'égalité (2.6) est vérifiée). Pour en déduire qu'elle est alors vraie au rang $n + 1$, il nous faut calculer $P\left(\bigcup_{i=1}^{n+1} A_i\right)$. On commence par appliquer (2.7) avec $A = \bigcup_{i=1}^n A_i$ et $B = A_{n+1}$. On obtient ainsi :

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P\left(\bigcup_{i=1}^n A_i\right) + P(A_{n+1}) - P\left(\left[\bigcup_{i=1}^n A_i\right] \cap A_{n+1}\right) \\ &= P\left(\bigcup_{i=1}^n A_i\right) + P(A_{n+1}) - P\left(\bigcup_{i=1}^n (A_i \cap A_{n+1})\right). \end{aligned}$$

On applique maintenant l'hypothèse de récurrence (formule de Poincaré (2.6)) d'abord avec les n événements A_1, \dots, A_n puis avec les n événements A'_1, \dots, A'_n , où l'on a posé

14. Il y a une autre méthode plus élégante utilisant l'écriture d'une probabilité comme espérance d'une fonction indicatrice. Elle pourra être vue ultérieurement en exercice.

$A'_i := A_i \cap A_{n+1}$. Il vient :

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= \sum_{i=1}^n P(A_i) + \sum_{k=2}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}) \\ &\quad + P(A_{n+1}) \\ &\quad - \sum_{i=1}^n P(A'_i) - \sum_{j=2}^n (-1)^{j+1} \sum_{1 \leq i_1 < i_2 < \dots < i_j \leq n} P(A'_{i_1} \cap \dots \cap A'_{i_j}) \\ &= \sum_{i=1}^{n+1} P(A_i) \end{aligned} \tag{2.8}$$

$$+ \sum_{k=2}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}) \tag{2.9}$$

$$+ (-1)^{2+1} \sum_{i=1}^n P(A_i \cap A_{n+1}) \tag{2.10}$$

$$+ \sum_{j=2}^n (-1)^{(j+1)+1} \sum_{1 \leq i_1 < i_2 < \dots < i_j \leq n} P(A_{i_1} \cap \dots \cap A_{i_j} \cap A_{n+1}) \tag{2.11}$$

Comparons ce résultat avec ce que l'on espère trouver, c'est-à-dire avec

$$\underbrace{\sum_{i=1}^{n+1} P(A_i) + \sum_{k=2}^{n+1} (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n+1} P(A_{i_1} \cap \dots \cap A_{i_k})}_{=: T_{n+1}}.$$

Cela revient à vérifier que T_{n+1} est égal à la somme des lignes (2.9) à (2.11) ci-dessus. Partageons T_{n+1} en deux blocs comme suit. Le premier bloc regroupe tous les termes tels que $i_k < n+1$ (et donc $i_k \leq n$ et $k \leq n$). On le retrouve exactement à la ligne (2.9). Le deuxième bloc regroupe tous les termes pour lesquels $i_k = n+1$. Dans ce bloc, la somme des termes pour lesquels $k = 2$ se retrouve ligne (2.10). Il reste alors la somme des termes pour lesquels $3 \leq k \leq n+1$ et $i_k = n+1$ (donc $i_{k-1} \leq n$). Cette somme est exactement le contenu de la ligne (2.11), comme on peut le voir en faisant le changement d'indice $k = j+1$ dans (2.11). Ceci achève la récurrence. \square

2.4 Exemples

Nous examinons maintenant quelques exemples d'espaces probabilisés et de calcul de probabilités d'événements.

Exemple 2.19. On effectue une partie de pile ou face en trois coups. Quelle est la probabilité d'obtenir pile aux premier et troisième lancers ?

On peut modéliser cette expérience en prenant $\Omega = \{\mathbf{f}, \mathbf{p}\}^3$ et pour famille d'événements observables $\mathcal{F} = \mathcal{P}(\Omega)$ l'ensemble de toutes les parties¹⁵ de Ω . La pièce étant

15. Lorsque Ω est fini, il est toujours possible de faire ce choix.

supposée symétrique, nous n'avons *a priori* pas de raison de supposer que l'un des 8 triplets de résultats possibles soit favorisé ou défavorisé par rapport aux autres. Nous choisissons donc P de sorte que tous les événements élémentaires aient même probabilité (hypothèse d'équiprobabilité), soit :

$$\forall \omega \in \Omega, \quad P(\{\omega\}) = \frac{1}{\text{Card } \Omega} = \frac{1}{2^3}.$$

L'événement B dont on veut calculer la probabilité s'écrit :

$$B = \{(\mathbf{p}, \mathbf{f}, \mathbf{p}); (\mathbf{p}, \mathbf{p}, \mathbf{p})\}.$$

D'où :

$$P(B) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.$$

Exemple 2.20. On fait remplir un questionnaire à 20 questions binaires. Quelle est la probabilité qu'un candidat répondant au hasard obtienne au moins 16 bonnes réponses ? On choisit ici :

$$\Omega = \{\text{oui}, \text{non}\}^{20}, \quad \mathcal{F} = \mathcal{P}(\Omega).$$

Si le candidat répond complètement au hasard, on peut considérer que chacune des 2^{20} grilles de réponses possibles a la même probabilité d'apparaître (hypothèse d'équiprobabilité sur Ω). Pour tout $B \subset \Omega$, on a alors :

$$P(B) = \frac{\text{Card } B}{\text{Card } \Omega}.$$

En particulier pour $B = \{\text{obtention d'au moins 16 bonnes réponses}\}$,

$$P(B) = \frac{C_{20}^{16} + C_{20}^{17} + C_{20}^{18} + C_{20}^{19} + C_{20}^{20}}{2^{20}} = \frac{6196}{2^{20}} \simeq 0,006.$$

Exemple 2.21 (contrôle de production). On prélève au hasard un échantillon de k pièces dans une production totale de N pièces comprenant en tout n pièces défectueuses. Le prélèvement est *sans remise* (donc $k \leq N$). Cherchons la probabilité de :

$$A_j = \{\text{il y a exactement } j \text{ pièces défectueuses dans l'échantillon}\}.$$

On prend pour Ω l'ensemble de toutes les parties à k éléments d'un ensemble à N éléments (ensemble de tous les échantillons possibles de taille k), $\mathcal{F} = \mathcal{P}(\Omega)$ et P l'équiprobabilité sur Ω . Il suffit alors de dénombrer tous les échantillons ayant exactement j pièces défectueuses. Un tel échantillon se construit en prenant j pièces dans le sous-ensemble des défectueuses (C_n^j choix possibles) et en complétant par $k - j$ pièces prises dans le sous-ensemble des non-défectueuses (C_{N-n}^{k-j} choix possibles). On en déduit :

$$P(A_j) = \frac{C_n^j C_{N-n}^{k-j}}{C_N^k} \quad \text{si} \quad \begin{cases} 0 \leq j \leq n, \\ 0 \leq j \leq k, \\ k - j \leq N - n. \end{cases}$$

Si l'une de ces trois conditions n'est pas vérifiée, $P(A_j) = 0$.

Remarque 2.22. Lorsque Ω est fini, la façon la plus simple de construire une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ est de choisir $P(\{\omega\}) = 1/\text{card } \Omega$. On parle alors d'*équiprobabilité* ou de *probabilité uniforme* sur $(\Omega, \mathcal{P}(\Omega))$. C'est la modélisation qui s'impose naturellement lorsqu'on n'a pas de raison de penser *a priori* qu'un résultat élémentaire de l'expérience soit favorisé ou défavorisé par rapport aux autres. La situation est radicalement différente lorsque Ω est infini *dénombrable*. Sur un tel ensemble, *il ne peut pas y avoir d'équiprobabilité*. Imaginons que l'on veuille tirer une boule *au hasard* dans une urne contenant une infinité de boules ayant chacune un numéro entier distinct (et une boule par entier). Soit $\{\omega_i\}$ l'événement *tirage de la boule numérotée i* ($i \in \mathbb{N}$) et p_i sa probabilité. Par σ -additivité, on doit toujours avoir :

$$\sum_{i \in \mathbb{N}} p_i = 1.$$

Mais si tous les p_i sont égaux, la série ci-dessus contient une infinité de termes tous égaux à p_0 . Sa somme vaut alors $+\infty$ si $p_0 > 0$ ou 0 si $p_0 = 0$, il y a donc une contradiction.

Voici maintenant une caractérisation de toutes les probabilités sur les espaces au plus dénombrables.

Proposition 2.23. *Soit $\Omega = \{\omega_i; i \in I\}$ un ensemble au plus dénombrable. La donnée d'une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ équivaut à la donnée d'une famille $(p_i)_{i \in I}$ dans \mathbb{R}_+ telle que :*

$$\sum_{i \in I} p_i = 1$$

et des égalités

$$P(\{\omega_i\}) = p_i, \quad i \in I.$$

La probabilité P s'écrit alors $P = \sum_{i \in I} p_i \delta_{\omega_i}$, où δ_{ω} désigne la masse de Dirac (ou mesure de Dirac) au point ω , définie sur $\mathcal{P}(\Omega)$ par $\delta_{\omega}(A) = 1$ si $\omega \in A$ et $\delta_{\omega}(A) = 0$ si $\omega \notin A$.

Preuve. En préliminaire, notons les faits suivants relatifs aux mesures de Dirac sur Ω .

$$\forall \omega \in \Omega, \quad \delta_{\omega}(\Omega) = 1. \quad (2.12)$$

$$\forall i, j \in I, \quad \delta_{\omega_i}(\{\omega_j\}) = \begin{cases} 0 & \text{si } i \neq j, \\ 1 & \text{si } i = j. \end{cases} \quad (2.13)$$

Soit P une probabilité sur $(\Omega, \mathcal{P}(\Omega))$. Comme la tribu $\mathcal{P}(\Omega)$ possède les singletons, $P(\{\omega\})$ est défini et fini (puisque majoré par 1). La mesure P est donc finie sur les singletons et d'après la remarque 2.10, c'est une mesure ponctuelle. Cela signifie qu'il existe $J \subset I$ (donc au plus dénombrable) et une famille $\{p_i\}_{i \in J}$ de \mathbb{R}_+ telle que $P = \sum_{i \in J} p_i \delta_{\omega_i}$. On peut compléter cette écriture en posant $p_i := 0$ pour $i \in I \setminus J$, pour obtenir

$$P = \sum_{i \in I} p_i \delta_{\omega_i}. \quad (2.14)$$

En écrivant que $P(\Omega) = 1$ et en utilisant (2.12), il vient $\sum_{i \in I} p_i = 1$. D'autre part on voit grâce à (2.13) que pour tout $i \in I$, $P(\{\omega_i\}) = p_i$.

Réciproquement, donnons nous une famille $(p_i)_{i \in I}$ de réels positifs de somme 1 et définissons la mesure P par (2.14). Grâce à (2.12), il est clair que $P(\Omega) = \sum_{i \in I} p_i = 1$, donc P est une probabilité. De plus pour tout $i \in I$, $P(\{\omega_i\}) = p_i$ en utilisant (2.13). \square

Exemple 2.24 (une probabilité définie sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$).

Soit a un réel strictement positif fixé. On pose :

$$\forall k \in \mathbb{N}, \quad p_k = \frac{e^{-a} a^k}{k!}.$$

On remarque que p_k est le terme général positif d'une série convergente :

$$\sum_{k=0}^{+\infty} \frac{e^{-a} a^k}{k!} = e^{-a} \sum_{k=0}^{+\infty} \frac{a^k}{k!} = e^{-a} e^a = 1.$$

Pour tout $A \subset \mathbb{N}$, on définit :

$$P(A) = \sum_{k \in A} p_k = \sum_{k \in \mathbb{N}} p_k \delta_k(A).$$

D'après la proposition 2.23, P est une probabilité sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. On l'appelle *loi de Poisson de paramètre a* . Calculons par exemple $P(2\mathbb{N})$ où $2\mathbb{N}$ désigne l'ensemble des entiers pairs.

$$P(2\mathbb{N}) = \sum_{k \in 2\mathbb{N}} p_k = \sum_{l=0}^{+\infty} \frac{e^{-a} a^{2l}}{(2l)!} = e^{-a} \operatorname{ch} a = \frac{1}{2}(1 + e^{-2a}).$$

Une conséquence de ce résultat est : si l'on tire un nombre entier au hasard suivant une loi de Poisson, la probabilité qu'il soit pair est strictement supérieure à $\frac{1}{2}$.

Exemple 2.25 (loi uniforme sur un segment). Prenons $\Omega = \mathbb{R}$, $\mathcal{F} = \operatorname{Bor}(\mathbb{R})$ et rappelons que λ_1 désigne la mesure de Lebesgue sur \mathbb{R} (cf. exemple 2.13). Soit $[a, b]$ un segment fixé de \mathbb{R} . On définit une probabilité P sur $(\mathbb{R}, \operatorname{Bor}(\mathbb{R}))$ en posant :

$$\forall A \in \operatorname{Bor}(\mathbb{R}), \quad P(A) = \frac{\lambda_1(A \cap [a, b])}{\lambda_1([a, b])} = \frac{\lambda_1(A \cap [a, b])}{b - a}. \quad (2.15)$$

D'après l'exemple 2.12, P est une mesure et comme $P(\mathbb{R}) = 1$, c'est une probabilité. On l'appelle *loi uniforme sur $[a, b]$* . Remarquons que pour cette probabilité, tout singleton est de probabilité nulle ($\forall x \in \mathbb{R}, P(\{x\}) = 0$), ce qui résulte de la propriété analogue de λ_1 . On voit sur cet exemple que la probabilité d'une union infinie *non dénombrable* d'événements deux à deux disjoints n'est pas forcément égale à la somme de la famille correspondante de probabilités d'événements. En effet,

$$1 = P([a, b]) = P\left(\bigcup_{x \in [a, b]} \{x\}\right) \neq \sum_{x \in [a, b]} P(\{x\}) = 0.$$

Exemple 2.26 (lois uniformes dans \mathbb{R}^d). On peut généraliser l'exemple précédent en prenant au lieu d'un segment un borélien B de \mathbb{R}^d , i.e. $B \in \text{Bor}(\mathbb{R}^d)$, tel que $0 < \lambda_d(B) < +\infty$. On définit alors une probabilité P sur $(\mathbb{R}^d, \text{Bor}(\mathbb{R}^d))$, en posant :

$$\forall A \in \text{Bor}(\mathbb{R}^d), \quad P(A) = \frac{\lambda_d(A \cap B)}{\lambda_d(B)}. \quad (2.16)$$

Cette probabilité est appelée *loi uniforme sur B* .

Nous allons donner maintenant une caractérisation de *toutes les probabilités* P sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$. Comme le montre l'exemple de la loi uniforme sur un segment, il est illusoire d'espérer caractériser ces probabilités par les $P(\{x\})$. La situation est donc radicalement différente du cas d'un espace Ω au plus dénombrable. Au lieu des $P(\{x\})$, nous allons utiliser les $P([a, b])$, ou les $P([-\infty, x])$. Nous aurons besoin pour cela de la notion de fonction de répartition.

Définition 2.27 (fonction de répartition). Soit P une probabilité sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$. On appelle *fonction de répartition de P* , l'application

$$F : \mathbb{R} \rightarrow [0, 1], \quad x \mapsto F(x) := P([-\infty, x]).$$

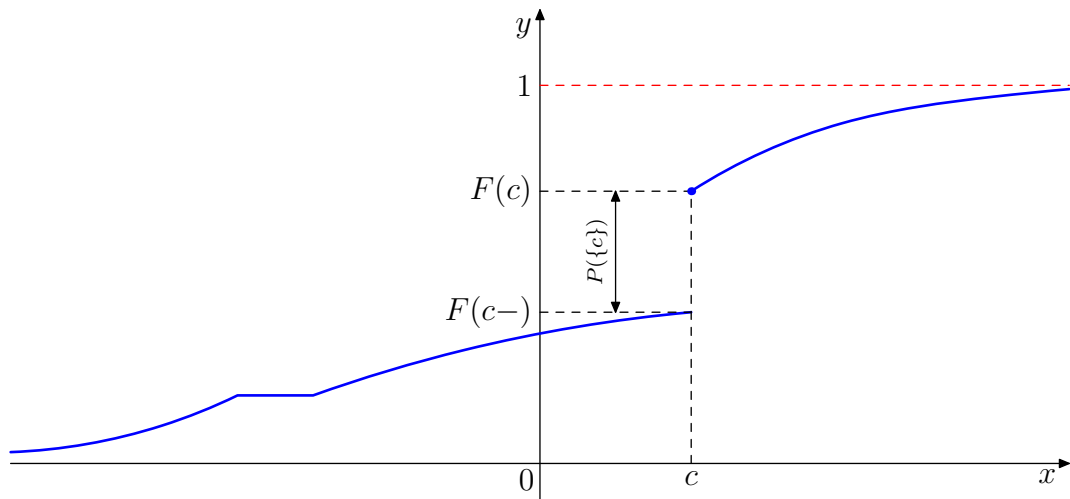


FIG. 2.3 – Une fonction de répartition avec une discontinuité au point $x = c$

Voici les propriétés générales des fonctions de répartition.

Proposition 2.28. La fonction de répartition F d'une probabilité P sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ a les propriétés suivantes.

- a) F est croissante sur \mathbb{R} .
- b) F a pour limite 0 en $-\infty$ et 1 en $+\infty$.
- c) F est continue à droite sur \mathbb{R} et a une limite à gauche en tout point $x \in \mathbb{R}$.

d) En notant $F(x-) := \lim_{\varepsilon \downarrow 0} F(x - \varepsilon)$ la limite à gauche¹⁶ au point x , on a

$$\forall x \in \mathbb{R}, \quad P(\{x\}) = F(x) - F(x-). \quad (2.17)$$

De plus l'ensemble des $x \in \mathbb{R}$ tels que $F(x) \neq F(x-)$ est au plus dénombrable.

e) Si deux probabilités P_1 et P_2 sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ ont même fonction de répartition F , elles sont égales, i.e. $P_1(B) = P_2(B)$ pour tout $B \in \text{Bor}(\mathbb{R})$.

Preuve.

Preuve du a). La croissance de F sur \mathbb{R} est une conséquence immédiate de la croissance de P pour l'inclusion (prop. 2.16 4). En effet si $x \leq x'$, $] - \infty, x] \subset] - \infty, x']$, d'où

$$F(x) = P(] - \infty, x]) \leq P(] - \infty, x']) = F(x').$$

Ainsi F est croissante sur \mathbb{R} . Il en résulte qu'elle possède une limite à gauche et une limite à droite en tout point x de \mathbb{R} (cf. cours d'analyse). \square

Preuve du b). Comme F est croissante, elle admet des limites en $-\infty$ et $+\infty$. On identifie la limite en $-\infty$ (dont on connaît l'existence) grâce à une suite particulière :

$$\lim_{x \rightarrow -\infty} F(x) = \lim_{n \rightarrow +\infty} F(-n) = \lim_{n \rightarrow +\infty} P(] - \infty, -n]).$$

On utilise la continuité séquentielle décroissante de P , i.e. la proposition 2.16 7 b) avec $B_n =] - \infty, -n]$, en vérifiant que :

$$\bigcap_{n \in \mathbb{N}^*}] - \infty, -n] = \emptyset. \quad (2.18)$$

En effet, soit x un élément de cette intersection. Alors $x \leq -n$ pour tout $n \in \mathbb{N}^*$ donc en passant à la limite, $x = -\infty$, ce qui est impossible puisque x est un réel. Donc cette intersection est vide et $\lim_{n \rightarrow +\infty} P(] - \infty, -n]) = P(\emptyset) = 0$. La preuve de $\lim_{x \rightarrow +\infty} F(x) = 1$ est analogue et est laissée au lecteur. \square

Preuve du c). Soit $x \in \mathbb{R}$ fixé. Comme on est assuré de l'existence de la limite à droite de F en ce point, pour montrer que cette limite vaut $F(x)$ et établir la continuité à droite de F en x , il suffit de vérifier que : $F(x) = \lim_{n \rightarrow +\infty} F(x + 1/n)$. Comme $F(x + 1/n) = P(] - \infty, x + 1/n])$, ceci résulte de la continuité séquentielle décroissante de P appliquée aux événements $B_n =] - \infty, x + 1/n]$ en remarquant que :

$$\bigcap_{n \in \mathbb{N}^*}] - \infty, x + \frac{1}{n}] =] - \infty, x]. \quad (2.19)$$

En effet, pour tout $y \in] - \infty, x]$, on a $y \leq x \leq x + 1/n$ pour tout $n \in \mathbb{N}^*$ et donc $y \in \bigcap_{n \in \mathbb{N}^*} B_n$ d'où $] - \infty, x]$ est inclus dans l'intersection des B_n ($n \in \mathbb{N}^*$). Réciproquement tout y dans cette intersection vérifie : $\forall n \in \mathbb{N}^*, y \leq x + 1/n$. Le passage à la limite quand n tend vers l'infini conservant cette inégalité large, on en déduit $y \leq x$ soit encore $y \in] - \infty, x]$. Ainsi l'intersection des B_n , ($n \in \mathbb{N}^*$) est incluse dans $] - \infty, x]$, ce qui achève la vérification de (2.19). \square

16. Cette notation est un peu abusive, puisqu'il ne s'agit pas forcément d'une valeur prise par la fonction F . Attention à ne pas confondre le « $x-$ » dans $F(x-)$ avec le « x^- » partie négative de x .

Preuve du d). On remarque que pour tout $x \in \mathbb{R}$ et tout $n \in \mathbb{N}^*$,

$$P\left(\left]x - \frac{1}{n}, x\right]\right) = P([-\infty, x]) - P\left(\left] - \infty, x - \frac{1}{n}\right]\right) = F(x) - F\left(x - \frac{1}{n}\right).$$

On vérifie d'autre part que :

$$\{x\} = \bigcap_{n \in \mathbb{N}^*} \left]x - \frac{1}{n}, x\right]. \quad (2.20)$$

En effet, le premier membre de (2.20) est clairement inclus dans le second. Réciproquement, si y est un élément quelconque de cette intersection, pour tout $n \in \mathbb{N}^*$, $x - 1/n < y \leq x$, donc en passant à la limite quand n tend vers l'infini, $x \leq y \leq x$ d'où $y = x$. Ceci termine la justification de (2.20). Comme on sait que F a une limite à gauche au point x , on en déduit par continuité de P pour la suite décroissante d'événements $B_n =]x - 1/n, x]$:

$$P(\{x\}) = \lim_{n \rightarrow +\infty} P\left(\left]x - \frac{1}{n}, x\right]\right) = \lim_{n \rightarrow +\infty} (F(x) - F(x - 1/n)) = F(x) - F(x-).$$

Enfin, considérons la famille de réels *positifs* $\{P(\{x\}), x \in \mathbb{R}\}$. Pour toute partie finie K de \mathbb{R} ,

$$S_K := \sum_{x \in K} P(\{x\}) = P(K) \leq 1.$$

Les sommes finies S_K étant ainsi uniformément bornées par 1, la famille est sommable par la proposition 1.83 a). On sait qu'alors l'ensemble des éléments non nuls de cette famille est au plus dénombrable, cf. proposition 1.85. \square

Nous admettons le e), dont la preuve sort du programme de ce cours. \square

Remarque 2.29. Les probabilités d'intervalles s'expriment facilement à l'aide de la fonction de répartition. Il faut prendre garde aux extrémités. En général on a la égalités suivantes pour une probabilité P sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$, de fonction de répartition F , avec a et b réels quelconques tels que $a < b$,

$$P([a, b]) = F(b) - F(a), \quad (2.21)$$

$$P([a, b]) = F(b) - F(a-), \quad (2.22)$$

$$P([a, b]) = F(b-) - F(a-), \quad (2.23)$$

$$P(]a, b]) = F(b-) - F(a), \quad (2.24)$$

$$P(]-\infty, a]) = F(a-), \quad (2.25)$$

$$P(]b, +\infty]) = 1 - F(b), \quad (2.26)$$

$$P([b, +\infty]) = 1 - F(b-). \quad (2.27)$$

Ces formules n'ont pas à être apprises par coeur. Elles se retrouvent avec un peu de réflexion. Elles se simplifient lorsque F est continue en a ou b . En particulier si F est continue aux points a et b , les 4 intervalles d'extrémités a et b ont même probabilité.

Nous admettrons le théorème suivant qui permet de construire toutes les probabilités sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$.

Théorème 2.30. *Soit F une fonction croissante et continue à droite sur \mathbb{R} , ayant pour limite 0 en $-\infty$ et 1 en $+\infty$. Il existe une unique probabilité P sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ telle que*

$$\forall a, b \in \mathbb{R}, \text{ avec } a \leq b, \quad P([a, b]) = F(b) - F(a).$$

Alors F est la fonction de répartition de P .

2.5 Remarques sur le choix d'un modèle

Envisageons la situation suivante « on jette deux dés ... ». Un événement élémentaire associé à cette expérience est la sortie de deux nombres entiers distincts ou non compris entre 1 et 6.

Une première modélisation consiste à choisir $\Omega = \{1, 2, 3, 4, 5, 6\}^2$, à prendre comme ensemble d'événements observables $\mathcal{F} = \mathcal{P}(\Omega)$ et à attribuer à chaque événement élémentaire $\{(i, j)\}$ la probabilité $1/36$ (modèle (1)). Il est commode de représenter Ω sous la forme d'un tableau à 36 cases correspondant chacune à un événement élémentaire. On peut alors définir la probabilité comme un rapport d'aires (modèle (1')) : $P(B) = \frac{\text{Aire de } B}{\text{Aire de } \Omega}$. Ce modèle (1) ou (1') est accepté d'autant plus facilement qu'on aura précisé que les deux dés sont distinguables (par exemple un dé rouge et un vert, ou deux dés blancs lancés chacun sur une table différente). On peut ainsi distinguer l'événement $A' = \{(2, 3)\}$ de l'événement $A'' = \{(3, 2)\}$ où la première composante désigne le résultat du dé rouge. Aux questions : quelle est la probabilité de l'événement $A := \{\text{obtenir un 2 et un 3}\}$, quelle est celle de l'événement $B := \{\text{obtenir un double 6}\}$? On répondra *naturellement* :

	1	2	3	4	5	6
1						
2			A''			
3		A'				
4						
5						
6						B

FIG. 2.4 – Modèle (1')

$$P(A) = P(\{A' \cup A''\}) = \frac{1}{36} + \frac{1}{36} = \frac{1}{18} \text{ et } P(B) = P(\{(6, 6)\}) = \frac{1}{36}.$$

Supposons maintenant que les deux dés ne sont plus distinguables, par exemple jet de deux dés blancs sur une même table ou jet de deux dés de couleurs différentes, l'observation étant notée sans tenir compte de la couleur. Voici une modélisation possible. On ne considère comme famille $\tilde{\mathcal{F}}$ d'événements observables que les parties *symétriques* de Ω (i.e. invariantes par $(i, j) \mapsto (j, i)$). Cela revient à remplacer $\Omega = \{1, \dots, 6\}^2$ par l'ensemble $\tilde{\Omega}$, dont les éléments (et donc les événements élémentaires) sont de deux types : les doubles et les paires de deux chiffres distincts. Alors $\tilde{\mathcal{F}} = \mathcal{P}(\tilde{\Omega})$ est constituée de 2^{21} événements, ce modèle est moins riche en information que le précédent (2^{36} événements). On peut *raisonnablement* penser que la couleur n'influe pas sur les résultats et attribuer pour rester cohérent avec notre première modélisation la probabilité $1/36$

aux doubles et $1/18$ aux paires de deux chiffres distincts. On voit ainsi un exemple de situation pourtant très simple où *il n'y a pas* équiprobabilité des événements élémentaires (modèle (2)). Remarquons là aussi que l'on peut donner une représentation géométrique de ce modèle en repliant le carré le long de sa diagonale $i = j$. Les événements élémentaires sont maintenant représentés par un carré ou un triangle et leur probabilité est définie comme un rapport d'aires, il s'agit de la loi uniforme sur le triangle rectangle de côté 6 (modèle (2')).

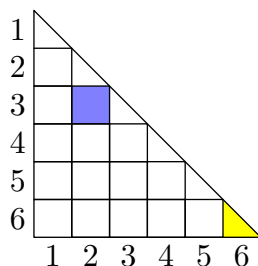


FIG. 2.5 – Modèle (2')

Un troisième modèle peut être proposé et il sera souvent adopté implicitement par des débutants à qui on aura dit : *on jette deux dés de même couleur...* On considère 21 événements élémentaires : les 6 doubles et les 15 paires à deux chiffres distincts¹⁷.

$$\Omega = \underbrace{\{1, 2, \dots, 6\}}_{\text{doubles}}, \underbrace{\{12, \dots, 16, 23, \dots, 26, 34, \dots, 56\}}_{\text{distincts}}$$

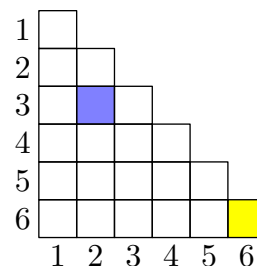


FIG. 2.6 – Modèle (3')

On prend comme ensemble d'événements observables $\mathcal{F} = \mathcal{P}(\Omega)$ ensemble des parties de Ω . On définit la probabilité P par l'hypothèse d'équiprobabilité (modèle (3)). On obtient une représentation géométrique équivalente à l'aide de la figure 2.6. Les événements élémentaires sont représentés par un carré et la probabilité est définie comme un rapport d'aires (modèle (3')). Avec ce modèle, la probabilité d'obtenir un double six est la même que celle d'obtenir un deux et un trois et vaut $\frac{1}{21}$.

On peut imaginer toute une liste d'excellents arguments qui militent en faveur des modèles (1) et (2) contre le modèle (3), par exemple : « on jette deux dés de couleurs différentes et on les filme avec deux caméras, une couleur et l'autre noir et blanc... », « il y a deux façons d'obtenir 2 et 3 et une seule de faire un double 6 ».

Tous ces arguments relèvent d'une analyse *a priori* de l'expérience et pas de la théorie mathématique des probabilités. Chacun des modèles présentés ci-dessus a sa cohérence comme objet mathématique. La question pertinente est : *parmi ces modèles, lequel (lesquels ?) représente(nt) le mieux la réalité ?*

Pour un physicien, ce qui fait la valeur d'un modèle est sa capacité à permettre la prévision du résultat d'expériences. Ainsi certaines expériences sont « expliquées » par

17. La paire « 1 et 2 » est notée « 12 »,...

la théorie de la relativité alors que leur résultat ne s'accorde pas avec la théorie de la mécanique classique.

Si l'on se place dans le cadre des modèles (1) et (2), la loi forte des grands nombres nous dit que si l'on jette les dés une infinité de fois, la fréquence d'apparition du double six va converger¹⁸ vers $\frac{1}{36}$ tandis qu'avec le modèle (3) cette fréquence convergera vers $\frac{1}{21}$. La réalisation d'un grand nombre de lancers permet donc ici de constater que les modèles (1) et (2) rendent mieux compte de la réalité que le (3). Une question importante est alors : que faut-il entendre par *grand nombre* ? Cette question sera discutée ultérieurement.

2.6 Probabilités conditionnelles

2.6.1 Introduction

Comment doit-on modifier la probabilité que l'on attribue à un événement lorsque l'on dispose d'une *information supplémentaire* ? Le concept de probabilité conditionnelle permet de répondre à cette question.

Par exemple, une fédération d'escrime regroupe N licenciés, dont N_H hommes et $N_F = N - N_H$ femmes. Il y a N_G gauchers (des deux sexes) parmi tous les licenciés. On choisit un individu au hasard. Notons :

$$\begin{aligned} G &= \{\text{l'individu choisi au hasard est gaucher}\} \\ H &= \{\text{l'individu choisi au hasard est un homme}\} \end{aligned}$$

On note $N_{G \cap H}$ le nombre d'escrimeurs hommes gauchers. On a bien sûr : $P(H) = N_H/N$ et $P(G \cap H) = N_{G \cap H}/N$. Quelle est la probabilité qu'un licencié *homme* choisi au hasard soit gaucher ? On dispose ici d'une information supplémentaire : on *sait* que l'individu choisi est un homme. En considérant les divers choix possibles d'un homme comme équiprobables, la probabilité cherchée est clairement : $N_{G \cap H}/N_H$. Ceci s'exprime aussi à l'aide de $P(H)$ et de $P(G \cap H)$ en remarquant que :

$$\frac{N_{G \cap H}}{N_H} = \frac{NP(G \cap H)}{NP(H)} = \frac{P(G \cap H)}{P(H)}.$$

Par analogie, nous donnons dans le cas général la définition formelle :

Définition 2.31 (probabilité conditionnelle). *Soit H un événement tel que $P(H) \neq 0$. Pour tout événement A , on définit :*

$$P(A | H) = \frac{P(A \cap H)}{P(H)},$$

appelée probabilité conditionnelle de l'événement A sous l'hypothèse H .

18. En un sens qui sera précisé dans le chapitre sur la loi des grands nombres.

Remarquons que pour l'instant, il ne s'agit que d'un jeu d'écriture. On a simplement défini un réel $P(A | H)$ pour que :

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

Ce qui fait l'intérêt du concept de probabilité conditionnelle, c'est qu'il est souvent bien plus facile d'attribuer *directement* une valeur à $P(A | H)$ en tenant compte des conditions expérimentales (liées à l'*information* H) et d'en déduire ensuite la valeur de $P(A \cap H)$. Le raisonnement implicite alors utilisé est : tout espace probabilisé modélisant correctement la réalité expérimentale devrait fournir telle valeur pour $P(A | H)$...

Exemple 2.32. Une urne contient r boules rouges et v boules vertes. On en tire deux l'une après l'autre, sans remise. Quelle est la probabilité d'obtenir deux rouges ?

Notons H et A les événements :

$$H = \{\text{rouge au 1}^{\text{er}} \text{ tirage}\}, \quad A = \{\text{rouge au 2}^{\text{e}} \text{ tirage}\}.$$

Un espace probabilisé (Ω, \mathcal{F}, P) modélisant correctement cette expérience devrait vérifier :

$$\begin{aligned} P(H) &= \frac{r}{r+v}, \\ P(A | H) &= \frac{r-1}{r+v-1}. \end{aligned}$$

En effet, si H est réalisé, le deuxième tirage a lieu dans une urne contenant $r+v-1$ boules dont $r-1$ rouges. On en déduit :

$$P(\text{deux rouges}) = P(A \cap H) = P(A | H)P(H) = \frac{r-1}{r+v-1} \times \frac{r}{r+v}.$$

On aurait pu arriver au même résultat en prenant pour Ω l'ensemble des arrangements de deux boules parmi $r+v$, muni de l'équiprobabilité et en faisant du dénombrement :

$$\text{card } \Omega = A_{r+v}^2 = (r+v)(r+v-1), \quad \text{card}(A \cap H) = A_r^2 = r(r-1).$$

d'où :

$$P(A \cap H) = \frac{r(r-1)}{(r+v)(r+v-1)}.$$

Notons d'ailleurs que $\text{card } H = r(r+v-1)$ d'où

$$P(H) = \frac{\text{card } H}{\text{card } \Omega} = \frac{r(r+v-1)}{(r+v)(r+v-1)} = \frac{r}{r+v}.$$

En appliquant la définition formelle de $P(A | H)$ on retrouve :

$$P(A | H) = \frac{P(A \cap H)}{P(H)} = \frac{r(r-1)}{(r+v)(r+v-1)} \times \frac{r+v}{r} = \frac{r-1}{r+v-1},$$

ce qui est bien la valeur que nous avons attribuée *a priori* en analysant les conditions expérimentales.

Remarque 2.33. Il importe de bien comprendre que l'écriture « $A \mid H$ » ne désigne pas un nouvel événement¹⁹ différent de A . Quand on écrit $P(A \mid H)$, ce que l'on a modifié, ce n'est pas l'événement A , mais la valeur numérique qui lui était attribuée par la fonction d'ensembles P . Il serait donc en fait plus correct d'écrire $P_H(A)$ que $P(A \mid H)$. On conservera néanmoins cette dernière notation essentiellement pour des raisons typographiques : $P(A \mid H_1 \cap H_2 \cap H_3)$ est plus lisible que $P_{H_1 \cap H_2 \cap H_3}(A)$.

2.6.2 Propriétés

Proposition 2.34. Soit (Ω, \mathcal{F}, P) un espace probabilisé et H un événement fixé tel que $P(H) \neq 0$. Alors la fonction d'ensembles $P(\cdot \mid H)$ définie par :

$$P(\cdot \mid H) : \mathcal{F} \rightarrow [0, 1] \quad B \mapsto P(B \mid H)$$

est une nouvelle probabilité sur (Ω, \mathcal{F}) .

La preuve a déjà été donnée à l'occasion de l'exemple 2.12. Une conséquence immédiate est que la fonction d'ensembles $P(\cdot \mid H)$ vérifie toutes les propriétés de la proposition 2.16.

Corollaire 2.35. La fonction d'ensembles $P(\cdot \mid H)$ vérifie :

1. $P(\emptyset \mid H) = 0$, $P(\Omega \mid H) = 1$ et si $A \supset H$, $P(A \mid H) = 1$.
2. Si les A_i sont deux à deux disjoints :

$$P\left(\bigcup_{i=1}^n A_i \mid H\right) = \sum_{i=1}^n P(A_i \mid H).$$

3. Pour tout $B \in \mathcal{F}$, $P(B^c \mid H) = 1 - P(B \mid H)$.
4. Pour tous $A \in \mathcal{F}$ et $B \in \mathcal{F}$, si $A \subset B$, $P(A \mid H) \leq P(B \mid H)$.
5. Pour tous $A \in \mathcal{F}$ et $B \in \mathcal{F}$,

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

6. Pour toute suite $(A_i)_{i \geq 1}$ d'événements :

$$P\left(\bigcup_{i \in \mathbb{N}^*} A_i \mid H\right) \leq \sum_{i=1}^{+\infty} P(A_i \mid H).$$

7. Si $B_n \uparrow B$, $P(B_n \mid H) \uparrow P(B \mid H)$, $(n \rightarrow +\infty)$.
8. Si $C_n \downarrow C$, $P(C_n \mid H) \downarrow P(C \mid H)$, $(n \rightarrow +\infty)$.

19. En fait cette écriture prise isolément (sans le P) n'a pas de sens et ne devrait jamais être utilisée. Le symbole \mid ne représente pas une opération sur les événements qui l'entourent.

Nous n'avons vu jusqu'ici aucune formule permettant de calculer la probabilité d'une intersection d'événements à l'aide des probabilités de ces événements. Une telle formule *n'existe pas dans le cas général*. Les probabilités conditionnelles fournissent une méthode générale tout à fait naturelle pour calculer une probabilité d'intersection.

Proposition 2.36 (règle des conditionnements successifs).

Si A_1, \dots, A_n sont n événements tels que $P(A_1 \cap \dots \cap A_{n-1}) \neq 0$, on a :

$$\begin{aligned} P(A_1 \cap \dots \cap A_n) &= P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) \times \dots \\ &\quad \dots \times P(A_n | A_1 \cap \dots \cap A_{n-1}). \end{aligned}$$

Preuve. Pour $1 \leq i \leq n-1$, on a $\bigcap_{j=1}^{n-1} A_j \subset \bigcap_{j=1}^i A_j$ d'où :

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

Donc $P\left(\bigcap_{j=1}^i A_j\right)$ n'est nul pour aucun $i \leq n-1$ et on peut conditionner par l'événement $\bigcap_{j=1}^i A_j$. Ceci légitime le calcul suivant :

$$\begin{aligned} &P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) \times \dots \times P(A_n | A_1 \cap \dots \cap A_{n-1}) \\ &= P(A_1) \times \frac{P(A_1 \cap A_2)}{P(A_1)} \times \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \times \dots \times \frac{P(A_1 \cap \dots \cap A_n)}{P(A_1 \cap \dots \cap A_{n-1})} \\ &= P(A_1 \cap \dots \cap A_n), \end{aligned}$$

après simplifications en chaîne de toutes ces fractions. □

Les probabilités conditionnelles permettent aussi de calculer la probabilité d'un événement en conditionnant par *tous les cas possibles*. Du point de vue ensembliste, ces *cas possibles* correspondent à une *partition* de Ω .

Définition 2.37 (partition). On dit qu'une famille $(H_i)_{i \in I}$ de parties de Ω est une partition de Ω si elle vérifie les trois conditions :

- $\forall i \in I, H_i \neq \emptyset$.
- $\Omega = \bigcup_{i \in I} H_i$.
- Les H_i sont deux à deux disjoints ($i \neq j \Rightarrow H_i \cap H_j = \emptyset$).

Proposition 2.38 (conditionnement par les cas possibles²⁰).

(i) Si H est tel que $P(H) \neq 0$ et $P(H^c) \neq 0$, on a

$$\forall A \in \mathcal{F}, \quad P(A) = P(A | H)P(H) + P(A | H^c)P(H^c).$$

20. ou formule des probabilités totales.

(ii) Si H_1, \dots, H_n est une partition finie de Ω en événements de probabilité non nulle,

$$\forall A \in \mathcal{F}, \quad P(A) = \sum_{i=1}^n P(A \mid H_i)P(H_i).$$

(iii) Si $(H_i)_{i \in \mathbb{N}}$ est une partition de Ω telle que $\forall i \in \mathbb{N}, P(H_i) \neq 0$:

$$\forall A \in \mathcal{F}, \quad P(A) = \sum_{i=0}^{+\infty} P(A \mid H_i)P(H_i).$$

Preuve. Il suffit de vérifier (iii), les deux premières propriétés se démontrant de façon analogue. Comme $(H_i)_{i \in \mathbb{N}}$ est une partition de Ω ,

$$A = A \cap \Omega = A \cap \left(\bigcup_{i \in \mathbb{N}} H_i \right) = \bigcup_{i \in \mathbb{N}} (A \cap H_i)$$

et cette réunion est disjointe car les H_i étant deux à deux disjoints, il en est de même pour les $(A \cap H_i)$. Par conséquent par σ -additivité :

$$P(A) = \sum_{i=0}^{+\infty} P(A \cap H_i) = \sum_{i=0}^{+\infty} P(A \mid H_i)P(H_i).$$

□

Lorsqu'on a une partition de Ω en n hypothèses ou cas possibles H_i et que l'on sait calculer les $P(H_i)$ et les $P(A \mid H_i)$, on peut se poser le problème inverse : calculer $P(H_j \mid A)$ à l'aide des quantités précédentes. La solution est donnée par la formule suivante quelquefois appelée (abusivement) formule des probabilités des causes.

Proposition 2.39 (formule de Bayes).

Soit A un événement de probabilité non nulle. Si les événements H_i ($1 \leq i \leq n$) forment une partition de Ω et si aucun $P(H_i)$ n'est nul, on a pour tout $j = 1, \dots, n$:

$$P(H_j \mid A) = \frac{P(A \mid H_j)P(H_j)}{\sum_{i=1}^n P(A \mid H_i)P(H_i)}.$$

Preuve. Par définition des probabilités conditionnelles on a :

$$P(H_j \mid A) = \frac{P(A \cap H_j)}{P(A)} = \frac{P(A \mid H_j)P(H_j)}{P(A)}.$$

Et il ne reste plus qu'à développer $P(A)$ en conditionnant par la partition $(H_i, 1 \leq i \leq n)$ comme à la proposition 2.38. □

La même formule se généralise au cas d'une partition dénombrable. Ces formules sont plus faciles à retrouver qu'à mémoriser. . .

2.6.3 Quelques exemples

Exemple 2.40. On considère deux urnes U_1 et U_2 . L'urne U_1 contient r_1 boules rouges et v_1 boules vertes. L'urne U_2 contient r_2 boules rouges et v_2 boules vertes. On lance un dé. S'il indique le chiffre 1, on choisit l'urne U_1 , sinon on choisit U_2 . Dans chaque cas on effectue deux tirages avec remise dans l'urne choisie. Quelle est la probabilité d'obtenir une rouge au premier tirage ? deux rouges en tout ?

Adoptons les notations d'événements suivantes :

$$R = \{\text{rouge au 1}^{\text{er}} \text{ tirage}\}, \quad R' = \{\text{rouge au 2}^{\text{e}} \text{ tirage}\},$$

$$H_1 = \{\text{choix de l'urne } U_1\}, \quad H_2 = H_1^c = \{\text{choix de l'urne } U_2\}.$$

Il est facile de calculer directement $P(R | H_i)$ et $P(R \cap R' | H_i)$ pour $i = 1, 2$. En effet, une fois l'urne U_i choisie, on a un problème classique de tirages avec remise dans la même urne que l'on peut traiter (par exemple) par le dénombrement. On a ainsi :

$$P(R | H_i) = \frac{r_i}{r_i + v_i}, \quad P(R \cap R' | H_i) = \left(\frac{r_i}{r_i + v_i} \right)^2.$$

La formule de conditionnement par la partition $\{H_1, H_2\}$ donne :

$$\begin{aligned} P(R) &= P(R | H_1)P(H_1) + P(R | H_2)P(H_2) \\ &= \frac{1}{6} \frac{r_1}{r_1 + v_1} + \frac{5}{6} \frac{r_2}{r_2 + v_2} \end{aligned}$$

et

$$\begin{aligned} P(R \cap R') &= P(R \cap R' | H_1)P(H_1) + P(R \cap R' | H_2)P(H_2) \\ &= \frac{1}{6} \left(\frac{r_1}{r_1 + v_1} \right)^2 + \frac{5}{6} \left(\frac{r_2}{r_2 + v_2} \right)^2. \end{aligned}$$

Exemple 2.41. Un questionnaire à choix multiples propose m réponses pour chaque question. Soit p la probabilité qu'un étudiant connaisse la réponse à une question donnée. S'il ignore la réponse, il choisit au hasard l'une des réponses proposées. Quelle est pour le correcteur la probabilité qu'un étudiant connaisse vraiment la bonne réponse lorsqu'il l'a donnée ?

Notons :

$$\begin{aligned} B &= \{\text{l'étudiant donne la bonne réponse}\} \\ C &= \{\text{l'étudiant connaît la bonne réponse}\} \end{aligned}$$

On cherche $P(C | B)$. Avec ces notations, les données de l'énoncé peuvent se traduire par :

$$P(C) = p, \quad P(C^c) = 1 - p, \quad P(B | C) = 1, \quad P(B | C^c) = \frac{1}{m}.$$

On en déduit :

$$\begin{aligned}
 P(C | B) &= \frac{P(B \cap C)}{P(B)} \\
 &= \frac{P(B | C)P(C)}{P(B | C)P(C) + P(B | C^c)P(C^c)} \\
 &= \frac{1 \times p}{1 \times p + \frac{1}{m}(1-p)} = \frac{mp}{1 + (m-1)p}.
 \end{aligned}$$

Pour p fixé, $P(C | B)$ est une fonction croissante de m , les deux bornes étant $P(C | B) = p$ (cas $m = 1$) et $P(C | B) \rightarrow 1$ ($m \rightarrow +\infty$). D'autre part pour m fixé, $P(C | B)$ est une fonction croissante de p . On a pour $p > 0$:

$$\frac{P(C | B)}{p} = \frac{m}{1 + (m-1)p} \geq 1,$$

l'égalité n'étant possible que pour $p = 1$. Tout ceci est conforme à l'intuition.

Exemple 2.42. Un test sanguin a une probabilité de 0,95 de détecter un certain virus lorsque celui-ci est effectivement présent. Il donne néanmoins un faux résultat positif pour 1% des personnes non infectées. Si 0,5% de la population est porteuse du virus, quelle est la probabilité qu'une personne ait le virus sachant qu'elle a un test positif? Notons

$$\begin{aligned}
 V &= \{\text{la personne testée a le virus}\}, \\
 T &= \{\text{la personne testée a un test positif}\}.
 \end{aligned}$$

On cherche $P(V | T)$. Or on sait que $P(V) = 0,005$, $P(T | V) = 0,95$ et $P(T | V^c) = 0,01$. On en déduit :

$$\begin{aligned}
 P(V | T) &= \frac{P(T \cap V)}{P(T)} = \frac{P(T | V)P(V)}{P(T | V)P(V) + P(T | V^c)P(V^c)} \\
 &= \frac{0,95 \times 0,005}{0,95 \times 0,005 + 0,01 \times 0,995} \simeq 0,323.
 \end{aligned}$$

On voit ainsi que contrairement à ce que l'on aurait pu croire, le test n'est pas fiable : si la personne présente un test positif, la probabilité qu'elle ne soit pas porteuse du virus est deux fois plus élevée que celle qu'elle le soit !

Exemple 2.43. Revenons sur le problème de dés étudié à la sous-section 2.2.3, i.e. attribution d'une probabilité à l'évènement E première obtention de la somme 9 avant celle de la somme 7. Les probabilités conditionnelles permettent d'en proposer une solution simple, sans calcul de série. En rappelant que F_1 , G_1 , H_1 désignent respectivement l'évènement *obtention au premier lancer d'une somme 9 (resp. 7, resp. ni 7 ni 9)*, la formule des probabilités totales nous donne :

$$P(E) = P(E | F_1)P(F_1) + P(E | G_1)P(G_1) + P(E | H_1)P(H_1). \quad (2.28)$$

On a clairement $P(E | F_1) = 1$ et $P(E | G_1) = 0$. Pour attribuer une valeur à $P(E | H_1)$, on considère que l'obtention d'une somme autre que 7 ou 9 au premier lancer ne devrait pas influencer sur l'apparition ultérieure du 7 ou du 9. Pour traduire cette idée, on pose $P(E | H_1) = P(E)$. En reportant ces valeurs dans (2.28), il vient

$$P(E) = P(F_1) + P(E)P(H_1) = \frac{1}{9} + \frac{13}{18}P(E), \quad \text{d'où} \quad \left(1 - \frac{13}{18}\right)P(E) = \frac{1}{9},$$

ce qui se résout en

$$P(E) = \frac{1}{9} \times \frac{18}{5} = \frac{2}{5}.$$

On retrouve ainsi la valeur obtenue à la sous-section 2.2.3, lorsque nous avons esquissé la construction d'un espace (Ω, \mathcal{F}, P) modélisant cette expérience aléatoire. Le point clé dans le raisonnement ci-dessus est l'égalité $P(E | H_1) = P(E)$, qui exprime *l'indépendance* des événements H_1 et E , une notion que nous allons étudier maintenant.

2.7 Indépendance

2.7.1 Indépendance de deux événements

Soient A et B deux événements de probabilité non nulle. Il arrive que la connaissance de la réalisation de A ne modifie pas notre information sur celle de B , autrement dit que $P(B | A) = P(B)$. C'est le cas par exemple lorsque l'on fait un tirage avec remise et que la réalisation de A ne dépend que du résultat du premier tirage, celle de B que du deuxième. Symétriquement on aura dans cet exemple $P(A | B) = P(A)$. Cette remarque se généralise :

Proposition 2.44. *Si A et B sont des événements de probabilité non nulle, les trois égalités suivantes sont équivalentes :*

- (i) $P(B | A) = P(B)$,
- (ii) $P(A | B) = P(A)$,
- (iii) $P(A \cap B) = P(A)P(B)$.

Preuve. Comme $P(A) \neq 0$ et $P(B) \neq 0$, on a la chaîne d'équivalences :

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

□

D'autre part la relation (iii) est toujours vérifiée dans le cas dégénéré où $P(A) = 0$ ou $P(B) = 0$. En effet, on a alors à la fois $P(A)P(B) = 0$ et $0 \leq P(A \cap B) \leq \min(P(A), P(B)) = 0$ d'où $P(A \cap B) = 0$. Ainsi la relation (iii) est un peu plus générale que (i) et (ii). Elle a aussi sur les deux autres l'avantage de la symétrie d'écriture. C'est elle que l'on retient pour définir l'indépendance.

Définition 2.45. Soit (Ω, \mathcal{F}, P) un espace probabilisé. Deux événements A et B de cet espace sont dits indépendants lorsque :

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

Exemple 2.46. On jette deux fois le même dé. Les événements

$$\begin{aligned} A &= \{\text{obtention d'un chiffre pair au premier lancer}\}, \\ B &= \{\text{obtention du 1 au deuxième lancer}\}, \end{aligned}$$

sont indépendants.

En effet, en prenant $\Omega = \{1, 2, \dots, 6\}^2$, $\mathcal{F} = \mathcal{P}(\Omega)$ et P l'équiprobabilité, on vérifie que :

$$\begin{aligned} P(A) &= \frac{3 \times 6}{36} = \frac{1}{2}, & P(B) &= \frac{6 \times 1}{36} = \frac{1}{6}, \\ P(A \cap B) &= \frac{3 \times 1}{36} = \frac{1}{12}, & P(A)P(B) &= \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}. \end{aligned}$$

Remarques 2.47.

- Si A est un événement tel que $P(A) = 0$ ou $P(A) = 1$, alors il est indépendant de tout événement, *y compris de lui-même* (c'est le cas en particulier pour Ω et \emptyset).
- Deux événements *incompatibles* A et B avec $P(A) > 0$ et $P(B) > 0$ ne sont *jamais indépendants*. En effet $A \cap B = \emptyset$ implique $P(A \cap B) = 0$ or $P(A)P(B) \neq 0$.
- L'indépendance de deux événements A et B n'est pas une propriété intrinsèque aux événements, elle est toujours relative au modèle (Ω, \mathcal{F}, P) que l'on a choisi. Voici un exemple pour l'illustrer.

Exemple 2.48. Une urne contient 12 boules numérotées de 1 à 12. On en tire une au hasard et on considère les événements :

$$A = \{\text{tirage d'un nombre pair}\}, \quad B = \{\text{tirage d'un multiple de 3}\}.$$

L'espace probabilisé qui s'impose naturellement ici est $\Omega = \{1, \dots, 12\}$ muni de l'équiprobabilité P . Les événements A et B s'écrivent :

$$A = \{2, 4, 6, 8, 10, 12\}, \quad B = \{3, 6, 9, 12\}, \quad A \cap B = \{6, 12\}.$$

Dans ce modèle, on a

$$\begin{aligned} P(A) &= \frac{6}{12} = \frac{1}{2}, & P(B) &= \frac{4}{12} = \frac{1}{3}, \\ P(A \cap B) &= \frac{2}{12} = \frac{1}{6}, & P(A)P(B) &= \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}, \end{aligned}$$

donc A et B sont indépendants.

On rajoute maintenant dans l'urne une boule numérotée treize et on recommence l'expérience. Les événements A et B restent les mêmes, mais le modèle a changé. On a maintenant l'équiprobabilité P' sur $\Omega' = \{1, \dots, 13\}$ et

$$P'(A) = \frac{6}{13}, \quad P'(B) = \frac{4}{13}, \quad P'(A \cap B) = \frac{2}{13},$$

mais

$$P'(A)P'(B) = \frac{6}{13} \times \frac{4}{13} = \frac{24}{169} \neq \frac{2}{13},$$

donc A et B ne sont plus indépendants. Un peu de réflexion permet de relier ces résultats calculatoires avec la notion intuitive d'indépendance présentée en introduction. Dans le premier cas, la proportion des multiples de trois parmi les pairs est la même que parmi les impairs. Le fait de savoir que la boule tirée est paire ne modifie en rien notre information sur B . Par contre dans le deuxième cas, l'ajout de la treizième boule modifie la proportion des multiples de trois : elle est plus élevée chez les pairs que chez les impairs. Donc le fait de savoir que la boule tirée est paire augmente un peu la probabilité que nous pouvons attribuer à B .

Proposition 2.49. *Si A et B sont indépendants, il en est de même pour les paires d'événements A et B^c , A^c et B , A^c et B^c .*

Preuve. Par hypothèse, $P(A \cap B) = P(A)P(B)$. En considérant la réunion disjointe $A = (A \cap B) \cup (A \cap B^c)$, nous avons : $P(A \cap B^c) = P(A) - P(A \cap B)$, d'où :

$$P(A \cap B^c) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c).$$

Donc A et B^c sont indépendants. L'échange des rôles de A et B dans ce raisonnement donne l'indépendance de A^c et B . En réutilisant le premier résultat avec A^c à la place de A , on obtient alors celle de A^c et B^c . \square

2.7.2 Indépendance mutuelle

On se propose de généraliser la notion d'indépendance à plus de deux événements. Examinons d'abord la situation suivante.

Exemple 2.50. Une urne contient quatre jetons : un bleu, un blanc, un rouge et un bleu-blanc-rouge. On en tire un au hasard. Considérons les trois événements

$$\begin{aligned} A &= \{\text{le jeton tiré contient du bleu}\}, \\ B &= \{\text{le jeton tiré contient du blanc}\}, \\ C &= \{\text{le jeton tiré contient du rouge}\}. \end{aligned}$$

Il est clair que $P(A) = P(B) = P(C) = 2/4 = 1/2$. D'autre part :

$$P(A \cap B) = P(\text{tricolore}) = \frac{1}{4} = P(A)P(B)$$

et de même $P(B \cap C) = 1/4 = P(B)P(C)$, $P(C \cap A) = 1/4 = P(C)P(A)$. Ainsi les événements A , B , C sont *deux à deux indépendants*.

D'autre part $P(A \mid B \cap C) = 1$ car $B \cap C = \{\text{tricolore}\}$. Donc la connaissance de la réalisation *simultanée* de B et C modifie notre information sur A . La notion d'indépendance deux à deux n'est donc pas suffisante pour traduire l'idée intuitive d'indépendance de plusieurs événements. Ceci motive la définition suivante.

Définition 2.51. *Trois événements A , B , C sont dits mutuellement indépendants lorsqu'ils vérifient les quatre conditions :*

$$\begin{aligned} P(A \cap B) &= P(A)P(B), \\ P(B \cap C) &= P(B)P(C), \\ P(C \cap A) &= P(C)P(A), \\ P(A \cap B \cap C) &= P(A)P(B)P(C). \end{aligned}$$

Avec cette définition de l'indépendance des événements A , B et C on a bien ²¹ $P(A \mid B) = P(A)$, $P(A \mid B \cap C) = P(A)$, ainsi que toutes les égalités qui s'en déduisent par permutation sur les lettres A , B , C . On peut généraliser cette définition comme suit.

Définition 2.52. *Les n événements A_1, \dots, A_n sont dits mutuellement indépendants si pour toute sous-famille A_{i_1}, \dots, A_{i_k} avec $1 \leq i_1 < \dots < i_k \leq n$, on a :*

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \times \dots \times P(A_{i_k}). \quad (2.29)$$

L'indépendance mutuelle implique évidemment l'indépendance deux à deux et la réciproque est fausse comme le montre l'exemple 2.50. Dans toute la suite, lorsque nous parlerons d'une famille de plusieurs événements indépendants sans autre précision, nous sous-entendrons systématiquement *mutuellement* indépendants.

Proposition 2.53. *Si $\{A_1, \dots, A_n\}$ est une famille de n événements indépendants, toute famille obtenue en remplaçant certains des A_i par leur complémentaire est encore indépendante.*

Preuve. Supposons la proposition démontrée dans le cas où l'on a remplacé *un seul* A_i par son complémentaire. Le cas général s'en déduit en utilisant cette propriété autant de fois qu'il y a de A_i changés en leur complémentaire. Dans le cas d'un seul A_i remplacé, on ne perd pas de généralité en supposant qu'il s'agit de A_1 (il suffit de changer l'indexation des événements, ce qui n'affecte pas leur indépendance mutuelle). Il nous reste alors à vérifier (2.29) dans le cas où $i_1 = 1$ avec $A_{i_1}^c$ à la place de A_{i_1} (dans le cas $i_1 > 1$, l'égalité ne fait intervenir que des éléments de la famille initiale et il n'y a donc rien à vérifier). Posons $B = A_{i_2} \cap \dots \cap A_{i_k}$. L'hypothèse (2.29) appliquée à la famille A_{i_2}, \dots, A_{i_k} nous donne $P(B) = P(A_{i_2}) \times \dots \times P(A_{i_k})$. La même hypothèse appliquée à A_{i_1}, \dots, A_{i_k} nous donne alors l'indépendance de A_{i_1} et de B . Par la proposition 2.49, on en déduit :

$$P(A_{i_1}^c \cap B) = P(A_{i_1}^c) \times P(B) = P(A_{i_1}^c) \times P(A_{i_2}) \times \dots \times P(A_{i_k}),$$

ce qui achève la preuve. □

21. Lorsque les probabilités conditionnelles existent.

Définition 2.54 (indépendance d'une suite d'événements). *Une suite infinie d'événements est dite indépendante si toute sous-suite finie est formée d'événements mutuellement indépendants.*

Remarque 2.55. Compte-tenu de la proposition 2.53, on voit immédiatement que si $(A_i)_{i \in \mathbb{N}}$ est une suite indépendante d'événements, toute suite formée en remplaçant certains des A_i (éventuellement tous) par leur complémentaire est encore indépendante.

2.7.3 Épreuves répétées

Considérons une suite d'épreuves réalisées dans les mêmes conditions expérimentales. Par exemple tirages avec remise dans la même urne, lancers successifs d'un dé, ... Il est alors raisonnable de supposer que les résultats de tout sous-ensemble fini d'épreuves n'ont aucune influence sur ceux des autres épreuves.

Définition 2.56. *On dit que les épreuves sont indépendantes si toute suite $(A_i)_{i \geq 1}$ telle que la réalisation de chaque A_i est déterminée uniquement par le résultat de la i^e épreuve est une suite indépendante d'événements.*

Exemple 2.57. On réalise une suite d'épreuves indépendantes. Chaque épreuve résulte en un succès avec probabilité $p \in]0, 1[$ ou en un échec avec probabilité $q = 1 - p$. Quelle est la probabilité des événements suivants :

- a) $A = \{\text{Au moins un succès au cours des } n \text{ premières épreuves}\},$
- b) $B = \{\text{Exactement } k \text{ succès au cours des } n \text{ premières épreuves}\},$
- c) $C = \{\text{Toutes les épreuves donnent un succès}\}?$

Notons pour tout $i \geq 1$: $R_i = \{\text{succès à la } i^e \text{ épreuve}\}$, R_i^c est alors l'échec à la i^e épreuve.

- a) $A = \bigcup_{i=1}^n R_i$, d'où $A^c = \bigcap_{i=1}^n R_i^c$. Les R_i^c étant indépendants, on a :

$$P(A^c) = \prod_{i=1}^n P(R_i^c) = (1 - p)^n = q^n.$$

On en déduit $P(A) = 1 - q^n$.

- b) Traitons d'abord le cas $0 < k < n$. L'événement B est la réunion disjointe de tous les événements du type :

$$B_I = \left(\bigcap_{i \in I} R_i \right) \cap \left(\bigcap_{j \in J} R_j^c \right),$$

où I est une partie de cardinal k de $\{1, \dots, n\}$ et J son complémentaire dans $\{1, \dots, n\}$. L'ensemble d'indices I représente un choix possible des k épreuves donnant un succès, les autres épreuves indexées par J donnant alors un échec. En considérant tous les choix possibles de l'ensemble I (il y en a C_n^k), on obtient une partition de B par les B_I . Par indépendance des épreuves, pour tout I on a :

$$P(B_I) = \prod_{i \in I} P(R_i) \times \prod_{j \in J} P(R_j^c) = p^k q^{n-k}.$$

On voit ainsi que $P(B_I)$ ne dépend pas de I . On en déduit :

$$P(B) = \sum_{\substack{I \subset \{1, \dots, n\} \\ \text{card } I = k}} P(B_I) = C_n^k p^k q^{n-k}.$$

La vérification de la validité de la formule $P(B) = C_n^k p^k q^{n-k}$ dans les cas $k = 0$ et $k = n$ est laissée au lecteur.

c) Pour $n \geq 1$, soit $C_n = \{\text{succès aux } n \text{ premières épreuves}\}$. Clairement C est inclus dans C_n donc $0 \leq P(C) \leq P(C_n)$. En utilisant l'indépendance des R_i on obtient :

$$P(C_n) = P\left(\bigcap_{i=1}^n R_i\right) = \prod_{i=1}^n P(R_i) = p^n.$$

donc pour tout $n \geq 1$, $0 \leq P(C) \leq p^n$. En faisant tendre n vers $+\infty$, on en déduit $P(C) = 0$.

Chapitre 3

Variables aléatoires

3.1 Introduction

Dans de nombreux jeux, on fait intervenir le hasard en observant la somme des points marqués par deux dés. Considérons le jet d'un dé bleu et d'un dé rouge et notons S la somme des points obtenus. On modélise cette expérience en prenant l'équiprobabilité sur :

$$\Omega = \{1, 2, 3, 4, 5, 6\}^2.$$

Un événement élémentaire ω est ici un couple (b, r) où b désigne le résultat du dé bleu et r celui du rouge et $S(\omega) = b + r$. Il est commode de décrire la situation par un tableau à 36 cases en écrivant la valeur de $S(\omega)$ dans la case représentant $\omega = (b, r)$ à l'intersection de la ligne b et de la colonne r .

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

FIG. 3.1 – Somme des points de deux dés

On a ainsi défini une application S de Ω dans l'ensemble des sommes de points possibles : $\{2, 3, \dots, 11, 12\}$. On dit que S est une *variable aléatoire* sur Ω . En fait, l'observation qui nous intéresse dans cette expérience, ce n'est pas ω , mais seulement $S(\omega)$. Ce que l'on aimerait connaître, c'est la probabilité que la somme des points prenne une valeur donnée, soit $P(S = k)$ pour k entier fixé entre 2 et 12. Ici la notation

« $P(S = k)$ » est un abus d'écriture commode pour désigner $P(\{\omega \in \Omega; S(\omega) = k\})$. En utilisant l'équiprobabilité sur Ω et la figure 3.1, nous obtenons le tableau 3.1.

k	2	3	4	5	6	7	8	9	10	11	12
$P(S = k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

TAB. 3.1 – Probabilités des valeurs de la somme des points

Cela revient à considérer un nouvel ensemble d'événements élémentaires :

$$\Omega' = S(\Omega) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

et à munir cet ensemble de la probabilité P_S définie par le tableau 3.1, plus précisément :

$$P_S := \sum_{k \in \Omega'} P(S = k) \delta_k. \quad (3.1)$$

Cette nouvelle probabilité s'appelle *loi de la variable aléatoire S* . En d'autres termes, nous avons réalisé *via S* un *transfert* de l'espace probabilisé $(\Omega, \mathcal{P}(\Omega), P)$ sur l'espace probabilisé $(\Omega', \mathcal{P}(\Omega'), P_S)$. Si $B \in \mathcal{P}(\Omega')$, en notant $\{S \in B\} := \{\omega \in \Omega; S(\omega) \in B\}$, on a ainsi puisque B est l'union finie de ses singletons,

$$P(S \in B) = \sum_{k \in B} P(S = k) = \sum_{k \in \Omega'} P(S = k) \delta_k(B) = P_S(B). \quad (3.2)$$

Remarquons maintenant que l'on peut facilement agrandir l'ensemble d'arrivée de S en remplaçant Ω' par $\Omega'' = \mathbb{R}$. On peut munir \mathbb{R} au choix de la tribu $\mathcal{P}(\mathbb{R})$ ou $\text{Bor}(\mathbb{R})$. En définissant encore P_S par (3.1), le calcul (3.2) s'adapte facilement pour tout $B \in \mathcal{P}(\mathbb{R})$, donc *a fortiori* pour tout B borélien :

$$P(S \in B) = P(S \in B \cap \Omega') = \sum_{k \in B \cap \Omega'} P(S = k) = \sum_{k \in \Omega'} P(S = k) \delta_k(B) = P_S(B). \quad (3.3)$$

Il y a de bonnes raisons à cet agrandissement de l'ensemble d'arrivée, ne serait-ce que pour pouvoir faire des opérations sur les variables aléatoires. Imaginons par exemple que l'on jette n fois la paire de dés et que l'on s'intéresse à la moyenne arithmétique M_n des sommes obtenues. Il ne serait guère commode de travailler avec un espace d'arrivée $\Omega'_n := M_n(\Omega'^n)$, surtout si l'on s'intéresse au comportement de M_n pour n tendant vers l'infini. Nous prendrons donc désormais comme ensemble d'arrivée \mathbb{R} pour les variables aléatoires que nous étudierons.

Si $X(\omega)$ est une grandeur physique (masse, température, pression, longueur, etc.) mesurée à partir du résultat ω d'une expérience, il n'est en général pas possible de la déterminer avec une précision absolue. Tout ce que l'on peut dire est que X appartient à un certain intervalle, dont la longueur dépend de la précision de l'instrument de mesure utilisé. Les quantités pertinentes pour identifier la *loi* de X sont alors les $P(X \in I)$

pour I intervalle¹ de \mathbb{R} , plutôt que les $P(X = x)$ qui pourraient être nulles pour tout x réel. Pour définir la loi de X , nous serons amenés à poser $P_X(B) = P(X \in B)$ pour B borélien quelconque de \mathbb{R} , sous réserve que cela ait un sens. Voyons cela de plus près.

Soit X une application $\Omega \rightarrow \mathbb{R}$. Pour tout $B \subset \mathbb{R}$, notons

$$\{X \in B\} := \{\omega \in \Omega; X(\omega) \in B\} =: X^{-1}(B).$$

Cette écriture « X^{-1} » ne suppose aucunement la bijectivité de X . Il s'agit seulement d'une notation commode pour « l'ensemble des antécédents des éléments de B par l'application X ». On aimerait pouvoir transporter par X la probabilité P , mesure définie sur (Ω, \mathcal{F}) en une probabilité P_X , définie sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ en posant :

$$\forall B \in \text{Bor}(\mathbb{R}), \quad P_X(B) := P(X^{-1}(B)) = P(X \in B).$$

Pour que cette écriture ait un sens, encore faut-il que $X^{-1}(B)$ soit un élément de la tribu \mathcal{F} sur laquelle est définie P . Nous supposons donc que X vérifie la condition suivante :

$$\forall B \in \text{Bor}(\mathbb{R}), \quad X^{-1}(B) \in \mathcal{F}. \quad (3.4)$$

On dit alors que X est *mesurable* pour les tribus \mathcal{F} et $\text{Bor}(\mathbb{R})$. Nous réserverons le nom de *variable aléatoire* aux applications $X : \Omega \rightarrow \mathbb{R}$ vérifiant (3.4). La mesurabilité est d'ailleurs une notion plus générale définie comme suit.

Définition 3.1 (mesurabilité). *Soit $h : \Omega_1 \rightarrow \Omega_2$, où Ω_1 et Ω_2 sont munis respectivement des tribus \mathcal{F}_1 et \mathcal{F}_2 . On dit que h est mesurable² $\mathcal{F}_1 - \mathcal{F}_2$ si pour tout $B \in \mathcal{F}_2$, $h^{-1}(B) \in \mathcal{F}_1$.*

On démontre en théorie de la mesure le résultat suivant.

Proposition 3.2. *Soient Ω_1 et Ω_2 deux ensembles munis respectivement des tribus \mathcal{F}_1 et \mathcal{F}_2 et \mathcal{C} une famille de parties de Ω_2 engendrant \mathcal{F}_2 ($\sigma(\mathcal{C}) = \mathcal{F}_2$). L'application $h : \Omega_1 \rightarrow \Omega_2$ est mesurable $\mathcal{F}_1 - \mathcal{F}_2$ si pour tout $B \in \mathcal{C}$, $h^{-1}(B) \in \mathcal{F}_1$.*

En particulier en prenant $\Omega_2 = \mathbb{R}$, $\mathcal{F}_2 = \text{Bor}(\mathbb{R})$ et \mathcal{C} la famille des intervalles $]a, b]$, on voit que pour que $X : \Omega \rightarrow \mathbb{R}$ soit une variable aléatoire, il suffit que

$$\forall a, b \in \mathbb{R}, \quad X^{-1}(]a, b]) \in \mathcal{F}.$$

La mesurabilité $\mathcal{F} - \text{Bor}(\mathbb{R})$ est conservée par toutes les opérations usuelles de l'algèbre (somme, produit, combinaisons linéaires, composition, ...) et de l'analyse³, pourvu que les familles d'applications concernées soient au plus dénombrables (inf, sup, lim inf, lim sup, limite si elle existe d'une suite de fonctions, série de fonctions, ...). Bref, il s'agit d'une notion très riche. Tellement riche en fait, que nous ne rencontrerons jamais dans ce cours d'application $X : \Omega \rightarrow \mathbb{R}$ qui ne soit pas une variable aléatoire. Ceci explique que dans les ouvrages de probabilités élémentaires, on appelle variable aléatoire n'importe quelle application $\Omega \rightarrow \mathbb{R}$.

1. Ceci est en accord avec la remarque faite p. 72 à propos de la caractérisation des probabilités sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$.

2. Ce langage est trompeur : la mesurabilité ne fait intervenir aucune mesure, elle concerne seulement h et les tribus.

3. Pour les énoncés précis, voir le cours d'IFP 2003-2004 chapitre 2
<http://math.univ-lille1.fr/~suquet/ens/IFP/Cours/cours04/CoursIFP04.html>

3.2 Généralités

3.2.1 Variables aléatoires réelles

Avant de formaliser la définition d'une variable aléatoire, commençons par donner deux propriétés utiles des inverses ensemblistes.

Proposition 3.3. *Soit $h : \Omega_1 \rightarrow \Omega_2$ une application. Pour tout $B \subset \Omega_2$, notons $h^{-1}(B) := \{\omega \in \Omega_1; h(\omega) \in B\}$. L'inverse ensembliste h^{-1} ainsi défini commute avec les unions et les intersections quelconques. Autrement dit, si $(B_i)_{i \in I}$ est une famille quelconque de parties de Ω_2 ,*

$$h^{-1}\left(\bigcup_{i \in I} B_i\right) = \bigcup_{i \in I} h^{-1}(B_i), \quad (3.5)$$

$$h^{-1}\left(\bigcap_{i \in I} B_i\right) = \bigcap_{i \in I} h^{-1}(B_i). \quad (3.6)$$

L'inverse ensembliste commute aussi avec le passage au complémentaire au sens suivant :

$$\forall B \subset \Omega_2, \quad h^{-1}(\Omega_2 \setminus B) = \Omega_1 \setminus h^{-1}(B). \quad (3.7)$$

Preuve. L'égalité d'ensembles (3.5) se vérifie par la chaîne d'équivalences logiques suivantes qui montre que l'appartenance au premier membre de (3.5) équivaut à l'appartenance à son deuxième membre :

$$\begin{aligned} \omega \in h^{-1}\left(\bigcup_{i \in I} B_i\right) &\Leftrightarrow h(\omega) \in \bigcup_{i \in I} B_i \\ &\Leftrightarrow \exists i \in I; h(\omega) \in B_i \\ &\Leftrightarrow \exists i \in I; \omega \in h^{-1}(B_i) \\ &\Leftrightarrow \omega \in \bigcup_{i \in I} h^{-1}(B_i). \end{aligned}$$

On procède de même pour vérifier (3.6) :

$$\begin{aligned} \omega \in h^{-1}\left(\bigcap_{i \in I} B_i\right) &\Leftrightarrow h(\omega) \in \bigcap_{i \in I} B_i \\ &\Leftrightarrow \forall i \in I, h(\omega) \in B_i \\ &\Leftrightarrow \forall i \in I; \omega \in h^{-1}(B_i) \\ &\Leftrightarrow \omega \in \bigcap_{i \in I} h^{-1}(B_i). \end{aligned}$$

Voici la vérification de (3.7) :

$$\begin{aligned} h^{-1}(\Omega_2 \setminus B) &= \{\omega \in \Omega_1; h(\omega) \in (\Omega_2 \setminus B)\} \\ &= \{\omega \in \Omega_1; h(\omega) \notin B\} \\ &= \{\omega \in \Omega_1; \omega \notin h^{-1}(B)\} \\ &= \Omega_1 \setminus h^{-1}(B). \end{aligned}$$

□

Définition 3.4 (variable aléatoire réelle). Soit (Ω, \mathcal{F}, P) un espace probabilisé. On appelle variable aléatoire réelle sur (Ω, \mathcal{F}) , ou plus simplement variable aléatoire, toute application X :

$$X : \Omega \longrightarrow \mathbb{R} \quad \omega \mapsto X(\omega),$$

mesurable \mathcal{F} - $\text{Bor}(\mathbb{R})$, i.e. vérifiant :

$$\forall B \in \text{Bor}(\mathbb{R}), \quad X^{-1}(B) \in \mathcal{F}.$$

En raison de la proposition 3.2, il suffit que la condition ci-dessus soit vérifiée pour $B =]a, b]$ avec a et b réels quelconques, pour que X soit une variable aléatoire.

Remarque 3.5. Il importe de noter que la mesure de probabilité P sur (Ω, \mathcal{F}) ne joue aucun rôle dans la définition de la notion de variable aléatoire. C'est pour cela que nous parlons de « variable aléatoire sur (Ω, \mathcal{F}) » plutôt que sur (Ω, \mathcal{F}, P) .

Définition 3.6 (variable aléatoire discrète). On appelle variable aléatoire discrète sur (Ω, \mathcal{F}) , toute application $X : \Omega \rightarrow \mathbb{R}$ vérifiant les deux conditions suivantes.

- (i) L'ensemble des images $X(\Omega) = \{X(\omega), \omega \in \Omega\}$ est une partie au plus dénombrable de \mathbb{R} . On peut donc numéroter ses éléments par des indices entiers⁴

$$X(\Omega) = \{x_0, x_1, \dots, x_k, \dots\}.$$

- (ii) X est mesurable \mathcal{F} - $\mathcal{P}(\mathbb{R})$, ce qui équivaut ici à

$$\forall x \in X(\Omega), \quad X^{-1}(\{x\}) \in \mathcal{F}. \quad (3.8)$$

Remarquons que si X est mesurable \mathcal{F} - $\mathcal{P}(\mathbb{R})$, elle est *a fortiori* mesurable \mathcal{F} - $\text{Bor}(\mathbb{R})$ puisque la condition $X^{-1}(B) \in \mathcal{F}$ doit être satisfaite dans ce cas seulement pour les B appartenant à la tribu $\text{Bor}(\mathbb{R})$ qui est une sous-famille de $\mathcal{P}(\mathbb{R})$. Par conséquent, toute variable aléatoire discrète est aussi une variable aléatoire réelle.

L'équivalence pour $X(\Omega)$ au plus dénombrable, entre la mesurabilité \mathcal{F} - $\mathcal{P}(\mathbb{R})$ et la condition (3.8) se justifie comme suit. D'abord il est clair que la mesurabilité \mathcal{F} - $\mathcal{P}(\mathbb{R})$ implique (3.8). Pour la réciproque, on suppose (3.8) vérifiée, on prend $B \in \mathcal{P}(\mathbb{R})$ quelconque et on montre qu'alors $X^{-1}(B) \in \mathcal{F}$. Notons $B' := B \cap X(\Omega)$ et $B'' := B \cap (\mathbb{R} \setminus X(\Omega))$. Remarquons que B' étant inclus dans $X(\Omega)$ est au plus dénombrable et que $X^{-1}(B'') = \emptyset$. Il suffit alors d'écrire en utilisant (3.5)

$$\begin{aligned} X^{-1}(B) = X^{-1}(B' \cup B'') &= X^{-1}(B') \cup X^{-1}(B'') \\ &= X^{-1}(B') \\ &= X^{-1}\left(\bigcup_{x \in B'} \{x\}\right) \\ &= \bigcup_{x \in B'} X^{-1}(\{x\}), \end{aligned} \quad (3.9)$$

pour faire apparaître $X^{-1}(B)$ comme union au plus dénombrable d'éléments de la tribu \mathcal{F} , ce qui implique son appartenance à \mathcal{F} .

4. Pour tous les exemples classiques que nous rencontrerons, il est possible de les numéroter de manière croissante : $x_0 < x_1 < x_2 \dots$. Mais ce n'est pas toujours le cas, car l'ensemble des valeurs possibles peut être par exemple, les décimaux (ou les rationnels) de $[0, 1]$.

3.2.2 Loi d'une variable aléatoire

Une variable aléatoire X permet de transporter la probabilité P définie sur (Ω, \mathcal{F}) en une probabilité P_X définie sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$.

Proposition 3.7. *Soit X une variable aléatoire sur (Ω, \mathcal{F}) et P une probabilité sur (Ω, \mathcal{F}) . La fonction d'ensembles $P_X = P \circ X^{-1}$ définie sur $\text{Bor}(\mathbb{R})$ par*

$$\forall B \in \text{Bor}(\mathbb{R}), \quad P_X(B) := P(X^{-1}(B)) = P(X \in B) \quad (3.10)$$

est une probabilité sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$.

Preuve. D'abord P_X est bien définie comme application $\text{Bor}(\mathbb{R}) \rightarrow [0, 1]$ en raison de la mesurabilité \mathcal{F} - $\text{Bor}(\mathbb{R})$ de la variable aléatoire X . Il est clair que $P_X(\mathbb{R}) = P(\Omega) = 1$. Montrons la σ -additivité de P_X . Soit $(B_k)_{k \geq 1}$ une suite quelconque de boréliens de \mathbb{R} deux à deux disjoints. Par (3.6), les $X^{-1}(B_k)$ sont des éléments deux à deux disjoints de la tribu d'événements \mathcal{F} . En combinant (3.5) avec la σ -additivité de la probabilité P , on en déduit :

$$\begin{aligned} P_X\left(\bigcup_{k \geq 1} B_k\right) &= P\left(X^{-1}\left(\bigcup_{k \geq 1} B_k\right)\right) = P\left(\bigcup_{k \geq 1} X^{-1}(B_k)\right) = \sum_{k \geq 1} P(X^{-1}(B_k)) \\ &= \sum_{k \geq 1} P_X(B_k). \end{aligned}$$

La fonction d'ensembles P_X est donc σ -additive, c'est une probabilité sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$. \square

Définition 3.8 (loi d'une variable aléatoire). *Soient (Ω, \mathcal{F}, P) un espace probabilisé et X une variable aléatoire sur (Ω, \mathcal{F}) . On appelle loi de X sous P , ou plus simplement loi de X , la probabilité P_X sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ définie par (3.10).*

Si μ est une mesure de probabilité sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$, on dit que X « suit la loi μ » si $P_X = P \circ X^{-1} = \mu$ (i.e. la loi de X sous P est la mesure μ).

Remarque 3.9. Dans les problèmes usuels de probabilités, on travaille souvent avec un seul (Ω, \mathcal{F}, P) et on se contente alors de l'appellation *loi de X* . Il n'en va pas de même en statistique où l'on met généralement en concurrence plusieurs modèles $(\Omega, \mathcal{F}, P_\theta)$, où θ est un paramètre inconnu et où on se propose de choisir un de ces modèles au vu des valeurs $X(\omega)$ observées. C'est là que l'appellation loi de X sous P_θ s'impose. Pour donner un exemple simple, considérons le problème du sondage d'un échantillon de 500 personnes avant le second tour d'une élection présidentielle opposant le candidat A au candidat B . Ici θ est la proportion *inconnue* d'électeurs votant A dans la population totale. Si X est le nombre de personnes interrogées favorables à A , la loi de X sous P_θ est la loi binomiale⁵ $\text{Bin}(500, \theta)$.

5. En fait c'est une loi hypergéométrique (tirages sans remise), mais en raison du théorème 3.34, on peut la remplacer en pratique par une binomiale.

Une autre situation où il est naturel de considérer plusieurs lois pour une même variable aléatoire est celle du *conditionnement*. Rappelons que si (Ω, \mathcal{F}, P) est un espace probabilisé et $H \in \mathcal{F}$ un évènement tel que $P(H) > 0$, on peut définir sur \mathcal{F} une nouvelle mesure de probabilité $P_H = P(\cdot | H)$ par

$$\forall A \in \mathcal{F}, \quad P_H(A) := P(A | H) = \frac{P(A \cap H)}{P(H)}.$$

Définition 3.10 (loi conditionnelle). Soient (Ω, \mathcal{F}, P) un espace probabilisé, $H \in \mathcal{F}$ tel que $P(H) > 0$, X une variable aléatoire sur (Ω, \mathcal{F}) . On appelle loi conditionnelle de X sachant H , la loi de X sous P_H . En la notant $P_{X|H}$, on a donc

$$\forall B \in \text{Bor}(\mathbb{R}), \quad P_{X|H}(B) = P_H(X^{-1}(B)) = P(X \in B | H).$$

Il importe de ne pas se laisser induire en erreur par la notation $P_{X|H}$, elle ne concerne pas une nouvelle variable aléatoire « $X | H$ » mais bien toujours la même variable aléatoire X . Ce qui a changé, c'est la probabilité dont on munit (Ω, \mathcal{F}) et sous laquelle on considère la loi de X .

Dans le cas d'une variable aléatoire discrète X , il est facile de donner une formule explicite pour la loi P_X .

Proposition 3.11. Soient (Ω, \mathcal{F}, P) un espace probabilisé et X une variable aléatoire discrète sur (Ω, \mathcal{F}) . La loi de X sous P est la probabilité

$$P_X = \sum_{x \in X(\Omega)} P(X = x) \delta_x, \quad (3.11)$$

que l'on peut considérer comme probabilité sur $(X(\Omega), \mathcal{P}(X(\Omega)))$, ou sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ ou sur $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$.

Preuve. Vérifions l'égalité de mesures (3.11) sur $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$, le résultat analogue pour $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ et $(X(\Omega), \mathcal{P}(X(\Omega)))$ s'en déduisant immédiatement par restriction.

D'abord P_X est aussi une probabilité sur $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$, par une adaptation immédiate⁶ de la preuve de la proposition 3.7. En utilisant la décomposition (3.9) on a pour tout $B \subset \mathbb{R}$,

$$P_X(B) = P\left(\bigcup_{x \in B \cap X(\Omega)} X^{-1}(\{x\})\right) = \sum_{x \in B \cap X(\Omega)} P(X^{-1}(\{x\})) = \sum_{x \in X(\Omega)} P(X = x) \delta_x(B).$$

L'égalité de mesures (3.11) sur $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ est ainsi vérifiée. \square

Définition 3.12 (loi discrète sur \mathbb{R}). On appelle loi discrète sur \mathbb{R} , toute mesure ponctuelle μ sur $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ qui est aussi une probabilité. Une telle loi admet donc une représentation sous la forme

$$\mu = \sum_{i \in I} p_i \delta_{x_i},$$

où I est un ensemble d'indices au plus dénombrable, $(x_i)_{i \in I}$ est une famille de nombres réels et $(p_i)_{i \in I}$ est une famille sommable de réels positifs, de somme $\sum_{i \in I} p_i = 1$.

6. Rappelons qu'une v.a. discrète X sur (Ω, \mathcal{F}) est mesurable \mathcal{F} - $\mathcal{P}(\mathbb{R})$ et pas seulement \mathcal{F} - $\text{Bor}(\mathbb{R})$.

Il est clair que par restriction, μ est aussi une probabilité sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$.

Remarque 3.13. Deux variables aléatoires peuvent avoir même loi sans être égales. Par exemple considérons le jet de deux dés, l'un bleu et l'autre rouge. Notons X le nombre de points indiqué par le dé bleu et Y celui du rouge. Les variables aléatoires X et Y sont définies sur le même espace probabilisé $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ muni de l'équiprobabilité. On a $X(\Omega) = Y(\Omega) = \{1, 2, 3, 4, 5, 6\}$ et :

$$\forall k \in \{1, 2, 3, 4, 5, 6\}, \quad P(X = k) = \frac{1}{6}, \quad P(Y = k) = \frac{1}{6}.$$

Donc X et Y ont même loi : $P_X = P_Y = \sum_{k=1}^6 \frac{1}{6} \delta_k$. Pour autant, on n'a pas l'égalité des variables aléatoires X et Y qui signifierait $X(\omega) = Y(\omega)$ pour tout $\omega \in \Omega$ (égalité de deux applications). Autrement dit, en lançant deux dés on obtiendrait à coup sûr un double. Par contre nous pouvons considérer l'événement $\{X = Y\}$ dont la réalisation n'est pas certaine et calculer sa probabilité :

$$P(X = Y) = P\left(\bigcup_{k=1}^6 \{(X, Y) = (k, k)\}\right) = \frac{6}{36} = \frac{1}{6}.$$

On en déduit : $P(X \neq Y) = 5/6$.

Remarque 3.14. Deux variables aléatoires peuvent avoir même loi en étant définies sur des espaces probabilisés différents (Ω, \mathcal{F}, P) et $(\Omega', \mathcal{F}', P')$. Prenons par exemple pour X les points du dé bleu comme ci-dessus et posons $Z = 0$ si X est pair, $Z = 1$ si X est impair. La variable aléatoire Z est définie sur $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ muni de la tribu $\mathcal{P}(\Omega)$ et de l'équiprobabilité P sur Ω . Sa loi est $P_Z = \frac{1}{2}(\delta_0 + \delta_1)$. Prenons maintenant $\Omega' = \{-1, 1\}$, muni de la tribu $\mathcal{P}(\Omega')$ et de l'équiprobabilité P' sur Ω' et posons pour $\omega' \in \Omega'$, $Z'(\omega') := (1 + \omega')/2$. Alors la loi de Z' est aussi $P'_{Z'} = \frac{1}{2}(\delta_0 + \delta_1) = P_Z$. Remarquons que si X et Y sont définies sur des espaces probabilisés différents, il n'y a pas d'événement $\{X = Y\}$, pas plus que de variable aléatoire « $X + Y$ ». Essayez d'en écrire la définition explicite pour vous en convaincre.

La remarque suivante peut être sautée en première lecture.

Remarque 3.15. Si X est une v.a. discrète sur (Ω, \mathcal{F}) , alors pour toute probabilité P sur (Ω, \mathcal{F}) , la loi de X sous P est une loi discrète au sens de la définition 3.12. C'est une conséquence de la proposition 3.11. Mais il peut aussi exister sur (Ω, \mathcal{F}) une variable aléatoire réelle Y et une probabilité P telles que Y ne soit pas discrète (i.e. $Y(\Omega)$ est un ensemble infini non dénombrable) mais que la loi de Y sous P soit une loi discrète. Voici un exemple simple. On prend $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{P}(\mathbb{R})$ et $Y : \omega \mapsto \omega$ l'application identité sur \mathbb{R} . Alors Y n'est pas une v.a. discrète puisque $Y(\Omega) = \mathbb{R}$. Notons au passage que Y est mesurable relativement à n'importe quelle tribu sur l'ensemble d'arrivée puisque l'ensemble de départ Ω est muni ici de la tribu $\mathcal{P}(\Omega)$. En particulier, Y est bien une variable aléatoire réelle. Munissons maintenant $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{P}(\mathbb{R}))$ de la mesure de probabilité $P = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ et cherchons la loi de Y sous P . Comme Y est l'identité sur \mathbb{R} , on a l'égalité $Y^{-1}(B) = B$ pour tout borélien B . Par conséquent

$P_Y(B) = P(Y^{-1}(B)) = P(B)$ pour tout $B \in \text{Bor}(\mathbb{R})$. On a donc $P_Y = P = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$, ce qui montre que la loi de Y sous P est la loi discrète $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. Bien sûr on a un résultat analogue en remplaçant P par n'importe quelle loi discrète Q sur \mathbb{R} , au sens de la définition 3.12.

Remarque 3.16. Pour toute probabilité Q sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$, il existe au moins un espace probabilisé (Ω, \mathcal{F}, P) et une variable aléatoire réelle X sur (Ω, \mathcal{F}) dont la loi sous P soit égale à Q . Il suffit de prendre $\Omega = \mathbb{R}$, $\mathcal{F} = \text{Bor}(\mathbb{R})$ et pour X l'application identité de $\mathbb{R} \rightarrow \mathbb{R}$. En prenant $P = Q$, on a clairement $P_X = Q$. Bien entendu, il y a une infinité d'autres solutions à ce problème.

Il y a donc identité entre les mesures de probabilité sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ et les lois des variables aléatoires réelles. Comme nous savons caractériser une probabilité sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ par sa fonction de répartition (théorème 2.30), ceci va nous permettre de classer les lois des variables aléatoires réelles.

3.2.3 Fonction de répartition

Définition 3.17 (f.d.r. d'une variable aléatoire). Soient (Ω, \mathcal{F}, P) un espace probabilisé et X une variable aléatoire sur (Ω, \mathcal{F}) . On appelle fonction de répartition (f.d.r.) de X , la fonction F_X définie sur \mathbb{R} par :

$$\forall x \in \mathbb{R}, \quad F_X(x) = P_X([-\infty, x]) = P(X \leq x).$$

La fonction F_X est la fonction de répartition de la probabilité P_X , au sens de la définition 2.27. Elle ne dépend donc que de la loi⁷ de X . Deux variables aléatoires de même loi ont même fonction de répartition. La proposition suivante donnant les propriétés générales des fonctions de répartition des variables aléatoires n'est qu'une simple réécriture de la proposition 2.28.

Proposition 3.18. La fonction de répartition F_X d'une variable aléatoire X est croissante sur \mathbb{R} , avec limite 0 en $-\infty$ et 1 en $+\infty$. Elle est continue à droite et limitée à gauche en tout point de \mathbb{R} et vérifie :

$$\forall x \in \mathbb{R}, \quad P(X = x) = F(x) - F(x-). \quad (3.12)$$

La fonction de répartition d'une variable aléatoire caractérise sa loi, autrement dit : $F_X = F_Y$ si et seulement si les variables aléatoires X et Y ont même loi.

On peut aussi traduire la remarque 2.29 pour obtenir les formules suivantes de calcul

7. Il serait plus correct, mais plus long, de parler de f.d.r. de la loi de X ou même de f.d.r. de la loi de X sous P .

à l'aide de F_X des $P(X \in I)$ pour I intervalle de \mathbb{R} :

$$P(a < X \leq b) = F_X(b) - F_X(a), \quad (3.13)$$

$$P(a \leq X \leq b) = F_X(b) - F_X(a-), \quad (3.14)$$

$$P(a \leq X < b) = F_X(b-) - F_X(a-), \quad (3.15)$$

$$P(a < X < b) = F_X(b-) - F_X(a), \quad (3.16)$$

$$P(X \leq a) = F_X(a), \quad (3.17)$$

$$P(X < a) = F_X(a-), \quad (3.18)$$

$$P(X > b) = 1 - F_X(b), \quad (3.19)$$

$$P(X \geq b) = 1 - F_X(b-). \quad (3.20)$$

Dans le cas particulier des variables aléatoires discrètes, on peut donner une formule explicite de calcul de la fonction de répartition.

Proposition 3.19 (f.d.r. d'une variable aléatoire discrète). *Soient (Ω, \mathcal{F}, P) un espace probabilisé et X une variable aléatoire discrète sur (Ω, \mathcal{F}) . Fixons une numérotation de l'ensemble au plus dénombrable $X(\Omega)$ par les entiers : $X(\Omega) = \{x_0, x_1, \dots, x_k, \dots\}$ et notons $p_k := P(X = x_k)$. La fonction de répartition F_X vérifie alors :*

$$\forall x \in \mathbb{R}, \quad F_X(x) = \sum_{x_k \in X(\Omega)} p_k \mathbf{1}_{[x_k, +\infty[}(x), \quad (3.21)$$

ce qui s'écrit aussi

$$\forall x \in \mathbb{R}, \quad F_X(x) = \sum_{\substack{x_k \in X(\Omega) \\ x_k \leq x}} P(X = x_k). \quad (3.22)$$

De plus, si on peut numéroter les éléments de $X(\Omega)$ de manière croissante (i.e. $\forall k, x_k < x_{k+1}$), la fonction F_X est constante sur chaque intervalle $[x_n, x_{n+1}[$ et vaut sur cet intervalle $F_X(x) = \sum_{k \leq n} p_k$.

Cette proposition se généralise au cas des variables aléatoires de loi discrète⁸, en remplaçant $X(\Omega)$ par $X_P(\Omega) = \{x \in \mathbb{R}; P(X = x) > 0\}$.

Preuve. En utilisant (3.11), on obtient en effet pour tout $x \in \mathbb{R}$,

$$F_X(x) = P_X([-\infty, x]) = \sum_{x_k \in X(\Omega)} p_k \delta_{x_k}([-\infty, x]) = \sum_{x_k \in X(\Omega)} p_k \mathbf{1}_{[x_k, +\infty[}(x), \quad (3.23)$$

en notant que

$$\delta_{x_k}([-\infty, x]) = \begin{cases} 1 & \text{si } x_k \in]-\infty, x] \\ 0 & \text{si } x_k \notin]-\infty, x] \end{cases} = \begin{cases} 1 & \text{si } x_k \leq x \\ 0 & \text{si } x_k > x \end{cases} = \mathbf{1}_{[x_k, +\infty[}(x).$$

Si la suite (x_k) est strictement croissante, le réel x appartient à un seul intervalle $[x_n, x_{n+1}[$. Pour $k \leq n$, on a alors $x_k \leq x_n \leq x$ et $\mathbf{1}_{[x_k, +\infty[}(x) = 1$, tandis que si $k > n$, $x_k \geq x_{n+1} > x$, donc $\mathbf{1}_{[x_k, +\infty[}(x) = 0$. On a donc $F_X(x) = \sum_{k \leq n} p_k$ et ceci étant valable pour tout $x \in [x_n, x_{n+1}[$, la fonction F_X est constante sur cet intervalle. \square

8. Voir la remarque 3.15.

À titre d'exemple, la figure 3.2 donne la représentation graphique de F_S où S est la variable aléatoire somme des points de deux dés.

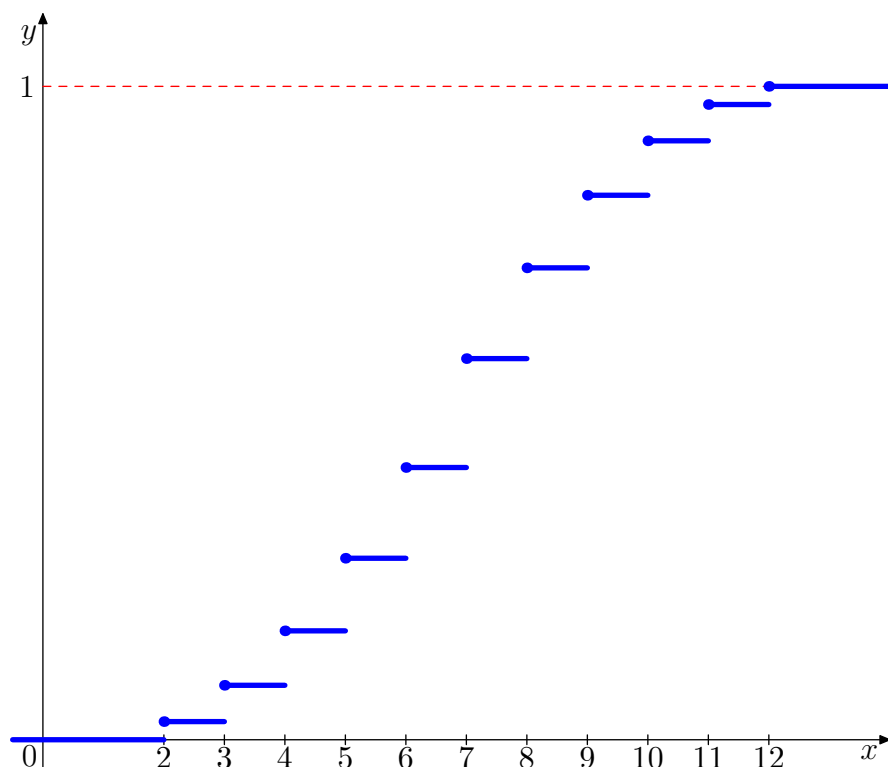


FIG. 3.2 – f.d.r. de S somme des points de deux dés

Si l'on veut esquisser une classification sommaire des lois des variables aléatoires, on peut commencer par les partager entre les lois à f.d.r. continue sur \mathbb{R} et les lois à f.d.r. non continue⁹ sur \mathbb{R} . On parle plus simplement de *lois continues* ou encore *lois diffuses* dans le premier cas et de lois non continues ou non diffuses dans le deuxième. Dans la famille des lois non continues, nous connaissons déjà la sous-famille des lois discrètes. Dans la famille des lois continues, une importante sous-famille est celle des *lois à densité* que nous allons examiner maintenant.

3.2.4 Lois à densité

La loi d'une variable aléatoire X est à densité f si pour tout intervalle de \mathbb{R} , la probabilité d'appartenance de X à cet intervalle peut s'écrire comme l'intégrale de f sur cet intervalle. L'apparente simplicité de cette définition informelle est trompeuse. Dans le cadre de ce cours, nous ne pouvons utiliser que l'intégration au sens de Riemann et il se trouve que cette notion n'est pas totalement satisfaisante pour les besoins de la théorie des probabilités. L'intégrale de Lebesgue donnerait une notion plus générale de densité,

9. Rappelons que l'ensemble des points de discontinuité d'une f.d.r. quelconque est au plus dénombrable.

permettant entre autres de caractériser les lois à densité comme celles dont la f.d.r. est *absolument continue*¹⁰. La définition plus restrictive que nous donnons ci-dessous est néanmoins suffisante pour la plupart des cas pratiques.

Définition 3.20 (densité de probabilité). *On appelle densité de probabilité sur \mathbb{R} toute fonction f vérifiant*

- a) *f est définie et positive sur $\mathbb{R} \setminus K$, où K est une partie finie (éventuellement vide) de \mathbb{R} ;*
- b) *f est Riemann intégrable sur tout intervalle $[a, b] \subset \mathbb{R} \setminus K$;*
- c) *l'intégrale généralisée de f sur $] -\infty, +\infty[$ converge et*

$$\int_{-\infty}^{+\infty} f(t) dt = 1.$$

Si f est une fonction positive définie seulement sur un intervalle $]a, b[$ de \mathbb{R} et telle que $\int_a^b f(t) dt = 1$, on peut en faire une densité en la prolongeant à tout \mathbb{R} en posant $f(t) := 0$ pour $t \notin]a, b[$. Voici quatre exemples simples de densités :

$$f_1(t) := \frac{1}{b-a} \mathbf{1}_{[a,b]}(t); \quad f_2(t) := \frac{1}{2\sqrt{t}} \mathbf{1}_{]0,1]}(t);$$

$$f_3(t) := e^{-t} \mathbf{1}_{[0,+\infty[}(t); \quad f_4(t) := \frac{1}{\pi(1+t^2)}.$$

Remarque 3.21 (usage des indicatrices dans les formules explicites). La définition de f_2 repose sur un *abus d'écriture* d'usage courant. En effet il y a en toute rigueur un problème pour calculer $f_2(t)$ lorsque $t \leq 0$, puisqu'alors il nous faut former le produit de l'expression $\frac{1}{2\sqrt{t}}$ non définie (du moins en tant que nombre réel) par 0. La convention adoptée est que si la formule de calcul d'une fonction contient le produit d'une indicatrice par une expression non définie lorsque cette indicatrice est nulle, le produit vaut 0 dans ce cas. Ceci permet de considérer que la « définition » de f_2 comme ci-dessus est un raccourci d'écriture commode pour :

$$f_2(t) := \begin{cases} \frac{1}{2\sqrt{t}} & \text{si } t \in]0, 1], \\ 0 & \text{si } t \notin]0, 1]. \end{cases}$$

Définition 3.22. Soient (Ω, \mathcal{F}, P) un espace probabilisé et X une variable aléatoire réelle sur (Ω, \mathcal{F}) . La loi de X sous P a pour densité f si :

$$\forall a \in \mathbb{R}, \forall b \geq a, \quad P(X \in]a, b]) = \int_a^b f(t) dt. \quad (3.24)$$

On dit aussi par abus de langage que X a pour densité f (lorsqu'il n'y a pas ambiguïté sur P).

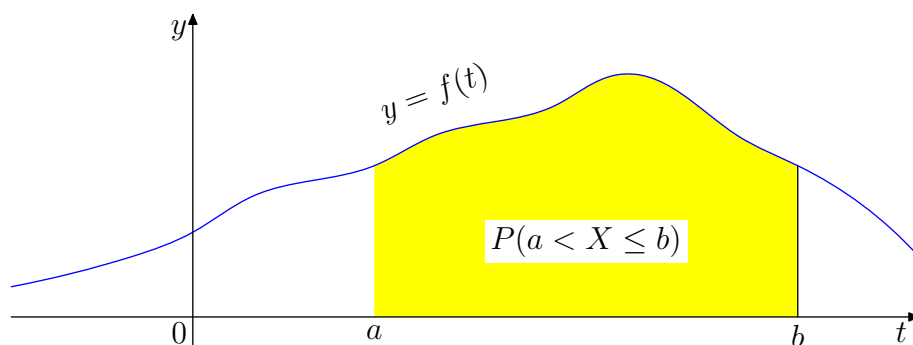


FIG. 3.3 – $P(a < X \leq b) = \int_a^b f(t) dt$ pour X de densité f

Remarque 3.23. Il est clair d'après cette définition que si Y est une autre variable aléatoire ayant même loi que X (donc mêmes probabilités d'appartenance aux intervalles), elle a aussi la densité f . D'autre part, il n'y a pas unicité de la densité d'une variable aléatoire. Par exemple $g_1 = \mathbf{1}_{[0,1]}$ et $g_2 = \mathbf{1}_{]0,1]}$ sont deux densités de probabilité qui donnent les mêmes intégrales : $\int_a^b g_1(t) dt = \int_a^b g_2(t) dt$ pour toute paire de réels a et b . Ces deux fonctions peuvent chacune être prise comme densité de la loi uniforme sur $[0, 1]$ (nous y reviendrons ci-dessous).

L'exemple ci-dessus montre qu'il ne suffit pas de vérifier que deux variables aléatoires ont des densités qui diffèrent en un point pour en déduire qu'elles n'ont pas même loi. Le lemme suivant donne une condition suffisante pratique pour que deux variables à densité n'aient pas même loi.

Lemme 3.24. Soient X et Y deux variables aléatoires admettant respectivement pour densité les fonctions f et g . On suppose qu'il existe un réel t_0 tel que $f(t_0) \neq g(t_0)$ et que de plus, f et g sont toutes deux continues au point t_0 . Alors X et Y n'ont pas même loi.

Preuve. On peut supposer sans perte de généralité que $f(t_0) < g(t_0)$. On va exploiter la continuité de f et g au point t_0 pour construire un intervalle $[a, b]$ voisinage de t_0 tel que $\int_a^b f(t) dt < \int_a^b g(t) dt$. Comme ces deux intégrales sont égales respectivement à $P(X \in]a, b])$ et $P(Y \in]a, b])$, ceci impliquera que X et Y n'ont pas même loi.

Fixons $\varepsilon > 0$ tel que $f(t_0) + \varepsilon < g(t_0) - \varepsilon$, par exemple $\varepsilon = (g(t_0) - f(t_0))/3$. Par continuité de f et g au point t_0 , il existe deux réels $\delta_1 > 0$ et $\delta_2 > 0$ tels que :

$$\forall t \in]t_0 - \delta_1, t_0 + \delta_1[, \quad f(t) < f(t_0) + \varepsilon \quad \text{et} \quad \forall t \in]t_0 - \delta_2, t_0 + \delta_2[, \quad g(t) > g(t_0) - \varepsilon.$$

L'intersection des deux intervalles ouverts $]t_0 - \delta_1, t_0 + \delta_1[$ et $]t_0 - \delta_2, t_0 + \delta_2[$ contient un intervalle $[a, b]$ avec $a < b$, on peut prendre par exemple $a = t_0 - \frac{1}{2} \min(\delta_1, \delta_2)$ et

10. La fonction $F : \mathbb{R} \rightarrow \mathbb{R}$ est dite *absolument continue* sur \mathbb{R} , si pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que pour toute famille finie d'intervalles $[a_k, b_k]$ ($k = 1, \dots, n$) deux à deux disjoints et vérifiant $\sum_{k=1}^n (b_k - a_k) \leq \delta$, on ait $\sum_{k=1}^n |F(b_k) - F(a_k)| \leq \varepsilon$. Cette propriété est plus forte que la continuité uniforme sur \mathbb{R} . Par ailleurs toute f.d.r. continue sur \mathbb{R} est *uniformément* continue sur \mathbb{R} (exercice).

$b = t_0 + \frac{1}{2} \min(\delta_1, \delta_2)$. On dispose alors des inégalités suivantes :

$$\forall t \in [a, b], \quad f(t) < f(t_0) + \varepsilon < g(t_0) - \varepsilon < g(t),$$

qui par intégration sur $[a, b]$ nous donnent

$$\int_a^b f(t) dt \leq (b-a)(f(t_0) + \varepsilon) < (b-a)(g(t_0) - \varepsilon) \leq \int_a^b g(t) dt,$$

d'où $\int_a^b f(t) dt \neq \int_a^b g(t) dt$, ce qui achève la preuve ¹¹. □

Examinons maintenant les relations entre densité (lorsqu'elle existe) et fonction de répartition (qui elle, existe toujours) d'une variable aléatoire.

Proposition 3.25. *Si la variable aléatoire X a pour densité f , sa fonction de répartition F vérifie :*

a) $\forall x \in \mathbb{R}, F(x) = \int_{-\infty}^x f(t) dt ;$

b) F est continue sur $\mathbb{R} ;$

c) si f est continue au point x_0 , alors F est dérivable en ce point et $F'(x_0) = f(x_0)$.

Corollaire 3.26. *Si la variable aléatoire X a pour densité f , on a pour $a, b \in \mathbb{R}$ quelconques,*

$$P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) = \int_a^b f(t) dt,$$

$$P(X < a) = P(X \leq a) = \int_{-\infty}^a f(t) dt,$$

$$P(X > b) = P(X \geq b) = \int_b^{+\infty} f(t) dt.$$

Preuve de la prop. 3.25 et du corollaire 3.26.

Preuve de a). Puisque X a pour densité f , on a pour tous réels $a < b$,

$$P(X \in]a, b]) = F(b) - F(a) = \int_a^b f(t) dt. \quad (3.25)$$

Il suffit d'appliquer (3.25) avec $b = x$ fixé et $a = -n$ pour chaque $n \in \mathbb{N}$ tel que $-n < x$. La suite d'événements

$$A_n := \{X \in]-n, x]\}, \quad n > -x,$$

est croissante pour l'inclusion et a pour réunion $A = \{X \in]-\infty, x]\}$. Par continuité monotone séquentielle (cf. proposition 2.16), on a $P(A_n) \uparrow P(A)$, d'où

$$F(x) = P(A) = \lim_{n \rightarrow +\infty} P(X \in]-n, x]) = \lim_{n \rightarrow +\infty} \int_{-n}^x f(t) dt = \int_{-\infty}^x f(t) dt,$$

en notant que l'intégrale généralisée de la densité f converge en $-\infty$. □

11. Parmi les inégalités de la ligne précédente, seule la deuxième est stricte et ce n'est pas une erreur typographique. Voyez-vous pourquoi ?

Preuve de b). Fixons $x_0 \in \mathbb{R}$ quelconque. On sait déjà que F est continue à droite en tout point comme *toute* fonction de répartition. Il suffit donc de montrer la continuité à gauche en x_0 . D'après le point b) de la définition 3.20, il existe $a < x_0$ tel que f soit définie et Riemann intégrable sur tout intervalle $[a, a'] \subset [a, x_0[$. On a alors

$$\lim_{x \uparrow x_0} \int_a^x f(t) dt = \int_a^{x_0} f(t) dt,$$

où la deuxième intégrale est soit une intégrale de Riemann ordinaire soit une intégrale généralisée convergente. Cette relation peut aussi s'écrire à l'aide de F :

$$\lim_{x \uparrow x_0} (F(x) - F(a)) = F(x_0) - F(a).$$

On en déduit par addition de $F(a)$ que $F(x)$ tend vers $F(x_0)$ quand x tend vers x_0 par valeurs inférieures. \square

Preuve de c). Puisque f est continue en x_0 , elle est définie sur tout un voisinage de x_0 et donc sur tout un intervalle $]a, b[$ contenant x_0 . La continuité de f en x_0 peut alors s'écrire :

$$\forall \varepsilon > 0, \exists]x_0 - \delta, x_0 + \delta[\subset]a, b[; \quad \forall t \in]x_0 - \delta, x_0 + \delta[, \quad |f(t) - f(x_0)| < \varepsilon. \quad (3.26)$$

Pour tout h tel que $0 < |h| < \delta$, on a alors $F(x_0 + h) - F(x_0) = \int_{x_0}^{x_0+h} f(t) dt$ d'où

$$|F(x_0 + h) - F(x_0) - hf(x_0)| = \left| \int_{x_0}^{x_0+h} (f(t) - f(x_0)) dt \right| \leq h\varepsilon.$$

En divisant par h on voit que F a bien une dérivée en x_0 et que celle ci vaut $f(x_0)$. \square

La proposition 3.25 est maintenant complètement démontrée. Pour le corollaire 3.26, il suffit de combiner les relations générales (3.13)–(3.20) avec (3.24) et les points a) et b) de la proposition 3.25. \square

Remarques 3.27.

1. Pour toute densité f (au sens de la définition 3.20), il existe une variable aléatoire X ayant f pour densité : il suffit d'appliquer le théorème 2.30 pour obtenir l'existence d'une mesure de probabilité μ sur $(\mathbb{R}, \text{Bor}(\mathbb{R}))$ ayant pour f.d.r F définie par a). La remarque 3.16 nous assure de l'existence d'une variable aléatoire X de loi μ , donc de f.d.r. F . La preuve du b) ci-dessus nous montre que F est continue sur \mathbb{R} . En particulier pour toute paire de réels $a \leq b$, on a $P(a < X \leq b) = F(b) - F(a) = \int_{-\infty}^b f(t) dt - \int_{-\infty}^a f(t) dt = \int_a^b f(t) dt$.
2. D'après b) toute variable aléatoire à densité a une fonction de répartition continue. La réciproque *est fausse* : il existe des lois à fonction de répartition continue sans densité.

3. Par ailleurs si X a une densité, sa fonction de répartition n'est pas forcément dérivable en tout point. Par exemple la densité f_2 ci-dessus a pour fonction de répartition associée $F_2(x) = \sqrt{x}\mathbf{1}_{]0,1]}(x) + \mathbf{1}_{]1,+\infty[}(x)$ (cette écriture condensée signifie que $F_2(x)$ est nul sur \mathbb{R}^- , vaut \sqrt{x} entre 0 et 1 et reste constant égal à 1 sur $]1, +\infty[$). F_2 est dérivable en tout point sauf en 0 et en 1.

La proposition suivante donne une règle pratique permettant de trouver la densité (lorsqu'elle existe!) à partir de la fonction de répartition dans les cas les plus courants.

Proposition 3.28. *On suppose que la fonction de répartition F de X est C^1 par morceaux au sens suivant : F est continue sur \mathbb{R} et dérivable sur \mathbb{R} privé (éventuellement) d'un ensemble fini de points $a_1 < \dots < a_n$. Sur chacun des intervalles ouverts $] -\infty, a_1[$, $]a_i, a_{i+1}[$ ($1 \leq i < n$), $]a_n, +\infty[$, la dérivée f de F est continue. Alors X a pour densité f .*

Preuve. Il est commode de poser $a_0 := -\infty$ et $a_{n+1} = +\infty$. Sur chacun des intervalles ouverts I découpés par les a_i , F est dérivable et sa dérivée f est continue. On sait alors que f a une infinité de primitives sur I et que si l'on fixe un α dans I , toute primitive H de f sur I est de la forme $H(x) = \int_{\alpha}^x f(t) dt + C$, avec C constante. Comme F est l'une des primitives de f sur I , en prenant $H = F$ et en faisant $x = \alpha$, on voit que la constante C vaut $F(\alpha)$. On a donc pour α et x quelconques dans I , $F(x) - F(\alpha) = \int_{\alpha}^x f(t) dt$. Fixons α et prenons $x \geq \alpha$. Faisons tendre x vers la borne supérieure a_i de I . Comme F est continue (ou dans le cas $a_i = +\infty$, F a une limite 1), l'intégrale généralisée $\int_{\alpha}^{a_i} f(t) dt$ converge et vaut $F(a_i) - F(\alpha)$ (ou $1 - F(\alpha)$ quand $a_i = +\infty$). De même en faisant tendre α vers a_{i-1} on voit que l'intégrale généralisée $\int_{a_{i-1}}^{a_i} f(t) dt$ converge et vaut $F(a_i) - F(a_{i-1})$ (ou $F(a_i)$ quand $a_{i-1} = -\infty$). Finalement soient a et $b > a$ quelconques dans \mathbb{R} . Si a et b sont dans le même intervalle I on a directement $F(b) - F(a) = \int_a^b f(t) dt$. Sinon on note $(a_i)_{i_0 \leq i \leq i_1}$ l'ensemble de tous les a_i qui sont dans $[a, b]$ et on écrit

$$F(b) - F(a) = F(a_{i_0}) - F(a) + \sum_{i=i_0}^{i_1-1} (F(a_{i+1}) - F(a_i)) + F(b) - F(a_{i_1}) = \int_a^b f(t) dt,$$

en utilisant la relation de Chasles pour les intégrales généralisées. On a donc toujours $P(X \in]a, b]) = F(b) - F(a) = \int_a^b f(t) dt$, ce qui montre que X a pour densité f . \square

3.3 Lois discrètes classiques

Dans toute la suite du chapitre, on fixe un espace probabilisé (Ω, \mathcal{F}, P) , on désigne par X une variable aléatoire sur (Ω, \mathcal{F}) et par P_X sa loi sous P . Cette clause sera implicite chaque fois que nous écrirons dans les définitions « La variable aléatoire X suit la loi...si... ». Pour X de loi discrète, nous utiliserons la notation :

$$X_P(\Omega) := \{x \in \mathbb{R}; P(X = x) > 0\}. \quad (3.27)$$

Bien sûr, $X_P(\Omega)$ est toujours inclus dans l'ensemble des valeurs « possibles » $X(\Omega)$ et dans la plupart des situations pratiques d'utilisation des variables aléatoires, ces deux ensembles sont égaux¹².

3.3.1 Lois de Bernoulli

Définition 3.29. *La variable aléatoire X suit la loi de Bernoulli de paramètre p ($p \in [0, 1]$) si $X_P(\Omega) = \{0, 1\}$ avec :*

$$P(X = 1) = p, \quad P(X = 0) = 1 - p = q.$$

On notera $X \sim \text{Bern}(p)$.

Si A est un événement de probabilité p , son indicatrice définie par

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases} = \begin{cases} 1 & \text{si } A \text{ est réalisé} \\ 0 & \text{si } A \text{ n'est pas réalisé} \end{cases}$$

est une variable aléatoire suivant la loi de Bernoulli de paramètre p . Réciproquement, si X est une v.a. de Bernoulli, on peut toujours écrire que $X = \mathbf{1}_A$ presque-sûrement, c'est-à-dire $P(X = \mathbf{1}_A) = 1$, en définissant $A = \{\omega \in \Omega, X(\omega) = 1\}$.

3.3.2 Loi uniforme sur un ensemble fini de réels

Définition 3.30. *La variable aléatoire X suit la loi uniforme sur l'ensemble de réels $\{x_1, \dots, x_n\}$ si P_X est l'équiprobabilité sur cet ensemble. Notation : $X \sim \text{Unif}\{x_1, \dots, x_n\}$.*

Autrement dit, l'ensemble des valeurs possibles de X est $X(\Omega) = \{x_1, \dots, x_n\}$ et :

$$\forall k = 1, \dots, n, \quad P(X = x_k) = \frac{1}{n}.$$

D'où

$$P_X = \frac{1}{n} \sum_{k=1}^n \delta_{x_k}.$$

Par exemple le nombre de points indiqué par un dé suit la loi uniforme sur $\{1, 2, 3, 4, 5, 6\}$.

3.3.3 Lois binomiales

Définition 3.31. *La variable aléatoire X suit la loi binomiale de paramètres n et p ($n \in \mathbb{N}^*$ et $p \in [0, 1]$) si $X_P(\Omega) = \{0, 1, \dots, n\}$ et*

$$\forall k = 0, 1, \dots, n, \quad P(X = k) = C_n^k p^k (1 - p)^{n-k}.$$

Notation : $X \sim \text{Bin}(n, p)$.

12. Pour comprendre l'utilité de ce distinguo entre $X(\omega)$ et $X_P(\Omega)$, (re)lisez la remarque 3.15.

La formule ci-dessus définit bien une loi de probabilité puisque les $C_n^k p^k (1-p)^{n-k}$ sont positifs et :

$$\sum_{k=0}^n C_n^k p^k (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1,$$

en appliquant la formule du binôme de Newton (d'où le nom de la loi). La loi binomiale $\text{Bin}(n, p)$ est la loi du nombre de succès obtenus en une suite de n épreuves répétées indépendantes avec pour chaque épreuve une probabilité de succès p . Ceci a été démontré dans l'exemple 2.57.

De même, soit A_1, \dots, A_n une famille de n événements mutuellement indépendants ayant tous même probabilité p et notons X_i la variable de Bernoulli indicatrice de A_i :

$$X_i(\omega) = \begin{cases} 1 & \text{si } \omega \in A_i, \\ 0 & \text{si } \omega \in A_i^c. \end{cases}$$

Alors la variable aléatoire $S_n = \sum_{i=1}^n X_i$ suit la loi binomiale $\text{Bin}(n, p)$.

3.3.4 Lois hypergéométriques

Alors que la loi binomiale intervient dans les tirages avec remise, la loi hypergéométrique correspond aux tirages sans remise.

Exemple 3.32. Dans une production totale de N objets dont M sont défectueux, on prélève au hasard un échantillon de n objets (tirage sans remise). Soit X le nombre aléatoire d'objets défectueux dans l'échantillon. Quelle est sa loi ?

On peut prendre comme espace Ω l'ensemble de tous les échantillons possibles (toutes les parties à n éléments d'un ensemble de cardinal N) muni de l'équiprobabilité. Chaque échantillon a ainsi une probabilité $1/C_N^n$ d'être choisi. Les échantillons (événements élémentaires) réalisant l'événement $\{X = k\}$ sont ceux qui contiennent k objets défectueux et $n-k$ objets non défectueux. Ceci n'est réalisable que si $0 \leq k \leq M$ et $0 \leq n-k \leq N-M$. Dénombrons ces échantillons. On les forme en choisissant k objets défectueux dans une sous-population de M et en complétant par $n-k$ objets non défectueux choisis dans une sous population de $N-M$. Il y en a donc $C_M^k \times C_{N-M}^{n-k}$. Finalement :

$$P(X = k) = \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} \quad \text{si} \quad \begin{cases} 0 \leq k \leq M, \\ 0 \leq n-k \leq N-M. \end{cases} \quad (3.28)$$

Définition 3.33. La loi définie par (3.28) s'appelle loi hypergéométrique de paramètres N , M et n . Notation : $X \sim \text{Hypg}(N, M, n)$. Le paramètre N est l'effectif de la population totale, M celui de la sous-population à laquelle on s'intéresse et n la taille de l'échantillon observé.

Pour une taille d'échantillon n fixée, plus N et M sont grands, moins les tirages sans remise diffèrent des tirages avec remise. Plus précisément, la loi hypergéométrique converge vers la loi binomiale au sens suivant.

Théorème 3.34 (convergence de l'hypergéométrique vers la binomiale). *On suppose que quand N tend vers $+\infty$, $M = M(N)$ tend vers $+\infty$ en vérifiant la condition :*

$$\lim_{N \rightarrow +\infty} \frac{M}{N} = p \quad \text{avec} \quad 0 < p < 1. \quad (3.29)$$

Alors, n restant fixé, la loi hypergéométrique $\text{Hypg}(N, M, n)$ converge vers la loi binomiale $\text{Bin}(n, p)$, ce qui signifie que si $(X_N)_{N \geq 1}$ est une suite de v.a. avec $X_N \sim \text{Hypg}(N, M, n)$ et Y est une v.a. de loi binomiale $\text{Bin}(n, p)$, alors :

$$\forall k = 0, 1, \dots, n, \quad \lim_{N \rightarrow +\infty} P(X_N = k) = P(Y = k), \quad (3.30)$$

autrement dit :

$$\forall k = 0, 1, \dots, n, \quad \lim_{N \rightarrow +\infty} \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} = C_n^k p^k (1-p)^{n-k}. \quad (3.31)$$

Preuve. Remarquons d'abord que comme p est strictement positif, l'hypothèse (3.29) implique que M tend vers $+\infty$ avec N ; il en va de même pour $N - M$ puisque $p < 1$.

Pour n et k fixés, posons :

$$\begin{aligned} p_N &= \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} \\ &= \frac{M!}{k!(M-k)!} \times \frac{(N-M)!}{(n-k)!((N-M)-(n-k))!} \times \frac{n!(N-n)!}{N!} \\ &= C_n^k \frac{M!}{(M-k)!} \times \frac{(N-M)!}{((N-M)-(n-k))!} \times \frac{(N-n)!}{N!}. \end{aligned} \quad (3.32)$$

Comme k est fixé et M tend vers $+\infty$, la première fraction dans (3.32) est le produit de k facteurs $M, (M-1), \dots, (M-k+1)$ tous équivalents¹³ à M d'où :

$$\frac{M!}{(M-k)!} \sim M^k, \quad N \rightarrow +\infty. \quad (3.33)$$

Par le même argument avec $n-k$ et $N-M$ au lieu de k et M :

$$\frac{(N-M)!}{((N-M)-(n-k))!} \sim (N-M)^{n-k}, \quad N \rightarrow +\infty. \quad (3.34)$$

Enfin :

$$\frac{(N-n)!}{N!} \sim \frac{1}{N^n}, \quad N \rightarrow +\infty. \quad (3.35)$$

13. Rappelons que deux suites (u_N) et (v_N) sont dites équivalentes lorsque $u_N = v_N(1 + \varepsilon_N)$ avec ε_N tendant vers 0 quand N tend vers $+\infty$ (notation : $u_N \sim v_N$).

En reportant ces équivalents dans (3.32), on voit que lorsque N tend vers $+\infty$:

$$p_N \sim C_n^k \frac{M^k (N-M)^{n-k}}{N^n} = C_n^k \left(\frac{M}{N}\right)^k \left(\frac{N-M}{N}\right)^{n-k}, \quad (3.36)$$

d'où : $\lim_{N \rightarrow +\infty} p_N = C_n^k p^k (1-p)^{n-k}$. □

3.3.5 Lois géométriques

Exemple 3.35 (un problème de temps d'attente).

Considérons une suite infinie d'épreuves répétées indépendantes avec même probabilité de succès $p \in]0, 1[$. Soit X le numéro (aléatoire) de la première épreuve où l'on obtient un succès. Si l'on n'obtient jamais de succès, on conviendra que $X = +\infty$. Calculer $P(X = k)$ pour tout $k \in \mathbb{N}^*$. En déduire les valeurs de $P(X \in \mathbb{N}^*)$ et $P(X = +\infty)$.

En notant $R_i = \{\text{succès à la } i\text{-ème épreuve}\}$, on a :

$$\begin{aligned} \{X = k\} &= \{\text{échec aux } (k-1) \text{ premières et succès à la } k\text{-ième}\} \\ &= \left(\bigcap_{i=1}^{k-1} R_i^c\right) \cap R_k. \end{aligned}$$

D'où par indépendance des épreuves :

$$P(X = k) = \left(\prod_{i=1}^{k-1} P(R_i^c)\right) \times P(R_k) = (1-p)^{k-1} p.$$

Posons $q = 1 - p$ et notons que $q \in]0, 1[$. La décomposition de l'événement $\{X \in \mathbb{N}^*\}$ en la réunion disjointe des $\{X = k\}$ ($k \in \mathbb{N}^*$) nous donne par σ -additivité :

$$\begin{aligned} P(X \in \mathbb{N}^*) &= \sum_{k \in \mathbb{N}^*} P(X = k) = \sum_{k \in \mathbb{N}^*} q^{k-1} p \\ &= p \sum_{l \in \mathbb{N}} q^l \quad (l = k-1) \\ &= p \frac{1}{1-q} = 1. \end{aligned}$$

Ainsi avec probabilité 1, le premier succès apparaît au bout d'un nombre *fini* d'épreuves¹⁴. Remarquons qu'on aurait pu arriver au même résultat en montrant que $P(X = +\infty) = 0$ par la méthode utilisée à l'exemple 2.57 c) en échangeant les rôles de succès et échec. En toute rigueur, X n'est pas une variable aléatoire discrète au sens de la définition 3.6 puisque $X(\Omega)$ est une partie dénombrable de $\overline{\mathbb{R}}$ au lieu de \mathbb{R} . Néanmoins $X' := X \mathbf{1}_{\{X < +\infty\}}$ est une variable aléatoire discrète et ce qui précède montre que X' a même loi¹⁵ que X . Cette loi est celle du temps d'attente du premier succès dans une suite d'épreuves répétées indépendantes, on l'appelle *loi géométrique de paramètre p* .

14. Mais pas borné par un nombre fixé choisi avant le début des épreuves...

15. Pour être tout à fait rigoureux, il faudrait avoir défini les v.a. à valeurs dans $\overline{\mathbb{R}}$ et leurs lois, ce qui nous aurait fait sortir du cadre de ce cours...

Définition 3.36. Une variable aléatoire X suit la loi géométrique de paramètre $p \in]0, 1[$, si $X_P(\Omega) = \mathbb{N}^*$ et :

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

Notation : $X \sim \text{Geom}(p)$.

Lorsque X suit une loi géométrique, les probabilités $P(X > n)$ ont une expression particulièrement simple en fonction de $q = 1 - p$. Calculons les de deux façons.

Première méthode. On calcule le reste d'une série géométrique :

$$\begin{aligned} P(X > n) &= \sum_{k=n+1}^{+\infty} q^{k-1}p = \sum_{l=n}^{+\infty} q^l p \\ &= pq^n \sum_{l=n}^{+\infty} q^{l-n} = pq^n \sum_{j=0}^{+\infty} q^j \\ &= \frac{pq^n}{1 - q} = q^n. \end{aligned}$$

Deuxième méthode. On se place dans la situation de l'exemple 3.35. L'événement $\{X > n\}$ se réalise si et seulement si les n premières épreuves donnent un échec.

$$\{X > n\} = \bigcap_{i=1}^n R_i^c.$$

En utilisant l'indépendance des R_i on en déduit :

$$P(X > n) = \prod_{i=1}^n P(R_i^c) = q^n.$$

3.3.6 Lois de Poisson

Définition 3.37. On dit que la variable aléatoire discrète X suit la loi de Poisson de paramètre $\alpha > 0$ si $X_P(\Omega) = \mathbb{N}$ et

$$\forall k \in \mathbb{N}, \quad P(X = k) = \frac{e^{-\alpha} \alpha^k}{k!}.$$

Notation : $X \sim \text{Pois}(\alpha)$.

On sait (cf. cours d'analyse) que la fonction exponentielle a un développement en série entière avec rayon de convergence infini. En particulier :

$$\forall \alpha > 0, \quad e^\alpha = \sum_{k=0}^{+\infty} \frac{\alpha^k}{k!}.$$

On a donc bien :

$$\sum_{k=0}^{+\infty} P(X = k) = e^{-\alpha} \sum_{k=0}^{+\infty} \frac{\alpha^k}{k!} = e^{-\alpha} e^\alpha = 1.$$

Une des raisons de l'importance de cette loi est le théorème de convergence de la loi binomiale vers la loi de Poisson.

Théorème 3.38. Si $(p_n)_{n \geq 1}$ est une suite de réels de $[0, 1]$ vérifiant

$$np_n \rightarrow \alpha \in]0, +\infty[, \quad \text{quand } n \rightarrow +\infty, \quad (3.37)$$

alors :

$$\forall k \in \mathbb{N}, \quad C_n^k p_n^k (1 - p_n)^{n-k} \longrightarrow \frac{e^{-\alpha} \alpha^k}{k!}, \quad \text{quand } n \rightarrow +\infty.$$

Preuve. L'hypothèse (3.37) peut s'écrire sous la forme plus maniable : $np_n = \alpha u_n$ avec u_n tendant vers 1 quand n tend vers $+\infty$. Ainsi $p_n = \alpha u_n / n$ et

$$C_n^k p_n^k (1 - p_n)^{n-k} = \frac{n!}{(n-k)!} \frac{1}{k!} \left(\frac{\alpha}{n}\right)^k u_n^k \left(1 - \frac{\alpha u_n}{n}\right)^{n-k}. \quad (3.38)$$

Pour obtenir la limite de cette expression lorsque n tend vers $+\infty$, k restant fixé, on remarque successivement que :

$$\lim_{n \rightarrow +\infty} \frac{n!}{(n-k)!} \frac{1}{n^k} = 1, \quad (3.39)$$

$$\lim_{n \rightarrow +\infty} u_n^k = 1, \quad (3.40)$$

$$\lim_{n \rightarrow +\infty} \left(1 - \frac{\alpha u_n}{n}\right)^{n-k} = e^{-\alpha}. \quad (3.41)$$

Pour justifier (3.41), on écrit :

$$\left(1 - \frac{\alpha u_n}{n}\right)^{n-k} = \exp\left[(n-k) \ln\left(1 - \frac{\alpha u_n}{n}\right)\right], \quad (3.42)$$

puis comme $\alpha u_n / n$ tend vers 0 :

$$(n-k) \ln\left(1 - \frac{\alpha u_n}{n}\right) \sim n\left(-\frac{\alpha u_n}{n}\right) \sim -\alpha, \quad (n \rightarrow +\infty).$$

Par continuité de la fonction exponentielle, la limite du second membre de (3.42) est donc bien $e^{-\alpha}$, ce qui prouve (3.41). On obtient alors la conclusion du théorème en passant à la limite dans (3.38). \square

Le théorème 3.38 sert de justification théorique à la règle pratique suivante : lorsque n est « grand » et np « petit », on peut remplacer la loi binomiale $\text{Bin}(n, p)$ par la loi de Poisson $\text{Pois}(\alpha)$ où $\alpha = np$. En général on considère que n de l'ordre de quelques centaines et np de l'ordre de quelques unités donnent une bonne approximation. Sous cette forme, cette règle relève plus de la cuisine que des mathématiques. Il est possible par des techniques élémentaires de contrôler l'erreur commise en utilisant cette approximation. Nous nous contenterons ici d'un exemple classique et d'une comparaison graphique pour illustrer la qualité de cette approximation.

Exemple 3.39. Le président d'un bureau de vote est né un 1^{er} avril. Il décide de noter le nombre X de personnes ayant leur anniversaire le même jour que lui parmi les 500 premiers électeurs qui se présentent.

La situation peut être assimilée à une suite d'épreuves répétées indépendantes et X est une variable aléatoire binomiale de paramètres $n = 500$ et $p = 1/365$ (en négligeant la question des années bissextiles sinon on prendrait $p = 4/(3 \times 365 + 366)$, ce qui ne changerait pas grand chose numériquement). Ainsi :

$$P(X = k) = C_{500}^k \left(\frac{1}{365}\right)^k \left(\frac{364}{365}\right)^{500-k}.$$

La règle énoncée ci-dessus nous conduit à approximer la loi de X par une loi de Poisson de paramètre :

$$\alpha = np = 500 \times \frac{1}{365}.$$

Voici une comparaison numérique pour les petites valeurs de k :

k	0	1	2	3	4	5
$P(X = k)$	0,253 7	0,348 4	0,238 8	0,108 9	0,037 2	0,010 1
$\frac{e^{-\alpha} \alpha^k}{k!}$	0,254 1	0,348 1	0,238 5	0,108 9	0,037 3	0,010 2

Remarquons que la probabilité d'observer plus de 5 anniversaires un 1^{er} avril, calculée par la loi exacte de X ou par son approximation poissonnienne est inférieure à 0,003.

Comparaison graphique :

Les *diagrammes en bâtons* ci-dessous représentent la loi binomiale $\text{Bin}(n, p)$ et la loi de Poisson approximante $\text{Pois}(\alpha)$ avec $\alpha = np$. Les segments verticaux (les bâtons) du diagramme représentant la loi d'une variable discrète X (à valeurs dans \mathbb{N}) ont une hauteur égale à $P(X = k)$ avec une extrémité inférieure au point d'abscisse k de l'axe horizontal. Pour la lisibilité, on a légèrement décalé vers la gauche les bâtons de la loi de Poisson (en bleu) et vers la droite ceux de la loi binomiale (en rouge). Bien que le diagramme en bâtons de la loi binomiale $\text{Bin}(n, p)$ soit constitué théoriquement de $n + 1$ bâtons (et que celui de la loi de Poisson en ait une infinité), seul un petit nombre de bâtons est visible sur les graphiques, les autres correspondant à des probabilités trop petites¹⁶. L'échelle verticale de chaque figure a été choisie de façon adaptative de façon que l'avant dernière graduation verticale donne la valeur de la plus grande probabilité binomiale. On constate que pour $n = 200$ (figure 3.7), la différence entre les deux diagrammes n'est pratiquement plus discernable *visuellement*.

16. En fait, on s'est contenté d'afficher les probabilités correspondant à k inférieur ou égal à la partie entière supérieure de $2\alpha + 4$. On peut vérifier que la somme des probabilités ainsi négligées est inférieure à 1%, pour chacune des deux lois.

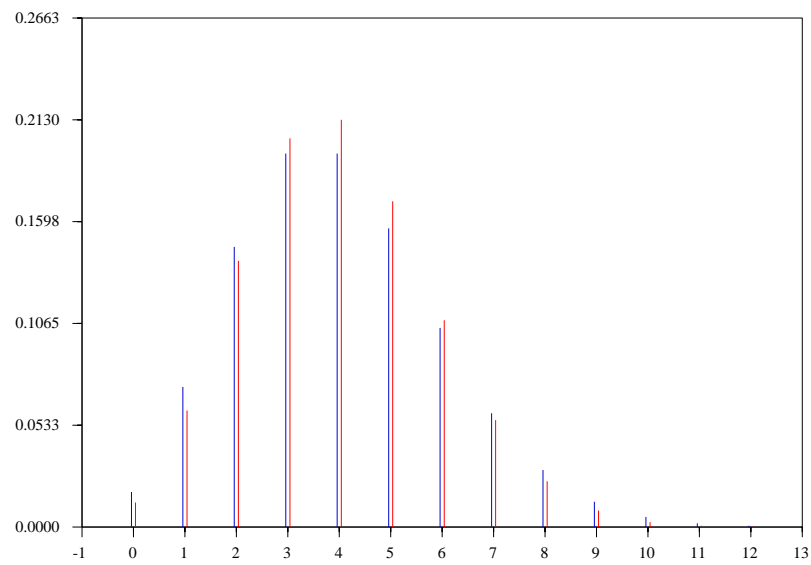


FIG. 3.4 – Lois $\text{Bin}(25; 0,16)$ et $\text{Pois}(4)$

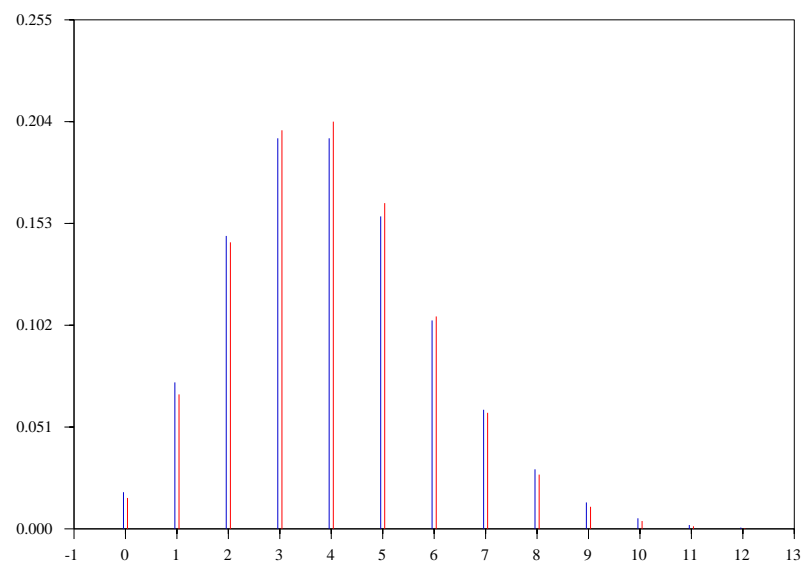


FIG. 3.5 – Lois $\text{Bin}(50; 0,08)$ et $\text{Pois}(4)$

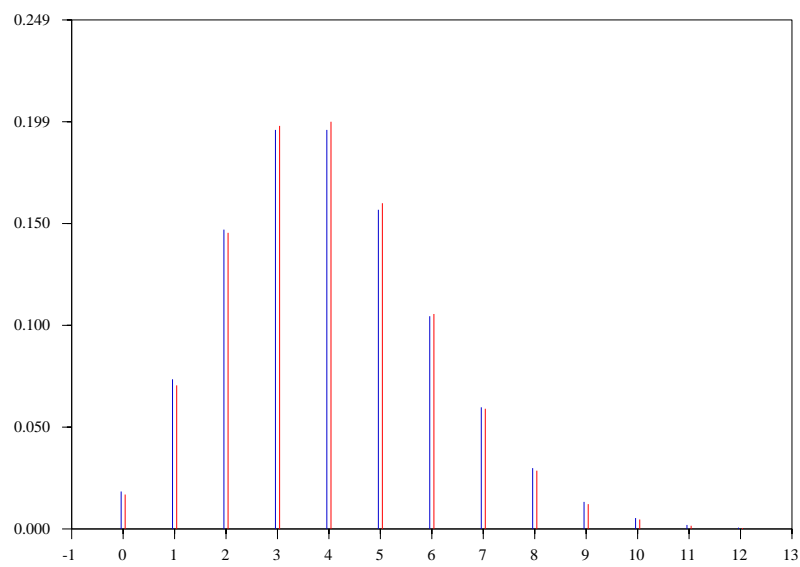


FIG. 3.6 – Lois Bin(100;0,04) et Pois(4)

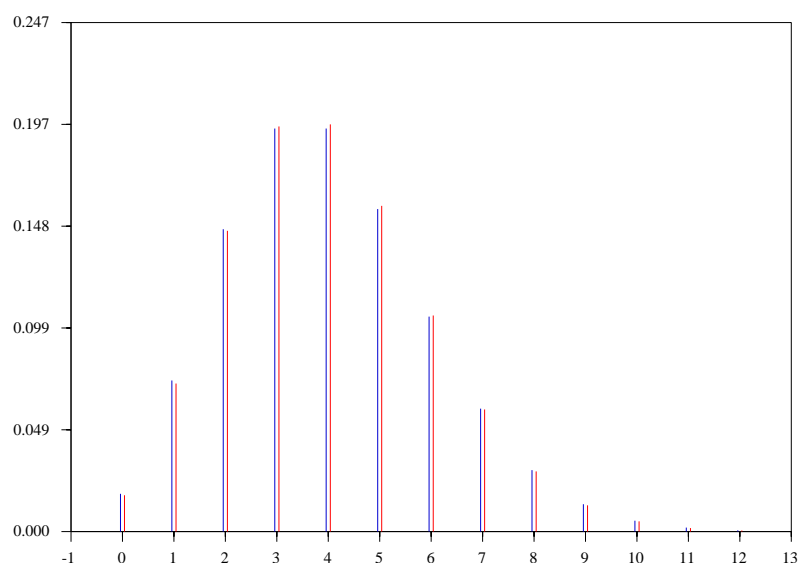


FIG. 3.7 – Lois Bin(200; 0,02) et Pois(4)

3.3.7 Sur le caractère universel de la loi de Poisson

L'étude qui suit a pour but de mieux faire saisir l'importance de la loi de Poisson, en justifiant au passage le bien fondé de l'hypothèse (3.37) du théorème de convergence de la loi binomiale vers la loi de Poisson.

Considérons un phénomène se traduisant par des observations (ou réalisations) aléatoires pendant un intervalle de temps $[0, 1[$ (exemples : désintégrations d'atomes, accidents d'avion, faux numéros de téléphone sur un standard, éruptions volcaniques, naissances de triplés, ...). On suppose que le phénomène vérifie les hypothèses suivantes :

- (a) Les observations dans des intervalles de temps disjoints sont indépendantes.
- (b) Pour tout réel t tel que $0 \leq t < t + T \leq 1$ la loi du nombre (aléatoire) d'observations dans l'intervalle $[t, t + T[$ ne dépend que de la durée T de cet intervalle.

Partageons l'intervalle de temps $[0, 1[$ en n intervalles disjoints

$$I_{n,k} = \left[\frac{k}{n}, \frac{k+1}{n} \right], \quad 0 \leq k < n.$$

Notons :

$$\begin{aligned} p_n &= P(\text{avoir exactement une observation dans } I_{n,k}) \\ r_n &= P(\text{avoir au moins une observation dans } I_{n,k}). \end{aligned}$$

D'après (b), p_n et r_n ne dépendent pas de k . En écrivant de deux façons la probabilité de n'avoir aucune observation dans $[0, 1[$ on obtient :

$$\begin{aligned} 1 - r_1 &= P\left(\bigcap_{k=0}^{n-1} \left\{ \text{aucune observation dans } \left[\frac{k}{n}, \frac{k+1}{n}\right[\right\}\right) \\ &= (1 - r_n)^n \quad \text{en utilisant (a) et (b).} \end{aligned}$$

D'où $1 - r_n = (1 - r_1)^{1/n}$. Un développement limité à l'ordre 1 de cette expression nous permet d'écrire :

$$(1 - r_1)^{1/n} = \exp\left(\frac{1}{n} \ln(1 - r_1)\right) = 1 + \frac{1}{n} \ln(1 - r_1) + \frac{\delta_n}{n}, \quad \text{où } \lim_{n \rightarrow +\infty} \delta_n = 0.$$

Nous noterons désormais :

$$-\ln(1 - r_1) = \alpha, \quad \alpha \in]0, +\infty[. \quad (3.43)$$

Il vient $r_n = \frac{\alpha}{n} - \frac{\delta_n}{n}$ d'où $\lim_{n \rightarrow +\infty} n r_n = \alpha$.

Pour le type de phénomène que nous envisageons, il est vraisemblable que l'on arrive à isoler les observations lorsque les intervalles de la subdivision sont assez petits : asymptotiquement, la probabilité d'avoir plus d'une observation dans $[k/n, (k+1)/n[$ est négligeable devant celle d'en avoir exactement une. Plus précisément, nous rajoutons à notre modèle l'hypothèse :

$$(c) \quad \varepsilon_n = \frac{r_n - p_n}{p_n} \longrightarrow 0, \text{ quand } n \rightarrow +\infty.$$

D'après (c), r_n/p_n converge vers 1, d'où $\lim_{n \rightarrow +\infty} n p_n = \alpha$.

Cherchons maintenant la probabilité d'avoir exactement l observations dans $[0, 1[$. Cette probabilité peut se décomposer en :

$$P(l \text{ observations dans } [0, 1[) = P(A_n) + P(B_n), \quad (3.44)$$

où

$$\begin{aligned} A_n &= \{l \text{ observations avec au plus une dans chaque } I_{n,k}\}, \\ B_n &= \{l \text{ observations avec au moins un } I_{n,k} \text{ en contenant plusieurs}\}. \end{aligned}$$

Calcul de $P(A_n)$: Notons

$$\begin{aligned} D_i &= \left\{ \text{exactement une observation dans } \left[\frac{i}{n}, \frac{i+1}{n}\right[\right\}, \\ E_i &= \left\{ \text{aucune observation dans } \left[\frac{i}{n}, \frac{i+1}{n}\right[\right\}, \quad 0 \leq i < n. \end{aligned}$$

L'événement A_n est la réunion disjointe de tous les événements du type :

$$\left(\bigcap_{i \in I} D_i\right) \cap \left(\bigcap_{j \in J} E_j\right),$$

où $I \subset \{1, \dots, n\}$, $\text{card } I = l$ et $J = \{1, \dots, n\} \setminus I$. D'après l'hypothèse d'indépendance (a), la probabilité de chacun de ces événements est $p_n^l(1 - r_n)^{n-l}$ d'où :

$$P(A_n) = C_n^l p_n^l (1 - r_n)^{n-l}.$$

Pour trouver la limite de $P(A_n)$ lorsque n tend vers l'infini, l restant fixé, il suffit d'adapter la preuve du théorème 3.38 : ici nous avons à trouver la limite de $(1 - r_n)^{n-l}$ au lieu de $(1 - p_n)^{n-l}$. Or

$$(n - l) \ln(1 - r_n) \sim -nr_n \sim -\alpha,$$

d'où $\lim_{n \rightarrow +\infty} (1 - r_n)^{n-l} = e^{-\alpha}$. On en déduit :

$$\lim_{n \rightarrow +\infty} P(A_n) = \frac{e^{-\alpha} \alpha^l}{l!}. \quad (3.45)$$

Majoration de $P(B_n)$: Le calcul de $P(B_n)$ étant trop compliqué, nous nous contenterons d'une majoration. La réalisation de l'événement B_n implique l'existence d'au moins deux observations dans au moins l'un des intervalles de longueur $1/n$. Autrement dit :

$$B_n \subset \bigcup_{k=0}^{n-1} \left\{ \text{au moins deux observations dans } \left[\frac{k}{n}, \frac{k+1}{n} \right] \right\}.$$

Par conséquent

$$\begin{aligned} P(B_n) &\leq P\left(\bigcup_{k=0}^{n-1} \left\{ \text{au moins deux observations dans } \left[\frac{k}{n}, \frac{k+1}{n} \right] \right\}\right) \\ &\leq \sum_{k=0}^{n-1} (r_n - p_n) = n(r_n - p_n) = np_n \varepsilon_n. \end{aligned}$$

D'après (c) et la convergence de np_n vers α , $np_n \varepsilon_n$ tend vers 0 quand n tend vers $+\infty$. Il en est donc de même pour $P(B_n)$.

Pour conclure, on remarque que (3.44) est vérifiée pour tout entier $n \geq 1$ et que le premier membre de cette égalité ne dépend pas de n . Cette égalité reste donc vraie à la limite :

$$P(l \text{ observations dans } [0, 1]) = \lim_{n \rightarrow +\infty} (P(A_n) + P(B_n)) = \frac{e^{-\alpha} \alpha^l}{l!},$$

d'après (3.45) et la majoration de $P(B_n)$. Ce résultat étant valable pour tout entier l , nous avons donc démontré :

Théorème 3.40. *Soit un phénomène donnant lieu à des observations aléatoires vérifiant les hypothèses :*

- (a) *Les observations dans des intervalles de temps disjoints sont indépendantes*
- (b) *Pour tout réel t tel que $0 \leq t < t + T \leq 1$ la loi du nombre (aléatoire) d'observations dans l'intervalle $[t, t + T[$ ne dépend que de la durée T de cet intervalle.*

- (c) En notant p_n la probabilité d'avoir exactement une observation dans un intervalle de temps de durée $1/n$ et r_n celle d'en avoir au moins une, $\varepsilon_n = \frac{r_n - p_n}{p_n} \longrightarrow 0$, quand $n \rightarrow +\infty$.

Alors le nombre aléatoire d'observations dans l'intervalle $[0, 1[$ suit la loi de Poisson de paramètre α défini par

$$\alpha = -\ln(1 - r_1).$$

Remarque 3.41. L'examen attentif de la démonstration ci-dessus montre que la structure d'ordre de l'intervalle $[0, 1[$ n'y joue aucun rôle. L'important est la possibilité de réaliser une partition de $[0, 1[$ en intervalles de même longueur tendant vers 0. Par conséquent en remplaçant la longueur par l'aire ou le volume, il est possible d'obtenir une version spatiale en dimension 2 ou 3 du théorème 3.40. Ceci permet de comprendre pourquoi la loi de Poisson fournit une bonne modélisation par exemple du nombre d'erreurs typographiques dans une page imprimée, du nombre d'impacts de météorites sur un territoire donné, du nombre d'accidents sur une portion d'autoroute pendant une période donnée, du nombre de raisins dans une portion de cake, du nombre d'étoiles dans une région de l'univers, ...

3.4 Lois à densité classiques

3.4.1 Lois uniformes

Définition 3.42. La variable aléatoire réelle X suit la loi uniforme sur l'intervalle $[a, b]$ ($-\infty < a < b < +\infty$) si

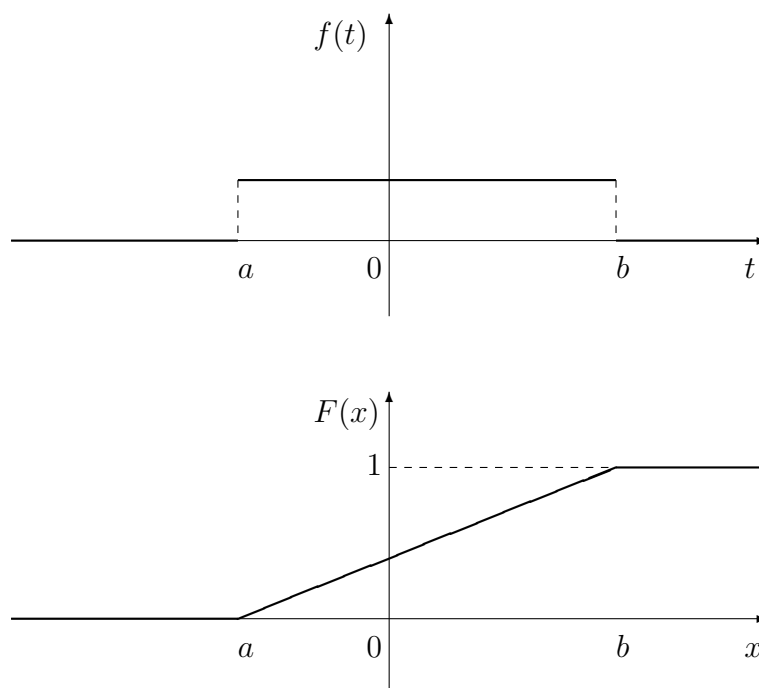
$$\forall B \in \text{Bor}(\mathbb{R}), \quad P(X \in B) = P_X(B) = \frac{\lambda_1([a, b] \cap B)}{\lambda_1([a, b])}, \quad (3.46)$$

où λ_1 désigne la mesure de Lebesgue sur \mathbb{R} (en particulier $\lambda_1([a, b]) = b - a$). Notation : $X \sim \text{Unif}[a, b]$.

Calculons la fonction de répartition F en prenant $B =]-\infty, x]$ pour x quelconque dans (3.46).

$$F(x) = P_X(]-\infty, x]) = \frac{\lambda_1([a, b] \cap]-\infty, x])}{\lambda_1([a, b])} = \begin{cases} 0 & \text{si } -\infty < x < a; \\ \frac{x-a}{b-a} & \text{si } a \leq x < b; \\ 1 & \text{si } b \leq x < +\infty. \end{cases}$$

La fonction de répartition F est affine par morceaux, donc aussi C^1 par morceaux au sens de la proposition 3.28, avec dérivabilité sur $\mathbb{R} \setminus \{a, b\}$ (figure 3.8). La loi a donc une densité f qui s'obtient par dérivation de F , ce qui nous donne $f(t) = 0$ si $t < a$, $f(t) = \frac{1}{b-a}$ si $a < t < b$ et $f(t) = 0$ si $t > b$. On complète la définition de f en la


 FIG. 3.8 – f.d.r. F et densité f de la loi $\text{Unif}[a, b]$

prolongeant en a et b , par exemple en posant $f(a) = f(b) = \frac{1}{b-a}$. La loi uniforme sur $[a, b]$ admet donc pour densité

$$f = \frac{1}{b-a} \mathbf{1}_{[a,b]}.$$

Dans les calculs faisant intervenir la loi uniforme sur $[a, b]$, il est vivement conseillé d'utiliser chaque fois que c'est possible la formule (3.46) de préférence aux calculs d'intégrales de f .

Remarque 3.43. Comme $\lambda_1(\{a\}) = \lambda_1(\{b\}) = 0$, la loi uniforme sur $[a, b]$ est aussi la loi uniforme sur $]a, b]$, $[a, b[$ ou $]a, b[$.

Une des raisons de l'importance de la loi uniforme sur $[0, 1]$ est le théorème suivant.

Théorème 3.44. *Si X est une variable aléatoire réelle de fonction de répartition continue strictement croissante F et si U est une variable aléatoire de loi uniforme sur $[0, 1]$, alors la variable aléatoire $Y := F^{-1}(U)$ a même loi que X .*

Rappelons qu'avoir même loi que X ne signifie aucunement être égale à X . Ce théorème permet de réduire la simulation informatique de la loi de X à celle de U . Nous verrons ultérieurement que ce résultat s'étend à toutes les fonctions de répartition, sans hypothèse de continuité ni de croissance stricte (avec une définition adaptée de F^{-1}).

Preuve. Comme F est continue strictement croissante, c'est une *bijection* de \mathbb{R} sur son image $]0, 1[$ (en raison de la stricte monotonie de F , les bornes 0 et 1 ne sont pas

atteintes). Par conséquent $F^{-1} :]0, 1[\rightarrow \mathbb{R}$ est bien définie et vérifie :

$$\forall u \in]0, 1[, \forall x \in \mathbb{R}, \quad F^{-1}(u) \leq x \text{ si et seulement si } u \leq F(x).$$

Comme $P(0 < U < 1) = 1$, on en déduit que les événements $\{F^{-1}(U) \leq x\}$ et $\{U \leq F(x)\}$ ont même probabilité. Pour obtenir la fonction de répartition de Y , on remarque alors que pour tout $x \in \mathbb{R}$,

$$P(Y \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = \frac{\lambda_1([0, F(x)])}{\lambda_1([0, 1])} = F(x).$$

Ainsi Y a pour fonction de répartition F donc a même loi que X . \square

3.4.2 Lois exponentielles

Définition 3.45. Soit a un réel strictement positif. La variable aléatoire réelle X suit la loi exponentielle de paramètre a si elle admet pour densité

$$f(t) = ae^{-at} \mathbf{1}_{[0, +\infty[}(t).$$

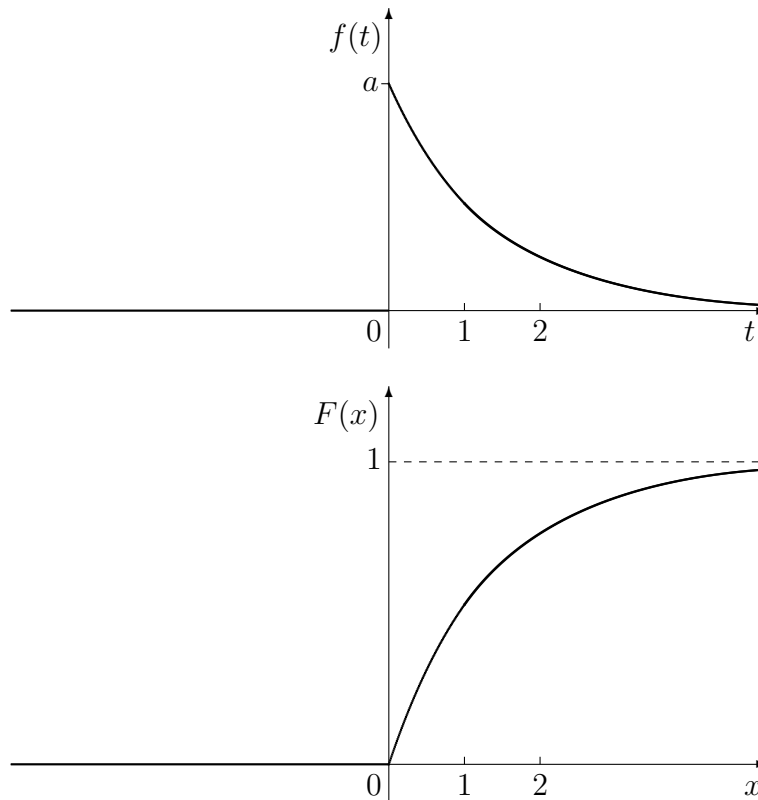


FIG. 3.9 – Densité et f.d.r. de la loi $\text{Exp}(a)$

En pratique, plutôt que de travailler avec la fonction de répartition d'une loi exponentielle, il est plus commode d'utiliser la *fonction de survie* G :

$$G(x) = P(X > x) = 1 - F(x) = \begin{cases} 1 & \text{si } x \leq 0, \\ e^{-ax} & \text{si } x > 0. \end{cases}$$

Les lois exponentielles sont souvent choisies pour modéliser des temps d'attente : temps d'attente à partir de maintenant du prochain tremblement de terre, du prochain faux numéro sur une ligne téléphonique, de la prochaine désintégration d'un atome de radium, etc.

La raison de ce choix est la propriété *d'absence de mémoire* en temps continu qui caractérise la famille des lois exponentielles.

Théorème 3.46 (absence de mémoire).

i) Si la variable aléatoire X suit une loi exponentielle, alors elle vérifie la propriété d'absence de mémoire :

$$\forall s \in \mathbb{R}^+, \forall t \in \mathbb{R}^+, \quad P(X > t + s \mid X > t) = P(X > s). \quad (3.47)$$

ii) Réciproquement si une variable aléatoire X vérifie (3.47), alors elle suit une loi exponentielle.

Comme la fonction de survie caractérise la loi, (3.47) signifie que la loi de $(X - t)$ conditionnelle à $\{X > t\}$ est la même que la loi de X .

En préliminaire à la preuve du théorème, remarquons que la probabilité conditionnelle dans (3.47) s'exprime commodément à l'aide de la fonction de survie G de la variable aléatoire X , définie par $G(x) := P(X > x)$. En effet, s étant positif, on a $t + s \geq t$ d'où l'inclusion de $\{X > t + s\}$ dans $\{X > t\}$ et l'égalité d'événements : $\{X > t + s\} \cap \{X > t\} = \{X > t + s\}$. On en déduit

$$P(X > t + s \mid X > t) = \frac{P(X > t + s)}{P(X > t)} = \frac{G(t + s)}{G(t)}. \quad (3.48)$$

Preuve de i). Si X suit la loi exponentielle de paramètre a , on a $G(x) = e^{-ax}$ pour tout x positif et (3.48) se traduit alors par :

$$P(X > t + s \mid X > t) = \frac{e^{-a(t+s)}}{e^{-at}} = e^{-as} = P(X > s).$$

Ainsi X de loi $\text{Exp}(a)$ vérifie la propriété d'absence de mémoire (3.47). □

Preuve de ii). Soit X une variable aléatoire dont la loi vérifie (3.47) et G sa fonction de survie. Comme $G = 1 - F$ (où F désigne la fonction de répartition de X), G est décroissante et continue à droite et tend vers 0 en $+\infty$. De plus l'écriture de (3.47) suppose implicitement que $G(t) > 0$ pour tout $t \geq 0$ car sinon $P(\cdot \mid X > t)$ ne serait

pas définie. Grâce à (3.48), on voit que la propriété d'absence de mémoire (3.47) équivaut à

$$\forall s \in \mathbb{R}^+, \forall t \in \mathbb{R}^+, \quad \frac{G(t+s)}{G(t)} = G(s).$$

La fonction de survie G doit donc être une solution décroissante, continue à droite, tendant vers 0 en $+\infty$ et telle que $0 < G(t) \leq 1$ de l'équation fonctionnelle¹⁷ :

$$\forall s \in \mathbb{R}^+, \forall t \in \mathbb{R}^+, \quad G(t+s) = G(t)G(s). \quad (3.49)$$

En faisant $s = t = 0$ dans (3.49), on obtient $G(0) = G(0)^2$ et comme $G(0) > 0$, on a

$$G(0) = 1. \quad (3.50)$$

En faisant $s = t$ dans (3.49), on obtient $G(2t) = G(t)^2$, puis de proche en proche

$$\forall n \in \mathbb{N}^*, \forall t \geq 0, \quad G(nt) = G(t)^n. \quad (3.51)$$

En particulier pour $t = 1/d$, $d \in \mathbb{N}^*$:

$$\forall n \in \mathbb{N}^*, \forall d \in \mathbb{N}^*, \quad G\left(\frac{n}{d}\right) = G\left(\frac{1}{d}\right)^n. \quad (3.52)$$

Lorsque $n = d$, (3.52) donne $G(1) = G(1/d)^d$ d'où

$$\forall d \in \mathbb{N}^*, \quad G\left(\frac{1}{d}\right) = G(1)^{1/d}. \quad (3.53)$$

Nous connaissons maintenant G sur l'ensemble des rationnels positifs puisque (3.50), (3.51), (3.52) et (3.53) nous donnent

$$\forall r \in \mathbb{Q}^+, \quad G(r) = G(1)^r. \quad (3.54)$$

Soit $x \in \mathbb{R}^+ \setminus \mathbb{Q}^+$, x est limite d'une suite décroissante (r_n) de rationnels. Comme G est continue à droite, $G(r_n)$ converge vers $G(x)$. D'autre part l'application $y \mapsto G(1)^y$ est continue sur \mathbb{R} . Ainsi en appliquant (3.54) à r_n et en faisant tendre n vers l'infini on obtient

$$\forall x \in \mathbb{R}^+, \quad G(x) = G(1)^x. \quad (3.55)$$

A priori la constante $G(1)$ est dans $]0, 1]$. On peut écarter la valeur $G(1) = 1$ car sinon d'après (3.55), la limite en $+\infty$ de G serait 1 alors qu'elle vaut 0.

Finalement, puisque $0 < G(1) < 1$, on peut poser $G(1) = e^{-a}$ pour un réel $a > 0$ (cela revient à prendre $a = -\ln G(1)$). On peut alors réécrire (3.55) sous la forme

$$\forall x \in \mathbb{R}^+, \quad G(x) = e^{-ax}.$$

La fonction de survie G est donc la même que celle de la loi exponentielle de paramètre a , donc X suit cette loi (puisque la fonction de survie caractérise la loi au même titre que la fonction de répartition). \square

17. Une équation fonctionnelle est une équation dont l'inconnue est...une fonction! Les équations différentielles sont des exemples bien connus d'équations fonctionnelles.

3.4.3 Lois gaussiennes

Ces lois jouent un rôle capital dans l'étude des lois limites de sommes de variables aléatoires indépendantes. Par exemple (théorème de de Moivre Laplace) si S_n suit la loi $\text{Bin}(n, p)$, alors pour tout $x \in \mathbb{R}$, $P(S_n - np \leq x\sqrt{np(1-p)})$ converge quand n tend vers l'infini vers $\Phi(x)$, où Φ est la f.d.r. de la loi gaussienne $\mathfrak{N}(0, 1)$.

Définition 3.47. On dit que la variable aléatoire X suit la loi gaussienne ou normale $\mathfrak{N}(m, \sigma)$ si elle a pour densité la fonction :

$$f_{m,\sigma} : \mathbb{R} \longrightarrow \mathbb{R}^+ \quad t \longmapsto \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right).$$

La loi $\mathfrak{N}(0, 1)$ est appelée loi normale standard.

Tous les calculs de probabilités concernant une variable aléatoire de loi $\mathfrak{N}(m, \sigma)$ peuvent se ramener à des calculs sur une variable de loi normale standard.

Proposition 3.48. Si la variable aléatoire X suit la loi $\mathfrak{N}(m, \sigma)$, alors $Y := (X - m)/\sigma$ suit la loi $\mathfrak{N}(0, 1)$. Autrement dit, toute v.a. gaussienne X de loi $\mathfrak{N}(m, \sigma)$ peut s'écrire $X = \sigma Y + m$ avec Y de loi $\mathfrak{N}(0, 1)$.

Preuve. On calcule $P(a < Y \leq b)$ pour a et b réels quelconques ($a < b$).

$$\begin{aligned} P\left(a < \frac{X - m}{\sigma} \leq b\right) &= P(\sigma a + m < X \leq \sigma b + m) \\ &= \int_{\sigma a + m}^{\sigma b + m} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - m)^2}{2\sigma^2}\right) dx. \end{aligned}$$

Il suffit alors de faire le changement de variable $y = (x - m)/\sigma$ pour obtenir

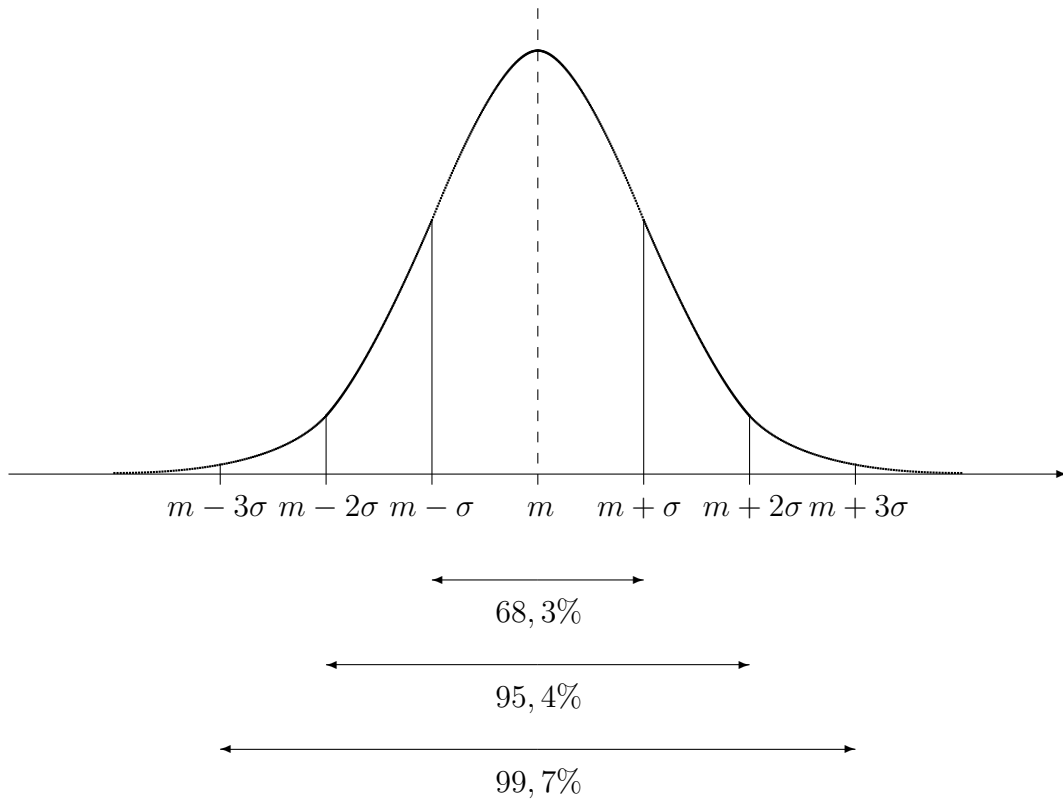
$$\forall a \in \mathbb{R}, \forall b > a, \quad P(a < Y \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy.$$

Donc Y a bien la densité $f_{0,1}$. □

Remarque 3.49. En adaptant le calcul fait dans la preuve de la proposition 3.48, on voit facilement que la famille des lois gaussiennes est *stable par transformations affines* : si X a pour loi $\mathfrak{N}(m, \sigma)$, alors pour tout $(\alpha, \beta) \in \mathbb{R}^* \times \mathbb{R}$, la v.a. $\alpha X + \beta$ est encore gaussienne, de loi $\mathfrak{N}(\alpha m + \beta, |\alpha|\sigma)$.

La figure 3.10 illustre la signification du paramètre de position m et du paramètre de dispersion σ pour la loi gaussienne $\mathfrak{N}(m, \sigma)$. Cette concentration de pratiquement toute la probabilité dans l'intervalle $[m - 3\sigma, m + 3\sigma]$ permet l'utilisation des lois gaussiennes pour modéliser des grandeurs aléatoires qui *a priori* prennent leurs valeurs seulement dans un petit intervalle de \mathbb{R}^+ : taille, poids, ..., même si théoriquement une variable gaussienne peut prendre toute valeur entre $-\infty$ et $+\infty$.

Il n'existe pas d'expression d'une primitive de la densité gaussienne $f_{m,\sigma}$ à l'aide des fonctions usuelles. Les valeurs de la fonction de répartition Φ de $\mathfrak{N}(0, 1)$ sont tabulées, cf. page 269. D'après la proposition 3.48, ceci suffit pour calculer numériquement n'importe quelle f.d.r. de loi gaussienne.

FIG. 3.10 – Concentration de la loi $\mathfrak{N}(m, \sigma)$ autour de m

3.4.4 Lois de Cauchy

Définition 3.50. La variable aléatoire X suit la loi de Cauchy (ou loi de Cauchy de paramètres 0 et 1) si elle admet pour densité :

$$f(t) = \frac{1}{\pi(1+t^2)}.$$

Notation : $X \sim \text{Cau}(0, 1)$.

Cette loi est *symétrique*, ce qui signifie que X et $-X$ ont même loi, ceci résultant ici de la parité de f . La fonction de répartition F est donnée par :

$$F(x) = \int_{-\infty}^x \frac{dt}{\pi(1+t^2)} = \frac{1}{\pi} \left(\frac{\pi}{2} + \arctan x \right),$$

où $\arctan x$ est l'unique réel $y \in]-\pi/2, \pi/2[$ tel que $\tan y = x$.

Si $Y = a + bX$, avec X de loi $\text{Cau}(0, 1)$, $a \in \mathbb{R}$ et $b \in \mathbb{R}_+^*$, on dit encore que Y suit une loi de Cauchy, de paramètres (a, b) , notation $Y \sim \text{Cau}(a, b)$. La densité est alors

$$f_{a,b}(t) = \frac{1}{\pi b} \frac{1}{1 + \left(\frac{t-a}{b}\right)^2}.$$

Chapitre 4

Espérance

4.1 Introduction

L'espérance d'une variable aléatoire est, lorsqu'elle existe, la *moyenne des valeurs de cette variable, pondérées par leurs probabilités de réalisation*. On voit bien comment traduire cette définition informelle dans le cas d'une variable aléatoire discrète X en posant :

$$\mathbf{E}X := \sum_{x \in X(\Omega)} xP(X = x). \quad (4.1)$$

Cette formule n'a de sens que si la famille de réels $\{xP(X = x); x \in X(\Omega)\}$ est sommable, ce qui se traduit par la condition suivante pour l'existence de l'espérance de la v.a. discrète X :

$$\sum_{x \in X(\Omega)} |x|P(X = x) < +\infty. \quad (4.2)$$

Tant que l'on reste dans le cadre des variables aléatoires discrètes, cette définition est satisfaisante et permet d'établir toutes les propriétés de l'espérance (cf. cours de Deug *Introduction au Calcul des Probabilités*, Chap. 5). En bonne place parmi ces propriétés figure *l'additivité* de l'espérance, *i.e.* si X et Y définies sur le même (Ω, \mathcal{F}, P) ont une espérance, il en va de même pour $X + Y$ et on a

$$\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y. \quad (4.3)$$

Essayons de traduire la définition informelle ci-dessus dans le cas d'une variable aléatoire à densité f . On part de (4.1) et on remplace $P(X = x)$ par $P(X \in [x, x + dx])$, probabilité « valant ¹ $f(x) dx$ » et on remplace la somme (ou série) par une intégrale, ce qui conduit à :

$$\mathbf{E}X := \int_{-\infty}^{+\infty} xf(x) dx, \quad (4.4)$$

1. Nous ne prétendons pas donner un sens rigoureux à cette probabilité d'appartenance à un « intervalle infinitésimal », il s'agit juste d'une approche intuitive.

la condition d'existence de l'espérance étant tout simplement la convergence absolue de cette intégrale généralisée, ce qui vu la positivité de f , se traduit par :

$$\int_{-\infty}^{+\infty} |x|f(x) dx < +\infty. \quad (4.5)$$

Cette définition malgré son analogie formelle avec (4.1) est loin d'offrir la même souplesse pour établir les propriétés de l'espérance. Par exemple la preuve de l'additivité est complètement hors de portée. En effet, si X et Y sont à densité, $X + Y$ n'est pas forcément à densité² et alors le premier membre de (4.3) n'est même pas défini pour la v.a. $Z = X + Y$.

La solution donnée à ce problème par la théorie moderne des probabilités est la définition dans le cas général, de l'espérance de X comme une intégrale abstraite sur Ω , relativement à la mesure P :

$$\mathbf{E}X := \int_{\Omega} X(\omega) dP(\omega), \quad \text{si } \int_{\Omega} |X(\omega)| dP(\omega) < +\infty. \quad (4.6)$$

On peut donner une première idée de ce qu'est cette intégrale abstraite en considérant le cas d'une variable aléatoire X telle que $X(\Omega) = \{x_1, \dots, x_n\}$. Alors en notant $A_k := \{X = x_k\} = \{\omega \in \Omega; X(\omega) = x_k\}$, on a :

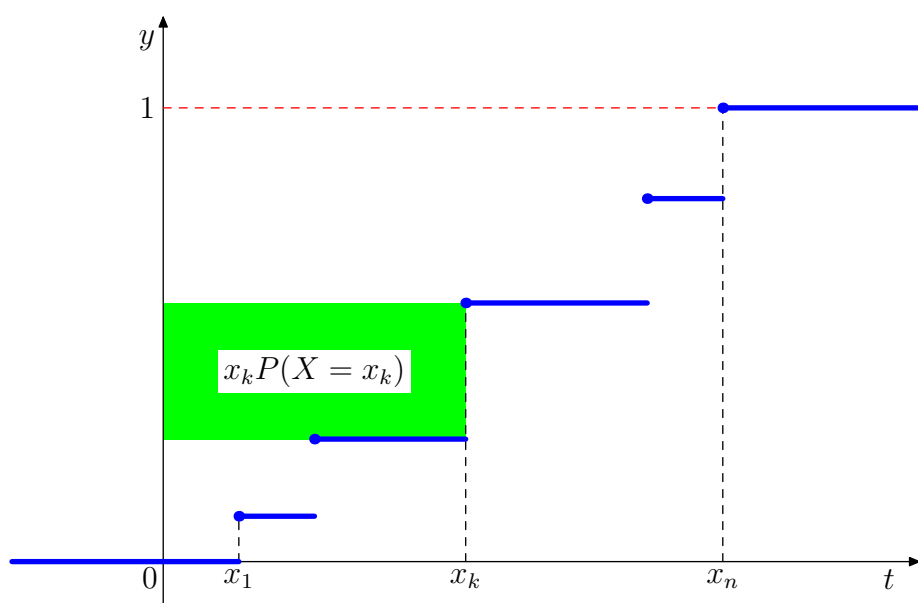
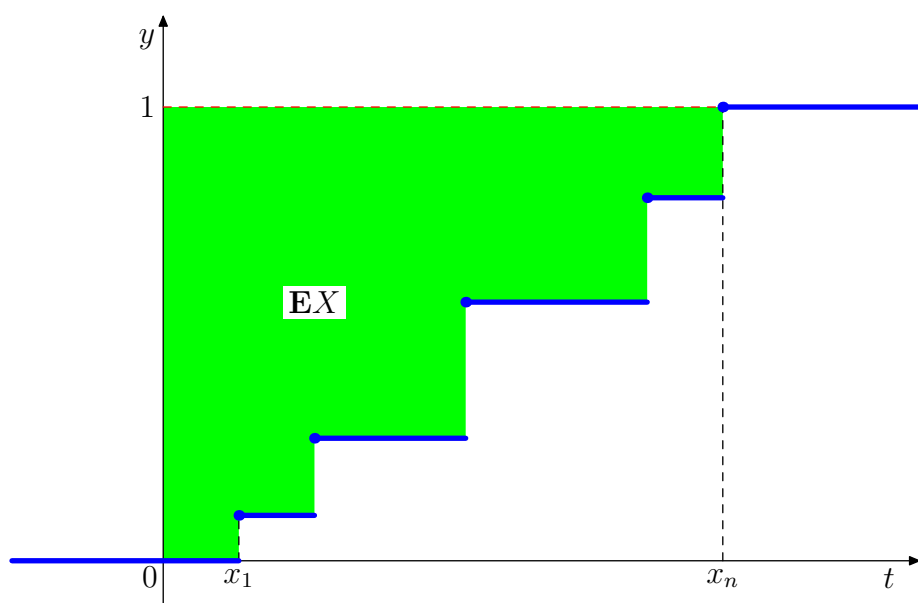
$$\int_{\Omega} X(\omega) dP(\omega) = \sum_{k=1}^n x_k P(A_k), \quad (4.7)$$

ce qui traduit bien la définition informelle de $\mathbf{E}X$ comme la moyenne des valeurs de X pondérées par leurs probabilités de réalisation. Le passage au cas d'une variable aléatoire X quelconque revient précisément à construire une intégrale au sens de Lebesgue sur (Ω, \mathcal{F}, P) et cette théorie sort du cadre de notre programme.

Il nous faut donc trouver une autre définition de $\mathbf{E}X$. Cette définition doit permettre un traitement unifié de toutes les lois³. Rappelons qu'il existe des lois qui ne sont ni discrètes ni à densité et que la description la plus générale des lois de variables aléatoires réelles est donnée par leur fonction de répartition, cf. le théorème 2.30 et la remarque 3.16. Il est donc naturel de chercher à définir $\mathbf{E}X$ à partir de la fonction de répartition $F : t \mapsto P(X \leq t)$. Nous allons motiver cette définition en nous restreignant au cas des variables aléatoires positives et en partant du cas simple où X est discrète avec $X(\Omega) = \{x_1, \dots, x_n\}$ partie finie de \mathbb{R}_+ . Dans ce cas, la définition informelle de $\mathbf{E}X$ se traduit par la formule $\mathbf{E}X = \sum_{k=1}^n x_k P(X = x_k)$. Les figures 4.1 et 4.2 nous montrent comment exprimer cette moyenne pondérée à l'aide de F . Rappelons que dans ce cas, F présente en chaque x_k un saut d'amplitude $P(X = x_k)$. L'interprétation graphique en terme d'aires donnée par la figure 4.2 nous permet d'écrire $\mathbf{E}X$ comme l'intégrale de Riemann ordinaire : $\mathbf{E}X = \int_0^{x_n} (1 - F(t)) dt$ et aussi comme la fausse intégrale généralisée $\int_0^{+\infty} (1 - F(t)) dt$.

2. Alors que la somme de deux variables aléatoires discrètes est *toujours* une variable aléatoire discrète.

3. La définition informelle de $\mathbf{E}X$ nous fait pressentir que $\mathbf{E}X$ ne doit dépendre que de la loi de X , ce qui est bien le cas dans les formules (4.1) et (4.4).

FIG. 4.1 – Interprétation graphique des $x_k P(X = x_k)$, pour $x_k \geq 0$ FIG. 4.2 – Interprétation graphique de $\mathbf{E}X = \sum_{k=1}^n x_k P(X = x_k)$, les $x_k \geq 0$.

Si on passe maintenant au cas d'une variable aléatoire positive quelconque, il paraît alors naturel de considérer que $\mathbf{E}X$ est l'aire (éventuellement infinie) délimitée par le segment vertical $t = 0, y \in [0, 1]$, la demi droite « asymptote » $y = 1, t \geq 0$ et le graphe de F , ce qui nous conduit à la formule

$$\mathbf{E}X := \int_0^{+\infty} (1 - F(t)) dt = \int_0^{+\infty} P(X > t) dt, \quad \text{pour toute v.a. positive } X.$$

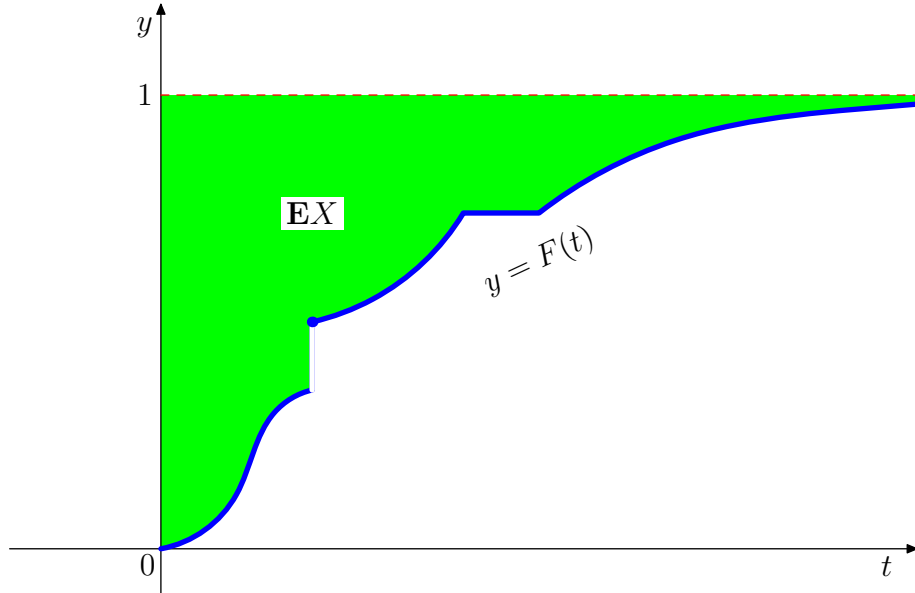


FIG. 4.3 – Interprétation graphique de $\mathbf{E}X$ via la f.d.r. de X v.a. positive.

Nous verrons que cette définition permet d'établir en toute généralité les propriétés de l'espérance. Bien sûr nous devons retrouver à partir de cette définition, les formules (4.1) et (4.4) pour X discrète ou à densité.

4.2 Espérance d'une variable aléatoire positive

Dans toute la suite de ce chapitre, on fixe un espace probabilisé (Ω, \mathcal{F}, P) . Toutes les variables aléatoires considérées seront, sauf mention explicite du contraire, définies sur cet espace et leur loi sera la loi sous P .

Définition 4.1 (espérance d'une v.a. positive). *Soit X une variable aléatoire positive⁴ sur (Ω, \mathcal{F}) . On appelle espérance de X (ou espérance de X sous P) la quantité*

$$\mathbf{E}X := \int_0^{+\infty} P(X > t) dt, \quad (4.8)$$

qui est un élément de $\overline{\mathbb{R}}_+$.

4. C'est-à-dire une application $\Omega \rightarrow \mathbb{R}_+$, mesurable \mathcal{F} - $\text{Bor}(\mathbb{R}_+)$.

Pour justifier l'existence de $\mathbf{E}X$, on commence par noter que l'application $G : \mathbb{R}_+ \rightarrow [0, 1]$, $t \mapsto G(t) := P(X > t)$ est *décroissante* sur \mathbb{R}_+ , donc *Riemann intégrable* sur $[0, b]$ pour tout $b \in \mathbb{R}_+^*$, cf. proposition A.9. L'intégrale $\int_0^b G(t) dt = \int_0^b P(X > t) dt$ existe donc bien et est un réel positif pour tout b . Comme c'est une fonction croissante de sa borne supérieure b , elle converge dans $\overline{\mathbb{R}}_+$ quand b tend vers $+\infty$.

Dans cette section, nous utiliserons l'interprétation graphique de $\mathbf{E}X$ *via* la fonction de survie $t \mapsto P(X > t)$ (cf. figure 4.4) plutôt que *via* la f.d.r. $F : t \mapsto P(X \leq t)$. On passe évidemment d'une représentation à l'autre en effectuant une symétrie orthogonale par rapport à la droite $y = 1/2$, puisque $G = 1 - F$. Cette symétrie conserve les aires, cf. prop. 2.14.

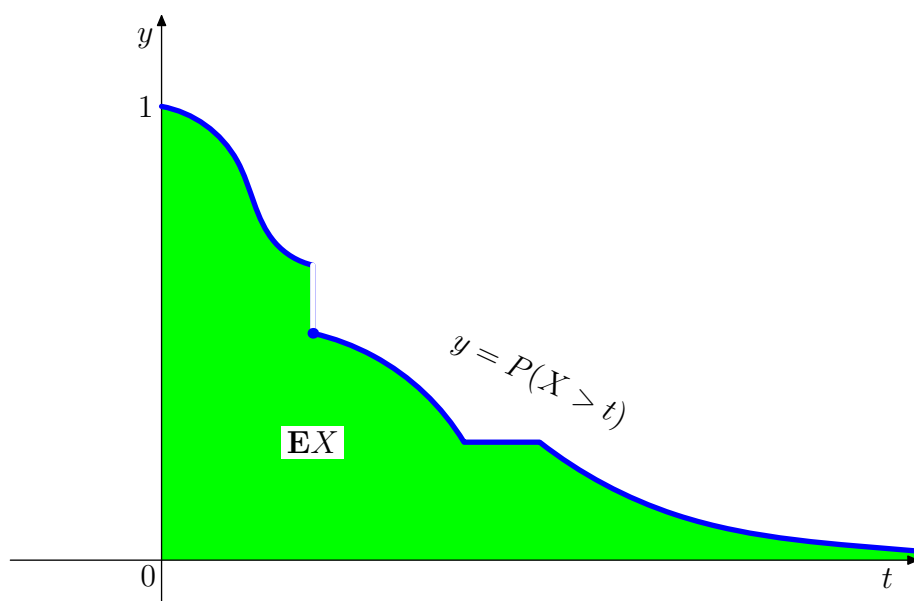


FIG. 4.4 – Interprétation graphique de $\mathbf{E}X$ *via* la fonction de survie de X v.a. positive.

Remarque 4.2. $\mathbf{E}X$ ne dépend que de la loi de X , il serait donc plus correct de parler de l'espérance de la loi de X sous P au lieu de l'espérance de X . L'usage donne néanmoins la préférence à cette dernière appellation *quand il n'y a pas d'ambiguïté sur P* .

Remarque 4.3 (espérance d'une v.a. presque sûrement positive). Dans les exercices, la variable aléatoire X n'est pas toujours donnée explicitement, il arrive assez souvent que l'on ne connaisse que sa loi P_X . Si $P_X(\mathbb{R}_+) = 1$, on s'autorisera une généralisation de la définition 4.1 en considérant que la formule (4.8) reste valable. Il s'agit bien d'une généralisation car on peut avoir $P(X \geq 0) = 1$ sans que $X(\omega)$ soit positif ou nul pour tout ω , par exemple si $\Omega = \mathbb{R}$, $\mathcal{F} = \text{Bor}(\mathbb{R})$, P est la loi uniforme sur $[0, 1]$ et $X : \omega \mapsto \omega$ est l'identité sur \mathbb{R} . On a alors $X(\omega) < 0$ pour une infinité non dénombrable de ω 's et $P(X \geq 0) = 1$. Cette généralisation est cohérente avec la remarque 4.2 car si X est une variable aléatoire réelle définie sur un espace probabilisé (Ω, \mathcal{F}, P) et telle que

$P(X \geq 0) = 1$, on peut toujours trouver un espace probabilisé $(\Omega', \mathcal{F}', P')$ et une variable aléatoire positive X' définie sur cet espace tels que X et X' aient même loi. Il suffit de prendre $\Omega' = \mathbb{R}_+$, $\mathcal{F}' = \text{Bor}(\mathbb{R}_+)$, $P' = P_X$ et $X' : \mathbb{R}_+ \rightarrow \mathbb{R}$, $\omega \mapsto \omega$ égale à l'identité sur \mathbb{R}_+ . On a alors pour tout borélien B de \mathbb{R} , $P'_{X'}(B) = P'(X' \in B) = P'(\{\omega \in \Omega'; X'(\omega) \in B\}) = P'(B \cap \mathbb{R}_+) = P_X(B \cap \mathbb{R}_+) = P_X(B)$ car $P_X(B \cap \mathbb{R}_-) \leq P_X(\mathbb{R}_-) = 0$. Ceci montre que X et X' ont même loi⁵.

Définition 4.4 (intégrabilité d'une v.a. positive). *On dit que la variable aléatoire positive X est intégrable si*

$$\int_0^{+\infty} P(X > t) dt < +\infty. \quad (4.9)$$

Exemple 4.5. Si la variable aléatoire positive X est bornée, i. e. s'il existe une constante c telle que pour tout $\omega \in \Omega$, $0 \leq X(\omega) \leq c$, alors elle est intégrable. En effet pour $t \geq c$, $P(X > t) = 0$, ce qui réduit l'intégrale généralisée définissant $\mathbf{E}X$ à une intégrale de Riemann ordinaire $\int_0^c P(X > t) dt$ donc finie (et majorée par c).

Plus généralement, si la loi de X , v.a. positive, vérifie $P(X > t) \leq Ct^{-\alpha}$ pour un certain $\alpha > 1$ et tout $t \geq t_0 > 0$, ou si $P(X > t) \leq t^{-1}(\ln t)^{-\beta}$ pour un $\beta > 1$ et tout $t \geq t_0 > 0$, alors X est intégrable. Réciproquement, l'intégrabilité de X nous donne un renseignement sur la vitesse de convergence⁶ vers 0 de $P(X > t)$ quand t tend vers $+\infty$. C'est l'inégalité de Markov que nous verrons ci-dessous (proposition 4.16).

Voyons maintenant quelques exemples simples de calcul d'espérance de variables aléatoires positives.

Exemple 4.6 (espérance d'une constante positive). Si la variable aléatoire X est une constante positive c , i.e. $X(\omega) = c$ pour tout $\omega \in \Omega$, alors $\mathbf{E}X = c$. En effet on a clairement :

$$P(X > t) = \begin{cases} 1 & \text{si } t < c \\ 0 & \text{si } t \geq c \end{cases} = \mathbf{1}_{]-\infty, c[}(t),$$

d'où

$$\mathbf{E}X = \int_0^{+\infty} \mathbf{1}_{]-\infty, c[}(t) dt = \int_0^c \mathbf{1}_{[0, c[}(t) dt = \int_0^c 1 dt = c.$$

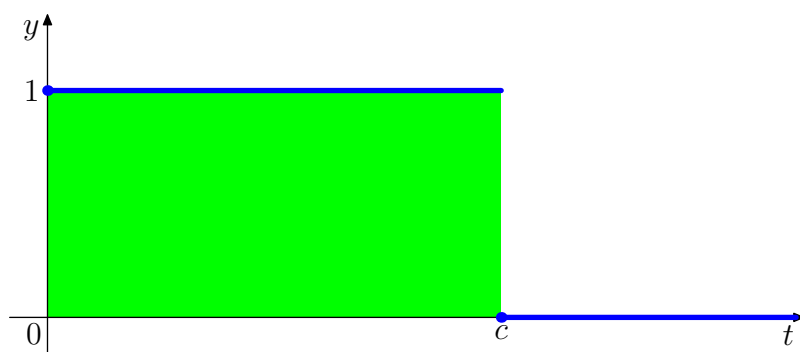
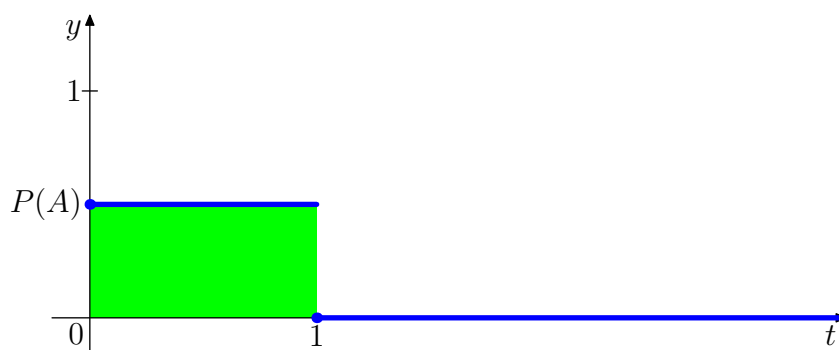
L'exemple suivant est d'une grande importance car il permet d'écrire toute probabilité d'évènement comme une espérance. Nous le formulons sous forme de proposition.

Proposition 4.7 (espérance d'une indicatrice d'évènement). *Pour tout évènement $A \in \mathcal{F}$,*

$$\mathbf{E}(\mathbf{1}_A) = P(A). \quad (4.10)$$

5. La tribu borélienne de \mathbb{R}_+ est la plus petite tribu contenant tous les ouverts de \mathbb{R}_+ . On peut vérifier qu'un sous-ensemble B' de \mathbb{R}_+ est un borélien de \mathbb{R}_+ si et seulement s'il s'écrit $B \cap \mathbb{R}_+$ où B est un borélien de \mathbb{R} .

6. Pour n'importe quelle variable aléatoire X , $P(X > t)$ tend vers 0 quand t tend vers $+\infty$, car $G(t) = 1 - F(t)$, où F est la f.d.r. de X qui tend toujours vers 1 en $+\infty$.


 FIG. 4.5 – Espérance de la v.a. constante $X = c$

 FIG. 4.6 – Espérance de la v.a. indicatrice $X = \mathbf{1}_A$

Preuve. La variable aléatoire positive $\mathbf{1}_A$ prend la valeur 1 sur l'évènement A et 0 sur A^c , elle suit la loi de Bernoulli de paramètre $P(A)$. L'évènement $\{\mathbf{1}_A > t\}$ est donc égal à A si $0 \leq t < 1$ et à l'ensemble vide si $t \geq 1$. On en déduit que :

$$P(\mathbf{1}_A > t) = \begin{cases} P(A) & \text{si } 0 \leq t < 1, \\ 0 & \text{si } t \geq 1. \end{cases}$$

Par conséquent,

$$\mathbf{E}(\mathbf{1}_A) = \int_0^1 P(A) dt = P(A).$$

□

Dans ce chapitre, les variables aléatoires discrètes X ne prenant qu'un nombre fini de valeurs jouent un rôle important car elles vont nous permettre d'établir par passage à la limite les principales propriétés de l'espérance. Il est commode de les dénommer comme suit.

Définition 4.8 (variable aléatoire simple). *On dit que la variable aléatoire réelle X définie sur (Ω, \mathcal{F}) est simple ou étagée si $X(\Omega)$ est fini. En notant $X(\Omega) = \{x_1, \dots, x_n\}$,*

X admet la décomposition

$$X = \sum_{k=1}^n x_k \mathbf{1}_{A_k}, \quad \text{où } A_k := \{X = x_k\}, \quad k = 1, \dots, n, \quad (4.11)$$

les évènements A_k formant une partition de Ω .

Proposition 4.9 (espérance d'une v.a. positive simple). *Si X est une variable aléatoire positive simple avec $X(\Omega) = \{x_1, \dots, x_n\}$,*

$$\mathbf{E}X = \sum_{k=1}^n x_k P(X = x_k). \quad (4.12)$$

On retrouve ainsi la formule (4.1) de l'introduction dans le cas particulier où $X(\Omega)$ est fini ; voir aussi (4.7).

Preuve. Notons en préliminaire qu'il nous faut résister ici à la tentation de dire « c'est immédiat en utilisant la décomposition (4.11), la proposition 4.7 et la linéarité de l'espérance », car nous n'avons pas encore prouvé que l'espérance est linéaire. En fait la proposition 4.9 est l'un des ingrédients de la preuve de la linéarité de l'espérance. Il nous faut donc vérifier (4.12) par un calcul direct basé sur la définition 4.1.

Quitte à réindexer, on peut toujours supposer que les x_k sont rangés par ordre croissant. Notons $p_i := P(X = x_i)$ et $s_k := \sum_{1 \leq i \leq k} p_i$. La fonction de répartition F peut alors s'écrire

$$F(t) = \sum_{k=1}^{n-1} s_k \mathbf{1}_{[x_k, x_{k+1}[}(t) + s_n \mathbf{1}_{[x_n, +\infty[}(t).$$

Notons que pour $t \geq x_n$, $F(t) = s_n = 1$, donc $P(X > t) = 1 - F(t) = 0$. Ainsi $\int_0^{+\infty} P(X > t) dt = \int_0^{x_n} P(X > t) dt$. On peut alors calculer $\mathbf{E}X$ comme suit en utilisant les propriétés de l'intégrale de Riemann sur l'intervalle fermé borné $[0, x_n]$.

$$\begin{aligned} \mathbf{E}X &= \int_0^{x_n} (1 - F(t)) dt = x_n - \int_0^{x_n} F(t) dt = x_n - \sum_{k=1}^{n-1} \int_{x_k}^{x_{k+1}} s_k dt \\ &= x_n - \sum_{k=1}^{n-1} (x_{k+1} - x_k) s_k \\ &= x_n - \sum_{j=2}^n x_j s_{j-1} + \sum_{j=1}^{n-1} x_j s_j \\ &= x_n - x_n s_{n-1} + \sum_{j=2}^{n-1} x_j (s_j - s_{j-1}) + x_1 s_1 \\ &= x_n p_n + \sum_{j=2}^{n-1} x_j p_j + x_1 p_1 = \sum_{j=1}^n p_j x_j. \end{aligned}$$

□

Proposition 4.10 (espérance d'une v.a. positive à densité). *Si la variable aléatoire positive X a pour densité f , on a*

$$\mathbf{E}X = \int_0^{+\infty} xf(x) dx. \quad (4.13)$$

Dans cette formule, $\mathbf{E}X$ peut prendre la valeur $+\infty$ si l'intégrale généralisée diverge.

Preuve. Si X admet pour densité f , on a pour tout t , $P(X > t) = \int_t^{+\infty} f(x) dx$. En reportant cette égalité dans la définition de $\mathbf{E}X$, on obtient :

$$\int_0^{+\infty} P(X > t) dt = \int_0^{+\infty} \left\{ \int_t^{+\infty} f(x) dx \right\} dt = \int_0^{+\infty} \left\{ \int_0^{+\infty} f(x) \mathbf{1}_{[t, +\infty[}(x) dx \right\} dt.$$

Notons que pour $t \geq 0$, on a $\mathbf{1}_{[t, +\infty[}(x) = \mathbf{1}_{[0, x]}(t)$. L'intégrande $(x, t) \mapsto \mathbf{1}_{[0, x]}(t)f(x)$ étant *positive*, le théorème de Fubini-Tonelli légitime l'interversion des intégrations⁷, ce qui donne :

$$\mathbf{E}X = \int_0^{+\infty} \left\{ \int_0^{+\infty} f(x) \mathbf{1}_{[0, x]}(t) dt \right\} dx = \int_0^{+\infty} f(x) \left\{ \int_0^{+\infty} \mathbf{1}_{[0, x]}(t) dt \right\} dx.$$

Comme pour $x \geq 0$, $\int_0^{+\infty} \mathbf{1}_{[0, x]}(t) dt = \int_0^x dt = x$, on en déduit (4.13). \square

Remarque 4.11. Notons que dans la démonstration ci-dessus, nous n'avons utilisé à aucun moment la positivité de la variable aléatoire X . On peut donc appliquer ce calcul à toute variable aléatoire réelle X ayant une densité f pour obtenir :

$$\int_0^{+\infty} P(X > t) dt = \int_0^{+\infty} xf(x) dx \quad (\text{égalité dans } \overline{\mathbb{R}}_+). \quad (4.14)$$

Attention à ne pas écrire $\mathbf{E}X$ au premier membre de (4.14), cette quantité n'étant pour l'instant définie que pour X *positive*. La vraie formule pour $\mathbf{E}X$ lorsque la v.a. réelle X est à densité est donnée à la proposition 4.32.

Proposition 4.12. *Si X est une variable aléatoire positive et c une constante réelle strictement positive, on a*

$$\mathbf{E}(cX) = c\mathbf{E}X.$$

Cette égalité reste vraie pour $c = 0$ si X est de plus intégrable.

Preuve. Puisque X est une variable aléatoire positive et c une constante positive, $cX : \omega \mapsto (cX)(\omega) := cX(\omega)$ est une variable aléatoire positive. En lui appliquant la définition 4.1, on obtient :

$$\mathbf{E}(cX) = \int_0^{+\infty} P(cX > t) dt = \int_0^{+\infty} P\left(X > \frac{t}{c}\right) dt.$$

7. Même si les intégrales valent $+\infty$.

Dans cette intégrale généralisée d'une fonction *positive* localement intégrable sur $[0, +\infty[$, on peut effectuer le changement de variable $s = t/c$, cf. proposition B.41-ii), qui nous donne :

$$\mathbf{E}(cX) = \int_0^{+\infty} P\left(X > \frac{t}{c}\right) dt = c \int_0^{+\infty} P(X > s) ds = c\mathbf{E}X.$$

Dans le cas particulier $c = 0$, cette méthode n'est plus valable (on ne peut déjà plus écrire « $P(cX > t) = P(X > t/c)$ ») mais la formule est vraie trivialement à condition que $\mathbf{E}X$ soit *finie*, puisqu'alors $\mathbf{E}(0 \times X) = \mathbf{E}(0) = 0$ et $0 \times \mathbf{E}X = 0$. \square

Proposition 4.13 (croissance de l'espérance). *Si X et Y sont deux variables aléatoires positives définies sur le même (Ω, \mathcal{F}, P) et si $X \leq Y$ i.e. $X(\omega) \leq Y(\omega)$ pour tout $\omega \in \Omega$, alors $\mathbf{E}X \leq \mathbf{E}Y$.*

Preuve. Si $X(\omega) > t$, alors comme $Y(\omega) \geq X(\omega)$, on a aussi $Y(\omega) > t$. Ceci justifie l'inclusion d'évènements $\{X > t\} \subset \{Y > t\}$, puis l'inégalité $P(X > t) \leq P(Y > t)$. Cette dernière inégalité étant vérifiée pour tout t , on peut l'intégrer entre 0 et $+\infty$, pour obtenir⁸ :

$$\mathbf{E}X = \int_0^{+\infty} P(X > t) dt \leq \int_0^{+\infty} P(Y > t) dt = \mathbf{E}Y.$$

\square

Proposition 4.14. *Pour toute variable aléatoire positive X , on a l'égalité (dans $\overline{\mathbb{R}}_+$) :*

$$\int_0^{+\infty} P(X > t) dt = \int_0^{+\infty} P(X \geq t) dt. \quad (4.15)$$

Avant d'en donner la preuve, notons que (4.15) n'a rien d'évident car $P(X > t)$ et $P(X \geq t)$ peuvent différer pour certaines valeurs de t (au plus pour une infinité dénombrable de valeurs de t).

Preuve. Notons respectivement I et J le premier et le deuxième membre de (4.15). On prouve leur égalité en montrant l'inégalité dans les deux sens. L'inégalité $I \leq J$ s'obtient par intégration de l'inégalité $P(X > t) \leq P(X \geq t)$ vraie pour tout t .

Pour montrer que $J \leq I$, fixons $\varepsilon > 0$ quelconque. L'intégrande dans J est une fonction *positive* localement Riemann intégrable sur $[0, +\infty[$. On peut donc effectuer dans J le changement de variable « translation » $t = s + \varepsilon$, cf. proposition B.41-i), qui nous donne :

$$J = \int_{-\varepsilon}^{+\infty} P(X \geq s + \varepsilon) ds = \int_{-\varepsilon}^0 P(X \geq s + \varepsilon) ds + \int_0^{+\infty} P(X \geq s + \varepsilon) ds.$$

8. La croissance de l'intégrale de Riemann (cf. prop. A.21 ii)) passe aux intégrales généralisées de fonctions *positives*. En effet, si f et g sont positives et localement Riemann intégrables sur $[0, +\infty[$ et telles que $f \leq g$ sur $[0, +\infty[$, alors on a pour tout $x \geq 0$, $\int_0^x f(t) dt \leq \int_0^x g(t) dt$ et cette inégalité entre deux fonctions *croissantes* de x se conserve dans \mathbb{R}_+ par passage à la limite quand x tend vers $+\infty$.

En majorant $P(X \geq s + \varepsilon)$ par 1 sur $[-\varepsilon, 0]$ et par $P(X > s)$ sur $[0, +\infty[$, on en déduit que

$$J \leq \varepsilon + \int_0^{+\infty} P(X > s) ds = \varepsilon + I.$$

L'inégalité $J \leq I + \varepsilon$ étant ainsi vérifiée pour tout $\varepsilon > 0$, on en déduit en faisant tendre ε vers 0 que $J \leq I$. \square

Remarque 4.15. Dans la démonstration ci-dessus, la positivité de X ne joue aucun rôle. Donc (4.15) reste valable pour n'importe quelle variable aléatoire réelle X . Une adaptation facile (faites la !) de la preuve ci-dessus montre que pour X variable aléatoire réelle, on a aussi

$$\int_{-\infty}^0 P(X < t) dt = \int_{-\infty}^0 P(X \leq t) dt. \quad (4.16)$$

Proposition 4.16 (inégalité de Markov). *Si X est une variable aléatoire positive, on a*

$$\forall x > 0, \quad P(X \geq x) \leq \frac{\mathbf{E}X}{x}. \quad (4.17)$$

Remarques 4.17. Cette inégalité n'a d'intérêt que lorsque le second membre est inférieur à 1, i.e. lorsque $\mathbf{E}X < +\infty$ et $x > \mathbf{E}X$.

D'autre part il peut sembler un peu incongru de vouloir contrôler $P(X \geq x)$ à l'aide de $\mathbf{E}X$, puisque le calcul de cette espérance par la définition 4.1 présuppose la connaissance des $P(X > t)$ pour $t \geq 0$ (dont on déduit facilement les $P(X \geq t)$). Il se trouve qu'il arrive souvent en pratique que l'on sache calculer $\mathbf{E}X$ sans connaître, ou sans avoir besoin de calculer, la loi de X . C'est le cas par exemple quand X est une somme finie de variables aléatoires d'espérances connues. On peut aussi savoir majorer $\mathbf{E}X$ sans connaître la loi de X . Dans ces situations, l'inégalité de Markov est très utile. Pour ne citer qu'un exemple, l'inégalité de Markov est l'un des outils pour établir des « lois des grands nombres ».

Voici maintenant 3 preuves de l'inégalité de Markov, libre au lecteur de choisir celle qu'il préfère.

Preuve n° 1. Voir la figure 4.7. \square

Preuve n° 2. Cette preuve ne fait que traduire explicitement la preuve graphique n° 1. Fixons $x > 0$, la quantité $P(X \geq x)$ devenant ainsi une constante. À partir de cette constante, définissons la fonction $h : [0, +\infty[\rightarrow \mathbb{R}_+$, $t \mapsto P(X \geq x) \mathbf{1}_{[0, x]}(t)$. Par décroissance de la fonction $t \mapsto P(X \geq t)$, on a $h(t) \leq P(X \geq t)$ pour tout $t \in [0, x]$. D'autre part cette inégalité est aussi vérifiée pour tout $t > x$ car alors $h(t) = 0$. En intégrant sur $[0, +\infty[$ l'inégalité $h(t) \leq P(X \geq t)$, on obtient compte-tenu de (4.15) :

$$\int_0^{+\infty} h(t) dt \leq \int_0^{+\infty} P(X \geq t) dt = \mathbf{E}X.$$

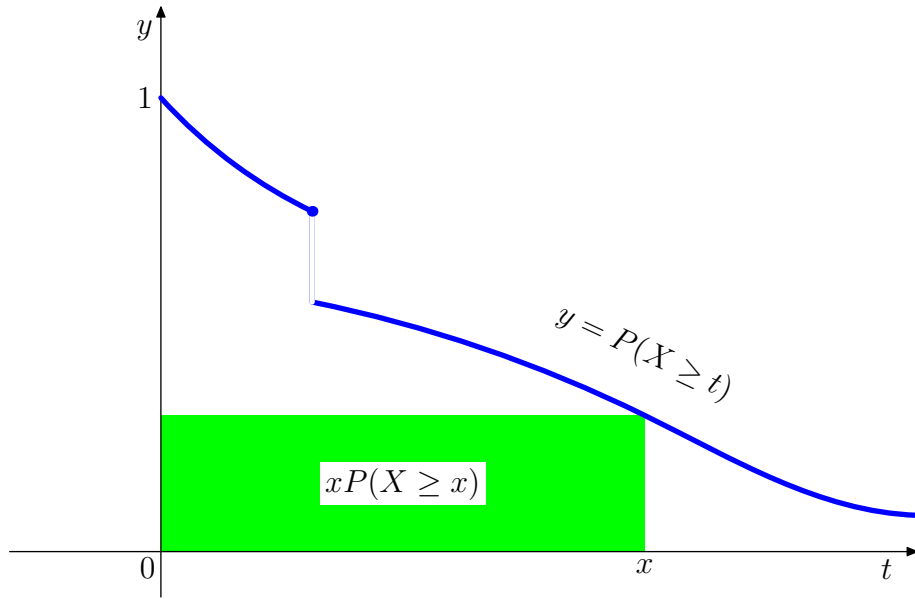


FIG. 4.7 – Inégalité de Markov : $xP(X \geq x) \leq \int_0^{+\infty} P(X \geq t) dt = \mathbf{E}X$.

D'autre part, puisque h est nulle sur $]x, +\infty[$,

$$\int_0^{+\infty} h(t) dt = \int_0^x P(X \geq x) dt = xP(X \geq x).$$

Par conséquent

$$xP(X \geq x) \leq \mathbf{E}X,$$

ce qui nous donne (4.17) puisque $x > 0$. □

Preuve n° 3. Cette preuve plus abstraite exploite les propriétés déjà connues de l'espérance des v.a. positives. On fixe x qui joue donc le rôle d'une constante dans toute la preuve. On part de l'inégalité entre v.a. positives : $x\mathbf{1}_{\{X \geq x\}} \leq X$ (vérifiez) dont on déduit par croissance de \mathbf{E} (proposition 4.13) :

$$\mathbf{E}(x\mathbf{1}_{\{X \geq x\}}) \leq \mathbf{E}X,$$

puis grâce aux propositions 4.12 et 4.7, $xP(X \geq x) \leq \mathbf{E}X$. On conclut en divisant par $x > 0$. □

Corollaire 4.18. *Si X est une variable aléatoire positive, on a l'équivalence*

$$\mathbf{E}X = 0 \Leftrightarrow P(X = 0) = 1,$$

autrement dit $\mathbf{E}X$ est nulle si et seulement si X est presque sûrement nulle.

Preuve. Supposons d'abord que $\mathbf{E}X = 0$. La v.a. X étant positive, l'égalité $P(X = 0) = 1$ équivaut à $P(X > 0) = 0$. Introduisons la suite des événements $A_n := \{X \geq 1/n\}$, $n \in \mathbb{N}^*$. Cette suite est croissante de réunion $A := \{X > 0\}$. Par continuité séquentielle croissante de P , $P(A) = \lim_{n \rightarrow +\infty} P(A_n)$. Or l'inégalité de Markov appliquée avec $x = 1/n$ nous montre que pour tout $n \in \mathbb{N}^*$, $0 \leq P(A_n) \leq n\mathbf{E}X = 0$. Ainsi $P(A_n) = 0$ pour tout n et $P(A) = 0$ comme limite de la suite nulle.

Réciproquement, si $P(X = 0) = 1$, alors pour tout $t \geq 0$, $P(X > t) = 0$, d'où $\mathbf{E}X = \int_0^{+\infty} 0 dt = 0$. \square

Proposition 4.19 (inégalité de Markov raffinée). *Si X est une variable aléatoire positive intégrable ($\mathbf{E}X < +\infty$),*

$$\lim_{x \rightarrow +\infty} xP(X \geq x) = 0. \quad (4.18)$$

Preuve. Voir la figure 4.8. \square

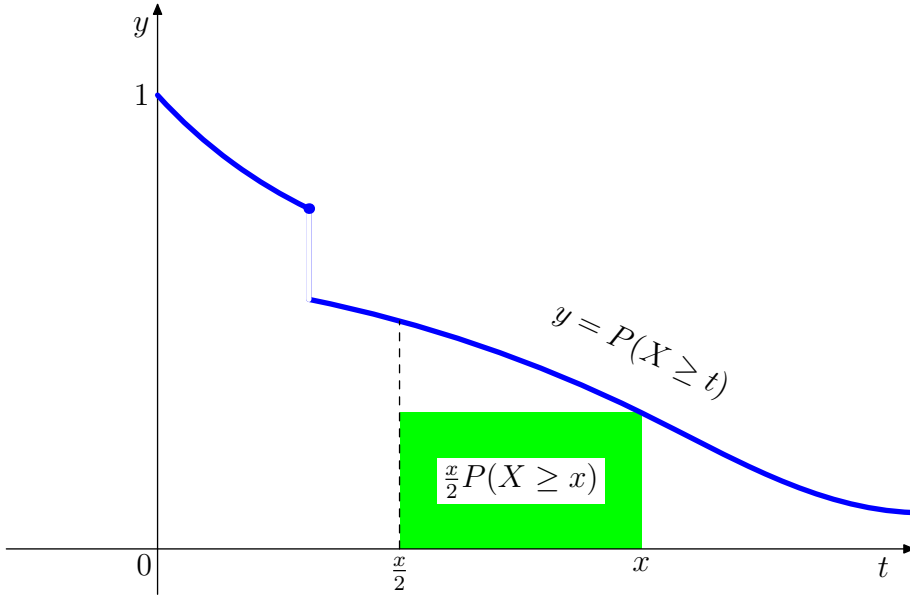


FIG. 4.8 – Preuve de (4.18) : $\frac{x}{2}P(X \geq x) \leq \int_{x/2}^{+\infty} P(X \geq t) dt = o(1)$ si $\mathbf{E}X < +\infty$.

Remarque 4.20. La réciproque de la proposition 4.19 est fausse. Il est possible que $xP(X \geq x) = o(1)$ sans que X soit intégrable. Un contre exemple élémentaire est obtenu avec X de fonction de survie $G(t) = P(X > t) = 1$ si $t \leq 1$ et $1/t$ si $t > 1$.

Théorème 4.21 (approximation par suite croissante de v.a. simples). *Toute variable aléatoire positive X sur (Ω, \mathcal{F}) est limite simple sur Ω d'une suite croissante $(X_n)_{n \geq 1}$ de variables aléatoires positives simples.*

Preuve. L'idée est d'utiliser pour construire X_n , les *valeurs approchées dyadiques par défaut* de X au *niveau de résolution* n . On peut procéder comme suit en définissant pour $n \in \mathbb{N}$, les ensembles

$$\begin{aligned} A_{n,k} &:= X^{-1}([k2^{-n}, (k+1)2^{-n}[), \quad 0 \leq k \leq n2^n - 1; \\ A_{n,n2^n} &:= X^{-1}([n, +\infty[). \end{aligned}$$

On prend alors

$$X_n := \sum_{k=0}^{n2^n} \frac{k}{2^n} \mathbf{1}_{A_{n,k}}.$$

Autrement dit,

$$X_n(\omega) = \begin{cases} n & \text{si } n \leq X(\omega), \\ k2^{-n} & \text{pour l'unique entier } k \text{ tel que } k2^{-n} \leq X(\omega) < (k+1)2^{-n} \text{ sinon.} \end{cases}$$

Comme X est mesurable, les $A_{n,k}$ sont dans \mathcal{F} , ce qui entraîne la mesurabilité de X_n (combinaison linéaire d'indicatrices d'éléments de la tribu \mathcal{F}). Comme $X_n(\Omega)$ est une partie finie de \mathbb{R}_+ , X_n est simple positive.

Il reste à vérifier que pour tout $\omega \in \Omega$, la suite de réels $(X_n(\omega))_{n \geq 1}$ est croissante et converge dans \mathbb{R}_+ vers $X(\omega)$. On note $[x]$ la partie entière du réel x , unique entier m tel que $m \leq x < m+1$. On voit que $X_n(\omega) = n$ pour $n \leq [X(\omega)]$ et que pour $n > [X(\omega)]$,

$$X_n(\omega) = \frac{k(n, \omega)}{2^n} = \max \left\{ \frac{l}{2^n}; \frac{l}{2^n} \leq X(\omega), l \in \mathbb{N} \right\} = 2^{-n} [2^n X(\omega)]. \quad (4.19)$$

La suite *finie* $(X_n(\omega))_{n \leq [X(\omega)]}$ est clairement croissante. Voyons la suite $(X_n(\omega))_{n > [X(\omega)]}$. D'après (4.19), on a

$$X_n(\omega) = \frac{k(n, \omega)}{2^n} = \frac{2k(n, \omega)}{2^{n+1}} \leq X(\omega) \Rightarrow \frac{2k(n, \omega)}{2^{n+1}} \leq \frac{k(n+1, \omega)}{2^{n+1}} = X_{n+1}(\omega),$$

d'où la croissance de la suite $(X_n(\omega))_{n > [X(\omega)]}$. Pour établir définitivement la croissance de *toute* la suite $(X_n(\omega))_{n \geq 1}$, il ne reste plus qu'à examiner le point de raccord des deux sous-suites, donc à comparer $X_n(\omega)$ et $X_{n+1}(\omega)$ pour $n = [X(\omega)]$. Il suffit de remarquer que $X_n(\omega) = n = (n2^{n+1})2^{-n-1} \leq X(\omega)$ et comme $X_{n+1}(\omega)$ est donné par (4.19), on a $(n2^{n+1})2^{-n-1} \leq k(n+1, \omega)2^{-n-1} = X_{n+1}(\omega)$.

La convergence est immédiate, puisque pour $n > [X(\omega)]$, on a d'après (4.19)

$$X_n(\omega) \leq X(\omega) < X_n(\omega) + \frac{1}{2^n},$$

d'où $0 \leq X(\omega) - X_n(\omega) < 2^{-n}$. Le théorème est démontré. On peut remarquer en bonus que la convergence est *uniforme* sur Ω si X est bornée (*i.e.* $M := \sup_{\omega \in \Omega} X(\omega) < +\infty$). En effet pour $n > M$ (constante indépendante de ω), on a pour tout $\omega \in \Omega$, $0 \leq X(\omega) - X_n(\omega) < 2^{-n}$. \square

Lemme 4.22. *Si $(X_n)_{n \geq 1}$ est une suite croissante de variables aléatoires positives ayant pour limite sur tout Ω la variable aléatoire positive X , on a*

$$\forall t \geq 0, \quad P(X_n > t) \uparrow P(X > t). \quad (4.20)$$

Preuve. Pour t fixé, la suite des évènements $A_n := \{X_n > t\}$ est croissante pour l'inclusion (puisque si $\omega \in A_n$, $X_{n+1}(\omega) \geq X_n(\omega) > t$, donc $\omega \in A_{n+1}$). Par continuité séquentielle croissante de P , $P(A_n)$ tend en croissant vers $P(A)$, où $A := \cup_{n \geq 1} A_n$. On prouve alors la convergence (4.20) en vérifiant que $A = \{X > t\}$. Pour cela on montre l'inclusion dans les deux sens.

Soit $\omega \in A$ quelconque. Cela signifie qu'il existe un $n_0 = n_0(\omega)$ tel que $\omega \in A_{n_0}$, i.e. $X_{n_0}(\omega) > t$. Alors par croissance de la suite $(X_n(\omega))_{n \geq 1}$, $X_n(\omega) \geq X_{n_0}(\omega)$ pour tout $n \geq n_0$, d'où en faisant tendre n vers l'infini, $X(\omega) \geq X_{n_0}(\omega) > t$. Donc en particulier, $X(\omega) > t$, d'où $\omega \in \{X > t\}$. Comme ω était quelconque, ceci établit l'inclusion $A \subset \{X > t\}$.

Pour l'inclusion inverse, soit ω quelconque dans $\{X > t\}$. Alors $X(\omega) > t$ et comme cette inégalité est stricte⁹ et $X_n(\omega)$ tend vers $X(\omega)$ quand n tend vers l'infini, on peut trouver un $n_0 = n_0(\omega)$ tel que pour tout $n \geq n_0$, $X_n(\omega) > t$ (si cela ne vous paraît pas évident, posez $\varepsilon = X(\omega) - t$ d'où $t = X(\omega) - \varepsilon$). Ainsi $\omega \in A_n$ pour tout $n \geq n_0$, donc *a fortiori* $\omega \in A$. L'inclusion $\{X > t\} \subset A$ est ainsi établie et ceci termine la preuve du lemme. \square

Théorème 4.23 (de Beppo Levi). *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires positives qui converge en croissant vers la variable aléatoire positive X , i.e. pour tout n $X_n \leq X_{n+1}$ et pour tout $\omega \in \Omega$, $\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)$. Alors la suite $\mathbf{E}X_n$ converge en croissant (dans $\overline{\mathbb{R}}_+$) vers $\mathbf{E}X$.*

Preuve. La croissance de la suite $(\mathbf{E}X_n)_{n \geq 1}$ dans $\overline{\mathbb{R}}_+$ est évidente par croissance de l'espérance (proposition 4.13). Pour étudier sa convergence, nous allons distinguer deux cas selon que X est ou n'est pas intégrable.

Cas 1, $\mathbf{E}X < +\infty$. Fixons $\varepsilon > 0$ quelconque. Par convergence de l'intégrale de Riemann généralisée $\int_0^{+\infty} P(X > t) dt$, on peut trouver un réel positif b (dépendant de ε) tel que :

$$\int_b^{+\infty} P(X > t) dt < \varepsilon. \quad (4.21)$$

De cette égalité et de la croissance de (X_n) qui implique que $X_n \leq X$ et donc que pour tout t , $P(X_n > t) \leq P(X > t)$, on déduit :

$$\forall n \geq 1, \quad \int_b^{+\infty} P(X_n > t) dt < \varepsilon. \quad (4.22)$$

9. C'est là qu'on voit pourquoi le lemme ne marcherait pas avec $P(X_n \geq t)$ et $P(X \geq t)$. Voici d'ailleurs un contre exemple élémentaire. On prend $X_n = 1 - 1/n$, v.a. constante et $X = 1$. Alors avec $t = 1$, on obtient $P(X_n \geq 1) = 0$ pour tout n , tandis que $P(X \geq 1) = 1$, ce qui empêche la convergence.

Posons $G_n(t) := P(X_n > t)$ et $G(t) := P(X > t)$. Chaque G_n est une fonction *décroissante* sur $[0, b]$ et par le lemme 4.22, $G_n(t)$ converge vers $G(t)$ pour tout $t \in [0, b]$. La suite (G_n) satisfait ainsi les hypothèses du théorème d'interversion limite et intégrale de Riemann pour une suite de fonctions décroissantes (cf. théorème A.35), donc :

$$\int_0^b P(X_n > t) dt \xrightarrow{n \rightarrow +\infty} \int_0^b P(X > t) dt. \quad (4.23)$$

En combinant (4.21), (4.22) et (4.23), on voit qu'il existe un entier $n_0(\varepsilon)$ tel que

$$\forall n \geq n_0(\varepsilon), \quad 0 \leq \int_0^{+\infty} P(X > t) dt - \int_0^{+\infty} P(X_n > t) dt < 2\varepsilon,$$

ce qui établit la convergence de $\mathbf{E}X_n$ vers $\mathbf{E}X$. □

Cas 2, $\mathbf{E}X = +\infty$. Il s'agit cette fois de montrer que $\mathbf{E}X_n$ tend vers $+\infty$. La divergence de l'intégrale généralisée $\int_0^{+\infty} P(X > t) dt$ signifie ici que $\int_0^x P(X > t) dt$ tend vers $+\infty$ quand x tend vers $+\infty$. Donc si on fixe $A > 0$ arbitraire, on peut trouver un $b \in]0, +\infty[$ tel que

$$\int_0^b P(X > t) dt > A. \quad (4.24)$$

En appliquant comme ci-dessus le théorème d'interversion limite intégrale pour la suite de fonctions décroissantes (G_n) , on obtient la convergence (dans \mathbb{R}_+) de $\int_0^b P(X_n > t) dt$ vers $\int_0^b P(X > t) dt$ quand n tend vers $+\infty$. Il existe donc un entier $n_0(A)$ tel que :

$$\forall n \geq n_0(A), \quad \int_0^b P(X_n > t) dt > \frac{A}{2}. \quad (4.25)$$

Ceci implique évidemment que

$$\forall n \geq n_0(A), \quad \mathbf{E}X_n = \int_0^{+\infty} P(X_n > t) dt > \frac{A}{2},$$

et comme A est arbitraire, on a ainsi prouvé que $\mathbf{E}X_n$ tend vers $+\infty$. □

La démonstration du théorème de Beppo Levi est achevée. □

Corollaire 4.24 (additivité de l'espérance). *Si X et Y sont deux variables aléatoires positives définies sur le même (Ω, \mathcal{F}, P) ,*

$$\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y \quad (\text{égalité dans } \overline{\mathbb{R}}_+). \quad (4.26)$$

Preuve. On montre (4.26) par un passage à la limite à partir de l'additivité de l'espérance des variables simples. Cette propriété, déjà vue en Deug, est reprise sous une forme *ad hoc* dans le lemme 4.25 ci-dessous.

Par le théorème 4.21, il existe deux suites (X_n) et (Y_n) de variables aléatoires simples positives telles que $X_n \uparrow X$ et $Y_n \uparrow Y$. Alors en posant $Z_n := X_n + Y_n$, on a aussi $Z_n \uparrow X + Y$. Par le lemme 4.25, on a

$$\forall n \in \mathbb{N}, \quad \mathbf{E}Z_n = \mathbf{E}(X_n + Y_n) = \mathbf{E}X_n + \mathbf{E}Y_n. \quad (4.27)$$

En appliquant le théorème de Beppo Levi à chacune des suites (Z_n) , (X_n) et (Y_n) , on peut passer à la limite quand n tend vers $+\infty$ dans (4.27) pour obtenir (4.26). \square

Lemme 4.25. *Si X et Y sont deux variables aléatoires positives simples, il en est de même pour $Z := X + Y$ et on a $\mathbf{E}Z = \mathbf{E}X + \mathbf{E}Y$.*

Preuve. Par hypothèse $X(\Omega)$ et $Y(\Omega)$ sont des sous-ensembles finis de \mathbb{R}_+ et il est clair qu'il en va de même pour $Z(\Omega)$. Donc Z est une variable aléatoire positive simple. Notons

$$X(\Omega) = \{x_1, \dots, x_l\}, \quad Y(\Omega) = \{y_1, \dots, y_m\}, \quad Z(\Omega) = \{z_1, \dots, z_n\}.$$

On voit facilement, en partitionnant Ω suivant toutes les valeurs possibles de Y , resp. de X , que

$$P(X = x_i) = \sum_{y_j \in Y(\Omega)} P(X = x_i \text{ et } Y = y_j), \quad (4.28)$$

$$P(Y = y_j) = \sum_{x_i \in X(\Omega)} P(X = x_i \text{ et } Y = y_j). \quad (4.29)$$

Pour $z_k \in Z(\Omega)$, la décomposition en union finie d'événements 2 à 2 disjoints¹⁰

$$\{Z = z_k\} = \bigcup_{x_i + y_j = z_k} \{X = x_i \text{ et } Y = y_j\},$$

nous donne

$$P(Z = z_k) = \sum_{x_i + y_j = z_k} P(X = x_i \text{ et } Y = y_j). \quad (4.30)$$

On obtient alors la conclusion souhaitée par le calcul suivant dans lequel tous les Σ sont des sommes d'un nombre fini de termes et où on utilise (4.28)–(4.30) et la proposition 4.9

10. Pour $z_k \in Z(\Omega)$ et certains couples (x_i, y_j) tels que $x_i + y_j = z_k$, il se peut qu'il n'existe aucun $\omega \in \Omega$ tel que $X(\omega) = x_i$ et $Y(\omega) = y_j$, l'événement $\{X = x_i \text{ et } Y = y_j\}$ est alors vide.

appliquée d'abord à Z , puis à X et Y .

$$\begin{aligned}
 \mathbf{E}Z &= \sum_{z_k \in Z(\Omega)} z_k P(Z = z_k) = \sum_{z_k \in Z(\Omega)} z_k \sum_{x_i + y_j = z_k} P(X = x_i \text{ et } Y = y_j) \\
 &= \sum_{z_k \in Z(\Omega)} \sum_{x_i + y_j = z_k} (x_i + y_j) P(X = x_i \text{ et } Y = y_j) \\
 &= \sum_{x_i \in X(\Omega), y_j \in Y(\Omega)} (x_i + y_j) P(X = x_i \text{ et } Y = y_j) \\
 &= \sum_{x_i \in X(\Omega), y_j \in Y(\Omega)} x_i P(X = x_i \text{ et } Y = y_j) \\
 &\quad + \sum_{x_i \in X(\Omega), y_j \in Y(\Omega)} y_j P(X = x_i \text{ et } Y = y_j) \\
 &= \sum_{x_i \in X(\Omega)} x_i \sum_{y_j \in Y(\Omega)} P(X = x_i \text{ et } Y = y_j) \\
 &\quad + \sum_{y_j \in Y(\Omega)} y_j \sum_{x_i \in X(\Omega)} P(X = x_i \text{ et } Y = y_j) \\
 &= \sum_{x_i \in X(\Omega)} x_i P(X = x_i) + \sum_{y_j \in Y(\Omega)} y_j P(Y = y_j) \\
 &= \mathbf{E}X + \mathbf{E}Y.
 \end{aligned}$$

□

Corollaire 4.26 (intersion série-espérance). *Soit $(X_k)_{k \in \mathbb{N}}$, une suite de variables aléatoires positives définies sur (Ω, \mathcal{F}) . On suppose que pour tout $\omega \in \Omega$, la série $\sum_{k=0}^{+\infty} X_k(\omega)$ converge¹¹ dans \mathbb{R}_+ . Alors l'application*

$$S : \Omega \rightarrow \mathbb{R}_+, \quad \omega \mapsto S(\omega) := \sum_{k=0}^{+\infty} X_k(\omega),$$

est une variable aléatoire positive sur (Ω, \mathcal{F}) et on a

$$\mathbf{E}S = \mathbf{E} \left(\sum_{k=0}^{+\infty} X_k \right) = \sum_{k=0}^{+\infty} \mathbf{E}X_k, \quad (4.31)$$

cette égalité ayant lieu dans $\overline{\mathbb{R}}_+$.

Preuve. L'application $S : \Omega \rightarrow \mathbb{R}_+$ est bien définie, grâce à l'hypothèse de convergence simple de la série sur tout Ω . Posons

$$\forall n \in \mathbb{N}, \quad S_n := \sum_{k=0}^n X_k.$$

11. Nous donnons ici une « version bridée » de l'intersion série-espérance. On pourrait se passer de cette hypothèse de convergence dans \mathbb{R}_+ , en considérant que cette série converge toujours dans $\overline{\mathbb{R}}_+$. Il faudrait alors considérer qu'une variable aléatoire positive est une application à valeurs dans $\overline{\mathbb{R}}_+$, ce que nous avons évité de faire dans ce cours...

Les S_n sont mesurables \mathcal{F} - $\text{Bor}(\mathbb{R}_+)$ comme sommes finies d'applications mesurables (les X_k). Comme S est limite simple sur tout Ω d'une suite d'applications mesurables, elle est mesurable pour les tribus \mathcal{F} - $\text{Bor}(\mathbb{R}_+)$, c'est donc bien une *variable aléatoire positive*.

Par additivité de l'espérance (corollaire 4.24, qui s'étend par une récurrence immédiate à toute somme d'un nombre *fini* de variables aléatoires positives¹²) on a

$$\forall n \in \mathbb{N}, \quad \mathbf{E}S_n = \sum_{k=0}^n \mathbf{E}X_k \quad (\text{égalité dans } \overline{\mathbb{R}}_+). \quad (4.32)$$

En raison de la positivité des X_k , la convergence de S_n vers S est *croissante* ($S_n \uparrow S$), donc par le théorème de Beppo Levi, $\mathbf{E}S_n$ converge vers $\mathbf{E}S$ quand n tend vers $+\infty$. En faisant tendre n vers $+\infty$ dans (4.32), on obtient alors la conclusion souhaitée (noter que le second membre de (4.32) converge dans $\overline{\mathbb{R}}_+$ vers le second membre de (4.31)). \square

Corollaire 4.27 (espérance d'une v.a. discrète positive). *Pour toute variable aléatoire discrète positive X ,*

$$\mathbf{E}X = \sum_{x \in X(\Omega)} xP(X = x), \quad (\text{égalité dans } \overline{\mathbb{R}}_+). \quad (4.33)$$

Preuve. Puisque X est discrète positive, $X(\Omega)$ est une partie finie ou dénombrable de \mathbb{R}_+ . Le cas où $X(\Omega)$ est finie est celui des variables aléatoires simples déjà traité ci-dessus (proposition 4.9). Supposons désormais $X(\Omega)$ dénombrable et indexons ses éléments par les entiers : $X(\Omega) = \{x_k; k \in \mathbb{N}\}$. On peut alors représenter X comme la somme d'une série de variables aléatoires positives en écrivant (la justification suit) :

$$X = \sum_{k=0}^{+\infty} x_k \mathbf{1}_{A_k}, \quad (A_k := \{X = x_k\}, \forall k \in \mathbb{N}). \quad (4.34)$$

En effet les événements A_k réalisent une partition de Ω , donc pour tout $\omega \in \Omega$, il existe un unique indice $j = j(\omega)$ tel que $\omega \in A_j$ et pour cet ω , $X(\omega) = x_j$. On a alors

$$\sum_{k=0}^{+\infty} x_k \mathbf{1}_{A_k}(\omega) = x_j \mathbf{1}_{A_j}(\omega) = x_j = X(\omega),$$

car dans cette série de réels positifs, il y a au plus un terme non nul¹³, celui d'indice j (pour $k \neq j$, $\omega \notin A_k$, d'où $\mathbf{1}_{A_k}(\omega) = 0$). Ceci prouve que la série de variables aléatoires positives (4.34) converge sur tout Ω vers la variable aléatoire positive X . Les hypothèses du corollaire 4.26 étant ainsi vérifiées, on a

$$\mathbf{E}X = \sum_{k=0}^{+\infty} \mathbf{E}(x_k \mathbf{1}_{A_k}) = \sum_{k=0}^{+\infty} x_k \mathbf{E}(\mathbf{1}_{A_k}) = \sum_{k=0}^{+\infty} x_k P(A_k) = \sum_{k=0}^{+\infty} x_k P(X = x_k),$$

en utilisant aussi les propositions 4.12 et 4.7. La formule (4.33) est ainsi démontrée. \square

12. Mais cette récurrence ne permet pas de traiter le cas d'une *série*.

13. Il se peut qu'ils soient tous nuls si $x_j = 0$.

4.3 Espérance d'une variable aléatoire réelle

L'extension de la notion d'espérance au cas des variables aléatoires réelles repose sur la décomposition $X = X^+ - X^-$ et la formule « $\mathbf{E}X := \mathbf{E}(X^+) - \mathbf{E}(X^-)$ », sous réserve que cette soustraction ait lieu dans \mathbb{R} . Avant d'expliquer cela, donnons quelques précisions sur les variables aléatoires positives X^+ et X^- . Elles sont définies par

$$X^+ := \max(0, X), \quad X^- := \max(0, -X).$$

Notons que X^+ et X^- sont des variables aléatoires¹⁴ positives. En se rappelant la définition 1.71, on a donc :

$$X^+ : \Omega \rightarrow \mathbb{R}_+, \quad \omega \mapsto X^+(\omega) = (X(\omega))^+, \quad X^- : \Omega \rightarrow \mathbb{R}_+, \quad \omega \mapsto X^-(\omega) = (X(\omega))^-.$$

Au risque d'insister lourdement, notons encore que pour tout $\omega \in \Omega$,

$$X^+(\omega) = \begin{cases} X(\omega) = |X(\omega)| & \text{si } X(\omega) \geq 0, \\ 0 & \text{si } X(\omega) \leq 0, \end{cases} \quad X^-(\omega) = \begin{cases} 0 & \text{si } X(\omega) \geq 0, \\ -X(\omega) = |X(\omega)| & \text{si } X(\omega) \leq 0. \end{cases}$$

On en déduit les égalités

$$X = X^+ - X^-, \quad |X| = X^+ + X^-.$$

Pour définir $\mathbf{E}X$ comme la différence $\mathbf{E}(X^+) - \mathbf{E}(X^-)$, il est clair qu'il nous faut interdire que ces deux espérances de variables aléatoires positives valent *simultanément* $+\infty$. On pourrait autoriser l'une des deux à valoir $+\infty$ à condition que l'autre soit finie, $\mathbf{E}X$ prenant alors la valeur $+\infty$ si $\mathbf{E}(X^+) = +\infty$ et $\mathbf{E}(X^-) < +\infty$, ou dans le cas inverse $\mathbf{E}X = -\infty$. Nous ne retiendrons pas cette option, car le problème de « $+\infty - \infty$ » réapparaîtrait de toutes façons pour la somme de deux variables aléatoires, ce qui empêcherait l'additivité de l'espérance pour les variables aléatoires réelles. Ainsi nous ne parlerons de l'espérance (sous-entendu dans \mathbb{R}) de la variable aléatoire réelle X que si $\mathbf{E}(X^+) < +\infty$ et $\mathbf{E}(X^-) < +\infty$, ou ce qui est équivalent si $\mathbf{E}|X| < +\infty$. Les propriétés de l'espérance établies dans cette section ne concernent que les variables aléatoires remplissant cette condition¹⁵. On dit que ces variables aléatoires réelles sont *intégrables*.

Définition 4.28 (v.a. réelle intégrable). *La variable aléatoire réelle X est dite intégrable si $\mathbf{E}|X| < +\infty$, autrement dit si :*

$$\int_0^{+\infty} P(|X| > t) dt < +\infty.$$

14. Elles héritent de la mesurabilité de X .

15. À titre exceptionnel, dans certaines situations, il peut être commode de dire qu'une variable aléatoire réelle a une espérance *infinie* si *une seule* des deux quantités $\mathbf{E}(X^+)$ et $\mathbf{E}(X^-)$ est infinie, par exemple dans la situation de la remarque 4.3, pour rester cohérent avec la définition de l'espérance d'une v.a. *positive* comme élément de \mathbb{R}_+ . Il convient alors de manipuler ces « espérances infinies » avec les plus grandes précautions. En tout état de cause, on ne parle *jamaïs* d'espérance ni finie ni infinie lorsque $\mathbf{E}X^+$ et $\mathbf{E}X^-$ valent tous deux $+\infty$.

Définition 4.29 (espérance d'une v.a. réelle). *Si X est une variable aléatoire réelle intégrable, on appelle espérance de X , le réel $\mathbf{E}X$ défini par*

$$\mathbf{E}X := \mathbf{E}(X^+) - \mathbf{E}(X^-). \quad (4.35)$$

Le second membre de (4.35) a bien un sens et représente un nombre réel, puisqu'en raison des inégalités $0 \leq X^+ \leq |X|$ et $0 \leq X^- \leq |X|$, l'intégrabilité de X implique la finitude de $\mathbf{E}(X^+)$ et de $\mathbf{E}(X^-)$.

Remarque 4.30. Une fois le réel $\mathbf{E}X$ défini comme ci-dessus, on peut parler de sa partie positive $(\mathbf{E}X)^+$ et de sa partie négative $(\mathbf{E}X)^-$. Il faut prendre garde à ne pas les confondre avec $\mathbf{E}(X^+)$ et $\mathbf{E}(X^-)$. Voici un exemple simple où ces quantités diffèrent. On prend X variable aléatoire de Rademacher, *i.e.* $X(\Omega) = \{-1, 1\}$ et $P(X = -1) = P(X = 1) = 1/2$. Comme X est bornée, elle est intégrable, donc $\mathbf{E}X$ existe. Comme la loi de X est symétrique, $\mathbf{E}X = 0$, donc $(\mathbf{E}X)^+ = 0^+ = 0$ et $(\mathbf{E}X)^- = 0^- = 0$. D'autre part, on vérifie (faites le !) que X^+ et X^- sont des variables de Bernoulli de paramètre $1/2$, donc $\mathbf{E}(X^+) = \mathbf{E}(X^-) = 1/2$.

La formule (4.35) ne semble pas très commode pour le calcul explicite de $\mathbf{E}X$. On aimerait pouvoir exprimer $\mathbf{E}X$, directement à l'aide de la loi de X , par exemple de sa f.d.r., sans passer par la loi de X^+ et celle de X^- . C'est l'objet de la proposition suivante.

Proposition 4.31 (calcul de $\mathbf{E}X$ à l'aide de la f.d.r.). *Soit X une variable aléatoire réelle intégrable de fonction de répartition F . Son espérance vérifie*

$$\mathbf{E}X = \int_0^{+\infty} P(X > t) dt - \int_{-\infty}^0 P(X < t) dt = \int_0^{+\infty} (1 - F(t)) dt - \int_{-\infty}^0 F(t) dt. \quad (4.36)$$

Preuve. Compte-tenu des remarques faites au début de cette section, on vérifie facilement pour tout $t \geq 0$, les équivalences logiques suivantes :

$$X^+(\omega) > t \Leftrightarrow X(\omega) > t, \quad (4.37)$$

$$X^-(\omega) > t \Leftrightarrow X(\omega) < -t. \quad (4.38)$$

Laissant (4.37) au lecteur, détaillons la deuxième. Comme $t \geq 0$, l'inégalité $X^-(\omega) > t$ implique que $X^-(\omega)$ est strictement positif, donc que $X(\omega)$ est strictement négatif et dans ce cas, $X^-(\omega) = -X(\omega)$, d'où $-X(\omega) > t$ et donc $X(\omega) < -t$. Ceci justifie l'implication \Rightarrow dans (4.38). Réciproquement, si $X(\omega) < -t$, $X(\omega)$ est strictement négatif, donc $X(\omega) = -X^-(\omega)$, d'où $-X^-(\omega) < -t$ et donc $X^-(\omega) > t$.

Les équivalences (4.37) et (4.38) nous donnent les égalités d'évènements $\{X^+ > t\} = \{X > t\}$ et $\{X^- > t\} = \{X < -t\}$, valables pour tout ¹⁶ $t \geq 0$, d'où :

$$\forall t \geq 0, \quad P(X^+ > t) = P(X > t), \quad P(X^- > t) = P(X < -t). \quad (4.39)$$

16. Elles ne seraient pas valables pour tout t réel, la positivité de t est utilisée dans la vérification de (4.37) et (4.38).

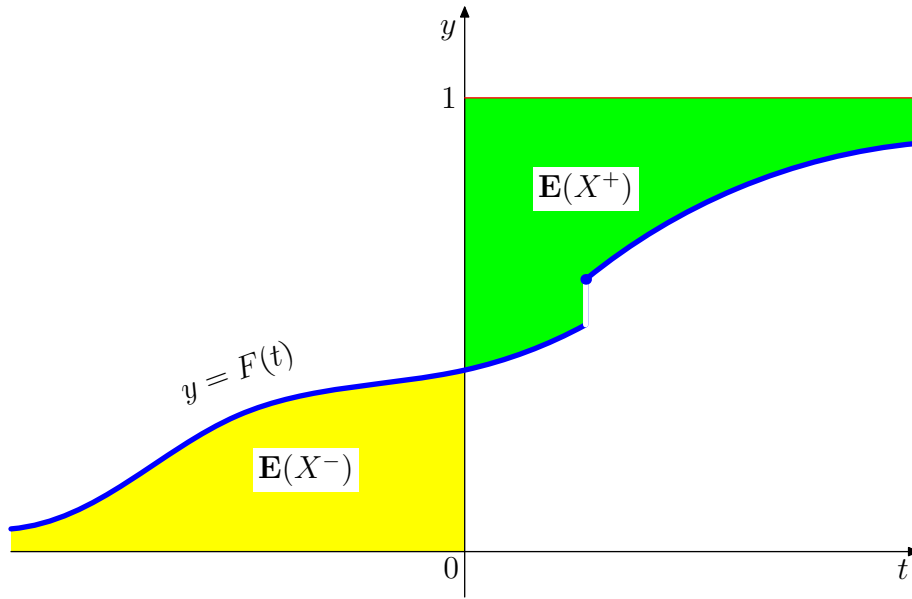


FIG. 4.9 – Espérance d’une v.a. réelle $\mathbf{E}X := \mathbf{E}(X^+) - \mathbf{E}(X^-)$

En reportant ces égalités dans la définition de $\mathbf{E}X$, on justifie le passage à (4.40) dans le calcul suivant :

$$\begin{aligned} \mathbf{E}X = \mathbf{E}(X^+) - \mathbf{E}(X^-) &= \int_0^{+\infty} P(X^+ > t) dt - \int_0^{+\infty} P(X^- > t) dt \\ &= \int_0^{+\infty} P(X > t) dt - \int_0^{+\infty} P(X < -t) dt \quad (4.40) \end{aligned}$$

$$= \int_0^{+\infty} P(X > t) dt - \int_{-\infty}^0 P(X < s) ds \quad (4.41)$$

$$= \int_0^{+\infty} (1 - F(t)) dt - \int_{-\infty}^0 F(t) dt. \quad (4.42)$$

Le passage de (4.40) à (4.41) résulte bien sûr du changement de variable $s = -t$ et justifie la première égalité dans (4.36). Le passage de (4.41) à (4.42) résulte de la remarque 4.15 et du remplacement de la « variable muette s » par t . \square

Proposition 4.32 (espérance d’une v.a. réelle à densité). *Soit X une variable aléatoire réelle de densité f . Alors X est intégrable si et seulement si*

$$\int_{-\infty}^{+\infty} |x|f(x) dx < +\infty \quad (4.43)$$

et dans ce cas,

$$\mathbf{E}X = \int_{-\infty}^{+\infty} xf(x) dx. \quad (4.44)$$

Preuve. Nous savons déjà par (4.39) et la remarque 4.11 que

$$\int_0^{+\infty} P(X^+ > t) dt = \int_0^{+\infty} P(X > t) dt = \int_0^{+\infty} xf(x) dx = \int_0^{+\infty} |x|f(x) dx, \quad (4.45)$$

sans aucune condition d'intégrabilité sur X .

Cherchons une formule analogue pour $\int_0^{+\infty} P(X^- > t) dt$. En utilisant (4.39), le passage de (4.40) à (4.41) et la remarque 4.15, on peut démarrer avec :

$$\int_0^{+\infty} P(X^- > t) dt = \int_{-\infty}^0 P(X < t) dt = \int_{-\infty}^0 P(X \leq t) dt.$$

On procède alors comme dans la preuve de la proposition 4.10 en injectant dans cette formule l'égalité $P(X \leq t) = \int_{-\infty}^t f(x) dx$ et en justifiant l'interversion des intégrations par le théorème de Fubini-Tonelli.

$$\begin{aligned} \int_{-\infty}^0 P(X \leq t) dt &= \int_{-\infty}^0 \left\{ \int_{-\infty}^t f(x) dx \right\} dt \\ &= \int_{-\infty}^0 \left\{ \int_{-\infty}^0 \mathbf{1}_{]-\infty, t]}(x) f(x) dx \right\} dt \\ &= \int_{-\infty}^0 \left\{ \int_{-\infty}^0 \mathbf{1}_{[x, 0]}(t) f(x) dt \right\} dx \\ &= \int_{-\infty}^0 \left\{ \int_{-\infty}^0 \mathbf{1}_{[x, 0]}(t) dt \right\} f(x) dx. \end{aligned}$$

Comme pour $x \leq 0$, $\int_{-\infty}^0 \mathbf{1}_{[x, 0]}(t) dt = \int_x^0 dt = -x$, on obtient finalement :

$$\int_0^{+\infty} P(X^- > t) dt = \int_{-\infty}^0 (-x)f(x) dx = \int_{-\infty}^0 |x|f(x) dx. \quad (4.46)$$

En appliquant l'additivité de l'espérance des v.a. positives à l'égalité $|X| = X^+ + X^-$ et en rassemblant (4.45) et (4.46), on voit que

$$\mathbf{E}|X| = \mathbf{E}(X^+) + \mathbf{E}(X^-) = \int_0^{+\infty} |x|f(x) dx + \int_{-\infty}^0 |x|f(x) dx = \int_{-\infty}^{+\infty} |x|f(x) dx,$$

ce qui établit la condition nécessaire et suffisante d'intégrabilité (4.43). Si cette condition est réalisée, $\mathbf{E}(X^+)$ et $\mathbf{E}(X^-)$ sont finis et

$$\mathbf{E}X = \mathbf{E}(X^+) - \mathbf{E}(X^-) = \int_0^{+\infty} xf(x) dx - \int_{-\infty}^0 (-x)f(x) dx = \int_{-\infty}^{+\infty} xf(x) dx,$$

ce qui établit (4.44). □

Proposition 4.33 (espérance d'une v.a. discrète). *La variable aléatoire discrète X est intégrable si et seulement si*

$$\sum_{x \in X(\Omega)} |x|P(X = x) < +\infty \quad (4.47)$$

et dans ce cas,

$$\mathbf{E}X = \sum_{x \in X(\Omega)} xP(X = x), \quad (4.48)$$

cette somme désignant une série absolument convergente dans le cas où $X(\Omega)$ est infini¹⁷.

Preuve. X étant discrète, $X(\Omega)$ est au plus dénombrable. Notons $B := X(\Omega) \cap]0, +\infty[$ et $B' := X(\Omega) \cap]-\infty, 0[$. Alors $X^+(\Omega)$ est égal à B si tous les $x \in X(\Omega)$ sont strictement positifs ou à $B \cup \{0\}$ si l'un au moins d'entre eux est négatif ou nul¹⁸. Quoiqu'il en soit, on a toujours en utilisant le corollaire 4.27 :

$$\mathbf{E}(X^+) = \sum_{x \in X^+(\Omega)} xP(X^+ = x) = \sum_{x \in X^+(\Omega)} xP(X = x) = \sum_{x \in B} xP(X = x), \quad (4.49)$$

puisque pour $x = 0$ le terme éventuel $xP(X = x)$ est nul. De même, $X^-(\Omega) = \{-x; x \in B'\} =: -B'$ si tous les $x \in X(\Omega)$ sont strictement négatifs ou à $(-B') \cup \{0\}$ si l'un au moins d'entre eux est positif ou nul. En appliquant le corollaire 4.27 à la v.a. positive discrète X^- , on a :

$$\mathbf{E}(X^-) = \sum_{y \in X^-(\Omega)} yP(X^- = y) = \sum_{x \in B'} (-x)P(X = x) = \sum_{x \in B'} |x|P(X = x). \quad (4.50)$$

En rassemblant (4.49) et (4.50), on obtient (noter que $B \cap B' = \emptyset$) :

$$\mathbf{E}|X| = \sum_{x \in B} xP(X = x) + \sum_{x \in B'} |x|P(X = x) = \sum_{x \in B \cup B'} |x|P(X = x) = \sum_{x \in X(\Omega)} |x|P(X = x),$$

ce qui nous donne la CNS d'intégrabilité (4.47). Si cette condition est réalisée, $\mathbf{E}(X^+)$ et $\mathbf{E}(X^-)$ sont finis et

$$\begin{aligned} \mathbf{E}X &= \mathbf{E}(X^+) - \mathbf{E}(X^-) = \sum_{x \in B} xP(X = x) - \sum_{x \in B'} (-x)P(X = x) \\ &= \sum_{x \in B \cup B'} xP(X = x) \\ &= \sum_{x \in X(\Omega)} xP(X = x), \end{aligned}$$

ce qui établit (4.48). □

17. $X(\Omega)$ est alors forcément dénombrable puisque X est discrète.

18. Il en résulte que $X^+(\Omega)$ est au plus dénombrable et donc que la v.a. positive X^+ est discrète.

Proposition 4.34 (Linéarité de l'espérance).

a) L'espérance des v.a. réelles intégrables est additive : si X et Y v.a. réelles définies sur le même (Ω, \mathcal{F}, P) sont intégrables, alors $X + Y$ l'est aussi et

$$\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y. \quad (4.51)$$

b) Si X est intégrable, cX l'est aussi pour toute constante réelle c et

$$\mathbf{E}(cX) = c\mathbf{E}X. \quad (4.52)$$

Preuve du a). Posons $Z := X + Y$. Alors Z est mesurable et compte-tenu de la croissance et de l'additivité de l'intégrale pour les v.a. positives, l'inégalité $|Z| \leq |X| + |Y|$ nous donne :

$$\mathbf{E}|Z| \leq \mathbf{E}(|X| + |Y|) = \mathbf{E}|X| + \mathbf{E}|Y| < +\infty.$$

Ainsi Z est intégrable et $\mathbf{E}Z$ bien définie.

L'égalité $Z^+ - Z^- = X^+ - X^- + Y^+ - Y^-$, peut se réécrire comme égalité entre sommes de v.a. positives :

$$Z^+ + X^- + Y^- = Z^- + X^+ + Y^+.$$

Par additivité de l'espérance pour les v.a. positives, on en déduit

$$\mathbf{E}(Z^+) + \mathbf{E}(X^-) + \mathbf{E}(Y^-) = \mathbf{E}(Z^-) + \mathbf{E}(X^+) + \mathbf{E}(Y^+).$$

La finitude de ces 6 espérances (due à l'intégrabilité de X , Y et Z) nous permet alors d'écrire :

$$\mathbf{E}(Z^+) - \mathbf{E}(Z^-) = \mathbf{E}(X^+) - \mathbf{E}(X^-) + \mathbf{E}(Y^+) - \mathbf{E}(Y^-),$$

ce qui donne bien $\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y$. □

Preuve du b). L'intégrabilité de cX résulte de l'intégrabilité de X et de la proposition 4.12 appliquée avec la constante positive $|c|$. En effet

$$\mathbf{E}|cX| = \mathbf{E}(|c||X|) = |c|\mathbf{E}|X| < +\infty.$$

Si $c > 0$, on voit que $(cX)^+ = cX^+$ et $(cX)^- = cX^-$. En utilisant la proposition 4.12, on en déduit :

$$\mathbf{E}(cX) = \mathbf{E}((cX)^+) - \mathbf{E}((cX)^-) = \mathbf{E}(cX^+) - \mathbf{E}(cX^-) = c\mathbf{E}(X^+) - c\mathbf{E}(X^-) = c\mathbf{E}X.$$

Si $c < 0$, $(cX)^+ = (-c)X^-$ et $(cX)^- = (-c)X^+$, d'où en utilisant la proposition 4.12 avec la constante positive $(-c)$:

$$\begin{aligned} \mathbf{E}(cX) &= \mathbf{E}((cX)^+) - \mathbf{E}((cX)^-) = \mathbf{E}((-c)X^-) - \mathbf{E}((-c)X^+) \\ &= (-c)\mathbf{E}(X^-) - (-c)\mathbf{E}(X^+) = c\mathbf{E}X. \end{aligned}$$

Le cas $c = 0$ est évident directement. □

Proposition 4.35 (espérance et ordre).

- a) L'espérance des v.a. réelles intégrables est croissante : si X et Y v.a. réelles définies sur le même (Ω, \mathcal{F}, P) sont intégrables et vérifient $X \leq Y$, i.e. pour tout $\omega \in \Omega$, $X(\omega) \leq Y(\omega)$, alors $\mathbf{E}X \leq \mathbf{E}Y$.
- b) Si X est intégrable, $|X|$ l'est aussi et

$$|\mathbf{E}X| \leq \mathbf{E}|X|. \quad (4.53)$$

Preuve. Pour le a), il suffit de noter que si $X \leq Y$, $Y - X$ est une v.a. positive et donc $\mathbf{E}(Y - X) \geq 0$. Comme X et Y sont intégrables, par linéarité de l'espérance, $\mathbf{E}(Y - X) = \mathbf{E}Y - \mathbf{E}X$. Ainsi la positivité de $Y - X$ et la linéarité de \mathbf{E} impliquent $\mathbf{E}Y - \mathbf{E}X \geq 0$ c'est-à-dire $\mathbf{E}Y \geq \mathbf{E}X$.

Une fois la croissance de \mathbf{E} ainsi établie, on en déduit le b) en partant de l'encadrement $-|X| \leq X \leq |X|$ qui implique $\mathbf{E}(-|X|) \leq \mathbf{E}X \leq \mathbf{E}|X|$. Par linéarité $\mathbf{E}(-|X|) = -\mathbf{E}|X|$, ce qui permet de réécrire l'encadrement ci-dessus sous la forme :

$$-\mathbf{E}|X| \leq \mathbf{E}X \leq \mathbf{E}|X|,$$

qui équivaut à (4.53). □

4.4 Moments

On étudie dans cette section les $\mathbf{E}h(X)$, où h est une fonction réelle. Pour que l'expression $\mathbf{E}h(X)$ ait un sens, il est nécessaire que $Y := h(X)$ soit une variable aléatoire réelle. Cette condition sera réalisée si $h : \mathbb{R} \rightarrow \mathbb{R}$ est borélienne, i.e. $B' = h^{-1}(B) \in \text{Bor}(\mathbb{R})$ pour tout $B \in \text{Bor}(\mathbb{R})$. En effet on a alors

$$Y^{-1}(B) = \{\omega \in \Omega; h(X(\omega)) \in B\} = \{\omega \in \Omega; X(\omega) \in h^{-1}(B)\} = X^{-1}(B') \in \mathcal{F},$$

en raison de la mesurabilité $\mathcal{F} - \text{Bor}(\mathbb{R})$ de la variable aléatoire X . Comme le borélien B ci-dessus est quelconque, ceci montre que Y est elle aussi mesurable $\mathcal{F} - \text{Bor}(\mathbb{R})$, i.e. que c'est une variable aléatoire sur (Ω, \mathcal{F}) . Ainsi si h est borélienne, $\mathbf{E}|h(X)|$ existe toujours comme élément de $\overline{\mathbb{R}}_+$ et si $\mathbf{E}|h(X)| < +\infty$, $\mathbf{E}h(X)$ existe (et $\mathbf{E}h(X) \in \mathbb{R}$). On pourra désigner $\mathbf{E}|h(X)|$ et $\mathbf{E}h(X)$ respectivement par l'appellation h -moment absolu de X et h -moment de X . Bien entendu si h est borélienne positive, h -moment absolu et h -moment sont confondus et ce dernier existe toujours dans $\overline{\mathbb{R}}_+$. Nous utiliserons aussi l'appellation générique de *moments fonctionnels* pour désigner les h -moments¹⁹.

Le cas le plus utile est celui où h est une fonction puissance, $h(x) = x^r$, on parle alors de moment d'ordre r de X .

Définition 4.36. Soit r un réel positif. On appelle moment absolu d'ordre r de la variable aléatoire réelle X la quantité $\mathbf{E}(|X|^r)$ (élément de $\overline{\mathbb{R}}_+$). Si r est entier et X^r

¹⁹. Attention, ces appellations ne sont pas standard, nous les adoptons pour des raisons de confort de rédaction.

intégrable (donc si le moment absolu d'ordre r de X est fini), on appelle moment d'ordre r de X le réel $\mathbf{E}(X^r)$. On notera $\mathbf{E}|X|^r$ pour $\mathbf{E}(|X|^r)$ et $\mathbf{E}X^r$ pour $\mathbf{E}(X^r)$ en prenant garde de ne pas confondre ces quantités avec $(\mathbf{E}|X|)^r$ et $(\mathbf{E}X)^r$ respectivement.

Proposition 4.37. *Si la variable aléatoire X a un moment absolu d'ordre r fini, elle a aussi un moment absolu d'ordre p fini pour tout $p \in [0, r]$.*

Preuve. Si $0 \leq p \leq r$, on a les inégalités suivantes entre variables aléatoires positives :

$$|X|^p \leq \mathbf{1}_{\{|X| \leq 1\}} + |X|^r \mathbf{1}_{\{|X| > 1\}} \leq 1 + |X|^r.$$

Par croissance de l'espérance des v.a. positives, on en déduit

$$\mathbf{E}(|X|^p) \leq \mathbf{E}(1) + \mathbf{E}(|X|^r) = 1 + \mathbf{E}(|X|^r) < +\infty.$$

□

L'existence d'un moment absolu d'ordre r fini donne un renseignement sur la vitesse de convergence vers 0 de $P(|X| \geq t)$ quand t tend vers $+\infty$. On a alors $P(|X| \geq t) = O(t^{-r})$ par le corollaire suivant de l'inégalité de Markov.

Proposition 4.38 (inégalité de Markov avec moment). *Pour toute variable aléatoire réelle X , on a pour tout $r > 0$,*

$$\forall t > 0, \quad P(|X| \geq t) \leq \frac{\mathbf{E}|X|^r}{t^r}. \quad (4.54)$$

Bien entendu, cette inégalité n'a d'intérêt que si $\mathbf{E}|X|^r < +\infty$ et pour t tel que $t^{-r}\mathbf{E}|X|^r < 1$.

Preuve. En utilisant la croissance de l'application $\mathbb{R}_+ \rightarrow \mathbb{R}_+$, $x \mapsto x^r$ et l'inégalité de Markov pour la v.a. positive $|X|^r$, on obtient :

$$P(|X| \geq t) \leq P(|X|^r \geq t^r) \leq \frac{\mathbf{E}|X|^r}{t^r}.$$

□

Proposition 4.39 (moments d'une v.a. discrète). *Si X est une variable aléatoire discrète, on a pour tout réel $r \geq 0$*

$$\mathbf{E}|X|^r = \sum_{x \in X(\Omega)} |x|^r P(X = x). \quad (4.55)$$

Si r est entier et $\mathbf{E}|X|^r < +\infty$, on a aussi

$$\mathbf{E}X^r = \sum_{x \in X(\Omega)} x^r P(X = x). \quad (4.56)$$

Cette proposition n'est qu'une application de la formule de calcul de $\mathbf{E}h(X)$ pour X discrète qui est établie dans toute sa généralité à la proposition 4.41 ci-dessous.

Proposition 4.40 (moments d'une v.a. à densité). *Si X est une variable aléatoire réelle à densité f , on a pour tout réel $r \geq 0$*

$$\mathbf{E}|X|^r = \int_{-\infty}^{+\infty} |x|^r f(x) dx. \quad (4.57)$$

Si r est entier et $\mathbf{E}|X|^r < +\infty$, on a aussi

$$\mathbf{E}X^r = \int_{-\infty}^{+\infty} x^r f(x) dx. \quad (4.58)$$

Là aussi il s'agit d'un cas particulier d'une formule générale pour $\mathbf{E}h(X)$ lorsque X est à densité, donnée ci-dessous. La démonstration de cette formule générale étant assez ardue, nous invitons le lecteur à démontrer directement en exercice la proposition 4.40. La preuve est grandement facilitée par la monotonie de $h : x \mapsto |x|^r$ sur chacun des ensembles \mathbb{R}_- et \mathbb{R}_+ .

Proposition 4.41 (calcul de $\mathbf{E}h(X)$, X discrète). *Soit X une variable aléatoire discrète et $h : \mathbb{R} \rightarrow \mathbb{R}$ une application borélienne (i.e. mesurable $\text{Bor}(\mathbb{R})$ - $\text{Bor}(\mathbb{R})$). Alors*

$$\mathbf{E}|h(X)| = \sum_{x \in X(\Omega)} |h(x)| P(X = x). \quad (4.59)$$

Si $\mathbf{E}|h(X)| < +\infty$, ce qui équivaut à $\sum_{x \in X(\Omega)} |h(x)| P(X = x) < +\infty$, on a de plus

$$\mathbf{E}h(X) = \sum_{x \in X(\Omega)} h(x) P(X = x). \quad (4.60)$$

Preuve. Posons $Y = |h(X)|$. L'ensemble $X(\Omega) = \{x_0, x_1, \dots\}$ est au plus dénombrable. L'application $|h|$ n'étant pas supposée injective, il peut y avoir des répétitions dans la suite des $|h(x_k)|$. L'ensemble $Y(\Omega)$ qui peut s'écrire en effaçant toutes les répétitions de cette suite est lui-même au plus dénombrable. La variable aléatoire $Y = |h(X)|$ étant discrète positive, la formule établie au corollaire 4.27 nous donne :

$$\mathbf{E}|h(X)| = \mathbf{E}Y = \sum_{y \in Y(\Omega)} y P(Y = y). \quad (4.61)$$

Pour chaque réel y de $Y(\Omega)$, notons B_y l'ensemble de ses antécédents par $|h|$:

$$B_y = \{x \in X(\Omega); |h(x)| = y\}, \quad y \in Y(\Omega).$$

Ce sous-ensemble de $X(\Omega)$ contient au moins un élément et est au plus dénombrable. On peut alors décomposer l'événement $\{Y = y\}$ en réunion disjointe (d'une famille au plus dénombrable) :

$$\{Y = y\} = \bigcup_{x \in B_y} \{X = x\}.$$

Le terme général de la série (4.61) peut donc s'écrire :

$$yP(Y = y) = y \sum_{x \in B_y} P(X = x) = \sum_{x \in B_y} yP(X = x) = \sum_{x \in B_y} |h(x)|P(X = x).$$

Comme les B_y forment une partition de $X(\Omega)$, on a en utilisant la propriété de sommation par paquets des séries à termes *positifs* :

$$\sum_{x \in X(\Omega)} |h(x)|P(X = x) = \sum_{y \in Y(\Omega)} \sum_{x \in B_y} |h(x)|P(X = x) = \sum_{y \in Y(\Omega)} yP(Y = y),$$

ce qui, compte-tenu de (4.61), établit la formule (4.59) pour $\mathbf{E}|h(X)|$.

Supposons maintenant que $\mathbf{E}|h(X)| < +\infty$. Alors grâce à (4.59), on voit que la famille de réels $\{h(x)P(X = x); x \in X(\Omega)\}$ est *sommable*. On peut alors reprendre le calcul fait ci-dessus en remplaçant partout $|h|$ par h , car les séries concernées sont absolument convergentes et on peut utiliser la propriété de sommation par paquets des familles sommables. On aboutit ainsi à la formule (4.60) pour $\mathbf{E}h(X)$. \square

Proposition 4.42 (calcul de $\mathbf{E}h(X)$, X à densité). *Soit X une variable aléatoire de densité f et $h : \mathbb{R} \rightarrow \mathbb{R}$ une application réglée sur tout intervalle fermé borné de \mathbb{R} . Alors*

$$\mathbf{E}|h(X)| = \int_{-\infty}^{+\infty} |h(x)|f(x) dx. \quad (4.62)$$

Si $\mathbf{E}|h(X)| < +\infty$, on a de plus

$$\mathbf{E}h(X) = \int_{-\infty}^{+\infty} h(x)f(x) dx. \quad (4.63)$$

Une application h est réglée sur $[a, b]$ si elle est limite uniforme sur $[a, b]$ d'une suite de fonctions en escaliers²⁰. On démontre en analyse que h est réglée sur $[a, b]$ si et seulement si elle admet en tout point de $]a, b[$ une limite à gauche et une limite à droite (finies) ainsi qu'une limite à droite finie en a et à gauche finie en b . La classe des fonctions réglées, sans être aussi grande que celle des fonctions boréliennes, devrait donc suffire à nos besoins. Elle contient en particulier les fonctions continues et les fonctions monotones par morceaux.

Schéma de la preuve. On procède en 5 étapes.

1. On examine d'abord le cas de h en escaliers et positive sur $[a, b]$, nulle en dehors de $[a, b]$. La fonction h peut alors s'écrire $h = \sum_{i=1}^m y_i \mathbf{1}_{A_i}$, où les y_i sont des réels positifs et les A_i sont des intervalles formant une partition de $[a, b]$. Par linéarité, on se ramène au cas encore plus simple où $h = \mathbf{1}_{A_i}$, lequel est évident, puisque $\int_{-\infty}^{+\infty} \mathbf{1}_{A_i}(x)f(x) dx = P(X \in A_i) = \mathbf{E}\mathbf{1}_{A_i}(X)$.

20. On en déduit ici la mesurabilité de h en montrant qu'elle est limite simple sur \mathbb{R} d'une suite de fonctions en escalier, donc boréliennes. Ainsi $h(X)$ est bien une variable aléatoire car mesurable \mathcal{F} – $\text{Bor}(\mathbb{R})$ par composition.

2. On traite ensuite le cas où h est réglée et positive sur un intervalle fixé $[a, b]$, et nulle en dehors de cet intervalle. Elle est donc limite uniforme sur $[a, b]$ d'une suite $(g_k)_{k \geq 1}$ de fonctions en escaliers positives sur $[a, b]$ et nulles en dehors de $[a, b]$. Comme les g_k et h sont nulles en dehors de $[a, b]$, cette convergence est uniforme sur \mathbb{R} et on en déduit (exercice) que $\mathbf{E}h(X) = \lim_{k \rightarrow \infty} \mathbf{E}g_k(X)$. Notons au passage que h est bornée, donc que la variable aléatoire positive $h(X)$ a une espérance finie. D'autre part la convergence uniforme sur $[a, b]$ de g_k vers h permet de légitimer l'interversion limite intégrale $\int_a^b h(x)f(x) dx = \int_a^b \lim_{k \rightarrow +\infty} g_k(x)f(x) dx = \lim_{k \rightarrow +\infty} \int_a^b g_k(x)f(x) dx = \lim_{k \rightarrow +\infty} \mathbf{E}g_k(X)$ et on obtient ainsi (4.63) pour h réglée positive sur $[a, b]$ et nulle en dehors²¹.
3. Pour h positive réglée sur tout $[a, b] \subset \mathbb{R}$, on pose $h_n := h\mathbf{1}_{[-n, n]}$ et on note que $h_n \uparrow h$. On étend alors le résultat du cas 2 (valable pour chaque h_n) en utilisant le théorème de Beppo Levi (Th. 4.23).
4. En appliquant ce qui précède avec $|h|$ au lieu de h , on obtient (4.62) pour h réglée sur tout $[a, b] \subset \mathbb{R}$ sans condition de positivité.
5. On obtient enfin (4.63) en écrivant $h = h^+ - h^-$, en utilisant le cas 3 et en recollant les morceaux, sans problème puisqu'ici les intégrales généralisées et les espérances concernées sont toutes finies.

□

Nous présentons maintenant une formule injustement méconnue, permettant de calculer des moments fonctionnels à partir de la fonction de survie. Son intérêt est de permettre un tel calcul pour des variables aléatoires qui ne sont ni discrètes ni à densité.

Proposition 4.43. *Soient X une variable aléatoire positive et g une application continue strictement croissante $\mathbb{R}_+ \rightarrow \mathbb{R}$, de classe C^1 sur \mathbb{R}_+^* . Alors*

$$\mathbf{E}g(X) = g(0) + \int_0^{+\infty} P(X > s)g'(s) ds. \quad (4.64)$$

La preuve de cette formule utilise le lemme suivant, valide sous des hypothèses plus générales pour g .

Lemme 4.44. *Soit $g : [0, +\infty[\rightarrow \mathbb{R}$ une application continue et strictement croissante sur $[0, +\infty[$. Notons ℓ la limite de g en $+\infty$. Alors pour toute variable aléatoire positive X ,*

$$\forall t \in [g(0), \ell[, \quad P(g(X) > t) = P(X > g^{-1}(t)), \quad (4.65)$$

Preuve du lemme. En raison de sa croissance, g a une limite ℓ en $+\infty$, qui est soit un réel, soit $+\infty$. En raison de sa continuité et de sa croissance stricte, g réalise une bijection de $[0, +\infty[$ sur $[g(0), \ell[$ et la réciproque g^{-1} de cette bijection est définie et strictement

21. Attention, l'interversion limite intégrale ne s'obtient pas directement par convergence uniforme de $g_k f$ vers $h f$, car f n'est pas forcément définie sur tout $[a, b]$ ni bornée.

croissante sur $[g(0), \ell]$. Pour prouver (4.65), nous allons établir l'égalité des évènements $A := \{g(X) > t\}$ et $B := \{X > g^{-1}(t)\}$ en vérifiant l'inclusion dans les deux sens.

Pour tout $\omega \in A$, $g(X(\omega)) > t$ et, comme les deux membres de cette inégalité appartiennent à $[g(0), \ell]$ qui est l'ensemble de définition de l'application strictement croissante g^{-1} , on en déduit que $g^{-1}(g(X(\omega))) > g^{-1}(t)$, ce qui s'écrit aussi $X(\omega) > g^{-1}(t)$. Ainsi tout ω dans A appartient aussi à B , autrement dit, $A \subset B$.

En échangeant les rôles de g et g^{-1} , le même argument montre que $B \subset A$. Finalement $A = B$ et $P(A) = P(B)$. \square

Preuve de la proposition 4.43. On ne perd pas de généralité en supposant que g est positive. En effet par croissance de g , $g - g(0)$ est positive et il est clair que si (4.64) est vraie pour $g - g(0)$, elle est vraie pour g .

Pour établir (4.64) avec g positive, on définit la variable aléatoire positive $Y = g(X)$ et on part de la définition $\mathbf{E}Y = \int_0^{+\infty} P(Y > t) dt$ en découpant $\int_0^{+\infty} = \int_0^{g(0)} + \int_{g(0)}^{\ell}$ si $\ell = +\infty$, ou $\int_0^{+\infty} = \int_0^{g(0)} + \int_{g(0)}^{\ell} + \int_{\ell}^{+\infty}$ si $\ell < +\infty$.

Remarquons d'abord que si $\ell < +\infty$, l'application strictement croissante g ne peut prendre que des valeurs strictement inférieures à ℓ et par conséquent pour tout $t \geq \ell$, $\{g(X) > t\} = \emptyset$, d'où $P(Y > t) = P(\emptyset) = 0$ et $\int_{\ell}^{+\infty} P(Y > t) dt = 0$. Que ℓ soit fini ou non, on a donc dans tous les cas :

$$\mathbf{E}Y = \int_0^{g(0)} P(Y > t) dt + \int_{g(0)}^{\ell} P(Y > t) dt.$$

Examinons maintenant la première intégrale. Puisque g est croissante, on a pour tout $\omega \in \Omega$, $Y(\omega) = g(X(\omega)) \geq g(0)$ et donc pour tout $t \in [0, g(0)[$, $1 \geq P(Y > t) \geq P(Y \geq g(0)) = 1$. La fonction $t \mapsto P(Y > t)$ est donc constante égale à 1 sur $[0, g(0)[$ et sa valeur en $g(0)$ est un réel²² de $[0, 1]$. L'intégrale de Riemann $\int_0^{g(0)} P(Y > t) dt$ vaut donc $\int_0^{g(0)} 1 dt = g(0)$. À ce stade nous avons donc établi la formule :

$$\mathbf{E}g(X) = g(0) + \int_{g(0)}^{\ell} P(Y > t) dt.$$

En utilisant (4.65) on peut écrire cette dernière intégrale sous la forme :

$$\int_{g(0)}^{\ell} P(Y > t) dt = \int_{g(0)}^{\ell} P(X > g^{-1}(t)) dt.$$

On peut alors appliquer le changement de variable C^1 et croissant²³ (cf. prop. A.33) $t = g(s)$ ou $s = g^{-1}(t)$ qui donne :

$$\int_{g(0)}^{\ell} P(Y > t) dt = \int_0^{+\infty} P(X > s)g'(s) ds$$

22. Cette valeur peut être strictement inférieure à 1 si $P(X = 0) \neq 0$. Mais en tout état de cause, elle n'a aucune influence sur la valeur de l'intégrale de Riemann $\int_0^{g(0)} P(Y > t) dt$.

23. Pour le justifier proprement, on peut d'abord faire le changement de variable sur $\int_{g(0)}^{g(b)}$, puis faire tendre b vers $+\infty$.

et achève la vérification de (4.64). \square

Le h -moment $\mathbf{E}h(X)$ pour $h : x \mapsto (x - \mathbf{E}X)^2$ occupe une place particulière dans la théorie des probabilités.

Définition 4.45 (variance et écart type). *Si X est de carré intégrable (i.e. $\mathbf{E}X^2 < +\infty$), on appelle variance de X le réel positif noté $\text{Var } X$ défini par*

$$\text{Var } X := \mathbf{E}(X - \mathbf{E}X)^2. \quad (4.66)$$

On appelle alors écart type de X le réel $\sigma(X) := (\text{Var } X)^{1/2}$.

Remarquons que si $\mathbf{E}X^2$ est fini, $\mathbf{E}|X|$ l'est aussi (proposition 4.37), donc $\mathbf{E}X$ est bien défini. De plus $(X - \mathbf{E}X)^2 = X^2 - 2(\mathbf{E}X)X + (\mathbf{E}X)^2$ apparaît alors comme une combinaison linéaire de trois variables²⁴ intégrables, donc est aussi intégrable. Ainsi la v.a. positive $(X - \mathbf{E}X)^2$ est intégrable et $\mathbf{E}(X - \mathbf{E}X)^2$ est bien un réel positif, ce qui justifie la définition 4.45. Notons aussi que si X représente une grandeur physique, X , $\mathbf{E}X$ et $\sigma(X)$ ont la même unité, mais pas $\text{Var } X$.

Lorsqu'elle existe, la variance de X est une façon de mesurer la *dispersion* de la loi de X autour de l'espérance. Les raisons de l'importance de la variance apparaîtront ultérieurement dans ce cours (inégalité de Tchebycheff, théorème limite central). L'application des propositions 4.41 et 4.42 nous donne (sous réserve d'intégrabilité de X^2) les formules respectives :

$$\text{Var } X = \sum_{x \in X(\Omega)} (x - \mathbf{E}X)^2 P(X = x) \quad (\text{cas } X \text{ discrète}), \quad (4.67)$$

$$\text{Var } X = \int_{-\infty}^{+\infty} (x - \mathbf{E}X)^2 f(x) dx \quad (\text{cas } X \text{ à densité } f). \quad (4.68)$$

Dans la pratique, ces formules sont rarement utilisées, on leur préfère la formule suivante qui simplifie les calculs.

Proposition 4.46 (formule de Koenig pour la variance). *Si la variable aléatoire X est de carré intégrable,*

$$\text{Var } X = \mathbf{E}X^2 - (\mathbf{E}X)^2. \quad (4.69)$$

Preuve. Rappelons que nous notons $\mathbf{E}X^2$ pour $\mathbf{E}(X^2)$ et que le second membre de la formule ci-dessus n'est donc généralement pas nul. On pose $c = \mathbf{E}X$.

$$\begin{aligned} \text{Var } X &= \mathbf{E}(X - c)^2 = \mathbf{E}[X^2 - 2cX + c^2] \\ &= \mathbf{E}X^2 - 2c\mathbf{E}X + \mathbf{E}c^2 \\ &= \mathbf{E}X^2 - 2c^2 + c^2 = \mathbf{E}X^2 - c^2, \end{aligned}$$

en utilisant la linéarité de l'espérance et l'espérance d'une constante. \square

24. À savoir X^2 , X et la v.a. constante $(\mathbf{E}X)^2$.

Proposition 4.47 (translation et changement d'échelle). *Si X a un moment d'ordre 2,*

$$\forall a \in \mathbb{R}, \forall b \in \mathbb{R}, \quad \text{Var}(aX + b) = a^2 \text{Var } X, \quad \sigma(aX + b) = |a|\sigma(X). \quad (4.70)$$

Preuve. On a en utilisant la définition 4.45, la linéarité de l'espérance et le fait que l'espérance de la constante b est b :

$$\begin{aligned} \text{Var}(aX + b) &= \mathbf{E}[aX + b - \mathbf{E}(aX + b)]^2 = \mathbf{E}[aX + b - a\mathbf{E}X - b]^2 \\ &= \mathbf{E}[a(X - \mathbf{E}X)]^2 = \mathbf{E}[a^2(X - \mathbf{E}X)^2] \\ &= a^2 \mathbf{E}[(X - \mathbf{E}X)^2] = a^2 \text{Var } X. \end{aligned}$$

□

Il est clair d'après la définition de la variance que la variance d'une constante est nulle. La réciproque est *presque* vraie :

Proposition 4.48 (nullité de la variance et constance p.s.).

$$\text{Var } X = 0 \Leftrightarrow X = \mathbf{E}X \text{ p.s.} \Leftrightarrow X \text{ est presque sûrement constante.} \quad (4.71)$$

Preuve. Les implications de droite à gauche dans (4.71) sont déjà acquises. On sait en effet que l'espérance d'une constante est cette constante et que si la v.a. $Y := (X - \mathbf{E}X)^2$ vaut 0 avec probabilité 1, son espérance est nulle²⁵.

Pour la première implication de gauche à droite, il suffit d'appliquer le corollaire 4.18 à la v.a. positive Y . La deuxième implication est triviale. □

25. En effet, pour tout $t \geq 0$, $P(Y > t) = 0$, donc Y étant positive, $\mathbf{E}Y = \int_0^{+\infty} P(Y > t) dt = 0$.

Chapitre 5

Vecteurs aléatoires et indépendance

L'information pertinente résultant d'une expérience aléatoire ne se résume pas toujours à la valeur prise par une seule variable aléatoire réelle. On a souvent besoin de connaître les valeurs d'une suite finie de variables aléatoires. Par exemple au jeu de 421, on lance trois dés et on a besoin de connaître les points affichés par chacun des dés, le résultat sera donc décrit par un vecteur $(X_1(\omega), X_2(\omega), X_3(\omega))$. Si on tire sur une cible, le résultat sera décrit par les coordonnées $(X(\omega), Y(\omega))$ du point d'impact. Si on étudie le fonctionnement d'un guichet en observant les n premiers clients, le résultat sera décrit par la suite des $X_1, Y_1, Z_1, X_2, Y_2, Z_2, \dots, X_n, Y_n, Z_n$ où X_i est le temps d'attente au guichet du i^{e} client, Y_i son temps de service et Z_i le temps s'écoulant entre le départ du i^{e} client et l'arrivée du $(i+1)^{\text{e}}$. Ces suites finies de variables aléatoires sont appelées des *vecteurs aléatoires*. De même qu'une variable aléatoire peut être vue comme un procédé de choix d'un nombre réel au hasard, un vecteur aléatoire de dimension d , $X = (X_1, \dots, X_d)$ est un procédé de choix au hasard d'un point de \mathbb{R}^d . Ses composantes X_1, \dots, X_d sont alors autant de variables aléatoires réelles. Arrivé là, l'étudiant inquiet du nombre de pages de ce document qu'il lui reste à lire avant l'examen se demande légitimement s'il y a un intérêt à consacrer tout un chapitre aux vecteurs aléatoires puisque ces objets ne sont que des suites finies de variables aléatoires et que ces dernières sont maintenant bien connues¹. L'intérêt de cette étude repose sur la remarque informelle suivante à laquelle nous donnerons bientôt un sens mathématique précis : la *connaissance probabiliste globale du vecteur* $X = (X_1, \dots, X_d)$ *apporte davantage d'information que la connaissance probabiliste individuelle de chacune de ses composantes* X_i . Au premier abord, cette idée peut paraître choquante car une lecture rapide de la phrase précédente laisse croire que la connaissance de $X(\omega)$ apporte quelque chose de plus que celle de tous les $X_i(\omega)$, $i = 1, \dots, d$, ce qui n'est évidemment pas vrai. La clé de l'énigme est dans l'expression « *connaissance probabiliste* » que nous remplacerons bientôt par connaissance de la *loi*, dès que nous aurons défini la loi d'un vecteur aléatoire. En attendant voici une image qui peut nous aider à comprendre ce dont il s'agit. Considérons un ensemble de 10 coureurs de fond, chacun muni d'un dossard numéroté de 1 à 10. Si on les rassemble sur une même piste pour une épreuve de 5 000 mètres, on peut représenter le résultat de la

1. Phrase écrite après une période de deux mois sans lecture d'un paquet de copies...

course par le vecteur (X_1, \dots, X_{10}) , où X_i désigne le temps mis par le coureur numéroté i pour parcourir les 5 000 mètres. Tout amateur d'athlétisme sait bien que cette expérience n'est pas équivalente à faire courir isolément un 5 000 mètres à chacun des 10 coureurs sur des stades séparés. La différence ici vient de la compétition, de la tactique de course, etc. Par contre dans d'autres situations, le comportement global du vecteur des d composantes se réduit au comportement individuel de chacune d'elles. On admet généralement que c'est le cas lorsqu'on lance trois dés en considérant qu'il revient au même de les lancer ensemble sur la même table ou séparément sur trois tables. On parle alors d'indépendance des composantes. Cette notion d'indépendance des composantes X_i du vecteur aléatoire X est liée à celle d'une suite d'évènements A_i , où la réalisation ou non de A_i ne dépend que des valeurs de X_i . L'étude des suites finies de variables aléatoires indépendantes prend donc place naturellement dans ce chapitre comme le cas particulier des vecteurs aléatoires à composantes indépendantes.

Sauf mention explicite du contraire, toutes les variables aléatoires et tous les vecteurs aléatoires considérés dans ce chapitre seront définis sur le même espace probabilisé (Ω, \mathcal{F}, P) .

5.1 Vecteurs aléatoires

5.1.1 Généralités

Définition 5.1 (vecteur aléatoire de \mathbb{R}^d). *On dit que l'application $X : \Omega \rightarrow \mathbb{R}^d$ est un vecteur aléatoire sur (Ω, \mathcal{F}) si c'est une application mesurable $\mathcal{F} - \text{Bor}(\mathbb{R}^d)$.*

Proposition 5.2. *Soit X un vecteur aléatoire sur (Ω, \mathcal{F}) et P une probabilité sur (Ω, \mathcal{F}) . La fonction d'ensembles $P_X = P \circ X^{-1}$ définie sur $\text{Bor}(\mathbb{R}^d)$ par*

$$\forall B \in \text{Bor}(\mathbb{R}^d), \quad P_X(B) := P(X^{-1}(B)) = P(X \in B) \quad (5.1)$$

est une probabilité sur $(\mathbb{R}^d, \text{Bor}(\mathbb{R}^d))$.

La preuve est exactement la même que celle de la proposition 3.7, en remplaçant les boréliens de \mathbb{R} par ceux de \mathbb{R}^d . Comme pour les variables aléatoires, cette proposition légitime la définition de la loi de X sous P .

Définition 5.3 (loi d'un vecteur aléatoire). *Soient (Ω, \mathcal{F}, P) un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}^d$ un vecteur aléatoire sur (Ω, \mathcal{F}) . On appelle loi de X sous P , ou plus simplement loi de X , la probabilité P_X sur $(\mathbb{R}^d, \text{Bor}(\mathbb{R}^d))$ définie par (5.1).*

Proposition 5.4 (lois marginales). *Si $X = (X_1, \dots, X_d)$ est un vecteur aléatoire sur (Ω, \mathcal{F}) , chacune de ses composantes X_i ($1 \leq i \leq d$) est une variable aléatoire réelle sur (Ω, \mathcal{F}) . La loi de X_i est appelée i^e loi marginale de X et est donnée par :*

$$\forall B_i \in \text{Bor}(\mathbb{R}), \quad P_{X_i}(B_i) = P(X_i \in B_i) = P(X \in \mathbb{R}^{i-1} \times B_i \times \mathbb{R}^{d-i}). \quad (5.2)$$

Preuve. La mesurabilité $\mathcal{F} - \text{Bor}(\mathbb{R})$ de X_i s'obtient par composition à partir de la mesurabilité $\mathcal{F} - \text{Bor}(\mathbb{R}^d)$ de X . En effet $X_i = \pi_i \circ X$, où $\pi_i : (x_1, \dots, x_d) \mapsto x_i$ est la i^{e} projection canonique de \mathbb{R}^d sur \mathbb{R} ; comme π_i est continue, elle est borélienne, c'est-à-dire ici mesurable $\text{Bor}(\mathbb{R}^d) - \text{Bor}(\mathbb{R})$. Donc X_i est bien une variable aléatoire. Pour vérifier (5.2), il suffit de remarquer que l'équivalence

$$X_i \in B_i \Leftrightarrow X \in \mathbb{R}^{i-1} \times B_i \times \mathbb{R}^{d-i}$$

entraîne l'égalité des événements correspondants et de leur probabilité. \square

Exemple 5.5. Soit $X = (X_1, X_2)$ un vecteur aléatoire de loi uniforme sur $[0, 1]^2$. Alors les lois marginales P_{X_1} et P_{X_2} sont égales à la loi uniforme sur $[0, 1]$. *Exercice !*

Remarque 5.6 (d'importance capitale). Une conséquence de la proposition 5.4 est que la connaissance de la loi du vecteur aléatoire X détermine complètement celle de ses lois marginales. La *reciproque est fausse*. On peut même affirmer sans hésiter qu'il est impossible de comprendre la notion de vecteur aléatoire tant que l'on n'a pas assimilé ce fait. La comparaison de l'exemple 5.7 ci-dessous avec l'exemple 5.5 permet de voir que la connaissance des lois marginales d'un vecteur ne détermine pas la loi du vecteur.

Exemple 5.7. Prenons une variable aléatoire réelle Y_1 de loi uniforme sur $[0, 1]$ et posons $Y_2 := Y_1$ et $Y := (Y_1, Y_2)$. Le vecteur aléatoire Y a par construction les mêmes lois marginales que le vecteur X de l'exemple 5.5. Notons $\Delta := \{(s, t) \in [0, 1]^2; s = t\}$ la première diagonale du carré unité. Il est clair par construction que $P(Y \in \Delta) = 1$.

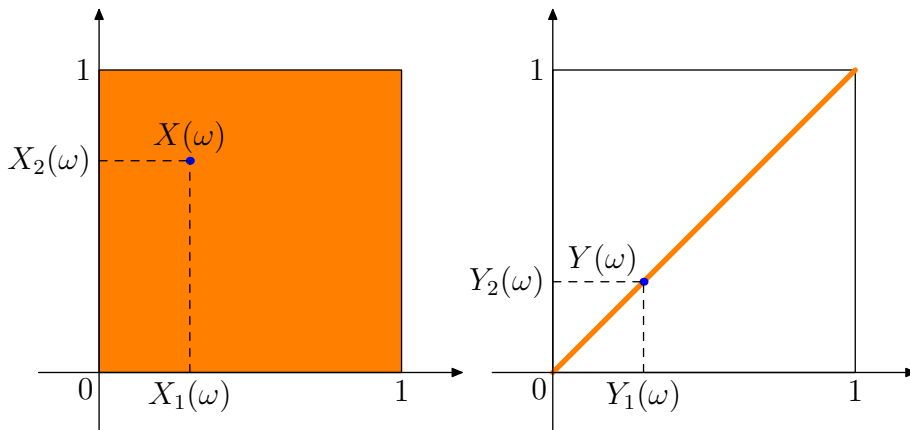
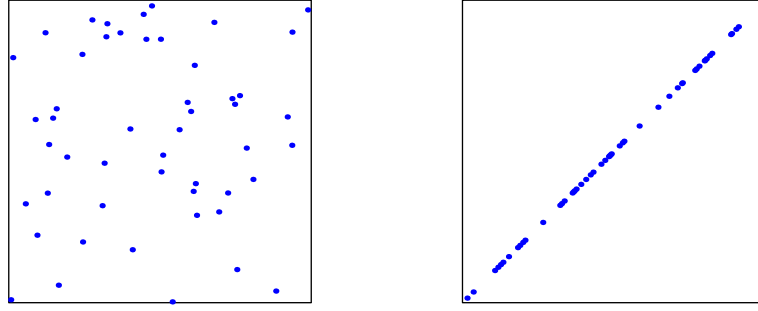


FIG. 5.1 – Ensembles de probabilité 1 pour les lois P_X et P_Y

D'un autre côté, $P(X \in \Delta) = 0$, car le segment Δ est de λ_2 mesure nulle. Ceci empêche que X et Y aient même loi. La figure 5.2 propose une illustration en affichant le résultat de la simulation du choix « au hasard » de 50 points suivant la loi P_X puis suivant la loi P_Y . La justification mathématique des méthodes de simulation sera étudiée dans le cours d'Initiation à la Statistique².

2. Ici on a utilisé le générateur de nombres aléatoires du logiciel METAPOST avec lequel la plupart des figures de ce document sont réalisées. Comme le code correspondant est inclus dans le source $\text{\LaTeX} 2_{\epsilon}$ de ce document, le choix des points aléatoires varie à chaque compilation. Ceci explique que la figure que vous voyez sur le document papier entre vos mains n'est pas forcément la même que sur le Web.


 FIG. 5.2 – 50 points choisis au hasard suivant la loi P_X , puis suivant la loi P_Y

Définition 5.8 (vecteur aléatoire discret). *Le vecteur aléatoire X de \mathbb{R}^d est dit discret si $X(\Omega)$ est une partie au plus dénombrable de \mathbb{R}^d .*

Il est clair que la loi de X s'écrit alors

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

Les variables aléatoires marginales X_i de $X = (X_1, \dots, X_d)$ sont alors des variables aléatoires discrètes. En effet, soit π_i , la restriction à $X(\Omega)$ de la projection canonique sur la i -ième composante de \mathbb{R}^d . Cette application réalise une surjection de $X(\Omega)$ sur $X_i(\Omega)$. Par la proposition 1.41, on en déduit que $X_i(\Omega)$ est au plus dénombrable.

Exemple 5.9 (lois multinomiales). Le vecteur aléatoire N suit la loi multinomiale de paramètres n et (p_1, \dots, p_d) où $n \in \mathbb{N}^*$ et les p_i sont strictement positifs et de somme 1 si pour tout d -uplet (j_1, j_2, \dots, j_d) d'entiers tels que $j_1 + j_2 + \dots + j_d = n$,

$$P\{N = (j_1, j_2, \dots, j_d)\} = \frac{n!}{j_1! j_2! \dots j_d!} p_1^{j_1} p_2^{j_2} \dots p_d^{j_d}.$$

Ici l'ensemble $N(\Omega) = \{(j_1, j_2, \dots, j_d) \in \mathbb{N}^d; j_1 + j_2 + \dots + j_d = n\}$ est fini et on vérifie grâce à la formule du multinôme, cf. prop. 1.23, que

$$\sum_{x \in N(\Omega)} P(N = x) = \sum_{j_1 + \dots + j_d = n} \frac{n!}{j_1! j_2! \dots j_d!} p_1^{j_1} p_2^{j_2} \dots p_d^{j_d} = (p_1 + \dots + p_d)^n = 1^n = 1.$$

La loi multinomiale est celle du vecteur des résultats d'une suite d'épreuves répétées indépendantes ayant chacune d issues possibles de probabilités respectives p_1, \dots, p_d . On pourra justifier cette affirmation en exercice. Par exemple considérons 20 tirages d'une boule avec remise dans une urne contenant 1 boule bleue, 3 jaunes, 4 rouges et 2 vertes. Notons $N = (N_1, N_2, N_3, N_4)$ où N_i est le nombre de boules de la couleur i en numérotant les couleurs par ordre alphabétique (b,j,r,v). On a $(p_1, p_2, p_3, p_4) = (\frac{1}{10}, \frac{3}{10}, \frac{4}{10}, \frac{2}{10})$. La probabilité d'obtenir en 20 tirages 3 bleues, 5 jaunes, 10 rouges et 2 vertes est

$$P(N = (3, 5, 10, 2)) = \frac{20!}{3! 5! 10! 2!} \left(\frac{1}{10}\right)^3 \left(\frac{3}{10}\right)^5 \left(\frac{4}{10}\right)^{10} \left(\frac{2}{10}\right)^2 \simeq 0,004\,745.$$

Certaines lois de vecteurs aléatoires ont une « densité » par rapport à la mesure de Lebesgue de \mathbb{R}^d . Nous donnons ci-dessous une définition volontairement restrictive des densités sur \mathbb{R}^d , restant dans le cadre de l'intégration au sens de Riemann. Même dans ce cadre, cette définition n'est ni la seule possible ni la plus générale. Elle devrait néanmoins suffire pour les exemples courants.

Définition 5.10 (densité de probabilité sur \mathbb{R}^d). *On appelle densité de probabilité sur \mathbb{R}^d toute fonction f vérifiant*

- a) *f est définie et positive sur $\mathbb{R}^d \setminus H$, où H est une réunion finie (éventuellement vide) d'hyperplans de \mathbb{R}^d ;*
- b) *f est localement Riemann intégrable³ sur $\mathbb{R}^d \setminus H$;*
- c) *l'intégrale généralisée de f sur \mathbb{R}^d converge⁴ et*

$$\int_{\mathbb{R}^d} f(t) dt = \int_{\mathbb{R}^d} f(t_1, \dots, t_d) dt_1 \dots dt_d = 1.$$

Définition 5.11 (vecteur aléatoire à densité). *Soit f une densité de probabilité sur \mathbb{R}^d . On dit que le vecteur aléatoire X de \mathbb{R}^d a pour densité f si pour tout pavé fermé borné $C = [a_1, b_1] \times \dots \times [a_d, b_d]$,*

$$P(X \in C) = \int_C f(t_1, \dots, t_d) dt_1 \dots dt_d.$$

Comme la classe des pavés fermés bornés engendre la tribu borélienne de \mathbb{R}^d , on peut montrer que si deux vecteurs aléatoires X et Y ont même densité f , ils ont même loi. Par ailleurs, si un vecteur aléatoire X admet une densité f , celle-ci n'est pas unique (si on la modifie en un point, on ne changera pas la valeur des ses intégrales sur les pavés fermés bornés).

Voici un premier exemple de vecteur aléatoire à densité. D'autres seront vus ultérieurement.

Exemple 5.12 (densité de la loi uniforme sur un borélien de \mathbb{R}^d). Soit B un borélien de \mathbb{R}^d tel que $0 < \lambda_d(B) < +\infty$. Si le vecteur aléatoire X de \mathbb{R}^d suit la loi uniforme sur B , cf. exemple 2.26, il admet pour densité la fonction

$$f = \frac{1}{\lambda_d(B)} \mathbf{1}_B.$$

3. Cela signifie qu'il existe une suite croissante pour l'inclusion (K_n) de compacts inclus dans $\mathbb{R}^d \setminus H$, que cette suite épuise $\mathbb{R}^d \setminus H$ et que f est Riemann intégrable sur chaque K_n . On dit que K_n épuise $\mathbb{R}^d \setminus H$ si pour tout compact $K \subset \mathbb{R}^d \setminus H$, il existe n_0 tel que $K \subset K_{n_0}$.

4. Cela signifie que pour toute suite (K_n) épuisant $\mathbb{R}^d \setminus H$, la suite $\int_{K_n} f(t) dt$ a une limite dans \mathbb{R} . On montre qu'alors cette limite ne dépend pas du choix de la suite (K_n) et on la note $\int_{\mathbb{R}^d} f(t) dt$. On montre aussi qu'une CNS pour la convergence de l'intégrale généralisée de f est que pour une suite (K_n) particulière, la suite de réels $\int_{K_n} |f(t)| dt$ soit bornée.

En effet, pour tout pavé fermé borné $C = [a_1, b_1] \times \cdots \times [a_d, b_d]$, on a

$$\begin{aligned} P(X \in C) &= \frac{\lambda_d(B \cap C)}{\lambda_d(B)} = \frac{1}{\lambda_d(B)} \int_{\mathbb{R}^d} \mathbf{1}_{B \cap C}(t_1, \dots, t_d) dt_1 \dots dt_d \\ &= \frac{1}{\lambda_d(B)} \int_{\mathbb{R}^d} \mathbf{1}_B(t_1, \dots, t_d) \mathbf{1}_C(t_1, \dots, t_d) dt_1 \dots dt_d \\ &= \frac{1}{\lambda_d(B)} \int_C \mathbf{1}_B(t_1, \dots, t_d) dt_1 \dots dt_d \\ &= \int_C f(t_1, \dots, t_d) dt_1 \dots dt_d. \end{aligned}$$

Le calcul ci-dessus utilise le fait que $\mathbf{1}_{B \cap C} = \mathbf{1}_B \mathbf{1}_C$ et la relation $\lambda_d(A) = \int_A dt_1 \dots dt_d = \int_{\mathbb{R}^d} \mathbf{1}_A(t_1, \dots, t_d) dt_1 \dots dt_d$ pour tout borélien A . Nous touchons ici aux limitations de ce cours sur la théorie de l'intégrale. En toute rigueur nous ne devrions écrire cette relation que si $\mathbf{1}_A$ vérifie les conditions a) et b) de la définition 5.10. En pratique nous n'utiliserons la loi uniforme que sur des boréliens assez simples pour lesquels ces conditions seront remplies.

Comme dans le cas $d = 1$, on dispose d'une condition suffisante pratique pour vérifier que deux vecteurs aléatoires à densité n'ont pas même loi.

Lemme 5.13. *Soient X et Y deux vecteurs aléatoires de \mathbb{R}^d admettant respectivement pour densité les fonctions f et g . On suppose qu'il existe $t_0 \in \mathbb{R}^d$ tel que $f(t_0) \neq g(t_0)$ et que de plus, f et g sont toutes deux continues au point t_0 . Alors X et Y n'ont pas même loi.*

La preuve est essentiellement la même que celle du lemme 3.24, en remplaçant les intervalles ouverts par des pavés ouverts et sera donc omise.

Proposition 5.14 (densités marginales). *Si le vecteur aléatoire X de \mathbb{R}^d est à densité f , ses lois marginales sont aussi à densité et pour $i = 1, \dots, d$, une densité de X_i est donnée par*

$$x_i \mapsto f_{X_i}(x_i) = \int_{\mathbb{R}^{i-1} \times \mathbb{R}^{d-i}} f(t_1, \dots, t_{i-1}, x_i, t_{i+1}, \dots, t_d) dt_1 \dots dt_{i-1} dt_{i+1} \dots dt_d.$$

Preuve. On calcule $P(X_i \in [a_i, b_i])$ en utilisant (5.2) :

$$\begin{aligned} P(X_i \in [a_i, b_i]) &= \int_{\mathbb{R}^{i-1} \times [a_i, b_i] \times \mathbb{R}^{d-i}} f(t_1, \dots, t_d) dt_1 \dots dt_d \\ &= \int_{[a_i, b_i]} \left\{ \int_{\mathbb{R}^{i-1} \times \mathbb{R}^{d-i}} f(t_1, \dots, t_d) dt_1 \dots dt_{i-1} dt_{i+1} \dots dt_d \right\} dt_i, \end{aligned}$$

par le théorème de Fubini-Tonnelli. Comme a_i et b_i sont quelconques, on en déduit que la fonction entre accolades ci-dessus est une densité de X_i . En toute rigueur, il y a ici un problème car le théorème de Fubini-Tonnelli nous dit que cette fonction est définie sauf peut-être sur un ensemble de mesure de Lebesgue nulle (où elle prend la valeur $+\infty$). En pratique nous ne rencontrerons pas de difficulté à ce sujet. \square

Exemple 5.15 (loi uniforme sur un disque). Soit D le disque unité de \mathbb{R}^2 et $X = (X_1, X_2)$ un vecteur aléatoire suivant la loi uniforme sur D . D'après l'exemple 5.12, nous savons qu'il admet pour densité $f = \pi^{-1} \mathbf{1}_D$. Calculons la densité marginale f_{X_1} fournie par la proposition 5.14.

$$f_{X_1}(x_1) = \int_{\mathbb{R}} \frac{1}{\pi} \mathbf{1}_D(x_1, t_2) dt_2 = \frac{1}{\pi} \int_{-\infty}^{+\infty} \mathbf{1}_D(x_1, t_2) dt_2. \quad (5.3)$$

Notons ici que l'appartenance à D du point (t_1, t_2) de \mathbb{R}^2 est caractérisée par l'inégalité $t_1^2 + t_2^2 \leq 1$. On a donc

$$\mathbf{1}_D(x_1, t_2) = \begin{cases} 1 & \text{si } x_1^2 + t_2^2 \leq 1 \\ 0 & \text{sinon} \end{cases} = \begin{cases} 1 & \text{si } x_1^2 \leq 1 \text{ et } t_2^2 \leq 1 - x_1^2 \\ 0 & \text{sinon} \end{cases}$$

D'où

$$\mathbf{1}_D(x_1, t_2) = \begin{cases} 1 & \text{si } |x_1| \leq 1 \text{ et } |t_2| \leq \sqrt{1 - x_1^2}, \\ 0 & \text{sinon,} \end{cases}$$

ce qui peut aussi s'écrire

$$\mathbf{1}_D(x_1, t_2) = \mathbf{1}_{[-1,1]}(x_1) \mathbf{1}_{[-(1-x_1^2)^{1/2}, (1-x_1^2)^{1/2}]}(t_2).$$

Géométriquement, ce petit calcul revient à chercher l'intersection du disque D avec la droite verticale $t_1 = x_1$, où x_1 joue le rôle d'une constante, et à projeter cette intersection sur le deuxième axe, cf figure 5.3. En reportant cette expression de $\mathbf{1}_D(x_1, t_2)$ dans (5.14), on obtient

$$\begin{aligned} f_{X_1}(x_1) &= \frac{1}{\pi} \int_{-\infty}^{+\infty} \mathbf{1}_{[-1,1]}(x_1) \mathbf{1}_{[-(1-x_1^2)^{1/2}, (1-x_1^2)^{1/2}]}(t_2) dt_2 \\ &= \frac{1}{\pi} \mathbf{1}_{[-1,1]}(x_1) \int_{-\infty}^{+\infty} \mathbf{1}_{[-(1-x_1^2)^{1/2}, (1-x_1^2)^{1/2}]}(t_2) dt_2 \\ &= \frac{1}{\pi} \mathbf{1}_{[-1,1]}(x_1) \int_{-(1-x_1^2)^{1/2}}^{(1-x_1^2)^{1/2}} dt_2. \end{aligned}$$

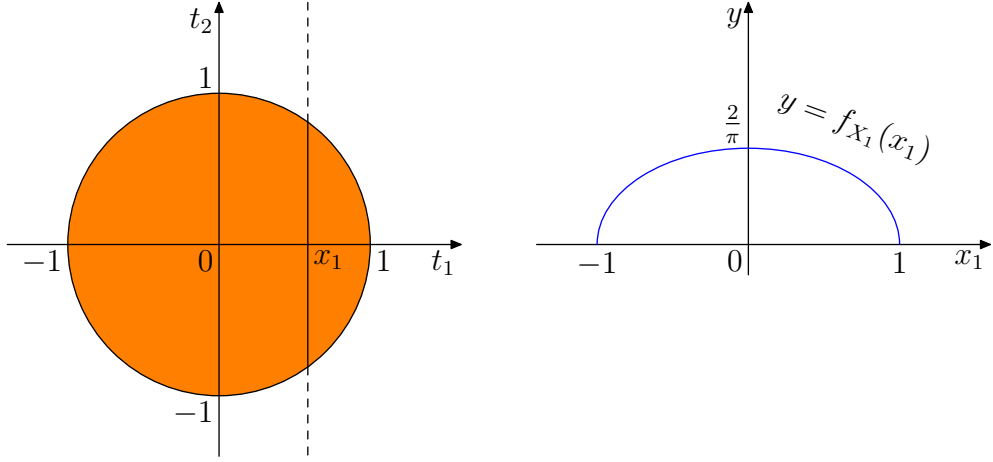
Cette dernière intégrale n'est rien d'autre que la longueur du segment défini par ses bornes (rappelons que x_1 est considéré comme une constante dans tout ce calcul). On aboutit ainsi à

$$f_{X_1}(x_1) = \frac{2}{\pi} (1 - x_1^2)^{1/2} \mathbf{1}_{[-1,1]}(x_1).$$

Par raison de symétrie il est clair que X_1 et X_2 ont même loi et donc même densité, d'où

$$f_{X_2}(x_2) = \frac{2}{\pi} (1 - x_2^2)^{1/2} \mathbf{1}_{[-1,1]}(x_2).$$

La construction du graphe de f_{X_1} est très simple : on l'obtient en appliquant au demi cercle unité supérieur d'équation $y = \sqrt{1 - x_1^2}$ la transformation $(x_1, y) \mapsto (x_1, \frac{2}{\pi}y)$. On obtient ainsi une moitié d'ellipse. Un coup d'oeil sur ce graphe devrait vous convaincre que la loi de X_1 , bien qu'ayant toute sa masse portée par le segment $[-1, 1]$, n'est pas la loi uniforme sur $[-1, 1]$.


 FIG. 5.3 – Densité marginale f_{X_1} de la loi uniforme sur le disque unité D

On peut définir la fonction de répartition F d'un vecteur aléatoire par

$$\forall x = (x_1, \dots, x_d), \quad F(x) = P(X \in]-\infty, x_1] \times \dots \times]-\infty, x_d]).$$

Comme en dimension 1, la f.d.r. caractérise la loi. Ceci est lié au fait que la tribu $\text{Bor}(\mathbb{R}^d)$ est engendrée par la classe des ensembles de la forme $]-\infty, x_1] \times \dots \times]-\infty, x_d]$. Néanmoins le rôle des f.d.r. en dimension $d > 1$ est bien moindre qu'en dimension 1. On préfère caractériser la loi d'un vecteur aléatoire par une collection de h -moments $\mathbf{E}h(X)$ ($h : \mathbb{R}^d \rightarrow \mathbb{R}$) au sens suivant.

Proposition 5.16 (caractérisation par les moments fonctionnels). *La loi d'un vecteur aléatoire X de \mathbb{R}^d est caractérisée par la famille de moments fonctionnels $\{\mathbf{E}h(X); h \in \mathcal{H}\}$, où \mathcal{H} est une classe « suffisamment riche » de fonctions boréliennes $\mathbb{R}^d \rightarrow \mathbb{R}$. Autrement dit, les deux vecteurs aléatoires X et Y ont même loi si et seulement si $\mathbf{E}h(X) = \mathbf{E}h(Y)$ pour toute $h \in \mathcal{H}$. Comme famille \mathcal{H} « suffisamment riche », on peut prendre :*

- l'ensemble des fonctions boréliennes positives $\mathbb{R}^d \rightarrow \mathbb{R}_+$,
- l'espace $C^b(\mathbb{R}^d)$ des fonctions continues bornées $\mathbb{R}^d \rightarrow \mathbb{R}$,
- l'espace $C^c(\mathbb{R}^d)$ des fonctions continues à support compact $\mathbb{R}^d \rightarrow \mathbb{R}$.

Preuve. L'espace $C^c(\mathbb{R}^d)$ étant inclus dans $C^b(\mathbb{R}^d)$, lui même inclus dans l'ensemble des fonctions boréliennes positives $\mathbb{R}^d \rightarrow \mathbb{R}_+$, il suffit de faire la preuve pour $\mathcal{H} = C^c(\mathbb{R}^d)$. La partie évidente est que l'égalité des lois des vecteurs aléatoires X et Y implique l'égalité des lois des variables aléatoires réelles bornées $h(X)$ et $h(Y)$ et donc l'égalité de leur espérance. Pour montrer la réciproque, on suppose désormais que

$$\forall h \in C^c(\mathbb{R}^d), \quad \mathbf{E}h(X) = \mathbf{E}h(Y). \quad (5.4)$$

Comme la loi d'un vecteur aléatoire Z de \mathbb{R}^d est caractérisée par les $P(Z \in C)$, où C décrit la famille des pavés ouverts de \mathbb{R}^d , il nous suffit de prouver que $P(X \in C) = P(Y \in C)$ pour tout pavé ouvert C .

Soit donc

$$C := \prod_{j=1}^d]a_j, b_j[,$$

un pavé ouvert non vide de \mathbb{R}^d (on a donc $a_j < b_j$ pour tout j). Définissons pour tout $n \geq n_0$, tel que $2/n_0 < \min_{1 \leq j \leq d} (b_j - a_j)$, la fonction continue à support compact h_n par

$$h_n(x) = h_n(x_1, \dots, x_d) := h_{n,1}(x_1) \dots h_{n,d}(x_d),$$

où $h_{n,j}$ est l'unique fonction continue $\mathbb{R} \rightarrow [0, 1]$, valant 1 sur $[a_j + 1/n, b_j - 1/n]$, 0 hors de $]a_j, b_j[$ et affine sur chacun des intervalles $[a_j, a_j + 1/n]$ et $[b_j - 1/n, b_j]$, voir la figure 5.4.

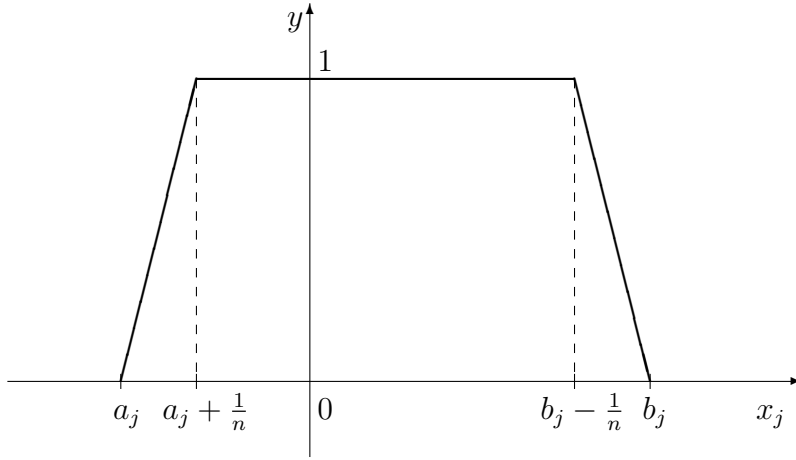


FIG. 5.4 – Fonction $h_{n,j}$

La fonction h_n est continue, à support compact (son support est exactement le pavé fermé \overline{C}). On remarque que pour tout $j = 1, \dots, d$, la suite de fonctions positives $(h_{n,j})_{n \geq n_0}$ est croissante et converge en tout point de \mathbb{R} vers l'indicatrice $\mathbf{1}_{]a_j, b_j[}$. On en déduit immédiatement que pour tout $x \in \mathbb{R}^d$, la suite de réels positifs $(h_n(x))_{n \geq n_0}$ converge en croissant vers $\mathbf{1}_C(x)$. Par conséquent $(h_n(X))_{n \geq n_0}$ et $(h_n(Y))_{n \geq n_0}$ sont deux suites croissantes de variables aléatoires positives qui convergent (sur tout Ω) respectivement vers $\mathbf{1}_C(X)$ et $\mathbf{1}_C(Y)$. Comme par l'hypothèse (5.4), $\mathbf{E}h_n(X) = \mathbf{E}h_n(Y)$ pour tout $n \geq n_0$, on en déduit en faisant tendre n vers $+\infty$ et en appliquant le théorème de Beppo Levi que $\mathbf{E}\mathbf{1}_C(X) = \mathbf{E}\mathbf{1}_C(Y)$, ce qui s'écrit aussi $P(X \in C) = P(Y \in C)$. \square

Il importe de savoir calculer les $\mathbf{E}h(X)$ quand on connaît la loi de X . Les formules sont analogues à celles déjà données en dimension 1.

Proposition 5.17 (calcul de $\mathbf{E}h(X)$).

a) Si le vecteur aléatoire X est discret et si h est une fonction borélienne $\mathbb{R}^d \rightarrow \mathbb{R}$,

$$\mathbf{E}|h(X)| = \sum_{x \in X(\Omega)} |h(x)|P(X = x).$$

Si de plus $\mathbf{E}|h(X)| < +\infty$,

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

b) Si X est à densité f et $h : \mathbb{R}^d \rightarrow \mathbb{R}$ est continue bornée sur \mathbb{R}^d ,

$$\mathbf{E}h(X) = \int_{\mathbb{R}^d} h(x)f(x) \, dx = \int_{\mathbb{R}^d} h(x_1, \dots, x_d)f(x_1, \dots, x_d) \, dx_1 \dots dx_d.$$

Cette formule se généralise au cas où h est continue non bornée sur \mathbb{R}^d , sous réserve que $\int_{\mathbb{R}^d} |h(x)|f(x) \, dx < +\infty$.

Nous donnons maintenant une formule de calcul de densité du vecteur aléatoire $g(X)$ où X est un vecteur aléatoire à densité.

Proposition 5.18 (densité d'un vecteur aléatoire image). Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de \mathbb{R}^d ayant une densité f_X . On suppose de plus que D est un ouvert de \mathbb{R}^d tel que $P(X \in D) = 1$ et que $g : D \rightarrow \mathbb{R}^d$ est C^1 , injective avec un déterminant jacobien $\text{Jac}(g)$ qui ne s'annule en aucun point de D . Alors g réalise une bijection C^1 d'inverse C^1 (autrement dit un C^1 -difféomorphisme) entre D et son image $D' := g(D)$. Le vecteur aléatoire $Y := g(X)$ admet pour densité

$$f_Y(y) = f_X(g^{-1}(y)) |\text{Jac}(g^{-1})(y)| \mathbf{1}_{D'}(y).$$

Nous ne démontrerons pas cette proposition. L'affirmation que g est un C^1 -difféomorphisme découle d'un théorème d'analyse (théorème d'inversion globale). Rappelons que le jacobien de g est donné par

$$\text{Jac}(g) = \det \left(\left[\frac{\partial g_i}{\partial x_j} \right]_{i,j=1,\dots,d} \right),$$

où les g_i sont les applications coordonnées de $g = (g_1, \dots, g_d)$, i.e. $g_i(x_1, \dots, x_d) = \pi_i(g(x_1, \dots, x_d))$, π_i étant la projection canonique sur la i^{e} coordonnée dans \mathbb{R}^d . Pour calculer $\text{Jac}(g^{-1})(y)$, on peut soit utiliser la formule ci-dessus en remplaçant g par g^{-1} et les x_j par les y_j , soit utiliser la relation

$$\text{Jac}(g^{-1})(y) = \frac{1}{\text{Jac}(g)(g^{-1}(y))}.$$

Un cas particulier facile et important est celui où g est une application linéaire bijective de \mathbb{R}^d sur \mathbb{R}^d . Dans ce cas, soit $A = [a_{i,j}]$ la matrice $d \times d$ de g relativement à

la base canonique de \mathbb{R}^d . Cette matrice est inversible, son déterminant est donc non nul. La i^{e} composante de $g(x)$ est ici $g_i(x_1, \dots, x_d) = \sum_{j=1}^d a_{i,j}x_j$, d'où

$$\forall i, j = 1, \dots, d, \quad \frac{\partial g_i}{\partial x_j} = a_{i,j}.$$

On en déduit que le jacobien de g comme celui de g^{-1} sont constants et valent

$$\text{Jac}(g) = \det A = \det(g), \quad \text{Jac}(g^{-1}) = \frac{1}{\det A} = \frac{1}{\det g}.$$

Nous obtenons ainsi le corollaire suivant de la proposition 5.18, avec ici $D = \mathbb{R}^d$ et $D' = g(D) = \mathbb{R}^d$.

Corollaire 5.19 (changement de variable linéaire). *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de \mathbb{R}^d ayant une densité f_X et $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$ une application linéaire bijective. Alors le vecteur aléatoire $Y := g(X)$ admet pour densité*

$$f_Y(y) = \frac{1}{|\det g|} f_X(g^{-1}(y)).$$

5.1.2 Covariance

Proposition 5.20 (inégalité de Cauchy-Schwarz). *Si les variables aléatoires réelles X et Y ont des moments d'ordre 2, alors la variable aléatoire XY est intégrable et*

$$|\mathbf{E}(XY)| \leq (\mathbf{E}X^2)^{1/2}(\mathbf{E}Y^2)^{1/2}. \quad (5.5)$$

Preuve. L'intégrabilité de la v.a. XY résulte de l'inégalité élémentaire $|XY| \leq X^2 + Y^2$. On remarque alors que la fonction trinômiale suivante de la variable réelle t ,

$$g : t \mapsto t^2 \mathbf{E}Y^2 + 2t \mathbf{E}(XY) + \mathbf{E}X^2 = \mathbf{E}(X + tY)^2,$$

est définie sur \mathbb{R} et positive sur \mathbb{R} . Ceci n'est possible que si son discriminant est négatif, ce qui s'écrit

$$\Delta' = (\mathbf{E}(XY))^2 - (\mathbf{E}X^2)(\mathbf{E}Y^2) \leq 0,$$

d'où l'inégalité (5.5). □

Définition 5.21 (covariance). *Si les variables aléatoires réelles X et Y ont des moments d'ordre 2, on appelle covariance du couple aléatoire (X, Y) la quantité :*

$$\text{Cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}X)(Y - \mathbf{E}Y)].$$

Remarquons que $\text{Cov}(X, X) = \text{Var } X$.

Proposition 5.22 (propriétés de la covariance). *Les propriétés suivantes sont vérifiées pour tout couple (X, Y) de v.a. réelles ayant des moments d'ordre 2.*

- (i) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
- (ii) Pour tous réels a, b, c, d : $\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y)$.
- (iii) $|\text{Cov}(X, Y)| \leq \sigma(X)\sigma(Y)$.

La vérification est laissée au lecteur.

Définition 5.23 (coefficient de corrélation). Si X et Y sont des variables aléatoires réelles non constantes ayant des moments d'ordre 2, on appelle coefficient de corrélation entre X et Y la quantité :

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

D'après (iii) on a toujours $-1 \leq \rho(X, Y) \leq 1$. D'autre part il résulte facilement du cas d'égalité dans l'inégalité de Cauchy-Schwarz que $|\rho|$ est maximal lorsque Y est une fonction affine de X : $Y = aX + b$. Quand $\rho = 0$ (ce qui arrive en particulier lorsque X et Y sont indépendantes), on dit que X et Y sont *non corrélées*.

Proposition 5.24 (formule de Koenig pour la covariance). Si la covariance de X et Y existe, elle peut se calculer par :

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

Preuve. La vérification est analogue à celle de la formule de Koenig pour la variance (qui n'est que le cas particulier $Y = X$) et est laissée en exercice. \square

Remarque 5.25 (calcul explicite de la covariance). Les formules de calcul des h -moments appliquées au vecteur aléatoire (X, Y) (prop. 5.17) et aux variables aléatoires réelles X et Y (prop. 4.41 et 4.42) nous donnent pour la covariance (lorsqu'elle existe) les formules explicites suivantes.

Si (X, Y) est discret,

$$\text{Cov}(X, Y) = \sum_{\substack{x \in X(\Omega) \\ y \in Y(\Omega)}} xyP(X = x, Y = y) - \sum_{x \in X(\Omega)} xP(X = x) \sum_{y \in Y(\Omega)} yP(Y = y). \quad (5.6)$$

Si (X, Y) est à densité f , en notant f_X et f_Y les densités marginales,

$$\text{Cov}(X, Y) = \int_{\mathbb{R}^2} xyf(x, y) \, dx \, dy - \int_{-\infty}^{+\infty} xf_X(x) \, dx \int_{-\infty}^{+\infty} yf_Y(y) \, dy. \quad (5.7)$$

Proposition 5.26 (variance d'une somme). Si X_1, \dots, X_n sont des variables aléatoires réelles ayant des moments d'ordre 2,

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i,j=1}^n \text{Cov}(X_i, X_j) \quad (5.8)$$

$$= \sum_{i=1}^n \text{Var} X_i + \sum_{\substack{i,j=1 \\ i \neq j}}^n \text{Cov}(X_i, X_j). \quad (5.9)$$

Dans le cas $n = 2$ (5.9) s'écrit :

$$\text{Var}(X + Y) = \text{Var } X + \text{Var } Y + 2 \text{Cov}(X, Y). \quad (5.10)$$

Preuve. Pour $n \geq 2$ quelconque, l'identité algébrique :

$$\left(\sum_{i=1}^n Y_i \right)^2 = \sum_{i,j=1}^n Y_i Y_j,$$

utilisée avec $Y_i = X_i - \mathbf{E}X_i$ et la linéarité de l'espérance nous donnent :

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^n X_i\right) &= \mathbf{E}\left\{\sum_{i=1}^n X_i - \mathbf{E}\left(\sum_{i=1}^n X_i\right)\right\}^2 \\ &= \mathbf{E}\left\{\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbf{E}X_i\right\}^2 \\ &= \mathbf{E}\left\{\sum_{i=1}^n Y_i\right\}^2 \\ &= \sum_{i,j=1}^n \mathbf{E}(Y_i Y_j) = \sum_{i,j=1}^n \text{Cov}(X_i, X_j). \end{aligned}$$

□

5.2 Indépendance de variables et vecteurs aléatoires

5.2.1 Suites indépendantes

Définition 5.27 (indépendance de n variables aléatoires). *Les variables aléatoires réelles X_1, \dots, X_n définies sur le même (Ω, \mathcal{F}, P) sont dites indépendantes si :*

$$\forall B_1, \dots, B_n \in \text{Bor}(\mathbb{R}), \quad P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \dots P(X_n \in B_n). \quad (5.11)$$

Remarques 5.28.

1. La condition (5.11) équivaut à l'indépendance mutuelle des événements $\{X_1 \in B_1\}, \dots, \{X_n \in B_n\}$ pour tout choix des boréliens B_i . En effet la liberté de choisir $B_i = \mathbb{R}$ permet « d'effacer » pour ce choix le rôle de X_i et ainsi d'obtenir à partir de (5.11) une égalité de même type pour l'intersection de toute sous-famille finie des événements considérés. La rédaction détaillée de la justification de cette équivalence est laissée en exercice. Le même argument montre que l'indépendance de X_1, \dots, X_n entraîne celle des X_{i_1}, \dots, X_{i_k} pour tout choix d'indices $1 \leq i_1 < i_2 < \dots < i_k \leq n$.
2. L'indépendance des X_i , $1 \leq i \leq n$ est une propriété plus forte que leur indépendance deux à deux.

3. Pour vérifier l'indépendance de X_1, \dots, X_n , il suffit de tester l'égalité dans (5.11) pour des B_i quelconques pris dans l'une des sous-familles suivantes de $\text{Bor}(\mathbb{R})$:
 - les intervalles $] -\infty, b_i]$;
 - les intervalles $] a_i, b_i]$;
 - les intervalles $] a_i, b_i[$;
 - les intervalles $[a_i, b_i]$;
 - les intervalles $] b_i, +\infty[$.

Cette réduction repose essentiellement sur le fait que chacune de ces familles d'intervalles engendre $\text{Bor}(\mathbb{R})$ et est stable par intersections finies (en autorisant $b_i < a_i$ pour avoir l'ensemble vide dans chaque famille). Nous nous contenterons de cette justification pour ne pas trop sortir du cadre de ce cours.

4. Clairement l'indépendance n'est pas affectée par une permutation sur les indices i des variables aléatoires. On peut donc parler d'indépendance d'une famille finie de variables aléatoires sans se préoccuper de l'ordre d'indexation.

La définition de l'indépendance se généralise comme suit à une famille finie de vecteurs⁵ aléatoires.

Définition 5.29 (indépendance de n vecteurs aléatoires). *Pour $i = 1, \dots, n$, soient $X_i : \Omega \rightarrow \mathbb{R}^{d_i}$ des vecteurs aléatoires sur le même espace probabilisé (Ω, \mathcal{F}, P) . On dit qu'ils sont indépendants si l'égalité*

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \dots P(X_n \in B_n) \quad (5.12)$$

est vérifiée pour tout choix des boréliens $B_i \in \text{Bor}(\mathbb{R}^{d_i})$, $i = 1, \dots, n$.

Notons qu'avec cette définition, la collection de vecteurs aléatoires considérée peut être complètement hétéroclite quant aux dimensions, y compris avec $d_i = 1$ pour certaines valeurs de i (les X_i correspondants étant alors des variables aléatoires réelles).

Une propriété bien commode de l'indépendance des v.a. est *l'hérédité*. Avant de l'énoncer formellement, voyons sa signification sur un exemple. Supposons que les v.a. réelles X_1, \dots, X_5 soient indépendantes. Alors les trois variables aléatoires Y_1, Y_2, Y_3 suivantes sont indépendantes

$$Y_1 := X_1 + X_2, \quad Y_2 := X_3 \sin X_4, \quad Y_3 := \exp(X_5^2 - X_5).$$

Il en va de même pour les vecteurs aléatoires Z_1, Z_2, Z_3

$$Z_1 := (X_1, X_2), \quad Z_2 := (X_3 + \cos X_4, X_4^2), \quad Z_3 := (X_5^3, 2X_5, X_5^2).$$

Nous énonçons le résultat à partir d'une suite de variables aléatoires indépendantes pour ne pas trop alourdir les notations, mais il se généralise à une famille finie hétéroclite de vecteurs aléatoires indépendants au sens de la définition 5.29.

5. Il va sans dire, mais mieux en le disant, que l'indépendance des vecteurs dont il est question ici n'a rien à voir avec l'indépendance linéaire !

Proposition 5.30 (hérédité de l'indépendance). *Soient X_1, \dots, X_n des variables aléatoires réelles indépendantes. Découpons $\{1, \dots, n\}$ en k blocs disjoints, en notant m_j le cardinal du j -ième bloc et $n_j = \sum_{1 \leq l \leq j} m_l$, avec $n_0 := 0$. Pour $j = 1, \dots, k$, associons au j -ième bloc une fonction borélienne $h_j : \mathbb{R}^{m_j} \rightarrow \mathbb{R}^{d_j}$. Posons enfin $Z_j := h_j(X_{n_{j-1}+1}, \dots, X_{n_j})$. Alors la suite finie $(Z_j)_{1 \leq j \leq k}$ de vecteurs aléatoires est indépendante (au sens de la définition 5.29).*

En prenant pour chaque j , $d_j = m_j$ et h_j égale à l'identité $\mathbb{R}^{m_j} \rightarrow \mathbb{R}^{m_j}$, on voit que la proposition 5.30 contient en particulier le résultat suivant.

Corollaire 5.31 (indépendance des blocs disjoints). *Si les X_1, \dots, X_n sont des variables aléatoires réelles indépendantes, les k « blocs disjoints » $Y_j := (X_{n_{j-1}+1}, \dots, X_{n_j})$, où $0 = n_0 < n_1 < n_2 < \dots < n_k = n$, sont des vecteurs aléatoires indépendants.*

Remarque 5.32. Compte-tenu de la remarque 5.28-4, on généralise facilement la proposition 5.30 et le corollaire 5.31 au cas où les « blocs » ne sont plus forcément indexés par des entiers consécutifs, pourvu que les blocs d'indices correspondants restent deux à deux disjoints. Ainsi par exemple si X_1, \dots, X_7 sont indépendantes, (X_3, X_1) , (X_2, X_5, X_7) et (X_6, X_4) sont des vecteurs aléatoires indépendants.

La démonstration détaillée de la proposition 5.30 sortirait du programme de ce cours où sont seulement exigibles l'énoncé et son utilisation pratique. Ce qui suit s'adresse aux lecteurs plus avancés ou désireux d'en savoir plus. Le lecteur débutant peut sauter ce passage et aller directement à la définition 5.34.

La preuve informelle⁶ de la proposition 5.30 que nous décrivons maintenant distingue deux idées dans cette proposition. La première est l'hérédité pour les blocs, *i.e.* le corollaire 5.31 et utilise la construction de « probabilités produits ». La deuxième est l'hérédité pour les images des blocs par les transformations boréliennes h_j et repose sur la notion d'indépendance de sous-tribus de \mathcal{F} .

Preuve informelle de l'hérédité pour les blocs. Partons donc de l'égalité (5.11) qui exprime notre hypothèse d'indépendance des X_i . Avec les notations pour les blocs introduites dans l'énoncé de la proposition 5.30, identifions \mathbb{R}^n avec $\mathbb{R}^{m_1} \times \dots \times \mathbb{R}^{m_k}$. Pour $j = 1, \dots, k$, notons \mathcal{C}_j la famille des produits cartésiens $B'_j = B_{n_{j-1}+1} \times \dots \times B_{n_j}$ de m_j boréliens quelconques de \mathbb{R} . En rappelant que $Y_j := (X_{n_{j-1}+1}, \dots, X_{n_j})$, ceci nous permet de réécrire (5.11) sous la forme suivante vérifiée pour tous choix des $B'_j \in \mathcal{C}_j$.

$$P(Y_1 \in B'_1, \dots, Y_k \in B'_k) = \prod_{j=1}^k \prod_{n_{j-1} < i \leq n_j} P(X_i \in B_i). \quad (5.13)$$

En particulier si on choisit ci-dessus pour un j fixé, $B_i = \mathbb{R}$ pour tous les i hors de $[n_{j-1}, n_j]$, chaque B'_l pour $l \neq j$ est égal à \mathbb{R}^{m_l} et on obtient après effacement des

6. Nous ne prétendons pas donner une preuve complète, mais plutôt pointer les difficultés et indiquer l'outillage mathématique utilisé pour les surmonter.

conditions inutiles du type $Y_l \in \mathbb{R}^{m_l}$ ou $X_i \in \mathbb{R}$,

$$\forall B'_j \in \mathcal{C}_j, \quad P(Y_j \in B'_j) = \prod_{n_{j-1} < i \leq n_j} P(X_i \in B_i). \quad (5.14)$$

Ceci étant vrai pour j quelconque, on obtient en reportant dans (5.13) :

$$\forall B'_1 \in \mathcal{C}_1, \dots, \forall B'_k \in \mathcal{C}_k, \quad P(Y_1 \in B'_1, \dots, Y_k \in B'_k) = \prod_{j=1}^k P(Y_j \in B'_j). \quad (5.15)$$

Pour en déduire que la suite finie de vecteurs aléatoires Y_1, \dots, Y_k est indépendante au sens de la définition 5.29, il faudrait pouvoir étendre (5.15) à tous les $B'_j \in \text{Bor}(\mathbb{R}^{m_j})$, ce qui est loin d'être trivial. Notons que si $m_j \geq 2$, \mathcal{C}_j n'est pas une tribu et est strictement incluse dans $\text{Bor}(\mathbb{R}^{m_j})$. Par exemple le disque unité ouvert de \mathbb{R}^2 est un borélien de \mathbb{R}^2 puisque c'est un ouvert et il est impossible de l'écrire comme produit cartésien de deux sous-ensembles de \mathbb{R} (exercice!). Notons $\mu := P_{Y_1, \dots, Y_k}$, la loi du vecteur aléatoire hétéroclite (Y_1, \dots, Y_k) de $\mathbb{R}^{m_1} \times \dots \times \mathbb{R}^{m_k}$, définie *via* l'indentification de cet espace avec \mathbb{R}^n par :

$$\forall D \in \text{Bor}(\mathbb{R}^n), \quad \mu(D) = P((Y_1, \dots, Y_k) \in D).$$

Définissons la tribu produit $\text{Bor}(\mathbb{R}^{m_1}) \otimes \dots \otimes \text{Bor}(\mathbb{R}^{m_k})$ comme la tribu sur $\mathbb{R}^{m_1} \times \dots \times \mathbb{R}^{m_k}$ engendrée par la famille des produits cartésiens $D_1 \times \dots \times D_k$, où chaque D_j est un borélien quelconque de \mathbb{R}^{m_j} . On peut montrer que cette tribu coïncide avec la tribu borélienne de $\mathbb{R}^{m_1} \times \dots \times \mathbb{R}^{m_k}$, donc de \mathbb{R}^n . On montre aussi que l'on peut définir sur cette tribu produit la *probabilité produit* $\nu = P_{Y_1} \otimes \dots \otimes P_{Y_k}$ comme l'unique mesure vérifiant

$$\forall D_1 \in \text{Bor}(\mathbb{R}^{m_1}), \dots, \forall D_k \in \text{Bor}(\mathbb{R}^{m_k}), \quad \nu(D_1 \times \dots \times D_k) = P_{Y_1}(D_1) \dots P_{Y_k}(D_k).$$

Notons enfin \mathcal{C} la famille des sous-ensembles de $\mathbb{R}^{m_1} \times \dots \times \mathbb{R}^{m_k}$ de la forme $B'_1 \times \dots \times B'_k$, où $B'_j \in \mathcal{C}_j$, $j = 1, \dots, k$. Muni de tout cet attirail, on peut réécrire (5.15) sous la forme :

$$\forall C \in \mathcal{C}, \quad \mu(C) = \nu(C).$$

Finalement, on vérifie que \mathcal{C} est une famille stable par intersections finies, qui engendre la tribu borélienne de $\mathbb{R}^{m_1} \times \dots \times \mathbb{R}^{m_k}$. Comme les mesures de probabilité μ et ν coïncident sur \mathcal{C} , elles coïncident sur la tribu engendrée⁷, *i.e.* $\mu(D) = \nu(D)$ pour tout $D \in \text{Bor}(\mathbb{R}^{m_1} \times \dots \times \mathbb{R}^{m_k})$. Comme cette tribu contient tous les produits cartésiens $D_1 \times \dots \times D_k$ de boréliens quelconques des \mathbb{R}^{m_j} , on en déduit l'extension de (5.15) avec les D_j au lieu des B'_j , obtenant ainsi l'indépendance de la suite de vecteurs aléatoires Y_1, \dots, Y_k . \square

7. Par le théorème d'unicité des mesures, appliqué dans le cas de mesures finies, cf. Cours IFP 2003–04, Chapitre 1, th. 1.34, disponible à l'URL <http://math.univ-lille1.fr/~suquet/>

Preuve informelle de l'hérédité par transformations boréliennes. Soient Ω et Ω' deux ensembles quelconques, \mathcal{F}' une tribu sur Ω' et $f : \Omega \rightarrow \Omega'$ quelconque. Notons $f^{-1}(\mathcal{F}')$ la famille $\{f^{-1}(B); B \in \mathcal{F}'\}$, en rappelant que si $B \subset \Omega'$, $f^{-1}(B) = \{\omega \in \Omega; f(\omega) \in B\}$. Il est facile de voir que $f^{-1}(\mathcal{F}')$ est une tribu sur Ω . Cela résulte du fait que la réunion et le passage au complémentaire commutent avec l'inverse ensembliste (cf. proposition 3.3) et que \mathcal{F}' est une tribu sur Ω' . Cette tribu $f^{-1}(\mathcal{F}')$ est appelée tribu engendrée par f (relativement à \mathcal{F}') et notée aussi $\sigma(f)$.

Si Ω est muni lui-même d'une tribu \mathcal{F} , la condition $f^{-1}(\mathcal{F}') \subset \mathcal{F}$ équivaut tout simplement à la mesurabilité de f pour les tribus $\mathcal{F} - \mathcal{F}'$. En particulier si X est une application $\Omega \rightarrow \mathbb{R}^d$, dire que X est un vecteur aléatoire sur (Ω, \mathcal{F}) s'écrit $X^{-1}(\text{Bor}(\mathbb{R}^d)) \subset \mathcal{F}$.

Arrivés là, il nous faut donner une définition qui jette un nouvel éclairage sur l'indépendance des variables ou vecteurs aléatoires.

Définition 5.33 (indépendance de n sous-tribus de \mathcal{F}). Soient (Ω, \mathcal{F}, P) un espace probabilisé et $\mathcal{F}_1, \dots, \mathcal{F}_n$ des sous-tribus⁸ de \mathcal{F} . On dit qu'elles sont indépendantes si

$$\forall A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n, \quad P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i). \quad (5.16)$$

Grâce à cette définition, on peut voir que l'indépendance des variables aléatoires X_1, \dots, X_n est exactement l'indépendance des tribus engendrées⁹ $\sigma(X_1), \dots, \sigma(X_n)$. On peut réécrire de la même façon l'indépendance des vecteurs aléatoires vue à la définition 5.29, la seule adaptation étant que $\sigma(X_i) = X_i^{-1}(\text{Bor}(\mathbb{R}^{d_i}))$ au lieu de $X_i^{-1}(\text{Bor}(\mathbb{R}))$.

Il est clair d'après la définition 5.33 que si les tribus $\mathcal{F}_1, \dots, \mathcal{F}_n$ sont des sous-tribus de \mathcal{F} indépendantes et si pour chaque i , \mathcal{G}_i est une sous-tribu de \mathcal{F}_i , alors $\mathcal{G}_1, \dots, \mathcal{G}_n$ sont aussi des sous-tribus indépendantes de \mathcal{F} . En d'autres termes, l'indépendance des tribus est héritée par leurs sous-tribus.

Disposant maintenant de tous les ingrédients, nous pouvons prouver la proposition 5.30 comme suit. D'abord l'indépendance de la suite X_1, \dots, X_n passe à celle des vecteurs blocs $Y_j = (X_{n_{j-1}+1}, \dots, X_{n_j})$, $j = 1, \dots, k$, comme nous l'avons établi ci-dessus. Cette indépendance des Y_j équivaut à celle des sous-tribus $\mathcal{F}_j := \sigma(Y_j)$ de \mathcal{F} . Notre but est d'établir l'indépendance des vecteurs images $Z_j = h_j(Y_j)$, qui équivaut à l'indépendance des tribus $\mathcal{G}_j := \sigma(Z_j)$. Celle-ci découle par héritage de l'indépendance des tribus $\sigma(Y_j)$, $j = 1, \dots, k$ en raison des inclusions $\sigma(Z_j) \subset \sigma(Y_j)$ que l'on peut justifier comme suit :

$$\begin{aligned} \sigma(Z_j) &= \{Z_j^{-1}(B); B \in \text{Bor}(\mathbb{R}^{d_j})\} \\ &= \{(h_j \circ Y_j)^{-1}(B); B \in \text{Bor}(\mathbb{R}^{d_j})\} \\ &= \{Y_j^{-1}(h_j^{-1}(B)); B \in \text{Bor}(\mathbb{R}^{d_j})\} \end{aligned}$$

et $h_j : \mathbb{R}^{m_j} \rightarrow \mathbb{R}^{d_j}$ étant borélienne, $h_j^{-1}(B) \in \text{Bor}(\mathbb{R}^{m_j})$ pour tout $B \in \text{Bor}(\mathbb{R}^{d_j})$, d'où

$$\sigma(Z_j) \subset \{Y_j^{-1}(C); C \in \text{Bor}(\mathbb{R}^{m_j})\} = \sigma(Y_j).$$

8. Une sous-tribu de \mathcal{F} est une sous-famille de \mathcal{F} qui possède encore la structure de tribu.

9. Qui sont bien des sous-tribus de \mathcal{F} , à cause de la mesurabilité des X_i .

□

Définition 5.34 (indépendance d'une suite infinie de v.a.). Soit $(X_i)_{i \in \mathbb{N}}$, une suite de variables aléatoires définies sur le même espace probabilisé (Ω, \mathcal{F}, P) . On dit qu'elles sont indépendantes si toute sous-suite finie est indépendante au sens de la définition 5.27 i.e. pour tout ensemble fini $\emptyset \neq K \subset \mathbb{N}$, et toute famille $(B_i)_{i \in K}$ de boréliens de \mathbb{R} ,

$$P\left(\bigcap_{i \in K} \{X_i \in B_i\}\right) = \prod_{i \in K} P(X_i \in B_i). \quad (5.17)$$

Cette définition se généralise au prix d'alourdissements d'écriture à une suite de vecteurs aléatoires.

5.2.2 Indépendance des composantes

Nous examinons maintenant la caractérisation de l'indépendance des composantes d'un vecteur aléatoire, en nous restreignant aux deux cas particuliers importants des vecteurs aléatoires discrets ou à densité.

Proposition 5.35 (vecteur aléatoire discret à composantes indépendantes). Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire discret. Ses composantes X_i sont indépendantes si et seulement si

$$\forall (x_1, \dots, x_d) \in X(\Omega), \quad P(X_1 = x_1, \dots, X_d = x_d) = P(X_1 = x_1) \dots P(X_d = x_d). \quad (5.18)$$

Preuve. L'indépendance des X_i implique clairement (5.18) en prenant $B_i = \{x_i\}$, pour $i = 1, \dots, d$ dans la définition 5.27. Pour la réciproque, on suppose que le vecteur aléatoire discret X vérifie (5.18) et il s'agit de montrer que $P_X(B) = P_{X_1}(B_1) \dots P_{X_d}(B_d)$, pour tout $B := B_1 \times \dots \times B_d$ où les B_i sont des boréliens quelconques de \mathbb{R} . Comme X est discret, chaque $X_i(\Omega)$ est au plus dénombrable (cf. le commentaire après la définition 5.8) donc $E := X_1(\Omega) \times \dots \times X_d(\Omega)$ est au plus dénombrable. De plus E contient $X(\Omega)$ et si $x \in E \setminus X(\Omega)$, $P(X = x) = 0$. Ces remarques nous permettent de démarrer le calcul de $P_X(B)$ ainsi :

$$P_X(B) = \left(\sum_{x \in X(\Omega)} P(X = x) \delta_x \right)(B) = \sum_{x \in X(\Omega)} P(X = x) \delta_x(B) = \sum_{x \in C} P(X = x),$$

en notant $C = E \cap B = C_1 \times \dots \times C_d$, où $C_i := (X_i(\Omega) \cap B_i)$ (vérifiez !). En utilisant

l'hypothèse (5.18) et le produit de d « séries » à termes positifs¹⁰ :

$$\begin{aligned}
 P_{(X_1, \dots, X_d)}(B) &= \sum_{(x_1, \dots, x_d) \in C} P_{(X_1, \dots, X_d)}(\{(x_1, \dots, x_d)\}) \\
 &= \sum_{(x_1, \dots, x_d) \in C} P(X_1 = x_1, \dots, X_d = x_d) \\
 &= \sum_{(x_1, \dots, x_d) \in C_1 \times \dots \times C_d} P(X_1 = x_1) \dots P(X_d = x_d) \quad (5.19)
 \end{aligned}$$

$$\begin{aligned}
 &= \left(\sum_{x_1 \in C_1} P(X_1 = x_1) \right) \times \dots \times \left(\sum_{x_d \in C_d} P(X_d = x_d) \right) \quad (5.20) \\
 &= P(X_1 \in B_1) \dots P(X_d \in B_d).
 \end{aligned}$$

Les boréliens B_i étant quelconques, ceci établit l'indépendance des variables aléatoires X_1, \dots, X_d . \square

Passons au cas d'un vecteur aléatoire à densité, pour lequel il est commode d'introduire la notation suivante. Si f_1, \dots, f_d sont des applications $\mathbb{R} \rightarrow \mathbb{R}$, on définit l'application « produit tensoriel » des f_i par

$$f_1 \otimes \dots \otimes f_d : \mathbb{R}^d \rightarrow \mathbb{R}, \quad (x_1, \dots, x_d) \mapsto f_1(x_1) \dots f_d(x_d). \quad (5.21)$$

On prendra bien garde de ne pas confondre ce produit tensoriel avec le produit ordinaire. Par exemple si f est l'identité sur \mathbb{R} , $f \otimes f$ est l'application $(s, t) \mapsto st$, tandis que $ff = f^2 : \mathbb{R} \rightarrow \mathbb{R}$, $s \mapsto s^2$.

Proposition 5.36 (vecteur à densité et indépendance des composantes). *Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de \mathbb{R}^d .*

- a) *Si les composantes de X sont indépendantes et si chaque X_i ($i = 1, \dots, d$) a une densité f_{X_i} , alors X admet pour densité la fonction $f_{X_1} \otimes \dots \otimes f_{X_d}$.*
- b) *Si X admet une densité de la forme $f = f_1 \otimes \dots \otimes f_d$, où les f_i sont des fonctions $\mathbb{R} \setminus K_i \rightarrow \mathbb{R}^+$ (les K_i étant finis éventuellement vides), alors les v.a. réelles X_i sont indépendantes et chaque X_i admet une densité $f_{X_i} = c_i f_i$ avec des constantes $c_i \in]0, +\infty[$ dont le produit $c_1 \dots c_d$ vaut 1.*

Preuve du a). Calculons $P_X(C)$ pour $C := [a_1, b_1] \times \dots \times [a_d, b_d]$ pavé fermé borné quelconque de \mathbb{R}^d . On commence par écrire l'évènement $\{X \in C\}$ comme une intersection d'évènements en notant que

$$\{X \in C\} = \{(X_1, \dots, X_d) \in [a_1, b_1] \times \dots \times [a_d, b_d]\} = \{\forall i = 1, \dots, d, X_i \in [a_i, b_i]\},$$

10. Certaines d'entre elles peuvent n'être que des sommes finies. La formule utilisée pour passer de (5.19) à (5.20) repose sur le théorème de sommation par paquets des familles à termes positifs et sa vérification est essentiellement la même que pour le produit de deux séries.

d'où, les X_i étant indépendantes et à densité,

$$P(X \in C) = P\left(\bigcap_{i=1}^d \{X_i \in [a_i, b_i]\}\right) = \prod_{i=1}^d P(X_i \in [a_i, b_i]) = \prod_{i=1}^d \int_{a_i}^{b_i} f_{X_i}(t_i) dt_i.$$

Par un corollaire classique du théorème de Fubini-Tonelli, ce produit d'intégrales de fonctions positives peut s'écrire comme une intégrale multiple :

$$\prod_{i=1}^d \int_{a_i}^{b_i} f_{X_i}(t_i) dt_i = \int_{[a_1, b_1] \times \dots \times [a_d, b_d]} f_{X_1}(t_1) \dots f_{X_d}(t_d) dt_1 \dots dt_d.$$

Finalement, nous obtenons ainsi

$$P(X \in C) = \int_C f_{X_1} \otimes \dots \otimes f_{X_d}(t_1, \dots, t_d) dt_1 \dots dt_d$$

et cette formule est vraie pour tout pavé fermé borné C . Pour en déduire en vertu de la définition 5.10 que le vecteur aléatoire X admet bien pour densité $f := f_{X_1} \otimes \dots \otimes f_{X_d}$, il nous reste seulement à vérifier que f ainsi construite est bien une densité de probabilité sur \mathbb{R}^d . D'abord chaque f_{X_i} est définie et positive sauf peut-être en un nombre fini de points. Chacun de ces points de non-définition génère pour f un ensemble de non-définition qui est un hyperplan de \mathbb{R}^d . Donc f est définie sauf peut-être sur une réunion finie H d'hyperplans de \mathbb{R}^d et positive sur son ensemble de définition. Ensuite f est localement Riemann intégrable car intégrable sur chaque pavé fermé borné inclus dans $\mathbb{R}^d \setminus H$ et on peut prendre comme suite de compacts K_n épuisant $\mathbb{R}^d \setminus H$ une suite croissante de réunions finies de pavés fermés bornés (un pavé dans chaque composante connexe de $\mathbb{R}^d \setminus H$). Pour vérifier la convergence de l'intégrale généralisée sur \mathbb{R}^d de f , on prend $2d$ suites $a_{i,n} \downarrow -\infty$ et $b_{i,n} \uparrow +\infty$, de façon à former une suite $C_n = [a_{1,n}, b_{1,n}] \times \dots \times [a_{d,n}, b_{d,n}]$ croissante pour l'inclusion et de réunion \mathbb{R}^d . Par continuité séquentielle croissante de la probabilité P_X , on obtient

$$1 = P_X(\mathbb{R}^d) = \lim_{n \rightarrow +\infty} P_X(C_n) = \lim_{n \rightarrow +\infty} \int_{[a_{1,n}, b_{1,n}] \times \dots \times [a_{d,n}, b_{d,n}]} f(t_1, \dots, t_d) dt_1 \dots dt_d.$$

On en déduit que l'intégrale généralisée $\int_{\mathbb{R}^d} f(t_1, \dots, t_d) dt_1 \dots dt_d$ converge et vaut 1. La fonction f est donc bien une densité de probabilité sur \mathbb{R}^d et ceci achève la preuve du a). \square

Preuve du b). Puisque X admet pour densité $f = f_1 \otimes \dots \otimes f_d$, on sait (proposition 5.14) que ses lois marginales sont toutes à densité et qu'une densité de X_i s'obtient en intégrant f par rapport à toutes les variables d'indice différent de i . Ceci nous donne :

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{i-1} \times \mathbb{R}^{d-i}} f(t_1) \dots f_{i-1}(t_{i-1}) f_i(x_i) f_{i+1}(t_{i+1}) \dots f_d(t_d) dt_1 \dots dt_{i-1} dt_{i+1} \dots dt_d.$$

Dans cette intégration avec x_i fixé et relativement aux variables muettes t_j , $j \neq i$, $f_i(x_i)$ est une constante que l'on peut sortir de l'intégrale multiple. On obtient ainsi la relation

$f_{X_i}(x_i) = c_i f_i(x_i)$, où c_i désigne l'intégrale multiple relative aux t_j ($j \neq i$) et ne dépend pas de x_i . Ce calcul étant valable pour tout réel x_i , on a bien $f_{X_i} = c_i f_i$, avec une constante c_i dans \mathbb{R}_+ (pour l'instant). En fait $c_i \neq 0$, car sinon f_{X_i} serait identiquement nulle sur son ensemble de définition (à l'exception peut-être d'un ensemble de mesure de Lebesgue nulle), ce qui l'empêcherait d'être une densité car cela impliquerait la nullité de $\int_{-\infty}^{+\infty} f_{X_i}(t) dt$.

D'après la remarque 5.28-3, pour établir l'indépendance des X_1, \dots, X_d , il suffit de tester l'égalité (5.11) pour des boréliens B_i de la forme $[a_i, b_i]$. Notons $C := [a_1, b_1] \times \dots \times [a_d, b_d]$. Compte-tenu de la forme particulière de f , on a ici en appliquant à nouveau le corollaire du théorème de Fubini-Tonnelli pour les produits $f_1 \otimes \dots \otimes f_d$,

$$P(X \in C) = \int_C f_1(t_1) \dots f_d(t_d) dt_1 \dots dt_d = \prod_{i=1}^d \int_{a_i}^{b_i} f_i(t_i) dt_i. \quad (5.22)$$

D'autre part, en utilisant les relations $f_{X_i} = c_i f_i$, on obtient :

$$\prod_{i=1}^d P(X_i \in [a_i, b_i]) = \prod_{i=1}^d \int_{a_i}^{b_i} c_i f_i(t_i) dt_i = (c_1 \dots c_d) \prod_{i=1}^d \int_{a_i}^{b_i} f_i(t_i) dt_i. \quad (5.23)$$

Par comparaison de (5.22) et (5.23), on voit alors que pour tout pavé fermé borné $C = [a_1, b_1] \times \dots \times [a_d, b_d]$,

$$\prod_{i=1}^d P(X_i \in [a_i, b_i]) = (c_1 \dots c_d) P(X \in C). \quad (5.24)$$

En prenant une suite $C_n = [a_{1,n}, b_{1,n}] \times \dots \times [a_{d,n}, b_{d,n}]$ croissante pour l'inclusion et de réunion \mathbb{R}^d , on en déduit par continuité séquentielle croissante de P_X et des P_{X_i} que

$$\prod_{i=1}^d P(X_i \in \mathbb{R}) = (c_1 \dots c_d) P(X \in \mathbb{R}^d),$$

d'où

$$c_1 \dots c_d = 1. \quad (5.25)$$

En réinjectant cette valeur dans (5.24), on conclut à l'indépendance des X_i . Notons en passant que (5.25) interdit que l'un des c_i vaille $+\infty$ (on a déjà vu qu'aucun ne peut être nul). \square

Exemple 5.37. Supposons que les variables aléatoires réelles X_1, \dots, X_d soient indépendantes et que pour $i = 1, \dots, d$, X_i soit gaussienne de loi $\mathfrak{N}(0, \sigma_i)$, avec $\sigma_i > 0$. La proposition 5.36 a) nous dit alors que le vecteur $X = (X_1, \dots, X_d)$ admet pour densité :

$$f(t_1, \dots, t_d) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sigma_1 \dots \sigma_d} \exp \left(-\frac{t_1^2}{2\sigma_1^2} - \dots - \frac{t_d^2}{2\sigma_d^2} \right). \quad (5.26)$$

Le vecteur X suit une loi gaussienne sur \mathbb{R}^d . Il s'agit ici d'un cas particulier de *vecteur gaussien*. Il existe des lois gaussiennes sur \mathbb{R}^d pour lesquelles les composantes ne sont pas indépendantes. Par ailleurs, il existe des lois gaussiennes sur \mathbb{R}^d n'ayant pas de densité¹¹.

Remarque 5.38. L'exemple 5.37 ne contredit pas la remarque 5.6 qui affirme que la seule connaissance des lois marginales ne suffit pas à reconstruire la loi du vecteur. En effet ici outre la connaissance des lois marginales, on dispose d'une information supplémentaire essentielle, l'indépendance des composantes. C'est d'ailleurs un exemple d'un principe tout à fait général illustré par les propositions 5.35 et 5.36 : si on connaît les lois marginales *et si on sait* que les composantes sont indépendantes, alors on peut reconstruire la loi du vecteur aléatoire¹².

Exemple 5.39. Soit $X = (X_1, X_2)$ un vecteur aléatoire de densité f donnée par

$$f(s, t) := \frac{1}{(2\pi^3)^{1/2}} \frac{1}{(1 + s^2)(te^t)^{1/2}} \mathbf{1}_{]0, +\infty[}(t),$$

avec la convention habituelle pour l'usage des indicatrices dans les formules explicites (cf. remarque 3.21). La densité est bien le produit d'une fonction de la seule variable s par une fonction de la seule variable t . On peut prendre par exemple $f_1(s) := (1 + s^2)^{-1}$ et $f_2(t) := (2\pi^3)^{-1/2}(te^t)^{-1/2} \mathbf{1}_{]0, +\infty[}(t)$. On en déduit par la proposition 5.36 b) que les v.a. X_1 et X_2 sont indépendantes et à densité. On sait de plus qu'alors X_i admet une densité $f_{X_i} = c_i f_i$ ($i = 1, 2$). On détermine c_1 en écrivant :

$$1 = \int_{-\infty}^{+\infty} f_{X_1}(s) ds = \int_{-\infty}^{+\infty} \frac{c_1}{1 + s^2} ds = c_1 \pi.$$

Donc $c_1 = 1/\pi$ et comme $c_1 c_2 = 1$, $c_2 = \pi$. Les densités marginales sont donc finalement :

$$f_{X_1} : s \mapsto \frac{1}{\pi(1 + s^2)}, \quad f_{X_2} : t \mapsto \frac{1}{\sqrt{2\pi t}} e^{-t/2} \mathbf{1}_{]0, +\infty[}(t).$$

On voit que X_1 suit la loi de Cauchy standard $\text{Cau}(0, 1)$. Quant à la loi de X_2 , c'est une loi $\chi^2(1)$, *i.e.* la loi de Z^2 où Z est gaussienne $\mathfrak{N}(0, 1)$. On vous laisse le soin de vérifier cette dernière affirmation en calculant la loi de Z^2 . Ceci vous permettra de vérifier *a posteriori* que f était bien une densité de probabilité sur \mathbb{R}^2 grâce à la proposition 5.36 a).

De même que la seule connaissance des lois des v.a. X et Y ne suffit pas en général pour déterminer la loi du vecteur aléatoire (X, Y) , elle ne suffit pas non plus pour connaître la loi de la v.a. $X + Y$. Voici un exemple presque trivial pour s'en convaincre. On prend d'abord X de loi $\text{Bern}(1/2)$, donc $P(X = 0) = P(X = 1) = 1/2$ et $Y := 1 - X$. Alors Y suit aussi la loi $\text{Bern}(1/2)$ car $P(Y = 0) = P(X = 1) = 1/2$ et $P(Y = 1) =$

11. Dans ce cas, il existe un sous-espace affine de \mathbb{R}^d , strictement inclus dans \mathbb{R}^d et de probabilité 1 pour cette loi gaussienne.

12. Pour les lecteurs n'ayant pas sauté la démonstration informelle de l'hérédité de l'indépendance, signalons que les composantes de X sont indépendantes si et seulement si $P_X = P_{X_1} \otimes \cdots \otimes P_{X_d}$, probabilité produit des lois marginales.

$P(X = 0) = 1/2$. La somme $X + Y$ est la variable constante 1, sa loi est donc la masse de Dirac δ_1 . Prenons maintenant $Y' := X$, alors Y' suit encore la loi Bern(1/2) et $X + Y' = 2X$ a pour loi $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_2$.

Supposons maintenant que les variables aléatoires réelles X et Y sont indépendantes. Comme $X + Y = s(X, Y)$, où s est la fonction borélienne $s : \mathbb{R}^2 \rightarrow \mathbb{R}, (x, y) \mapsto x + y$, on devrait pouvoir déterminer la loi de $X + Y$, à partir des lois de X et de Y puisque l'indépendance permet de reconstruire la loi de (X, Y) à partir de ses lois marginales. Nous allons traiter ce problème dans les deux cas particuliers où X et Y sont toutes deux discrètes ou toutes deux à densité. On pourra voir en exercice un autre cas particulier, X discrète et Y à densité.

Proposition 5.40 (loi de la somme de deux v.a. discrètes). *Si X et Y sont deux v.a. discrètes définies sur le même espace probabilisé (Ω, \mathcal{F}, P) , la loi de la v.a. discrète $Z := X + Y$ est donnée par :*

$$\forall z \in Z(\Omega), \quad P(Z = z) = \sum_{x \in X(\Omega)} P(X = x, Y = z - x) \quad (5.27)$$

$$= \sum_{y \in Y(\Omega)} P(X = z - y, Y = y). \quad (5.28)$$

Si de plus X et Y sont indépendantes, P_Z peut se calculer explicitement à partir des lois P_X et P_Y par les formules :

$$\forall z \in Z(\Omega), \quad P(Z = z) = \sum_{x \in X(\Omega)} P(X = x)P(Y = z - x) \quad (5.29)$$

$$= \sum_{y \in Y(\Omega)} P(X = z - y)P(Y = y). \quad (5.30)$$

Preuve. Vérifions d'abord que Z est discrète, c'est-à-dire que $Z(\Omega)$ est au plus dénombrable. Pour cela, on note que l'ensemble $A := \{(X(\omega), Y(\omega)); \omega \in \Omega\}$ est au plus dénombrable parce qu'inclus dans $X(\Omega) \times Y(\Omega)$, produit cartésien de deux ensembles au plus dénombrables. L'application $s : A \rightarrow Z(\Omega), (x, y) \mapsto x + y$ est *surjective*, donc $Z(\Omega)$ est au plus dénombrable, cf. proposition 1.41. Nous reviendrons ci-dessous (remarque 5.41) sur la description de $Z(\Omega)$ que nous n'avons pas besoin d'explicitier davantage à ce stade.

Justifions (5.27). On commence par découper Ω suivant les valeurs de X en union au plus dénombrable d'évènements deux à deux disjoints :

$$\Omega = \bigcup_{x \in X(\Omega)} \{X = x\}.$$

On en déduit que pour tout $z \in Z(\Omega)$,

$$\begin{aligned} \{Z = z\} &= \left(\bigcup_{x \in X(\Omega)} \{X = x\} \right) \cap \{Z = z\} = \bigcup_{x \in X(\Omega)} (\{X = x\} \cap \{Z = z\}) \\ &= \bigcup_{x \in X(\Omega)} \{X = x, Y = z - x\}. \end{aligned}$$

Cette dernière union est au plus dénombrable et les évènements concernés deux à deux disjoints (certains pouvant être vides). Par σ -additivité de P on en déduit (5.27). Cette formule montre qu'il est toujours possible de calculer la loi de $X + Y$ si l'on connaît la loi $P_{(X,Y)}$ du couple (X, Y) . Si de plus X et Y sont indépendantes, alors

$$P(X = x, Y = z - x) = P(X = x)P(Y = z - x),$$

d'où (5.29).

Les formules (5.28) et (5.30) s'obtiennent en échangeant les rôles joués par X et Y dans la preuve ci-dessus. \square

Remarque 5.41. Si l'on connaît seulement la loi de X et celle de Y , on ne peut même pas déterminer $Z(\Omega)$. Pour vous en convaincre, revoyez le contre exemple avec les lois Bern(1/2) donné ci-dessus où $Z(\Omega)$ peut être selon le cas $\{1\}$ ou $\{0, 2\}$. D'une manière générale, il est clair que l'on a toujours

$$Z(\Omega) \subset \{x + y; x \in X(\Omega), y \in Y(\Omega)\}, \quad (5.31)$$

l'inclusion pouvant être stricte. Cette inclusion est une égalité dans le cas où X et Y sont indépendantes *et* vérifient la condition supplémentaire suivante

$$\forall x \in X(\Omega), \quad P(X = x) \neq 0, \quad \forall y \in Y(\Omega), \quad P(Y = y) \neq 0. \quad (5.32)$$

Pour le voir, montrons que si (x, y) est un couple quelconque de $X(\Omega) \times Y(\Omega)$, il existe au moins un $\omega \in \Omega$ tel que $X(\omega) = x$ et $Y(\omega) = y$, donc pour cet ω , $Z(\omega) = X(\omega) + Y(\omega) = x + y$. En effet par indépendance de X et Y et la condition (5.32),

$$P(X = x, Y = y) = P(X = x)P(Y = y) \neq 0,$$

donc l'évènement $\{\omega \in \Omega; X(\omega) = x, Y(\omega) = y\}$ n'est pas vide.

Exemple 5.42 (somme de deux v.a. de Poisson indépendantes). Si X et Y sont indépendantes et suivent la loi de Poisson de paramètre α et β respectivement, $Z := X + Y$ suit la loi de Poisson de paramètre $\gamma = \alpha + \beta$.

Vérification. Ici $X(\Omega) = Y(\Omega) = \mathbb{N}$ et

$$\forall i \in \mathbb{N}, \quad P(X = i) = \frac{e^{-\alpha} \alpha^i}{i!}, \quad \forall j \in \mathbb{N}, \quad P(Y = j) = \frac{e^{-\beta} \beta^j}{j!},$$

donc (5.32) est vérifiée et $Z(\Omega) = \{k = i + j; i \in \mathbb{N}, j \in \mathbb{N}\} = \mathbb{N}$. Appliquons maintenant (5.29) en notant que pour $i > k$, $k - i \notin Y(\Omega)$, donc $P(Y = k - i) = 0$:

$$\begin{aligned} \forall k \in \mathbb{N}, \quad P(Z = k) &= \sum_{i \in \mathbb{N}} P(X = i)P(Y = k - i) = \sum_{i=0}^k \frac{e^{-\alpha} \alpha^i}{i!} \frac{e^{-\beta} \beta^{k-i}}{(k-i)!} \\ &= e^{-(\alpha+\beta)} \frac{1}{k!} \sum_{i=0}^k \frac{k!}{i!(k-i)!} \alpha^i \beta^{k-i} \\ &= \frac{e^{-(\alpha+\beta)} (\alpha + \beta)^k}{k!}, \end{aligned}$$

ce qui montre que Z suit la loi $\text{Pois}(\gamma)$ pour $\gamma = \alpha + \beta$. \square

Proposition 5.43 (somme de deux v.a. à densité indépendantes). *Si les variables aléatoires réelles indépendantes X et Y admettent pour densités respectives f et g , leur somme $S = X + Y$ admet pour densité le produit de convolution $f * g$ défini sur \mathbb{R} par*

$$(f * g)(s) = \int_{-\infty}^{+\infty} f(s-t)g(t) dt = \int_{-\infty}^{+\infty} f(t)g(s-t) dt. \quad (5.33)$$

Preuve. Notons d'abord que les deux intégrales généralisées figurant dans (5.33) existent toujours comme éléments de $\overline{\mathbb{R}}_+$ et que l'on passe de la première à la deuxième par le changement de variable $t \mapsto s - t$ (pour s fixé).

Calculons $\mathbf{E}h(S)$ pour h continue bornée quelconque à l'aide de la loi du couple (X, Y) . On sait par la proposition 5.36 a) que cette loi $P_{(X,Y)}$ admet pour densité la fonction $p = f \otimes g$. En utilisant la proposition 5.17 appliquée à la fonction composée $(x, y) \mapsto h(x + y)$, on obtient :

$$\mathbf{E}h(S) = \mathbf{E}h(X + Y) = \int_{\mathbb{R}^2} h(x + y)p(x, y) dx dy = \int_{\mathbb{R}^2} h(x + y)f(x)g(y) dx dy.$$

En effectuant dans cette dernière intégrale double le changement de variable linéaire :

$$(s, t) = (x + y, y), \quad \text{d'inverse} \quad (x, y) = (s - t, t),$$

on obtient (noter que le déterminant du changement de variable vaut ici 1) :

$$\mathbf{E}h(S) = \int_{\mathbb{R}^2} h(s)f(s-t)g(t) ds dt = \int_{-\infty}^{+\infty} h(s) \left\{ \int_{-\infty}^{+\infty} f(s-t)g(t) dt \right\} ds.$$

Ainsi pour toute $h : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée sur \mathbb{R} , on a

$$\mathbf{E}h(S) = \int_{-\infty}^{+\infty} h(s)(f * g)(s) ds. \quad (5.34)$$

Si on prend en particulier $h = 1$ constante sur \mathbb{R} , (5.34) nous donne $1 = \int_{-\infty}^{+\infty} (f * g)(s) ds$, ce qui nous montre¹³ que $f * g$ est une densité de probabilité sur \mathbb{R} . Soit Z une variable aléatoire de densité $f * g$. Alors par la proposition 4.42,

$$\mathbf{E}h(Z) = \int_{-\infty}^{+\infty} h(s)(f * g)(s) ds = \mathbf{E}h(S)$$

pour toute h continue bornée sur \mathbb{R} . Par la proposition 5.16 appliquée en dimension 1, on en déduit que S et Z ont même loi, donc que S admet pour densité $f * g$. \square

Exemple 5.44 (somme de deux v.a. exponentielles indépendantes). Soient X et Y deux variables aléatoires indépendantes suivant des lois exponentielles de paramètres respectifs

13. En toute rigueur, il faudrait aussi vérifier que $f * g$ est bien Riemann intégrable sur tout intervalle fermé borné de son ensemble de définition (*i.e.* de l'ensemble des s tels que $(f * g)(s) < +\infty$).

a et b . Calculons la densité $h = f * g$ de la loi de $X + Y$, avec ici $f(t) = a e^{-at} \mathbf{1}_{\mathbb{R}_+}(t)$ et $g(t) = b e^{-bt} \mathbf{1}_{\mathbb{R}_+}(t)$.

$$h(s) = \int_{-\infty}^{+\infty} a e^{-a(s-t)} b e^{-bt} \mathbf{1}_{\mathbb{R}_+}(s-t) \mathbf{1}_{\mathbb{R}_+}(t) dt = ab e^{-as} \int_{-\infty}^{+\infty} e^{(a-b)t} \mathbf{1}_{\mathbb{R}_+}(s-t) \mathbf{1}_{\mathbb{R}_+}(t) dt.$$

On peut réécrire le produit d'indicatrices comme suit :

$$\mathbf{1}_{\mathbb{R}_+}(s-t) \mathbf{1}_{\mathbb{R}_+}(t) = \begin{cases} 1 & \text{si } t \geq 0 \text{ et } s-t \geq 0 \\ 0 & \text{sinon} \end{cases} = \begin{cases} \mathbf{1}_{[0,s]}(t) & \text{si } s \geq 0 \\ 0 & \text{sinon} \end{cases} = \mathbf{1}_{\mathbb{R}_+}(s) \mathbf{1}_{[0,s]}(t).$$

En reportant ceci dans l'intégrale ci-dessus, il vient

$$h(s) = ab e^{-as} \mathbf{1}_{\mathbb{R}_+}(s) \int_{-\infty}^{+\infty} e^{(a-b)t} \mathbf{1}_{[0,s]}(t) dt = ab e^{-as} \mathbf{1}_{\mathbb{R}_+}(s) \int_0^s e^{(a-b)t} dt.$$

Cette dernière intégrale est une intégrale de Riemann ordinaire qui se calcule par primitive en distinguant les cas $a \neq b$ et $a = b$.

Si $a \neq b$, on obtient ainsi

$$h(s) = ab e^{-as} \mathbf{1}_{\mathbb{R}_+}(s) \left[\frac{e^{(a-b)t}}{a-b} \right]_0^s = \frac{ab e^{-as}}{a-b} (e^{(a-b)s} - 1) \mathbf{1}_{\mathbb{R}_+}(s) = \frac{ab}{a-b} (e^{-bs} - e^{-as}) \mathbf{1}_{\mathbb{R}_+}(s).$$

À titre de précaution, on peut vérifier que cette densité h est bien positive en notant que pour $s \geq 0$, $(e^{-bs} - e^{-as})/(b-a)$ est le taux d'accroissement entre a et b de la fonction décroissante $u \mapsto e^{-su}$, donc que ce quotient est négatif.

Si $a = b$, on obtient

$$h(s) = a^2 e^{-as} \mathbf{1}_{\mathbb{R}_+}(s) \int_0^s dt = a^2 s e^{-as} \mathbf{1}_{\mathbb{R}_+}(s).$$

5.2.3 Indépendance et espérance de produits

Une propriété essentielle de l'indépendance des suites de variables aléatoires est que l'espérance d'un produit est égale au produit des espérances. La formulation précise, d'une portée encore plus générale, est la suivante.

Théorème 5.45. *Pour $i = 1, \dots, n$, soient $X_i : \Omega \rightarrow \mathbb{R}^{d_i}$ des vecteurs aléatoires indépendants et $h_i : \mathbb{R}^{d_i} \rightarrow \mathbb{R}$, des fonctions boréliennes telles que les variables aléatoires réelles $h_i(X_i)$ soient intégrables. Alors la v.a. $h_1(X_1) \dots h_n(X_n)$ est intégrable et*

$$\mathbf{E}(h_1(X_1) \dots h_n(X_n)) = \prod_{i=1}^n \mathbf{E}h_i(X_i). \quad (5.35)$$

En particulier si pour tout i , $d_i = 1$ et h_i est l'identité sur \mathbb{R} , on obtient l'interversion espérance produit.

Corollaire 5.46. Si X_1, \dots, X_n sont des variables aléatoires réelles indépendantes et intégrables, leur produit $X_1 \dots X_n$ est intégrable et

$$\mathbf{E}(X_1 \dots X_n) = (\mathbf{E}X_1) \dots (\mathbf{E}X_n). \quad (5.36)$$

Preuve. Par hérédité de l'indépendance, plus précisément par la version de la proposition 5.30 où la suite de départ X_1, \dots, X_n est une famille hétéroclite de vecteurs aléatoires indépendants, le problème se réduit à la preuve de l'intégrabilité de $Y = Y_1 \dots Y_n$ et de l'égalité

$$\mathbf{E}(Y_1 \dots Y_n) = (\mathbf{E}Y_1) \dots (\mathbf{E}Y_n), \quad (5.37)$$

pour toute suite finie Y_1, \dots, Y_n ($n \geq 2$) de v.a. intégrables et indépendantes.

Bien qu'il soit possible de traiter le cas n quelconque d'un coup, nous allons réduire le problème au cas $n = 2$, pour alléger les notations. Admettons donc pour un instant que nous ayons établi (5.37) dans le cas $n = 2$. On fait alors une récurrence finie sur $2 \leq k < n$ en prenant pour hypothèse que (5.37) est vérifiée pour les produits de k facteurs. Alors l'indépendance des Y_i ($1 \leq i \leq n$) implique celle de (Y_1, \dots, Y_k) et de Y_{k+1} , par hérédité pour les blocs disjoints cf. corollaire 5.31 et remarque 5.28-1. Par l'hypothèse de récurrence, $Y_1 \dots Y_k$ est intégrable et comme Y_{k+1} l'est aussi, leur produit l'est encore par le cas $n = 2$. On a donc

$$\begin{aligned} \mathbf{E}(Y_1 \dots Y_k Y_{k+1}) &= \mathbf{E}((Y_1 \dots Y_k) Y_{k+1}) \\ &= (\mathbf{E}(Y_1 \dots Y_k)) (\mathbf{E}Y_{k+1}) \quad \text{par le cas } n = 2 \\ &= (\mathbf{E}Y_1) \dots (\mathbf{E}Y_k) (\mathbf{E}Y_{k+1}) \quad \text{par l'hypothèse de récurrence.} \end{aligned}$$

Ainsi la validité de (5.37) au rang k implique sa validité au rang $k + 1$, ce qui achève la partie « hérédité » de notre récurrence finie.

Il nous reste maintenant à prouver que pour toutes v.a. réelles Y_1 et Y_2 indépendantes et intégrables,

$$Y_1 Y_2 \text{ est intégrable et } \mathbf{E}(Y_1 Y_2) = (\mathbf{E}Y_1)(\mathbf{E}Y_2). \quad (5.38)$$

Nous allons traiter successivement les cas suivants.

1. $Y_i = \mathbf{1}_{A_i}$, $i = 1, 2$ où les $A_i \in \mathcal{F}$ sont des événements ;
2. Y_1 et Y_2 variables aléatoires simples ;
3. Y_1 et Y_2 variables aléatoires positives ;
4. Y_1 et Y_2 variables aléatoires réelles intégrables.

Cas 1. Si $Y_i = \mathbf{1}_{A_i}$, pour un $A_i \in \mathcal{F}$, les v.a. Y_i et leur produit sont intégrables car bornées. L'indépendance des v.a. Y_1 et Y_2 implique¹⁴ celle des événements A_1 et A_2 car $A_i = \{\mathbf{1}_{A_i} \in \{1\}\}$. D'autre part, on vérifie facilement (faites-le) que

$$\mathbf{1}_{A_1} \mathbf{1}_{A_2} = \mathbf{1}_{A_1 \cap A_2}.$$

14. En fait *équivalent* à ... (exercice).

Grâce à la formule $\mathbf{E}(\mathbf{1}_A) = P(A)$ et à l'indépendance de A_1 et A_2 , on en déduit

$$\mathbf{E}(Y_1 Y_2) = \mathbf{E}(\mathbf{1}_{A_1 \cap A_2}) = P(A_1 \cap A_2) = P(A_1)P(A_2) = \mathbf{E}(\mathbf{1}_{A_1})\mathbf{E}(\mathbf{1}_{A_2}) = (\mathbf{E}Y_1)(\mathbf{E}Y_2),$$

ce qui achève la vérification de (5.38) dans le cas des variables aléatoires indicatrices. \square

Cas 2. Supposons maintenant que Y_1 et Y_2 sont deux variables aléatoires simples indépendantes. Elles sont alors bornées donc intégrables ainsi que leur produit. Notons $Y_1(\Omega) = \{y_{1,1}, \dots, y_{1,l}\}$, les $y_{1,j}$ étant tous distincts et faisons de même pour $Y_2(\Omega) = \{y_{2,1}, \dots, y_{2,m}\}$. Alors on a les décompositions

$$Y_1 = \sum_{j=1}^l y_{1,j} \mathbf{1}_{A_{1,j}}, \quad Y_2 = \sum_{k=1}^m y_{2,k} \mathbf{1}_{A_{2,k}},$$

avec $A_{1,j} = \{Y_1 = y_{1,j}\} = \{Y_1 \in \{y_{1,j}\}\}$ et $A_{2,k} = \{Y_2 = y_{2,k}\} = \{Y_2 \in \{y_{2,k}\}\}$. On en déduit que pour $1 \leq j \leq l$ et $1 \leq k \leq m$, les deux événements $A_{1,j}$ et $A_{2,k}$ héritent de l'indépendance de Y_1 et Y_2 . En utilisant la linéarité de l'espérance, cette indépendance d'évènements et le cas 1, on obtient :

$$\begin{aligned} \mathbf{E}(Y_1 Y_2) &= \mathbf{E}\left(\sum_{j=1}^l \sum_{k=1}^m y_{1,j} y_{2,k} \mathbf{1}_{A_{1,j}} \mathbf{1}_{A_{2,k}}\right) = \sum_{j=1}^l \sum_{k=1}^m y_{1,j} y_{2,k} \mathbf{E}(\mathbf{1}_{A_{1,j}} \mathbf{1}_{A_{2,k}}) \\ &= \sum_{j=1}^l \sum_{k=1}^m y_{1,j} y_{2,k} \mathbf{E}(\mathbf{1}_{A_{1,j}}) \mathbf{E}(\mathbf{1}_{A_{2,k}}) \\ &= \left(\sum_{j=1}^l y_{1,j} \mathbf{E}(\mathbf{1}_{A_{1,j}})\right) \left(\sum_{k=1}^m y_{2,k} \mathbf{E}(\mathbf{1}_{A_{2,k}})\right) \\ &= \mathbf{E}\left(\sum_{j=1}^l y_{1,j} \mathbf{1}_{A_{1,j}}\right) \mathbf{E}\left(\sum_{k=1}^m y_{2,k} \mathbf{1}_{A_{2,k}}\right) \\ &= \mathbf{E}Y_1 \mathbf{E}Y_2. \end{aligned}$$

Ainsi (5.38) est vérifiée dans le cas des variables aléatoires simples. \square

Cas 3. Soient Y_1 et Y_2 deux variables aléatoires positives indépendantes (on n'a pas besoin de les supposer intégrables ici). On sait (cf. théorème 4.21) que la v.a. positive Y_i est limite d'une suite croissante de v.a. positives simples : $Y_{i,n} \uparrow Y_i$, $i = 1, 2$. Plus précisément, on peut prendre (revoyez la preuve du théorème 4.21 et vérifiez la formule proposée ci-après)

$$Y_{i,n} = h_n(Y_i), \quad h_n : \mathbb{R}_+ \rightarrow \mathbb{R}_+, \quad x \mapsto \min(n, 2^{-n} \lceil 2^n x \rceil).$$

La fonction h_n est croissante donc borélienne. La proposition 5.30 nous dit alors que pour chaque n , les v.a. $Y_{1,n}$ et $Y_{2,n}$ héritent de l'indépendance de Y_1 et Y_2 . Comme ce sont des v.a. simples, le cas 2 nous donne :

$$\forall n \in \mathbb{N}^*, \quad \mathbf{E}(Y_{1,n} Y_{2,n}) = (\mathbf{E}Y_{1,n})(\mathbf{E}Y_{2,n}). \quad (5.39)$$

En notant que $Y_{1,n}Y_{2,n}$ converge en croissant vers Y_1Y_2 , un passage à la limite dans (5.39) et une triple application du théorème de Beppo Levi nous donnent $\mathbf{E}(Y_1Y_2) = \mathbf{E}Y_1 \mathbf{E}Y_2$. \square

Cas 4. Si Y_1 et Y_2 sont deux variables aléatoires réelles indépendantes et intégrables, $|Y_1|$ et $|Y_2|$ sont des v.a. positives indépendantes et par le cas 3,

$$\mathbf{E}|Y_1Y_2| = \mathbf{E}(|Y_1||Y_2|) = (\mathbf{E}|Y_1|)(\mathbf{E}|Y_2|) < +\infty,$$

ce qui établit l'intégrabilité de Y_1Y_2 . Cette intégrabilité implique aussi celle des produits de v.a. positives $Y_1^+Y_2^+$, $Y_1^+Y_2^-$, $Y_1^-Y_2^+$ et $Y_1^-Y_2^-$. D'autre part les facteurs de chacun de ces produits héritent leur indépendance de celle de Y_1 et Y_2 , puisque $Y_i^+ = \max(Y_i, 0)$ et $Y_i^- = \max(-Y_i, 0)$. En utilisant le cas 3 on obtient alors (notez que toutes les espérances intervenant dans ce calcul sont finies) :

$$\begin{aligned} \mathbf{E}(Y_1Y_2) &= \mathbf{E}((Y_1^+ - Y_1^-)(Y_2^+ - Y_2^-)) \\ &= \mathbf{E}(Y_1^+Y_2^+) - \mathbf{E}(Y_1^+Y_2^-) - \mathbf{E}(Y_1^-Y_2^+) + \mathbf{E}(Y_1^-Y_2^-) \\ &= \mathbf{E}Y_1^+\mathbf{E}Y_2^+ - \mathbf{E}Y_1^+\mathbf{E}Y_2^- - \mathbf{E}Y_1^-\mathbf{E}Y_2^+ + \mathbf{E}Y_1^-\mathbf{E}Y_2^- \\ &= (\mathbf{E}Y_1^+ - \mathbf{E}Y_1^-)(\mathbf{E}Y_2^+ - \mathbf{E}Y_2^-) \\ &= (\mathbf{E}Y_1)(\mathbf{E}Y_2), \end{aligned}$$

ce qui achève la vérification de (5.38). \square

La preuve du théorème 5.45 est maintenant complète. \square

Comme sous-produit de la démonstration ci-dessus et notamment du cas 3, on a le résultat suivant.

Proposition 5.47. *Le théorème 5.45 s'applique sans condition d'intégrabilité des $Y_i = h_i(X_i)$ si les fonctions boréliennes h_i sont positives. Le corollaire 5.46 est vrai sans condition d'intégrabilité pour des variables aléatoires positives.*

Une application importante du théorème 5.45 est que l'indépendance de deux v.a. implique la nullité de leur covariance (lorsqu'elle existe).

Proposition 5.48 (indépendance et covariance).

- a) Si X et Y sont des v.a. réelles de carré intégrable et indépendantes, leur covariance est nulle.
- b) Si X_1, \dots, X_n sont des v.a. réelles de carré intégrable et deux à deux indépendantes, en notant $S_n := \sum_{k=1}^n X_k$, on a

$$\text{Var } S_n = \sum_{k=1}^n \text{Var } X_k. \quad (5.40)$$

Preuve. Le a) est une conséquence immédiate du corollaire 5.46 et de la formule de Koenig puisque $\text{Cov}(X, Y) = \mathbf{E}(XY) - (\mathbf{E}X)(\mathbf{E}Y) = (\mathbf{E}X)(\mathbf{E}Y) - (\mathbf{E}X)(\mathbf{E}Y)$. On peut le voir aussi à partir de la définition de la covariance puisque $X - \mathbf{E}X$ et $Y - \mathbf{E}Y$ héritent de l'indépendance de X et Y et sont deux v.a. d'espérance nulle. Notons aussi que pour des v.a. indépendantes, on n'a pas besoin de supposer que X et Y soient de carré intégrable pour définir leur covariance, leur intégrabilité suffit.

Le b) est une conséquence immédiate du a) *via* la formule 5.9. Notons qu'il suffit d'avoir ici l'indépendance deux à deux qui est plus faible que l'indépendance mutuelle¹⁵. \square

Remarque 5.49. La nullité de la covariance de deux v.a. réelles X et Y *n'implique pas* leur indépendance. Voici un contre exemple. Prenons X de loi uniforme sur $[-1, +1]$ et $Y := X^2$. On a alors en notant que X a pour densité $\frac{1}{2}\mathbf{1}_{[-1,1]}$,

$$\mathbf{E}X = \frac{1}{2} \int_{-1}^{+1} x \, dx = 0 \quad \text{et} \quad \mathbf{E}(XY) = \mathbf{E}(X^3) = \frac{1}{2} \int_{-1}^{+1} x^3 \, dx = 0,$$

d'où $\text{Cov}(X, Y) = \mathbf{E}(XY) - (\mathbf{E}X)(\mathbf{E}Y) = 0$. Il est clair intuitivement que X et Y ne sont pas indépendantes puisque Y est une fonction déterministe de X . Pour vérifier cette non-indépendance par le calcul, on peut remarquer que d'une part

$$\begin{aligned} P(X \in [0, 1/2] \text{ et } Y \in [0, 1/4]) &= P(X \in [0, 1/2] \text{ et } X \in [-1/2, 1/2]) \\ &= P(X \in [0, 1/2]) = \frac{1}{4} \end{aligned}$$

et d'autre part

$$P(X \in [0, 1/2])P(Y \in [0, 1/4]) = \frac{1}{4}P(X \in [-1/2, 1/2]) = \frac{1}{8}.$$

15. En fait il suffit d'avoir la non-corrélation deux à deux, c'est-à-dire $\text{Cov}(X_i, X_j) = 0$ pour $i \neq j$, ce qui est encore plus faible.

Chapitre 6

Théorèmes limites

Nous commençons dans ce chapitre l'étude du comportement asymptotique de suites de variables aléatoires. Après avoir vu les différents modes de convergence de ces suites, nous abordons la loi des grands nombres. Ce résultat essentiel nous dit que les moyennes arithmétiques d'une suite de v.a. X_i indépendantes et de même loi ayant une espérance, convergent en un certain sens, vers cette espérance :

$$M_n := \frac{S_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow +\infty]{} \mathbf{E}X_1. \quad (6.1)$$

Cette convergence est très utile en statistique pour *estimer* des paramètres d'une loi inconnue, sur la base de l'observation d'un *échantillon* X_1, \dots, X_n de grande taille. Le prolongement naturel de ce chapitre est l'étude du théorème limite central qui donne une sorte de vitesse de convergence pour la loi des grands nombres, permettant notamment de construire des « intervalles de confiance » pour l'estimation d'un paramètre. Ce théorème sera vu au début du cours d'*Initiation à la Statistique*.

6.1 Convergences de suites de v.a.

6.1.1 Convergence presque sûre et en probabilité

Quand on envisage la question de la convergence d'une suite de v.a. (Y_n) vers une v.a. Y , la première notion de convergence qui vient à l'esprit est la convergence simple sur tout Ω , au sens de l'analyse¹ :

$$\forall \omega \in \Omega, \quad Y_n(\omega) \xrightarrow[n \rightarrow +\infty]{} Y(\omega).$$

On voit immédiatement que cette notion n'est pas satisfaisante pour la convergence de la suite (M_n) donnée en (6.1). En effet considérons le modèle probabiliste infini le plus simple possible, à savoir le jeu de pile ou face infini. On peut prendre ici $\Omega = \{\mathbf{f}, \mathbf{p}\}^{\mathbb{N}^*}$ et ω est une suite infinie $\omega = (u_i)_{i \geq 1}$, avec $u_i \in \{\mathbf{f}, \mathbf{p}\}$ pour tout i . En prenant pour X_i

1. Ceci suppose que les Y_n et Y sont définies sur le même Ω .

l'indicatrice de l'évènement obtention de pile au i^{e} lancer, M_n est la fréquence d'apparition de pile au cours des n premiers lancers. Si la pièce est équilibrée, on s'attend à ce que M_n converge vers $1/2$. Or il est clair qu'il y a une infinité d'évènements élémentaires ω pour lesquels $M_n(\omega)$ ne converge pas vers $1/2$. On peut même construire facilement une infinité de ω pour lesquels $M_n(\omega)$ n'a *aucune* limite². Ce simple exemple montre que la notion de convergence simple n'est pas pertinente en théorie des probabilités. Pour dépasser ce problème, on introduit la notion de convergence presque sûre, *i.e.* la convergence simple de Y_n vers Y sur un sous-ensemble Ω' de probabilité 1 de Ω .

Définition 6.1 (convergence presque sûre). *Soient $(Y_n)_{n \geq 1}$ et Y des variables aléatoires définies sur le même espace probabilisé (Ω, \mathcal{F}, P) . On dit que Y_n converge presque sûrement vers Y , notation $Y_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} Y$, si $P(\Omega') = 1$, en notant*

$$\Omega' := \left\{ \omega \in \Omega; \lim_{n \rightarrow +\infty} Y_n(\omega) = Y(\omega) \right\}. \quad (6.2)$$

Cette définition soulève immédiatement une question. Pour que l'égalité $P(\Omega') = 1$ ait un sens, encore faut-il que Ω' appartienne à la tribu \mathcal{F} sur laquelle est définie la fonction d'ensembles P . Pour établir l'appartenance à \mathcal{F} de Ω' , il est naturel de s'appuyer sur la mesurabilité des Y_n et de Y dont héritent les v.a. positives $|Y_n - Y|$. Pour cela, on commence par écrire avec des quantificateurs l'appartenance à Ω' :

$$\omega \in \Omega' \quad \Leftrightarrow \quad \forall \varepsilon > 0, \exists j = j(\omega, \varepsilon) \in \mathbb{N}, \forall k \geq j, |Y_k(\omega) - Y(\omega)| < \varepsilon. \quad (6.3)$$

En utilisant la traduction automatique des quantificateurs en opérations ensemblistes, cf. p. 6, on en déduit que

$$\Omega' = \bigcap_{\varepsilon > 0} \bigcup_{j \in \mathbb{N}} \bigcap_{k \geq j} \{|Y_k - Y| < \varepsilon\}. \quad (6.4)$$

En lisant (6.4) de droite à gauche, on obtient les appartenances successives à \mathcal{F} , d'abord des ensembles $\{|Y_k - Y| < \varepsilon\}$ en raison de la mesurabilité des v.a. $|Y_n - Y|$, puis de l'intersection dénombrable sur k , puis de l'union dénombrable sur j . Arrivés là, on est coincés car la dernière intersection sur ε a pour ensemble d'indexation $]0, +\infty[$ qui est infini *non*-dénombrable. On ne peut donc pas en déduire l'appartenance à \mathcal{F} de Ω' . Pour franchir cet obstacle, il suffit de revenir à (6.3) et d'écrire une version « discrétisée » de la convergence de $Y_n(\omega)$ vers $Y(\omega)$. Pour cela on choisit une suite de réels strictement positifs $(\varepsilon_i)_{i \geq 1}$, tendant vers 0 et on écrit que

$$\omega \in \Omega' \quad \Leftrightarrow \quad \forall i \geq 1, \exists j = j(\omega, i) \in \mathbb{N}, \forall k \geq j, |Y_k(\omega) - Y(\omega)| < \varepsilon_i. \quad (6.5)$$

La traduction automatique des quantificateurs nous donne maintenant

$$\Omega' = \bigcap_{i \in \mathbb{N}^*} \bigcup_{j \in \mathbb{N}} \bigcap_{k \geq j} \{|Y_k - Y| < \varepsilon_i\}. \quad (6.6)$$

2. Pour approfondir cette question, voir la section « 6.5 Discussion » dans [ICP].

Sous cette forme, il est maintenant clair que Ω' appartient à \mathcal{F} et ceci légitime la définition 6.1. Nous avons établi au passage que

$$Y_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} Y \Leftrightarrow P\left(\bigcap_{i \in \mathbb{N}^*} \bigcup_{j \in \mathbb{N}} \bigcap_{k \geq j} \{|Y_k - Y| < \varepsilon_i\}\right) = 1, \quad (6.7)$$

pour une suite de réels $\varepsilon_i > 0$, tendant vers 0.

Pour l'instant, (6.7) n'est pas directement exploitable comme méthode pratique pour montrer une convergence presque sûre, mais on peut progresser dans cette direction en « faisant sortir le $\forall i$ » de la probabilité. Cette opération est légitimée par le lemme suivant.

Lemme 6.2. *Si $(B_i)_{i \geq 1}$ est une suite d'événements, on a l'équivalence*

$$P\left(\bigcap_{i \geq 1} B_i\right) = 1 \Leftrightarrow \forall i \geq 1, P(B_i) = 1. \quad (6.8)$$

Preuve. L'implication « \Rightarrow » est évidente car pour tout $j \geq 1$, on a $\bigcap_{i \geq 1} B_i \subset B_j$, d'où

$$1 = P\left(\bigcap_{i \geq 1} B_i\right) \leq P(B_j) \leq 1.$$

Pour la réciproque, on montre que si tous les $P(B_i)$ valent 1, alors $P((\bigcap_{i \geq 1} B_i)^c) = 0$. Ceci s'obtient par sous- σ -additivité de P :

$$P\left(\left(\bigcap_{i \geq 1} B_i\right)^c\right) = P\left(\bigcup_{i \geq 1} B_i^c\right) \leq \sum_{i=1}^{+\infty} P(B_i^c) = 0,$$

puisque tous les termes de cette série sont nuls. \square

En appliquant le lemme 6.2 aux événements $B_i := \bigcup_{j \in \mathbb{N}} \bigcap_{k \geq j} \{|Y_k - Y| < \varepsilon_i\}$, nous obtenons à partir de (6.7) l'équivalence entre la convergence presque sûre de Y_n vers Y et la condition

$$\forall i \in \mathbb{N}^*, P\left(\bigcup_{j \in \mathbb{N}} \bigcap_{k \geq j} \{|Y_k - Y| < \varepsilon_i\}\right) = 1, \quad (6.9)$$

Maintenant que nous avons réussi à faire sortir le « $\forall \varepsilon$ » de la probabilité, on peut laisser tomber la discrétisation pour aboutir à la caractérisation suivante de la convergence presque sûre.

Proposition 6.3 (une c.n.s. de convergence p.s.). *La suite de variables aléatoires $(Y_n)_{n \geq 1}$ converge presque sûrement vers la variable aléatoire Y si et seulement si*

$$\forall \varepsilon > 0, P\left(\bigcup_{j \in \mathbb{N}} \bigcap_{k \geq j} \{|Y_k - Y| < \varepsilon\}\right) = 1, \quad (6.10)$$

ou encore

$$\forall \varepsilon > 0, P\left(\bigcap_{j \in \mathbb{N}} \bigcup_{k \geq j} \{|Y_k - Y| \geq \varepsilon\}\right) = 0. \quad (6.11)$$

Preuve. L'équivalence entre (6.10) et (6.11) est évidente par passage au complémentaire. Vérifions l'équivalence entre (6.10) et la c.n.s. (6.9) de convergence presque sûre. Il est clair que (6.10) implique (6.9). Pour la réciproque, il suffit de remarquer que la fonction $t \mapsto P(\bigcup_{j \in \mathbb{N}} \bigcap_{k \geq j} \{|Y_k - Y| < t\})$ est *croissante*. En effet si $s < t$, l'inclusion d'évènements $\{|Y_k - Y| < s\} \subset \{|Y_k - Y| < t\}$ se propage à l'intersection sur $k \geq j$, puis à l'union sur j . Supposons (6.9) vraie et fixons $\varepsilon > 0$ quelconque. Comme la suite $(\varepsilon_i)_{i \geq 1}$ tend vers 0, on peut trouver un i_0 tel que $\varepsilon_{i_0} < \varepsilon$ et alors

$$1 = P\left(\bigcup_{j \in \mathbb{N}} \bigcap_{k \geq j} \{|Y_k - Y| < \varepsilon_{i_0}\}\right) \leq P\left(\bigcup_{j \in \mathbb{N}} \bigcap_{k \geq j} \{|Y_k - Y| < \varepsilon\}\right) \leq 1,$$

d'où (6.10). \square

Intéressons nous maintenant à (6.11). En notant $A_k := \{|Y_k - Y| \geq \varepsilon\}$, on aimerait trouver une condition qui nous assure que $P(\bigcap_{j \in \mathbb{N}} \bigcup_{k \geq j} A_k) = 0$. En pratique, on a souvent une majoration des probabilités $P(A_k)$ *via* une inégalité du type inégalité de Markov et il est donc naturel de chercher une condition portant sur les $P(A_k)$. Ce problème a un intérêt propre, indépendant de la définition des A_k , provenant de l'interprétation suivante de l'évènement $\bigcap_{j \in \mathbb{N}} \bigcup_{k \geq j} A_k$:

$$\begin{aligned} \bigcap_{j \in \mathbb{N}} \bigcup_{k \geq j} A_k &= \{\omega \in \Omega; \omega \text{ appartient à une infinité de } A_k\} \\ &= \{\omega \in \Omega; \omega \text{ réalise une infinité de } A_k\} \\ &= \{\text{réalisation d'une infinité de } A_k\}. \end{aligned}$$

L'étude de la probabilité de cet évènement conduit aux deux lemmes de Borel Cantelli.

Lemme 6.4 (Borel Cantelli I). *Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'évènements telle que*

$$\sum_{n=0}^{+\infty} P(A_n) < +\infty. \quad (6.12)$$

Alors

$$P(\text{réalisation d'une infinité de } A_n) = 0.$$

Preuve. Introduisons la notation³ $A_\infty := \{\text{réalisation d'une infinité de } A_n\}$ et posons $C_j := \bigcup_{k \geq j} A_k$. La suite $(C_j)_{j \in \mathbb{N}}$ est décroissante pour l'inclusion et d'intersection A_∞ . Par continuité séquentielle décroissante de P , on a donc

$$P(C_j) \xrightarrow{j \rightarrow +\infty} P(A_\infty). \quad (6.13)$$

Par sous- σ -additivité de P , on a d'autre part

$$\forall j \in \mathbb{N}, \quad 0 \leq P(C_j) \leq \sum_{k \geq j} P(A_k) =: r(j). \quad (6.14)$$

3. Cette écriture est commode, mais non standard, donc pensez à en expliciter la définition si vous êtes tenté de la réutiliser dans un exercice. Vous trouverez parfois dans la littérature la notation $\limsup_{n \rightarrow +\infty} A_n$ pour l'évènement « réalisation d'une infinité de A_n ». Cette notation est proscrite de ce cours par choix pédagogique.

Grâce à l'hypothèse (6.12), $r(j)$ est le reste d'une série *convergente*, donc tend vers 0 quand j tend vers $+\infty$. Cette convergence combinée avec la majoration (6.14) nous donne

$$P(C_j) \xrightarrow{j \rightarrow +\infty} 0. \quad (6.15)$$

La conclusion $P(A_\infty) = 0$ découle alors de (6.13) et (6.15). \square

Le lemme de Borel Cantelli I est d'une portée très générale, puisqu'on obtient la conclusion $P(A_\infty) = 0$ sous la seule hypothèse de convergence de la série $\sum_{n \in \mathbb{N}} P(A_n)$, *sans rien supposer sur la structure de dépendance* de la suite (A_n) . Pour les suites d'événements indépendants, on a le résultat complémentaire suivant.

Lemme 6.5 (Borel Cantelli II). *Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements indépendants telle que*

$$\sum_{n=0}^{+\infty} P(A_n) = +\infty. \quad (6.16)$$

Alors

$$P(\text{réalisation d'une infinité de } A_n) = 1.$$

Preuve. Notons encore $A_\infty := \{\text{réalisation d'une infinité de } A_n\}$ et posons

$$C_{j,l} := \bigcup_{j \leq k \leq l} A_k, \quad C_j := \bigcup_{k \geq j} A_k,$$

d'où

$$A_\infty = \bigcap_{j \in \mathbb{N}} C_j.$$

Les A_k^c héritant de l'indépendance des A_k , cf. remarque 2.55, on a

$$P(C_{j,l}) = 1 - P(C_{j,l}^c) = 1 - P\left(\bigcap_{k=j}^l A_k^c\right) = 1 - \prod_{k=j}^l (1 - P(A_k)).$$

On utilise alors l'inégalité de convexité⁴ $e^{-x} \geq 1 - x$ avec $x = P(A_k)$ pour obtenir la minoration

$$1 \geq P(C_{j,l}) \geq 1 - \prod_{k=j}^l \exp(-P(A_k)) = 1 - \exp\left(-\sum_{k=j}^l P(A_k)\right). \quad (6.17)$$

En laissant j fixe et faisant tendre l vers l'infini dans (6.17), on en déduit grâce à l'hypothèse (6.16) que $P(C_{j,l})$ tend vers 1. D'autre part C_j est limite croissante pour l'inclusion des $C_{j,l}$, donc par continuité croissante séquentielle de P ,

$$P(C_j) = \lim_{l \rightarrow +\infty} P(C_{j,l}) = 1.$$

4. La représentation graphique de la fonction convexe $x \mapsto e^{-x}$ est toujours au-dessus de sa tangente à l'origine d'où $e^{-x} \geq 1 - x$ pour tout $x \in \mathbb{R}$.

Cette égalité étant vraie pour tout $j \in \mathbb{N}$, on en déduit par le lemme 6.2 que

$$P\left(\bigcap_{j \in \mathbb{N}} C_j\right) = 1.$$

Comme l'intersection de tous les C_j est l'évènement A_∞ , le lemme est démontré. \square

Après cette longue, mais utile, digression sur les lemmes de Borel Cantelli, revenons à notre quête d'une condition pratique de convergence presque sûre. Le premier lemme de Borel Cantelli nous permet d'aboutir au résultat suivant.

Proposition 6.6 (condition suffisante de convergence p.s.). *Soient Y et $(Y_n)_{n \geq 1}$ des variables aléatoires réelles définies sur le même espace probabilisé (Ω, \mathcal{F}, P) et vérifiant*

$$\forall \varepsilon > 0, \quad \sum_{n=1}^{+\infty} P(|Y_n - Y| \geq \varepsilon) < +\infty. \quad (6.18)$$

Alors Y_n converge presque sûrement vers Y .

On dit d'une suite $(Y_n)_{n \geq 1}$ vérifiant (6.18) qu'elle converge *presque complètement* vers Y .

Preuve. Fixons $\varepsilon > 0$ quelconque et posons $A_n := \{|Y_n - Y| \geq \varepsilon\}$. Par le premier lemme de Borel-Cantelli, on déduit de (6.18) que la probabilité de réalisation d'une infinité de A_n est nulle, ce qui s'écrit encore

$$P\left(\bigcap_{j \in \mathbb{N}} \bigcup_{k \geq j} \{|Y_k - Y| \geq \varepsilon\}\right) = 0.$$

Ceci étant vérifié pour tout $\varepsilon > 0$, la c.n.s. (6.11) de la proposition 6.3 nous donne la convergence presque sûre de Y_n vers Y . \square

Nous introduisons maintenant un nouveau mode de convergence de Y_n vers Y , la convergence en probabilité. Cette notion nous sera utile pour la loi faible des grands nombres.

Définition 6.7. *La suite Y_n converge en probabilité vers Y (notation $Y_n \xrightarrow[n \rightarrow +\infty]{\text{Pr}} Y$) si*

$$\forall \varepsilon > 0, \quad P(|Y_n - Y| \geq \varepsilon) \xrightarrow[n \rightarrow +\infty]{} 0.$$

Notons la différence de point de vue par rapport à la convergence presque sûre. Dans la convergence presque-sûre, on reste proche de la notion de convergence simple de l'analyse. Il s'agit de la convergence de la suite de réels $(Y_n(\omega))_{n \geq 1}$, pour tous les ω d'un même évènement de probabilité 1. La convergence en probabilité ne concerne pas le comportement asymptotique individuel de chaque suite $(Y_n(\omega))_{n \geq 1}$, mais plutôt celui de la suite d'évènements $D_{n,\varepsilon} := \{|Y_n - Y| < \varepsilon\}$ dont la probabilité $P(D_{n,\varepsilon})$ doit tendre vers 1, pour tout ε . En pratique, établir la convergence en probabilité de Y_n vers Y est souvent un travail préliminaire pour prouver la convergence presque-sûre de Y_n .

vers Y . En effet si on utilise la condition suffisante de convergence p.s. (6.18), pour que la série converge, il faut déjà que son terme général tende vers 0 et cette convergence vers 0 (pour tout ε) est précisément la convergence en probabilité de Y_n vers Y . Bien sûr, comme (6.18) n'est qu'une condition suffisante, cette remarque ne nous permet pas d'affirmer que la convergence en probabilité est une notion plus faible que la convergence presque-sûre. Ce qui suit va nous montrer qu'il en est pourtant bien ainsi.

Proposition 6.8. *La convergence presque-sûre implique la convergence en probabilité.*

Preuve. Fixons $\varepsilon > 0$. L'hypothèse de convergence presque sûre de Y_n vers Y signifie que l'événement

$$\Omega' := \{\omega \in \Omega; \lim_{n \rightarrow +\infty} Y_n(\omega) = Y(\omega)\}$$

a pour probabilité 1. Définissons

$$\Omega'_\varepsilon := \{\omega \in \Omega; \exists k_0 = k_0(\omega), \forall n \geq k_0, |Y_n(\omega) - Y(\omega)| < \varepsilon\}.$$

C'est bien un événement (*i.e.* $\Omega'_\varepsilon \in \mathcal{F}$) puisqu'il s'écrit

$$\Omega'_\varepsilon = \bigcup_{k \in \mathbb{N}^*} \bigcap_{n \geq k} \{|Y_n - Y| < \varepsilon\}.$$

De plus Ω'_ε contient Ω' , donc $P(\Omega'_\varepsilon) = 1$. Pour tout $k \geq 1$, notons

$$A_k := \{\omega \in \Omega; \forall n \geq k, |Y_n(\omega) - Y(\omega)| < \varepsilon\} = \bigcap_{n \geq k} \{|Y_n - Y| < \varepsilon\}.$$

La suite $(A_k)_{k \geq 1}$ est clairement croissante pour l'inclusion et sa réunion est Ω'_ε . Par continuité séquentielle croissante de P , on a donc $P(A_k) \uparrow P(\Omega'_\varepsilon) = 1$ ($k \rightarrow +\infty$). Par conséquent,

$$\forall \delta > 0, \exists k_1, \quad P(A_{k_1}) > 1 - \delta.$$

Pour tout $n \geq k_1$, l'événement $\{|Y_n - Y| < \varepsilon\}$ contient A_{k_1} , d'où

$$\forall n \geq k_1, \quad P(|Y_n - Y| < \varepsilon) > 1 - \delta.$$

En passant à l'événement complémentaire, on obtient finalement

$$\forall \delta > 0, \exists k_1 \in \mathbb{N}, \forall n \geq k_1, \quad P(|Y_n - Y| \geq \varepsilon) < \delta.$$

Ceci établit la convergence vers 0 de $P(|Y_n - Y| \geq \varepsilon)$. Comme ε était quelconque, on a bien convergence en probabilité de Y_n vers Y . \square

Remarque 6.9. La convergence en probabilité n'implique pas la convergence presque-sûre. Voici un contre exemple. On prend comme espace probabilisé $(]0, 1], \text{Bor}(]0, 1]), \lambda)$, où λ est la restriction à $]0, 1]$ de la mesure de Lebesgue sur \mathbb{R} . On définit sur cet espace les Y_n comme suit :

$$\begin{aligned} Y_1 &= \mathbf{1}_{]0,1]}, \\ Y_2 &= \mathbf{1}_{]0,1/2]}, Y_3 = \mathbf{1}_{]1/2,1]}, \\ Y_4 &= \mathbf{1}_{]0,1/4]}, Y_5 = \mathbf{1}_{]1/4,1/2]}, Y_6 = \mathbf{1}_{]1/2,3/4]}, Y_7 = \mathbf{1}_{]3/4,1]}, \\ Y_8 &= \mathbf{1}_{]0,1/8]}, Y_9 = \mathbf{1}_{]1/8,1/4]}, \dots, Y_{15} = \mathbf{1}_{]7/8,1]}, \\ Y_{16} &= \mathbf{1}_{]0,1/16]}, \dots \end{aligned}$$

Le lecteur qui ne se satisferait pas de cette définition informelle peut toujours s'exercer à trouver une formule explicite pour Y_n . Sans entrer dans ces détails techniques, on peut facilement se convaincre de deux choses :

1. Pour tout $\omega \in]0, 1]$, la suite de « bits » $(Y_n(\omega))_{n \geq 1}$ est formée d'une infinité de 0 et d'une infinité de 1. Elle ne peut donc converger (sa limite inférieure vaut 0 et sa limite supérieure 1). Ainsi non seulement on n'a pas de convergence presque sûre de Y_n , mais en plus $Y_n(\omega)$ ne converge pour *aucun* $\omega \in \Omega$.
2. Pour $0 < \varepsilon < 1$, $P(|Y_n - 0| > \varepsilon) = P(Y_n = 1) = \lambda(I_n)$, en notant I_n l'intervalle dyadique dont Y_n est l'indicatrice. La longueur $\lambda(I_n)$ de cet intervalle tend vers zéro quand n tend vers l'infini (à la même vitesse que l'inverse du logarithme en base deux de n). Donc Y_n converge vers 0 en probabilité.

Remarque 6.10. Il est toutefois possible d'obtenir la convergence presque sûre à partir de la convergence en probabilité, à condition d'avoir une *bonne vitesse* de convergence en probabilité. Le sens précis de cette affirmation est donné par la proposition 6.6.

La proposition 6.6 permet aussi, même sans bonne vitesse de convergence en probabilité, d'obtenir de la convergence p.s. pour une *sous-suite*.

Proposition 6.11 (convergence p.s. d'une sous-suite). *Si la suite $(Y_n)_{n \geq 1}$ converge en probabilité vers Y , on peut en extraire une sous-suite $(Y_{n_i})_{i \geq 1}$ qui converge presque sûrement vers Y .*

Preuve. Notons $\varepsilon_i = 2^{-i}$. La convergence en probabilité de Y_n vers Y implique pour tout $i \geq 1$, la convergence de $P(|Y_n - Y| > \varepsilon_i)$ vers 0 quand n tend vers l'infini. On en déduit l'existence d'une suite strictement croissante d'indices n_i telle que

$$\forall i \geq 1, \quad P(|Y_{n_i} - Y| \geq \varepsilon_i) \leq \frac{1}{i^2}.$$

Vérifions maintenant que

$$\forall \varepsilon > 0, \quad \sum_{i=1}^{+\infty} P(|Y_{n_i} - Y| \geq \varepsilon) < +\infty.$$

En effet la convergence vers 0 de ε_i nous assure de l'existence d'un $i_0 = i_0(\varepsilon)$ tel que pour tout $i \geq i_0$, $\varepsilon_i < \varepsilon$. Pour $i \geq i_0$, on a donc $\{|Y_{n_i} - Y| \geq \varepsilon\} \subset \{|Y_{n_i} - Y| \geq \varepsilon_i\}$ et cette inclusion d'événements nous permet de majorer le terme général de la série ci-dessus par i^{-2} à partir du rang i_0 . Ainsi la suite (Y_{n_i}) converge presque complètement vers Y donc aussi presque sûrement. \square

6.1.2 Convergence en moyenne d'ordre p

Nous introduisons maintenant un nouveau mode de convergence, utile notamment dans les problèmes d'interversion limite espérance.

Définition 6.12. Soit $p \geq 1$ un réel et $(Y_n)_{n \geq 1}$ une suite de v.a. ayant un moment absolu d'ordre p fini. On dit que cette suite converge en moyenne d'ordre p (ou au sens L^p) vers la v.a. Y si

$$\lim_{n \rightarrow +\infty} \mathbf{E}(|Y_n - Y|^p) = 0. \quad (6.19)$$

Notation : $Y_n \xrightarrow[n \rightarrow +\infty]{L^p} Y$.

Remarque 6.13. Si Y_n converge vers Y en moyenne d'ordre p , Y a nécessairement un moment absolu d'ordre p fini. Pour le voir, on utilise la *convexité* de la fonction $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, $x \mapsto x^p$, pour $p \geq 1$. Ceci signifie que la « corde » entre deux points de son graphe est « au dessus » de l'arc de courbe correspondant, ou encore que « l'image du barycentre » est majorée par « le barycentre des images », voir la figure 6.1. Cette convexité implique notamment que pour tous $a, b \geq 0$, $\varphi\left(\frac{a+b}{2}\right) \leq \frac{1}{2}\varphi(a) + \frac{1}{2}\varphi(b)$, ce qui s'écrit encore

$$\forall a, b \geq 0, \quad (a+b)^p \leq 2^{p-1}(a^p + b^p). \quad (6.20)$$

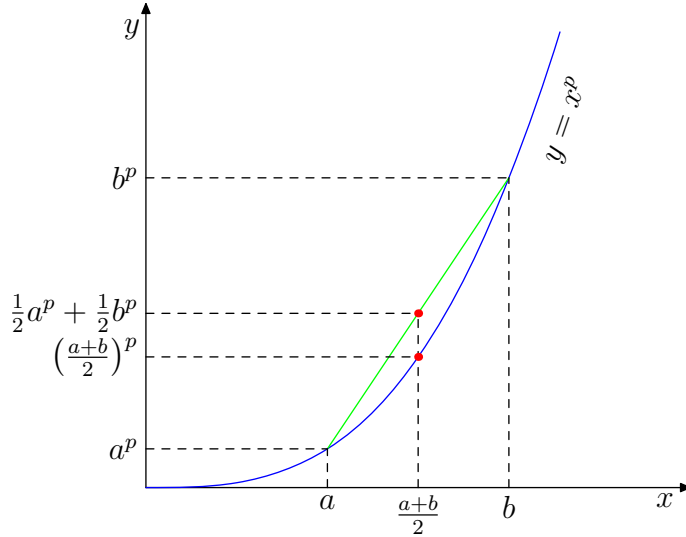


FIG. 6.1 – Convexité de $x \mapsto x^p$ et inégalité $\left(\frac{a+b}{2}\right)^p \leq \frac{1}{2}(a^p + b^p)$

Par croissance de φ , inégalité triangulaire et (6.20) appliquée avec $a = |Y_n(\omega)|$ et $b = |Y(\omega) - Y_n(\omega)|$, on voit que pour tout $\omega \in \Omega$,

$$|Y(\omega)|^p \leq (|Y_n(\omega)| + |Y(\omega) - Y_n(\omega)|)^p \leq 2^{p-1}|Y_n(\omega)|^p + 2^{p-1}|Y(\omega) - Y_n(\omega)|^p.$$

Par croissance de l'espérance, on en déduit :

$$\mathbf{E}|Y|^p \leq 2^{p-1}\mathbf{E}|Y_n|^p + 2^{p-1}\mathbf{E}|Y - Y_n|^p.$$

Ce majorant est fini car $\mathbf{E}|Y_n|^p$ est fini par hypothèse et $\mathbf{E}|Y - Y_n|^p$ tend vers 0 donc est fini au moins pour n assez grand.

Proposition 6.14. *La convergence en moyenne d'ordre p implique la convergence en probabilité.*

Preuve. C'est une conséquence immédiate de l'inégalité de Markov avec moment, cf. proposition 4.38, puisque

$$\forall \varepsilon > 0, \quad P(|Y_n - Y| \geq \varepsilon) \leq \frac{\mathbf{E}(|Y_n - Y|^p)}{\varepsilon^p}.$$

□

Remarque 6.15. Par contre, il n'y a aucune implication entre convergence p.s. et convergence en moyenne d'ordre p . Pour voir que la convergence L^p n'implique pas la convergence p.s., le contre exemple déjà vu à la remarque 6.9 fait l'affaire car on vérifie (exercice) que la suite Y_n de cet exemple converge au sens L^p vers 0 (pour tout $p \geq 1$). Voici maintenant un contre exemple montrant que la convergence p.s. n'implique pas la convergence L^p . On prend U de loi uniforme sur $[0, 1]$ et on pose $Y_n := n^2 \mathbf{1}_{[0, 1/n]}(U)$. On voit facilement que

$$P\left(\left\{\omega \in \Omega; \lim_{n \rightarrow +\infty} Y_n(\omega) = 0\right\}\right) = P(U \in]0, 1]) = 1,$$

d'où la convergence presque-sûre de Y_n vers la v.a. constante 0. D'autre part, Y_n est une v.a. discrète ne prenant que les valeurs 0 ou n^2 et on a immédiatement $\mathbf{E}|Y_n|^p = n^{2p-1}$. Le fait que ce moment absolu d'ordre p tende vers $+\infty$ avec n interdit la convergence L^p de la suite (Y_n) . Pour s'en convaincre, supposons que Y_n converge vers Y au sens L^p . Alors $\mathbf{E}|Y|^p$ doit être fini et par l'inégalité de convexité (6.20), on obtient

$$\mathbf{E}|Y_n|^p \leq 2^{p-1} \mathbf{E}|Y|^p + 2^{p-1} \mathbf{E}|Y_n - Y|^p,$$

ce qui est impossible puisque le premier membre tend vers l'infini avec n tandis que le second membre reste borné.

Il y a une hiérarchie entre les convergences L^p pour diverses valeurs de p et c'est la même que celle établie à la proposition 4.37 pour l'existence des moments absolus d'ordre p .

Proposition 6.16 (hiérarchie des convergences L^p). *Si $1 \leq p < r < +\infty$ et si $Y_n \xrightarrow[n \rightarrow +\infty]{L^r} Y$, alors $Y_n \xrightarrow[n \rightarrow +\infty]{L^p} Y$.*

Preuve. Posons $Z_n := |Y_n - Y|$. Il s'agit de montrer que la convergence vers 0 de $\mathbf{E}(Z_n^r)$ implique celle de $\mathbf{E}(Z_n^p)$ vers 0. Fixons ε arbitraire dans $]0, 1[$. On commence par le découpage

$$\mathbf{E}(Z_n^p) = \int_0^{+\infty} P(Z_n^p > t) dt = \int_0^1 P(Z_n^p > t) dt + \int_1^{+\infty} P(Z_n^p > t) dt. \quad (6.21)$$

Pour tout réel $x \geq 1$, on a $x^p \leq x^r$ puisque $p < r$. On en déduit que l'inclusion d'évènements $\{Z_n^p > t\} \subset \{Z_n^r > t\}$ est vraie pour tout $t \geq 1$, d'où

$$\int_1^{+\infty} P(Z_n^p > t) dt \leq \int_1^{+\infty} P(Z_n^r > t) dt \leq \mathbf{E}(Z_n^r). \quad (6.22)$$

Comme ce dernier majorant tend vers 0 quand n tend vers $+\infty$, on peut trouver un entier $n_1 = n_1(\varepsilon)$ tel que

$$\forall n \geq n_1, \quad \int_1^{+\infty} P(Z_n^p > t) dt < \varepsilon. \quad (6.23)$$

Rappelant que $\varepsilon < 1$, on a d'autre part

$$\begin{aligned} \int_0^1 P(Z_n^p > t) dt &= \int_0^\varepsilon P(Z_n^p > t) dt + \int_\varepsilon^1 P(Z_n^p > t) dt \\ &\leq \int_0^\varepsilon dt + \int_\varepsilon^1 P(Z_n^p > t) dt \\ &= \varepsilon + \int_\varepsilon^1 P(Z_n > t^{1/p}) dt \\ &\leq \varepsilon + (1 - \varepsilon)P(Z_n > \varepsilon^{1/p}), \end{aligned}$$

d'où finalement

$$\forall n \in \mathbb{N}^*, \quad \int_0^1 P(Z_n^p > t) dt \leq \varepsilon + P(Z_n > \varepsilon^{1/p}). \quad (6.24)$$

Par la proposition 6.14, la convergence de Z_n vers 0 au sens L^r implique sa convergence en probabilité⁵. Par conséquent $P(Z_n > \varepsilon^{1/p})$ converge vers 0 quand n tend vers $+\infty$, ce qui nous assure de l'existence d'un entier $n_2 = n_2(\varepsilon)$ tel que

$$\forall n \geq n_2, \quad \int_0^1 P(Z_n^p > t) dt \leq 2\varepsilon. \quad (6.25)$$

On déduit de (6.21), (6.23) et (6.25) que

$$\forall n \geq n_0 := \max(n_1, n_2), \quad \mathbf{E}(Z_n^p) < 3\varepsilon.$$

Comme ε était arbitraire dans $]0, 1[$, $\mathbf{E}(Z_n^p)$ tend bien vers 0 quand n tend vers $+\infty$. \square

Remarque 6.17. La démonstration classique de la proposition 6.16 est plus rapide que celle présentée ci-dessus puisqu'elle est une conséquence immédiate de l'inégalité

$$(\mathbf{E}|X|^p)^{1/p} \leq (\mathbf{E}|X|^r)^{1/r}, \quad \forall 1 \leq p < r < +\infty.$$

On pourra éventuellement voir cette inégalité en exercice. La preuve proposée ci-dessus, certes moins élégante, a néanmoins l'avantage de donner comme sous-produit le résultat suivant qui a son intérêt propre.

5. Remarquer que $\mathbf{E}(Z_n^r) = \mathbf{E}|Z_n - 0|^r$ et donc que la convergence L^r de Z_n vers 0 équivaut à la convergence vers 0 de $\mathbf{E}(Z_n^r)$. Cette simplification est particulière à la limite 0, on n'a bien sûr pas équivalence en général entre convergence de Z_n vers Z au sens L^r et convergence de $\mathbf{E}|Z_n|^r$ vers $\mathbf{E}|Z|^r$.

Proposition 6.18. *Si la suite de variables aléatoires Y_n est bornée par une constante positive c (ou plus largement si pour tout n , $P(|Y_n| \leq c) = 1$) et converge en probabilité vers Y , alors Y_n converge au sens L^p vers Y , pour tout $p \geq 1$. En particulier pour $p = 1$, on en déduit l'interversion limite espérance : $\lim_{n \rightarrow +\infty} \mathbf{E}Y_n = \mathbf{E}Y$.*

L'argument de la preuve est essentiellement le même que ci-dessus, *modulo* quelques adaptations mineures que l'on vous laisse le soin de rédiger en exercice.

Nous allons maintenant établir le célèbre théorème de *convergence dominée* qui est très utile pour l'interversion limite espérance.

Théorème 6.19 (convergence dominée). *On suppose que les variables aléatoires réelles Y_n ($n \geq 1$), Y et Z définies sur le même espace probabilisé (Ω, \mathcal{F}, P) vérifient*

- a) Y_n converge presque-sûrement vers Y quand n tend vers $+\infty$;
- b) pour tout $n \geq 1$, $|Y_n| \leq Z$ p.s. ;
- c) Z est intégrable.

Dans ces conditions,

- 1. les Y_n et Y sont intégrables ;
- 2. Y_n converge vers Y au sens L^1 , i.e. $\mathbf{E}|Y_n - Y| \rightarrow 0$;
- 3. on a l'interversion limite espérance : $\lim_{n \rightarrow +\infty} \mathbf{E}Y_n = \mathbf{E}Y$.

Preuve. En utilisant a), b) et le lemme 6.2, on vérifie facilement l'existence d'un événement $\Omega' \in \mathcal{F}$ de probabilité 1 tel que

$$\forall \omega \in \Omega', \forall n \geq 1, \quad |Y_n(\omega)| \leq Z(\omega) \text{ et } |Y(\omega)| \leq Z(\omega). \quad (6.26)$$

Remarquons que pour tout événement A , $P(A \cap \Omega'^c) \leq P(\Omega'^c)$, d'où :

$$\text{quand } P(\Omega') = 1, \quad \forall A \in \mathcal{F}, \quad P(A \cap \Omega'^c) = 0 \quad \text{et} \quad P(A \cap \Omega') = P(A). \quad (6.27)$$

On en déduit de (6.26) et (6.27) que

$$\forall t \geq 0, \quad P(|Y_n| > t) = P(\{|Y_n| > t\} \cap \Omega') \leq P(\{Z > t\} \cap \Omega') = P(Z > t)$$

et de même avec Y à la place de Y_n . En intégrant sur \mathbb{R}_+ relativement à t , on obtient grâce à c) les inégalités

$$\mathbf{E}|Y_n| \leq \mathbf{E}Z < +\infty, \quad \mathbf{E}|Y| \leq \mathbf{E}Z < +\infty,$$

qui établissent l'intégrabilité des Y_n et de Y .

Une fois établie cette intégrabilité et donc l'existence de $\mathbf{E}Y_n$ et $\mathbf{E}Y$, on remarque que l'interversion limite espérance est une conséquence immédiate de la convergence L^1 en raison de l'inégalité :

$$|\mathbf{E}Y_n - \mathbf{E}Y| = |\mathbf{E}(Y_n - Y)| \leq \mathbf{E}|Y_n - Y|.$$

Il nous reste alors à prouver la convergence L^1 , autrement dit, en posant $Z_n := |Y_n - Y|$, la convergence vers 0 de $\mathbf{E}Z_n$.

Pour cela on utilise comme dans la preuve de la proposition 6.16 un découpage en 3 de l'intégrale définissant $\mathbf{E}Z_n$ en écrivant pour $\varepsilon > 0$ arbitraire et $b = b(\varepsilon)$ convenablement choisi, $\int_0^{+\infty} = \int_0^\varepsilon + \int_\varepsilon^b + \int_b^{+\infty}$. Voyons d'abord le choix de b . Par inégalité triangulaire, $Z_n \leq 2Z$ sur Ω' , ce qui grâce à (6.27) nous donne pour tout $t > 0$, $P(Z_n > t) \leq P(2Z > t)$. La variable aléatoire Z étant intégrable par hypothèse c), il en est de même pour $2Z$, ce qui implique la convergence dans \mathbb{R}_+ de l'intégrale de Riemann généralisée $\int_0^{+\infty} P(2Z > t) dt$. On peut donc choisir b assez grand (et supérieur à ε) pour que $\int_b^{+\infty} P(2Z > t) dt < \varepsilon$. On a alors

$$\forall n \in \mathbb{N}^*, \quad \int_b^{+\infty} P(Z_n > t) dt \leq \int_b^{+\infty} P(2Z > t) dt < \varepsilon. \quad (6.28)$$

Ensuite on contrôle l'intégrale $\int_0^b P(Z_n > t) dt$ en écrivant

$$\int_0^b P(Z_n > t) dt \leq \int_0^\varepsilon dt + \int_\varepsilon^b P(Z_n > t) dt \leq \varepsilon + (b - \varepsilon)P(Z_n > \varepsilon) \leq \varepsilon + bP(Z_n > \varepsilon).$$

Par l'hypothèse a), Z_n converge presque-sûrement vers 0, donc converge aussi en probabilité vers 0. On peut donc trouver un entier $n_0 = n_0(\varepsilon)$ tel que $bP(Z_n > \varepsilon) < \varepsilon$ pour tout $n \geq n_0$. On a alors

$$\forall n \geq n_0, \quad \int_0^b P(Z_n > t) dt < 2\varepsilon. \quad (6.29)$$

En recollant les morceaux à partir de (6.28) et (6.29), on obtient

$$\forall n \geq n_0, \quad \mathbf{E}Z_n = \int_0^b P(Z_n > t) dt + \int_b^{+\infty} P(Z_n > t) dt < 3\varepsilon.$$

Comme $\varepsilon > 0$ était arbitraire, ceci établit la convergence vers 0 de $\mathbf{E}Z_n$ et achève la preuve. \square

6.1.3 Bilan sur les convergences de v.a.

Arrivé à ce stade, il est bon de faire le point sur les différents modes de convergence étudiés. Le diagramme de la figure 6.2 résume les relations entre ces modes de convergence. Les flèches en trait plein représentent des implications (si la suite converge selon le mode de la case de départ, alors elle converge aussi selon celui de la case d'arrivée). Les flèches en tirets signifient l'existence d'une sous-suite convergente selon le mode de la case d'arrivée. La convergence en loi sera étudiée ultérieurement.

6. Noter que b dépend de ε et peut très bien tendre vers $+\infty$ quand ε tend vers 0, mais que l'on travaille avec un ε arbitraire fixé, donc aussi avec b fixé.

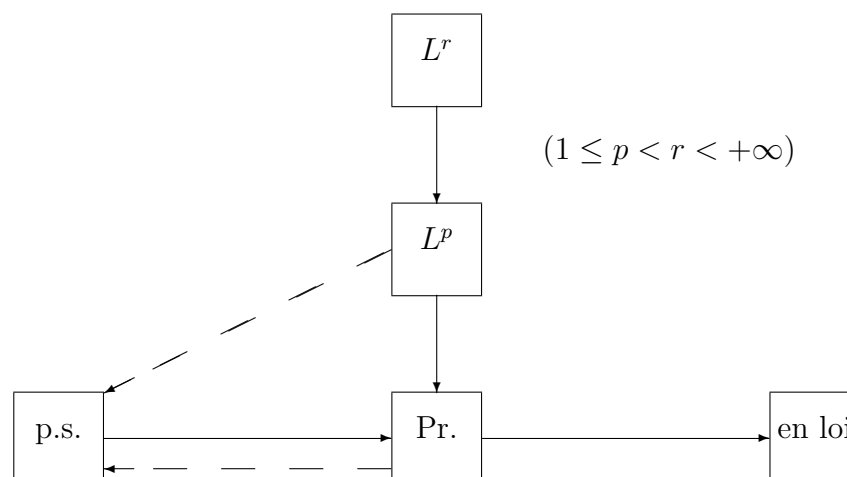


FIG. 6.2 – Diagramme des convergences des suites de v.a.

Signalons deux propriétés communes aux trois modes de convergence étudiés jusqu'ici (vérification laissée au lecteur).

D'abord la limite pour ces modes de convergence n'est pas à strictement parler unique. Elle l'est *modulo* l'égalité presque sûre. Notons (m) , (m') l'un des trois modes de convergence (p.s., en probabilité ou L^p). On peut vérifier que si Y_n converge vers Y au sens (m) et vers Y' au sens (m) alors $Y = Y'$ presque-sûrement (la réciproque est évidente). D'autre part si Y_n converge au sens (m) vers Y et au sens (m') vers Y' , ces deux convergences impliquent la convergence en probabilité de Y_n vers Y et vers Y' , donc l'égalité p.s. de Y et Y' .

Chacun des trois modes de convergence est compatible avec la structure d'espace vectoriel. Si $X_n \xrightarrow{(m)} X$ et $Y_n \xrightarrow{(m)} Y$, $X_n + Y_n \xrightarrow{(m)} X + Y$ et $aX_n \xrightarrow{(m)} aX$ pour tout $a \in \mathbb{R}$.

6.2 Loi des grands nombres

Nous abordons maintenant les lois des grands nombres. Il s'agit d'étudier la convergence des moyennes arithmétiques S_n/n construites à partir d'une suite de variables aléatoires $(X_i)_{i \geq 1}$ en posant $S_n := X_1 + \dots + X_n$. Si S_n/n converge en probabilité, on parle d'une loi faible des grands nombres, tandis que si elle converge presque-sûrement on parle d'une loi forte.

6.2.1 Loi faible des grands nombres

Rappelons que si X est de carré intégrable ($\mathbf{E}X^2 < +\infty$), sa variance est définie par

$$\text{Var } X := \mathbf{E}(X - \mathbf{E}X)^2.$$

Les X_k sont dites deux à deux *non-corrélées* si $\text{Cov}(X_i, X_j) = 0$ pour tous i, j distincts. Ceci se produit en particulier lorsque les X_k sont deux à deux indépendantes, cf. prop. 5.48. Pour des X_k deux à deux non-corrélées, on déduit immédiatement de (5.9) l'égalité

$$\text{Var } S_n = \sum_{k=1}^n \text{Var } X_k. \quad (6.30)$$

Proposition 6.20 (inégalité de Bienaymé-Tchebycheff). *Si les X_k sont de carré intégrable et deux à deux non-corrélées,*

$$\forall t > 0, \quad P\left(\left|\sum_{k=1}^n (X_k - \mathbf{E}X_k)\right| \geq t\right) \leq \frac{1}{t^2} \sum_{k=1}^n \text{Var } X_k. \quad (6.31)$$

Preuve. Il suffit d'écrire :

$$\begin{aligned} P(|S_n - \mathbf{E}S_n| \geq t) &= P(|S_n - \mathbf{E}S_n|^2 \geq t^2) \leq \frac{1}{t^2} \mathbf{E}(S_n - \mathbf{E}S_n)^2 \\ &= \frac{1}{t^2} \text{Var } S_n, \end{aligned} \quad (6.32)$$

où l'inégalité dans (6.32) est l'inégalité de Markov appliquée à la variable aléatoire positive $|S_n - \mathbf{E}S_n|^2$. On conclut avec (6.30). \square

Théorème 6.21 (loi faible des grands nombres). *Si les X_k sont de même loi, de carré intégrable et deux à deux non-corrélées, on a la convergence en probabilité :*

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow[n \rightarrow +\infty]{\text{Pr}} \mathbf{E}X_1. \quad (6.33)$$

Preuve. Comme les X_k ont même loi, on a pour tout k les égalités $\mathbf{E}X_k = \mathbf{E}X_1$ et $\text{Var } X_k = \text{Var } X_1$. Comme elles sont aussi deux à deux non-corrélées, (6.30) nous donne $\text{Var } S_n = n \text{Var } X_1$. Par linéarité de l'espérance on a aussi $\mathbf{E}S_n = n\mathbf{E}X_1$. L'inégalité de Bienaymé-Tchebycheff nous dit alors que :

$$\forall t > 0, \quad P(|S_n - \mathbf{E}S_n| \geq t) = P(|S_n - n\mathbf{E}X_1| \geq t) \leq \frac{n \text{Var } X_1}{t^2}.$$

Posant $t = n\varepsilon$, on en déduit :

$$\forall \varepsilon > 0, \quad \forall n \in \mathbb{N}^*, \quad P\left(\left|\frac{S_n}{n} - \mathbf{E}X_1\right| \geq \varepsilon\right) \leq \frac{n \text{Var } X_1}{n^2 \varepsilon^2} = \frac{\text{Var } X_1}{n \varepsilon^2}.$$

Pour tout $\varepsilon > 0$ fixé, on a ainsi

$$P\left(\left|\frac{S_n}{n} - \mathbf{E}X_1\right| \geq \varepsilon\right) \leq \frac{\text{Var } X_1}{n \varepsilon^2} \xrightarrow[n \rightarrow +\infty]{} 0,$$

ce qui établit la convergence en probabilité de la suite de variables aléatoires S_n/n vers la variable aléatoire constante $\mathbf{E}X_1$. \square

6.2.2 Loi forte des grands nombres

Dans le cadre de ce cours, nous limiterons notre étude des lois fortes des grands nombres au cas où les X_k sont i.i.d. (indépendantes identiquement distribuées), c'est-à-dire indépendantes et de même loi.

Théorème 6.22 (loi forte des grands nombres de Khintchine). *On suppose les X_k indépendantes, de même loi et $\mathbf{E}|X_1| < +\infty$. Alors*

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbf{E}X_1. \quad (6.34)$$

Ce résultat est le meilleur possible en raison du théorème suivant.

Théorème 6.23 (réciproque de la l.f.g.n. de Khintchine). *Soit $(X_k)_{k \geq 1}$ une suite de variables aléatoires indépendantes et de même loi telle que S_n/n converge presque sûrement. Alors $\mathbf{E}|X_1| < +\infty$ et la limite p.s. de S_n/n est la constante $\mathbf{E}X_1$.*

La démonstration de ces deux théorèmes sort du cadre de ce cours. Nous nous contenterons de prouver le théorème 6.22 sous l'hypothèse plus restrictive $\mathbf{E}(X_1^2) < +\infty$.

Preuve de (6.34) pour des X_k i.i.d. de carré intégrable. Puisque les X_k ont même loi, elles sont intégrables comme X_1 et de même espérance. On en déduit que S_n et S_n/n sont intégrables et que par linéarité de l'espérance,

$$\mathbf{E}\left(\frac{S_n}{n}\right) = \frac{1}{n} \mathbf{E}S_n = \frac{1}{n} (n \mathbf{E}X_1) = \mathbf{E}X_1.$$

Posons $X'_k := X_k - \mathbf{E}X_k$ et $S'_n := \sum_{k=1}^n X'_k = S_n - n \mathbf{E}X_1$. Alors $S'_n/n = S_n/n - \mathbf{E}X_1$, donc la convergence p.s. de S_n/n vers $\mathbf{E}X_1$ équivaut à la convergence p.s. de S'_n/n vers 0 (noter aussi que les X'_k sont i.i.d., propriété héritée des X_k). On ne perd donc pas de généralité en supposant désormais pour le confort d'écriture que $\mathbf{E}X_1 = 0$. On a alors $\text{Var } X_1 = \mathbf{E}(X_1^2) =: \sigma^2$ et il s'agit maintenant de prouver que

$$M_n := \frac{1}{n} S_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0. \quad (6.35)$$

Montrons dans un premier temps que la sous-suite $(M_{n^2})_{n \geq 1}$ converge p.s. vers 0. En effet l'inégalité de Bienaymé-Tchebycheff nous donne pour tout $\varepsilon > 0$,

$$P(|M_{n^2}| \geq \varepsilon) = P(|S_{n^2}| \geq n^2 \varepsilon) \leq \frac{\text{Var}(S_{n^2})}{\varepsilon^2 n^4} = \frac{n^2 \text{Var } X_1}{\varepsilon^2 n^4} = \frac{\sigma^2}{\varepsilon^2 n^2}.$$

On en déduit que

$$\forall \varepsilon > 0, \quad \sum_{n=1}^{+\infty} P(|M_{n^2}| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \sum_{n=1}^{+\infty} \frac{1}{n^2} < +\infty,$$

ce qui implique par la proposition 6.6

$$M_{n^2} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0. \quad (6.36)$$

Pour tout $n \in \mathbb{N}^*$, notons $r(n)$ la partie entière de $n^{1/2}$, ce qui nous donne l'encadrement

$$r(n)^2 \leq n < (r(n) + 1)^2. \quad (6.37)$$

La suite d'entiers $(r(n))_{n \geq 1}$ tend vers l'infini comme \sqrt{n} , mais avec des blocs de valeurs répétées de plus en plus longs : $r(n) = 4$ pour $n \in [16, 24[$, $r(n) = 5$ pour $n \in [25, 35[$, etc. Pour cette raison $(M_{r(n)^2})_{n \geq 1}$ n'est pas à proprement parler une sous-suite de $(M_n)_{n \geq 1}$. Nous allons néanmoins voir que l'on peut raccrocher le comportement asymptotique de $(M_n)_{n \geq 1}$ à celui de $(M_{r(n)^2})_{n \geq 1}$ en commençant par écrire :

$$M_n = \frac{1}{n} \sum_{1 \leq j \leq r(n)^2} X_j + \frac{1}{n} \sum_{r(n)^2 < j \leq n} X_j = \frac{r(n)^2}{n} \frac{1}{r(n)^2} \sum_{1 \leq j \leq r(n)^2} X_j + T_n = \frac{r(n)^2}{n} M_{r(n)^2} + T_n,$$

où l'on a posé

$$T_n := \frac{1}{n} \sum_{r(n)^2 < j \leq n} X_j.$$

Par indépendance des X_j , on a

$$\text{Var } T_n = \frac{1}{n^2} (n - r(n)^2) \text{Var } X_1. \quad (6.38)$$

En utilisant (6.37),

$$\frac{1}{n^2} (n - r(n)^2) \leq \frac{(r(n) + 1)^2 - r(n)^2}{n^2} = \frac{2r(n) + 1}{n^2} \leq \frac{2n^{1/2} + 1}{n^2} \leq \frac{3}{n^{3/2}}.$$

En reportant cette majoration dans (6.38) et en utilisant l'inégalité de Markov avec moment d'ordre 2 (noter que $\mathbf{E}T_n = 0$), on obtient

$$\forall \varepsilon > 0, \quad P(|T_n| > \varepsilon) \leq \frac{\text{Var } T_n}{\varepsilon^2} \leq \frac{3\sigma^2}{\varepsilon^2} \frac{1}{n^{3/2}}.$$

Ce majorant étant le terme général d'une série convergente, on en déduit par une nouvelle application de la proposition 6.6 :

$$T_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0. \quad (6.39)$$

Pour conclure, on écrit

$$M_n = T_n + \frac{r(n)^2}{n} M_{r(n)^2}.$$

Comme $r(n)^2/n$ est toujours dans $[0, 1]$, on déduit de (6.36) et (6.39) que M_n converge presque sûrement vers zéro⁷. \square

7. Noter que les suites $(M_{r(n)^2})_{n \geq 1}$ et $(M_{n^2})_{n \geq 1}$ ne sont pas les mêmes. La première a des séquences de termes consécutifs répétés de plus en plus longues et c'est en effaçant ces répétitions que l'on retrouve la deuxième. On voit ainsi que ces deux suites ont même limite. À vous de d'écrire proprement la justification de ce point.

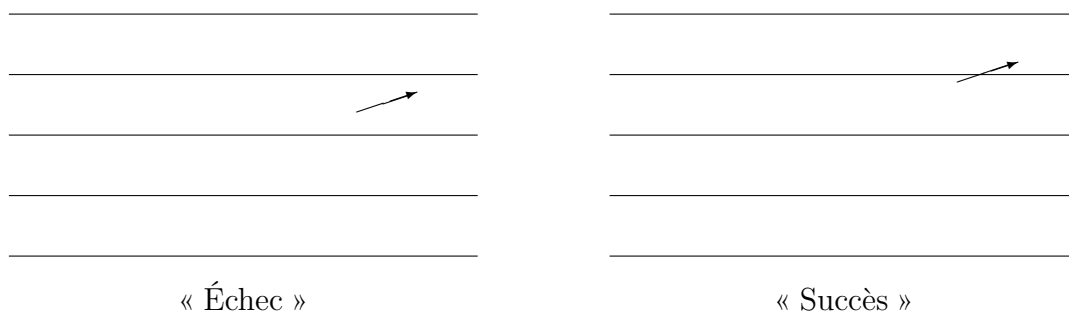
L'application la plus simple et aussi une des plus importantes du théorème 6.22 est la convergence des fréquences de succès dans une suite d'épreuves répétées de Bernoulli indépendantes. Ce résultat explique *a posteriori* l'approche « fréquentiste » dans la définition d'une probabilité. En effet si $(X_k)_{k \geq 1}$ est une suite de variables aléatoires de Bernoulli indépendantes et de même paramètre p , le théorème 6.22 nous donne la convergence presque sûre de $n^{-1}S_n$ vers $\mathbf{E}X_1 = p$. Soit maintenant $A \in \mathcal{F}$ un évènement et $(A_k)_{k \geq 1}$ une suite d'évènements indépendants de même probabilité que A . En prenant $X_k = \mathbf{1}_{A_k}$ et en notant que $\mathbf{E}\mathbf{1}_{A_k} = P(A_k) = P(A)$, on obtient

$$\frac{1}{n} \sum_{k=1}^n \mathbf{1}_{A_k} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} P(A).$$

Par exemple si A est l'évènement « obtention du cinq lors du lancer d'un dé équilibré », ceci nous dit que la *fréquence* d'obtention du cinq en n lancers converge presque sûrement vers $1/6$ lorsque n tend vers l'infini.

6.2.3 L'aiguille de Buffon

À titre d'illustration historique de la loi forte des grands nombres, nous présentons maintenant une méthode expérimentale pour obtenir une approximation numérique du nombre π dont l'intérêt est essentiellement d'ordre culturel⁸. Cette méthode a été proposée en 1777 par le célèbre naturaliste Buffon⁹. On trace sur une surface plane horizontale des droites parallèles équidistantes, séparées par une distance a (on peut par exemple utiliser les rainures d'un parquet). On laisse tomber sur cette surface une aiguille de longueur $\ell \leq a$ et une fois l'aiguille immobilisée, on observe si elle coupe l'une des droites du réseau. On répète l'expérience en notant la fréquence des intersections. Lorsque le nombre d'expériences augmente indéfiniment, cette fréquence converge selon Buffon vers $p = \frac{2\ell}{\pi a}$ permettant ainsi d'obtenir une estimation expérimentale du nombre π .

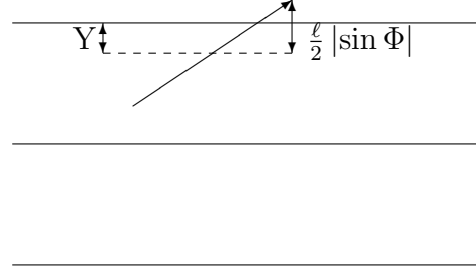


Cherchons une modélisation de cette expérience. On note Y la distance du milieu de l'aiguille à la droite du réseau la plus proche. Y prend ses valeurs dans $[0, \frac{a}{2}]$. On note Φ

8. C'est aussi un excellent problème de révision.

9. Georges Louis Leclerc, comte de Buffon (1707–1788), auteur de l'*Histoire naturelle*, organisateur du Jardin des Plantes de Paris.

une mesure de l'angle entre les droites du réseau (toutes orientées dans le même sens) et l'aiguille orientée du chas vers la pointe. Φ prend ses valeurs dans $[0, 2\pi]$ (par exemple)¹⁰.



Y et Φ sont des variables aléatoires. La connaissance du couple $(Y(\omega), \Phi(\omega))$ suffit pour savoir s'il y a ou non intersection.

Nous ferons les hypothèses suivantes sur les variables aléatoires Y et Φ :

- (H_1) Y suit la loi uniforme sur $[0, \frac{a}{2}]$.
- (H_2) Φ suit la loi uniforme sur $[0, 2\pi]$.
- (H_3) Y et Φ sont indépendantes.

On note E l'événement « l'aiguille coupe l'une des droites du réseau ». La longueur de la projection de la demi-aiguille sur une droite orthogonale au réseau est $Z = \frac{\ell}{2} |\sin \Phi|$. Il y a donc intersection si et seulement si la distance Y du centre de l'aiguille à la droite du réseau la plus proche est inférieure ou égale à Z . Ceci nous permet d'écrire l'évènement E sous la forme :

$$E = \left\{ Y \leq \frac{\ell}{2} |\sin \Phi| \right\}.$$

Comme Y et Φ sont indépendantes, la loi du couple est le produit des lois marginales : $P_{(Y, \Phi)} = P_Y \otimes P_\Phi$. Comme ces lois marginales sont à densités par rapport à λ_1 , on en déduit que $P_{(Y, \Phi)}$ est à densité $f_Y \otimes f_\Phi$ par rapport à λ_2 . D'où

$$f_{(Y, \Phi)}(y, t) = f_Y(y) f_\Phi(t) = \frac{2}{a} \mathbf{1}_{[0, a/2]}(y) \frac{1}{2\pi} \mathbf{1}_{[0, 2\pi]}(t) = \frac{1}{a\pi} \mathbf{1}_{[0, a/2] \times [0, 2\pi]}(y, t).$$

On voit ainsi que le couple (Y, Φ) suit la loi uniforme sur le rectangle $[0, a/2] \times [0, 2\pi]$.

Notons D le borélien de \mathbb{R}^2 défini par

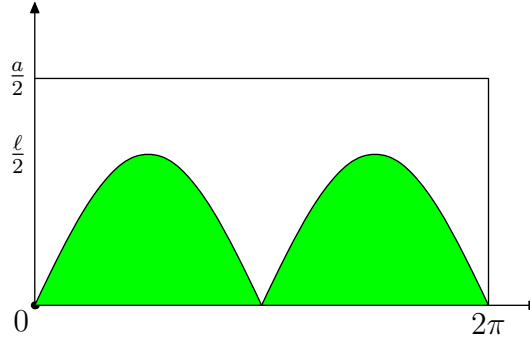
$$D := \left\{ (t, y) \in [0, 2\pi] \times [0, a/2]; y \leq \frac{\ell}{2} |\sin t| \right\}.$$

Comme l'évènement E s'écrit aussi $\{(\Phi, Y) \in D\}$, on peut calculer $P(E)$ en utilisant la loi du couple (Φ, Y) qui est la loi uniforme sur $[0, 2\pi] \times [0, a/2]$:

$$\begin{aligned} P(E) = P((\Phi, Y) \in D) &= P_{(\Phi, Y)}(D) \\ &= \int_D f_{(\Phi, Y)}(y, t) \, dy \, dt \\ &= \frac{1}{a\pi} \lambda_2(D \cap [0, 2\pi] \times [0, a/2]) \\ &= \frac{1}{a\pi} \lambda_2(D), \end{aligned}$$

10. On pourrait aussi utiliser les angles de droites, Φ serait alors à valeurs dans un intervalle de longueur π .

en remarquant que $D \subset [0, 2\pi] \times [0, a/2]$. Le calcul de $P(E)$ se réduit ainsi à celui de l'aire de l'hypographe de la fonction $g : t \mapsto \frac{\ell}{2} |\sin t| \mathbf{1}_{[0, 2\pi]}(t)$.



Par conséquent

$$\lambda_2(D) = \int_0^{2\pi} \frac{\ell}{2} |\sin t| dt = \ell \int_0^\pi \sin t dt = 2\ell.$$

Finalement,

$$P(E) = \frac{2\ell}{a\pi}.$$

On effectue une suite de lancers de l'aiguille et on note E_i l'évènement « lors du i ème lancer, l'aiguille intersecte une des droites du réseau ». On pose $X_i = \mathbf{1}_{E_i}$ et

$$F_n := \frac{1}{n} \sum_{i=1}^n X_i.$$

Les X_i sont des variables aléatoires de Bernoulli, de paramètre $p := P(E_i) = P(E)$. Elles sont clairement intégrables, puisque bornées. Les E_i forment une suite d'évènements mutuellement indépendants et de même probabilité p . Il en résulte que les X_i forment une suite de variables aléatoires indépendantes et de même loi $\text{Bern}(p)$. Par la loi forte des grands nombres pour des variables i.i.d. et intégrables,

$$F_n := \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbf{E}X_1 = P(E).$$

Compte-tenu du calcul de $P(E)$, on peut réécrire ce résultat sous la forme :

$$\frac{2\ell}{aF_n} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \pi.$$

L'interprétation physique est la suivante. Si on réalise *une* série de lancers avec n grand, la valeur $F_n(\omega)$ observée nous fournira l'approximation

$$\frac{2\ell}{aF_n(\omega)} \simeq \pi.$$

On considère ainsi qu'il est physiquement impossible d'observer un ω n'appartenant pas à l'évènement de probabilité 1 $\{F_n \text{ converge vers } p\}$.

Le document de la page 212 représente les résultats de 1200 lancers réalisés avec une allumette et un réseau tracé sur une feuille de format A4. On a ici $\ell = a = 4,5$ cm et $p = \frac{2}{\pi} \simeq 0,637$. Les lancers sont regroupés par dizaine. Les bits 0 ou 1 sont les valeurs observées pour $X_i(\omega)$. Après chaque dizaine on a noté le nombre d'intersections observées sur la dizaine et le nombre d'intersections cumulé depuis le début des lancers. La table 6.1 présente les fréquences observées F_{10k} pour $k = 1, \dots, 120$.

$10k$	0	10	20	30	40	50	60	70	80	90
		0,600	0,650	0,600	0,625	0,600	0,583	0,571	0,575	0,611
100	0,600	0,627	0,633	0,623	0,629	0,607	0,606	0,606	0,617	0,616
200	0,615	0,624	0,627	0,630	0,629	0,628	0,623	0,619	0,618	0,610
300	0,610	0,613	0,609	0,606	0,609	0,603	0,600	0,600	0,605	0,610
400	0,610	0,612	0,614	0,614	0,609	0,609	0,615	0,615	0,619	0,614
500	0,618	0,618	0,619	0,621	0,620	0,622	0,621	0,619	0,621	0,624
600	0,622	0,621	0,626	0,627	0,625	0,628	0,624	0,621	0,622	0,623
700	0,626	0,625	0,626	0,629	0,628	0,625	0,628	0,632	0,635	0,637
800	0,637	0,640	0,635	0,635	0,636	0,634	0,634	0,636	0,637	0,638
900	0,638	0,635	0,635	0,632	0,633	0,633	0,633	0,630	0,628	0,629
1000	0,629	0,628	0,630	0,630	0,626	0,623	0,623	0,621	0,620	0,622
1100	0,622	0,622	0,622	0,623	0,626	0,626	0,628	0,628	0,628	0,627
1200	0,627									

TAB. 6.1 – Tableau des fréquences observées

Cette expérience permet de proposer l'estimation :

$$\pi \simeq \frac{2}{0,627} \simeq 3,1898.$$

Si vous avez la patience et le loisir de réaliser votre propre expérience, vous trouverez probablement une valeur légèrement différente...

Bien entendu, cette méthode pour calculer π n'est pas très performante. On peut montrer que sa vitesse de convergence est en $O(n^{-1/2})$. Son intérêt est essentiellement d'ordre culturel et historique.

Résultats de 1200 lancers

0111110001	6	6	1111111011	9	69	1011111110	8	131
0101111011	7	13	1110100111	7	76	1001111011	7	138
0011100110	5	18	0111100010	5	81	1101101110	7	145
1110011101	7	25	1110111100	7	88	1001011110	6	151
1110000011	5	30	0000100101	3	91	0111110001	6	157
0001111001	5	35	0001111011	6	97	1101001001	5	162
1101000101	5	40	1110011001	6	103	1101001100	5	167
1101111000	6	46	1011111110	8	111	1110100101	6	173
1111011111	9	55	1101010101	6	117	0110001100	4	177
1000100111	5	60	1100100111	6	123	1011010110	6	183
1011101011	7	190	0111110011	7	251	1101100110	6	315
1101100010	5	195	0110110111	7	258	1011011101	7	322
1100000111	5	200	1011110010	6	264	1100101111	7	329
0001111111	7	207	0000111001	4	268	1111010010	6	335
0010001101	4	211	0111100101	6	274	1110001111	7	342
0101001011	5	216	1111111101	9	283	0101001111	6	348
0100111011	6	222	1101010101	6	289	1001100101	5	353
0111111101	8	230	0111111110	8	297	1101111001	7	360
1111111010	8	238	0011101000	4	301	1010111111	8	368
1110001011	6	244	0111111101	8	309	0111101000	5	373
1001110011	6	379	1011011100	6	444	0111111110	8	518
1111011111	9	388	1011111010	7	451	0100010010	3	521
0110011111	7	395	1011110111	8	459	0101101011	6	527
1000011101	5	400	0011001111	6	465	0011111101	7	534
1111100111	8	408	1000001101	4	469	1100100101	5	539
1010010001	4	412	1011011111	8	477	1010110110	6	545
0100001110	4	416	1111111111	10	487	1111101101	8	553
1111011010	7	423	1011111101	8	495	1110111101	8	561
1010111110	7	430	1011111101	8	503	1110110110	7	568
1101011111	8	438	1110011101	7	510	0111001101	6	574
1000010110	4	578	1101001001	5	634	1100110011	6	690
0110110110	6	584	1110111111	9	643	1010111101	7	697
1000101100	4	588	0110110110	6	649	1010111110	7	704
1011011101	7	595	1000001000	2	651	1111111111	10	714
1100110110	6	601	1000011000	3	654	0011011110	6	720
1111011010	7	608	0111001011	6	660	0111111110	8	728
0001000110	3	611	1100010100	4	664	1001111101	7	735
0001110001	4	615	0101110110	6	670	0100111101	6	741
1111100111	8	623	1111111001	8	678	0100110101	5	746
0010111110	6	629	0000111111	6	684	1110101010	6	752

Annexe A

Intégrale de Riemann sur $[a, b]$

A.1 Construction

Soit $[a, b]$ un intervalle *fermé borné* de \mathbb{R} . On appelle subdivision de $[a, b]$ toute suite finie du type $\Delta = \{x_0 = a < x_1 < \cdots < x_n = b\}$. Pour une fonction *bornée* $f : [a, b] \rightarrow \mathbb{R}$, $(-\infty < a < b < +\infty)$, on définit ses sommes de Darboux inférieure $S_\Delta(f)$ et supérieure $S^\Delta(f)$ par

$$S_\Delta(f) := \sum_{k=1}^n (x_k - x_{k-1}) \inf_{[x_{k-1}, x_k]} f, \quad S^\Delta(f) := \sum_{k=1}^n (x_k - x_{k-1}) \sup_{[x_{k-1}, x_k]} f.$$

Pour une illustration, voir les figures [A.1](#) et [A.2](#).

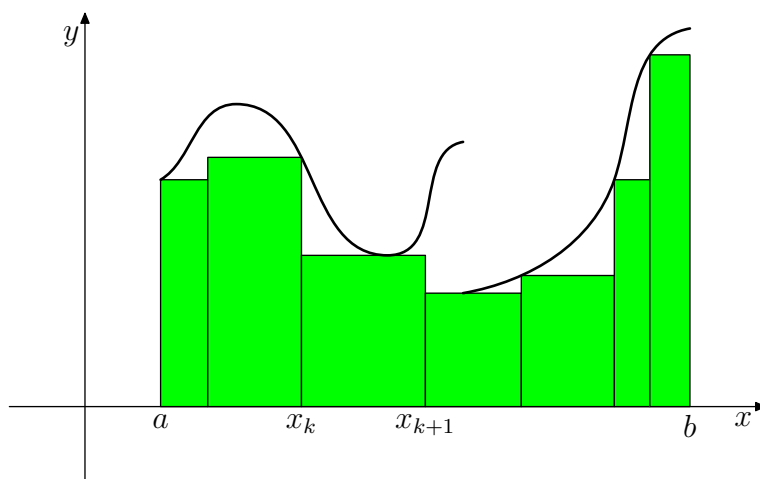


FIG. A.1 – $S_\Delta(f)$

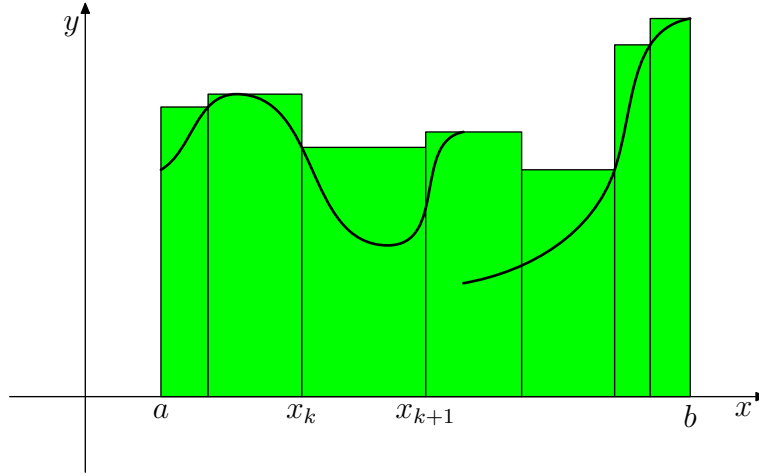


FIG. A.2 – $S^\Delta(f)$

On dit que la subdivision Δ' est un raffinement de Δ si l'ensemble des valeurs de la suite finie Δ est inclus dans celui des valeurs de la suite Δ' , ce que nous noterons avec un léger abus $\Delta \subset \Delta'$. Il est facile de vérifier que

$$\Delta \subset \Delta' \Rightarrow S_\Delta(f) \leq S_{\Delta'}(f) \text{ et } S^\Delta(f) \geq S^{\Delta'}(f).$$

Les figures A.3 et A.4 illustrent l'effet de l'adjonction à la subdivision Δ des figures A.1 et A.2 de deux nouveaux points.

Les intégrales de Riemann *inférieure* $I_*(f)$ et *supérieure* $I^*(f)$ sont définies par

$$I_*(f) := \sup_{\Delta} S_\Delta(f), \quad I^*(f) := \inf_{\Delta} S^\Delta(f),$$

le supremum et l'infimum étant pris sur toutes les subdivisions Δ de $[a, b]$.

Pour Δ_1 et Δ_2 subdivisions de $[a, b]$ on a clairement

$$S_{\Delta_1}(f) \leq S_{\Delta_1 \cup \Delta_2}(f) \leq S^{\Delta_1 \cup \Delta_2}(f) \leq S^{\Delta_2}(f),$$

d'où $S_{\Delta_1}(f) \leq S^{\Delta_2}(f)$. En prenant successivement le sup sur tous les Δ_1 , puis l'inf sur tous les Δ_2 , on en déduit

$$I_*(f) \leq I^*(f),$$

inégalité vérifiée par toute fonction bornée $f : [a, b] \rightarrow \mathbb{R}$.

Définition A.1. On dit que f bornée $[a, b] \rightarrow \mathbb{R}$ est Riemann intégrable si avec les notations ci-dessus, $I_*(f) = I^*(f)$. Dans ce cas on définit son intégrale au sens de Riemann notée $\int_a^b f(x) dx$ par

$$\int_a^b f(x) dx := I_*(f) = I^*(f).$$

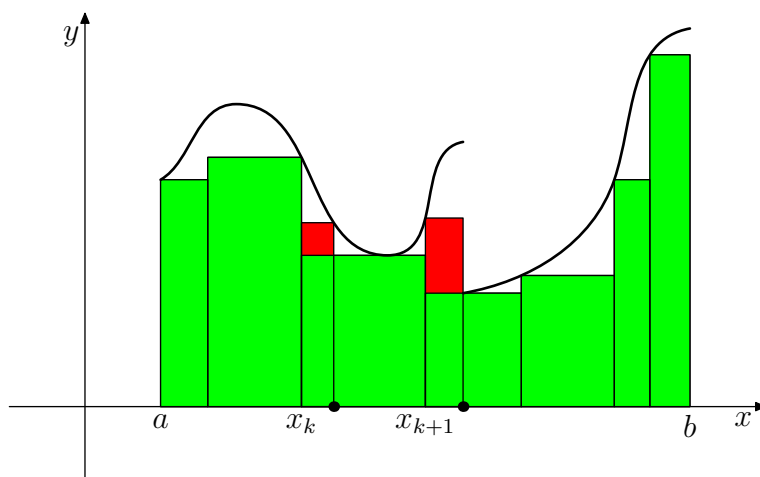


FIG. A.3 – Si $\Delta \subset \Delta'$, $S_{\Delta}(f) \leq S_{\Delta'}(f)$

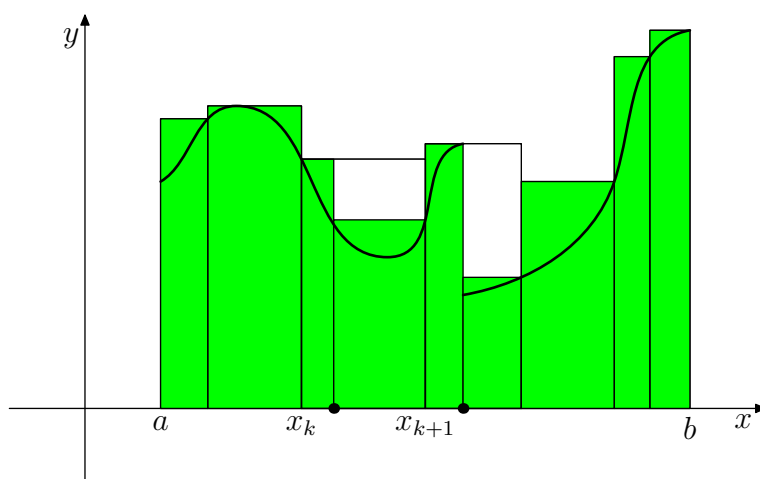


FIG. A.4 – Si $\Delta \subset \Delta'$, $S^{\Delta}(f) \geq S^{\Delta'}(f)$

Il est commode de donner aussi une définition de $\int_a^b f(x) dx$ lorsque $b < a$. Cette définition peut se justifier en reprenant toute l'étude précédente avec des subdivisions Δ de $[b, a]$ par des suites finies décroissantes¹ $a = x_0 > x_1 > \dots > x_{n-1} > x_n = b$. En conservant les mêmes définitions de $S^\Delta(f)$ et $S_\Delta(f)$, le seul changement par rapport aux subdivisions croissantes de $[b, a]$ est que les $(x_k - x_{k-1})$ sont négatifs ce qui implique une inversion des inégalités entre $S^\Delta(f)$ et $S_\Delta(f)$, on a maintenant $S^\Delta(f) \leq S_\Delta(f)$. Associons à chaque subdivision décroissante Δ de $[b, a]$, la subdivision *retournée* $\Delta' = \{x'_0, x'_1, \dots, x'_n\}$ définie par $x'_0 = x_n, x'_1 = x_{n-1}, \dots, x'_n = x_0$. Alors Δ' est une subdivision croissante de $[b, a]$ et $S^{\Delta'}(f) = -S^\Delta(f)$, $S_{\Delta'}(f) = -S_\Delta(f)$. On en déduit immédiatement que

$$I_*(f, a, b) := \sup_{\Delta} S^\Delta(f) = -\inf_{\Delta'} S^{\Delta'}(f) =: -I^*(f, b, a), \quad (\text{A.1})$$

$$I^*(f, a, b) := \inf_{\Delta} S_\Delta(f) = -\sup_{\Delta'} S_{\Delta'}(f) =: -I_*(f, b, a), \quad (\text{A.2})$$

les infima et suprema indexés par Δ s'entendant pour toute subdivision décroissante de a à b et ceux indexés par Δ' pour toute subdivision croissante de $[b, a]$. On définit alors l'intégrabilité de f de a à b par la condition $I_*(f, a, b) = I^*(f, a, b)$, dont on voit par (A.1) et (A.2) qu'elle équivaut à $I^*(f, b, a) = I_*(f, b, a)$, c'est-à-dire à l'intégrabilité de f sur $[b, a]$. En définissant enfin $\int_a^b f(x) dx$ comme la valeur commune de $I_*(f, a, b)$ et $I^*(f, a, b)$, on obtient $\int_a^b f(x) dx = -\int_b^a f(x) dx$, cette dernière intégrale relevant de la définition A.1. Tout ceci légitime la définition formelle suivante.

Définition A.2. Si $-\infty < b < a < +\infty$, on dit que f est Riemann intégrable de a à b si elle est Riemann intégrable sur $[b, a]$ et on pose dans ce cas :

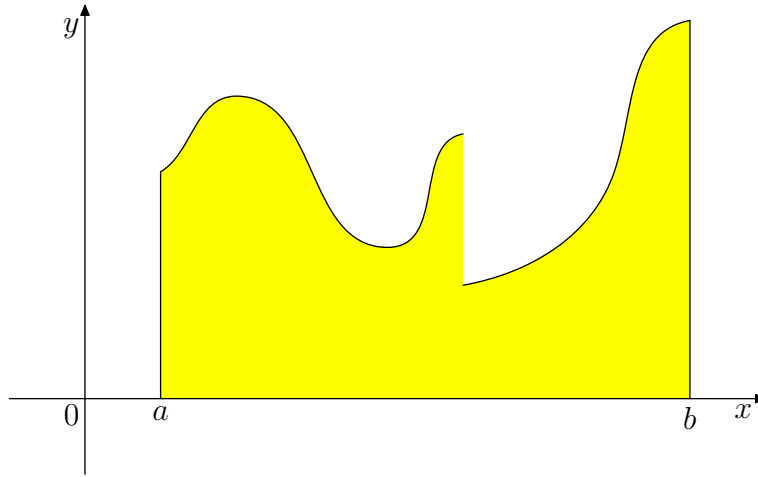
$$\int_a^b f(x) dx := -\int_b^a f(x) dx. \quad (\text{A.3})$$

Remarque A.3 (variable d'intégration). Dans l'écriture $\int_b^a f(x) dx$, la « variable d'intégration » x est « muette », on peut la remplacer par n'importe quelle autre lettre (sauf ici a, b ou f). Cette variable joue le même rôle que l'indice i de sommation dans $\sum_{i=1}^n u_i$ qui est lui aussi muet.

Remarque A.4 (intégrale de Riemann et aire). Soit f une fonction positive et Riemann intégrable sur $[a, b]$. On interprète classiquement $\int_a^b f(x) dx$ comme *l'aire de l'hypographe* de f entre a et b , i.e. de la région du plan délimitée par l'axe des abscisses, les droites verticales d'équation $x = a$ ou $x = b$ et le graphe² de f , la courbe d'équation $y = f(x)$, $x \in [a, b]$. Voici une justification informelle de cette affirmation, dont on pourra se contenter en première lecture. Reprenons la fonction f des figures A.1 et A.2. L'hypographe H de f est représenté figure A.5. On peut se convaincre « visuellement », cf. figure A.1,

1. Rappelons que pour tous réels a et b , $[a, b]$ est défini comme l'ensemble des x réels tels que $a \leq x \leq b$. Ainsi pour $b < a$, $[a, b]$ est l'ensemble vide. C'est pour cela que l'on subdivise ici $[b, a]$ et non $[a, b]$. De même on parlera d'intégrale de f de a à b mais pas d'intégrale de f sur $[a, b]$ quand $b < a$.

2. D'où le nom « hypographe », littéralement ce qui est sous le graphe.

FIG. A.5 – Hypographe H de f entre a et b

que pour toute somme de Darboux inférieure l'aire des rectangles coloriés égale à $S_{\Delta}(f)$ est inférieure à l'aire de l'hypographe de f . De même cf. figure A.2, pour toute somme de Darboux supérieure, l'aire des rectangles coloriés égale à $S^{\Delta}(f)$ est supérieure à l'aire de l'hypographe. L'aire de H est donc un majorant de toute S_{Δ} et un minorant de toute S^{Δ} . D'où

$$I_*(f) = \sup_{\Delta} S_{\Delta}(f) \leq \text{aire}(H) \leq \inf_{\Delta} S_{\Delta}(f) = I^*(f).$$

Par Riemman intégrabilité de f , $I_*(f) = I^*(f) = \int_a^b f(x) dx$, d'où $\text{aire}(H) = \int_a^b f(x) dx$.

Pour les lecteurs exigeants que la remarque A.4 laisserait insatisfaits, nous proposons et démontrons ci-dessous un énoncé plus précis. Pour cela, il convient d'abord de s'interroger sur la définition mathématique de l'aire de H . Dans le cadre de ce cours, nous avons admis l'existence de la mesure de Lebesgue λ_2 sur \mathbb{R}^2 , définie comme l'unique mesure μ sur la tribu borélienne de \mathbb{R}^2 vérifiant $\mu([x_1, x_2] \times [y_1, y_2]) = (x_2 - x_1)(y_2 - y_1)$ pour tout pavé semi-ouvert $[x_1, x_2] \times [y_1, y_2]$ de \mathbb{R}^2 , voir l'exemple 2.13 p. 56. C'est cette mesure de Lebesgue qui donne un sens mathématique précis à la notion d'aire. On se propose donc de montrer que $\lambda_2(H) = \int_a^b f(x) dx$. Pour que cela ait un sens, encore faut-il que H soit un borélien de \mathbb{R}^2 . Une condition suffisante pour que H soit un borélien de \mathbb{R}^2 est que la fonction f soit borélienne, c'est-à-dire mesurable (voir def. 3.1 93) pour les tribus boréliennes de $[a, b]$ et de \mathbb{R} . La preuve de cette affirmation sort du programme de ce cours³. Signalons simplement que tous les exemples de fonctions Riemman intégrables donnés dans la suite de ce document — fonctions monotones, continues, réglées — sont boréliennes. Les seules propriétés de λ_2 utilisées dans ce qui suit sont la croissance et l'additivité finie — propriétés vérifiées par toute mesure, voir p. 51 — et le fait que les frontières des pavés sont de λ_2 -mesure nulle⁴, voir la proposition 2.14 v).

3. Voir le cours d'IFP 2003-2004 chapitre 5.

4. En réalité on a seulement besoin de savoir que si J est un segment vertical $\{a\} \times [y_1, y_2]$, $\lambda_2(J) = 0$ et de même pour un segment horizontal. Ceci se démontre facilement en exercice en utilisant la croissance

Proposition A.5. Soit $f : [a, b] \rightarrow \mathbb{R}$ une fonction positive et Riemann intégrable sur $[a, b]$. On suppose de plus que son hypographe entre a et b

$$H := \{(x, y) \in \mathbb{R}^2; a \leq x \leq b \text{ et } 0 \leq y \leq f(x)\} \quad (\text{A.4})$$

est un borélien de \mathbb{R}^2 . Alors

$$\lambda_2(H) = \int_a^b f(x) dx, \quad (\text{A.5})$$

autrement dit l'intégrale de f entre a et b est l'aire de l'hypographe de f entre a et b .

Preuve. Soit $\Delta = \{x_0 = a < x_1 < \dots < x_n = b\}$ une subdivision quelconque de $[a, b]$. Notons pour $k = 1, \dots, n$,

$$m_k := \inf_{[x_{k-1}, x_k]} f, \quad M_k := \sup_{[x_{k-1}, x_k]} f.$$

Définissons les « rectangles » $R_{\Delta, k}$ par

$$R_{\Delta, 1} := [a, x_1] \times [0, m_1], \quad R_{\Delta, k} :=]x_{k-1}, x_k] \times [0, m_k] \text{ pour } k = 2, \dots, n,$$

et notons $R^{\Delta, k}$ les rectangles obtenus en remplaçant m_k par M_k , $k = 1, \dots, n$. Posons enfin

$$R_{\Delta} := \bigcup_{k=1}^n R_{\Delta, k}, \quad R^{\Delta} := \bigcup_{k=1}^n R^{\Delta, k}.$$

Commençons par justifier la double inclusion

$$\forall \Delta, \quad R_{\Delta} \subset H \subset R^{\Delta}. \quad (\text{A.6})$$

Soit (x', y') un élément quelconque de la réunion R_{Δ} . Il appartient donc à un $R_{\Delta, k'}$ d'où $x_{k'-1} \leq x' \leq x_{k'}$ et $0 \leq y' \leq m_{k'} \leq f(x')$ car $m_{k'}$ est l'infimum de f sur $[x_{k'-1}, x_{k'}]$. Le couple (x', y') vérifie ainsi les inégalités $a \leq x_{k'-1} \leq x' \leq x_{k'} \leq b$ et $0 \leq y' \leq f(x')$, donc appartient à H . Ceci justifie la première inclusion dans (A.6). Soit maintenant (x'', y'') un élément quelconque de H , donc vérifiant $a \leq x'' \leq b$ et $0 \leq y'' \leq f(x'')$. La subdivision Δ induit la partition de $[a, b]$ en les intervalles $J_1 := [a, x_1]$, $J_k :=]x_{k-1}, x_k]$, $k = 2, \dots, n$. Il existe donc un unique indice k'' entre 1 et n tel que $J_{k''}$ contienne x'' . On a alors $0 \leq y'' \leq f(x'') \leq M_{k''}$ car $M_{k''}$ est le supremum de f sur $[x_{k''-1}, x_{k''}]$. Ainsi (x'', y'') appartient à $R^{\Delta, k''}$, donc aussi à R^{Δ} , ce qui justifie la deuxième inclusion dans (A.6).

Comme R_{Δ} , H et R^{Δ} sont des boréliens de \mathbb{R}^2 , on déduit de (A.6) par croissance de λ_2 que

$$\forall \Delta, \quad \lambda_2(R_{\Delta}) \leq \lambda_2(H) \leq \lambda_2(R^{\Delta}). \quad (\text{A.7})$$

Calculons maintenant $\lambda_2(R_{\Delta})$. Les $R_{\Delta, k}$ étant deux à deux disjoints, on a par additivité finie de λ_2 :

$$\lambda_2(R_{\Delta}) = \sum_{k=1}^n \lambda_2(R_{\Delta, k}). \quad (\text{A.8})$$

de λ . Faites le !

Par la proposition 2.14 v) ou par la note 4 p. 217, on a pour tout $k = 1, \dots, n$, $\lambda_2(R_{\Delta,k}) = (x_k - x_{k-1})m_k$, d'où en reportant dans (A.8), $\lambda_2(R_\Delta) = S_\Delta(f)$. De même il est clair que $\lambda_2(R^\Delta) = S^\Delta(f)$. On déduit alors de (A.7) que

$$\forall \Delta, \quad S_\Delta(f) \leq \lambda_2(H) \leq \lambda_2(R^\Delta). \quad (\text{A.9})$$

La première inégalité dans (A.9) nous dit que le réel $\lambda_2(H)$ qui ne dépend pas de Δ majore toutes les sommes de Darboux inférieures $S_\Delta(f)$. Il majore donc aussi leur supremum $I_*(f)$. Par la deuxième inégalité, $\lambda_2(H)$ minore toutes les $S^\Delta(f)$, donc minore aussi leur infimum $I^*(f)$. Nous obtenons ainsi l'encadrement

$$I_*(f) \leq \lambda_2(H) \leq I^*(f).$$

Comme f est Riemann intégrable, $I_*(f) = I^*(f)$, donc $\lambda_2(H) = I_*(f) = I^*(f) = \int_a^b f(x) dx$. \square

Remarque A.6. La mesure λ_2 étant invariante par la symétrie $(x, y) \mapsto (x, -y)$ cf. prop. 2.14 ii), on obtient immédiatement une version de la proposition A.5 pour une fonction g négative sur $[a, b]$ en remplaçant H par

$$H' := \{(x, y) \in \mathbb{R}^2; a \leq x \leq b \text{ et } g(x) \leq y \leq 0\}.$$

En effet en posant $f = -g$, il vient

$$\lambda_2(H') = \lambda_2(H) = \int_a^b f(x) dx = \int_a^b |g(x)| dx. \quad (\text{A.10})$$

A.2 Riemann intégrabilité

Dans cette section nous examinons la Riemann intégrabilité de certaines familles de fonctions. Les deux plus importantes en pratique sont celle des fonctions monotones et celle des fonctions continues. On généralise la Riemann intégrabilité des fonctions continues au cas des fonctions bornées continues sur $[a, b]$ sauf en un nombre fini de points, comme celle de la figure A.1. Enfin nous établissons que toute limite uniforme sur $[a, b]$ d'une suite de fonctions Riemann intégrables sur $[a, b]$ est encore Riemann intégrable. Ceci nous donne notamment la Riemann intégrabilité de toutes les fonction *réglées*, *i.e.* limites uniformes de fonctions en escalier. Les fonctions en escaliers sont les plus simples de toutes les fonctions Riemann intégrables et c'est par elles que nous commençons cette étude.

Définition A.7 (fonction en escalier). *Une application $f : [a, b] \rightarrow \mathbb{R}$ est appelée fonction en escalier sur $[a, b]$, s'il existe une subdivision $\Delta_0 = \{t_0 = a < t_1 < \dots < t_j = b\}$ telle que f soit constante sur chaque intervalle ouvert $]t_{i-1}, t_i[$, $i = 1, \dots, j$.*

Il est clair que Δ_0 n'est pas unique, en particulier pour tout raffinement Δ'_0 de Δ_0 , f est constante sur chacun des intervalles ouverts ayant pour extrémités deux points

consécutifs de Δ'_0 . Il y a donc une infinité de subdivisions Δ telles que f en escalier soit constante sur chacun des intervalles ouverts de Δ (*i.e.* les intervalles ayant pour extrémités deux points consécutifs de Δ). Nous appellerons *subdivision* associée à f en escalier, toute subdivision Δ telle que f soit constante sur chacun des intervalles ouverts de Δ . La moins fine des subdivisions associées à f en escalier est constituée des points a et b et des points de discontinuité de f dans $]a, b[$.

Proposition A.8 (intégrabilité d'une fonction en escalier). *Soit f une fonction en escalier sur $[a, b]$ et $\Delta_0 = \{t_0 = a < t_1 < \dots < t_j = b\}$ une subdivision associée à f , la valeur constante de f sur $]t_{i-1}, t_i[$ étant notée c_i . Alors f est Riemann intégrable sur $[a, b]$ et on a*

$$\int_a^b f(x) dx = \sum_{i=1}^j (t_i - t_{i-1}) c_i. \quad (\text{A.11})$$

Remarquons que comme $\int_a^b f(x) dx$ ne dépend, lorsqu'elle existe, que de f , (A.11) implique que si $\Delta_1 = \{s_0 = a < s_1 < \dots < s_l = b\}$ est une autre subdivision associée à f et en notant d_k la valeur constante de f sur $]s_{k-1}, s_k[$, $k = 1, \dots, l$, on a

$$\sum_{i=1}^j (t_i - t_{i-1}) c_i = \sum_{k=1}^l (s_k - s_{k-1}) d_k.$$

Preuve. D'abord, f est bornée puisque $f([a, b]) = \{c_1, \dots, c_j\} \cup \{f(t_0), \dots, f(t_j)\}$ qui est fini (de cardinal au plus $2j + 1$) donc borné dans \mathbb{R} . Pour chaque δ vérifiant

$$0 < \delta < \frac{1}{2} \min_{1 \leq i \leq j} (t_i - t_{i-1}), \quad (\text{A.12})$$

notons Δ_δ la subdivision construite en adjoignant à Δ_0 les points $t_0 + \delta, t_1 - \delta, t_1 + \delta, t_2 - \delta, t_2 + \delta, \dots, t_{j-1} + \delta, t_j - \delta$. Notons en outre

$$m := \inf_{x \in [a, b]} f(x), \quad M := \sup_{x \in [a, b]} f(x),$$

$$m'_i := \inf_{|t_i - x| \leq \delta} f(x) = \min(c_i, c_{i+1}, f(t_i)), \quad M'_i := \sup_{|t_i - x| \leq \delta} f(x) = \max(c_i, c_{i+1}, f(t_i)),$$

avec l'adaptation évidente pour $i = j$. On a bien sûr $M'_i \leq M$ et $m'_i \geq m$ pour tout i . Avec ces notations on a

$$\begin{aligned} S^{\Delta_\delta}(f) &= \sum_{i=1}^j (t_i - t_{i-1} - 2\delta) c_i + \delta M'_0 + \delta M'_j + 2\delta \sum_{i=1}^{j-1} M'_i \\ &\leq \sum_{i=1}^j (t_i - t_{i-1}) c_i + 2j\delta(M - m). \end{aligned} \quad (\text{A.13})$$

De même avec les m'_i à la place de M'_i on obtient

$$S_{\Delta_\delta}(f) \geq \sum_{i=1}^j (t_i - t_{i-1}) c_i - 2j\delta(M - m). \quad (\text{A.14})$$

Soit $\varepsilon > 0$ quelconque, en choisissant $\delta = \delta(\varepsilon)$ vérifiant à la fois (A.12) et $2j\delta(M-m) < \varepsilon$, on dispose ainsi par (A.13) et (A.14) d'une subdivision $\Delta_{\delta(\varepsilon)}$ telle que

$$\sum_{i=1}^j (t_i - t_{i-1})c_i - \varepsilon < S_{\Delta_{\delta(\varepsilon)}} \leq S^{\Delta_{\delta(\varepsilon)}} < \sum_{i=1}^j (t_i - t_{i-1})c_i + \varepsilon.$$

On en déduit pour tout $\varepsilon > 0$ l'encadrement :

$$\sum_{i=1}^j (t_i - t_{i-1})c_i - \varepsilon < I_*(f) \leq I^*(f) < \sum_{i=1}^j (t_i - t_{i-1})c_i + \varepsilon,$$

puis en faisant tendre ε vers 0 que

$$I_*(f) = I^*(f) = \sum_{i=1}^j (t_i - t_{i-1})c_i,$$

ce qui établit l'intégrabilité de f et (A.11). \square

Proposition A.9 (intégrabilité d'une fonction monotone). *Si $f : [a, b] \rightarrow \mathbb{R}$ est monotone sur $[a, b]$, elle est Riemann intégrable sur $[a, b]$.*

Preuve. Supposons pour fixer les idées que f est décroissante, l'adaptation de ce qui suit au cas f croissante étant immédiate. Alors f est bornée puisque pour tout $x \in [a, b]$, $f(b) \leq f(x) \leq f(a)$. Pour $\Delta = \{x_0 = a < x_1 < \dots < x_n = b\}$ subdivision quelconque de $[a, b]$, notons m_k et M_k les bornes inférieure et supérieure de f sur $[x_{k-1}, x_k]$ et remarquons que par décroissance de f , $m_k = f(x_k)$ et $M_k = f(x_{k-1})$. On a alors

$$\begin{aligned} 0 \leq S^\Delta(f) - S_\Delta(f) &= \sum_{k=1}^n (x_k - x_{k-1})(M_k - m_k) \\ &= \sum_{k=1}^n (x_k - x_{k-1})(f(x_{k-1}) - f(x_k)) \\ &\leq \max_{1 \leq k \leq n} (x_k - x_{k-1}) \sum_{k=1}^n (f(x_{k-1}) - f(x_k)) \\ &= \max_{1 \leq k \leq n} (x_k - x_{k-1})(f(a) - f(b)). \end{aligned}$$

Soit $\varepsilon > 0$ quelconque. En choisissant une subdivision Δ de pas au plus ε , i.e. telle que $\max_{1 \leq k \leq n} (x_k - x_{k-1}) \leq \varepsilon$, on a

$$0 \leq S^\Delta(f) - S_\Delta(f) \leq \varepsilon(f(a) - f(b)).$$

On en déduit que $I^*(f) - I_*(f) \leq \varepsilon(f(a) - f(b))$, puis comme ε est quelconque que $I^*(f) - I_*(f) = 0$. La fonction f est donc Riemann intégrable. \square

Proposition A.10 (intégrabilité d'une fonction continue). *Si $f : [a, b] \rightarrow \mathbb{R}$ est continue, elle est Riemann intégrable sur $[a, b]$. De plus si F est une primitive de f sur $[a, b]$,*

$$\int_a^b f(x) dx = F(b) - F(a). \quad (\text{A.15})$$

Preuve. Sur le compact $[a, b]$, la fonction f est bornée et *uniformément* continue :

$$\forall \varepsilon > 0, \quad \exists \delta > 0, \quad \forall x, x' \in [a, b], \quad |x - x'| < \delta \Rightarrow |f(x) - f(x')| < \varepsilon.$$

Pour chaque $\varepsilon > 0$, on peut trouver une subdivision $\Delta = \{x_0 = a < x_1 < \dots < x_n = b\}$ (dépendant de δ) telle que pour $k = 1, \dots, n$, $x_k - x_{k-1} < \delta$. Comme les bornes inférieure m_k et supérieure M_k de f sur le compact $[x_{k-1}, x_k]$ sont atteintes, on a pour tout k $0 \leq M_k - m_k < \varepsilon$. On a alors

$$0 \leq S^\Delta(f) - S_\Delta(f) = \sum_{k=1}^n (x_k - x_{k-1})(M_k - m_k) \leq \varepsilon \sum_{k=1}^n (x_k - x_{k-1}) = \varepsilon(b - a).$$

En raison de l'encadrement $S_\Delta(f) \leq I_*(f) \leq I^*(f) \leq S^\Delta(f)$, nous avons ainsi établi que

$$\forall \varepsilon > 0, \quad 0 \leq I^*(f) - I_*(f) \leq \varepsilon(b - a).$$

Comme $I^*(f) - I_*(f)$ ne dépend pas de ε , on en déduit que $I^*(f) - I_*(f) = 0$, d'où la Riemann intégrabilité de f sur $[a, b]$.

Rappelons que F est une primitive de f sur $[a, b]$ si elle est dérivable en tout point de $[a, b]$ (à droite en a et à gauche en b) et pour tout $x \in [a, b]$, $F'(x) = f(x)$. Soit $\Delta = \{x_0 = a < x_1 < \dots < x_n = b\}$ une subdivision *quelconque* de $[a, b]$. Par le théorème des accroissements finis⁵, il existe dans chaque $]x_{k-1}, x_k[$ un c_k tel que $F(x_k) - F(x_{k-1}) = (x_k - x_{k-1})F'(c_k) = (x_k - x_{k-1})f(c_k)$. En écrivant

$$F(b) - F(a) = \sum_{k=1}^n (F(x_k) - F(x_{k-1})) = \sum_{k=1}^n (x_k - x_{k-1})f(c_k)$$

et en encadrant $f(c_k)$ entre les bornes inférieure et supérieure de f sur $[x_{k-1}, x_k]$, on en déduit

$$S_\Delta(f) \leq F(b) - F(a) \leq S^\Delta(f).$$

Cet encadrement est valide pour toute subdivision Δ et $F(b) - F(a)$ ne dépend pas de Δ . Par conséquent

$$I_*(f) \leq F(b) - F(a) \leq I^*(f)$$

et comme nous savons déjà que f est Riemann intégrable on en déduit $F(b) - F(a) = I_*(f) = I^*(f)$, ce qui établit (A.15). \square

5. Appelé aussi *formule des accroissements finis* : si f est continue sur $[a, b]$ et dérivable sur $]a, b[$, il existe $c \in]a, b[$ tel que $f(b) - f(a) = f'(c)(b - a)$.

La preuve de (A.15) s'étend immédiatement au cas où F est dérivable sur $]a, b[$, continue à droite en a et à gauche en b et $F' = f$ sur $]a, b[$. D'autre part on peut utiliser l'intégrale de Riemann pour montrer que toute fonction continue sur $[a, b]$ admet des primitives sur $[a, b]$.

Proposition A.11. *Toute fonction $f : [a, b] \rightarrow \mathbb{R}$, bornée sur $[a, b]$ et continue sur $[a, b]$, sauf en un nombre fini de points est Riemann intégrable sur $[a, b]$.*

Preuve. Nous nous contenterons de le montrer dans le cas où f présente un seul point de discontinuité $c \in]a, b[$, la généralisation ne coûtant qu'un alourdissement de notations. L'adaptation de ce qui suit au cas $c = a$ ou $c = b$ est aussi immédiate.

Fixons $\varepsilon > 0$ arbitraire et soit $\eta > 0$ assez petit pour que $[c - \eta, c + \eta] \subset]a, b[$ et dont le choix en fonction de ε sera précisé ultérieurement. Soit Δ une subdivision de $[a, b]$ ayant comme points consécutifs $c - \eta$ et $c + \eta$ (i. e. $x_{k_0} = c - \eta$ et $x_{k_0+1} = c + \eta$ pour un certain indice k_0). Cette subdivision Δ peut se construire comme réunion d'une subdivision quelconque Δ_1 de $[a, c - \eta]$ et d'une subdivision quelconque Δ_2 de $[c + \eta, b]$. Comme f est continue sur $[a, c - \eta]$ et $[c + \eta, b]$, elle est Riemann intégrable sur chacun de ces deux segments (prop. A.10), ce qui nous autorise à choisir Δ_1 et Δ_2 telles que

$$0 \leq S^{\Delta_1}(f) - S_{\Delta_1}(f) \leq \frac{\varepsilon}{3}, \quad 0 \leq S^{\Delta_2}(f) - S_{\Delta_2}(f) \leq \frac{\varepsilon}{3}. \quad (\text{A.16})$$

Notons m et M , m_η et M_η les bornes inférieure et supérieure de f sur respectivement $[a, b]$ et $[c - \eta, c + \eta]$. On a clairement $m \leq m_\eta \leq M_\eta \leq M$, d'où $2\eta(M_\eta - m_\eta) \leq 2\eta(M - m)$, de sorte qu'en choisissant

$$\eta < \frac{\varepsilon}{6(M - m)},$$

on ait

$$2\eta(M_\eta - m_\eta) < \frac{\varepsilon}{3}. \quad (\text{A.17})$$

Avec le choix de Δ opéré ci-dessus, nous avons

$$\begin{aligned} S^\Delta(f) &= S^{\Delta_1}(f) + 2\eta M_\eta + S^{\Delta_2}(f) \\ S_\Delta(f) &= S_{\Delta_1}(f) + 2\eta m_\eta + S_{\Delta_2}(f), \end{aligned}$$

d'où compte-tenu de (A.16) et (A.17),

$$0 \leq S^\Delta(f) - S_\Delta(f) \leq S^{\Delta_1}(f) - S_{\Delta_1}(f) + 2\eta(M_\eta - m_\eta) + S^{\Delta_2}(f) - S_{\Delta_2}(f) < \varepsilon.$$

On en déduit que $0 \leq I^*(f) - I_*(f) < \varepsilon$, puis par arbitrarité de ε que $I^*(f) = I_*(f)$, i. e. que f est Riemann intégrable sur $[a, b]$. \square

Nous allons maintenant établir que la Riemann intégrabilité se conserve par convergence uniforme sur $[a, b]$. Le lemme suivant nous sera utile.

Lemme A.12. Soit E une partie quelconque de \mathbb{R} . On suppose que chaque fonction f_n est définie et bornée sur E et que la suite $(f_n)_{n \geq 1}$ converge vers f uniformément sur E . Alors f est bornée sur E et

$$\begin{aligned} m_n(E) &:= \inf_{x \in E} f_n(x) \xrightarrow{n \rightarrow +\infty} \inf_{x \in E} f(x) =: m(E), \\ M_n(E) &:= \sup_{x \in E} f_n(x) \xrightarrow{n \rightarrow +\infty} \sup_{x \in E} f(x) =: M(E). \end{aligned}$$

Plus précisément, si pour tout $n \geq n_\varepsilon$, on a pour tout $x \in E$, $|f_n(x) - f(x)| < \varepsilon$, alors

$$\forall n \geq n_\varepsilon, \quad |m_n(E) - m(E)| \leq \varepsilon \quad \text{et} \quad |M_n(E) - M(E)| \leq \varepsilon. \quad (\text{A.18})$$

Preuve. La convergence uniforme de (f_n) vers f sur E , s'écrit

$$\forall \varepsilon > 0, \exists n_\varepsilon \in \mathbb{N}, \forall n \geq n_\varepsilon, \forall x \in E, \quad |f_n(x) - f(x)| < \varepsilon. \quad (\text{A.19})$$

En réécrivant cette inégalité sous la forme $f_n(x) - \varepsilon < f(x) < f_n(x) + \varepsilon$ on en déduit :

$$\forall n \geq n_\varepsilon, \forall x \in E, \quad m_n(E) - \varepsilon < f(x) < M_n(E) + \varepsilon,$$

puis en prenant l'infimum et le supremum en $x \in E$ dans cette double inégalité⁶ :

$$\forall n \geq n_\varepsilon, \quad m_n(E) - \varepsilon \leq m(E) \quad \text{et} \quad M(E) \leq M_n(E) + \varepsilon. \quad (\text{A.20})$$

En choisissant un n particulier, par exemple $n = n_\varepsilon$, on en déduit que f est bornée sur E ($-\infty < m_n(E) - \varepsilon \leq m(E) \leq M(E) \leq M_n(E) + \varepsilon < +\infty$).

En réécrivant l'inégalité (A.19) sous la forme $f(x) - \varepsilon < f_n(x) < f(x) + \varepsilon$, on obtient de la même façon :

$$\forall n \geq n_\varepsilon, \quad m(E) - \varepsilon \leq m_n(E) \quad \text{et} \quad M_n(E) \leq M(E) + \varepsilon. \quad (\text{A.21})$$

En regroupant (A.20) et (A.21), on voit ainsi que pour tout $n \geq n_\varepsilon$,

$$m(E) - \varepsilon \leq m_n(E) \leq m(E) + \varepsilon \quad \text{et} \quad M(E) - \varepsilon \leq M_n(E) \leq M(E) + \varepsilon,$$

ce qui nous donne (A.18) et donc les convergences de $m_n(E)$ et $M_n(E)$ vers respectivement $m(E)$ et $M(E)$ puisque $\varepsilon > 0$ est ici arbitraire. \square

Proposition A.13. Si f est limite uniforme sur $[a, b]$ d'une suite $(f_n)_{n \geq 1}$ de fonctions Riemann intégrables sur $[a, b]$, alors f est elle-même Riemann intégrable sur $[a, b]$.

Preuve. D'abord, f est bornée sur $[a, b]$ comme limite uniforme d'une suite de fonctions bornées (lemme A.12 avec $E = [a, b]$). On peut donc bien définir les sommes de Darboux $S_\Delta(f)$ et $S^\Delta(f)$ pour toute subdivision Δ de $[a, b]$.

Notons qu'il y a une difficulté supplémentaire dans cette démonstration par rapport aux preuves de la Riemann intégrabilité des fonctions monotones ou continues. Dans

6. Noter ici le passage des inégalités strictes aux inégalité larges.

ces deux cas, f atteignait ses bornes inférieure et supérieure sur chaque intervalle de la subdivision, ce qui facilitait le traitement des sommes de Darboux. Ici, nous n'avons plus ce confort et c'est le lemme A.12 qui arrange les choses.

Par convergence uniforme de f_n vers f sur $[a, b]$, pour tout $\varepsilon > 0$, il existe un entier n_ε tel que

$$\forall n \geq n_\varepsilon, \forall x \in [a, b], \quad |f_n(x) - f(x)| < \frac{\varepsilon}{b-a}.$$

En appliquant le lemme A.12, on a alors avec le même n_ε ,

$$\forall n \geq n_\varepsilon, \forall E \subset [a, b], \quad |m_n(E) - m(E)| \leq \frac{\varepsilon}{b-a}, \quad |M_n(E) - M(E)| \leq \frac{\varepsilon}{b-a}. \quad (\text{A.22})$$

Soit $\Delta = \{x_0 = a < x_1 < \dots < x_j = b\}$ une subdivision quelconque de $[a, b]$. En appliquant (A.22) avec pour E chacun des intervalles $[x_{k-1}, x_k]$ de la subdivision, on vérifie immédiatement que :

$$\forall \Delta, \forall n \geq n_\varepsilon, \quad S_\Delta(f_n) - \varepsilon \leq S_\Delta(f) \leq S^\Delta(f) \leq S^\Delta(f_n) + \varepsilon. \quad (\text{A.23})$$

La fonction f_{n_ε} étant par hypothèse Riemann intégrable sur $[a, b]$, il existe une subdivision Δ_ε telle que

$$S_{\Delta_\varepsilon}(f_{n_\varepsilon}) > S^{\Delta_\varepsilon}(f_{n_\varepsilon}) - \varepsilon. \quad (\text{A.24})$$

En choisissant dans (A.23) $n = n_\varepsilon$ et $\Delta = \Delta_\varepsilon$ et en combinant l'encadrement ainsi obtenu avec l'inégalité (A.24), il vient :

$$S^{\Delta_\varepsilon}(f_{n_\varepsilon}) - 2\varepsilon < S_{\Delta_\varepsilon}(f) \leq S^{\Delta_\varepsilon}(f) \leq S^{\Delta_\varepsilon}(f_{n_\varepsilon}) + \varepsilon,$$

d'où l'on tire $0 \leq S^{\Delta_\varepsilon}(f) - S_{\Delta_\varepsilon}(f) < 3\varepsilon$, puis $0 \leq I^*(f) - I_*(f) < 3\varepsilon$. Par arbitrarité de ε , on en déduit $I_*(f) = I^*(f)$, ce qui établit la Riemann intégrabilité de f . \square

Définition A.14 (fonction réglée). *On dit que f est réglée sur $[a, b]$ si elle est limite uniforme sur $[a, b]$ d'une suite de fonctions en escalier.*

Corollaire A.15 (intégrabilité d'une fonction réglée). *Toute fonction réglée $[a, b] \rightarrow \mathbb{R}$ est Riemann intégrable sur $[a, b]$.*

Preuve. C'est une conséquence immédiate des propositions A.8 et A.13. \square

Pour finir cette section, nous donnons un exemple de fonction Riemann intégrable qui ne soit pas réglée (les fonctions monotones ou continues sont toutes réglées, exercice!) et un exemple de fonction bornée et borélienne qui ne soit pas Riemann intégrable.

Exemple A.16 (une fonction intégrable non réglée). Soit $E := \{2^{-k}; k \in \mathbb{N}^*\}$ et $f := \mathbf{1}_E$. La fonction f est bornée et Riemann intégrable sur $[0, 1]$, mais pas réglée. Vérifions ces deux affirmations. Soit $\Delta_n := \{0, 2^{-n}\} \cup \{2^{-k} \pm 2^{-2n}; 1 \leq k < n\}$. On voit immédiatement que pour tout $n \geq 2$:

$$0 = S_{\Delta_n}(f) \leq I_*(f) \leq I^*(f) \leq S^{\Delta_n}(f) = 2^{-n} + 2(n-1)2^{-2n}.$$

En faisant tendre n vers $+\infty$, on en déduit que $I_*(f) = I^*(f) = 0$, ce qui prouve que f est Riemann intégrable sur $[0, 1]$ et que $\int_0^1 f(x) dx = 0$. Notons au passage qu'on a ainsi un exemple de fonction f positive d'intégrale de Riemann nulle sur $[0, 1]$ sans que f soit identiquement nulle sur $[0, 1]$. Cette situation ne pourrait pas se produire avec une f continue (exercice).

Supposons que $f = \mathbf{1}_E$ soit réglée. Ceci implique que pour tout $\varepsilon > 0$, il existe une fonction en escalier g telle que $|f(x) - g(x)| < \varepsilon$ pour tout $x \in [0, 1]$. Prenons $\varepsilon = 1/3$, choisissons une telle g , notons $\Delta_0 = \{x_0 = 0 < x_1 < \dots < x_j = 1\}$ une subdivision associée à g (cf. la définition A.7) et c_1 la valeur constante de g sur $]0, x_1[$. Comme f prend au moins une fois la valeur 1 (en fait une infinité de fois) sur $]0, x_1[$, on a $|1 - c_1| < 1/3$ et de même f prenant au moins une fois la valeur 0 (en fait une infinité de fois) sur $]0, x_1[$, on a $|0 - c_1| < 1/3$. Ces deux inégalités sont incompatibles, donc f ne peut pas être réglée.

Exemple A.17 (une fonction bornée non intégrable). Soit $E := [0, 1] \cap \mathbb{Q}$ et $f := \mathbf{1}_E$. La fonction f est bornée et borélienne (comme indicatrice d'un ensemble borélien de \mathbb{R}), mais n'est pas Riemann intégrable sur $[0, 1]$. En effet en notant que dans tout intervalle ouvert non vide de \mathbb{R} il y a au moins un rationnel et un irrationnel, on vérifie facilement que pour toute subdivision Δ de $[0, 1]$,

$$S_\Delta(f) = 0 \quad \text{et} \quad S^\Delta(f) = 1.$$

On en déduit que $I_*(f) = 0$ et $I^*(f) = 1$, donc f n'est pas Riemann intégrable sur $[0, 1]$.

A.3 Propriétés de l'intégrale de Riemann

Cette section regroupe les propriétés générales de l'intégrale de Riemann, à l'exception de celles relatives à l'interversion limite intégrale. Nous étudions d'abord les propriétés relatives aux fonctions à intégrer (les intégrandes), autrement dit la structure de l'ensemble $\mathcal{R}[a, b]$. Nous verrons ensuite les propriétés concernant l'intervalle d'intégration.

A.3.1 Propriétés de l'ensemble $\mathcal{R}[a, b]$

Proposition A.18 (additivité). *Si f et g sont Riemann intégrables sur $[a, b]$, $f + g$ l'est aussi et*

$$\int_a^b (f + g)(x) dx = \int_a^b f(x) dx + \int_a^b g(x) dx. \quad (\text{A.25})$$

Preuve. Notons en préliminaire que si f et g sont bornées sur l'intervalle I , $f + g$ l'est aussi et on a

$$\inf_{x \in I} f(x) + \inf_{x \in I} g(x) \leq \inf_{x \in I} (f + g)(x), \quad \sup_{x \in I} (f + g)(x) \leq \sup_{x \in I} f(x) + \sup_{x \in I} g(x),$$

ces inégalités pouvant être strictes⁷.

Fixons $\varepsilon > 0$ quelconque. La Riemann intégrabilité de f et g nous fournissent des subdivisions Δ_1 et Δ_2 telles que $\int_a^b f(x) dx - \varepsilon < S_{\Delta_1}(f) \leq S^{\Delta_1}(f) < \int_a^b f(x) dx + \varepsilon$ et $\int_a^b g(x) dx - \varepsilon < S_{\Delta_2}(g) \leq S^{\Delta_2}(g) < \int_a^b g(x) dx + \varepsilon$. Avec leur raffinement commun $\Delta := \Delta_1 \cup \Delta_2$, on a ainsi :

$$\int_a^b f(x) dx - \varepsilon < S_{\Delta}(f) \leq S^{\Delta}(f) < \int_a^b f(x) dx + \varepsilon, \quad (\text{A.26})$$

$$\int_a^b g(x) dx - \varepsilon < S_{\Delta}(g) \leq S^{\Delta}(g) < \int_a^b g(x) dx + \varepsilon. \quad (\text{A.27})$$

Notons x_i , $0 \leq i \leq n$ les points de Δ , m_i , m'_i , m''_i , M_i , M'_i , M''_i les infima et suprema respectifs de $f + g$, f et g sur $[x_{i-1}, x_i]$ pour $i = 1, \dots, n$. Par la remarque faite en préliminaire, on a pour tout $i = 1, \dots, n$,

$$m'_i + m''_i \leq m_i \leq M_i \leq M'_i + M''_i.$$

On en déduit que

$$S_{\Delta}(f) + S_{\Delta}(g) \leq S_{\Delta}(f + g) \leq S^{\Delta}(f + g) \leq S^{\Delta}(f) + S^{\Delta}(g). \quad (\text{A.28})$$

En combinant (A.26), (A.27) et (A.28), on obtient

$$\int_a^b f(x) dx + \int_a^b g(x) dx - 2\varepsilon < S_{\Delta}(f + g) \leq S^{\Delta}(f + g) < \int_a^b f(x) dx + \int_a^b g(x) dx + 2\varepsilon,$$

d'où

$$\int_a^b f(x) dx + \int_a^b g(x) dx - 2\varepsilon < I_*(f + g) \leq I^*(f + g) < \int_a^b f(x) dx + \int_a^b g(x) dx + 2\varepsilon.$$

Ce dernier encadrement étant vérifié pour tout $\varepsilon > 0$, on peut y faire tendre ε vers 0 pour obtenir finalement

$$I_*(f + g) = I^*(f + g) = \int_a^b f(x) dx + \int_a^b g(x) dx,$$

ce qui établit l'intégrabilité de $f + g$ et (A.25). □

Proposition A.19. *Si f est intégrable sur $[a, b]$ et $c \in \mathbb{R}$ est une constante, cf est intégrable sur $[a, b]$ et*

$$\int_a^b cf(x) dx = c \int_a^b f(x) dx. \quad (\text{A.29})$$

7. Par exemple $f : x \mapsto x$ et $g : x \mapsto 1 - x$ sur $I = [0, 1]$.

Preuve. Le résultat est trivial si $c = 0$, puisqu'alors cf est la fonction identiquement nulle sur $[a, b]$, Riemann intégrable d'intégrale 0. Supposons $c \neq 0$. Si f est bornée sur l'ensemble E , cf est bornée sur E . On vérifie facilement que

$$\text{si } c > 0, \quad \inf_{x \in E} (cf)(x) = c \inf_{x \in E} f(x), \quad \sup_{x \in E} (cf)(x) = c \sup_{x \in E} f(x), \quad (\text{A.30})$$

et que

$$\text{si } c < 0, \quad \inf_{x \in E} (cf)(x) = c \sup_{x \in E} f(x), \quad \sup_{x \in E} (cf)(x) = c \inf_{x \in E} f(x). \quad (\text{A.31})$$

Pour $c > 0$, on déduit de (A.30) que pour toute subdivision Δ de $[a, b]$,

$$S_{\Delta}(cf) = cS_{\Delta}(f), \quad S^{\Delta}(cf) = cS^{\Delta}(f),$$

d'où $I_*(cf) = cI_*(f)$ et $I^*(cf) = cI^*(f)$. Ces deux égalités sont valables pour n'importe quelle fonction bornée f sur $[a, b]$. Ici f est de plus intégrable sur $[a, b]$, donc $I_*(f) = I^*(f) = \int_a^b f(x) dx$, d'où $I_*(cf) = I^*(cf) = c \int_a^b f(x) dx$, ce qui prouve l'intégrabilité de cf et établit (A.29). Pour $c < 0$, on déduit de (A.31) que pour toute subdivision Δ de $[a, b]$,

$$S_{\Delta}(cf) = cS^{\Delta}(f), \quad S^{\Delta}(cf) = cS_{\Delta}(f),$$

d'où $I_*(cf) = cI^*(f)$ et $I^*(cf) = cI_*(f)$. Par intégrabilité de f , on en déduit comme ci-dessus que $I_*(cf) = I^*(cf) = c \int_a^b f(x) dx$, ce qui complète la preuve. \square

On peut synthétiser les propositions A.18 et A.19 dans l'énoncé suivant.

Proposition A.20 (linéarité). *L'ensemble $\mathcal{R}[a, b]$ des applications $f : [a, b] \rightarrow \mathbb{R}$, Riemann intégrables sur $[a, b]$ est un \mathbb{R} -espace vectoriel et l'application $\Psi : \mathcal{R}[a, b] \rightarrow \mathbb{R}$, $f \mapsto \int_a^b f(x) dx$ est une forme linéaire sur cet espace.*

Proposition A.21 (croissance de l'intégrale). *L'intégrale de Riemann sur $[a, b]$ possède les trois propriétés suivantes relativement à la relation d'ordre partiel \leq définie sur $\mathcal{R}[a, b]$ par $f \leq g$ si $\forall x \in [a, b], f(x) \leq g(x)$.*

i) *Positivité :*

$$\text{si } f \in \mathcal{R}[a, b] \text{ et } f \geq 0 \text{ sur } [a, b], \quad \int_a^b f(x) dx \geq 0. \quad (\text{A.32})$$

ii) *Croissance :*

$$\text{si } f, g \in \mathcal{R}[a, b] \text{ et } f \leq g \text{ sur } [a, b], \quad \int_a^b f(x) dx \leq \int_a^b g(x) dx. \quad (\text{A.33})$$

iii) *Si $f \in \mathcal{R}[a, b]$, l'application $|f| : x \mapsto |f(x)|$ est elle aussi Riemann intégrable sur $[a, b]$ et*

$$\left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx. \quad (\text{A.34})$$

Preuve. Rappelons que lorsqu'on parle d'intégrale sur $[a, b]$, on suppose implicitement $a \leq b$. Dans le cas $a > b$, on aurait $\int_a^b f(x) dx \leq 0$ dans (A.32) et $\int_a^b g(x) dx \leq \int_a^b f(x) dx$ dans (A.33).

Pour prouver i), on remarque que si $f \geq 0$ sur $[a, b]$ et Δ est une subdivision (croissante) quelconque de $[a, b]$, on a $m_k \geq 0$ pour chaque intervalle $[x_{k-1}, x_k]$ de Δ , donc $S_\Delta(f) \geq 0$ et par conséquent $I^*(f) \geq I_*(f) \geq 0$. Ce raisonnement est valable pour toute f positive bornée sur $[a, b]$. Comme ici f est de plus Riemann intégrable sur $[a, b]$, $I_*(f) = I^*(f) = \int_a^b f(x) dx$ et (A.32) est vérifiée.

On vérifie ii) en notant que si f, g sont dans $\mathcal{R}[a, b]$, $h := g - f$ aussi (prop. A.20) et $h \geq 0$. En utilisant i) et la proposition A.20, on obtient

$$0 \leq \int_a^b h(x) dx = \int_a^b g(x) dx - \int_a^b f(x) dx,$$

ce qui nous donne (A.33).

Admettons un instant que l'intégrabilité de f implique celle de $|f|$. En appliquant ii) avec l'encadrement $-|f| \leq f \leq |f|$, il vient⁸ :

$$-\int_a^b |f(x)| dx \leq \int_a^b f(x) dx \leq \int_a^b |f(x)| dx,$$

ce qui équivaut à (A.34).

Il reste à montrer que $|f|$ hérite de l'intégrabilité de f . Pour toute subdivision $\Delta = \{x_0 = a < x_1 < \dots < x_n = b\}$, notons

$$m_k := \inf_{x \in [x_{k-1}, x_k]} f(x), \quad m'_k := \inf_{x \in [x_{k-1}, x_k]} |f(x)|,$$

$$M_k := \sup_{x \in [x_{k-1}, x_k]} f(x), \quad M'_k := \sup_{x \in [x_{k-1}, x_k]} |f(x)|.$$

Par le lemme A.22 ci-dessous, on a pour tout $k = 1, \dots, n$, $0 \leq M'_k - m'_k \leq M_k - m_k$, d'où

$$0 \leq I^*(|f|) - I_*(|f|) \leq S^\Delta(|f|) - S_\Delta(|f|) \leq S^\Delta(f) - S_\Delta(f)$$

Comme f est Riemann intégrable sur $[a, b]$, on peut choisir pour tout ε une subdivision Δ telle que $S^\Delta(f) - S_\Delta(f) < \varepsilon$ et l'encadrement ci-dessus appliqué à cette subdivision nous donne $0 \leq I^*(|f|) - I_*(|f|) < \varepsilon$, d'où $I^*(|f|) = I_*(|f|)$ par arbitrarité de ε . \square

Lemme A.22. *Si f est bornée sur $E \subset \mathbb{R}$, alors $|f|$ est bornée sur E et en notant $m := \inf_E f$, $m' := \inf_E |f|$, $M := \sup_E f$, $M' := \sup_E |f|$, on a $0 \leq M' - m' \leq M - m$, l'inégalité pouvant être stricte.*

Preuve. Si $M' - m' = 0$, c'est trivial⁹. Supposons désormais que $M' - m' > 0$. Alors pour tout ε tel que $0 < \varepsilon < (M' - m')/2$, on peut trouver $x_1, x_2 \in E$, dépendants de ε , tels que :

$$m' \leq |f(x_1)| < m' + \varepsilon < M' - \varepsilon < |f(x_2)| \leq M'. \quad (\text{A.35})$$

8. En utilisant aussi $\int_a^b -|f(x)| dx = -\int_a^b |f(x)| dx$ par linéarité.

9. Cette situation peut se produire, par exemple $E = [-1, 1]$ et $f(x) = \mathbf{1}_{[-1, 0]}(x) - \mathbf{1}_{[0, 1]}(x)$, on a dans ce cas $0 = M' - m' < M - m = 2$.

Par inégalité triangulaire, $|f(x_2)| - |f(x_1)| \leq |f(x_2) - f(x_1)|$. D'autre part $f(x_1)$ et $f(x_2)$ sont des réels du segment $[m, M]$, d'où $|f(x_2) - f(x_1)| \leq M - m$ et donc

$$|f(x_2)| - |f(x_1)| \leq M - m.$$

En combinant cette dernière inégalité avec (A.35), il vient :

$$M' - m' - 2\varepsilon \leq M - m.$$

Faisant tendre ε vers 0 dans cette inégalité *large*, on obtient $M' - m' \leq M - m$. \square

Remarque A.23. La Riemann intégrabilité de $|f|$ n'implique pas celle de f . Voici un contre exemple avec $[a, b] = [0, 1]$, $f(x) = 1$ si $x \in \mathbb{Q}$ et -1 si $x \notin \mathbb{Q}$. Alors, comme dans l'exemple A.17, f n'est pas Riemann intégrable sur $[0, 1]$, tandis que $|f|$ l'est (comme fonction constante).

Voici une conséquence immédiate de l'implication $f \in \mathcal{R}[a, b] \Rightarrow |f| \in \mathcal{R}[a, b]$ dans la proposition A.21 iii).

Corollaire A.24. Si f est Riemann intégrable sur $[a, b]$, les fonctions $f^+ := \max(f, 0)$ et $f^- := \max(-f, 0) = -\min(f, 0)$ le sont aussi. On a de plus

$$\int_a^b f(x) dx = \int_a^b f^+(x) dx - \int_a^b f^-(x) dx. \quad (\text{A.36})$$

Preuve. On commence par remarquer que

$$f^+ = \frac{1}{2}(|f| + f), \quad f^- = \frac{1}{2}(|f| - f).$$

Par la proposition A.21 iii) l'intégrabilité de f implique celle de $|f|$. L'ensemble $\mathcal{R}[a, b]$ étant un espace vectoriel (prop. A.20), on en déduit la Riemann intégrabilité sur $[a, b]$ de f^+ et f^- . La linéarité de l'intégrale et l'égalité $f = f^+ - f^-$ nous donnent (A.36). \square

Remarque A.25 (semi norme sur $\mathcal{R}[a, b]$). Grâce au point iii) de la proposition A.21, on peut définir l'application

$$N : \mathcal{R}[a, b] \rightarrow \mathbb{R}_+, \quad f \mapsto N(f) := \int_a^b |f(x)| dx.$$

Cette application est une *semi norme* sur $\mathcal{R}[a, b]$ car elle vérifie $N(cf) = |c|N(f)$ pour toute constante c et $N(f + g) \leq N(f) + N(g)$. Elle n'est pas une norme car on peut avoir $\int_a^b |f(x)| dx = 0$ sans que f soit la fonction nulle sur $[a, b]$, voir l'exemple A.16.

Proposition A.26 (Intégrabilité d'un produit). Si f et g sont Riemann intégrables sur $[a, b]$, leur produit fg l'est aussi.

Preuve. En écrivant $fg = (f^+ - f^-)(g^+ - g^-) = f^+g^+ - f^-g^+ - f^+g^- + f^-g^-$ et en utilisant le corollaire A.24 et la linéarité de l'intégrale de Riemann, on voit qu'il suffit de traiter le cas où f et g sont toutes deux *positives* sur $[a, b]$.

Fixons $\varepsilon > 0$ quelconque. La Riemann intégrabilité de f et g nous fournissent des subdivisions Δ_1 et Δ_2 telles que $\int_a^b f(x) dx - \varepsilon < S_{\Delta_1}(f) \leq S^{\Delta_1}(f) < \int_a^b f(x) dx + \varepsilon$ et $\int_a^b g(x) dx - \varepsilon < S_{\Delta_2}(g) \leq S^{\Delta_2}(g) < \int_a^b g(x) dx + \varepsilon$. Avec leur raffinement commun $\Delta := \Delta_1 \cup \Delta_2$, on a ainsi :

$$0 \leq S^\Delta(f) - S_\Delta(f) \leq 2\varepsilon, \quad (\text{A.37})$$

$$0 \leq S^\Delta(g) - S_\Delta(g) \leq 2\varepsilon. \quad (\text{A.38})$$

Rappelons ici que f et g Riemann intégrables sur $[a, b]$ sont *ipso facto* bornées sur cet intervalle, donc fg est aussi bornée sur $[a, b]$. Notons x_i , $0 \leq i \leq n$ les points de Δ , m_i , m'_i , m''_i , M_i , M'_i , M''_i les infima et suprema respectifs de fg , f et g sur $[x_{i-1}, x_i]$ pour $i = 1, \dots, n$. Par positivité on a pour tout $i = 1, \dots, n$,

$$\forall x \in [x_{i-1}, x_i], \quad 0 \leq m'_i m''_i \leq f(x)g(x) \leq M'_i M''_i,$$

d'où

$$0 \leq m'_i m''_i \leq m_i \leq M_i \leq M'_i M''_i.$$

Notons c un majorant commun sur $[a, b]$ aux fonctions positives bornées f et g . On alors on a pour tout $i = 1, \dots, n$,

$$\begin{aligned} 0 \leq M_i - m_i \leq M'_i M''_i - m'_i m''_i &= (M'_i - m'_i) M''_i + m'_i (M''_i - m''_i) \\ &\leq c(M'_i - m'_i) + c(M''_i - m''_i). \end{aligned}$$

En reportant cette majoration dans le calcul de $S^\Delta(fg) - S_\Delta(fg)$ et en tenant compte de (A.37) et (A.38), on obtient

$$0 \leq S^\Delta(fg) - S_\Delta(fg) \leq 4c\varepsilon.$$

Comme ε était arbitraire, on en déduit la Riemann intégrabilité de fg . \square

Contrairement à ce qui se passe pour la Riemann intégrabilité d'une somme $f + g$, il n'y a pas de formule permettant de calculer $\int_a^b f(x)g(x) dx$ en fonction de $\int_a^b f(x) dx$ et $\int_a^b g(x) dx$. La formule « $\int_a^b f(x)g(x) dx = \int_a^b f(x) dx \int_a^b g(x) dx$ » est *grossièrement fausse*. Voici un contre exemple élémentaire avec des fonctions en escalier. Prenons $a = 0$, $b = 2$, $f = \mathbf{1}_{[0,1]}$, $g = \mathbf{1}_{[1,2]}$. Alors fg est la fonction nulle sur $[0, 2]$ et donc $\int_0^2 f(x)g(x) dx = 0$, alors que $\int_0^2 f(x) dx \int_0^2 g(x) dx = 1$ parce que $\int_0^2 f(x) dx$ et $\int_0^2 g(x) dx$ valent chacune 1.

Proposition A.27 (inégalité de Cauchy-Schwarz dans $\mathcal{R}[a, b]$). *Si f et g sont Riemann intégrables sur $[a, b]$, on a l'inégalité*

$$\left| \int_a^b f(x)g(x) dx \right| \leq \left\{ \int_a^b f(x)^2 dx \right\}^{1/2} \left\{ \int_a^b g(x)^2 dx \right\}^{1/2}. \quad (\text{A.39})$$

Preuve. Soit t un réel quelconque. Par les propositions A.20 et A.26, les fonctions f^2 , g^2 , fg et $(tf + g)^2$ héritent de la Riemann intégrabilité de f et g . Posons

$$P(t) := \int_a^b (tf(x) + g(x))^2 dx.$$

Il est clair que $P(t)$ est positif ou nul pour tout t réel. Or en développant le carré $(tf + g)^2$ et en utilisant la linéarité de l'intégrale, on obtient

$$P(t) = \left\{ \int_a^b f(x)^2 dx \right\} t^2 + 2 \left\{ \int_a^b f(x)g(x) dx \right\} t + \int_a^b g(x)^2 dx.$$

On reconnaît là un trinôme du second degré $At^2 + Bt + C$ dont les coefficients A, B, C sont des intégrales. Ce trinôme ne peut avoir de signe constant, celui de $A = \int_a^b f(x)^2 dx \geq 0$, que si son discriminant $\Delta = B^2 - 4AC$ est négatif ou nul. Remplaçant A, B et C par leurs expressions sous forme d'intégrales, on en déduit (A.39). \square

A.3.2 Propriétés relatives à l'intervalle d'intégration

L'intégrale de Riemann se laisse volontiers découper en morceaux. Voici les énoncés précis dont la vérification est laissée au lecteur.

Proposition A.28 (additivité relative aux intervalles). *Soit $f : [a, b] \rightarrow \mathbb{R}$ et $c \in]a, b[$. Pour que f soit Riemann intégrable sur $[a, b]$, il faut et il suffit qu'elle soit Riemann intégrable sur $[a, c]$ et sur $[c, b]$. On a alors*

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx. \quad (\text{A.40})$$

En combinant la proposition A.28 avec la définition A.2, on obtient la formule classique suivante.

Proposition A.29 (relation de Chasles). *Pour tous réels a, b, c , on a*

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx,$$

pourvu que f soit Riemann intégrable sur $[\min(a, b, c), \max(a, b, c)]$.

Une autre application de l'additivité relative aux intervalles est la généralisation de la formule (A.10) pour le calcul de l'aire du domaine H délimité par le graphe de f , l'axe des abscisses et les deux droites verticales d'équations $x = a$ et $x = b$, voir figure A.6. Plus précisément, H est défini par

$$H := \{(x, y) \in \mathbb{R}^2; a \leq x \leq b \text{ et } -f^-(x) \leq y \leq f^+(x)\}, \quad (\text{A.41})$$

en notant que $-f^-(x) = f(x)$ ou 0 selon que $f(x) < 0$ ou non et que $f^+(x) = f(x)$ ou 0 selon que $f(x) > 0$ ou non. En combinant la proposition A.5, la remarque A.6, l'additivité relative aux intervalles de l'intégrale de Riemann et l'additivité finie de λ_2 , on obtient le résultat suivant pour les fonctions n'ayant qu'un nombre fini de changements de signe sur $[a, b]$.

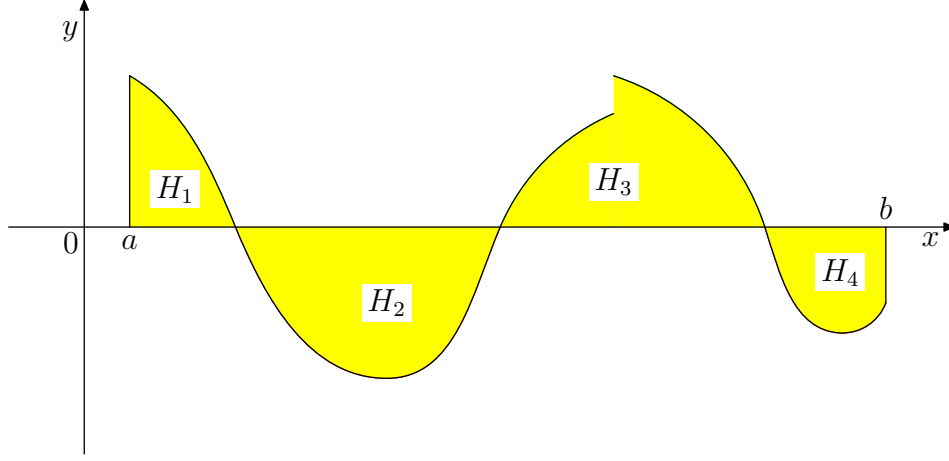


FIG. A.6 – Domaine H délimité par f entre a et b

Proposition A.30. Soit f une fonction borélienne sur $[a, b]$ et Riemann intégrable sur $[a, b]$. On suppose qu'il existe une subdivision $x_0 = a < x_1 < \dots < x_n = b$ telle que le signe de f soit constant sur chacun des $[x_{i-1}, x_i]$, $1 \leq i \leq n$. Alors l'aire du domaine H défini par (A.41) est donnée par

$$\lambda_2(H) = \int_a^b |f(x)| \, dx. \quad (\text{A.42})$$

Dans cet énoncé, « signe constant » s'entend au sens large : ou bien $f(x) \geq 0$ pour tout $x \in [x_{i-1}, x_i]$ ou bien $f(x) \leq 0$ pour tout $x \in [x_{i-1}, x_i]$. L'hypothèse f borélienne assure que sa restriction à chacun des $[x_{i-1}, x_i]$ est encore borélienne (pour les tribus adéquates) et donc que chaque $H_i := \{(x, y) \in \mathbb{R}^2; x_{i-1} \leq x \leq x_i \text{ et } -f^-(x) \leq y \leq f^+(x)\}$ est un borélien de \mathbb{R}^2 (admis). Bien entendu en écrivant cet énoncé, on a en tête le cas où le signe de f change à la traversée de chaque x_i , $0 < i < n$, mais la formule (A.42) reste évidemment vraie sans cette hypothèse. Si f a le même signe sur deux intervalles consécutifs, on peut trouver une subdivision « plus économique » en les fusionnant.

Nous pouvons maintenant donner une interprétation géométrique de l'intégrale de Riemann, au moins pour les fonctions f vérifiant les hypothèses de la proposition A.30. Pour cela on appelle « aire algébrique », la somme des $\lambda_2(H_i)$, *chacun étant compté avec le signe de f sur l'intervalle correspondant*. L'aire algébrique du domaine H représenté à la figure A.6 vaut ainsi $\lambda_2(H_1) - \lambda_2(H_2) + \lambda_2(H_3) - \lambda_2(H_4)$. Plus formellement, posons

$$s_i := \begin{cases} +1 & \text{si } f(t) > 0 \text{ pour au moins un } t \in]x_{i-1}, x_i[, \\ -1 & \text{si } f(t) < 0 \text{ pour au moins un } t \in]x_{i-1}, x_i[, \\ 0 & \text{si } f(t) = 0 \text{ pour tout } t \in]x_{i-1}, x_i[. \end{cases}$$

L'argumentation esquissée pour la proposition A.30 nous donne

$$\text{aire algébrique}(H) := \sum_{i=1}^n s_i \lambda_2(H_i) = \int_a^b f(x) \, dx. \quad (\text{A.43})$$

Nous regroupons dans le théorème suivant les propriétés de « l'intégrale indéfinie », c'est à dire de la fonction $x \mapsto \int_a^x f(t) dt$.

Théorème A.31. *Soit f Riemann intégrable sur $[a, b]$. Alors elle est aussi Riemann intégrable sur $[a, x]$ pour tout $x \in [a, b]$, ce qui permet de définir l'application $F : [a, b] \rightarrow \mathbb{R}$ par*

$$F(x) := \int_a^x f(t) dt.$$

- i) F est continue sur $[a, b]$ et même lipschitzienne.
- ii) Si f a une limite ℓ au point c de $[a, b]$ (resp. une limite à droite, resp. à gauche), F est dérivable au point c (resp. à gauche, resp. à droite) et $F'(c) = \ell$.
- iii) Si f est continue sur $[a, b]$, F est dérivable sur $[a, b]$ et a pour fonction dérivée f . C'est l'unique primitive de f sur $[a, b]$ qui s'annule au point a .

Preuve de i). Notons $C := \sup_{t \in [a, b]} |f(t)|$. En utilisant la relation de Chasles et la proposition A.21, on a pour tous $a \leq x \leq y \leq b$,

$$|F(y) - F(x)| = \left| \int_x^y f(t) dt \right| \leq \int_x^y |f(t)| dt \leq \int_x^y C dt = C|y - x|.$$

Ceci montre que F est lipschitzienne de rapport C sur $[a, b]$, donc *a fortiori* continue. \square

Preuve de ii). Commençons par noter que pour tout $x \neq c$ dans $[a, b]$,

$$\frac{F(x) - F(c)}{x - c} = \frac{1}{x - c} \int_c^x f(t) dt, \quad \text{et} \quad \ell = \frac{1}{x - c} \int_c^x \ell dt. \quad (\text{A.44})$$

Comme f a pour limite ℓ au point c , on a :

$$\forall \varepsilon > 0, \exists \delta > 0, 0 \leq |t - c| \leq |x - c| < \delta \Rightarrow |f(t) - \ell| < \varepsilon. \quad (\text{A.45})$$

En combinant (A.44) et (A.45), on voit que pour tout $x \in [a, b]$ vérifiant $|x - c| < \delta$,

$$\left| \frac{F(x) - F(c)}{x - c} - \ell \right| = \left| \frac{1}{x - c} \int_c^x (f(t) - \ell) dt \right| \leq \frac{1}{|x - c|} \int_c^x |f(t) - \ell| dt \leq \varepsilon,$$

ce qui montre que F est dérivable au point c , de nombre dérivé $F'(c) = \ell$. L'adaptation au cas d'une limite à droite ou à gauche (avec dérivée à droite ou à gauche) est immédiate. \square

Preuve de iii). Si f est continue sur $[a, b]$, elle a pour limite $f(c)$ en tout point c de $[a, b]$ et donc d'après ii), F est dérivable sur $[a, b]$ et $F'(c) = f(c)$. Cette dernière égalité ayant lieu maintenant pour tout $c \in [a, b]$, on a $F' = f$, autrement dit F est une primitive de f sur $[a, b]$. On sait que toutes les primitives de f sur l'intervalle $[a, b]$ diffèrent entre elles d'une constante¹⁰ Il y en a donc une seule qui s'annule au point a , c'est F . \square

10. C'est une conséquence de la formule des accroissements finis (cf. p.222) : si une fonction continue sur $[a, b]$ a une dérivée nulle sur $]a, b[$, elle est constante sur $[a, b]$ et on applique ceci à la différence de deux primitives quelconques de f sur $[a, b]$.

Proposition A.32 (changement de variable).

i) *Translation.* Soit $c \in \mathbb{R}$. Pour toute f Riemann intégrable sur $[a+c, b+c]$, l'application $g : [a, b] \rightarrow \mathbb{R}$, $x \mapsto f(x+c)$ est Riemann intégrable sur $[a, b]$ et

$$\int_a^b f(x+c) dx = \int_{a+c}^{b+c} f(y) dy. \quad (\text{A.46})$$

ii) *Changement d'échelle.* Soit $c \in \mathbb{R}^*$. Pour toute f Riemann intégrable sur l'intervalle fermé d'extrémités¹¹ ac et bc , l'application $h : [a, b] \rightarrow \mathbb{R}$, $x \mapsto f(cx)$ est Riemann intégrable sur $[a, b]$ et

$$\int_a^b f(cx) dx = \frac{1}{c} \int_{ac}^{bc} f(y) dy. \quad (\text{A.47})$$

iii) *Classique.* Soit $\varphi : [a, b] \rightarrow \mathbb{R}$, une fonction ayant une dérivée continue sur $[a, b]$ (autrement dit $\varphi \in C^1[a, b]$). Pour toute fonction f continue sur l'intervalle fermé borné $\varphi([a, b])$, on a

$$\int_a^b f(\varphi(x)) \varphi'(x) dx = \int_{\varphi(a)}^{\varphi(b)} f(y) dy. \quad (\text{A.48})$$

Bien sûr i) et ii) sont contenus dans iii) si f est continue, mais l'intérêt de ces deux énoncés séparés est qu'ils sont valables avec *n'importe quelle* fonction f Riemann intégrable.

Preuve de i). À chaque subdivision $\Delta = \{x_0 = a < x_1 < \dots < x_n = b\}$ de $[a, b]$, associons la subdivision translatée $\Delta' = \{y_0 = a+c < \dots < y_k = x_k+c < \dots < y_n = b+c\}$. Comme f est bornée sur $[a+c, b+c]$, g est bornée sur $[a, b]$ avec mêmes bornes. De plus en notant m_k, m'_k les infima respectifs de f sur $[x_{k-1}, x_k]$ et de g sur $[y_{k-1}, y_k]$, et en définissant de même M_k et M'_k pour les suprema, on $m_k = m'_k$ et $M_k = M'_k$ pour $k = 1, \dots, n$. Par conséquent $S_\Delta(g) = S_{\Delta'}(f)$ et $S^\Delta(g) = S^{\Delta'}(f)$, pour toute subdivision Δ de $[a, b]$. Comme la transformation $\Delta \mapsto \Delta'$ réalise une bijection entre l'ensemble des subdivisions de $[a, b]$ et l'ensemble des subdivisions de $[a+c, b+c]$, on en déduit que

$$I^*(g) = \inf_{\Delta} S^\Delta(g) = \inf_{\Delta'} S^{\Delta'}(f) = I^*(f) \quad (\text{A.49})$$

et de même

$$I_*(g) = \sup_{\Delta} S_\Delta(g) = \sup_{\Delta'} S_{\Delta'}(f) = I_*(f). \quad (\text{A.50})$$

Comme on sait de plus que f est Riemann intégrable sur $[a+c, b+c]$, on a $I^*(f) = I_*(f)$. Compte-tenu de (A.49)–(A.50), on en déduit $I^*(g) = I_*(g) = I^*(f) = \int_{a+c}^{b+c} f(y) dy$, ce qui nous donne la Riemann intégrabilité de g sur $[a, b]$ et l'égalité (A.46). \square

11. Il s'agit de $[ac, bc]$ si $c > 0$ et de $[bc, ac]$ si $c < 0$.

Preuve de ii). La méthode étant essentiellement la même que pour le i), nous nous contenterons d'indiquer les adaptations nécessaires. Si $c > 0$, on associe à $\Delta = \{x_0 = a < x_1 < \dots < x_n = b\}$ la subdivision Δ' dont les points sont les $y_k = cx_k$. Alors Δ' est une subdivision croissante de $[ac, bc]$ et comme $y_k - y_{k-1} = c(x_k - x_{k-1})$, on voit que $S_\Delta(h) = \frac{1}{c}S_{\Delta'}(f)$ et $S^\Delta(h) = \frac{1}{c}S^{\Delta'}(f)$. On en déduit comme ci-dessus l'intégrabilité de h et (A.47).

Si $c < 0$, $bc \leq ac$ et on prend pour subdivision croissante Δ' associée à Δ la subdivision de $[bc, ac]$ ayant pour points les $y_k = cx_{n-k}$. Alors pour $k = 1, \dots, n$, $y_k - y_{k-1} = c(x_{n-k} - x_{n-k+1}) = -c(x_{n-k+1} - x_{n-k})$. On en déduit que $S_\Delta(h) = \frac{-1}{c}S_{\Delta'}(f)$ et $S^\Delta(h) = \frac{-1}{c}S^{\Delta'}(f)$ puis, que h est Riemann intégrable et

$$\int_a^b h(x) dx = \frac{-1}{c} \int_{bc}^{ac} f(y) dy = \frac{1}{c} \int_{ac}^{bc} f(y) dy,$$

en utilisant la définition A.2. □

Preuve de iii). D'abord, φ étant continue, l'image J de $[a, b]$ par φ est un *intervalle* (théorème des valeurs intermédiaires) et comme $[a, b]$ est *compact*, $J = \varphi([a, b])$ est aussi compact (l'image d'un compact par une application continue est un compact). Ainsi J est un intervalle compact, donc un intervalle fermé borné. Cet intervalle contient évidemment l'intervalle I d'extrémités $\varphi(a)$ et $\varphi(b)$ (pas forcément dans cet ordre), l'inclusion pouvant être stricte. La fonction f étant continue sur J l'est aussi par restriction sur I et l'intégrale au second membre de (A.48) est donc bien définie. L'intégrale du premier membre l'est tout autant puisque $(f \circ \varphi)\varphi'$ est continue sur $[a, b]$.

Introduisons les fonctions F, G, H suivantes :

$$F : J \rightarrow \mathbb{R}, s \mapsto \int_{\varphi(a)}^s f(y) dy, \quad G : [a, b] \rightarrow \mathbb{R}, s \mapsto \int_a^s f(\varphi(x))\varphi'(x) dx, \quad H := F \circ \varphi.$$

Par le théorème A.31, F est dérivable sur J et $F' = f$. De même G est dérivable sur $[a, b]$ et $G' = (f \circ \varphi)\varphi'$. D'autre part H est dérivable comme fonction composée et

$$H' = (F' \circ \varphi)\varphi' = (f \circ \varphi)\varphi' = G'.$$

Les fonctions H et G ont ainsi même dérivée sur $[a, b]$, leur différence est donc constante sur $[a, b]$. Or $H(a) = 0$ et $G(a) = 0$, donc $H = G$. En particulier, $H(b) = G(b)$, ce qui établit (A.48). □

Voici maintenant une version moins « classique » du changement de variable.

Proposition A.33 (changement de variable). *Si φ est C^1 et monotone sur $[a, b]$ et f est Riemann intégrable sur $\varphi([a, b])$,*

$$\int_a^b f(\varphi(x))\varphi'(x) dx = \int_{\varphi(a)}^{\varphi(b)} f(y) dy. \quad (\text{A.51})$$

La différence avec le changement de variable « classique » (Proposition A.32 iii) est qu'on ne suppose plus f continue. On paie cette plus grande généralité pour f avec une hypothèse plus restrictive de monotonie pour φ . Comme cas particuliers, on retrouve les points i) et ii) de la proposition A.32.

Preuve. Notons d'abord que l'existence de l'intégrale au second membre de (A.51) fait partie des hypothèses, mais que celle de l'intégrale du premier membre reste à établir. Pour alléger les écritures, convenons que les notations $I, I^*, I_*, S_\Delta, S^\Delta$, sont relatives aux bornes $\varphi(a)$ et $\varphi(b)$ lorsqu'elles s'appliquent à f et relatives aux bornes a et b lorsqu'elles s'appliquent à $(f \circ \varphi)\varphi'$. Nous détaillons la preuve dans le cas où φ est croissante, laissant au lecteur l'adaptation au cas où f décroît.

Il suffit de démontrer (A.51) pour f positive sur $[\varphi(a), \varphi(b)]$. En effet (A.51) est clairement vérifiée si on remplace f par une fonction constante c et comme f est bornée inférieurement sur $[\varphi(a), \varphi(b)]$ car Riemann intégrable, on peut toujours trouver c telle que $g = f + c$ soit une fonction Riemann intégrable positive. Si on arrive à démontrer (A.51) pour g , alors on en déduit facilement la même égalité pour f par soustraction. Désormais, nous supposons donc sans perte de généralité que f est positive.

Fixons un $\varepsilon > 0$ arbitraire. Comme f est Riemann intégrable sur $[\varphi(a), \varphi(b)]$, il existe une subdivision

$$\Delta = \{y_0 = \varphi(a) < y_1 < \dots < y_n = \varphi(b)\}$$

telle que

$$I(f) - \varepsilon < S_\Delta(f) = \sum_{k=1}^n (y_k - y_{k-1})m_k \leq S^\Delta(f) = \sum_{k=1}^n (y_k - y_{k-1})M_k < I(f) + \varepsilon.$$

Rappelons que $m_k = \inf\{f(y); y \in [y_{k-1}, y_k]\}$ et $M_k = \sup\{f(y); y \in [y_{k-1}, y_k]\}$. Comme φ est croissante et continue, l'image de $[a, b]$ par φ est le segment $[\varphi(a), \varphi(b)]$ et φ est une surjection¹² de $[a, b]$ sur $[\varphi(a), \varphi(b)]$. On peut donc trouver pour chaque y_k au moins un antécédent x_k par φ et quelque soit le choix de cet antécédent, on aura nécessairement $x_{k-1} < x_k$ car y_{k-1} est *strictement* inférieur à y_k et φ est croissante au sens large.

La fonction φ' étant continue sur le compact $[a, b]$ est *uniformément* continue sur cet intervalle. On peut donc subdiviser chaque $[x_{k-1}, x_k]$ en $x_{k,0} = x_{k-1} < \dots < x_{k,j_k} = x_k$ de sorte qu'en notant pour $j = 0, \dots, j_k$, $m'_{k,j} = \inf\{\varphi'(x); x \in [x_{k,j-1}, x_{k,j}]\}$ et $M'_{k,j} = \sup\{\varphi'(x); x \in [x_{k,j-1}, x_{k,j}]\}$, on ait :

$$M'_{k,j} - m'_{k,j} < \frac{\varepsilon}{M(b-a)} =: \varepsilon', \quad \text{où } M := \sup_{[\varphi(a), \varphi(b)]} f.$$

12. Elle n'est pas forcément injective car on n'a pas supposé la croissance stricte.

En retenant de cette inégalité que $m'_{k,j} > M'_{k,j} - \varepsilon'$, on en déduit la minoration suivante :

$$\begin{aligned} y_k - y_{k-1} &= \int_{x_{k-1}}^{x_k} \varphi'(x) \, dx = \sum_{j=1}^{j_k} \int_{x_{k,j-1}}^{x_{k,j}} \varphi'(x) \, dx \\ &\geq \sum_{j=1}^{j_k} m'_{k,j} (x_{k,j} - x_{k,j-1}) \\ &\geq \sum_{j=1}^{j_k} (M'_{k,j} - \varepsilon') (x_{k,j} - x_{k,j-1}), \end{aligned}$$

d'où en remarquant que M_k hérite de la positivité de f ,

$$M_k(y_k - y_{k-1}) \geq \sum_{j=1}^{j_k} M_k M'_{k,j} (x_{k,j} - x_{k,j-1}) - \varepsilon' M_k (x_k - x_{k-1}).$$

En notant $M_{k,j}$ le supremum de $f \circ \varphi$ sur $[x_{k,j-1}, x_{k,j}]$, on a

$$M_k M'_{k,j} \geq M_{k,j} M'_{k,j} \geq \sup_{[x_{k,j-1}, x_{k,j}]} (f \circ \varphi) \varphi'.$$

Finalement en désignant par Δ' la subdivision de $[a, b]$ formée par tous les $x_{k,j}$, $k = 1, \dots, n$, $j = 0, \dots, j_k$, on aboutit à

$$S^\Delta(f) \geq S^{\Delta'}((f \circ \varphi)\varphi') - \varepsilon \geq I^*((f \circ \varphi)\varphi') - \varepsilon.$$

Comme ce dernier minorant ne dépend pas de Δ , on en déduit que

$$I(f) \geq I^*((f \circ \varphi)\varphi') - \varepsilon.$$

Enfin, comme ε est arbitraire, il vient $I(f) \geq I^*((f \circ \varphi)\varphi')$.

On laisse le soin au lecteur d'adapter l'argumentation ci-dessus pour montrer de même que $I(f) \leq I_*((f \circ \varphi)\varphi') + \varepsilon$, puis $I(f) \leq I_*((f \circ \varphi)\varphi')$. En rassemblant ces inégalités et en se souvenant que $I_* \leq I^*$, on en déduit que

$$I(f) = I^*((f \circ \varphi)\varphi') = I_*((f \circ \varphi)\varphi'),$$

ce qui prouve (A.51) en établissant au passage la Riemann intégrabilité de $(f \circ \varphi)\varphi'$. \square

A.4 Intersion limite intégrale

Théorème A.34. Soit $(f_n)_{n \geq 1}$ une suite de fonctions Riemann intégrables sur $[a, b]$. On suppose que cette suite converge uniformément vers f sur $[a, b]$. Alors f est Riemann intégrable sur $[a, b]$ et on a

$$\int_a^b f_n(t) \, dt \xrightarrow{n \rightarrow +\infty} \int_a^b f(t) \, dt. \quad (\text{A.52})$$

Preuve. Dans la preuve de ce théorème la partie difficile est d'établir la Riemann intégrabilité de f , mais nous l'avons déjà vue par la proposition A.13. Une fois que l'on sait que f est Riemann intégrable, on peut écrire :

$$\forall n \in \mathbb{N}^*, \quad \left| \int_a^b f(t) dt - \int_a^b f_n(t) dt \right| \leq \int_a^b |f(t) - f_n(t)| dt. \quad (\text{A.53})$$

Par convergence uniforme, on a pour tout $\varepsilon > 0$ un entier n_ε tel que

$$\forall n \geq n_\varepsilon, \quad \forall t \in [a, b], \quad |f(t) - f_n(t)| = |f - f_n|(t) < \frac{\varepsilon}{b-a}.$$

En reportant cette inégalité dans (A.53), en utilisant la linéarité de l'intégrale et la proposition A.21 iii), on en déduit que pour tout $n \geq n_\varepsilon$,

$$\left| \int_a^b f(t) dt - \int_a^b f_n(t) dt \right| = \left| \int_a^b (f - f_n)(t) dt \right| \leq \int_a^b |f - f_n|(t) dt \leq \int_a^b \frac{\varepsilon}{b-a} dt = \varepsilon.$$

Ceci étant valable pour tout $\varepsilon > 0$, la convergence (A.52) est établie. \square

Théorème A.35. Soit $(f_n)_{n \geq 1}$ une suite de fonctions toutes décroissantes sur $[a, b]$. On suppose de plus que :

$$\forall t \in [a, b], \quad f_n(t) \xrightarrow{n \rightarrow +\infty} f(t) \in \mathbb{R}.$$

Alors la fonction f ainsi définie est Riemann intégrable sur $[a, b]$ et

$$\int_a^b f_n(t) dt \xrightarrow{n \rightarrow +\infty} \int_a^b f(t) dt.$$

Il est clair que le théorème reste vrai si toutes les f_n d'indice $n \geq n_0$ sont décroissantes ou si elles sont croissantes pour tout $n \geq n_0$. Attention à ne pas confondre « suite de fonctions décroissantes sur $[a, b]$ » avec « suite décroissante de fonctions définies sur $[a, b]$ ». Ici on est dans le premier cas et on ne suppose rien sur le sens de variation des suites de réels $(f_n(t))_{n \geq 1}$, $t \in [a, b]$.

Preuve. Remarquons d'abord que la fonction limite f est décroissante sur $[a, b]$ comme limite d'une suite de fonctions décroissantes puisque le passage à la limite conserve les inégalités larges. Par la proposition A.9, f est donc elle aussi Riemann intégrable.

Cette Riemann intégrabilité de f nous assure de l'existence pour $\varepsilon > 0$ arbitraire fixé d'une subdivision $\Delta = \{t_0 = a < t_1 < \dots < t_j = b\}$ telle que

$$S^\Delta(f) - \varepsilon < \int_a^b f(t) dt < S_\Delta(f) + \varepsilon. \quad (\text{A.54})$$

Notons que par *décroissance* de f et de f_n , ces fonctions *atteignent* sur $[t_{k-1}, t_k]$ leur supremum au point t_{k-1} et leur infimum¹³ au point t_k . On peut donc expliciter comme suit pour f et f_n les sommes de Darboux supérieures

$$S^\Delta(f) = \sum_{k=1}^j f(t_{k-1})(t_k - t_{k-1}), \quad S^\Delta(f_n) = \sum_{k=1}^j f_n(t_{k-1})(t_k - t_{k-1}),$$

13. Donc le supremum et l'infimum sont ici respectivement un maximum et un minimum.

et les sommes de Darboux inférieures :

$$S_{\Delta}(f) = \sum_{k=1}^j f(t_k)(t_k - t_{k-1}), \quad S_{\Delta}(f_n) = \sum_{k=1}^j f_n(t_k)(t_k - t_{k-1}).$$

Par convergence simple de f_n vers f sur $[a, b]$, on peut trouver n_{ε} tel que

$$\forall n \geq n_{\varepsilon}, \forall k = 0, 1, \dots, j, \quad |f(t_k) - f_n(t_k)| < \frac{\varepsilon}{b-a}. \quad (\text{A.55})$$

En effet on n'a qu'un nombre fini $j+1$ d'écarts à contrôler (rappelons que pour l'instant ε est fixé et donc j aussi) et par convergence de $f_n(t_k)$ vers $f(t_k)$, on trouve pour chaque $k = 0, 1, \dots, j$ un rang $n_{\varepsilon,k}$ à partir duquel l'inégalité ci-dessus est toujours réalisée. On prend alors $n_{\varepsilon} = \max_{0 \leq k \leq j} n_{\varepsilon,k}$. En utilisant (A.55) et la positivité des $(t_k - t_{k-1})$, on en déduit immédiatement que

$$S^{\Delta}(f) > S^{\Delta}(f_n) - \varepsilon, \quad S_{\Delta}(f) < S_{\Delta}(f_n) + \varepsilon. \quad (\text{A.56})$$

En reportant ces inégalités dans (A.54), on obtient

$$S^{\Delta}(f_n) - 2\varepsilon < \int_a^b f(t) dt < S_{\Delta}(f_n) + 2\varepsilon,$$

puis

$$\int_a^b f_n(t) dt - 2\varepsilon < \int_a^b f(t) dt < \int_a^b f_n(t) dt + 2\varepsilon.$$

Autrement dit, nous avons trouvé un entier n_{ε} tel que

$$\forall n \geq n_{\varepsilon}, \quad \left| \int_a^b f_n(t) dt - \int_a^b f(t) dt \right| < 2\varepsilon.$$

Comme ε était arbitraire, ceci exprime précisément la convergence quand n tend vers l'infini, de $\int_a^b f_n(t) dt$ vers $\int_a^b f(t) dt$. \square

Annexe B

Intégrale généralisée

L'intégrale de Riemann étudiée dans l'annexe A concerne des fonctions f définies en tout point d'un intervalle *fermé borné* $[a, b]$ et *bornées* sur cet intervalle. Il est utile de généraliser cette notion au cas de fonctions définies sur un intervalle quelconque, sauf peut-être en un nombre fini de points, et pas forcément bornées. Dans le cadre de ce cours, les principales applications de cette notion d'intégrale généralisée concernent les lois à densité, l'espérance et les moments de variables aléatoires.

B.1 Construction

Commençons par examiner cette généralisation de l'intégrale de Riemann sur quelques exemples simples.

Exemple B.1. *Peut-on définir $\int_0^{+\infty} e^{-x} dx$?*

Ici l'intervalle d'intégration est $[0, +\infty[$ et l'intégrande $f : x \mapsto e^{-x}$ est définie et continue sur cet intervalle, donc en particulier Riemann intégrable sur tout sous-intervalle fermé borné de la forme $[0, b]$. L'intégrale $\int_0^b e^{-x} dx$ a bien un sens. Elle se calcule d'ailleurs immédiatement par primitivation de f et vaut $[-e^{-x}]_0^b = 1 - e^{-b}$. Cette valeur a pour limite 1 quand b tend vers l'infini et il est naturel de prendre cette limite comme définition de l'intégrale généralisée $\int_0^{+\infty} e^{-x} dx$. L'interprétation géométrique de ce résultat est que l'aire de l'hypographe $H := \{(x, y) \in \mathbb{R}^2; x \in \mathbb{R}_+, 0 \leq y \leq e^{-x}\}$ vaut 1, cf. figure B.1. Pour le justifier, on remarque que H est la réunion de la suite croissante pour l'inclusion $(H_n)_{n \geq 1}$, où H_n est l'hypographe de f entre 0 et n . Par continuité séquentielle croissante de la mesure de Lebesgue λ_2 , cf. proposition 2.16 page 63, on a $\lambda_2(H) = \lim_{n \rightarrow +\infty} \lambda_2(H_n) = \lim_{n \rightarrow +\infty} \int_0^n e^{-x} dx = 1$.

Exemple B.2. *Peut-on définir $\int_0^{+\infty} \frac{x}{1+x^2} dx$?*

Comme dans l'exemple B.1, l'intégrale sur $[0, b]$ a un sens pour tout $b \in \mathbb{R}_+$. Elle se calcule par changement de variable $u = x^2$ et primitivation :

$$\int_0^b \frac{x}{1+x^2} dx = \frac{1}{2} \int_0^{b^2} \frac{du}{1+u} = \frac{1}{2} [\ln(1+u)]_0^{b^2} = \frac{1}{2} \ln(1+b^2).$$

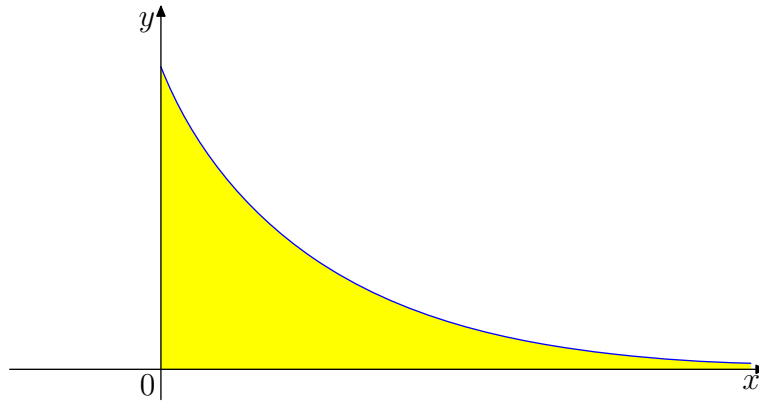


FIG. B.1 – Hypographe H de $f : x \mapsto e^{-x}$ entre 0 et $+\infty$

Cette quantité tend vers $+\infty$ lorsque b tend vers $+\infty$. On conviendra donc que

$$\int_0^{+\infty} \frac{x}{1+x^2} dx = +\infty.$$

L'interprétation géométrique est que l'aire de l'hypographe de $x \mapsto x(1+x^2)^{-1}$ entre 0 et $+\infty$ est infinie. La justification repose sur la continuité séquentielle croissante de λ_2 comme pour l'exemple B.1.

Exemple B.3. *Peut-on définir l'intégrale $\int_0^{+\infty} \cos x dx$?*

La fonction cosinus est continue sur \mathbb{R} , donc Riemann intégrable sur tout intervalle $[0, b]$. Le calcul par primitivation de cette intégrale donne

$$\forall b > 0, \quad \int_0^b \cos x dx = [\sin x]_0^b = \sin b.$$

Lorsque b tend vers $+\infty$, $\sin b$ n'a pas de limite, même dans $\overline{\mathbb{R}}$, on ne peut donc pas définir $\int_0^{+\infty} \cos x dx$. Géométriquement, le domaine H délimité par le demi-axe des abscisses positives, l'axe des ordonnées et la courbe $y = \cos x$, $x \geq 0$, n'a pas d'aire algébrique.

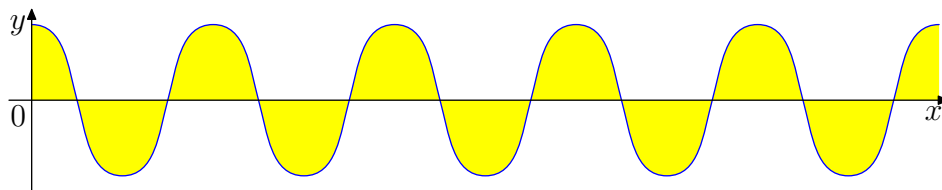


FIG. B.2 – Domaine H délimité par le graphe de $f : x \mapsto \cos x$ entre 0 et $+\infty$

Exemple B.4. *Peut-on définir les intégrales $\int_0^1 \frac{dt}{t^\alpha}$, $\alpha > 0$?*

Notons en préalable que si $\alpha \leq 0$, la réponse est immédiate puisque sur $[0, 1]$ la fonction $f : t \mapsto t^{-\alpha}$ est continue donc Riemann intégrable. Par contre si $\alpha > 0$, f n'est pas définie en 0, est continue sur $]0, 1]$ et tend vers $+\infty$ à droite en 0.

Puisque f est continue sur $]0, 1]$, elle est Riemann intégrable sur tout intervalle $[a, 1]$ pour $a > 0$. On va donc regarder la convergence éventuelle de $I(a) := \int_a^1 f(t) dt$ lorsque a tend vers 0 par valeurs supérieures. L'intégrale $I(a)$ se calcule par primitivation. Si $\alpha \neq 1$, on obtient

$$\forall a > 0, \quad I(a) = \int_a^1 \frac{dt}{t^\alpha} = \left[\frac{t^{-\alpha+1}}{-\alpha+1} \right]_a^1 = \frac{1 - a^{-\alpha+1}}{-\alpha+1}$$

Quand a tend vers 0 par la droite, $I(a)$ tend vers une limite finie $1/(1-\alpha)$ si $-\alpha+1 > 0$, *i.e.* si $\alpha < 1$. Par contre si $-\alpha+1 < 0$, $I(a)$ tend vers $+\infty$, compte tenu du signe négatif du dénominateur constant $-\alpha+1$. Dans le cas particulier $\alpha = 1$, une primitive de f est la fonction logarithme népérien, d'où

$$\forall a > 0, \quad I(a) = \int_a^1 \frac{dt}{t} = [\ln t]_a^1 = -\ln a,$$

ce qui tend vers $+\infty$ quand a tend vers 0 par la droite.

Finalement nous pouvons écrire

$$\int_0^1 \frac{dt}{t^\alpha} = \begin{cases} \frac{1}{1-\alpha} & \text{si } \alpha < 1, \\ +\infty & \text{si } \alpha \geq 1. \end{cases}$$

À nouveau on peut interpréter géométriquement ce résultat. Soit $H_\alpha := \{(t, y) \in \mathbb{R}^2; 0 < t \leq 1 \text{ et } 0 \leq y \leq t^{-\alpha}\}$ l'hypographe de f entre 0 et 1, cf. figure B.3. Si $\alpha < 1$, son aire est finie et vaut $(1-\alpha)^{-1}$. Si $\alpha \geq 1$, son aire est infinie. Pour la justification, on peut utiliser la suite, croissante pour l'inclusion, des hypographes de f entre $1/n$ et 1.

Exemple B.5. Peut-on définir $\int_{-1}^1 \frac{dt}{t}$?

L'intégrande $f : t \mapsto t^{-1}$ est définie et continue sur l'intervalle « troué » $[-1, 1] \setminus \{0\}$. Elle est donc Riemann intégrable sur chacun des intervalles fermés bornés $[-1, a]$ et $[b, +1]$ pour $-1 < a < 0$ et $0 < b < 1$. Ceci nous amène à étudier la limite quand a et b tendent vers respectivement 0 à gauche ou à droite de

$$I(a, b) := \int_{-1}^a \frac{dt}{t} + \int_b^1 \frac{dt}{t} = [\ln |t|]_{-1}^a + [\ln |t|]_b^1 = \ln |a| - \ln b.$$

Quand a et b tendent vers $0-$ et $0+$ respectivement, $\ln |a|$ et $\ln b$ tendent tous deux vers $+\infty$, donc leur différence $I(a, b)$ n'a pas de limite. Si vous n'en êtes pas convaincu, regardez ce problème de convergence avec les suites $a_n := 1/n$, $b_n = 1/n^2$, puis avec les suites $a'_n = -1/n^2$ et $b'_n = 1/n$. Dans le premier cas $I(a_n, b_n) = \ln n$ tend vers $+\infty$, tandis que dans le second, $I(a'_n, b'_n) = -\ln n$ tend vers $-\infty$. Ceci interdit à la fonction de deux variables $(a, b) \mapsto I(a, b)$ d'avoir une limite en $(0, 0)$. On ne peut donc pas définir l'intégrale généralisée $\int_{-1}^1 \frac{dt}{t}$, même comme élément de $\overline{\mathbb{R}}$. Soit H le domaine délimité par le graphe de f et l'axe des abscisses entre -1 et 1 , *i.e.*

$$H := \{(t, y) \in \mathbb{R}^2; t \in [-1, 1] \setminus \{0\}, -(1/t)^- \leq y \leq (1/t)^+\},$$

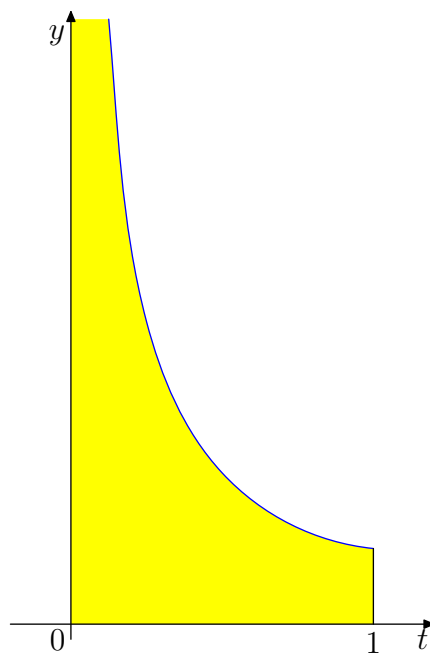


FIG. B.3 – Hypographe H_α de $f : t \mapsto t^{-\alpha}$ entre 0 et 1

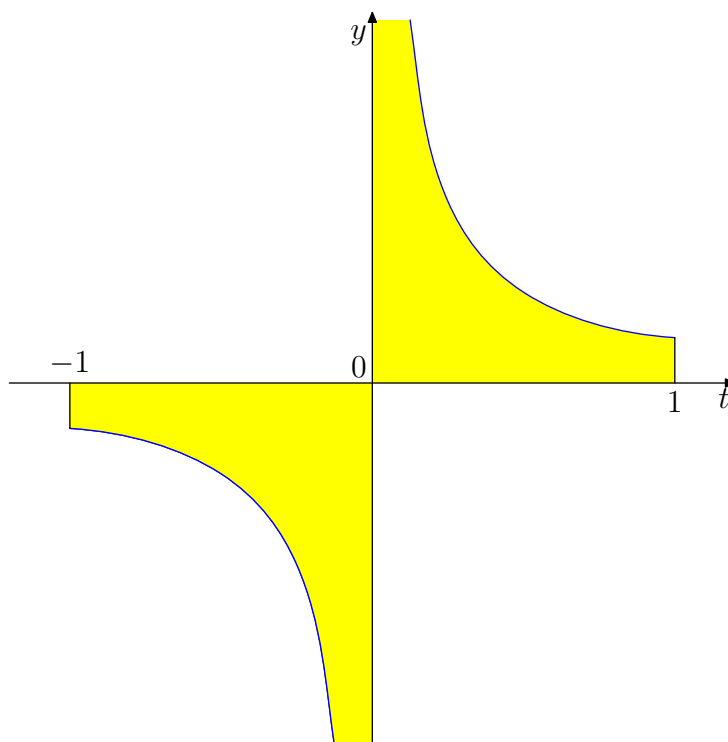


FIG. B.4 – $H := \{(t, y) \in \mathbb{R}^2; t \in [-1, 1] \setminus \{0\}, -(1/t)^- \leq y \leq (1/t)^+\}$

voir figure B.4. Le fait que l'on ne puisse définir $\int_{-1}^1 \frac{dt}{t}$ signifie que H n'a pas d'aire algébrique¹.

Attention, il peut paraître tentant au vu de l'impairité de f de définir $\int_{-1}^1 \frac{dt}{t}$ comme valant 0 et de dire que l'aire algébrique de H est nulle. Il faut absolument résister à cette tentation. En effet, ceci reviendrait à dire que $I(a, b)$ tend vers 0 quand (a, b) tend vers $(0, 0)$ simplement parce que $I(a, -a) = 0$ pour tout $a \in [-1, 0[$.

Après ces exemples introductifs, nous allons formaliser la définition de l'intégrale de Riemann généralisée. Il est commode de désigner les ensembles d'intégration considérés sous le nom² d'« intervalle troué ».

Définition B.6 (intervalle troué). *Soit I un intervalle quelconque de \mathbb{R} et $T = \{t_1, \dots, t_d\}$ une partie de cardinal d de I , l'indexation des t_i vérifiant :*

$$-\infty \leq t_0 := \inf I < t_1 < \dots < t_d < t_{d+1} := \sup I \leq +\infty.$$

On appelle intervalle troué I_T l'ensemble $I \setminus T$. On a alors

$$I_T = \bigcup_{i=0}^d I_i, \quad (\text{B.1})$$

avec $I_i :=]t_i, t_{i+1}[$ pour $1 \leq i < d$, I_0 a pour bornes t_0 et t_1 et est ouvert à droite, I_d a pour bornes t_d et t_{d+1} et est ouvert à gauche. Nous engloberons dans cette définition et ces notations le cas particulier $d = 0$ où il n'y a pas de trous, la réunion ci-dessus se réduisant à $I_\emptyset = I = I_0$.

Les ensembles d'intégration utilisés dans les exemples B.1–B.5 sont ainsi des intervalles troués avec $I = [0, +\infty[$ et $d = 0$ pour les exemples B.1, B.2 et B.3, $I =]0, 1]$ et $d = 0$ pour l'exemple B.4, $I = [-1, 1]$, $d = 1$, $t_1 = 0$ pour l'exemple B.5.

Définition B.7 (fonction localement Riemann intégrable). *Soient I_T un intervalle troué et f une fonction $I_T \rightarrow \mathbb{R}$. On dit que f est localement Riemann intégrable sur I_T si elle est Riemann intégrable sur tout intervalle fermé borné $[\alpha, \beta]$ inclus dans I_T , donc nécessairement inclus dans l'un des intervalles I_i de la décomposition (B.1).*

Définition B.8 (intégrale généralisée). *Soient I un intervalle de \mathbb{R} , de bornes $a, b \in \overline{\mathbb{R}}$, $T = \{t_1, \dots, t_d\}$ un ensemble de trous dans I et f localement Riemann intégrable sur l'intervalle troué I_T . On dit que l'intégrale généralisée de f entre a et b converge si on peut trouver une suite finie c_0, c_1, \dots, c_d avec $c_i \in I_i$ pour $i = 0, \dots, d$ telle que chacune des limites suivantes existe et soit finie :*

$$\forall i = 0, \dots, d, \quad \lim_{\substack{x_i \rightarrow t_{i+1}, \\ x_i < t_{i+1}}} \int_{c_i}^{x_i} f(t) dt =: \ell_i, \quad \lim_{\substack{x'_i \rightarrow t_i, \\ x'_i > t_i}} \int_{x'_i}^{c_i} f(t) dt =: \ell'_i. \quad (\text{B.2})$$

1. Par contre il a une aire $\lambda_2(H) = +\infty$, ce qui correspond au fait que $\int_{-1}^1 \frac{dt}{|t|} = +\infty$ (exercice).

2. non standard.

On dit alors que l'intégrale généralisée $\int_a^b f(t) dt$ converge et on définit $\int_a^b f(t) dt$ comme le réel

$$\int_a^b f(t) dt := \sum_{i=0}^d (\ell'_i + \ell_i). \quad (\text{B.3})$$

Si l'une au moins des conditions (B.2) n'est pas vérifiée, i.e. il n'y a pas de limite ou une limite infinie, on dit que l'intégrale généralisée $\int_a^b f(t) dt$ diverge. Dans ce cas l'écriture $\int_a^b f(t) dt$ ne représente pas un nombre réel.

L'utilisation du symbole $\int_a^b f(t) dt$ est analogue à celle de $\sum_{k=0}^{+\infty} u_k$ qui désigne à la fois une série et lorsqu'elle converge dans \mathbb{R} , sa somme qui est le réel limite de la suite des sommes partielles $S_n = \sum_{k=0}^n u_k$. La série peut diverger parce que la suite $(S_n)_{n \geq 1}$ tend vers $+\infty$ (resp. $-\infty$), auquel cas on s'autorise l'écriture $\sum_{k=0}^{+\infty} u_k = +\infty$ (resp. $-\infty$). Mais elle peut aussi diverger parce que $(S_n)_{n \geq 1}$ n'a pas de limite, même infinie. Dans ce cas l'écriture $\sum_{k=0}^{+\infty} u_k$ est purement formelle et ne représente pas un élément de $\overline{\mathbb{R}}$. Pour l'intégrale généralisée divergente $\int_a^b f(t) dt$, on s'autorisera l'écriture $\int_a^b f(t) dt = +\infty$ si certains des ℓ_i , ou ℓ'_i valent $+\infty$, les autres étant finis. De même on écrira $\int_a^b f(t) dt = -\infty$ si certains des ℓ_i , ou ℓ'_i valent $-\infty$, les autres étant finis. Dans tous les autres cas de divergence³, le symbole $\int_a^b f(t) dt$ est seulement une écriture formelle et ne représente pas un élément de $\overline{\mathbb{R}}$.

Le lecteur attentif n'aura pas manqué de noter que la définition B.8 pose un problème de cohérence car l'existence et la valeur de $\int_a^b f(t) dt$ semblent dépendre du choix de la suite c_0, c_1, \dots, c_d . Le lemme suivant répond à cette légitime inquiétude.

Lemme B.9. Avec les notations de la définition B.8, on suppose que la suite finie c_0, c_1, \dots, c_d vérifie (B.2). Soit $\tilde{c}_0, \tilde{c}_1, \dots, \tilde{c}_d$ une suite telle que $\tilde{c}_i \in I_i$ pour $i = 0, \dots, d$. Alors on a

$$\forall i = 0, \dots, d, \quad \lim_{\substack{x_i \rightarrow t_{i+1}, \\ x_i < t_{i+1}}} \int_{\tilde{c}_i}^{x_i} f(t) dt =: \tilde{\ell}_i = \ell_i + \int_{\tilde{c}_i}^{c_i} f(t) dt \quad (\text{B.4})$$

et

$$\forall i = 0, \dots, d, \quad \lim_{\substack{x'_i \rightarrow t_i, \\ x'_i > t_i}} \int_{x'_i}^{\tilde{c}_i} f(t) dt =: \tilde{\ell}'_i = \ell'_i + \int_{c_i}^{\tilde{c}_i} f(t) dt. \quad (\text{B.5})$$

Une conséquence immédiate de ce lemme est que

$$\sum_{i=0}^d (\ell'_i + \ell_i) = \sum_{i=0}^d (\tilde{\ell}'_i + \tilde{\ell}_i)$$

puisque $\ell'_i + \ell_i = \tilde{\ell}'_i + \tilde{\ell}_i$, la somme de $\int_{\tilde{c}_i}^{c_i}$ et $\int_{c_i}^{\tilde{c}_i}$ s'annihilant par la relation de Chasles. Ainsi ni la convergence de l'intégrale généralisée de f entre a et b ni la définition de sa valeur par (B.3) ne dépendent du choix de la suite c_0, c_1, \dots, c_d .

3. i.e. si l'une au moins des intégrales de (B.2) n'a pas de limite même dans $\overline{\mathbb{R}}$ ou si elles ont toutes une limite dans \mathbb{R} , mais avec au moins une des limites valant $-\infty$ et au moins une valant $+\infty$.

Preuve du lemme B.9. Puisque l'on fait tendre x_i vers t_{i+1} par valeurs inférieures, on peut toujours supposer que $\max(c_i, \tilde{c}_i) < x_i < t_{i+1}$. La fonction f est alors Riemann intégrable sur $[\min(c_i, \tilde{c}_i), x_i]$ et la relation de Chasles combinée avec (B.2) nous donne

$$\int_{\tilde{c}_i}^{x_i} f(t) dt = \int_{\tilde{c}_i}^{c_i} f(t) dt + \int_{c_i}^{x_i} f(t) dt \xrightarrow[x_i < t_{i+1}]{x_i \rightarrow t_{i+1}} \int_{\tilde{c}_i}^{c_i} f(t) dt + \ell_i.$$

La vérification de (B.5) est analogue. \square

Remarque B.10. Dans le cas où l'intervalle I de la définition B.8 est fermé en b (donc $b \in I$), la condition

$$\lim_{\substack{x_d \rightarrow b, \\ x_d < b}} \int_{c_d}^{x_d} f(t) dt =: \ell_d \in \mathbb{R}$$

est automatiquement vérifiée. En effet, $[c_d, b]$ est inclus dans I_T , donc f est Riemann intégrable sur $[c_d, b]$ et la fonction $x_d \mapsto \int_{c_d}^{x_d} f(t) dt$ est continue sur cet intervalle, cf. théorème A.31 i), donc continue à gauche au point b , ce qui nous donne l'existence de la limite finie ℓ_d et sa valeur $\ell_d = \int_{c_d}^b f(t) dt$. En pratique on s'abstiendra donc de revérifier l'existence de cette limite. Par exemple si f est localement Riemann intégrable sur $]c, b]$, il faut seulement regarder ce qui se passe au voisinage de c .

Bien entendu si $a \in I$, on a une situation analogue, à savoir la convergence automatique quand $x'_0 \rightarrow a+$ de $\int_{x'_0}^{c_0} f(t) dt$ vers l'intégrale de Riemann ordinaire $\int_a^{c_0} f(t) dt$.

Nous définissons aussi les intégrales généralisées \int_a^b avec $a > b$ en cohérence avec la définition A.2.

Définition B.11. Si $a > b$ et si l'intégrale généralisée $\int_b^a f(t) dt$ converge, on pose

$$\int_a^b f(t) dt := - \int_b^a f(t) dt. \quad (\text{B.6})$$

Remarque B.12 (réduction du problème). Une fois payé notre tribut au formalisme avec la définition B.8, il convient de se simplifier la vie en notant que l'étude de la convergence d'une intégrale généralisée se ramène à l'étude de limites du type $\int_{x'}^c f(t) dt$ quand $x' \rightarrow a+$ avec f localement Riemann intégrable sur $]a, c]$ ou $\int_c^x f(t) dt$ quand $x \rightarrow b-$ avec f localement Riemann intégrable sur $[c, b[$. Nous nous contenterons la plupart du temps, d'énoncer des résultats relatifs au deuxième type, en laissant au lecteur le soin d'écrire leur adaptation immédiate au premier type et de recoller les morceaux par (B.3).

Définition B.13 (notation $\mathcal{R}^{\text{loc}}[a, b]$). Soient a et b tels que $-\infty < a < b \leq +\infty$. Nous notons $\mathcal{R}^{\text{loc}}[a, b]$ l'ensemble des fonctions localement Riemann intégrables sur $[a, b]$.

Pour étudier la convergence d'une intégrale généralisée, on a souvent recours à une technique de comparaison avec une intégrale de référence. Les intégrales généralisées des fonctions puissances $t \mapsto t^{-\alpha}$ au voisinage de 0 ou de $+\infty$ sont les intégrales de référence les plus utilisées.

Proposition B.14. Soit α un réel.

- a) $\int_1^{+\infty} \frac{dt}{t^\alpha}$ converge si $\alpha > 1$ et diverge si $\alpha \leq 1$.
- b) $\int_0^1 \frac{dt}{t^\alpha}$ converge si $\alpha < 1$ et diverge si $\alpha \geq 1$.

Preuve. Le a) a déjà été traité à l'exemple B.4 en rappelant que si $\alpha \leq 0$, on a affaire à l'intégrale de Riemann ordinaire d'une fonction continue sur $[0, 1]$. Pour le b), on note que $f : t \mapsto t^{-\alpha}$ est continue sur $[1, +\infty[$ donc localement Riemann intégrable sur cet intervalle, donc Riemann intégrable sur tout intervalle $[1, x]$ pour $x \geq 1$. Par primitivation on a

$$\int_1^x \frac{dt}{t^\alpha} = \begin{cases} \frac{x^{-\alpha+1}}{-\alpha+1} - \frac{1}{1-\alpha} & \text{si } \alpha \neq 1, \\ \ln x & \text{si } \alpha = 1. \end{cases}$$

On en déduit que pour $\alpha \leq 1$, $\int_1^x t^{-\alpha} dt$ tend vers $+\infty$ quand x tend vers $+\infty$ et que pour $\alpha > 1$, $\int_1^x t^{-\alpha} dt$ tend vers la limite finie $1/(\alpha - 1)$ quand x tend vers $+\infty$. \square

Corollaire B.15. Si $-\infty < a < b < +\infty$ et si α est un réel,

$$\int_a^b \frac{dt}{(t-a)^\alpha} \quad \text{converge si et seulement si } \alpha < 1, \quad (\text{B.7})$$

$$\int_a^b \frac{dt}{(b-t)^\alpha} \quad \text{converge si et seulement si } \alpha < 1. \quad (\text{B.8})$$

Preuve. C'est une adaptation immédiate de la preuve du a) ci-dessus, voir exemple B.4, via les changements de variable $u = t - a$ et $v = b - t$ respectivement. \square

Dans la situation de la proposition B.14, l'intégrale diverge lorsque f tend trop vite vers $+\infty$ en 0 dans le cas a) et tend trop lentement vers 0 en $+\infty$ dans le cas b)⁴. Dans le cas b), on peut penser à l'analogie avec la série de terme général $k^{-\alpha}$, cf. théorème 1.55 p. 26. Pour autant il faut se garder de tirer des conclusions hâtives de cette analogie. Si la convergence d'une série implique toujours la convergence vers 0 de son terme général, la situation est plus compliquée pour les intégrales de la forme $\int_a^{+\infty} f(t) dt$.

Remarque B.16. La convergence de l'intégrale $\int_a^{+\infty} f(t) dt$ n'implique pas que $f(t)$ ait pour limite 0 en $+\infty$. Voici un contre exemple. Prenons $a = 0$ et pour f la fonction continue affine par morceaux dont l'hypographe se réduit (en dehors des segments où f est nulle) à la réunion de la suite $(T_n)_{n \geq 1}$ des triangles isocèles de sommet principal $(n, 2^n)$ et de base $[n - 4^{-n}, n + 4^n]$, cf. figure⁵ B.5. La fonction f vérifie les deux propriétés suivantes.

4. On pourrait d'ailleurs déduire le b) du a) par un argument géométrique sur les hypographe en notant que les fonctions $f : t \mapsto t^{-\alpha}$ et $g : t \mapsto t^{-1/\alpha}$ définies sur $]0, +\infty[$ sont réciproques l'une de l'autre, donc que leurs graphes en repère orthonormé se correspondent par la symétrie relativement à la première bissectrice du repère. Par conservation de λ_2 , on en déduit que l'hypographe de g entre 1 et $+\infty$ a même aire que l'hypographe de f entre 0 et 1 privé du carré unité (faites le dessin!).

5. Pour des raisons de lisibilité on a muni l'axe des ordonnées d'une échelle logarithmique et on a fortement exagéré la base de chaque triangle isocèle.

1. $\int_a^{+\infty} f(t) dt$ converge (et vaut 1).
2. Pour tout $n \in \mathbb{N}^*$, $f(n) = 2^n$ et $f(n+1/2) = 0$. Par conséquent f n'a pas de limite en $+\infty$.

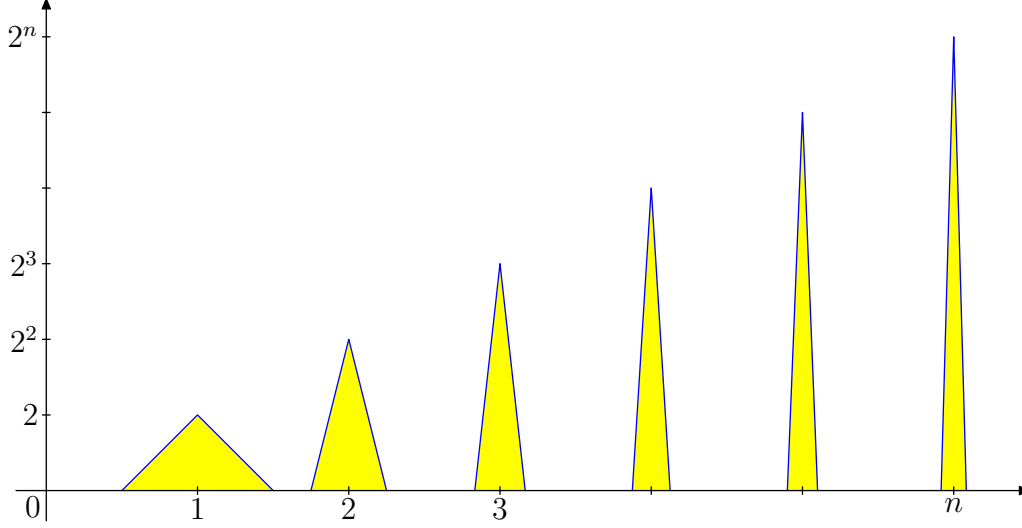


FIG. B.5 – $\int_0^{+\infty} f(t) dt$ peut converger sans que f ait une limite en $+\infty$

Preuve. Comme fonction continue, f appartient à $\mathcal{R}^{\text{loc}}[0, +\infty[$ donc est Riemann intégrable sur tout intervalle $[0, x]$, $x \in \mathbb{R}_+$. Les triangles T_n sont deux à deux disjoints et l'aire $\lambda_2(T_n)$ se calcule par la formule classique demi-produit de la base par la hauteur ⁶, d'où

$$\lambda_2(T_n) = 4^{-n} \times 2^n = 2^{-n}.$$

On en déduit immédiatement que la série de terme général $\lambda_2(T_k)$ est géométrique convergente. Le calcul de sa somme partielle S_n est bien connu :

$$S_n := \sum_{k=1}^n \lambda_2(T_k) = \sum_{k=1}^n 2^{-k} = 2^{-1} \sum_{j=0}^{n-1} 2^{-j} = 2^{-1} \frac{1 - 2^{-n}}{1 - 2^{-1}} = 1 - 2^{-n}.$$

Nous allons montrer la convergence de $\int_0^{+\infty} f(t) dt$ en comparant $\int_0^x f(t) dt$ et S_n pour $n = [x]$, la partie entière de x . En effet, pour n fixé, $g_n : x \mapsto \int_0^x f(t) dt - S_n$ est une fonction croissante puisque f est positive. Sur l'intervalle $[n, n+1]$, cette fonction a pour minimum $g_n(n) = -\frac{1}{2}\lambda_2(T_n)$ et pour maximum $g_n(n+1) = \frac{1}{2}\lambda_2(T_{n+1})$. Comme $\lambda_2(T_n) > \lambda_2(T_{n+1})$, on en déduit

$$\forall x \in [n, n+1], \quad \left| \int_0^x f(t) dt - S_n \right| \leq \frac{1}{2} \lambda_2(T_n) < 2^{-n}.$$

6. Si vous êtes sceptiques vous pouvez toujours chercher une expression analytique pour la restriction de f à $[n - 4^{-n}, n + 4^{-n}]$ et l'intégrer sur ce segment pour voir si vous trouvez le même résultat.

Compte-tenu du calcul de S_n rappelé ci-dessus, ceci nous permet d'écrire

$$\forall x \geq 1, \quad \left| \int_0^x f(t) dt - (1 - 2^{-[x]}) \right| < 2^{-[x]}.$$

En faisant tendre x vers $+\infty$, on en déduit que $\int_0^x f(t) dt$ tend vers 1. Donc $\int_0^{+\infty} f(t) dt$ converge et vaut 1. Le point 2 est évident. \square

Ce genre de pathologie n'est pas réservé aux intégrales généralisées de la forme $\int_0^{+\infty} f(t) dt$. À titre d'exercice, on vous laisse le soin de construire une fonction f continue sur $[0, 1[$ telle que $\int_0^1 f(t) dt$ converge et qu'il existe 3 suites $(u_n)_{n \geq 1}$, $(v_n)_{n \geq 1}$ et $(w_n)_{n \geq 1}$ convergentes vers 1 dans $[0, 1[$ telles que $f(u_n)$ tende vers $+\infty$, $f(v_n)$ tende vers 0 et $f(w_n)$ tende vers $-\infty$. Voici une suggestion parmi les multiples solutions possibles. Découper $[0, 1[$ en trois segments de même longueur $[0, 1/3[$, $[1/3, 2/3[$ et $[2/3, 1[$. Sur les deux premiers prendre pour graphe de f les triangles isocèles de base ces segments et de « hauteurs » respectives $+2$ et -2 . Itérer ce partage en trois sur $[2/3, 1[$ avec sur les deux premiers segments du partage des triangles isocèles de hauteur $+4$ et -4 et ainsi de suite jusqu'à l'infini.

Après ces contre exemples, voyons ce que l'on peut dire dans des situations moins pathologiques sur la relation entre convergence de $\int_a^{+\infty} f(t) dt$ et comportement de f au voisinage de $+\infty$.

Proposition B.17. Soit $f \in \mathcal{R}^{\text{loc}}[a, +\infty[$.

- i) On suppose qu'il existe $A \in [a, +\infty[$ et $m > 0$ tels que $f(t) \geq m$ pour tout $t \geq A$. Alors $\int_a^{+\infty} f(t) dt$ diverge.
- ii) Cette divergence a lieu aussi s'il existe $m' < 0$ et $A \geq a$ tels que $f(t) \leq m'$ pour tout $t \geq A$.
- iii) En conséquence, si f a une limite non nulle $\ell \in \overline{\mathbb{R}}$ en $+\infty$, $\int_a^{+\infty} f(t) dt$ diverge.

Preuve. Pour i), il suffit de remarquer que pour tout $x \geq A$,

$$\int_a^x f(t) dt = \int_a^A f(t) dt + \int_A^x f(t) dt \geq \int_a^A f(t) dt + m(x - A),$$

par croissance de l'intégrale de Riemann sur $[A, x]$, voir la proposition A.21 ii) et la figure B.6. En faisant tendre x vers $+\infty$, on en déduit que $\int_a^x f(t) dt$ tend vers $+\infty$, d'où la divergence de $\int_a^{+\infty} f(t) dt$.

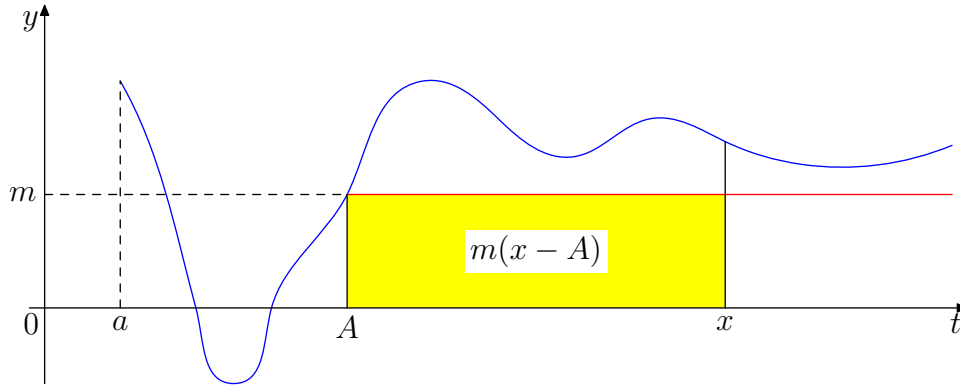
De même pour ii), on obtient la minoration $\int_a^x f(t) dt \leq \int_a^A f(t) dt + m'(x - A)$ et donc $\int_a^x f(t) dt$ tend vers $-\infty$ quand x tend vers $+\infty$.

Supposons maintenant que f ait une limite non nulle $\ell \in \overline{\mathbb{R}}$ en $+\infty$. On peut distinguer 4 cas.

Cas 1. $\ell \in]0, +\infty[$, alors il existe un $A \geq a$ tel que pour tout $t \geq A$, $f(t) > \frac{\ell}{2}$. On applique i) avec $m = \frac{\ell}{2} > 0$.

Cas 2. $\ell = +\infty$, alors il existe un $A \geq a$ tel que pour tout $t \geq A$, $f(t) \geq 1$, on pourrait bien sûr remplacer ce minorant 1 par n'importe quel réel $B > 0$ choisi à l'avance⁷. On applique i) avec $m = 1$.

7. Mais pas par $\frac{\ell}{2}$ qui a ici le mauvais goût de valoir $+\infty$!


 FIG. B.6 – $\int_A^x f(t) dt \geq m(x - A)$

Cas 3. $\ell \in]-\infty, 0[$, alors il existe un $A \geq a$ tel que pour tout $t \geq A$, $f(t) < \frac{\ell}{2}$. On applique ii) avec $m' = \frac{\ell}{2} < 0$.

Cas 4. $\ell = -\infty$, alors il existe un $A \geq a$ tel que pour tout $t \geq A$, $f(t) \leq -1$. On applique ii) avec $m = -1$.

□

Corollaire B.18. Soit a un réel et f une fonction positive et décroissante sur $[a, +\infty[$. Si $\int_a^{+\infty} f(t) dt$ converge, alors f tend vers 0 en $+\infty$.

Preuve. Une fonction décroissante sur $[A, +\infty[$ est localement Riemann intégrable sur cet intervalle, cf. proposition A.9. D'autre part comme fonction monotone, elle admet toujours en $+\infty$ une limite $\ell \in \overline{\mathbb{R}}$. Par décroissance et positivité de f , cette limite est nécessairement dans $[0, +\infty[$. Si $\ell > 0$, alors par le cas 1 ci-dessus, $\int_a^{+\infty} f(t) dt$ diverge, ce qui contredit l'hypothèse de convergence de cette intégrale. Donc $\ell = 0$. □

B.2 Critère de Cauchy pour intégrales généralisées

L'intérêt du critère de Cauchy — dans un espace complet — est de permettre d'établir l'existence d'une limite sans connaître *a priori* sa valeur. Nous allons voir une version de ce critère pour la convergence des intégrales généralisées. Auparavant, il n'est peut-être pas superflu de rappeler quelques versions du critère de Cauchy pour l'existence de limites de suites ou de fonctions.

Proposition B.19 (critères de Cauchy).

1. La suite de réels $(u_n)_{n \geq 1}$ converge dans \mathbb{R} si et seulement si

$$\forall \varepsilon > 0, \exists N \in \mathbb{N}^*, \quad \forall n, p \geq N, \quad |u_n - u_p| < \varepsilon. \quad (\text{B.9})$$

2. Soit F une fonction à valeurs réelles ou complexes, définie sur $D \subset \mathbb{R}$ et a un point adhérent à D . Alors F a une limite finie au point a si et seulement si :

$$\forall \varepsilon > 0, \exists \delta > 0, \quad \forall x, x' \in D \cap]a - \delta, a + \delta[\setminus \{a\}, \quad |F(x) - F(x')| < \varepsilon. \quad (\text{B.10})$$

3. Soit F une fonction à valeurs réelles ou complexes, définie sur $D \subset \mathbb{R}$ et $a \in \mathbb{R}$ tel que pour tout réel $\eta > 0$, $D \cap]a, a + \eta[$ soit non vide. Alors F a une limite à droite finie au point a si et seulement si :

$$\forall \varepsilon > 0, \exists \delta > 0, \quad \forall x, x' \in D \cap]a, a + \delta[, \quad |F(x) - F(x')| < \varepsilon. \quad (\text{B.11})$$

4. Soit F une fonction à valeurs réelles ou complexes, définie sur $D \subset \mathbb{R}$ et $a \in \mathbb{R}$ tel que pour tout réel $\eta > 0$, $D \cap]a - \eta, a[$ soit non vide. Alors F a une limite à gauche finie au point a si et seulement si :

$$\forall \varepsilon > 0, \exists \delta > 0, \quad \forall x, x' \in D \cap]a - \delta, a[, \quad |F(x) - F(x')| < \varepsilon. \quad (\text{B.12})$$

5. Soit F une fonction à valeurs réelles ou complexes, définie sur un intervalle $[a, +\infty[$, $a \in \mathbb{R}$. Alors F a une limite finie en $+\infty$ si et seulement si :

$$\forall \varepsilon > 0, \exists A > 0, \quad \forall x, x' \geq A, \quad |F(x) - F(x')| < \varepsilon. \quad (\text{B.13})$$

De même que l'application du critère de Cauchy (B.9) à la suite des sommes partielles d'une série conduit au critère de Cauchy pour les séries, cf. théorème 1.50, les critères (B.10)–(B.13) nous fournissent des critères de Cauchy pour la convergence d'intégrales généralisées. Nous les énoncerons seulement pour la convergence en b des intégrales $\int_a^b f(t) dt$ en utilisant (B.12) ou (B.13) selon que b est fini ou non. Au lecteur de compléter.

Théorème B.20 (critère de Cauchy pour les intégrales).

1. Soit $f \in \mathcal{R}^{\text{loc}}[a, +\infty[$. L'intégrale généralisée $\int_a^{+\infty} f(t) dt$ converge si et seulement si :

$$\forall \varepsilon > 0, \exists A > 0, \quad \forall x, x' \geq A, \quad \left| \int_x^{x'} f(t) dt \right| < \varepsilon. \quad (\text{B.14})$$

2. Soit $f \in \mathcal{R}^{\text{loc}}[a, b[$, avec $b < +\infty$. Alors $\int_a^b f(t) dt$ converge si et seulement si :

$$\forall \varepsilon > 0, \exists \delta \in]0, b - a[, \quad \forall x, x' \in]b - \delta, b[, \quad \left| \int_x^{x'} f(t) dt \right| < \varepsilon. \quad (\text{B.15})$$

Preuve. Il suffit d'appliquer le critère de Cauchy (B.12) ou (B.13) à la fonction $F : [a, b[\rightarrow \mathbb{R}$, $x \mapsto F(x) := \int_a^x f(t) dt$ qui est bien définie sur $[a, b[$ puisque f est Riemann intégrable sur $[a, x]$ pour tout $x \in [a, b[$. \square

Corollaire B.21. Soit $f \in \mathcal{R}^{\text{loc}}[a, b[$ avec b fini.

1. Si f est bornée sur $[a, b[$, alors $\int_a^b f(t) dt$ converge.
2. Si f a une limite à gauche finie en b , alors $\int_a^b f(t) dt$ converge.

Preuve. Vérifions le point 1. Par hypothèse, il existe $M \in \mathbb{R}_+$ tel que pour tout $t \in [a, b]$, $|f(t)| \leq M$. D'autre part f étant localement Riemann intégrable sur $[a, b]$ est Riemann intégrable sur tout segment $[x, x'] \subset [a, b]$. Par croissance de l'intégrale de Riemann, cf. proposition A.21, on en déduit :

$$\forall x, x' \in [a, b] \text{ avec } x < x', \quad \left| \int_x^{x'} f(t) dt \right| \leq \int_x^{x'} |f(t)| dt \leq M(x' - x). \quad (\text{B.16})$$

Cette inégalité nous permet de vérifier le critère de Cauchy (B.15). En effet, soit $\varepsilon > 0$ arbitraire. Posons $\delta := \min(\varepsilon/M, b - a)$. Pour tous $x, x' \in]b - \delta, b]$, on a clairement $M(x' - x) < \varepsilon$, donc compte-tenu de (B.16), $|\int_x^{x'} f(t) dt| < \varepsilon$.

Pour le point 2, il suffit de noter que si f a une limite finie ℓ à gauche en b , alors sur un intervalle $]b - \delta', b]$ suffisamment petit, on a $|f(t)| \leq |\ell| + 1$. Comme f est aussi bornée sur $[a, b - \delta']$ car Riemann intégrable sur ce segment, elle est bornée sur la réunion des deux intervalles, *i.e.* sur $[a, b]$ et on conclut en appliquant le point 1. \square

Exemple B.22. L'intégrale généralisée $\int_{-1/\pi}^0 \sin\left(\frac{1}{t}\right) dt$ converge.

En effet, $f : t \mapsto \sin(1/t)$ est définie et continue sur $[-1/\pi, 0[$, donc $f \in \mathcal{R}^{\text{loc}}[-1/\pi, 0[$. Elle est bornée sur cet intervalle puisque $|\sin(1/t)| \leq 1$ pour tout $t \in \mathbb{R}^*$. Le point 1 du corollaire B.21 nous donne la convergence de $\int_{-1/\pi}^0 f(t) dt$. Notons au passage que f n'a pas de limite à gauche en zéro, car elle oscille une infinité de fois entre les valeurs -1 et 1 sur tout voisinage à gauche de 0 , aussi petit soit-il, cf. figure B.7.

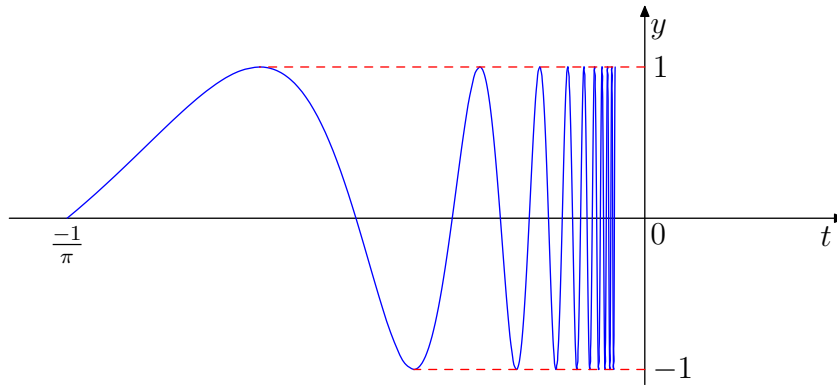


FIG. B.7 – Graphe de $t \mapsto \sin(1/t)$ pour $t \in [-\frac{1}{\pi}, -\frac{2}{39\pi}]$

Corollaire B.23. Soit $f \in \mathcal{R}^{\text{loc}}[a, b]$. Si $\int_a^b |f(t)| dt$ converge, alors $\int_a^b f(t) dt$ converge.

Preuve. Par le théorème B.20, la convergence de $\int_a^b |f(t)| dt$ implique le critère de Cauchy (B.14) si $b \in \mathbb{R}$ ou (B.15) si $b = +\infty$, avec $|f|$ à la place de f . L'inégalité

$$\left| \int_x^{x'} f(t) dt \right| \leq \int_x^{x'} |f(t)| dt$$

montre que le critère de Cauchy correspondant est aussi vérifié par f , d'où la convergence de $\int_a^b f(t) dt$ par une nouvelle invocation du théorème B.20. \square

Remarque B.24. La réciproque du corollaire B.23 est fausse. Nous verrons un peu plus tard que par (contre) exemple, $\int_1^{+\infty} \frac{\sin t}{t} dt$ converge, alors que $\int_1^{+\infty} \left| \frac{\sin t}{t} \right| dt$ diverge.

Définition B.25 (convergence absolue). Soit $f \in \mathcal{R}^{\text{loc}}[a, b]$. Si $\int_a^b |f(t)| dt$ converge, on dit que $\int_a^b f(t) dt$ est absolument convergente.

Une intégrale absolument convergente est toujours convergente (cor. B.23), la réciproque est fausse (rem. B.24).

Le théorème suivant permet entre autres de montrer la convergence d'intégrales de la forme $\int_a^{+\infty} f(t) \sin t dt$ avec f positive décroissante et tendant vers 0 en $+\infty$. Sa preuve combine le critère de Cauchy et la deuxième formule de la moyenne que nous n'avons pas vue. Nous admettrons ce théorème.

Théorème B.26 (critère d'Abel). Soient $f, g \in \mathcal{R}^{\text{loc}}[a, +\infty[$ et vérifiant

- i) f est positive et décroissante sur $[a, +\infty[$ et a pour limite 0 en $+\infty$.
- ii) Il existe une constante M telle que

$$\forall x, y \in [a, +\infty[, \quad \left| \int_x^y g(t) dt \right| \leq M. \quad (\text{B.17})$$

Alors $\int_a^{+\infty} f(t)g(t) dt$ converge.

Si on prend en particulier $g(t) = \sin t$, il est facile de vérifier (B.17). En effet

$$\int_x^y \sin t dt = [-\cos t]_x^y = \cos x - \cos y \quad \text{et} \quad |\cos x - \cos y| \leq 2.$$

Il en va de même avec $g(t) = \cos t$, ou $g(t) = \sin(ct)$, ou $g(t) = \cos(ct)$.

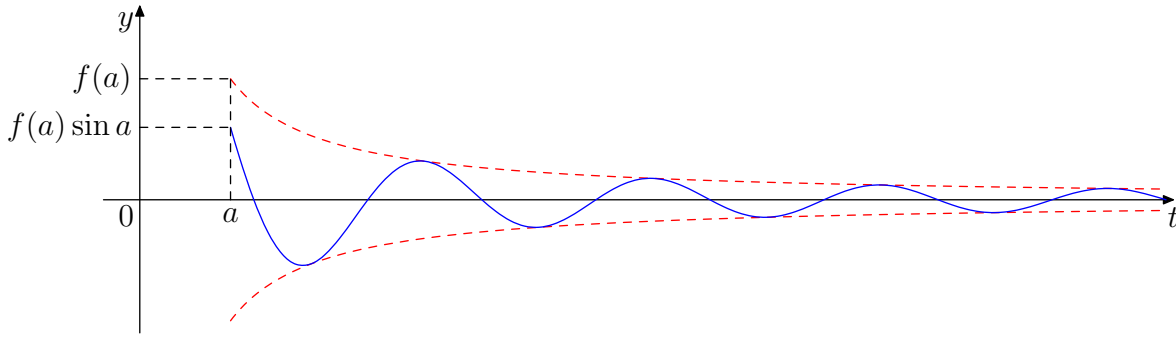
Énonçons séparément ce cas particulier du théorème B.26 avant d'en proposer une démonstration directe.

Proposition B.27. Si f est positive et décroissante sur $[a, +\infty[$ et a pour limite 0 en $+\infty$, les intégrales généralisées $\int_a^{+\infty} f(t) \sin t dt$ et $\int_a^{+\infty} f(t) \cos t dt$ convergent.

Preuve. Nous montrerons simplement la convergence de $\int_a^{+\infty} f(t) \sin t dt$, l'adaptation de la méthode au cas de $\int_a^{+\infty} f(t) \cos t dt$ étant immédiate. Remarquons d'abord que l'intégrande $h : t \mapsto f(t) \sin t$ est localement Riemann intégrable sur $[a, +\infty[$. En effet, les restrictions des fonctions f et \sin au segment $[\alpha, \beta] \subset [a, +\infty[$ sont respectivement monotone et continue donc Riemann intégrables. Leur produit h est donc aussi Riemann intégrable sur $[\alpha, \beta]$, cf. prop. A.26. Ceci étant vrai pour tout $[\alpha, \beta] \subset [a, +\infty[$, h est dans $\mathcal{R}^{\text{loc}}[a, +\infty[$.

Notons $I_k := [k\pi, (k+1)\pi]$, $k \in \mathbb{N}^*$. Pour k assez grand, disons $k \geq k_0$, $I_k \subset [a, +\infty[$. Aux bornes de I_k , $\sin t$ s'annule et pour tout t intérieur à I_k , $\sin t$ a même signe que $(-1)^k$. On a donc pour tout $t \in I_k$, $\sin t = (-1)^k |\sin t|$ d'où

$$\int_{k\pi}^{(k+1)\pi} f(t) \sin t dt = (-1)^k \int_{k\pi}^{(k+1)\pi} f(t) |\sin t| dt =: (-1)^k v_k. \quad (\text{B.18})$$


 FIG. B.8 – Graphe de $t \mapsto f(t) \sin(t)$ avec $f \downarrow 0$, graphes de f et $-f$ en pointillés

On vérifie maintenant que la convergence quand x tend vers $+\infty$ de $\int_a^x h(t) dt$ se réduit à la convergence de la série de terme général $(-1)^k v_k$. Pour $x > k_0\pi$, il existe un unique entier n tel que $(n+1)\pi \leq x < (n+2)\pi$. Cet entier dépendant de x tend évidemment vers $+\infty$ avec x . On peut alors écrire

$$\int_a^x h(t) dt = \int_a^{k_0\pi} h(t) dt + \int_{k_0\pi}^{(n+1)\pi} h(t) dt + \int_{(n+1)\pi}^x h(t) dt. \quad (\text{B.19})$$

Traitions d'abord le terme « résiduel »

$$\varepsilon(x) := \int_{(n+1)\pi}^x h(t) dt,$$

en notant que par la majoration $|\sin t| \leq 1$ et la décroissance de f ,

$$|\varepsilon(x)| \leq \int_{(n+1)\pi}^x |h(t)| dt \leq \int_{(n+1)\pi}^x f(t) dt \leq (x - (n+1)\pi) f((n+1)\pi) \leq \pi f((n+1)\pi).$$

Or f tend vers 0 en $+\infty$ et $n = n(x)$ tend vers l'infini avec x , donc $\varepsilon(x)$ tend vers 0 en $+\infty$. En notant C la constante $\int_a^{k_0\pi} h(t) dt$, en utilisant la relation de Chasles et (B.18), nous pouvons ainsi réécrire (B.19) sous la forme

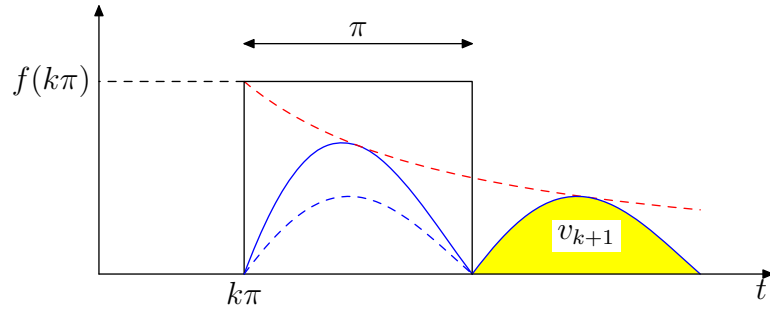
$$\int_a^x h(t) dt = C + \sum_{k=k_0}^n (-1)^k v_k + \varepsilon(x), \quad \lim_{x \rightarrow +\infty} \varepsilon(x) = 0.$$

Il est alors clair que la convergence en $+\infty$ de $\int_a^x h(t) dt$ vers une limite finie équivaut à la convergence de la série de terme général $(-1)^k v_k$.

La convergence de cette série résultera du théorème des séries *alternées* (th. 1.59) si l'on montre que la suite $(v_k)_{k \geq k_0}$ tend vers 0 en décroissant.

Les mêmes majorations que celles utilisées pour $\varepsilon(x)$ nous donnent

$$v_k = \int_{k\pi}^{(k+1)\pi} f(t) |\sin t| dt \leq \pi f(k\pi) \xrightarrow{k \rightarrow +\infty} 0.$$


 FIG. B.9 – $v_k \downarrow 0$ car $\pi f(k\pi) \geq v_k \geq v_{k+1}$

Pour vérifier la décroissance de $(v_k)_{k \geq k_0}$, comparons v_k et v_{k+1} grâce au changement de variable $s = t + \pi$ qui transforme I_{k+1} en I_k :

$$v_{k+1} = \int_{(k+1)\pi}^{(k+2)\pi} f(s) |\sin s| \, ds = \int_{k\pi}^{(k+1)\pi} f(t + \pi) |\sin t| \, dt.$$

Par décroissance de f on a pour tout $t \in I_k$, $f(t) |\sin t| \geq f(t + \pi) |\sin t|$ d'où par intégration de cette inégalité sur I_k , $v_k \geq v_{k+1}$. \square

Exemple B.28. L'intégrale $\int_1^{+\infty} \frac{\sin t}{t} \, dt$ est convergente, mais pas absolument.

La convergence résulte de la proposition B.27 avec $f(t) = 1/t$. Vérifions que l'intégrale n'est pas absolument convergente. Pour $x \geq \pi$, il existe un unique entier n tel que $(n+1)\pi \leq x < (n+2)\pi$ et cet entier dépendant de x tend vers $+\infty$ avec x . En posant $|h(t)| := t^{-1} |\sin t|$ et $v_k := \int_{k\pi}^{(k+1)\pi} |h(t)| \, dt$, on a

$$\int_1^x |h(t)| \, dt = \int_1^\pi |h(t)| \, dt + \sum_{k=1}^n v_k + \int_{(n+1)\pi}^x |h(t)| \, dt \geq \sum_{k=1}^n v_k.$$

Pour prouver que $\int_1^x |h(t)| \, dt$ tend vers $+\infty$ avec x , il suffit donc de montrer la divergence de la série de terme général positif v_k . Ceci résulte de la minoration suivante qui utilise la décroissance de $t \mapsto 1/t$ et la π -périodicité de $|\sin t|$:

$$v_k = \int_{k\pi}^{(k+1)\pi} \frac{|\sin t|}{t} \, dt \geq \int_{k\pi}^{(k+1)\pi} \frac{|\sin t|}{(k+1)\pi} \, dt = \frac{1}{(k+1)\pi} \int_0^\pi \sin t \, dt = \frac{2}{(k+1)\pi}.$$

Comme la série de terme général positif $u_k := \frac{2}{(k+1)\pi}$ diverge (cor. 1.56), on en déduit que

$$\sum_{k=1}^n v_k \geq \sum_{k=1}^n u_k \xrightarrow{n \rightarrow +\infty} +\infty.$$

Ainsi $\int_1^x |h(t)| \, dt$ tend bien vers $+\infty$ avec x .

Nous avons choisi pour cet exemple d'intégrer entre 1 et $+\infty$, mais le résultat reste valable en intégrant la même fonction entre 0 et $+\infty$. On vous laisse en exercice la justification de la convergence en 0.

B.3 Intégrales généralisées de fonctions positives

La fonction f définie sur $[a, b[$ est dite positive sur $[a, b[$ si pour tout $t \in [a, b[$, $f(t) \geq 0$. Nous énoncerons tous les résultats de cette section avec des fonctions positives sur $[a, b[$, mais il est clair qu'ils s'étendent au cas plus général des fonctions f définies sur $[a, b[$ et positives *au voisinage de b* , i.e. il existe un $c \in [a, b[$ tel que $t \in [c, b[$, $f(t) \geq 0$. Ils s'étendent aussi au cas des fonctions de signe constant au voisinage de b , modulo une adaptation laissée au lecteur.

Soit $f \in \mathcal{R}^{\text{loc}}[a, b[$ et positive sur $[a, b[$. Alors la fonction F définie par

$$F(x) := \int_a^x f(t) dt, \quad x \in [a, b[,$$

est *croissante* sur $[a, b[$. En effet si $a \leq x' \leq x'' < b$,

$$F(x'') - F(x') = \int_{x'}^{x''} f(t) dt \geq 0, \quad (\text{B.20})$$

par positivité de f sur $[x', x'']$. Il n'y a donc *que deux possibilités* pour le comportement de $F(x)$ quand x tend vers b à gauche.

1. La fonction croissante F est *majorée* sur $[a, b[$, i.e. il existe $M \in \mathbb{R}_+$ tel que pour tout $x \in [a, b[$, $F(x) \leq M < +\infty$. Alors F a une limite finie ℓ à gauche en b ($\ell \leq M$), autrement dit l'intégrale généralisée $\int_a^b f(t) dt$ *converge* et $\int_a^b f(t) dt = \ell$.
2. La fonction croissante F *n'est pas majorée* sur $[a, b[$. Alors F tend vers $+\infty$ à gauche en b , l'intégrale généralisée $\int_a^b f(t) dt$ *diverge* et $\int_a^b f(t) dt = +\infty$.

Rappelons ici qu'il existe des intégrales divergentes auxquelles on ne peut attribuer aucune valeur, pas même infinie, voir l'exemple B.3. Comme nous venons de le voir, ce type de divergence ne peut se produire lorsque f est de signe constant.

Théorème B.29 (comparaison). *Soient $f, g \in \mathcal{R}^{\text{loc}}[a, b[$ telles que $0 \leq f \leq g$ sur $[a, b[$. Alors*

$$\int_a^b g(t) dt \text{ converge} \implies \int_a^b f(t) dt \text{ converge}, \quad (\text{B.21})$$

$$\int_a^b f(t) dt \text{ diverge} \implies \int_a^b g(t) dt \text{ diverge}. \quad (\text{B.22})$$

Preuve. Soit x quelconque dans $[a, b[$. Alors f et g sont Riemann intégrables sur $[a, x]$ et en intégrant sur cet intervalle l'inégalité $f \leq g$, on voit que

$$\forall x \in [a, b[, \quad F(x) := \int_a^x f(t) dt \leq \int_a^x g(t) dt =: G(x).$$

Comme f et g sont positives, les fonctions F et G sont croissantes d'après (B.20).

Si $\int_a^x g(t) dt$ converge, cela signifie que G a une limite à gauche finie L en b . La fonction *croissante* G est donc majorée sur $[a, b[$ par L et F l'est aussi puisque $F \leq G$

sur $[a, b]$. Étant croissante sur $[a, b]$ et majorée par $L < +\infty$ sur cet intervalle, F a aussi une limite à gauche finie $\ell \leq L$ en b . Autrement dit, $\int_a^b f(t) dt$ converge. Ceci établit l'implication (B.21).

Si $\int_a^x f(t) dt$ diverge, le point 2 de l'alternative ci-dessus entre en vigueur⁸. Autrement dit, $F(x)$ tend vers $+\infty$ à gauche en b . Il en va de même pour G puisque $F \leq G$. Donc $\int_a^b g(t) dt = +\infty$ et cette intégrale diverge. L'implication (B.21) est ainsi vérifiée. \square

Exemple B.30. Les intégrales « gaussiennes » $\int_{-\infty}^{+\infty} e^{-ct^2} dt$, $c > 0$, convergent.

Par réduction du problème, cf. remarque B.12, on se ramène à l'étude séparée de $\int_{-\infty}^0$ et de $\int_0^{+\infty}$. Par parité de l'intégrande $f : t \mapsto \exp(-ct^2)$, il est clair qu'il suffit d'étudier la convergence de l'intégrale généralisée de f sur $[0, +\infty[$. Notons au passage que f est continue sur \mathbb{R} , donc clairement membre de $\mathcal{R}^{\text{loc}}[-\infty, 0]$ et de $\mathcal{R}^{\text{loc}}[0, +\infty[$. Puisque f est en particulier Riemann intégrable sur $[0, 1]$, le découpage $\int_0^x f(t) dt = \int_0^1 f(t) dt + \int_1^x f(t) dt$ nous ramène finalement à l'étude de la convergence de $\int_1^{+\infty} f(t) dt$.

Cette dernière réduction est motivée par l'inégalité $t^2 \geq t$ pour $t \geq 1$. Par positivité de c et croissance de la fonction exponentielle, on en déduit $-ct^2 \leq -ct$ et $f(t) = \exp(-ct^2) \leq \exp(-ct) =: g(t)$. Nous pouvons alors appliquer l'implication (B.21) pour conclure à la convergence de $\int_1^{+\infty} f(t) dt$. En effet

$$\int_1^x \exp(-ct) dt = \left[\frac{\exp(-ct)}{-c} \right]_1^x = \frac{\exp(-c)}{c} - \frac{\exp(-cx)}{c} \xrightarrow{x \rightarrow +\infty} \frac{\exp(-c)}{c},$$

donc $\int_1^{+\infty} g(t) dt$ converge.

Corollaire B.31. Soient $f, g \in \mathcal{R}^{\text{loc}}[a, b]$. On suppose que $|f| \leq g$ sur $[c, b]$ pour un $c \in [a, b]$ et que $\int_c^b g(t) dt$ converge. Alors l'intégrale généralisée $\int_a^b f(t) dt$ est absolument convergente.

Preuve. Les fonctions f et g (donc aussi leur valeur absolue) sont Riemann intégrables sur $[a, c]$. Le découpage $\int_a^x = \int_a^c + \int_c^x$ montre alors que la convergence absolue de $\int_a^b f(t) dt$ équivaut à celle de $\int_c^b f(t) dt$. Cette dernière convergence découle immédiatement de celle de $\int_c^b g(t) dt$ par (B.21) appliqué sur $[c, b]$ au lieu de $[a, b]$, avec $|f|$ à la place de f . \square

Exemple B.32. L'intégrale $\int_1^{+\infty} \frac{\sin(t \cos t)}{t^2} dt$ converge absolument.

C'est une application immédiate du corollaire B.31 avec $a = c = 1$ et $g(t) = t^{-2}$.

Théorème B.33 (intégrandes équivalentes). Soient $f, g \in \mathcal{R}^{\text{loc}}[a, b]$, positives au voisinage à gauche de b . Si elles sont équivalentes en $b-$, $\int_a^b f(t) dt$ et $\int_a^b g(t) dt$ sont de même nature.

8. Parce que f est positive sur $[a, b]$.

Preuve. Rappelons que « f équivalente à g en $b-$ » noté encore $f \sim_{b-} g$ signifie qu'il existe un réel $c \in [a, b[$ et une fonction h définie que $[c, b[$ telle que

$$\forall t \in [c, b[, \quad f(t) = g(t)h(t) \quad \text{et} \quad \lim_{t \rightarrow b-} h(t) = 1.$$

Cette limite à gauche de h en b nous permet de trouver un $d \in [c, b[$ tel que l'encadrement $1/2 \leq h(t) \leq 3/2$ soit vérifié⁹ pour tout $t \in [d, b[$. Par positivité de f et g au voisinage de b et quitte à remplacer d par $d' \in [d, b[$, on se ramène au cas où g est positive sur $[d, b[$. On a alors

$$\forall t \in [d, b[, \quad 0 \leq \frac{g(t)}{2} \leq f(t) = g(t)h(t) \leq \frac{3g(t)}{2}, \quad (\text{B.23})$$

d'où l'on tire

$$\forall x \in [d, b[, \quad \frac{1}{2} \int_d^x g(t) dt \leq \int_d^x f(t) dt \leq \frac{3}{2} \int_d^x g(t) dt. \quad (\text{B.24})$$

Supposons que $\int_a^b g(t) dt$ diverge. Comme $\int_a^d g(t) dt$ est une constante finie, $\int_d^b g(t) dt$ diverge aussi. Par *positivité* de g sur $[d, b[$, $\int_d^x g(t) dt$ tend alors vers $+\infty$ quand x tend vers b à gauche. Il en va de même pour $\int_d^x f(t) dt$ à cause de la première inégalité dans (B.24). Par addition de la constante $\int_a^d f(t) dt$ on voit finalement que $\int_a^x f(t) dt$ tend vers $+\infty$ en $b-$. Ainsi la divergence de $\int_a^b g(t) dt$ implique celle de $\int_a^b f(t) dt$.

Si $\int_a^b g(t) dt$ converge, l'ensemble $\{ \int_d^x g(t) dt; x \in [d, b[\}$ est majoré et la deuxième inégalité dans (B.24), montre qu'il en va de même pour $\{ \int_d^x f(t) dt; x \in [d, b[\}$. On en déduit facilement que $\int_a^b f(t) dt$ converge en utilisant la *positivité* sur $[d, b[$ de f qui résulte de (B.23). \square

Corollaire B.34. Soient $f, g \in \mathcal{R}^{\text{loc}}[a, b[$ telles que g soit strictement positive sur un voisinage à gauche de b et que

$$\lim_{t \rightarrow b-} \frac{f(t)}{g(t)} = K, \quad K \in]0, +\infty[. \quad (\text{B.25})$$

Alors $\int_a^b f(t) dt$ converge si et seulement si $\int_a^b g(t) dt$ converge.

Preuve. La positivité stricte de g sur un voisinage à gauche de b et (B.25) impliquent que f et Kg sont toutes deux positives sur un même voisinage $[c, b[$ de b et que f et Kg sont équivalentes en $b-$. On conclut en appliquant le théorème B.33 à f et Kg . \square

Exemple B.35. $I := \int_0^1 t^\alpha (1-t)^\beta dt$ converge si et seulement si $\alpha > -1$ et $\beta > -1$.

L'intégrande $f : t \mapsto t^\alpha (1-t)^\beta$ est toujours continue au moins sur $]0, 1[$, donc f est localement Riemann intégrable sur $]0, 1[$.

9. Appliquer la définition de la limite avec $\varepsilon = 1/2$.

Pour $\alpha \geq 0$ et $\beta \geq 0$, f est continue sur $[0, 1]$ donc Riemann intégrable sur $[0, 1]$. I est alors une intégrale ordinaire. Si $\alpha < 0$ ou $\beta < 0$, I est une intégrale généralisée. Pour étudier sa convergence, on regarde séparément $\int_0^{1/2}$ et $\int_{1/2}^1$. Notons que f est strictement positive sur $]0, 1[$, ce qui nous permet d'utiliser le théorème B.33. On voit ainsi que

$$t^\alpha(1-t)^\beta \underset{0+}{\sim} t^\alpha \quad \text{et} \quad t^\alpha(1-t)^\beta \underset{1-}{\sim} (1-t)^\beta.$$

Compte-tenu du corollaire B.15, on en déduit que I converge si et seulement si $\alpha > -1$ et $\beta > -1$.

Exemple B.36. $I := \int_0^{+\infty} \frac{dt}{1+t^3}$ converge.

On a envie de dire que c'est une application immédiate du théorème B.33 puisque $f(t) := (1+t)^{-3} \simeq t^{-3} =: g(t)$ en $+\infty$. Mais alors on introduit artificiellement un problème en zéro pour g et $\int_0^{+\infty} g(t) dt$ diverge (à cause de la borne 0). En y regardant de plus près, on voit que les hypothèses du théorème B.33 ne sont pas toutes vérifiées puisque g est localement intégrable sur $]0, +\infty[$, mais pas sur $[0, +\infty[$. Notons que f elle, est bien dans $\mathcal{R}^{\text{loc}}[0, +\infty[$ comme fonction continue sur $[0, +\infty[$. On se sort de ce mauvais pas en remarquant que I et $\int_1^{+\infty} f(t) dt$ sont de même nature et en appliquant le théorème B.33 sur l'intervalle $[1, +\infty[$ avec les restrictions de f et g à cet intervalle. En effet $\int_1^{+\infty} t^{-3} dt$ converge par la proposition B.14 a).

On aurait pu aussi prouver par cette méthode la convergence de $\int_0^{+\infty} \frac{dt}{1+t^2}$, mais dans ce cas, il y a bien plus simple puisque $\int_0^x \frac{dt}{1+t^2} = \arctan(x)$ qui tend vers $\pi/2$ lorsque x tend vers $+\infty$. Donc on voit directement que $\int_0^{+\infty} \frac{dt}{1+t^2}$ converge et vaut $\pi/2$.

Exemple B.37. $I := \int_0^{+\infty} \frac{3t+2}{5t^3+t^2+4} dt$ converge.

L'intégrande $f : t \mapsto (3t+2)(5t^3+t^2+4)^{-1}$ est positive et continue sur $[0, +\infty[$ comme quotient de deux fonctions continues car le dénominateur qui est minoré par 4 sur cet intervalle ne s'y annule pas. Ainsi f appartient à $\mathcal{R}^{\text{loc}}[0, +\infty[$. D'autre part en $+\infty$, $f(t)$ est équivalent à $\frac{3}{5}t^{-2} =: g(t)$. On est confronté au même piège qu'à l'exemple B.36 et il faut éviter d'introduire artificiellement un problème en 0 à cause de la divergence de $\int_0^\varepsilon g(t) dt$. Là encore il suffit de se ramener à $\int_1^{+\infty} f(t) dt$ qui converge par comparaison avec $\int_1^{+\infty} t^{-2} dt$.

Exemple B.38. $I := \int_0^{\frac{\pi}{2}} \frac{dt}{(\cos t)^\alpha}$ converge si et seulement si $\alpha < 1$.

Sur $[0, \pi/2]$ la fonction continue cosinus ne s'annule qu'au point $\pi/2$ et est positive ailleurs. Pour $\alpha \leq 0$, l'intégrande $f : t \mapsto (\cos t)^{-\alpha}$ est continue sur $[0, \pi/2]$ et I est une intégrale de Riemann ordinaire. Pour $\alpha > 0$, f est continue sur $[0, \pi/2[$ et tend vers $+\infty$ à gauche en $\pi/2$. Dans ce cas $f \in \mathcal{R}^{\text{loc}}[0, \pi/2[$ et I est une véritable intégrale généralisée. Un développement limité à l'ordre 1 du cosinus au point $\pi/2$ s'écrit :

$$\cos t = \cos \frac{\pi}{2} + \left(-\sin \frac{\pi}{2} \right) \left(t - \frac{\pi}{2} \right) + \left(t - \frac{\pi}{2} \right) \varepsilon \left(t - \frac{\pi}{2} \right), \quad \varepsilon(u) \xrightarrow{u \rightarrow 0} 0,$$

d'où

$$\cos t = \left(\frac{\pi}{2} - t\right) \left(1 - \varepsilon\left(t - \frac{\pi}{2}\right)\right),$$

autrement dit, $\cos t$ a pour équivalent $\pi/2 - t$ en $\pi/2$. On en déduit que

$$f(t) \underset{\frac{\pi}{2}}{\sim} \left(\frac{\pi}{2} - t\right)^{-\alpha} =: g(t).$$

La fonction g étant comme f , continue et positive sur $[0, \pi/2[$, le théorème B.33 s'applique et nous dit que I et $\int_0^{\pi/2} (\pi/2 - t)^{-\alpha} dt$ sont de même nature. Cette dernière intégrale converge si et seulement si $\alpha < 1$ par le corollaire B.15.

Remarque B.39. Le théorème B.33 s'adapte immédiatement au cas où f et g sont toutes deux négatives au voisinage à gauche de b . Par contre et même si f et g sont de même signe au voisinage à gauche de b , le théorème n'est plus valable si f n'a pas un signe constant au voisinage de b -. Le contre exemple suivant devrait vous en convaincre.

Exemple B.40 (à méditer). Définissons $f, g : [1, +\infty[\rightarrow \mathbb{R}$ par

$$f(t) := \frac{\sin t}{\sqrt{t}}, \quad g(t) := \frac{\sin t}{\sqrt{t}} + \frac{\sin^2 t}{t}. \quad (\text{B.26})$$

Alors $f(t) \sim g(t)$ en $+\infty$, f change de signe une infinité de fois au voisinage de $+\infty$, f et g sont de même signe au voisinage de $+\infty$. L'intégrale généralisée $\int_1^{+\infty} f(t) dt$ converge mais $\int_1^{+\infty} g(t) dt$ diverge.

Justifications. Les fonctions f et g sont continues sur $[1, +\infty[$ donc dans $\mathcal{R}^{\text{loc}}[1, +\infty[$. On note d'abord que pour tout $t \in [1, +\infty[$, $g(t) = f(t)h(t)$ avec

$$g(t) = f(t)h(t) \quad \text{avec} \quad h(t) = 1 + \frac{\sin t}{\sqrt{t}} \xrightarrow[t \rightarrow +\infty]{} 1. \quad (\text{B.27})$$

Ceci établit l'équivalence de f et g en $+\infty$.

La fonction f ayant le signe du sinus change de signe une infinité de fois au voisinage de $+\infty$. Il en va de même pour g à cause de (B.27) puisque $h(t)$ est strictement positif¹⁰ sur $[1, +\infty[$. De plus, f et g ont même signe et mêmes zéros sur tout l'intervalle $[1, +\infty[$.

L'intégrale généralisée $\int_1^{+\infty} f(t) dt$ converge par le théorème d'Abel ou la proposition B.27.

Supposons que $\int_1^{+\infty} g(t) dt$ converge, alors nécessairement $\int_1^{+\infty} t^{-1} \sin^2 t dt$ doit converger. En effet

$$\int_1^x \frac{\sin^2 t}{t} dt = \int_1^x g(t) dt - \int_1^x f(t) dt$$

et le second membre doit avoir une limite finie quand x tend vers $+\infty$ en raison de la convergence des deux intégrales $\int_1^{+\infty} f(t) dt$ et $\int_1^{+\infty} g(t) dt$. Nous allons montrer que

10. Il suffirait que h soit strictement positive sur un voisinage $[c, +\infty[$ de $+\infty$, ce qui découle du fait que h a une limite strictement positive en $+\infty$. Mais ici il est plus simple de remarquer que $t^{-1/2} \sin t > -1$ pour $t > 1$ et que $h(1) = \sin 1 > 0$.

l'on aboutit à une *contradiction* en vérifiant *directement* que $\int_1^{+\infty} t^{-1} \sin^2 t \, dt$ diverge. En effet, l'identité $\sin^2 t = \frac{1}{2}(1 - \cos 2t)$ nous donne

$$\int_1^x \frac{\sin^2 t}{t} \, dt = \int_1^x \left(\frac{1}{2t} - \frac{\cos 2t}{2t} \right) \, dt = \left[\frac{1}{2} \ln t \right]_1^x - \frac{1}{2} \int_1^x \frac{\cos 2t}{t} \, dt = \frac{1}{2} \ln x - \frac{1}{2} \int_1^x \frac{\cos 2t}{t} \, dt.$$

Quand x tend vers $+\infty$, $\frac{1}{2} \ln x$ tend vers $+\infty$, tandis que $\frac{1}{2} \int_1^x \frac{\cos 2t}{t} \, dt$ tend vers une limite finie car $\frac{1}{2} \int_1^{+\infty} \frac{\cos 2t}{t} \, dt$ converge grâce au théorème d'Abel ou à la proposition B.27 (poser $s = 2t$). Donc $\int_1^x t^{-1} \sin^2 t \, dt$ tend vers $+\infty$ avec x , autrement dit $\int_1^{+\infty} t^{-1} \sin^2 t \, dt$ diverge, ce qui établit la contradiction annoncée et impose la divergence de $\int_1^{+\infty} g(t) \, dt$. \square

B.4 Divers

B.4.1 Changements de variable

Nous examinons l'extension des formules de changement de variable au cas des intégrales généralisées. *Grosso modo* tout se passe bien lorsque l'on utilise un changement de variable *monotone*. Si ce n'est pas le cas, il convient d'être prudent et de revenir à la définition de l'intégrale généralisée $\int_a^b = \lim_{x \rightarrow b-} \int_a^x$ pour appliquer le changement de variable aux intégrales de Riemann ordinaires \int_a^x avant de faire tendre x vers $b-$.

Proposition B.41 (translation et changement d'échelle).

- i) *Translation.* Soient $c \in \mathbb{R}$ et f localement Riemann intégrable sur $[a+c, b+c]$, avec $b+c := +\infty$ si $b = +\infty$. Alors l'application $g : [a, b] \rightarrow \mathbb{R}$, $t \mapsto f(t+c)$ est localement Riemann intégrable sur $[a, b]$. Les intégrales $\int_a^b g(t) \, dt$ et $\int_{a+c}^{b+c} f(s) \, ds$ sont de même nature. Si l'une des deux converge on a

$$\int_a^b f(t+c) \, dt = \int_{a+c}^{b+c} f(s) \, ds. \quad (\text{B.28})$$

Cette égalité reste vraie dans $\overline{\mathbb{R}}_+$ sans hypothèse de convergence si f ou g est positive sur son intervalle d'intégration.

- ii) *Changement d'échelle.* Soient $c \in \mathbb{R}^*$ et f localement Riemann intégrable sur l'intervalle d'extrémités¹¹ ac et bc , semi-fermé en ac , avec dans le cas où $b = +\infty$, $bc := +\infty$ si $c > 0$, $bc := -\infty$ si $c < 0$. Alors l'application $h : [a, b] \rightarrow \mathbb{R}$, $t \mapsto f(ct)$ est localement Riemann intégrable sur $[a, b]$. Les intégrales $\int_a^b h(t) \, dt$ et $\int_{ac}^{bc} f(s) \, ds$ sont de même nature. Si l'une des deux converge on a

$$\int_a^b f(ct) \, dt = \frac{1}{c} \int_{ac}^{bc} f(s) \, ds. \quad (\text{B.29})$$

Cette égalité reste vraie dans $\overline{\mathbb{R}}_+$ sans hypothèse de convergence si f ou h est positive sur son intervalle d'intégration.

11. Il s'agit de $[ac, bc[$ si $c > 0$ et de $]bc, ac]$ si $c < 0$.

Remarquons qu'il n'y a ici aucune hypothèse de continuité sur f pour ces formules de changement de variable par translation ou changement d'échelle dans les intégrales généralisées. On peut donc les appliquer notamment avec des fonctions décroissantes positives qui peuvent avoir une infinité de discontinuités, mais sont toujours localement Riemann intégrables.

Preuve de i). Puisque $f \in \mathcal{R}^{\text{loc}}[a+c, b+c[$, elle est Riemann intégrable sur le segment $[a+c, x+c]$ pour tout $x \in [a, b[$. Alors par la proposition A.32 i), g est Riemann intégrable sur $[a, x]$ et ceci valant pour tout $x \in [a, b[$, g est bien dans $\mathcal{R}^{\text{loc}}[a, b[$. De plus on a par la formule de changement de variable (A.46) :

$$\forall x \in [a, b[, \quad \int_a^x g(t) dt = \int_a^x f(t+c) dt = \int_{a+c}^{x+c} f(s) ds.$$

En faisant tendre x vers $b-$, on en déduit que $\int_a^b g(t) dt$ et $\int_{a+c}^{b+c} f(s) ds$ sont de même nature. Si l'une des deux intégrales généralisées converge, cela signifie que l'intégrale de Riemann ordinaire correspondante ci-dessus a une limite dans \mathbb{R} quand x tend vers $b-$. En raison de l'égalité, il en va de même pour l'autre intégrale et les limites sont égales, ce qui nous donne (B.28). Si f ou g est positive, ces deux intégrales dépendant de x ont toujours une limite dans $\overline{\mathbb{R}}_+$ et les limites sont égales. \square

Preuve de ii). La preuve est analogue à celle de i) à quelques alourdissements d'écriture près que l'auteur abandonne lâchement au lecteur. \square

Voici maintenant une extension partielle¹² aux intégrales généralisées du changement de variable « classique » de la proposition A.32 iii).

Proposition B.42 (changement de variable C^1 monotone). *Soit $\varphi : [a, b[\rightarrow \mathbb{R}$, une fonction monotone ayant une dérivée continue sur $[a, b[$. On suppose de plus que φ est strictement monotone au voisinage à gauche de b . Pour toute fonction f continue sur l'intervalle $\varphi([a, b[)$, les deux intégrales généralisées ci-dessous sont de même nature et si l'une converge on a*

$$\int_a^b f(\varphi(t)) \varphi'(t) dt = \int_{\varphi(a)}^{\varphi(b-)} f(s) ds, \quad (\text{B.30})$$

où $\varphi(b-) \in \overline{\mathbb{R}}$ désigne la limite à gauche de φ en b . Cette égalité demeure vraie dans $\overline{\mathbb{R}}$ sans condition de convergence si f est de signe constant sur l'intervalle $\varphi([a, b[)$.

Notons $h := (f \circ \varphi) \varphi'$. La condition sur la stricte monotonie de φ à gauche de b est là pour écarter un cas artificiel où l'intégrale $\int_a^b h(t) dt$ est une intégrale de Riemann ordinaire d'une fonction continue sur un intervalle $[a, c]$. En effet si φ est constante sur $[c, b[$ pour un $c \in [a, b[$, on voit que $\int_a^x h(t) dt = \int_a^c h(t) dt$ pour tout $x \in [c, b[$ en raison de la nullité de φ' sur $[c, b[$. En faisant tendre x vers $b-$, cette égalité nous

12. On notera l'hypothèse plus restrictive sur le changement de variable φ .

donne $\int_a^b h(t) dt = \int_a^c h(t) dt$. Comme h est continue sur $[a, c]$, le changement de variable classique prop. A.32 iii) nous donne $\int_a^c h(t) dt = \int_{\varphi(a)}^{\varphi(c)} f(s) ds$ et comme $\varphi(b-) = \varphi(c)$, on obtient bien (B.30). On voit ainsi que dans ce cas les deux intégrales dans (B.30) sont de fausses intégrales généralisées puisqu'elles peuvent s'écrire comme intégrales de fonctions continues sur $[a, c]$.

Preuve de la proposition B.42. La fonction φ étant monotone admet une limite à gauche finie ou infinie en b que nous notons $\varphi(b-)$. En raison de la continuité et de la monotonie de φ sur $[a, b[$, stricte au voisinage de $b-$, $\varphi([a, b[)$ est l'intervalle de bornes $\varphi(a)$ et $\varphi(b-)$, fermé en $\varphi(a)$ et ouvert en $\varphi(b-)$. Nous traitons le cas où φ est décroissante, l'adaptation au cas où elle est croissante étant immédiate. On a alors $\varphi([a, b[) =]\varphi(b-), \varphi(a)]$. Par application de la prop. A.32 iii), on a

$$\forall x \in [a, b[, \quad \int_a^x f(\varphi(t))\varphi'(t) dt = \int_{\varphi(a)}^{\varphi(x)} f(s) ds = - \int_{\varphi(x)}^{\varphi(a)} f(s) ds. \quad (\text{B.31})$$

Les fonctions h et f étant continues l'une sur $[a, b[$ et l'autre sur $]\varphi(b-), \varphi(a)]$ appartiennent respectivement à $\mathcal{R}^{\text{loc}}[a, b[$ et $\mathcal{R}^{\text{loc}}]\varphi(b-), \varphi(a)]$. En faisant tendre x vers $b-$ dans (B.31), et en notant que par continuité et décroissance de φ , $\varphi(x)$ tend alors vers $\varphi(b-)$ par la droite, on voit que les intégrales généralisées $\int_a^b h(t) dt$ et $\int_{\varphi(a)}^{\varphi(b-)} f(s) ds$ sont de même nature. Si l'une des deux converge, on en déduit en se souvenant de la définition B.11 :

$$\int_a^b f(\varphi(t))\varphi'(t) dt = - \int_{\varphi(b-)}^{\varphi(a)} f(s) ds = \int_{\varphi(a)}^{\varphi(b-)} f(s) ds,$$

ce qui nous donne (B.30). D'autre part si f est positive, h est négative car φ est décroissante donc $\varphi' \leq 0$. Alors dans (B.31) les intégrales $\int_a^x h(t) dt$ et $-\int_{\varphi(x)}^{\varphi(a)} f(s) ds$ sont des fonctions négatives et décroissantes de x donc convergent dans $\overline{\mathbb{R}}$ quand x tend vers $b-$, soit vers un réel négatif, soit vers $-\infty$ et l'égalité (B.31) se conserve par passage à la limite. Si f est négative, h est positive car $\varphi' \leq 0$. Alors les intégrales de (B.31) sont des fonctions positives et croissantes de la variable x et elles restent égales à la limite (dans $\overline{\mathbb{R}}_+$) quand x tend vers $b-$. \square

Exemple B.43. L'intégrale généralisée $I := \int_0^{+\infty} \sin(t^2) dt$ converge¹³.

En effet par le changement de variable croissant et C^1 , $\varphi : t \mapsto t^2$, l'intervalle $[0, +\infty[$ a pour image $[0, +\infty[$ et I est de même nature que

$$J := \int_0^{+\infty} \frac{\sin s}{2\sqrt{s}} ds.$$

L'intégrale J converge par la proposition B.27, donc I converge. De plus on a alors $I = J$ par l'égalité (B.30).

13. Je sais, cela surprend, surtout si on compare avec l'exemple B.3.

Exemple B.44 (un changement de variable illicite). Voici un exemple où un changement de variable C^1 non monotone appliqué sans précaution conduit à une erreur. Dans l'intégrale généralisée $I := \int_0^{2\pi} \frac{\sin t}{\cos t} dt$, on pose $s = \cos t$. On obtient alors l'intégrale de Riemann ordinaire $J := \int_1^{-1} \frac{-ds}{s} = 0$. L'égalité $I = J$ ne peut être valide ici car I diverge. En effet la fonction tangente est localement Riemann intégrable sur l'intervalle troué $[0, 2\pi] \setminus \{\pi/2, 3\pi/2\}$ et nous devons considérer séparément chacune des intégrales $\int_0^{\pi/2}$, $\int_{\pi/2}^{3\pi/2}$ et $\int_{3\pi/2}^{2\pi}$. L'intégrale $\int_0^{\pi/2} \tan t dt$ diverge car la fonction tangente est positive sur $[0, \pi/2[$ et équivalente à $1/\cos t$ au voisinage à gauche de $\pi/2$. Or on sait par l'exemple B.38 que $\int_0^{\pi/2} \frac{dt}{\cos t}$ diverge¹⁴.

B.4.2 Intégration par parties

Il n'y a pas d'extension automatique de la règle d'intégration par parties (i.p.p.) des intégrales de Riemann ordinaires (calculables par primitivation) aux intégrales généralisées. Lorsque l'on effectue une intégration par parties sur une intégrale généralisée, tout peut arriver :

1. transformation d'une intégrale absolument convergente en intégrale absolument convergente ;
2. transformation d'une intégrale absolument convergente en intégrale convergente mais pas absolument et *vice versa* ;
3. transformation d'une intégrale absolument convergente en intégrale divergente.

En pratique pour effectuer une intégration par parties sur $\int_a^b f(t) dt$ avec $f \in \mathcal{R}^{\text{loc}}[a, b[$ (en fait avec $f \in C[a, b[$), on l'effectue d'abord sur l'intégrale de Riemann ordinaire $\int_a^x f(t) dt$ pour x quelconque dans $[a, b[$ avant de regarder ce qui se passe lorsque l'on fait tendre x vers $b-$.

Voici quelques exemples illustrant les différentes situations possibles.

Exemple B.45. Calcul de $I := \int_0^{+\infty} te^{-t} dt$ par i.p.p.

L'intégrande est une fonction positive et continue sur $[0, +\infty[$. Comme $te^{-t/2}$ tend vers 0 en $+\infty$, cette quantité est majorée pour $t \geq t_0$ par une constante M . On a alors pour $t \geq t_0$, $te^{-t} \leq Me^{-t/2} =: g(t)$ et comme $\int_0^{+\infty} g(t) dt$ converge (évident par primitivation), le théorème de comparaison (th. B.29) nous donne la convergence de I .

On effectue l'i.p.p. sur $I(x) := \int_0^x te^{-t} dt$ en posant

$$u(t) = t, \quad v'(t) = e^{-t}, \quad u'(t) = 1, \quad v(t) = -e^{-t},$$

d'où

$$\forall x \in [0, +\infty[, \quad I(x) = [-te^{-t}]_0^x - \int_0^x (-e^{-t}) dt = xe^{-x} + \int_0^x e^{-t} dt.$$

14. On pourrait contester cet exemple, car dans le contexte de la proposition B.42, avant d'envisager un changement de variable dans l'intégrale I , il convient de vérifier que l'intégrande est localement intégrable sur $[0, 2\pi[$, ce qui n'est pas le cas ici.

Quand x tend vers $+\infty$, xe^{-x} tend vers 0 et $\int_0^x e^{-t} dt$ tend vers l'intégrale généralisée convergente $\int_0^{+\infty} e^{-t} dt$, laquelle se calcule d'ailleurs immédiatement par primitivation (exemple B.1) et vaut 1. Finalement l'i.p.p. nous permet ici de retrouver la convergence de I et de calculer sa valeur : $I = 1$.

Exemple B.46. *Intégration par parties de $I := \int_{\pi}^{+\infty} \frac{\sin^2 t}{t^2} dt$.*

Ici l'intégrande f est localement Riemann intégrable sur $[\pi, +\infty[$, comme fonction continue sur cet intervalle. La convergence de I résulte de l'inégalité $0 \leq f(t) \leq t^{-2}$ et de la convergence de $\int_{\pi}^{+\infty} t^{-2} dt$, par le théorème de comparaison. Ainsi I est une intégrale généralisée absolument convergente.

On effectue l'i.p.p. sur $I(x) := \int_0^x f(t) dt$ en posant

$$u(t) = \sin^2 t, \quad v'(t) = t^{-2}, \quad u'(t) = 2 \sin t \cos t = \sin(2t), \quad v(t) = -t^{-1},$$

ce qui nous donne

$$\forall x \geq \pi, \quad I(x) = \left[-\frac{\sin^2 t}{t} \right]_{\pi}^x + \int_{\pi}^x \frac{\sin(2t)}{t} dt = -\frac{\sin^2 x}{x} + \int_{2\pi}^{2x} \frac{\sin s}{s} ds.$$

Faisons tendre x vers $+\infty$, alors au premier membre $I(x)$ tend vers I puisque l'on sait déjà que I converge. Au second membre $-x^{-1} \sin^2 x$ tend vers 0. On en déduit que $J(x) := \int_{2\pi}^{2x} s^{-1} \sin s ds$ tend vers une limite finie égale à I . Ceci prouve que l'intégrale généralisée

$$J := \int_{2\pi}^{+\infty} \frac{\sin s}{s} ds$$

converge et est égale à I . On sait par ailleurs que J n'est pas absolument convergente, voir l'exemple B.28. Ici l'intégration par parties a transformé une intégrale absolument convergente I en une intégrale J convergente mais pas absolument. Une autre i.p.p. partant de J donnerait

$$J = \frac{1}{2\pi} + \int_{2\pi}^{+\infty} \frac{\cos s}{s^2} ds,$$

nous fournissant un exemple de transformation d'une intégrale convergente mais pas absolument, en intégrale absolument convergente.

Exemple B.47. *Tentative d'i.p.p. sur $I := \int_0^{\pi/2} \frac{\sin t}{t^{3/2}} dt$*

L'intégrande f est dans $\mathcal{R}^{\text{loc}}]0, \pi/2]$ et positive. L'intégrale converge en 0 grâce à la majoration $0 \leq \sin t \leq t$ d'où $f(t) \leq t^{-1/2}$ valable pour tout $t \in]0, \pi/2]$. L'intégrale généralisée I est donc absolument convergente. Intégrons par parties $I(x) := \int_x^{\pi/2} f(t) dt$ en posant :

$$u(t) = t^{-3/2}, \quad v'(t) = \sin t, \quad u'(t) = -\frac{3}{2}t^{-5/2}, \quad v(t) = -\cos t,$$

d'où

$$I(x) = \left[\frac{-\cos t}{t^{3/2}} \right]_x^{\pi/2} - \frac{3}{2} \int_x^{\pi/2} \frac{\cos t}{t^{5/2}} dt = \frac{\cos x}{x^{3/2}} - \frac{3}{2} \int_x^{\pi/2} \frac{\cos t}{t^{5/2}} dt.$$

Si on fait tendre x vers $0+$, on obtient une forme indéterminée du type « $\infty - \infty$ », car l'intégrale généralisée

$$J := \int_0^{\pi/2} \frac{\cos t}{t^{5/2}} dt$$

est divergente et vaut $+\infty$ (justifiez !). On a là un exemple d'une intégration par parties sur une intégrale absolument convergente qui fait apparaître une intégrale divergente.

Ceci dit l'i.p.p. ci-dessus n'est pas complètement inutile. Elle permet en effet de donner une « vitesse de divergence » de J . En effet puisque $I(x)$ a une limite finie I en $0+$, on en déduit que

$$J(x) = \int_x^{\pi/2} \frac{\cos t}{t^{5/2}} dt \underset{0+}{\sim} \frac{2 \cos x}{3x^{3/2}} \underset{0+}{\sim} \frac{2}{3x^{3/2}}.$$

B.4.3 Comparaison des intégrales ordinaires et généralisées

On examine maintenant quelles sont les propriétés de l'intégrale de Riemann ordinaire qui passent à l'intégrale généralisées.

L'intégrale généralisée hérite des propriétés suivantes de l'intégrale de Riemann ordinaire, à condition de remplacer l'hypothèse « $f, g \in \mathcal{R}[a, b]$ » par « $f, g \in \mathcal{R}^{\text{loc}}[a, b[$ et les intégrales généralisées $\int_a^b f(t) dt$ et $\int_a^b g(t) dt$ convergent ». Nous laissons le soin au lecteur de vérifier ces propriétés en les appliquant d'abord à \int_a^x avant de faire tendre x vers $b-$.

- Additivité, voir prop. A.18.
- Linéarité, voir prop. A.20.
- Positivité et croissance, voir les points i) et ii) de la proposition A.21.
- L'additivité relative aux intervalles, cf. prop. A.28 en notant que la convergence de $\int_a^b f(t) dt$ implique celle de $\int_a^c f(t) dt$ et de $\int_c^b f(t) dt$, pour tout $c \in]a, b[$.
- La relation de Chasles prop. A.29, à condition que chacune des trois intégrales concernées soit convergente. En effet en prenant c extérieur à $[a, b]$, on risque de faire apparaître une intégrale divergente.

Voyons maintenant les propriétés qui ne passent pas de l'intégrale ordinaire à l'intégrale généralisée. Il s'agit essentiellement de ce qui concerne la valeur absolue et le produit.

Rappelons d'abord que la Riemann intégrabilité de f implique celle de $|f|$ et que la réciproque est fautive, cf. remarque A.23. On en déduit immédiatement que si f est *localement* Riemann intégrable sur $[a, b[$, $|f|$ l'est aussi. Par contre pour $f \in \mathcal{R}^{\text{loc}}[a, b[$, la convergence de l'intégrale généralisée $\int_a^b f(t) dt$ *n'implique pas* celle de $\int_a^b |f|(t) dt$, voir l'exemple B.28. Dans le même ordre d'idées, la Riemann intégrabilité locale de f sur $[a, b[$ implique celle de f^+ et de f^- , mais la convergence de $\int_a^b f(t) dt$ *n'implique pas* celle de $\int_a^b f^+(t) dt$ et $\int_a^b f^-(t) dt$. L'exemple B.28 avec $f(t) = t^{-1} \sin t$ sert aussi de

contre exemple ici (vérification laissée en exercice). Par contre si $\int_a^b f(t) dt$ est *absolument* convergente, $\int_a^b f^+(t) dt$ et $\int_a^b f^-(t) dt$ sont convergentes et on a

$$\int_a^b |f|(t) dt = \int_a^b f^+(t) dt + \int_a^b f^-(t) dt, \quad \int_a^b f(t) dt = \int_a^b f^+(t) dt - \int_a^b f^-(t) dt,$$

ainsi que

$$\left| \int_a^b f(t) dt \right| \leq \int_a^b |f(t)| dt < +\infty. \quad (\text{B.32})$$

Regardons maintenant le produit. Si $f, g \in \mathcal{R}^{\text{loc}}[a, b[$, il découle immédiatement de la proposition A.26 que leur produit fg est lui aussi dans $\mathcal{R}^{\text{loc}}[a, b[$. Par contre, la convergence, même absolue, des intégrales généralisées $\int_a^b f(t) dt$ et $\int_a^b g(t) dt$ *n'implique pas* celle de $\int_a^b f(t)g(t) dt$. Un contre exemple immédiat est avec $[a, b[= [0, 1[$, $f(t) = g(t) = (1 - t)^{-1/2}$. On peut néanmoins obtenir la convergence de $\int_a^b f(t)g(t) dt$ à partir de celle des intégrales généralisées non pas de f et g mais de f^2 et g^2 .

Proposition B.48 (inégalité de Cauchy-Schwarz). *Si $f, g \in \mathcal{R}^{\text{loc}}[a, b[$ sont telles que $\int_a^b f(t)^2 dt$ et $\int_a^b g(t)^2 dt$ convergent, $\int_a^b f(t)g(t) dt$ converge absolument et vérifie*

$$\left| \int_a^b f(t)g(t) dt \right| \leq \left\{ \int_a^b f(t)^2 dt \right\}^{1/2} \left\{ \int_a^b g(t)^2 dt \right\}^{1/2}. \quad (\text{B.33})$$

Preuve. D'abord, puisque f, g sont dans $\mathcal{R}^{\text{loc}}[a, b[$, il en va de même pour $|f|$, $|g|$ et leur produit $|f| \cdot |g| = |fg|$. En appliquant l'inégalité de Cauchy Schwarz pour les intégrales ordinaires, on en déduit que

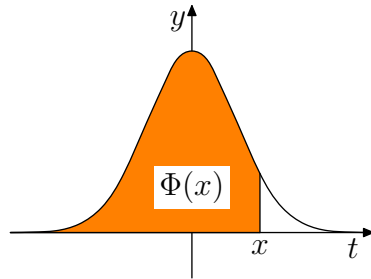
$$\forall x \in [a, b[, \quad \int_a^x |f(t)g(t)| dt \leq \left\{ \int_a^x f(t)^2 dt \right\}^{1/2} \left\{ \int_a^x g(t)^2 dt \right\}^{1/2}.$$

Toutes ces intégrales sont des fonctions croissantes de x . En faisant tendre x vers $b-$, elles convergent toutes dans $\overline{\mathbb{R}}_+$. Compte tenu de l'hypothèse convergence de $\int_a^b f(t)^2 dt$ et $\int_a^b g(t)^2 dt$, on en déduit que

$$\int_a^b |f(t)g(t)| dt \leq \left\{ \int_a^b f(t)^2 dt \right\}^{1/2} \left\{ \int_a^b g(t)^2 dt \right\}^{1/2} < +\infty.$$

Ceci montre que $\int_a^b f(t)g(t) dt$ converge absolument et on conclut en appliquant l'inégalité (B.32) à la fonction fg . \square

Table des valeurs de Φ , f.d.r. de la loi normale standard $\mathfrak{N}(0, 1)$



x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5754
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6627	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7122	0.7156	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7356	0.7389	0.7421	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7793	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8079	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8414	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8622
1.1	0.8643	0.8665	0.8687	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9083	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9193	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9485	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9648	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9895	0.9898	0.9901	0.9903	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9926	0.9928	0.9930	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9944	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Table pour les grandes valeurs de x

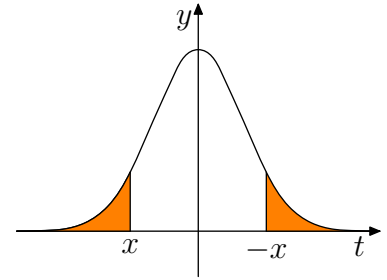
x	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.8	4.0	4.5
$\Phi(x)$	0.99865	0.99904	0.99931	0.99952	0.99966	0.99976	0.999841	0.999928	0.999968	0.999997

La table donne les valeurs de $\Phi(x)$ pour x positif. Lorsque x est négatif, on utilise la relation

$$\Phi(x) = 1 - \Phi(-x)$$

qui résulte de la parité de la densité gaussienne $\mathfrak{N}(0, 1)$.

Exemple : pour $x = -1,8$, on trouve :
 $\Phi(x) = 1 - 0,9641 = 0,0359$.



Pour les « très grandes valeurs de x », (i.e. $|x| \geq 4$), on dispose du résultat suivant qui donne une évaluation de la « queue » de la loi normale.

Pour tout $x > 0$, on a l'encadrement :

$$\left(\frac{1}{x} - \frac{1}{x^3} \right) \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{x^2}{2} \right) < 1 - \Phi(x) < \frac{1}{x} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{x^2}{2} \right).$$

Index

- absence de mémoire, 122
- additivité
 - d'une mesure, 51
 - de l'espérance, 151
 - de l'intégrale de Riemann, 226
- aiguille de Buffon, 208
- aire, 56, 216
 - algébrique, 233
 - d'hypographe, 218, 233
 - et intégrale de Riemann, 218
- arrangement, 12
- Bayes (formule de), 81
- Beppo Levi (théorème de), 141
- bijection, 8
 - réciproque, 8
- borélien, 53
- Borel Cantelli, 194, 195
- Buffon (aiguille de), 208
- cardinal, 8, 16
 - d'un produit cartésien, 10
- coefficient de corrélation, 172
- coefficients
 - binomiaux, 13
 - multinomiaux, 15
- combinaison, 12
- complémentaire, 6
- conditionnement
 - par les cas possibles, 80
- conditionnements successifs, 80
- continuité monotone séquentielle, 63
- convergence
 - dominée, 202
 - en moyenne d'ordre p , 199, 200
 - en probabilité, 196, 197, 202
 - presque complète, 196
 - presque sûre, 192, 193, 196, 198
- convergence en loi
 - binomiale vers Poisson, 112
 - hypergéométrique vers binomiale, 109
- convolution
 - de deux densités, 185
- corrélation, 172
- covariance, 171
 - changement d'échelle, 171
 - de v.a. indépendantes, 189
 - formule de Koenig, 172
 - formules de calcul, 172
 - translation, 171
- critère de Cauchy, 251
 - famille sommable, 41
 - intégrales généralisées, 252
 - séries, 25
- croissance
 - d'une mesure, 51
 - de l'espérance, 136, 152
 - de l'intégrale de Riemann, 228
- dénombrabilité, 16
 - d'un produit cartésien, 18
 - d'une union, 22
 - de \mathbb{N}^2 , 17
 - de $\mathbb{N}^d, \mathbb{Z}^d$, 19
 - de \mathbb{Q} , 19
 - de \mathbb{Z} , 16
 - et famille sommable, 38
 - par image surjective, 23
- dénombrable, 16
- densité
 - marginale, 166
 - somme de v.a.r. indépendantes, 185
 - vecteur aléatoire, 165
 - vecteur aléatoire image, 170

- densité de probabilité
 - sur \mathbb{R} , 102
 - sur \mathbb{R}^d , 165
- dyadique (nombre), 21
- écart type, 158
- ensemble
 - au plus dénombrable, 16
 - dénombrable, 16
 - fini, 8
- épreuves indépendantes, 88
- équiprobabilité, 70
- espace
 - de Banach, 25
- espérance
 - croissance, 136
 - d'une constante positive, 132
 - d'une indicatrice, 132
 - d'une v.a. discrète, 150
 - d'une v.a. discrète positive, 145
 - d'une v.a. positive, 130
 - d'une v.a. positive à densité, 135
 - d'une v.a. positive simple, 134
 - d'une v.a. réelle, 147
 - d'une v.a. réelle à densité, 148
 - intersion série-espérance, 144
 - linéarité, 151
 - produit de v.a. indépendantes, 187
- famille sommable, 35
 - absolument, 43
 - combinaison linéaire, 36
 - critère de Cauchy, 41
 - normalement, 43
 - permutation d'indices, 36
 - sommation par paquets, 44, 45
- fonction
 - absolument continue, 103
 - convexe, 199
 - en escalier, 219
 - localement Riemann intégrable, 245
 - réglée, 225
 - Riemann intégrable, 214
- fonction de répartition
 - d'une probabilité sur \mathbb{R} , 72
 - d'une v.a. discrète, 100
 - d'une variable aléatoire, 99
- fonction de survie, 122
- formule
 - de Bayes, 81
 - de Koenig
 - covariance, 172
 - variance, 158
 - des accroissements finis, 222
 - du binôme, 15
 - du multinôme, 13
- hypographe, 218
- i.i.d., 206
- inclusion, 6
- indépendance
 - conservation par complémentaire, 87
 - covariance, 189
 - d'une suite d'évènements, 88
 - d'une suite de v.a., 178
 - de n variables aléatoires, 173
 - de n vecteurs aléatoires, 174
 - de deux évènements, 85
 - de sous-tribus, 177
 - des composantes (cas à densité), 179
 - des composantes (cas discret), 178
 - deux à deux, 87
 - espérance de produits, 186
 - hérédité, 175
 - mutuelle, 87
 - somme de v.a., 183, 185
- indicatrice, 11
 - dans formules explicites, 102
- inégalité de Bienaymé-Tchebycheff, 205
- inégalité de Cauchy-Schwarz
 - espérance, 171
 - intégrale de Riemann, 231
 - intégrale généralisée, 268
- inégalité de Markov, 137
 - avec moment, 153
- injection, 7
- intégrabilité

- d'une v.a. discrète, 150
- d'une v.a. réelle, 146
- d'une v.a. réelle à densité, 148
- intégrable
 - variable aléatoire positive, 132
 - variable aléatoire réelle, 146
- intégrale de Riemann
 - additivité, 226
 - changement de variable, 235, 236
 - croissance, 228
 - de a à b ($a > b$), 216
 - et primitive, 222
 - inférieure, 214
 - linéarité, 228
 - positivité, 228
 - relation de Chasles, 232
 - supérieure, 214
 - sur $[a, b]$ ($a < b$), 214
- intégrale généralisée, 245
 - absolument convergente, 253, 254, 258
 - additivité, 267
 - changement de variable, 262, 263
 - comparaison, 257
 - convergente, 246
 - critère d'Abel, 254
 - critère de Cauchy, 252
 - croissance, 267
 - divergente, 246
 - équivalents, 258, 261
 - inégalité de Cauchy-Schwarz, 268
 - linéarité, 267
 - positivité, 267
 - produit, 268
 - relation de Chasles, 267
- intégrale indéfinie, 234
- intersection, 6
- intervalle troué, 245
- intersion
 - limite espérance, 141, 202
 - limite intégrale, 238, 239
 - produit espérance, 187
 - série espérance, 144
- inverse ensembliste, 94
- jacobien, 170
- lemme
 - Borel Cantelli, 194, 195
- loi
 - à densité, 102
 - binomiale, 88, 107
 - conditionnelle, 97
 - continue, 101
 - d'un vecteur aléatoire, 162
 - d'une variable aléatoire discrète, 97
 - d'une variable aléatoire réelle, 96
 - de Bernoulli, 107
 - de Cauchy, 125
 - de Poisson, 111, 184
 - diffuse, 101
 - discrète sur \mathbb{R} , 97
 - exponentielle, 121
 - géométrique, 111
 - gaussienne $\mathfrak{N}(m, \sigma)$, 124
 - hypergéométrique, 108
 - marginale (vecteur aléatoire), 162
 - multinomiale, 164
 - normale, 124
 - tables, 269
 - somme de v.a. discrètes, 183
 - indépendantes, 183
 - uniforme sur un borélien de \mathbb{R}^d , 72
 - uniforme sur un ensemble fini, 107
 - uniforme sur un segment, 71
- loi faible des grands nombres, 205
- loi forte des grands nombres, 206
- masse de Dirac, 54
- mesurable (application), 93
- mesure, 52
 - aire, 56
 - de comptage, 55
 - de Dirac, 54
 - de Lebesgue sur \mathbb{R}^d , 56
 - longueur, 56
 - ponctuelle, 55
 - propriétés, 67
 - série de mesures, 54

- volume, 56
- moment, 152
 - h -moment d'une v.a., 152
 - absolu, 152
 - d'une v.a. à densité, 154
 - d'une v.a. discrète, 153
 - fonctionnel, 152
- n -uplet, 7
- non dénombrabilité
 - de $\mathcal{P}(\mathbb{N})$, $[0, 1]$, \mathbb{R} , \mathbb{C} , 20
 - de $\{0, 1\}^{\mathbb{N}}$, 19
- partie négative (d'un réel), 31
- partie positive (d'un réel), 31
- partition, 80
- permutation, 12
- primitive, 222
- probabilité
 - conditionnelle, 77
 - d'un intervalle, 74
 - produit, 176
- probabilités
 - des causes, 81
 - totales (formule des), 80
- produit \otimes de fonctions, 179
- produit \otimes de probabilités, 176
- produit cartésien, 6
- quantificateurs, 6, 192
- réunion, 6
- Riemann intégrabilité, 214
 - d'un produit, 230
 - de $|f|$, 228
 - de f^+ et f^- , 230
 - locale, 245
 - par convergence uniforme, 224
- Riemann intégrabilité de f
 - bornée continue par morceaux, 223
 - continue, 222
 - en escalier, 220
 - monotone, 221
 - réglée, 225
- semi norme, 230
- série, 24
 - absolument convergente, 25
 - alternée, 28
 - commutativement convergente, 29
 - convergence normale, 25
 - de Bertrand, 27
 - de Riemann, 27
 - double, 46
 - géométrique, 25
 - harmonique, 24
 - produit, 49
 - reste (série convergente), 24
 - somme partielle, 24
 - vectorielle, 24
- série double, 46
 - intersion des sommations, 48
- σ -additivité, 52
- σ -algèbre, 52
- sommation par paquets, 44
 - dans $\overline{\mathbb{R}}_+$, 45
- sommes de Darboux, 213
- subdivision de $[a, b]$, 213
- surjection, 7
 - et dénombrabilité, 23
- temps d'attente, 110
- théorème
 - absence de mémoire, 122
 - caractérisation probabilités sur \mathbb{R} , 75
 - comparaison série intégrale, 26
 - comparaison séries à termes positifs, 26
 - de B. Levi, 141
 - de convergence dominée, 202
 - des séries alternées, 28
 - intersion des séries doubles, 48
 - règle des équivalents (séries), 26
 - série produit, 49
 - sommation par paquets, 44
 - sommation par paquets dans $\overline{\mathbb{R}}_+$, 45
- tribu, 52
 - borélienne, 53
 - engendrée, 53
 - engendrée par une application, 177
 - produit, 176

- sous-tribu, [177](#)
- variable aléatoire
 - étagée, [133](#)
 - de Rademacher, [147](#)
 - discrète, [95](#)
 - positive, [130](#)
 - positive intégrable, [132](#)
 - réelle, [95](#)
 - réelle intégrable, [146](#)
 - simple, [133](#)
- variance, [158](#)
 - changement d'échelle, [159](#)
 - d'une somme, [172](#)
 - formule de Koenig, [158](#)
 - translation, [159](#)
- vecteur aléatoire, [162](#)
 - à densité, [165](#)
 - discret, [164](#)
 - lois marginales, [162](#)