

Intégration et probabilités
L3 MIASHS, Université de Paris 1 Panthéon-Sorbonne

Pascal Gourdel¹

26 décembre 2023

1. Je remercie l'équipe pédagogique de cette année mais également un certain nombre de collègues de l'université Paris 1, pour leurs remarques, discussions, commentaires.
Une partie de ces notes de cours s'inspire librement des notes de Sylvain Rubenthaler de l'université de Nice.

Table des matières

1	Espaces mesurables et espaces mesurés	1
1.1	Cardinaux	1
1.1.1	Rappels	1
1.1.2	Ensembles dénombrables	1
1.1.3	Compléments	3
1.2	Introduction aux espaces mesurables	3
1.2.1	Tribus d'information	4
1.2.2	Tribus au sens mathématique	5
1.2.3	Tribu finies	6
1.2.4	Tribu des boréliens de \mathbb{R}	7
1.2.5	Fonctions mesurables	7
1.3	Mesures	7
1.3.1	Définition	7
1.3.2	Théorèmes d'impossibilité	9
1.3.3	Mesure de Borel sur \mathbb{R}	10
1.4	Application aux fondements de la théorie des probabilités	10
1.4.1	Définitions générales	10
1.4.2	Fonction de répartition	11
1.4.3	Conséquence de la mesurabilité d'une v.a.r.	12
1.5	Compléments sur les tribus et les mesures	12
1.5.1	Retour sur la tribu des boréliens de \mathbb{R}	12
1.5.2	Tribu image et tribu engendrée par une fonction	13
1.5.3	Tribu trace	14
1.5.4	Tribu produit	14
1.5.5	Ensembles négligeables	15
1.5.6	Tribu de Lebesgue	16
1.5.7	Mesure et tribu de Lebesgue dans \mathbb{R}^p	17
1.5.8	Loi de probabilité sur les vecteurs aléatoires	17
1.5.9	Construction de fonctions mesurables	18
1.5.10	Résultats sur $\overline{\mathbb{R}}$	20
2	Intégration	23
2.1	Fonctions étagées	23
2.1.1	Définition	23
2.1.2	Approximation par des fonctions étagées	24
2.1.3	Intégrale d'une fonction étagée positive	24
2.1.4	Le cas particulier de $\Omega = \mathbb{N}$	25
2.1.5	Applications aux variables aléatoires réelles sur Ω fini	25
2.2	Fonctions mesurables et intégrales	26
2.2.1	Intégrales des fonctions mesurables positives	26
2.2.2	Intégrales des fonctions mesurables de signe quelconque.	27
2.2.3	Lien avec l'intégrale de Riemann	29
2.2.4	Ensembles négligeables	31
2.3	Applications aux variables aléatoires discrètes	32
2.3.1	Définition	32
2.3.2	Cas des variables aléatoires discrètes raisonnables	32

2.3.3	Lien entre théorie de la mesure et théorie des probabilités	33
2.3.4	Définition de l'espérance	33
2.4	Cas d'une variable aléatoire réelle générale	34
2.4.1	Généralités	34
2.4.2	Vecteur aléatoire à densité	35
2.4.3	Densités	35
2.4.4	Le théorème de Lebesgue-Radon-Nikodym	36
2.4.5	Le théorème de transfert	37
2.4.6	Inégalités	38
3	Théorèmes limites et applications	41
3.1	Convergence presque partout.	41
3.2	Théorèmes de convergence pour les intégrales.	42
3.2.1	Théorème de convergence monotone.	42
3.2.2	Lemme de Fatou	43
3.2.3	Théorème de convergence dominée	44
3.3	Conséquences des théorèmes de convergence déjà connues	45
3.3.1	Lien avec les intégrales impropres dans le cas positif	45
3.3.2	Retour sur les fonctions dénombrablement étagées	45
3.4	Intégrales dépendant d'un paramètre	46
3.4.1	Position du problème	46
3.4.2	Continuité sous l'intégrale	46
3.4.3	Dérivation sous l'intégrale	47
3.5	Applications	48
3.5.1	Fonctions caractéristiques	48
3.5.2	Fonctions génératrices (partie non traitée cette année)	49
4	Mesure produit et applications	51
4.1	Mesure produit et théorèmes de Fubini	51
4.1.1	Mesure produit	51
4.1.2	Théorèmes de Fubini	52
4.2	Changement de variable	53
4.2.1	Théorème de changement de variable	53
4.2.2	Coordonnées polaires	55
4.2.3	Convolution	56
4.3	Variables indépendantes (vu seulement en cours)	57
4.3.1	Événements et variables indépendantes	57
4.3.2	Conséquences immédiates de l'indépendance	58
4.3.3	Loi jointes, lois marginales	58
4.3.4	Exemples	59

Chapitre 1

Espaces mesurables et espaces mesurés

Afin de ne pas laisser d'ambiguïté sur la notion de dénombrabilité qui est fondamentale en théorie de la mesure, on commence par un approfondissement de ce point.

1.1 Cardinaux

1.1.1 Rappels

Définition 1.1.1. Soit E, F des ensembles, $f : E \rightarrow F$ est une injection si $\forall x, y \in E$, $f(x) = f(y) \Rightarrow x = y$.

Définition 1.1.2. Soit E, F des ensembles, $f : E \rightarrow F$ est une surjection si $\forall z \in F$, $\exists x \in E$ tel que $f(x) = z$.

Définition 1.1.3. Soit E, F des ensembles, $f : E \rightarrow F$ est une bijection si f est une injection et une surjection.

Proposition 1.1.4. Soient E, F, G des ensembles. Soient $f : E \rightarrow F$, $g : F \rightarrow G$. Alors $[f \text{ et } g \text{ injectives}] \Rightarrow [g \circ f \text{ injective}]$.

Définition 1.1.5. Soit E un ensemble

- S'il est vide, on dit qu'il est de fini de cardinal zéro.
- S'il est non vide, on dit que E est de cardinal fini un entier strictement positif et $f : \{1, 2, \dots, n\} \rightarrow E$ qui soit une bijection. Dans ce cas, on dit que $\text{card}(E) = n$.

Définition 1.1.6. Soit E un ensemble non fini, on dit que E est (de cardinal) infini.

Proposition 1.1.7. Soit E un ensemble, on a l'équivalence entre

- E est infini
- il existe un sous ensemble strict F tel que E et F soient en bijection.

Par exemple, on peut penser à une bijection entre $2\mathbb{N}$ et \mathbb{N} , à une bijection entre $]0, 1[$ et \mathbb{R} , à une bijection entre $]0, +\infty[$ et \mathbb{R} .

1.1.2 Ensembles dénombrables

Définition 1.1.8. On dit qu'un ensemble E est dénombrable s'il existe une bijection entre \mathbb{N} et E . Un ensemble est dit au plus dénombrable s'il est fini ou dénombrable. Un ensemble qui n'est pas au plus dénombrable sera dit non dénombrable.

Si l'ensemble est au plus dénombrable, on peut numéroter un à un les éléments de E .

Attention! Certains auteurs utilisent le vocabulaire "dénombrable" pour "au plus dénombrable", et le vocabulaire "infini dénombrable" pour "dénombrable".

Exemple 1.1.9. \mathbb{Z} est dénombrable car l'application

$$f : \mathbb{Z} \rightarrow \mathbb{N}$$

$$k \mapsto \begin{cases} 2n & \text{si } n \geq 0 \\ -2n + 1 & \text{si } n < 0 \end{cases}$$

est bijective (donc injective).

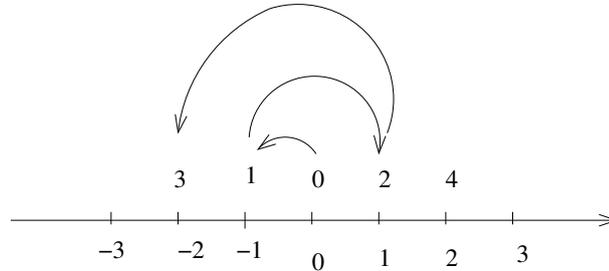


FIGURE 1.1 – énumération des éléments de \mathbb{Z} .

Exemple 1.1.10. $\mathbb{N} \times \mathbb{N}$ est dénombrable car l'application

$$f : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$$

$$(p, q) \mapsto \frac{(p+q)(p+q+1)}{2} + q$$

est bijective (donc injective).

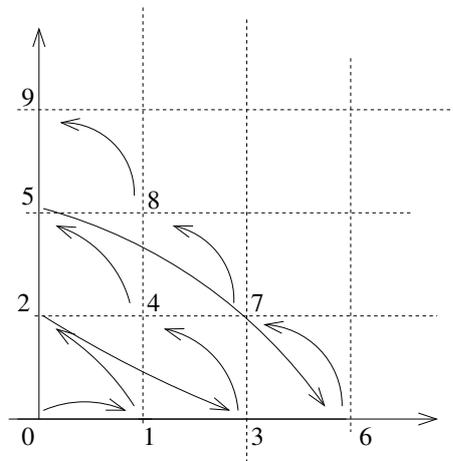


FIGURE 1.2 – énumération des éléments de $\mathbb{N} \times \mathbb{N}$.

Exemple 1.1.11. L'ensemble \mathbb{Q} est dénombrable. L'ensemble \mathbb{R} n'est pas dénombrable.

On utilise fréquemment la remarque suivante.

Remarque 1.1.12. Soit X , un sous ensemble au plus dénombrable d'un ensemble E , et $A \subset X$ alors, A est au plus dénombrable.

Proposition 1.1.13. Si X et Y sont deux ensembles tels qu'il existe une injection de X dans Y , et si Y est au plus dénombrable alors X est au plus dénombrable.

Proposition 1.1.14. Si X et Y sont deux ensembles tels qu'il existe une surjection de X dans Y , alors

- i) Si X est fini alors Y finie.
- ii) Si X est au plus dénombrable alors Y au plus dénombrable.

Remarque 1.1.15. *Si X est un ensemble. Il est au plus dénombrable si et seulement si il existe une surjection de \mathbb{N} dans X .*

Proposition 1.1.16. *Soit E_1, \dots, E_n une famille finie d'ensembles dénombrables alors $E = E_1 \times E_2 \times \dots \times E_n$ est un ensemble dénombrable.*

Théorème 1.1.17. *Soit $(A_i)_{i \in I}$ une famille indexée par I de parties d'un ensemble X , alors*

- i) Si I est fini et si chacun des A_i est fini, alors $\cup_{i \in I} A_i$ est fini.*
- ii) Si I est au plus dénombrable et si chacun des A_i est au plus dénombrable, alors $\cup_{i \in I} A_i$ est au plus dénombrable.*
- iii) Si I est dénombrable et si chacun des A_i est dénombrable, alors $\cup_{i \in I} A_i$ est dénombrable.*

Remarque 1.1.18. *On peut tout à fait envisager des familles non dénombrables, par exemple $A_x = \{x\}$. Donc*

$$\mathbb{R} = \cup_{x \in \mathbb{R}} A_x.$$

On peut même remarquer qu'il s'agit d'une union disjointe. Si on munit \mathbb{R} d'une (loi de) probabilité $\mathcal{N}(0, 1)$, alors la probabilité d'un singleton vaut 0. On constate que la formule usuelle "en cas d'union disjointe, la probabilité de l'union est la somme des probabilités" ne s'applique pas ici, puisque la probabilité de l'union vaut 1, tandis que la somme des probabilités vaut $0 + 0 + \dots = 0$. La conclusion, est que cette formule ne s'applique que pour les unions disjointes DÉNOMBRABLES. D'où l'importance de maîtriser le concept de dénombrabilité.

1.1.3 Compléments

La bonne intuition pour comprendre ces questions est de généraliser le concept de cardinal. Tout ensemble possède un cardinal et les règles sont simples mais ne seront pas approfondies.

- Il existe une bijection de I dans J si et seulement si $\text{card}(I) = \text{card}(J)$.
- S'il existe une injection de I dans J , alors $\text{card}(I) \leq \text{card}(J)$.
- S'il existe une surjection de I dans J , alors $\text{card}(I) \geq \text{card}(J)$.
- $\text{card}(I) < \text{card}(\mathcal{P}(I))$.
- $\text{card}(\{1, 2, \dots, n\}) < \text{card}(\mathbb{N})$.
- $\text{card}(\mathbb{N}) < \text{card}(\mathbb{R})$.

En cohérence avec ces règles, on peut énoncer.

Théorème 1.1.19. *S'il existe une injection d'un ensemble X vers un ensemble Y et une injection de Y vers X , alors il existe une bijection entre eux.*

Théorème 1.1.20. *Soit X un ensemble non vide, il n'existe pas de bijection entre X et $\mathcal{P}(X)$. Plus précisément il n'existe pas de surjection entre X et $\mathcal{P}(X)$.*

Par exemple, le cas fini, $\mathcal{P}(\mathbb{N})$ est non dénombrable, $\mathcal{P}(\mathbb{R})$ est non dénombrable.

Théorème 1.1.21. *Soit $(X_i)_{i \in I}$ une famille d'ensemble non vide indicés par un ensemble dénombrable, si aucun des X_i n'est un singleton, alors $\prod_{i \in I} X_i$ est un ensemble infini non dénombrable.*

En d'autres termes, le produit cartésien de deux ensembles dénombrables est dénombrable tandis que $\{0, 1\}^{\mathbb{N}}$ est non dénombrable (car en bijection avec $\mathcal{P}(\mathbb{N})$).

Théorème 1.1.22. *L'ensemble \mathbb{R} est un ensemble infini non dénombrable.*

1.2 Introduction aux espaces mesurables

La théorie de la mesure est un outil théorique fondamental qui permet de modéliser correctement le hasard en unifiant la théorie des variables aléatoires. Elle repose sur le concept de tribus dont l'introduction paraît initialement artificiellement compliquée. On va commencer par introduire le concept de tribus à partir d'un exemple.

1.2.1 Tribus d'information

Supposons que « l'univers » soit l'ensemble $\Omega = \{0, 1, \dots, 36\}$. On a tiré au hasard un élément ω dans Ω , mais on ne peut savoir précisément quel est le tirage. Par contre, on dispose d'une certaine information. Pour rendre les choses concrètes, je sais que l'on connaît la couleur du tirage (voir figure 1.3).



FIGURE 1.3 – la roulette française.

La question que l'on se pose est « quels sont les événements que l'on peut manipuler ? ». Formellement, on appelle \mathcal{A} l'ensemble de tels événements A tandis que $\omega \in \Omega$ est une réalisation de l'aléa.

- Par exemple $A_{\text{vert}} := \{0\} \in \mathcal{A}$ car si la couleur tirée est vert, je suis sûr que le 0 est sorti.
- L'événement « le tirage est rouge », correspond au sous-ensemble

$$A_{\text{rouge}} := \{1, 3, 5, 7, 9, 12, 14, 16, 18, 19, 21, 23, 25, 27, 30, 32, 34, 36\}, A_{\text{rouge}} \in \mathcal{A}$$

Cette situation ne correspond pas à une information parfaite, car si la couleur tirée est rouge, je suis sûr d'être dans cet ensemble mais je n'ai rien de plus précis.

- De même, on peut envisager

$$A_{\text{noir}} := \{2, 4, 6, 8, 10, 11, 13, 15, 17, 20, 22, 24, 26, 28, 29, 31, 33, 35\} \in \mathcal{A}$$

car si la couleur tirée est noire, je suis sûr d'être dans l'ensemble A_{noir} .

- Ou bien $A_{\text{rouge ou noir}} := A_{\text{rouge}} \cup A_{\text{noir}} \in \mathcal{C}$ car si la couleur tirée est rouge ou noire, je suis sûr d'être dans l'ensemble $A_{\text{rouge ou noir}}$.
- Il y a également l'événement « le tirage n'est ni rouge, ni vert, ni noir », qui correspond à l'ensemble $A_{\text{ni rouge ni noir ni vert}} := \emptyset$ car il s'agit d'une situation impossible. On a donc $\emptyset \in \mathcal{A}$.
- ...

A l'arrivée, avec des notations évidentes, on obtient que l'ensemble \mathcal{A} des événements est l'ensemble

$$\{\emptyset, A_{\text{vert}}, A_{\text{rouge}}, A_{\text{noir}}, A_{\text{rouge ou noir}}, A_{\text{rouge ou vert}}, A_{\text{vert ou noir}}, \Omega\}. \quad (1.2.1)$$

Observons les propriétés de ce que l'on va qualifier de « tribu d'information ».

Remarque 1.2.1. Dans cet exemple, l'ensemble \mathcal{A} de parties de Ω vérifie

1. $\Omega \in \mathcal{A}$
2. stabilité par passage au complémentaire
3. stabilité par union finie

Cet exemple introductif possède l'avantage et l'inconvénient d'être simple. En effet, il est un peu trop simple, puisque Ω étant fini, il n'y a pas lieu de considérer des unions dénombrables alors que c'est une notion que l'on manipule fréquemment : par exemple si Y suit une variable de Poisson. On sait que

$$P(Y \text{ soit pair}) = P((Y = 0) \cup (Y = 2) \cup \dots) \text{ (union dénombrable)}.$$

Au passage, il s'agit d'une union disjointe et on notera $A \sqcup B$ pour désigner l'ensemble $A \cup B$ mais en rappelant qu'on manipule une union disjointe.

Après cet exemple, on peut maintenant donner la définition formelle d'une tribu.

1.2.2 Tribus au sens mathématique

Dans la suite, on utilisera un ensemble Ω que l'on appellera « univers ». Il contient tous les aléas possibles. Les éléments de Ω sont appelés réalisation. Un événement peut coïncider avec une réalisation mais en général, un événement n'est pas un singleton de réalisation (exemple, "le dé est tombé sur une face paire" est un événement mais pas une réalisation).

Définition 1.2.2. Une famille \mathcal{A} de parties de Ω est une tribu (sur Ω) si elle vérifie

1. $\Omega \in \mathcal{A}$
2. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$ (stabilité par passage au complémentaire)
3. $A_0, A_1, A_2, \dots \in \mathcal{A} \Rightarrow \bigcup_{n \geq 0} A_n \in \mathcal{A}$ (une réunion dénombrable d'éléments de \mathcal{A} est dans \mathcal{A}).

Si \mathcal{A} de parties de Ω est une tribu (sur Ω), on dit que l'ensemble (Ω, \mathcal{A}) est un espace mesurable. On rencontre parfois le vocabulaire espace mesurable également appelé espace probabilisable.

Remarque 1.2.3. On remarque que :

- Une tribu est un ensemble de parties. Ces parties sont appelées « événements ».
- $\emptyset \in \mathcal{A}$
- Le mot tribu est synonyme en anglais de σ -algèbre (vocabulaire alternatif). Le terme σ renvoie à de la dénombrabilité. On vérifie donc qu'une tribu (σ -algèbre) est a fortiori une algèbre¹.
- Il existe d'autres propriétés sur les ensembles de parties (algèbre appelée aussi clan, les π -systèmes, les λ -systèmes... qui sont fondamentaux dans les approfondissements de la théorie de la mesure mais qui seront omis dans ce cours).

Remarque 1.2.5. On vérifie trivialement qu'une tribu est stable par union finie, intersection finie et par différence ensembliste.

Proposition 1.2.6. Stabilité par intersection dénombrable.

Soient \mathcal{A} une tribu et $A_0, A_1, A_2, \dots \in \mathcal{A}$, alors $\bigcap_{n \geq 0} A_n \in \mathcal{A}$.

Démonstration. On note pour tout n , $B_n = A_n^c$. Donc, par définition d'une tribu, $B_n \in \mathcal{A}, \forall n$ et $\bigcup_{n \geq 0} B_n \in \mathcal{A}$.

$$\begin{aligned} \bigcap_{n \geq 0} A_n &= \bigcap_{n \geq 0} B_n^c \\ &= \left(\bigcup_{n \geq 0} B_n \right)^c \\ (\text{par définition}) &\in \mathcal{A}. \end{aligned}$$

□

Exemple 1.2.7. Pour n'importe quel ensemble Ω , $\mathcal{A} = \{\emptyset, \Omega\}$ est une tribu.

Exemple 1.2.8. Pour n'importe quel ensemble Ω , $\mathcal{A} = \mathcal{P}(\Omega)$ (les parties de Ω) est une tribu.

1.

Définition 1.2.4. Une famille \mathcal{A} de parties de Ω est une algèbre sur Ω si elle vérifie

1. $\Omega \in \mathcal{A}$
2. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$ (stabilité par passage au complémentaire)
3. Si A_1 et A_2 sont dans \mathcal{A} , alors $A_1 \cup A_2$ (stabilité par union finie).
4. Si A_1 et A_2 sont dans \mathcal{A} , alors $A_1 \cap A_2$ (stabilité par intersection finie).

Définition 1.2.9. Soit \mathcal{A}_1 et \mathcal{A}_2 deux tribus, on dit que \mathcal{A}_1 est plus fine que \mathcal{A}_2 si $\mathcal{A}_2 \subset \mathcal{A}_1$.

Exemple 1.2.10. Si considère la tribu \mathcal{A} du paragraphe 1.2.1 et la tribu d'information \mathcal{E} associée au fait de connaître à la fois la parité et la couleur du tirage. Le vocabulaire employé est cohérent avec l'idée que l'on possède une information plus riche : la tribu d'information \mathcal{E} est plus fine que \mathcal{A} , dit autrement, \mathcal{E} correspond à un « raffinement » de l'information par rapport à \mathcal{A} .

Exemple 1.2.11. Pour n'importe quel ensemble Ω (par exemple \mathbb{R}), $\mathcal{P}(\Omega)$ est la tribu la plus fine (tribu discrète), on peut envisager n'importe quel événement. On peut alternativement considérer $\{\emptyset, \Omega\}$ qui est la tribu la moins fine (tribu grossière).

Remarque 1.2.12. L'intersection d'une famille de tribus de Ω est une tribu mais en général l'union de deux tribus n'est pas une tribu.

Exemple 1.2.13. Pour $\Omega = \{a, b, c\}$,

$$\mathcal{A} = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$$

$$\tilde{\mathcal{A}} = \{\emptyset, \{a, b\}, \{c\}, \Omega\}$$

sont des tribus tandis que $\mathcal{A} \cup \tilde{\mathcal{A}}$ n'est pas une tribu.

Proposition 1.2.14. Soit $\mathcal{C} \subset \mathcal{P}(\Omega)$, il existe une tribu notée $\sigma(\mathcal{C})$ telle que si \mathcal{B} est une tribu telle que $\mathcal{C} \subset \mathcal{B}$ alors $\sigma(\mathcal{C}) \subset \mathcal{B}$.

On dira que $\sigma(\mathcal{C})$ est la plus petite tribu contenant \mathcal{C} , ou encore que $\sigma(\mathcal{C})$ est la tribu engendrée par \mathcal{C} .

Remarque 1.2.15. On peut définir $\sigma(\mathcal{C})$ en termes d'intersection de toutes les tribus contenant \mathcal{C} .

Remarque 1.2.16. Si \mathcal{A} et $\tilde{\mathcal{A}}$ sont deux tribus d'informations $\sigma(\mathcal{A}, \tilde{\mathcal{A}})$ est la tribu correspondant à la réunion des deux informations.

Remarque 1.2.17. Il existe une grande analogie en termes de manipulations entre le concept de tribu engendrée et le concept d'espace vectoriel engendré.

1.2.3 Tribu finies

Un cas fondamental à maîtriser correspond au cas des tribus de cardinal fini, même si potentiellement Ω est de cardinal infini et par analogie au cas d'un système complet d'événement et qu'il n'existe pas de tribu dénombrable.

Proposition et définition 1.2.18. Soit I un ensemble fini (respectivement dénombrable) de $\mathcal{P}(\Omega)$, et $(\Omega_i)_{i \in I}$, une partition finie (respectivement dénombrable) de Ω . Alors l'ensemble $\mathcal{A} \subset \mathcal{P}(\Omega)$ défini par

$$A \in \mathcal{A} \text{ si et seulement si il existe } J \subset I \text{ tel que } A = \cup_{i \in J} \Omega_i.$$

est une tribu finie sur Ω . On appelle système complet d'événement $(\Omega_i)_{i \in I}$. Réciproquement toute tribu finie est de cette forme.

Exemple 1.2.19. Si A est un sous-ensemble non vide de Ω et différent de Ω alors $\{\emptyset, A, A^c, \Omega\}$ est une tribu.

Exemple 1.2.20. A la partition $(] - \infty, -1],] - 1, 1],] 1, +\infty[)$ est associée la tribu

$$\mathcal{A} = \{\emptyset,] - \infty, -1],] - \infty, 1],] - 1, 1],] - 1, +\infty[,] - \infty, -1] \cup] 1, +\infty, -1],] 1, \infty[, \mathbb{R}\}$$

Remarque 1.2.21. On peut remarquer que la tribu d'information du paragraphe 1.2.1 est bien de cette forme. De plus une conséquence immédiate de cette proposition est que le cardinal d'une tribu finie est de la forme 2^n .

1.2.4 Tribu des boréliens de \mathbb{R}

Définition 1.2.22. Soit l'ensemble de parties de \mathbb{R} suivant :

$$\mathcal{C} = \{]a, b[: a, b \in \mathbb{R} \cup \{+\infty, -\infty\}\}$$

(c'est l'ensemble des intervalles ouverts). On remarque que \mathcal{C} n'est pas une tribu. La tribu $\tau(\mathcal{C})$ s'appelle la tribu des boréliens² et se note $\mathcal{B}(\mathbb{R})$.

Exemple 1.2.24. Soit $[a, b]$ intervalle fermé de \mathbb{R} . Les intervalles $] - \infty, a[$, $]b, +\infty[$ sont dans $\mathcal{B}(\mathbb{R})$. Par propriété de tribu de $\mathcal{B}(\mathbb{R})$, on a donc $] - \infty, a[\cup]b, +\infty[\in \mathcal{B}(\mathbb{R})$ (stabilité par réunion finie), et donc aussi $(] - \infty, a[\cup]b, +\infty[)^c = [a, b] \in \mathcal{B}(\mathbb{R})$ (stabilité par passage au complémentaire).

De même, on peut montrer que tous les intervalles de \mathbb{R} sont dans $\mathcal{B}(\mathbb{R})$, en particulier tous les singletons (les ensembles de la forme $\{x\}$, $x \in \mathbb{R}$).

Exemple 1.2.25. L'ensemble des rationnels est donc un borélien car union dénombrable de singletons.

1.2.5 Fonctions mesurables

Définition 1.2.26. Soit (Ω, \mathcal{A}) et (Y, \mathcal{F}) deux espaces mesurables. On dit que l'application $f : \Omega \rightarrow Y$ est mesurable de (Ω, \mathcal{A}) dans (Y, \mathcal{F}) et on note $f \in \mathcal{L}^0((\Omega, \mathcal{A}), (Y, \mathcal{F}))$ si pour tout $B \in \mathcal{F}$, $f^{-1}(B) \in \mathcal{A}$.

On utilisera la notation réduite $\mathcal{L}^0(\Omega, \mathcal{A}, \mathbb{R})$ pour $\mathcal{L}^0((\Omega, \mathcal{A}), (\mathbb{R}, \mathcal{B}(\mathbb{R})))$.

Proposition 1.2.27. Soit (Ω, \mathcal{A}) et (Y, \mathcal{F}) deux espaces mesurés et soit $f : \Omega \rightarrow Y$. On suppose que la tribu \mathcal{F} sur Y est la tribu grossière $\{\emptyset, Y\}$. Alors, f est mesurable.

Proposition 1.2.28. Soit (Ω, \mathcal{A}) et (Y, \mathcal{F}) deux espaces mesurés et soit $f : \Omega \rightarrow Y$. On suppose que la tribu sur \mathcal{A} sur Ω est la tribu la plus fine $\mathcal{P}(\Omega)$. Alors, f est mesurable.

Proposition 1.2.29. Soit (Ω, \mathcal{A}) et (Y, \mathcal{F}) deux espaces mesurés et soit $f : \Omega \rightarrow Y$ une application mesurable de (Ω, \mathcal{A}) dans (Y, \mathcal{F}) .

- i) Si \mathcal{A}' est une tribu sur Ω plus fine que \mathcal{A} , alors f est mesurable de (Ω, \mathcal{A}') dans (Y, \mathcal{F}) .
- ii) Si \mathcal{F}' est une tribu sur Y moins fine que \mathcal{F} , alors f est mesurable de (Ω, \mathcal{A}) dans (Y, \mathcal{F}') .

Théorème 1.2.30. Soit $(\Omega_1, \mathcal{A}_1)$, $(\Omega_2, \mathcal{A}_2)$ et $(\Omega_3, \mathcal{A}_3)$ trois espaces mesurables. On considère $f : \Omega_1 \rightarrow \Omega_2$ et $g : \Omega_2 \rightarrow \Omega_3$, si f est mesurable de $(\Omega_1, \mathcal{A}_1)$ dans $(\Omega_2, \mathcal{A}_2)$ et g est mesurable de $(\Omega_2, \mathcal{A}_2)$ dans $(\Omega_3, \mathcal{A}_3)$ alors $g \circ f$ est mesurable de $(\Omega_1, \mathcal{A}_1)$ dans $(\Omega_3, \mathcal{A}_3)$.

1.3 Mesures

Pour alléger le vocabulaire, on appellera mesure une mesure positive. L'étude des mesures générales dépasse le cadre de ce cours.

1.3.1 Définition

Notation 1.3.1. Dans le contexte des mesures positives, on rappelle les conventions usuelles de calcul $\forall \lambda \in \overline{\mathbb{R}}$, $\lambda > 0 \Rightarrow \lambda \times (+\infty) = +\infty$, et on ajoute la condition suivante (qui n'est pas valable ailleurs) : $0 \times (+\infty) = 0$.

L'intérêt de cette notation est de pouvoir écrire par exemple la relation suivante en incluant le cas $\lambda = 0$

$$\forall \lambda \geq 0, \sum_{n=1}^{+\infty} \frac{\lambda}{n} = \lambda \sum_{n=1}^{+\infty} \frac{1}{n}.$$

2. On verra plus tard que c'est compatible avec la définition générale des boréliens

Définition 1.2.23. Plus généralement si Ω est un espace métrique, on considère \mathcal{O} l'ensemble des ouverts. La tribu $\tau(\mathcal{O})$ notée $\mathcal{B}(\Omega)$ s'appelle la tribu des boréliens de Ω (ou sur Ω).

Définition 1.3.2. Soit Ω un ensemble muni d'une tribu \mathcal{A} . On dit que μ est une mesure (positive) sur (Ω, \mathcal{A}) si :

1. $\mu : \mathcal{A} \rightarrow [0, +\infty]$ (elle peut prendre la valeur $+\infty$)
2. $\mu(\emptyset) = 0$
3. si $A_0, A_1, A_2, \dots \in \mathcal{A}$ et sont deux à deux disjoints alors $\mu(\bigsqcup_{n \geq 0} A_n) = \sum_{n \geq 0} \mu(A_n)$.

On dit alors que le triplet $(\Omega, \mathcal{A}, \mu)$ constitue un espace mesuré.

Quand μ est une mesure sur (Ω, \mathcal{A}) est telle que $\mu(\Omega) = 1$, on dit que μ est une mesure de probabilité. La tribu \mathcal{A} contient tous les événements possibles et, pour $A \in \mathcal{A}$, $\mu(A)$ est la probabilité que A se produise.

Définition 1.3.3. Quand μ est telle que $\mu(\Omega) < \infty$, on dit que μ est une mesure finie. Dans ce cas, on peut normaliser en posant $\mathbb{P}(A) = \mu(A)/\mu(\Omega)$ afin d'obtenir une (mesure de) probabilité.

Définition 1.3.4. Quand μ est telle qu'il existe une famille dénombrable $(\Omega_n)_{n \in \mathbb{N}}$ telle que pour tout n , $\mu(\Omega_n) < \infty$ et $\Omega = \cup_n \Omega_n$, on dit que μ est une mesure σ -finie.

Remarque 1.3.5. Soit (Ω, \mathcal{A}) un espace mesurable. On peut munir (Ω, \mathcal{A}) de deux mesures

- la mesure nulle
- $\mu(A) = +\infty$ si $A \neq \emptyset$ et $\mu(\emptyset) = 0$.

Ce qui justifie le vocabulaire d'espace mesurable.

Exemple 1.3.6. Le triplet $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu_C)$ est un espace mesuré. Nous avons vu (exemple 1.2.8) que $\mathcal{P}(\mathbb{N})$ est une tribu sur \mathbb{N} . De plus :

1. Pour $A \in \mathcal{P}(\mathbb{N})$, $\mu_C(A)$ (= le nombre d'éléments de A , μ_C est la mesure de comptage) est bien dans $[0, +\infty]$.
2. La partie \emptyset est de cardinal 0, donc $\mu_C(\emptyset) = 0$.
3. Si $A_0, A_1, \dots \in \mathcal{P}(\mathbb{N})$ sont deux à deux disjoints, $\mu_C(\bigsqcup_{n \geq 0} A_n) = \sum_{n \geq 0} \mu_C(A_n)$ (somme qui peut comporter le terme $+\infty$ si l'ensemble n'est pas fini).

Exemple 1.3.7. Soit (Ω, \mathcal{A}) un espace mesurable. On peut définir la mesure de comptage :

1. Pour $A \in \mathcal{A}$, $\mu_C(A)$ (= le nombre d'éléments de A) est bien dans $[0, +\infty]$.
2. La partie \emptyset est de cardinal 0, donc sa mesure de comptage vaut 0.
3. Si $A_0, A_1, \dots \in \mathcal{A}$ sont deux à deux disjoints, $\mu_C(\bigsqcup_{n \geq 0} A_n) = \sum_{n \geq 0} \mu_C(A_n)$.

$$\mu_C(\{1, 5, 7\}) = 3, \quad \mu_C(\mathbb{N}) = +\infty, \quad \mu_C([0, 1]) = +\infty,$$

Exemple 1.3.8. Soit (Ω, \mathcal{A}) un espace mesurable et $a \in \Omega$. On peut définir la mesure de Dirac au point a notée δ_a : pour $A \in \mathcal{A}$,

$$\delta_a(A) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{si } a \notin A \end{cases}$$

1. La mesure de Dirac est bien à valeur dans $\overline{\mathbb{R}}_+$.
2. La mesure de Dirac de la partie \emptyset vaut 0.
3. Si $A_0, A_1, \dots \in \mathcal{A}$ sont deux à deux disjoints, $\delta_a(\bigsqcup_{n \geq 0} A_n) = \sum_{n \geq 0} \delta_a(A_n)$.

Pour $a = 0$,

$$\delta_a(\{1, 5, 7\}) = 0, \quad \delta_a(\mathbb{N}) = 1, \quad \delta_a([0, 1]) = 1, \quad \delta_a(]0, 1]) = 0$$

Proposition 1.3.9. Croissance et mesure d'une différence

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Soit $A, B \in \mathcal{A}$ tels que $B \subset A$.

- $\mu(B) \leq \mu(A)$.
- $\mu(A \setminus B) + \mu(B) = \mu(A)$.
- Si de plus le terme de droite existe dans $\overline{\mathbb{R}}$, alors $\mu(A \setminus B) = \mu(A) - \mu(B)$.

(Rappel : $A \setminus B = \{x \in A \mid x \notin B\}$.)

Démonstration. On a $\mu(A) = \mu(A \setminus B) + \mu(B)$ (car $A \setminus B$ et B sont disjoints). Donc $\mu(B) \leq \mu(A)$. Si de plus le terme de droite existe dans $\overline{\mathbb{R}}$, nous avons alors $\mu(A \setminus B) = \mu(A) - \mu(B)$. \square

Proposition 1.3.10. *Sous-additivité.*

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Si $A_0, A_1, A_2, \dots \in \mathcal{A}$ (pas forcément deux à deux disjoints). Alors $\mu(\bigcup_{n \geq 0} A_n) \leq \sum_{n \geq 0} \mu(A_n)$.

Proposition 1.3.11. *Mesure d'une réunion croissante.*

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Soient $A_0, A_1, \dots \in \mathcal{A}$ tels que $A_0 \subset A_1 \subset \dots \subset A_n \subset A_{n+1} \subset \dots$. Alors $\mu(\bigcup_{k \geq 0} A_k) = \lim_{n \rightarrow \infty} \mu(A_n) = \sup_{n \in \mathbb{N}} \mu(A_n)$.

Proposition 1.3.12. *Mesure d'une intersection décroissante.*

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Soient $A_0, A_1, \dots \in \mathcal{A}$ tels que $A_0 \supset A_1 \supset \dots \supset A_n \supset A_{n+1} \supset \dots$ et tels que $\mu(A_0) < +\infty$. Alors $\mu(\bigcap_{k \geq 0} A_k) = \lim_{n \rightarrow +\infty} \mu(A_n) = \inf_{n \rightarrow +\infty} \mu(A_n)$.

Remarque 1.3.13. Ces deux propositions constituent ce que l'on appelle un principe de continuité monotone séquentiel COMOSEQ. De plus,

- On peut se contenter de l'hypothèse, il existe un indice k tel que $\mu(A_k) < +\infty$.
- l'hypothèse $\mu(A_0) < +\infty$ est automatiquement vérifiée dans le cas d'une mesure de probabilité.

Exemple 1.3.14. On peut regarder le cas où $A_n = [n, +\infty[$, le cas où $B_n = [0, n]$ et le cas où $C_n = [0, 1 - 1/n]$ le cas où $D_n = [0, 1 + 1/n]$.

1.3.2 Théorèmes d'impossibilité

Le plus naturel serait de ne pas utiliser les tribus et de se contenter de travailler avec $\mathcal{P}(\Omega)$ mais historiquement on s'est rapidement rendu compte que ce n'était tout simplement pas possible.

Théorème 1.3.15 (Vitali, 1905). *Il n'existe pas de mesure $\mu : \mathcal{P}(\mathbb{R}) \rightarrow \overline{\mathbb{R}}_+$ telle que*

- Si I est un intervalle, $\mu(I)$ est égal à sa longueur.
- Si B est un translaté de A , alors $\mu(A) = \mu(B)$

On peut trouver une explication de cette propriété ici

<http://carolinevernier.website/compl/non-lebesgue-mesurable.pdf> On pourrait penser que le problème vient de la σ -additivité mais il n'en est rien.

Théorème 1.3.16 (Banach Tarski, 1922). *Si $n \geq 3$, il n'existe pas de fonction $\mu : \mathcal{P}(\mathbb{R}^n) \rightarrow \overline{\mathbb{R}}_+$ telle que*

- $\mu([0, 1]^n) = 1$.
- Si A et B sont des ensembles, $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$.
- μ est invariante par translation/rotation.

L'idée de la démonstration repose sur le résultat clé suivant.

Théorème 1.3.17 (Banach Tarski, 1922). *Il existe cinq ensembles notés $(A_1, A_2, A_3, A_4, A_5)$ et et 5 ensembles notés $(E_1, E_2, E_3, E_4, E_5)$ tels que*

- chaque E_i est l'image de A_i par un déplacement³ noté d_i , c'est-à-dire $E_i = d_i(A_i)$
- La famille $(A_1, A_2, A_3, A_4, A_5)$ constitue une partition de la boule unité B .
- La famille (E_1, E_2) constitue une partition de la boule unité B .
- La famille (E_3, E_4, E_5) constitue une partition de la boule unité B .

Si comme on le voudrait intuitivement, le volume se conserve par déplacement, on doit avoir $rmVol(A_i) = Vol(E_i)$. Mais c'est incompatible avec

$$\begin{cases} Vol(B) = Vol(A_1) + Vol(A_2) + Vol(A_3) + Vol(A_4) + Vol(A_5) \\ Vol(B) = Vol(E_1) + Vol(E_2) \\ Vol(B) = Vol(E_3) + Vol(E_4) + Vol(E_5) \end{cases}$$

puisque'on arrive à $Vol(B) = 0$ ou $Vol(B) = +\infty$.

3. On appelle déplacement une isométrie affine qui conserve l'orientation, par exemple une translation, une rotation mais aussi leurs composées.

1.3.3 Mesure de Borel sur \mathbb{R}

Puisque on ne peut construire de mesure sur $\mathcal{P}(\mathbb{R})$, on va commencer par construire une mesure sur les boréliens.

Théorème 1.3.18. *Mesure de Borel.*

Il existe une unique mesure μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ vérifiant

1. pour tout intervalle de la forme $[a, b[$, $\mu([a, b[) = b - a$
2. $\forall A \in \mathcal{B}(\mathbb{R}), \forall x \in \mathbb{R}, \{y : y - x \in A\} \in \mathcal{B}(\mathbb{R})$ et $\mu(\{y : y - x \in A\}) = \mu(A)$ (stabilité par translation).

Cette mesure μ ou μ_1 ou $\mu_{\mathbb{R}}$ s'appelle la mesure de Borel de \mathbb{R} ou la mesure de Borel sur \mathbb{R} .

Exemple 1.3.19. 1. *Mesure de Borel d'un singleton.*

Soit $x \in \mathbb{R}, \forall n \geq 1, \{x\} \subset]x - 1/n, x + 1/n[$. Donc $\forall n \geq 1, \mu(\{x\}) \leq \mu(]x - 1/n, x + 1/n[) = 2/n$. Donc $\mu(\{x\}) = 0$.

2. *Mesure de Borel d'un segment.*

Soit $a \leq b$ deux nombres réels, $\forall n \geq 1, [a, b[\subset [a, b] \subset [a, b + 1/n[$. Donc $\forall n \geq 1, \mu([a, b]) \leq \mu([a, b]) \leq \mu([a, b + 1/n[)$. Donc $\mu([a, b]) = b - a$.

3. *Mesure de Borel d'un intervalle ouvert.*

Soit $a < b$ deux nombres réels, $\forall n \geq 1, [a + 1/n, b[\subset [a, b[\subset [a, b]$. Donc $\forall n \geq 1, \mu([a + 1/n, b]) \leq \mu([a, b]) \leq \mu([a, b])$. Donc $\mu([a, b]) = 0$.

Exemple 1.3.20 (admis). Soit $A \in \mathcal{B}(\mathbb{R})$ et α une constante, on pose $B = \alpha A$ (image par une homothétie). Alors $B \in \mathcal{B}(\mathbb{R})$ et $\mu(B) = |\alpha|\mu(A)$.

1.4 Application aux fondements de la théorie des probabilités

1.4.1 Définitions générales

Définition 1.4.1. On appelle espace probabilisé un espace mesuré $(\Omega, \mathcal{A}, \mathbb{P})$ où la mesure \mathbb{P} est telle que $\mathbb{P}(\Omega) = 1$. On dit alors que \mathbb{P} est une mesure de probabilité (et c'est pour cela qu'on la note \mathbb{P} plutôt que μ). Les éléments de \mathcal{A} sont appelés événements.

Définition 1.4.2. On appelle variable aléatoire toute application mesurable X d'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ dans un espace mesurable (E, \mathcal{E}) . On dit alors que X est à valeurs dans E .

On notera v.a. pour « variable aléatoire » et v.a.r. pour « variable aléatoire réelle » (variable aléatoire à valeurs dans \mathbb{R}).

Remarque 1.4.3. Si par exemple, on considère l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ égal à $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \delta_5)$, l'espace d'arrivée $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et la fonction $X(t) = 2t$. Il s'agit d'une variable aléatoire alors que ce n'est pas une variable et que cette fonction n'a rien d'aléatoire. Il faut comprendre le terme variable aléatoire comme un bloc insécable « variablaléatoire ».

Dans toute la suite du document, si on ne précise rien, $(\Omega, \mathcal{A}, \mathbb{P})$ sera utilisé pour désigner un espace probabilisé.

Exemple 1.4.4. On peut vérifier que les variables aléatoires telles qu'on les pensait dans les classes antérieures s'inscrivent dans cette définition mathématique. Par exemple, soit X le gain de pile dans un jeu de pile ou face. Formellement $\Omega = \{P, F\}$ muni de la tribu $\mathcal{P}(\Omega)$ (sauf exception, on prendra systématiquement $\mathcal{P}(\Omega)$ comme tribu si Ω est fini) et de la mesure de probabilité uniforme \mathbb{P} . On considère l'espace mesuré $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et la fonction $X(P) = 1$ et $X(F) = -1$. Puisque la tribu sur Ω est la tribu la plus fine, la variable aléatoire X est mesurable, c'est donc bien une variable aléatoire. Pour autant, la fonction X n'est pas aléatoire, l'aléa se situe au niveau de Ω via la distribution de probabilité \mathbb{P} qui constitue une information fondamentale de la nature de X . En particulier, on peut jouer à pile ou face avec une pièce truquée, en tant que fonction X ne va pas changer mais en tant que variable aléatoire, ce n'est plus le même X .

Exemple 1.4.5. Toutes les variables aléatoires ne sont pas réelles. Soit $\Omega = \{1, 2, \dots, 6\}$ muni de la tribu $\mathcal{P}(\Omega)$ et de la mesure de probabilité uniforme \mathbb{P} . On considère l'espace mesuré $(E, \mathcal{E}) = (\{\text{pair}, \text{impair}\}, \mathcal{P}(E))$ et la fonction $X : \Omega \rightarrow \{\text{pair}, \text{impair}\}$. Puisque la tribu sur Ω est la tribu la plus fine, la variable aléatoire X est mesurable.

Proposition et définition 1.4.6. Soit $X : \Omega \rightarrow (E, \mathcal{E})$ une variable aléatoire. On pose pour tout $A \in \mathcal{E}$,

$$\begin{aligned}\mathbb{P}_X(A) &= \mathbb{P}(X \in A) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) \\ &= \mathbb{P}(X^{-1}(A)).\end{aligned}$$

La fonction \mathbb{P}_X est une mesure de probabilité définie sur (E, \mathcal{E}) . Si $E = \mathbb{R}^d$, on appelle loi de probabilité de X la mesure \mathbb{P}_X .

Remarque 1.4.7. On rappelle que, par définition, $\{\omega \in \Omega : X(\omega) \in A\} = X^{-1}(A)$. On notera $\mathbb{P}_X(A) = \mathbb{P}(X \in A)$. C'est un abus de notation très courant. Dans le cas général, la mesure \mathbb{P}_X s'appelle la mesure image de la mesure \mathbb{P} par la fonction X car c'est un cas particulier d'un concept plus général. Il est facile de vérifier que $\mathbb{P}_X(E) = \mathbb{P}(\Omega) = 1$, donc c'est une mesure de probabilité.

Remarque 1.4.8. Dans le cas où X est une variable aléatoire réelle (respectivement à valeurs dans $\overline{\mathbb{R}}$, respectivement à valeurs dans \mathbb{R}^d), on utilise systématiquement sur \mathbb{R} , la tribu de Lebesgue de \mathbb{R} qui sera définie ultérieurement (respectivement la tribu de Lebesgue de $\overline{\mathbb{R}}$, respectivement la tribu de Lebesgue de \mathbb{R}^d), \mathbb{P}_X est une loi sur \mathbb{R} . Il ne faut pas confondre Ω avec \mathbb{R} , ni \mathbb{P} avec \mathbb{P}_X .

1.4.2 Fonction de répartition

Dans un cadre général, on peut maintenant définir le concept de fonction de répartition d'une variable aléatoire réelle.

Définition 1.4.9. Soit X une v.a.r.⁴. On appelle fonction de répartition de X la fonction de répartition associée à la mesure \mathbb{P}_X , c'est à dire la fonction

$$\begin{aligned}F &: \mathbb{R} \rightarrow \mathbb{R}_+ \\ t &\mapsto F_X(t) = \mathbb{P}_X(]-\infty, t]) = \mathbb{P}(X \leq t)\end{aligned}$$

Le théorème suivant a déjà été vu aux travers d'exemples.

Théorème 1.4.10. Soit X une v.a.r.. Alors

- (i) F_X est croissante
- (ii) $\lim_{t \rightarrow -\infty} F_X(t) = 0$, $\lim_{t \rightarrow +\infty} F_X(t) = 1$
- (iii) F_X est càdlàg et $\lim_{t \rightarrow x, t < x} F_X(t) = \mathbb{P}(X < x)$
 F_X est continue en x si, et seulement si, $\mathbb{P}(X = x) = 0$.

Dans le cas de la dimension 2, on énonce (la généralisation à \mathbb{R}^d étant triviale).

Définition 1.4.11. Soit X une v.a. à valeurs dans \mathbb{R}^2 . Formellement, $X = (X_1, X_2)$ est définie sur espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans l'espace mesurable $(\mathbb{R}^2, \mathcal{P}(\mathbb{R}^2))$. On appelle fonction de répartition de X la fonction de répartition associée à la mesure \mathbb{P}_X , c'est à dire la fonction

$$\begin{aligned}F &: \mathbb{R}^2 \rightarrow \mathbb{R}_+ \\ t &\mapsto F_X(a_1, a_2) = \mathbb{P}_X(]-\infty, a_1] \times]-\infty, a_2]) = \mathbb{P}(X_1 \leq a_1, X_2 \leq a_2)\end{aligned}$$

4. On rappelle que formellement, X est définie sur espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans l'espace mesurable $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ et que cette fonction est mesurable

1.4.3 Conséquence de la mesurabilité d'une v.a.r.

Dans la définition d'une v.a.r., la condition de mesurabilité est essentielle mais difficile à appréhender ; elle est liée à des questions d'information. Reprenons l'exemple initial de la page 4. On va munir l'espace mesuré (Ω, \mathcal{A}) de la probabilité naturelle, c'est-à-dire que que \mathbb{P} est définie sur \mathcal{A} et $\mathbb{P}(A_{vert}) = 1/37$, $\mathbb{P}(A_{rouge}) = 18/37$, etc... on a ainsi construit un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Attention $\mathbb{P}(\{2\})$ n'existe pas car $\{2\} \notin \mathcal{A}$.

La fonction $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ correspondant au gain (0 ou 1) si le tirage est pair (au sens de la roulette) n'est pas une variable aléatoire. En effet, $f^{-1}(\{1\})$ qui vaut $\{2, 4, \dots, 36\}$ n'est pas un événement (il n'appartient pas à la tribu que nous avons considéré sur Ω). Cette non mesurabilité traduit le fait que si on a uniquement une information sur la couleur, on ne peut en déduire la parité (afin de procéder au paiement).

Attention : par contre, si on considère le même ensemble Ω MAIS muni de la tribu $\mathcal{P}(\Omega)$, et de la loi uniforme noté $\tilde{\mathbb{P}}$, $f : \Omega \rightarrow \mathbb{R}$ sera alors une variable aléatoire. Les lois de probabilité \mathbb{P} et $\tilde{\mathbb{P}}$ sont différentes car $\tilde{\mathbb{P}}(\{2, 4, \dots, 36\}) = 18/37$ alors que $\mathbb{P}(\{2, 4, \dots, 36\})$ n'existe pas.

Proposition 1.4.12. *On considère un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et X est une v.a.r. définie sur (Ω, \mathcal{A}) , (dit autrement $X \in \mathcal{L}((\Omega, \mathcal{A}), (\mathbb{R}, \mathcal{B}(\mathbb{R})))$) alors*

1. *Si le cardinal de \mathcal{A} est égal à 2, alors X est une va constante.*
2. *Si le cardinal de \mathcal{A} est égal à 4, alors il existe deux constantes a et b , il existe Z une v.a.r. qui suit une loi de Bernoulli tel que $X = xZ + b$.*

Quand on a manipulé jusqu'en L2 une variable aléatoire réelle donnée, on a considéré fréquemment des variables aléatoires $Y = \varphi(X)$ où $\varphi : \mathbb{R} \rightarrow \mathbb{R}$, par exemple $Y = X^2$, ou $Y = e^X \dots$

Compte tenu de la définition et du théorème de composition 1.2.30, il FAUT vérifier que Y soit mesurable mais il SUFFIT pour cela que φ soit mesurable de $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ d'où l'importance du concept de fonction mesurable.

1.5 Compléments sur les tribus et les mesures

Il faut approfondir les notions de mesurabilité car même sur \mathbb{R} , on ne possède pas de critères pour comprendre ce qu'est un borélien.

1.5.1 Retour sur la tribu des boréliens de \mathbb{R}

Dans la définition du cours, on a expliqué que la tribu des boréliens était la tribu engendré par les intervalles ouverts, ce qui ne correspond pas à la définition générale. C'est une conséquence du fait qu'il existe une base dénombrable d'intervalles ouverts pour la topologie de \mathbb{R} . De la même façon qu'un espace vectoriel possède plusieurs bases, une tribu peut avoir plusieurs familles génératrices.

Remarque 1.5.1. *Si \mathcal{C}_1 et \mathcal{C}_2 sont des familles de parties de Ω telles que $\mathcal{C}_1 \subset \tau(\mathcal{C}_2)$ et $\mathcal{C}_2 \subset \tau(\mathcal{C}_1)$, alors les tribus engendrées sont égales.*

Lemme 1.5.2. *Soit O un ouvert de \mathbb{R} , alors il existe une suite d'intervalles ouverts $(]a_n, b_n[)_{n \in \mathbb{N}}$ tels que $-\infty < a_n \leq b_n < +\infty$ et $O = \cup_{n \in \mathbb{N}}]a_n, b_n[$.*

Démonstration. Pour tout $x \in O$, on sait qu'il existe un intervalle non trivial $]x - \varepsilon, x + \varepsilon[$ contenu dans O . On sait également que les intervalles $]x - \varepsilon, x[$ et $]x, x + \varepsilon[$ contiennent un rationnel (car \mathbb{Q} est dense dans \mathbb{R}). Donc il existe (r, t) un couple de rationnels tels que par construction $x \in]r, t[\subset O$. Mais pour chaque x , il existe un rationnel, on a en fait une fonction $\varphi : O \rightarrow \mathbb{Q} \times \mathbb{Q}$ telle que $x \in]\varphi_1(x), \varphi_2(x)[$ et $] \varphi_1(x), \varphi_2(x)[\subset O$. Cette application n'est ni injective (pour des raisons de dénombrabilité) ni surjective (car $(1, 0)$ ne sera jamais atteint) et on est sûr que $\varphi(O)$ est dénombrable. Donc (quitte à faire des répétition) que tout les éléments de $\varphi(O)$ sont de la forme $]a_n, b_n[$. Une telle suite répond à la question. En

effet, si $x \in O$, il existe un entier $n(x)$ tel $\varphi(x) = (a_{n(x)}, b_{n(x)})$, donc $x \in \cup_{n \in \mathbb{N}}]a_n, b_n[$ mais par construction $]a_n, b_n[\subset O$. D'où l'égalité. \square

On va considérer les notations suivantes

- \mathcal{F} désigne l'ensemble des fermés.
- \mathcal{IB} désigne l'ensemble des intervalles bornés.
- \mathcal{I}_g désigne $\{[a, b[\text{ tel que } (a, b) \in \mathbb{R}^2\}$.
- \mathcal{I}_d désigne $]a, b] \text{ tel que } (a, b) \in \mathbb{R}^2\}$.
- \mathcal{IO} désigne l'ensemble des intervalles ouverts bornés.
- \mathcal{IF} désigne l'ensemble des intervalles fermés bornés.
- $\mathcal{IO}_{-\infty}$ désigne l'ensemble des demi droites ouvertes $] - \infty, b[$ pour un b dans \mathbb{R} .
- \mathcal{IO}_{∞} désigne l'ensemble des demi droites ouvertes $]a, +\infty[$ pour un a dans \mathbb{R} .
- $\mathcal{IF}_{-\infty}$ désigne l'ensemble des demi droites fermées $] - \infty, b]$ pour un b dans \mathbb{R} .
- \mathcal{IF}_{∞} désigne l'ensemble des demi droites fermées $]a, +\infty]$ pour un a dans \mathbb{R} .

Théorème 1.5.3.

$$\begin{aligned} \mathcal{B}(\mathbb{R}) &= \tau(\mathcal{O}) = \tau(\mathcal{F}) = \tau(\mathcal{IB}) = \tau(\mathcal{I}_g) = \tau(\mathcal{I}_d) = \tau(\mathcal{IO}) = \tau(\mathcal{IF}) \\ &= \tau(\mathcal{IO}_{-\infty}) = \tau(\mathcal{IO}_{\infty}) = \tau(\mathcal{IF}_{-\infty}) = \tau(\mathcal{IF}_{\infty}). \end{aligned}$$

1.5.2 Tribu image et tribu engendrée par une fonction

On présente ici deux notions, une anecdotique de tribu image et l'autre qui sera fondamentale de tribu engendrée.

Proposition et définition 1.5.4. *Si (Ω, \mathcal{A}) est un espace mesurable, Y un ensemble quelconque, et f une application de Ω dans Y , alors on définit l'ensemble \mathcal{S} par*

$$\mathcal{S} = \{S \in \mathcal{P}(Y) \mid f^{-1}(S) \in \mathcal{A}\}$$

Alors \mathcal{S} est une tribu sur Y , de plus $f \in \mathcal{L}^0((\Omega, \mathcal{A}), (Y, \mathcal{S}))$. On dit que \mathcal{A} est la tribu image de \mathcal{A} par f .

Pour rendre f mesurable, on aurait pu utiliser $\mathcal{P}(Y)$ la tribu grossière mais \mathcal{S} la tribu la moins grossière sur Y qui rende f mesurable.

Exemple 1.5.5. *On peut illustrer ce concept en prenant par exemple la fonction couleur (cf. page 4). $f : \Omega \rightarrow Y$ où $\Omega = \{0, 1, \dots, 36\}$ et où $Y = \{\text{vert, rouge, noir}\}$ tandis que \mathcal{A} est définie par (1.2.1). On vérifie que la tribu image de \mathcal{A} par f est bien la tribu discrète sur Y .*

Proposition et définition 1.5.6. *Si Ω est un ensemble quelconque, (Y, \mathcal{U}) un espace mesurable, et f une application de Ω dans Y , alors on définit l'ensemble \mathcal{A} (qui est contenu dans $\mathcal{P}(\Omega)$) par*

$$A \in \mathcal{A} \Leftrightarrow \exists B \in \mathcal{U} \text{ tel que } A = f^{-1}(B).$$

Alors \mathcal{A} est une tribu, de plus $f \in \mathcal{L}^0((\Omega, \mathcal{A}), (Y, \mathcal{U}))$. On dit que \mathcal{A} est la tribu engendrée par f et \mathcal{U} (ou simplement engendrée par f si \mathcal{U} est suffisamment implicite). On la note souvent $\sigma(f)$. Si X est une variable aléatoire alors $Y = \mathbb{R}$ (ou parfois $\overline{\mathbb{R}}$) muni de la tribu des boréliens, on note $\tau(X)$ la tribu engendrée par X .

Pour rendre f mesurable, on aurait pu utiliser $\mathcal{P}(\Omega)$ mais \mathcal{A} la plus petite tribu sur Ω qui rende f mesurable.

Remarque 1.5.7. *Dans le cadre ci-dessus, on peut énoncer l'équivalence entre*

- i) $\mathcal{A} = \sigma(f)$
- ii) $\left\{ \begin{array}{l} f \in \mathcal{L}^0((\Omega, \mathcal{A}), (Y, \mathcal{U})) \\ \text{Pour toute tribu } \mathcal{T} \text{ de } \Omega, f \in \mathcal{L}^0((\Omega, \mathcal{T}), (Y, \mathcal{U})) \Rightarrow \mathcal{A} \subset \mathcal{T} \end{array} \right.$

Exemple 1.5.8. *On peut illustrer ce concept en prenant par exemple la fonction couleur (cf. page 4). $f : \Omega \rightarrow Y$ où $\Omega = \{0, 1, \dots, 36\}$ et où $Y = \{\text{vert, rouge, noir}\}$. On considère sur Y la tribu $\mathcal{P}(Y)$. On vérifie que \mathcal{A} dans (1.2.1) est bien la tribu engendrée par f .*

1.5.3 Tribu trace

Comme pour la pour notion de topologie induite sur un sous ensemble, on peut construire une tribu sur un sous-ensemble.

Proposition et définition 1.5.9. *Soit (Ω, \mathcal{A}) un espace mesurable, soit $X_0 \subset \Omega$ un sous-ensemble non vide de Ω . On définit la famille de sous ensembles \mathcal{A}_{X_0} en prenant l'intersection avec X_0 :*

$$B \in \mathcal{A}_{X_0} = \{A \cap X_0 \mid A \in \mathcal{A}\}.$$

La famille \mathcal{A}_{X_0} constitue une tribu sur X_0 , appelée tribu trace (ou tribu induite) sur X_0

Exemple 1.5.10. *Si $(\Omega, \mathcal{A}) = (\{0, 1, \dots, 36\}, \{\emptyset, A_{\text{vert}}, A_{\text{rouge}}, A_{\text{noir}}, A_{\text{rouge ou noir}}, A_{\text{rouge ou vert}}, A_{\text{vert ou noir}}, \Omega\})$ (cf. page 4). Si on s'intéresse à $\Gamma = \{1, \dots, 36\}$, alors la tribu trace est simplement*

$$\mathcal{A}_\Gamma = \{\emptyset, A_{\text{rouge}}, A_{\text{noir}}, A_{\text{rouge ou noir}}\} = \{\emptyset, A_{\text{rouge}}, A_{\text{noir}}, \Gamma\}$$

Remarque 1.5.11. *Par exemple, sur \mathbb{R}_+ , on peut donc définir deux tribus naturelles :*

- *La tribu trace sur \mathbb{R}_+ de la tribu des boréliens de \mathbb{R} .*
- *La tribu des boréliens de la topologie trace sur \mathbb{R}_+ .*

Mais les deux coïncident.

Théorème 1.5.12. *Soit X un espace métrique définissant une topologie \mathcal{O} et X_0 une partie non vide de X . Si on munit X_0 de la topologie \mathcal{O}_{X_0} induite par \mathcal{O} sur X_0 , c'est-à-dire*

$$A \in \mathcal{O}_{X_0} \Leftrightarrow \text{il existe } O \in \mathcal{O} \text{ tel que } A = X_0 \cap O.$$

alors $\mathcal{B}(X_0) = \{X_0 \cap B \mid B \in \mathcal{B}(X)\}$.

Si de plus X_0 est mesurable alors les sous-ensembles mesurables de X_0 sont les sous-ensembles de X_0 qui sont mesurables, $\mathcal{B}(X_0) = \mathcal{P}(X_0) \cap \mathcal{B}(X)$.

1.5.4 Tribu produit

Un autre exemple fondamental de construction de tribus est fourni lorsqu'on manipule un produit cartésien de deux espaces mesurables. Par exemple, si $\Omega_1 = \{1, 2, 3\}$ et $\Omega_2 = \{a, b\}$ que l'on munit respectivement des tribus

$$\mathcal{A}_1 = \{\emptyset, \{1\}, \{2, 3\}, \{1, 2, 3\}\}$$

$$\mathcal{A}_2 = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}.$$

En termes d'interprétation, on ne sait pas distinguer 2 et 3. On peut écrire

$$\Omega_1 \times \Omega_2 = \{(1, a), (2, a), (3, a), (1, b), (2, b), (3, b)\},$$

Un autre exemple fondamental de construction de tribus est fourni lorsqu'on manipule un produit cartésien de deux espaces mesurables. Par exemple, $\Omega_1 = \{1, 2, 3\}$ muni de la tribu

$$\mathcal{A}_1 = \{\emptyset, \{1\}, \{2, 3\}, \{1, 2, 3\}\}$$

puis $\Omega_2 = \{a, b\}$ muni de la tribu

$$\mathcal{A}_2 = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}.$$

En termes d'interprétation, on ne sait pas distinguer 2 et 3. On peut écrire

$$\Omega_1 \times \Omega_2 = \{(1, a), (2, a), (3, a), (1, b), (2, b), (3, b)\},$$

L'ensemble $\mathcal{A}_1 \times \mathcal{A}_2$ constitué de parties de $\Omega_1 \times \Omega_2$ possède (a priori) au plus $4 \times 4 = 16$ éléments mais en fait moins, car l'ensemble vide apparaît plusieurs fois. Il possède en fait seulement 10 éléments qui sont

- $E_1 = \emptyset$, que l'on peut interpréter comme “ $X \neq 1$ et $X \neq 2$ et $X \neq 3$ ”
- $E_2 = \{(1, a)\}$, “ $X = 1$ et $Y = a$ ”

- $E_3 = \{(1, b)\}$, “ $X = 1$ et $Y = b$ ”
- $E_4 = \{(1, a), (1, b)\}$, “ $X = 1$ ”
- $E_5 = \{(2, a), (3, a)\}$, “($X = 2$ ou $X = 3$) et $Y = a$ ”
- $E_6 = \{(2, b), (3, b)\}$, “($X = 2$ ou $X = 3$) et $Y = b$ ”
- $E_7 = \{(2, a), (3, a), (2, b), (3, b)\}$, “ $X = 2$ ou $X = 3$ ”
- $E_8 = \{(1, a), (2, a), (3, a)\}$, “ $Y = a$ ”
- $E_9 = \{(1, b), (2, b), (3, b)\}$, “ $Y = b$ ”
- $E_{10} = \{(1, a), (2, a), (3, a), (1, b), (2, b), (3, b)\}$, “($X = 1$ ou $X = 2$ ou $X = 3$)”

Ce n'est pas une tribu, il manque par exemple le complémentaire de E_2 . On va donc considérer non pas $\mathcal{A}_1 \times \mathcal{A}_2$ mais $\tau(\mathcal{A}_1 \times \mathcal{A}_2)$. On a donc $(\mathcal{A}_1 \times \mathcal{A}_2)$ est inclu dans la tribu produit mais en général (sauf si l'une tribu est la tribu grossière) il n'y a pas égalité. Ce choix permettra ultérieurement de raisonner composante par composante.

Définition 1.5.13. Soit $(\Omega_1, \mathcal{A}_1)$ et $(\Omega_2, \mathcal{A}_2)$ deux espaces mesurables. On définit la tribu produit sur $\Omega_1 \times \Omega_2$ comme la tribu engendré par

$$\{A_1 \times A_2 \mid A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}.$$

On la note $\mathcal{A}_1 \otimes \mathcal{A}_2$.

Exemple 1.5.14. Il suffit de considérer $\Omega_1 = \{Haut, Bas\}$ et $\Omega_2 = \{Droite, Gauche\}$ muni respectivement de $\mathcal{A}_i = \mathcal{P}(\Omega_i)$. On a $\{(Haut, Droite), (Bas, Gauche)\} \notin \mathcal{A}_1 \times \mathcal{A}_2$ mais $\mathcal{A}_1 \otimes \mathcal{A}_2 = \mathcal{P}(\Omega_1 \times \Omega_2)$.

Théorème 1.5.15 (admis). Soit $(\Omega_1, \mathcal{O}_1)$ et $(\Omega_2, \mathcal{O}_2)$ deux espaces topologiques que l'on munit de la tribu des boréliens $\mathcal{B}(\Omega_1)$ et $\mathcal{B}(\Omega_2)$. Si les topologies sont à base dénombrables,

$$\mathcal{B}(\Omega_1) \otimes \mathcal{B}(\Omega_2) = \mathcal{B}(\Omega_1 \times \Omega_2).$$

Application : la tribu $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R})$.

1.5.5 Ensembles négligeables

Définition 1.5.16. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Un sous-ensemble A de Ω est dit négligeable (pour la mesure μ) si il existe $B \in \mathcal{A}$ tel que $B \subset A$ et $\mu(B) = 0$.

Définition 1.5.17. On considère $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Soit $f : \Omega \rightarrow \mathbb{R}$ une fonction mesurable. On dit que qu'une propriété est vraie μ -presque partout si l'ensemble des points qui ne vérifient pas la propriété est négligeable. Si la mesure $\mu = \mathbb{P}$ est une mesure de probabilité, on utilise plutôt le vocabulaire μ -presque sûrement. (Dans la pratique, ces deux termes sont interchangeables).

Par exemple, f est dite μ -presque partout nulle si $\exists A \in \mathcal{A}$ négligeable tel que $x \in A^c \Rightarrow f(x) = 0$. On dira aussi que f est presque partout nulle, p.p. nulle, si la mesure μ est suffisamment implicite.

Proposition 1.5.18. On considère $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Toute réunion dénombrable d'ensemble négligeable pour la mesure μ est un ensemble μ -négligeable.

Corollaire 1.5.19. Toute conjonction dénombrable de On considère $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Toute réunion dénombrable de propriétés μ -presque partout est une propriété μ -presque partout.

Par exemple, Si pour tout entier n , la fonction f_n est presque partout strictement positive, alors presque partout, toutes les f_n sont strictement positives.

Définition 1.5.20. Un espace mesuré $(\Omega, \mathcal{A}, \mu)$ pour lequel

$$[N \subset A \text{ avec } A \in \mathcal{A}, \mu(A) = 0] \Rightarrow [N \in \mathcal{A}]$$

est appelé un espace mesuré complet.

Si l'espace mesuré n'est pas complet, il existe un ensemble négligeable N qui n'appartient pas à \mathcal{A} . Mais il est interdit d'écrire que $\mu(N) = 0$ puisque formellement $\mu(N)$ n'existe pas. Il est désagréable de travailler avec un espace mesuré qui n'est pas complet.

Remarque 1.5.21. Par exemple, si A est un sous-ensemble de \mathbb{R} non mesurable au sens de Borel ($A \in \mathcal{P}(\mathbb{R}) \setminus \mathcal{B}(\mathbb{R})$), en conséquence la fonction indicatrice de A notée $\mathbf{1}_A$ n'est pas mesurable (elle n'appartient pas à $\mathcal{L}^0(\mathbb{R}, \mathcal{B}(\mathbb{R}), (\mathbb{R}, \mathcal{B}(\mathbb{R})))$ bien qu'elle soit μ -presque partout à la fonction nulle qui elle est mesurable. Ce n'est pas très pratique car fondamentalement cela représente le même comportement.

Mais on peut toujours étendre (compléter) le couple (tribu, mesure) à l'aide du résultat suivant.

Théorème 1.5.22. [Complétude d'une mesure]

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Il existe une tribu $\overline{\mathcal{A}}$ sur Ω et une mesure ν sur $\overline{\mathcal{A}}$ telles que

- Si $A \in \mathcal{A}$, alors $\nu(A) = \mu(A)$.
- Si N est μ -négligeable, on a $N \in \overline{\mathcal{A}}$ et $\nu(N) = 0$.

La tribu $\overline{\mathcal{A}}$ est alors appelée tribu complétée de \mathcal{A} et ν (parfois notée $\overline{\mu}$) est appelée mesure complétée de μ .

Proposition 1.5.23. Soit (Ω, \mathcal{A}) et (Y, \mathcal{F}) deux espaces mesurables. On suppose que la mesure μ rend l'espace mesuré $(\Omega, \mathcal{A}, \mu)$ complet alors si $f \in \mathcal{L}^0((\Omega, \mathcal{A})(Y, \mathcal{F}))$ et si $f=g$ μ -p.p. alors $g \in \mathcal{L}^0((\Omega, \mathcal{A})(Y, \mathcal{F}))$.

Démonstration. On considère $B \in \mathcal{F}$, on veut montrer que $X_2 := g^{-1}(B) \in \mathcal{A}$. On sait que $X_1 := f^{-1}(B) \in \mathcal{A}$. En général, ces deux ensembles sont différents car f diffère de g . Mais on sait que $N := \{\omega \in \Omega \mid f(\omega) \neq g(\omega)\}$ est de mesure nulle. Il suffit de remarquer que $X_2 \subset X_1 \cup N$ car si $\omega \in X_2$, $g(\omega) \in B$. On a l'alternative

$$\begin{cases} f(\omega) = g(\omega) & \Rightarrow f(\omega) \in B & \Rightarrow \omega \in X_1 \\ \text{ou} \\ f(\omega) \neq g(\omega) & & \Rightarrow \omega \in N \end{cases}$$

Puis on peut conclure à l'aide de la complétude de la tribu \mathcal{A} car

$$X_2 = (X_1 \cap N^C) \cup (N \cap X_2).$$

□

1.5.6 Tribu de Lebesgue

Au vu de la remarque 1.5.21, on n'utilisera pas seulement la mesure de Borel car elle n'est pas complète. En comparaison, $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda)$ est un espace mesuré complet. Pour autant, il existe des ensemble non mesurable au sens de Lebesgue : $\mathcal{L}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})$.

En appliquant le résultat général, on peut énoncer.

Définition 1.5.24. Soit \mathcal{N} l'ensembles des négligeables au sens de Borel. On appelle tribu de Lebesgue sur \mathbb{R} la tribu $\tau(\mathcal{N} \cup \mathcal{B}(\mathbb{R}))$. On la note $\mathcal{L}(\mathbb{R})$ (ou parfois $\overline{\mathcal{B}}(\mathbb{R})$).

Proposition 1.5.25. Soit A une partie de \mathbb{R} . On a l'équivalence entre

1. $A \in \mathcal{L}(\mathbb{R})$
2. il existe $N \in \mathcal{N}$ et $B \in \mathcal{B}(\mathbb{R})$ tels que $A = N \cup B$ (décomposition non unique).

Ces résultats existent dans un cadre plus général.

Proposition 1.5.26. Soit A une partie de \mathbb{R} . On a l'équivalence entre

1. A est μ -négligeable.
2. A est λ -négligeable.

Théorème 1.5.27. Mesure de Lebesgue.

Il existe une mesure λ sur $(\mathbb{R}, \mathcal{L}(\mathbb{R}))$ vérifiant

1. pour tout intervalle $]a, b[$, $\lambda(]a, b[) = b - a$
2. $\forall A \in \mathcal{B}(\mathbb{R})$, $\lambda(A) = \mu(A)$.
3. $\forall A \in \overline{\mathcal{B}}(\mathbb{R})$, $\forall x \in \mathbb{R}$, $\{y : y - x \in A\} \in \overline{\mathcal{B}}(\mathbb{R})$ et $\lambda(\{y : y - x \in A\}) = \lambda(A)$.

Cette mesure λ s'appelle la mesure de Lebesgue.

Proposition 1.5.28. Soit $A \in \mathcal{L}(\mathbb{R})$, on suppose qu'il existe deux décompositions $A = N_1 \cup B_1 = N_2 \cup B_2$ où les N_i sont négligeables, et B_i sont des boréliens, alors

1. $\mu(B_1) = \mu(B_2)$
2. cette valeur commune vaut $\lambda(A)$.

Exemple 1.5.29. Comme ces ensembles sont des boréliens, la mesure de Borel et la mesure de Lebesgue coïncident.

- Mesure de Lebesgue d'un singleton. Rappelons que $\{x\} \in \mathcal{B}(\mathbb{R})$, $\lambda(\{x\}) = \mu(\{x\}) = 0$.
- Mesure de Lebesgue d'un intervalle quelconque $\lambda([a, b]) = \mu([a, b]) = b - a$
- Mesure de Lebesgue de \mathbb{Q} . Puisque \mathbb{Q} est dénombrable. $\lambda(\mathbb{Q}) = \mu(\mathbb{Q}) = 0$.

Proposition 1.5.30. Soit $A \in \mathcal{L}(\mathbb{R})$, soient u et α deux constantes. On pose $B = \{u\} + A$ (ensemble translaté) et $C = \alpha A$, alors B et C sont dans $\mathcal{L}(\mathbb{R})$ et $\lambda(B) = \lambda(A)$, $\lambda(C) = |\alpha|\lambda(A)$.

1.5.7 Mesure et tribu de Lebesgue dans \mathbb{R}^n

Le cas $n = 1$ a déjà été traité. On se consacre au cas $n \geq 2$.

Proposition 1.5.31. La tribu $\mathcal{B}(\mathbb{R}^n)$ peut-être engendrée par

- les boules ouvertes de \mathbb{R}^n , $B(x, r)$
- les pavés fermés de \mathbb{R}^n , $\prod_{i=1}^n [a_i, b_i]$.
- les ensembles de la forme $\prod_{i=1}^n [a_i, b_i[$.
- les ensembles de la forme $\prod_{i=1}^n] - \infty, a_i]$.

Définition 1.5.32. (mesure de Borel sur \mathbb{R}^n). Sur la tribu $\mathcal{B}(\mathbb{R}^n)$, il existe une unique mesure μ_n telle que $\mu_n(\prod_{i=1}^n [a_i, b_i]) = \prod_{i=1}^n (b_i - a_i)$.

Définition 1.5.33. On appelle tribu de Lebesgue sur \mathbb{R}^n , la tribu complétée $\mathcal{L}(\mathbb{R}^n) = \overline{\mathcal{B}(\mathbb{R}^n)}$ et on note λ_n mesure de Lebesgue sur \mathbb{R}^n , la mesure complétée de la mesure de Borel en dimension n .

Proposition 1.5.34. La mesure de Lebesgue est invariante par translation. Soit $(a_1, \dots, a_n) \in \mathbb{R}_+^n$, alors

$$\lambda_n \left(\prod_{i=1}^n [0, a_i[\right) = \prod_{i=1}^n a_i.$$

Remarque 1.5.35. La complétée de la tribu produit n'est pas la tribu produit des complétés.

$$\overline{\mathcal{B}(\mathbb{R}^2)} \neq \overline{\mathcal{B}(\mathbb{R})} \otimes \overline{\mathcal{B}(\mathbb{R})}$$

Proposition 1.5.36. Soit φ un endomorphisme de \mathbb{R}^n alors pour tout $B \in \mathcal{L}(\mathbb{R}^n)$, $\varphi(B) \in \mathcal{L}(\mathbb{R}^n)$ et

$$\lambda_n(\varphi(B)) = |\det(\varphi)|\lambda_n(B).$$

en particulier si φ est une homothétie, ($\varphi = aId$), alors $\lambda_n(aB) = |a^n|\lambda_n(B)$.

En particulier, on en déduit que tout hyperplan est de mesure nulle.

1.5.8 Loi de probabilité sur les vecteurs aléatoires

Définition 1.5.37. On considère X_1 une variable aléatoire réelle définie sur un espace probabilisé $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$ et X_2 une variable aléatoire réelle définie sur un espace probabilisé $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$. On dit que X_1 et X_2 suivent la même loi si $\mathbb{P}_{X_1} = \mathbb{P}_{X_2}$, c'est-à-dire si pour tout borélien B de \mathbb{R} ,

$$\mathbb{P}_{X_1}(B) = \mathbb{P}_1(X_1 \in B) = \mathbb{P}_2(X_2 \in B) = \mathbb{P}_{X_2}(B).$$

Proposition 1.5.38. On considère X_1 et X_2 deux variables aléatoires réelles. Les variables X_1 et X_2 suivent la même loi si et seulement si $F_{X_1} = F_{X_2}$.

Ce résultat se généralise en dimension finie, on va l'énoncer dans le cas $n = 2$. La démonstration est similaire et se base là encore sur le lemme des classes monotones.

Proposition 1.5.39. *On considère $X = (X_1, X_2)$ et $Y = (Y_1, Y_2)$ deux variables aléatoires à valeurs dans \mathbb{R}^2 . Les variables X et Y suivent la même loi si et seulement si $F_X = F_Y$ (ce sont deux fonctions de $\mathbb{R}^2 \rightarrow [0, 1]$).*

Théorème 1.5.40 (admis⁵). *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ l'espace de probabilité $(]0, 1[, \mathcal{B}(]0, 1[, \mu_{\mathcal{B}(]0, 1[})$. On considère F une fonction croissante continue à droite telle que $\lim_{t \rightarrow -\infty} F(t) = 0$ et $\lim_{t \rightarrow +\infty} F(t) = 1$ alors il existe une variable aléatoire réelle définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ (et à valeurs dans $] -\infty, +\infty[$ telle que F soit la fonction de répartition de la loi X . On peut donner une construction explicite de X :*

$$X(\omega) = \inf\{x \in \mathbb{R} \mid F(x) \geq \omega\}.$$

Attention, $(\Omega =]0, 1[, \mathcal{A}, \mathbb{P})$ n'est pas le seul espace possible. La fonction X mentionnée est aussi appelée fonction quantile de la loi de X , fonction quantile de la fonction de répartition. En particulier, si F est continue et bijective. La fonction X est la fonction réciproque de F et $X(1/10)$ est le premier décile, $X(1/2)$ la médiane et $X(9/10)$ le dernier décile.

1.5.9 Construction de fonctions mesurables

Résultats généraux

Exemple 1.5.41. *Soit (Ω, \mathcal{A}) un espace mesurable. Alors l'application Identité est mesurable de (Ω, \mathcal{A}) dans (Ω, \mathcal{A}) .*

Exemple 1.5.42. *Soit (Ω, \mathcal{A}) et (Y, \mathcal{U}) deux espaces mesurables. Alors toute application constante est mesurable de (Ω, \mathcal{A}) dans (Y, \mathcal{U}) .*

Définition 1.5.43. *Soit $A \subset \Omega$. La fonction indicatrice de A est la fonction*

$$\begin{aligned} \mathbf{1}_A &: \Omega \rightarrow \{0, 1\} \\ x &\mapsto \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A. \end{cases} \end{aligned}$$

Il existe d'autres notations. Par exemple si $A = [0, 1] \subset \mathbb{R}$, on peut écrire $\mathbf{1}_A(x) = \mathbf{1}_{x \in [0, 1]} = \mathbf{1}_{0 \leq x \leq 1}$.

Exemple 1.5.44. *Soit (Ω, \mathcal{A}) un espace mesurable. Alors $A \in \mathcal{A}$ si et seulement si la fonction indicatrice de A est une fonction mesurable de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Si de plus, on considère un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ alors $\mathbf{1}_A$ est une variable aléatoire qui suit une loi de Bernoulli.*

Proposition 1.5.45. *Soit (Ω, \mathcal{A}) et (Y, \mathcal{U}) des espaces mesurables. Soit \mathcal{C} une partie génératrice de la tribu \mathcal{U} . Soit $f : \Omega \rightarrow Y$ telle que pour tout $C \in \mathcal{C}$, $f^{-1}(C) \in \mathcal{A}$, alors f est mesurable de (Ω, \mathcal{A}) dans (Y, \mathcal{U}) .*

Exemple 1.5.46. *On considère f continue de \mathbb{R} dans \mathbb{R} . Puisque les boréliens sont engendré par les ouverts, il suffit de vérifier que pour tout ouvert O de l'espace d'arrivée \mathbb{R} , $f^{-1}(O)$ est un borélien de l'espace de départ \mathbb{R} . C'est trivial, puisque cet ensemble est ouvert. Donc toute fonction continue de \mathbb{R} dans \mathbb{R} est mesurable de $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*

Proposition 1.5.47. *Soit (Ω, \mathcal{A}) un espace mesurable et X_0 un sous-ensemble non vide. On considère \mathcal{A}_0 la tribu trace de \mathcal{A} sur X_0 . Alors l'injection canonique est mesurable de (X_0, \mathcal{A}_0) dans (Ω, \mathcal{A}) .*

Corollaire 1.5.48. *Soit (Ω, \mathcal{A}) et (Y, \mathcal{U}) des espace mesurables. Soit X_0 un sous-ensemble non vide de Ω . On considère \mathcal{A}_0 la tribu trace de \mathcal{A} sur X_0 . Si f est mesurable de (Ω, \mathcal{A}) dans (Ω', \mathcal{A}') alors $f|_{X_0}$ est mesurable de (X_0, \mathcal{A}_0) dans (Y, \mathcal{U}) .*

5. Ce résultat est lié à la théorie de l'intégration de Stieltjes donc la preuve est omise.

Proposition 1.5.49. Soit (Ω, \mathcal{A}) et (Y, \mathcal{U}) des espace mesurables. Soit Y_0 un sous-ensemble non vide de Y . On considère \mathcal{U}_{Y_0} la tribu trace de \mathcal{U} sur Y_0 . Si f est mesurable de (Ω, \mathcal{A}) dans (Y_0, \mathcal{U}_{Y_0}) alors (avec un abus de notation) f est mesurable de (Ω, \mathcal{A}) dans (Y, \mathcal{U}) .

Exemple 1.5.50. Si f est mesurable de (Ω, \mathcal{A}) dans $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$. On peut considérer l'application g qui "coïncide avec f " mais qui va de Ω dans \mathbb{R} . Alors g est mesurable de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Lemme 1.5.51 (lemme de recollement). Soit (Ω, \mathcal{A}) et (Y, \mathcal{U}) des espaces mesurables et une fonction $f : \Omega \rightarrow Y$. On considère un ensemble $I \subset \mathbb{N}$ (donc au plus dénombrable) et une partition de Ω indexée par I dont les termes sont mesurables, $(X_n)_{n \in I}$. On note \mathcal{A}_n la tribu trace de \mathcal{A} sur X_n telle que pour chaque n dans I , la restriction de f à X_n soit mesurable de (X_n, \mathcal{A}_n) dans (Y, \mathcal{U}) alors f est mesurable de (Ω, \mathcal{A}) dans (Y, \mathcal{U}) .

Exemple 1.5.52. On considère la fonction partie entière de \mathbb{R} dans \mathbb{R} est mesurable.

Exemple 1.5.53. On considère la fonction f et g_α de \mathbb{R} dans \mathbb{R} définie par

$$f(x) = \begin{cases} x^2 & \text{si } x > 0; \\ -4 & \text{si } x = 0; \\ e^x & \text{si } x < 0. \end{cases} \quad g_\alpha(x) = \begin{cases} 1/x & \text{si } x \neq 0; \\ \alpha & \text{si } x = 0; \end{cases}$$

Proposition 1.5.54. Soient $(Y_1, \mathcal{U}_1), \dots, (Y_n, \mathcal{U}_n)$ des espaces mesurables. On considère le produit cartésien $Y = Y_1 \times \dots \times Y_n$ muni de la tribu produit $\mathcal{U} = \mathcal{U}_1 \otimes \dots \otimes \mathcal{U}_n$. Alors la jeme projection π_j est mesurable de (Y, \mathcal{U}) dans (Y_j, \mathcal{U}_j) .

Théorème 1.5.55. Soit (Ω, \mathcal{A}) un espace mesurable. Soient $(Y_1, \mathcal{U}_1), \dots, (Y_n, \mathcal{U}_n)$ des espaces mesurables. On considère le produit cartésien $Y = Y_1 \times \dots \times Y_n$ muni de la tribu produit $\mathcal{U} = \mathcal{U}_1 \otimes \dots \otimes \mathcal{U}_n$. Soit $f = (f_1, \dots, f_n)$ de Ω dans Y . Alors on a l'équivalence entre

- f est mesurable de (Ω, \mathcal{A}) dans (Y, \mathcal{U}) .
- pour tout j , f_j est mesurable de (Ω, \mathcal{A}) dans (Y_j, \mathcal{U}_j) .

Résultats dans le cas d'une fonction à valeurs dans un métrique

On peut généraliser la démarche de l'exemple 1.5.46.

Proposition 1.5.56. Soit (Ω, \mathcal{A}) un espace mesurable et Y un espace métrique muni de $\mathcal{B}(Y)$. Alors on a l'équivalence entre

- f est mesurable de (Ω, \mathcal{A}) dans $(Y, \mathcal{B}(Y))$.
- Pour tout ouvert O de Y , $f^{-1}(O) \in \mathcal{A}$.
- Pour tout fermé F de Y , $f^{-1}(F) \in \mathcal{A}$.

Définition 1.5.57. Soit Ω et Y deux espaces métriques. Si $f : \Omega \rightarrow Y$ on dit que f est une fonction borélienne si f est mesurable de $(\Omega, \mathcal{B}(\Omega))$ dans $(Y, \mathcal{B}(Y))$.

Théorème 1.5.58. Soit Ω et Y deux espaces métriques. Si $f : \Omega \rightarrow Y$ est continue alors elle est borélienne.

Résultats dans le cas d'une fonction numérique

Proposition 1.5.59. Soit (Ω, \mathcal{A}) un espace mesurable et $f : \Omega \rightarrow \mathbb{R}$ alors on a l'équivalence entre

- f est mesurable de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$;
- pour tout $(a, b) \in \mathbb{R}^2$, $f^{-1}(]a, b]) \in \mathcal{A}$;
- pour tout $(a, b) \in \mathbb{R}^2$, $f^{-1}([a, b]) \in \mathcal{A}$;
- pour tout $(a, b) \in \mathbb{R}^2$, $f^{-1}(]a, b[) \in \mathcal{A}$;
- pour tout $(a, b) \in \mathbb{R}^2$, $f^{-1}([a, b[) \in \mathcal{A}$;
- pour tout $a \in \mathbb{R}$, $f^{-1}([a, +\infty[) \in \mathcal{A}$;
- pour tout $a \in \mathbb{R}$, $f^{-1}(]a, +\infty]) \in \mathcal{A}$;
- pour tout $b \in \mathbb{R}$, $f^{-1}(]-\infty, b]) \in \mathcal{A}$;
- pour tout $b \in \mathbb{R}$, $f^{-1}(]-\infty, b[) \in \mathcal{A}$.

Avec cette caractérisation, on peut établir le résultat suivant.

Théorème 1.5.60. Soit (f, g) appartenant à $\mathcal{L}^0((\Omega, \mathcal{A}), (\mathbb{R}, \mathcal{B}(\mathbb{R})))$ et soit φ une fonction mesurable au sens Borel (par exemple continue) de $\mathbb{R} \rightarrow \mathbb{R}$, alors $\varphi(f)$ est mesurable de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Soit Ψ une fonction mesurable (par exemple continue) de $\mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, alors $\Psi(f, g)$ est mesurable de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Exemple 1.5.61. Soit (f, g) appartenant à $\mathcal{L}^0((\Omega, \mathcal{A}), (\mathbb{R}, \mathcal{B}(\mathbb{R})))$. On peut en déduire la mesurabilité des fonctions

$$\begin{array}{cccc} |f| & c.f \text{ si } c \in \mathbb{R} & f^+ & f^- \\ f+g & f.g & \max(f, g) & \min(f, g) \end{array}$$

Si de plus pour tout $\omega \in \Omega$, $f(\omega) \neq 0$ alors $1/f$ est mesurable de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ en utilisant la fonction mesurable (α est un nombre réel arbitraire)

$$\varphi(t) = \begin{cases} 1/t & \text{si } t \neq 0; \\ \alpha & \text{si } t = 0. \end{cases}$$

Exemple 1.5.62. Si X et Y sont deux variables aléatoires réelles définies sur le même espace de probabilité, alors $|X|$ est une v.a.r., $X + Y$ est une v.a.r., ...

Résultats asymptotiques

On donnera un énoncé volontairement simplifié.

Théorème 1.5.63. Soit $(f_n)_{n \in \mathbb{N}} \in (\mathcal{L}^0(\Omega, \mathcal{A}, \mathbb{R}))^{\mathbb{N}}$. On suppose qu'il existe M tel que pour tout n , et pour tout ω $|f_n(\omega)| \leq M$. Alors on a

- $\inf_n f_n \in \mathcal{L}^0(\Omega, \mathcal{A}, \mathbb{R})$;
- $\sup_n f_n \in \mathcal{L}^0(\Omega, \mathcal{A}, \mathbb{R})$;
- $\limsup_n f_n \in \mathcal{L}^0(\Omega, \mathcal{A}, \mathbb{R})$;
- $\liminf_n f_n \in \mathcal{L}^0(\Omega, \mathcal{A}, \mathbb{R})$;
- si la suite de fonctions $(f_n)_n$ converge simplement vers f sur Ω , alors $f \in \mathcal{L}^0(\Omega, \mathcal{A}, \mathbb{R})$;

Exemple 1.5.64. Transcription en termes de variables aléatoires.

1.5.10 Résultats sur $\overline{\mathbb{R}}$

En fait, dans la pratique, le dernier résultat peut être difficilement applicable car en général la suite (f_n) n'est pas bornée, et donc la limite supérieure peut prendre la valeur $+\infty$. Pour un résultat rigoureux, on peut considérer $\overline{\mathbb{R}}$ en le munissant d'une métrique compatible avec la topologie usuelle de \mathbb{R} . On considère par exemple la fonction de répartition Φ de la loi normale centrée réduite (qui est strictement croissante) que l'on prolonge en posant $\Phi(-\infty) = 0$ et $\Phi(+\infty) = 1$. On définit une distance sur $\overline{\mathbb{R}}$ par pour tout (α, β) de $\overline{\mathbb{R}}$,

$$d(x, y) = |\Phi(\alpha) - \Phi(\beta)|.$$

Au sens de $\overline{\mathbb{R}}$, $[-\infty, a[$ est un voisinage ouvert du point $-\infty$, de même $]b, +\infty]$ est un voisinage ouvert du point $+\infty$. On vérifie par exemple qu'une suite $(u_n)_{n \in \mathbb{N}}$ à valeurs dans $\overline{\mathbb{R}}$ (ou même dans \mathbb{R}) converge vers $-\infty$ si et seulement si $d(u_n, -\infty) \rightarrow 0$.

Cette compatibilité a pour conséquence que la tribu des boréliens sur $\overline{\mathbb{R}}$ est la tribu induite par la tribu des boréliens sur \mathbb{R} , $\mathcal{B}(\overline{\mathbb{R}}) = \{\mathbb{R} \cap B \mid B \in \mathcal{B}(\mathbb{R})\}$.

On va considérer les notations suivantes

- $\overline{\mathcal{I}\mathcal{O}}_{-\infty}$ désigne l'ensemble des $[-\infty, b[$ pour un b dans \mathbb{R} .
- $\overline{\mathcal{I}\mathcal{O}}_{\infty}$ désigne l'ensemble des $]a, +\infty]$ pour un a dans \mathbb{R} .
- $\overline{\mathcal{I}\mathcal{F}}_{-\infty}$ désigne l'ensemble des $[-\infty, b]$ pour un b dans \mathbb{R} .
- $\overline{\mathcal{I}\mathcal{F}}_{\infty}$ désigne l'ensemble des $[a, +\infty]$ pour un a dans \mathbb{R} .

Théorème 1.5.65.

$$\mathcal{B}(\overline{\mathbb{R}}) = \tau(\overline{\mathcal{I}\mathcal{O}}_{-\infty}) = \tau(\overline{\mathcal{I}\mathcal{F}}_{-\infty}) = \tau(\overline{\mathcal{I}\mathcal{O}}_{\infty}) = \tau(\overline{\mathcal{I}\mathcal{F}}_{\infty})$$

On utilisera la notation réduite $\mathcal{L}^0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ pour $\mathcal{L}^0((\Omega, \mathcal{A}), (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})))$.

Exemple 1.5.66. Soit $\alpha \in \overline{\mathbb{R}}$ fixé. On considère la fonction de $\overline{\mathbb{R}} \times \overline{\mathbb{R}}$ dans $\overline{\mathbb{R}}$ définie par

$$\varphi_\alpha(u, v) = \begin{cases} \alpha & \text{si } (u, v) \in \{(-\infty, +\infty)\{-\infty\} \cup \{(+\infty, -\infty)\} \\ u + v & \text{sinon} \end{cases}$$

Il suffit d'appliquer le lemme de recollement pour montrer que la fonction φ_α est mesurable de $(\overline{\mathbb{R}} \times \overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}^2))$ dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ puisque chacun des "morceaux" est mesurable et les restrictions sont continues.

Proposition 1.5.67. Soit (Ω, \mathcal{A}) un espace mesurable et $f : \Omega \rightarrow \mathbb{R}$ alors on a (modulo un abus de notation) l'équivalence entre

- f est mesurable de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$;
- f est mesurable de (Ω, \mathcal{A}) dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$.

Proposition 1.5.68. Soit (f, g) appartenant à $\mathcal{L}((\Omega, \mathcal{A}, \overline{\mathbb{R}})$ et soit φ une fonction mesurable (par exemple continue) de $\overline{\mathbb{R}} \rightarrow \overline{\mathbb{R}}$, alors $\varphi(f)$ est mesurable de (Ω, \mathcal{A}) dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$. Soit Ψ une fonction mesurable (par exemple continue) de $\overline{\mathbb{R}} \times \overline{\mathbb{R}} \rightarrow \overline{\mathbb{R}}$, alors $\Psi(f, g)$ est mesurable de (Ω, \mathcal{A}) dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$.

Exemple 1.5.69. Donc si (f, g) appartenant à $\mathcal{L}((\Omega, \mathcal{A}, \overline{\mathbb{R}})$, on a

- $|f| \in \mathcal{L}^0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$;
- fixons arbitrairement $\alpha \in \overline{\mathbb{R}}$ alors la fonction de Ω dans $\overline{\mathbb{R}}$ définie par

$$\omega \mapsto \begin{cases} f(\omega) + g(\omega) & \text{si } (f(\omega), g(\omega)) \notin \{(-\infty, +\infty), (+\infty, -\infty)\} \\ \alpha & \text{sinon} \end{cases}$$

appartient à $\mathcal{L}^0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$.

- fixons arbitrairement β et δ dans $\overline{\mathbb{R}}$ alors la fonction de Ω dans $\overline{\mathbb{R}}$ définie par

$$\omega \mapsto \begin{cases} f(\omega)g(\omega) & \text{si } (f(\omega), g(\omega)) \notin \{(0, +\infty), (+\infty, 0), (0, -\infty), (-\infty, 0)\} \\ \beta & \text{si } (f(\omega), g(\omega)) \in \{(0, +\infty), (+\infty, 0)\} \\ \delta & \text{si } (f(\omega), g(\omega)) \notin \{(0, -\infty), (-\infty, 0)\} \end{cases}$$

appartient à $\mathcal{L}^0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$;

- fixons arbitrairement γ dans $\overline{\mathbb{R}}$ alors la fonction de Ω dans $\overline{\mathbb{R}}$ définie par

$$\omega \mapsto \begin{cases} 1/f(\omega) & \text{si } f(\omega) \neq 0 \\ \gamma & \text{si } f(\omega) = 0 \end{cases}$$

appartient à $\mathcal{L}^0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$;

- $c \cdot f \in \mathcal{L}^0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$ si $c \in \mathbb{R}$;
- $f^+ \in \mathcal{L}^0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$;
- $f^- \in \mathcal{L}^0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$;
- $\max(f, g) \in \mathcal{L}^0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$;
- $\min(f, g) \in \mathcal{L}^0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$;

Résultats asymptotiques

Théorème 1.5.70. Soit $(f_n)_{n \in \mathbb{N}} \in (\mathcal{L}^0(\Omega, \mathcal{A}, \overline{\mathbb{R}}))^{\mathbb{N}}$. Alors on a

- $\inf_n f_n \in \mathcal{L}^0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$;
- $\sup_n f_n \in \mathcal{L}^0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$;
- $\limsup_n f_n \in \mathcal{L}^0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$;
- $\liminf_n f_n \in \mathcal{L}^0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$;
- si la suite de fonctions $(f_n)_n$ converge simplement vers f sur Ω , alors $f \in \mathcal{L}^0(\Omega, \mathcal{A}, \overline{\mathbb{R}})$;

Exemple 1.5.71. Transcription en termes de variables aléatoires.

Chapitre 2

Intégration

2.1 Fonctions étagées

2.1.1 Définition

On se donne un espace mesuré $(\Omega, \mathcal{A}, \mu)$.

Définition et notation 2.1.1. Soit $f : \Omega \rightarrow \mathbb{R}$. On dit que f est étagée s'il existe une famille finie (A_1, \dots, A_n) telle que

- les A_i forment une partition \mathcal{A} -mesurable finie, (A_1, \dots, A_n) de \mathcal{A} telle que de Ω (ce qui veut dire que A_1, \dots, A_n sont deux à deux disjoints et que $\Omega = \bigsqcup_{1 \leq i \leq n} A_i$)
- $\forall i \in \{1, \dots, n\}, \exists a_i$ tel que $f(x) = a_i, \forall x \in A_i$.

Si f est étagée sur une partition mesurable, on dira que cette partition est adaptée à f . On note $PMF_{\mathcal{A}}(A)$ l'ensemble des partition \mathcal{A} -mesurable finie, où $PMF(A)$, partition mesurable finie, s'il n'y a pas d'ambiguïtés sur \mathcal{A} .

Remarque 2.1.2. Si f est une fonction étagée définie avec une partition A_1, \dots, A_n , il peut exister une autre partition B_1, \dots, B_m (différente de A_1, \dots, A_n) telle que f est constante sur chacun des B_i . Le concept de fonction étagée ne fait pas intervenir la mesure μ dans la définition mais dans son utilisation.

Proposition 2.1.3. Si $f : \Omega \rightarrow \mathbb{R}$ alors on a l'équivalence (voir TD).

- f est étagée ;
- $f \in \mathcal{L}(\Omega, \mathcal{A}, \mathbb{R})$ et $f(\Omega)$ fini.

En particulier, les fonctions indicatrices d'ensemble mesurable, par exemple $\mathbf{1}_{\mathbb{Q}}$ (voir TD).

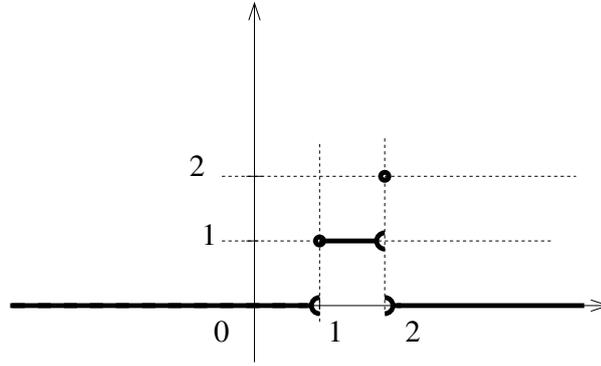
Lemme 2.1.4. Si f (respectivement g) est étagée par rapport à la partition (A_1, \dots, A_n) (respectivement la partition (B_1, \dots, B_p)), il existe une partition commune (C_1, \dots, C_q) adaptée à f et à g .

Corollaire 2.1.5. L'ensemble des fonctions étagées constitue un espace vectoriel. De plus le produit de deux fonctions étagées est étagée.

Exemple 2.1.6. La fonction partie entière noté Ent n'est pas étagée. La fonction

$$f : \mathbb{R} \rightarrow \mathbb{R} \\ x \mapsto \begin{cases} 0 & \text{si } x < 0 \\ \text{Ent}(x) & \text{si } x \in [0, 2] \\ 0 & \text{sinon} \end{cases}$$

est une fonction positive étagée. En effet, elle est constante sur $] -\infty, 0[$, $[0, 1[$, $[1, 2[$, $\{2\}$, $]2, +\infty[$.

FIGURE 2.1 – Dessin de f .

Avec des fonctions indicatrice, nous pouvons écrire f de manière plus compacte :

$$f(x) = \text{Ent}(x) \times \mathbf{1}_{[0,2]}(x) = \mathbf{1}_{[1,2]}(x) + 2 \times \mathbf{1}_{\{2\}}(x).$$

2.1.2 Approximation par des fonctions étagées

Les fonctions étagées permettent d'approximer les fonctions mesurables

Proposition 2.1.7. Soit $f \in \mathcal{L}^0(\Omega, \mathcal{A}, \overline{\mathbb{R}}_+)$ une fonction mesurable. Alors il existe une suite croissante de fonctions (mesurables) étagées qui converge simplement (ponctuellement) vers f sur Ω .

Démonstration. Pour tout $n \in \mathbb{N}$, on commence par définir $\varphi_n : \overline{\mathbb{R}}_+ \rightarrow \overline{\mathbb{R}}_+$, par

$$\varphi_n(t) = \begin{cases} 2^{-n} \text{Ent}(2^n t) & \text{si } t \in [0, n] \\ n & \text{si } t > n \end{cases}$$

où $E(t)$ désigne la partie entière de t . Il est clair que φ_n est étagée telle que $0 \leq \varphi_n(t) \leq \varphi_{n+1}(t)$, pour tout t de $\overline{\mathbb{R}}$ et pour tout n . On pose pour chaque n , $f_n = \varphi_n \circ f$, f_n alors est étagée, pour tout n , $f_n \leq f_{n+1}$ et pour tout $\omega \in \Omega$, $f_n(\omega) \rightarrow f(\omega)$.

En effet, si $f(\omega) = +\infty$, on a pour tout n , $f_n = n$. D'autre part, si $f(\omega) < +\infty$, à partir d'un certain rang $f_n(\omega) = 2^{-n} \text{Ent}(2^n f(\omega))$ donc $0 \leq f(\omega) - f_n(\omega) \leq 2^{-n}$. \square

2.1.3 Intégrale d'une fonction étagée positive

Définition et notation 2.1.8. Soit f une fonction positive étagée associée à une partition A_1, \dots, A_n . On va noter a_i la valeur commune de f sur A_i . On appelle intégrale de f sur Ω par rapport à la mesure μ le nombre suivant

$$\int_{\Omega} f(x) d\mu(x) := \sum_{i=1}^n a_i \mu(A_i).$$

Ce nombre peut être $+\infty$ et la fonction est dite non intégrable tandis que si ce nombre est fini ($\int_{\Omega} f(x) d\mu(x) < +\infty$) la fonction f est dite intégrable sur Ω .

De plus, si $A \subset \Omega$, alors appelle intégrale de f sur A par rapport à μ le nombre suivant

$$\int_A f(x) d\mu(x) := \int_{\Omega} \mathbf{1}_A(x) f(x) d\mu(x).$$

Exemple 2.1.9. Si $A \in \mathcal{A}$, alors $f = \mathbf{1}_A$ est étagée et de plus

$$\int_{\Omega} f(x) d\mu(x) = 0 \times \mu(\Omega \setminus A) + 1 \times \mu(A) = \mu(A) = \int_{\Omega} \mathbf{1}_A(x) d\mu(x).$$

Attention, on a utilisé la convention $0 \times +\infty = 0$.

Remarque 2.1.10. La valeur de $\int_{\Omega} f(x)d\mu(x)$ est indépendante de la partition associée à f . On la note également $\int_{\Omega} f d\mu$ voire même $\int_{\Omega} f(x) d\mu(x)$ ou $\int_{\Omega} f(\omega) d\mu(\omega)$.

Par exemple, si $\Omega_1 \subset \Omega$ peut se décomposer $\Omega_1 = \Omega_2 \sqcup \Omega_3$ alors on peut envisager les partitions $(\Omega_1, \Omega \setminus \Omega_1)$ et $(\Omega_2, \Omega_3, \Omega \setminus \Omega_1)$ de Ω qui sont deux partitions adaptées pour la fonction f qui vaut $\alpha \mathbf{1}_{\Omega_1}$. Puisque on a à la fois $f = \alpha \mathbf{1}_{\Omega_1}$ et $f = \alpha \mathbf{1}_{\Omega_2} + \alpha \mathbf{1}_{\Omega_3}$ et

$$\int_{\Omega} f(x)d\mu(x) = \alpha\mu(\Omega_1) = \alpha\mu(\Omega_2) + \alpha\mu(\Omega_3).$$

Remarque 2.1.11. Si f est étagée et si X et Y sont disjoints, il est trivial de vérifier que

$$\int_{X \sqcup Y} f(x)d\mu(x) = \int_X f(x)d\mu(x) + \int_Y f(x)d\mu(x).$$

Proposition 2.1.12. En généralisant la démarche ci-dessus, si f et g sont deux fonctions étagées et si $c \in \mathbb{R}$, il est trivial de vérifier que

$$\begin{aligned} \int_{\Omega} cf(x)d\mu(x) &= c \int_{\Omega} f(x)d\mu(x) \\ \int_{\Omega} (f+g)(x)d\mu(x) &= \int_{\Omega} f(x)d\mu(x) + \int_{\Omega} g(x)d\mu(x) \\ \text{si } f \leq g, \int_{\Omega} f(x)d\mu(x) &\leq \int_{\Omega} g(x)d\mu(x) \end{aligned}$$

2.1.4 Le cas particulier de $\Omega = \mathbb{N}$

Sur \mathbb{N} , on travaille systématiquement en considérant par défaut le cas de l'espace mesuré $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu_C)$ où μ_C désigne la mesure de comptage. Puisque la tribu est $\mathcal{P}(\mathbb{N})$ toute fonction est mesurable. De plus, on rappelle que toute suite positive s'identifie avec une fonction f positive définie sur \mathbb{N} , $(u_n)_n = (f(n))_n$.

Une fonction de \mathbb{N} dans \mathbb{R} , n'est en général pas étagée mais pour toute est une partie finie I de \mathbb{N} , $f\mathbf{1}_I$ est une fonction étagée. Si on note $I = \{n_1, \dots, n_p\}$. Puisque l'on peut décomposer en singletons $I = \{n_1, \dots, n_p\} = A_1 \sqcup A_2 \sqcup \dots \sqcup A_p$. Puisque

$$(f\mathbf{1}_I)(n) = \begin{cases} a_1 & := u_{n_1} = f(n_1) & \text{si } n \in A_1 \\ \dots & \dots & \dots \\ a_p & := u_{n_p} = f(n_p) & \text{si } n \in A_p \\ a_{p+1} & := 0 & \text{si } n \in A_{p+1} \text{ où } A_{p+1} := \mathbb{N} \setminus I \end{cases}$$

Par définition, $\int_I f d\mu_C = \int_{\mathbb{N}} (f\mathbf{1}_I)(x) d\mu_C(x) := \sum_{i=1}^{p+1} a_i \text{card}(A_i)$. Puisque a_{p+1} est nul, et que les ensembles A_1, \dots, A_p sont des singletons

$$\int_{\{n_1, \dots, n_p\}} f(x)\mu_C(dx) = \sum_{i=1}^p u_{n_i} = \sum_{n \in I} u_n$$

Dans le cas particulier où f est étagée positive, il existe $(b_1, \dots, b_p) \in \mathbb{R}^p$ et une partition (A_1, \dots, A_p) de \mathbb{N} telle que $f|_{A_i} = b_i$, alors on a dans $\overline{\mathbb{R}}_+$ un lien entre série et intégrale (cf. partie permutation du cours sur les séries)

$$\int_{\mathbb{N}} f(x)d\mu_C(x) = \int_{\mathbb{N}} f(x) \text{card}(dx) := \sum_{i=1}^p b_i \text{card}(B_i) = \sum_{n=0}^{+\infty} f(n).$$

2.1.5 Applications aux variables aléatoires réelles sur Ω fini

On regarde ici le cas classique où $\Omega = \{1, \dots, n\}$ est muni de la tribu discrète $\mathcal{A} = \mathcal{P}(\Omega)$. Dans ce cas, \mathbb{P} une mesure de probabilité qui est entièrement déterminée par le n -uplet (p_1, \dots, p_n) (dont les composantes valent respectivement $p_i = \mathbb{P}(\{i\})$) dont la connaissance

suffit à connaître la valeur sur la tribu. Pour insister sur l'interprétation variable aléatoire réelle, on va appeler X plutôt que f la fonction. Là encore, toute fonction $X : \Omega \rightarrow \mathbb{R}_+$ est mesurable et même étagée et en utilisant la même argumentation qu'au paragraphe précédent, on peut calculer l'intégrale en appliquant la définition,

$$\int_{\Omega} X d\mathbb{P} := \sum_{i=1}^n X(i)\mathbb{P}(\{i\}) = \sum_{i=1}^n p_i X(i).$$

On a tout de suite une interprétation de cette intégrale qui est un cas particulier de l'espérance, comme une moyenne pondérée des valeurs prises par X .

De plus, si on appelle $\tilde{x}_1, \dots, \tilde{x}_p$ les valeurs prises par X , on peut analyser X en regroupant les différents ω pour lesquels $X(\omega)$ est le même. Si on pose $A_i = \{\omega \in \Omega \mid X(\omega) = \tilde{x}_i\}$, on peut remarquer que $X = \tilde{x}_1 \mathbf{1}_{A_1} + \dots + \tilde{x}_p \mathbf{1}_{A_p}$ donc l'intégrale vaut également

$$\mathbb{E}(x) := \int_{\Omega} X d\mathbb{P} := \sum_{i=1}^p \tilde{x}_i \mathbb{P}(A_i) = \sum_{i=1}^p \tilde{x}_i \mathbb{P}(X = \tilde{x}_i).$$

On se rend compte que cette intégrale vaut l'espérance (au sens usuel) de la variable aléatoire.

2.2 Fonctions mesurables et intégrales

2.2.1 Intégrales des fonctions mesurables positives

Définition 2.2.1. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Si $f : \Omega \rightarrow [0, +\infty]$ est mesurable (par rapport aux tribus \mathcal{A} et $\mathcal{B}(\mathbb{R})$) positive, l'intégrale de f sur Ω par rapport à la mesure μ est définie par

$$\int_{\Omega} f d\mu = \int_{\Omega} f(x) d\mu(x) := \sup_{\phi \in \mathcal{E}(f)} \int_{\Omega} \phi(x) d\mu(x) \quad (2.2.1)$$

où $\mathcal{E}(f)$ désigne l'ensemble non vide $\{\phi \text{ étagée positive} \mid \phi(x) \leq f(x), \forall x \in \Omega\}$.

Pour $B \in \mathcal{A}$, on note

$$\int_B f(x) d\mu(x) = \int_{\Omega} f(x) \mathbf{1}_B(x) d\mu(x).$$

Définition 2.2.2. Cette intégrale peut prendre sa valeur dans $[0, +\infty]$. Une fonction mesurable positive f est dite intégrable si

$$\int_{\Omega} f(x) d\mu(x) < +\infty.$$

Exemple 2.2.3. On vérifie :

- La fonction $\mathbf{1}_{[0,1]}$ est intégrable mais elle n'est pas étagée.
- La fonction $x \mapsto x \mathbf{1}_{\mathbb{R}_+}(x)$ n'est pas intégrable.
- La fonction $\mathbf{1}_{\mathbb{Q}}$ est intégrable.
- La fonction $\mathbf{1}_{\mathbb{R} \setminus \mathbb{Q}}$ n'est pas intégrable.

Remarque 2.2.4. La définition 2.2.1 est cohérente car si f est étagée, alors le terme de droite dans l'équation 2.2.1 vaut bien $\int_{\Omega} f d\mu$. En général, les fonctions étagées ne sont pas intégrables, (par exemple, $f = \mathbf{1}_{\mathbb{R}_+}$ pour la mesure de Borel) mais elles le sont si μ est une mesure de probabilité (voire même une mesure finie).

Proposition 2.2.5. Croissance de l'intégrale.

Soient f, g deux fonctions positives mesurables sur $(\Omega, \mathcal{A}, \mu)$. Si $f \leq g$ (ce qui veut dire $f(\omega) \leq g(\omega), \forall \omega \in \Omega$) alors $\int_{\Omega} f(x) d\mu(x) \leq \int_{\Omega} g(x) d\mu(x)$.

Démonstration. Nous avons $\mathcal{E}(f) \subset \mathcal{E}(g)$ car $f \leq g$. Donc

$$\sup_{\phi \in \mathcal{E}(f)} \int_{\Omega} \phi(x) d\mu(x) \leq \sup_{\phi \in \mathcal{E}(g)} \int_{\Omega} \phi(x) d\mu(x).$$

□

Cette proposition admet comme corollaire le théorème suivant.

Théorème 2.2.6. *Théorème de comparaison.*

Soient f, g deux fonctions positives mesurables sur $(\Omega, \mathcal{A}, \mu)$. Si $f \leq g$ et g est intégrable alors f est intégrable.

Théorème 2.2.7. *Pseudo-linéarité de l'intégrale des fonctions positives.*

Soient f, g deux fonctions positives mesurables sur $(\Omega, \mathcal{A}, \mu)$ et $a \geq 0$, alors :

$$\int_{\Omega} (f(x) + g(x)) d\mu(x) = \int_{\Omega} f(x) d\mu(x) + \int_{\Omega} g(x) d\mu(x)$$

et

$$\int_{\Omega} a f(x) d\mu(x) = a \int_{\Omega} f(x) d\mu(x) .$$

En particulier, si f et g sont intégrables alors $f + g$ aussi.

Théorème 2.2.8. *Inégalité de Markov.*

Soit f une fonction positive mesurable sur $(\Omega, \mathcal{A}, \mu)$. Soit $a > 0$. Alors :

$$\mu(\{x \in \Omega : f(x) \geq a\}) \leq \frac{1}{a} \int_{\Omega} f(x) d\mu(x) .$$

Démonstration. On a $\mathbf{1}_{\{y: f(y) \geq a\}} \leq f$ donc par théorème de comparaison (théorème 2.2.6) :

$$\int_{\Omega} a \mathbf{1}_{\{y: f(y) \geq a\}}(x) d\mu(x) \leq \int_{\Omega} f(x) d\mu(x) .$$

La fonction $a \mathbf{1}_{\{y: f(y) \geq a\}}$ est une fonction étagée et on calcule son intégrale :

$$\int_{\Omega} a \mathbf{1}_{\{y: f(y) \geq a\}}(x) d\mu(x) = a \times \mu(\{y : f(y) \geq a\}) + 0 \times \mu(\{y : f(y) < a\}) .$$

D'où le résultat. □

Proposition 2.2.9. Soit f une fonction positive mesurable sur $(\Omega, \mathcal{A}, \mu)$. Soit X, Y disjoints de Ω :

$$\int_{X \sqcup Y} f(x) d\mu(x) = \int_X f(x) d\mu(x) + \int_Y f(x) d\mu(x) .$$

Définition 2.2.10. Soit ν une mesure sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. La mesure ν est dite avoir pour densité¹ la fonction $f \geq 0$ sur \mathbb{R}^n (par rapport à λ_n) si $\forall B \in \mathcal{B}(\mathbb{R}^n)$,

$$\nu(B) = \int_B f(x) d\lambda_n .$$

2.2.2 Intégrales des fonctions mesurables de signe quelconque.

Soit un espace mesuré $(\Omega, \mathcal{A}, \mu)$. Soit $f : \Omega \rightarrow \mathbb{R}$ mesurable. Elle peut toujours s'écrire $f = f^+ - f^-$ avec f^+ et f^- mesurables positives :

$$f^+(x) = \begin{cases} f(x) & \text{si } f(x) \geq 0 \\ 0 & \text{sinon} \end{cases} \quad f^-(x) = \begin{cases} 0 & \text{si } f(x) \geq 0 \\ -f(x) & \text{sinon.} \end{cases}$$

1.

Remarque 2.2.11. Si on utilise la proposition 2.1.7, le théorème de convergence monotone permet de déduire que $\forall \phi$ mesurable positive $\mathbb{R}^n \rightarrow \mathbb{R}$,

$$\int_{\mathbb{R}^n} \phi(x) d\nu(x) = \int_{\mathbb{R}^n} \phi(x) f(x) d\lambda_n(x) .$$

Définition 2.2.12. Une fonction f mesurable sur un espace mesuré $(\Omega, \mathcal{A}, \mu)$ est dite intégrable si f^+ et f^- le sont (voir définition 2.2.2 de l'intégrabilité des fonctions mesurables positives) et dans ce cas, on définit l'intégrale de f (sur Ω par rapport à μ) par

$$\int_{\Omega} f(x) d\mu(x) := \int_{\Omega} f^+(x) d\mu(x) - \int_{\Omega} f^-(x) d\mu(x)$$

et, $\forall A \in \mathcal{A}$, l'intégrale de f sur A par

$$\int_A f(x) d\mu(x) := \int_{\Omega} f(x) \mathbf{1}_A(x) d\mu(x) .$$

Définition 2.2.13. Soit une fonction f mesurable sur un espace mesuré $(\Omega, \mathcal{A}, \mu)$ telle que f^- soit intégrable mais pas f^+ . La fonction f n'est pas intégrable mais par un léger abus de notation (abus de sens), on dira quand même que l'intégrale a un sens et on définit sa valeur par

$$\int_{\Omega} f(x) d\mu(x) := \int_{\Omega} f^+(x) d\mu(x) - \int_{\Omega} f^-(x) d\mu(x) = +\infty$$

Lemme 2.2.14. Soit f une fonction mesurable sur un espace mesuré $(\Omega, \mathcal{A}, \mu)$. Alors f est intégrable si et seulement si $|f|$ est intégrable et dans ce cas.

$$\left| \int_{\Omega} f(x) d\mu(x) \right| \leq \int_{\Omega} |f(x)| d\mu(x)$$

Remarque 2.2.15. Dans le cas de la définition 2.2.13, l'inégalité du lemme précédent est satisfaite (mais dans \mathbb{R}).

Démonstration.

$$\begin{aligned} \left| \int_{\Omega} f(x) d\mu(x) \right| &= \left| \int_{\Omega} f^+(x) d\mu(x) - \int_{\Omega} f^-(x) d\mu(x) \right| \\ &\leq \left| \int_{\Omega} f^+(x) d\mu(x) \right| + \left| \int_{\Omega} f^-(x) d\mu(x) \right| \\ &= \int_{\Omega} f^+(x) d\mu(x) + \int_{\Omega} f^-(x) d\mu(x) \\ &= \int_{\Omega} |f(x)| d\mu(x) . \end{aligned}$$

□

Ce lemme peut aussi être vu comme une conséquence de l'inégalité de Jensen (cf. exercice 4 du chapitre 3 et théorème 2.4.24).

Théorème 2.2.16. *Linéarité.*

Soient f, g deux fonctions positives intégrables sur $(\Omega, \mathcal{A}, \mu)$ et $a \in \mathbb{R}$, alors leur somme est intégrable ainsi que la fonction af et on a :

$$\int_{\Omega} (f(x) + g(x)) d\mu(x) = \int_{\Omega} f(x) d\mu(x) + \int_{\Omega} g(x) d\mu(x)$$

et

$$\int_{\Omega} af(x) d\mu(x) = a \int_{\Omega} f(x) d\mu(x) .$$

Un cas particulier de fonctions est fourni par une généralisation du cas fonction étagée.

Proposition 2.2.17. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et soit $f : \Omega \rightarrow \mathbb{R}$ une fonction telle qu'il existe une partition dénombrable $(A_k)_{k \in \mathbb{N}}$ telle que $f|_{A_k} = \alpha_k$ soit constante et pour tout k , $A_k \in \mathcal{A}$. On peut parler de fonction "dénombrablement étagée" et de "partition mesurable dénombrable".

Alors f est intégrable si et seulement si $\sum_{k=0}^{+\infty} |\alpha_k| \mu(A_k) < +\infty$ et alors

$$\int_{\Omega} f(x) d\mu(x) = \sum_{k=0}^{+\infty} \alpha_k \mu(A_k)$$

Remarque 2.2.18. *Attention, cette intégrale doit être comprise comme une théorie de Lebesgue de l'intégration pour la mesure μ , qui elle, est une mesure quelconque et non nécessairement la mesure de Lebesgue. D'ailleurs Ω est quelconque. C'est non seulement valable pour le résultat précédent, mais pour tout le chapitre sauf mention explicite.*

2.2.3 Lien avec l'intégrale de Riemann

Quand $(\Omega, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda_1)$, l'intégrale $\int_{\Omega} f(x) d\mu(x) = \int_{\mathbb{R}} f(x) \lambda_1(dx)$ que nous venons de définir s'appelle l'intégrale de Lebesgue sur \mathbb{R} . Vu la définition 2.2.12, l'intégrale de Lebesgue sur le segment $[a, b]$ est donnée par

$$\int_{[a,b]} f(x) d\lambda_1(x) := \int_{\mathbb{R}} f(x) \mathbf{1}_{[a,b]}(x) d\lambda_1(x) = \sup_{\phi \in \mathcal{E}(f \mathbf{1}_{[a,b]})} \int_{\Omega} \phi(x) d\mu(x).$$

On va supposer que la fonction f est Riemann-intégrable. Puisque f est bornée sur $[a, b]$, alors $f \mathbf{1}_{[a,b]}$ est intégrale au sens de Lebesgue. On rappelle que l'intégrale de Riemann se calcule avec des fonctions en escalier. On note ici $Esc(f, [a, b])$ l'ensemble des fonctions en escalier h telle que $h \leq f$ sur $[a, b]$ et l'intégrale de Riemann vaut

$$\int_a^b f(x) dx = \sup_{h \in Esc(f, [a, b])} \int_a^b h(x) dx$$

On a donc deux intégrales, l'une au sens de Riemann et une au sens de Lebesgue, et on va les comparer en commençant par un cas particulier.

Cas des fonctions en escalier

Si la fonction f est en escalier sur $[a, b]$, elle est associée à une subdivision ($a = \alpha_0 < \alpha_1 < \dots < \alpha_n = b$) (elle vaut β_i sur $]\alpha_i, \alpha_{i+1}[$ pour $i = 0, \dots, p-1$). Donc, $f \mathbf{1}_{[a,b]}$ est étagée sur \mathbb{R} et si on calcule les deux intégrales en appliquant les définitions, on constate qu'elles valent la même chose :

$$\begin{aligned} \int_a^b f(x) dx &= \sum_{i=0}^{p-1} (\alpha_{i+1} - \alpha_i) \beta_i \\ &= \sum_{i=0}^{p-1} \lambda_1 \mu(] \alpha_i - \alpha_{i+1} [) \beta_i \\ &= \sum_{i=0}^{p-1} \lambda_1 \mu(] \alpha_i - \alpha_{i+1} [) \beta_i + \sum_{i=0}^p \lambda_1 \mu(\{ \alpha_i \}) f(\alpha_i) + 0 \times \lambda_1(\mathbb{R} \setminus [a, b]) \\ &= \int_{\mathbb{R}} f(x) \mathbf{1}_{[a,b]} d\lambda_1(x). \\ &= \int_{[a,b]} f(x) d\lambda_1(x). \end{aligned}$$

Le cas général

On sait que si f est Riemann-intégrable, alors pour tout $\varepsilon = 1/n > 0$, il existe φ_n et ψ_n en escalier telles que $\varphi_n \leq f \leq \psi_n$ et

$$\int_a^b \psi_n(x) dx - \frac{1}{n} \leq \int_a^b \varphi_n(x) dx \leq \int_a^b \psi_n(x) dx$$

En conséquence $\lim_n \int_a^b \varphi_n(x) dx = \int_a^b f(x) dx = \lim_n \int_a^b \psi_n(x) dx$.

D'après ce que l'on sait sur les fonctions en escalier, cela revient à dire que $\lim_n \int_{[a,b]} \varphi_n d\lambda_1 = \lim_n \int_{[a,b]} \psi_n d\lambda_1$. Mais par monotonie de l'intégrale de Lebesgue, on peut en déduire que pour chaque n

$$\int_{[a,b]} \varphi_n d\lambda_1 \leq \int_{[a,b]} f d\lambda_1 \leq \int_{[a,b]} \psi_n d\lambda_1.$$

Par unicité de la limite, on peut conclure que

$$\int_{[a,b]} f(x) d\lambda_1(x) = \int_a^b f(x) dx.$$

On peut conclure que l'intégrale de Lebesgue prolonge la notion d'intégration de Riemann à d'autres fonctions.

Remarque 2.2.19. Soit f continue sur $[a, b]$ (de primitive F), alors f est Lebesgue intégrable sur $[a, b]$ et

$$\int_{[a,b]} f d\lambda_1 = F(b) - F(a)$$

Proposition 2.2.20. Soit $f : [a, b] \rightarrow \mathbb{R}$. La fonction f est intégrable au sens de Riemann sur $[a, b]$ si et seulement si f est bornée et continue en dehors d'un ensemble de mesure nulle. Dans ce cas,

$$\int_{]a,b[} f d\lambda = \int_{[a,b]} f d\lambda = \int_{]a,b]} f d\lambda = \int_{[a,b[} f d\lambda = \int_a^b f(x) dx.$$

Démonstration. Admis □

Remarque 2.2.21. Attention, si on considère la fonction bornée $f : [a, b] \rightarrow \mathbb{R}$ définie par $f(x) = \mathbf{1}_{\mathbb{Q}}$. La fonction n'est continue en aucun point donc la fonction f n'est intégrable au sens de Riemann sur $[a, b]$.

Pour autant $f|_{[a,b] \setminus \mathbb{Q}}$ est continue car constante donc il ne faut pas confondre

- “en dehors d'un ensemble de mesure nulle, f est continue” : $f|_{[a,b] \setminus N} : [a, b] \setminus N \rightarrow \mathbb{R}$ est continue.
- f est continue en dehors d'un ensemble de mesure nulle : $\{x \in [a, b] \mid f \text{ discontinue au point } x\}$ est de mesure nulle.

Remarque 2.2.22. Soit f une fonction localement Riemann intégrable sur \mathbb{R}_+ . Lorsque f ne garde pas un signe constant, on peut² avoir convergence de l'intégrale généralisée $\int_0^{+\infty} f(x) dx$ tout en ayant f non intégrable sur \mathbb{R}_+ .

Proposition 2.2.23. Soit f une fonction localement Riemann intégrable sur \mathbb{R}_+ de signe constant. Alors on a l'équivalence entre la convergence de l'intégrale généralisée $\int_0^{+\infty} f(x) dx$ et l'intégrabilité de f sur \mathbb{R}_+ . De plus, on peut toujours écrire

$$\int_0^{+\infty} f(x) dx = \int_{\mathbb{R}_+} f(x) d\lambda_1(x)$$

Démonstration. La preuve sera donnée ultérieurement (voir le corollaire 3.3.2 page 45). □

Exemple 2.2.24. La fonction $f(t) = e^{-t^2}$ est intégrable sur \mathbb{R} et

$$\int_{\mathbb{R}} e^{-t^2} d\lambda_1(t) = \int_{-\infty}^{+\infty} e^{-t^2} dt.$$

Pour simplifier les écritures, on va se contenter de démontrer que la fonction $h(t) = e^{-t^2}$ est intégrable sur \mathbb{R}_+ et l'argument que nous allons utiliser est basé sur la monotonie

2. par exemple pour $f(x) = \sin(x+1)/x+1$.

$$\int_0^{+\infty} \frac{\sin(x+1)}{x+1} dx = \int_1^{+\infty} \frac{\sin(x)}{x} dx$$

est une intégrale semi-convergente. Dans la mesure où

$$\int_0^{+\infty} \left| \frac{\sin(x+1)}{x+1} \right| dx = \int_1^{+\infty} \left| \frac{\sin(x)}{x} \right| dx = +\infty,$$

la fonction f n'est pas intégrable sur \mathbb{R}_+ au sens de Lebesgue.

de f sur \mathbb{R}_+ et de la comparaison série intégrale. Il suffit de considérer la fonction φ (dénombrablement étagée) définies par

$$\varphi(t) = \begin{cases} f(1) & \text{si } t \in [0, 1[\\ f(2) & \text{si } t \in [1, 2[\\ \dots & \end{cases}$$

On sait que $f \leq \varphi$, or d'après la proposition 2.2.17, on sait que φ est intégrable sur \mathbb{R}_+ et donc f aussi. On conclut avec la proposition 2.2.23.

2.2.4 Ensembles négligeables

Remarque 2.2.25. Une fonction positive intégrable est finie p.p.

Si f est une fonction mesurable positive à valeurs éventuellement infinies, $f : \Omega \rightarrow \overline{\mathbb{R}}_+$.

Il suffit de poser $A = f^{-1}(\{+\infty\})$ qui est mesurable, et de considérer pour $M > 0$, la fonction étagée $\phi_M(x) = M \times \mathbf{1}_A(x)$. Cette fonction vérifie $\phi_M \leq f$, donc par définition $M\mu(A) = \int_{\Omega} \phi_M(x) d\mu(x) \leq \int_{\Omega} f(x) d\mu(x) < +\infty$. Puisque c'est vrai pour tout M , on en déduit que $\mu(A) = 0$.

En contraposant, on peut énoncer soit f une fonction positive, telle que $\mu(f = +\infty) > 0$ alors $\int_{\Omega} f(x) d\mu(x) = +\infty$.

Théorème 2.2.26. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et f fonction mesurable sur cet espace à valeurs dans $\overline{\mathbb{R}}_+$. Alors f est p.p. nulle si et seulement si $\int_{\Omega} |f(x)| d\mu(x) = 0$.

Démonstration. Il suffit d'appliquer l'inégalité de Markov (Théorème 2.2.8). La réciproque est laissée en exercice.

$$\mu(\{x \in \Omega : |f(x)| \geq a\}) \leq \frac{1}{a} \int_{\Omega} |f(x)| d\mu(x).$$

En particulier, la famille dénombrable $A_n = \{x \in \Omega : |f(x)| \geq 1/n\}$ est négligeable donc $\bigcup_n A_n$ aussi. Donc f est nulle p.p. \square

Proposition 2.2.27. Intégrale sur un ensemble négligeable.

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Soit $A \in \mathcal{A}$ négligeable. Soit $f, g : \Omega \rightarrow \overline{\mathbb{R}}$ mesurables. On suppose que $\int_{\Omega} f(x) d\mu(x)$ est définie ainsi que $\int_{\Omega} g(x) d\mu(x)$. On suppose que $f(x) = g(x)$ si $x \notin A$ (donc f et g sont presque partout égales). Alors

$$\int_A f(x) d\mu(x) = 0,$$

$$\int_{\Omega} f(x) d\mu(x) = \int_{\Omega} g(x) d\mu(x).$$

Démonstration. On peut écrire

- Par définition,

$$\int_A f(x) d\mu(x) = \int_{\Omega} f(x) \mathbf{1}_A(x) d\mu(x).$$

Donc par le théorème précédent, $\int_A f(x) d\mu(x) = 0$ et $\int_A g(x) d\mu(x) = 0$.

- Donc $\int_{\Omega} f(x) d\mu(x) = \int_{\Omega \setminus A} f(x) d\mu(x) = \int_{\Omega \setminus A} g(x) d\mu(x) = \int_{\Omega} g(x) d\mu(x)$.

\square

On retient de la proposition précédente que deux fonctions égales presque partout ont la même intégrale.

Exemple 2.2.28. Si on considère l'ensemble $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda_1)$,

Soit $f :]a, b[\rightarrow \mathbb{R}$ une fonction mesurable. On considère un prolongement $g : [a, b] \rightarrow \overline{\mathbb{R}}$ alors g est mesurable et de plus l'intégrale $\int_{]a, b[} f(x) d\lambda_1(x)$ est définie si et seulement si $\int_{[a, b]} g(x) d\lambda_1(x)$ est définie. Les deux intégrales sont égales et en particulier

$$\int_{]a, b[} f(x) d\lambda_1(x) = \int_{[a, b]} g(x) d\lambda_1(x)$$

2.3 Applications aux variables aléatoires discrètes

2.3.1 Définition

Définition 2.3.1. La variable aléatoire réelle définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ est dite discrète si il existe une partie au plus dénombrable \mathcal{D} de \mathbb{R} , telle que $X(\Omega) \subset \mathcal{D}$.

Remarque 2.3.2. Si la variable aléatoire réelle définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ est discrète alors il existe un sous-ensemble I de \mathbb{N} qui vaut soit $\{1, \dots, n\}$ soit \mathbb{N} , tel que \mathcal{D} s'écrive $\{x_i \mid i \in I\}$ pour des x_i tous distincts. On peut poser $p_i = \mathbb{P}_X(\{x_i\}) = \mathbb{P}(X = x_i)$. On a alors $\sum_{i \in I} p_i = 1$ et $\mathbb{P}_X = \sum_{i \in I} p_i \delta_{x_i}$. Dans le cas discret, la loi d'une variable aléatoire discrète est entièrement décrite par la donnée de $(x_k, p_k)_{k \in I}$.

Démonstration. si B est un borélien de \mathbb{R} , alors on peut considérer le sous-ensemble I_B de \mathbb{N} , $I_B = B \cap \mathcal{D}$.

$$\begin{aligned} i) \quad & \mathbb{P}_X(B) = \mathbb{P}(X \in B) \\ & = \mathbb{P}(X \in B \cap \mathcal{D}) + \mathbb{P}(X \in B \cap (\mathbb{R} \setminus \mathcal{D})) \\ & = \mathbb{P}(X \in B \cap \mathcal{D}) + \mathbb{P}(\emptyset) \\ & = \sum_{k \in I_B} \mathbb{P}(X = x_k) \\ & = \sum_{k \in I_B} p_k. \\ ii) \quad & \text{Tandis que puisque } \delta_{x_n}(B) \text{ vaut } \mathbf{1}_{n \in I_B}, \text{ on peut écrire} \\ & (\sum_{n \in \mathbb{N}} p_n \delta_{x_n})(B) = \sum_{n \in \mathbb{N}} p_n \delta_{x_n}(B) \\ & = \sum_{n \in \mathbb{N}} p_n \mathbf{1}_{n \in I_B}. \\ & = \sum_{n \in I_B} p_n. \end{aligned}$$

□

Exemple 2.3.3. Une variable aléatoire X qui suit une loi de Poisson mais aussi e^X qui n'est plus à valeurs dans \mathbb{N} . Un autre exemple, est le temps Y d'apparition du premier "pile" dans le cas d'un tirage pile/face répété, Y est en général un entier mais si on n'a pas de chance, Y peut valoir $+\infty$ même si $\mathbb{P}(Y = +\infty) = 0$.

Pour simplifier le vocabulaire une loi discrète finie rentre dans ce cadre exactement dénombrable, en "rajoutant" des points à \mathcal{D} pour obtenir une partie dénombrable.

La proposition suivante nous montre que pour les variables aléatoires discrètes (par exemple finies), il n'y a pas de précautions à prendre sur la mesurabilité.

Proposition 2.3.4. Soit X une variable aléatoire réelle discrète définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ associée à une partie dénombrable \mathcal{D} de $\overline{\mathbb{R}}$. Soit $\varphi : \mathcal{D} \rightarrow \overline{\mathbb{R}}$ quelconque alors $Y = \varphi(X)$, qui est donc bien définie, est une variable aléatoire discrète.

Démonstration. Admis

□

2.3.2 Cas des variables aléatoires discrètes raisonnables

Remarque 2.3.5. Le concept de variable aléatoire discrète ainsi défini comporte des cas particuliers plus que contre-intuitifs. Dans la pratique, on se contentera souvent³ de manipuler des variables aléatoires discrètes raisonnables pour lesquelles \mathcal{D} est de la forme $\{x_k \mid k \in \mathbb{N}\} = \{x_0, x_1, \dots\}$ où $x_k < x_{k+1}$.

Remarque 2.3.6. Dans le cas raisonnable, on remarque que la fonction de répartition s'écrit $F_X(t) = \sum_{k \in N_t} p_k$ où N_t désigne l'ensemble éventuellement vide $\{n \in \mathbb{N} \mid x_n \leq t\}$. Ces ensembles N_t sont de la forme

- \emptyset et dans ce cas $F_X(t) = 0$;
- $\{0, 1, \dots, k-1, k\}$ et dans ce cas $x_k \leq t < x_{k+1}$ et $F_X(t) = p_0 + \dots + p_k$;
- \mathbb{N} et dans ce cas $F_X(t) = 1$.

3. On pourrait considérer aussi comme raisonnable le cas \mathcal{D} est de la forme $\{x_k \mid k \in \mathbb{Z}\}$ où $x_k < x_{k+1}$ mais cela complexifie l'écriture.

On peut décrire F_X par

$$F_X(t) = \begin{cases} 0 & \text{si } t < x_0 \\ p_0 & \text{si } t \in [x_0, x_1[\\ p_0 + p_1 & \text{si } t \in [x_1, x_2[\\ \dots & \dots \end{cases}$$

2.3.3 Lien entre théorie de la mesure et théorie des probabilités

On va utiliser ici de façon indifférenciée

- soit f pour se projeter dans un contexte théorie de la mesure
- soit X pour se projeter dans un contexte théorie des probabilités.

De toute façon $f = X$. On va d'abord alterner les deux regards puis ne garder que la notation X dans cette partie.

- Dire que X variable aléatoire réelle définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ se traduit en f est une fonction mesurable de $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
- Dire « la variable aléatoire X est discrète finie » se traduit en « la fonction f est étagée ».
- Dire « la variable aléatoire X est discrète » se traduit en « la fonction f est dénombrablement étagée ».

2.3.4 Définition de l'espérance

Si la variable aléatoire est discrète, on va reprendre la démarche vue page 25. Si on considère une v.a.r. discrète X de $(\Omega, \mathcal{A}, \mathbb{P})$ et avec les notations précédentes, on peut supposer que \mathcal{D} s'écrit $\{x_0, x_1, \dots\}$ où les x_n sont tous distincts. On remarque que le complémentaire de \mathcal{D} dans \mathbb{R} est \mathbb{P}_X négligeable⁴. On rappelle que la variable aléatoire X est dénombrablement étagée par rapport à la partition mesurable $(A_i)_{i \in \mathbb{N}}$ où A_i est l'événement $X = x_i$.

Si la fonction X est à valeurs positives, ou si X est intégrable, on peut calculer l'intégrale $\int_{\Omega} X d\mathbb{P}$. En utilisant la proposition 2.2.17, on peut écrire

$$\int_{\Omega} X d\mathbb{P} := \sum_{i=0}^{+\infty} x_i \mathbb{P}(A_i) = \sum_{i=0}^{+\infty} p_i x_i.$$

On a d'une part une interprétation de cette intégrale comme une moyenne pondérée des valeurs prise par X et d'autre part une définition de l'espérance en tant qu'intégrale

$$\mathbb{E}_{\mathbb{P}}(X) := \int_{\Omega} X d\mathbb{P} = \int_{\Omega} X(\omega) d\mathbb{P}(\omega).$$

En utilisant la probabilité \mathbb{P}_X sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on peut énoncer une autre interprétation.

Exemple 2.3.7. Soit X une v.a.r. discrète positive X de $(\Omega, \mathcal{A}, \mathbb{P})$ alors

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} x d\mathbb{P}_X(x)$$

Démonstration. Il faut calculer le terme de droite. On remarque que la fonction $g : \overline{\mathbb{R}}_+ \rightarrow \overline{\mathbb{R}}_+ = x$ définie par $g(x) = x$ est borélienne et que $g\mathbf{1}_{\mathcal{D}}$ est dénombrablement étagée. De plus g est \mathbb{P}_X -p.p. égale à $g\mathbf{1}_{\mathcal{D}}$.

$$\int_{\mathbb{R}} g(t) d\mathbb{P}_X(t) = \int_{\mathcal{D}} g(t) d\mathbb{P}_X(t) = \sum_{i=0}^{+\infty} g(x_i) \mathbb{P}_X(x_i) = \sum_{i=0}^{+\infty} x_i p_i$$

□

En généralisant, on peut même écrire la version discrète du théorème de transfert. Un cas important est le cas où φ est l'identité.

4. Par exemple, si X est une loi de Poisson $\mathbb{P}_X(\mathbb{R} \setminus \mathbb{N}) = \mathbb{P}(X \notin \mathbb{N}) = 0$.

Théorème 2.3.8. [Version discrète du théorème de transfert]

i) Soit une v.a.r. discrète positive X de $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\overline{\mathbb{R}}_+$ déterminée par $\mathbb{P}(X = x_i) = p_i$ pour $i \in I$ et $\varphi : \overline{\mathbb{R}}_+ \rightarrow \overline{\mathbb{R}}_+$. Alors $\varphi \in \mathcal{L}^0(\overline{\mathbb{R}}_+, \mathcal{L}(\overline{\mathbb{R}}_+), \overline{\mathbb{R}}_+)$ et est \mathbb{P}_X -p.s. discrète. En posant $Y = \varphi(X)$, alors

$$\mathbb{E}_{\mathbb{P}}(Y) = \mathbb{E}_{\mathbb{P}}(\varphi(X)) = \int_{\Omega} (\varphi \circ X)(\omega) d\mathbb{P}(\omega) = \sum_{i \in I} p_i \varphi(x_i) = \int_{\mathbb{R}} \varphi(x) d\mathbb{P}_X(x) = \mathbb{E}_{\mathbb{P}_X}(\varphi).$$

ii) Soit une v.a.r. discrète X de $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\overline{\mathbb{R}}$ et $\varphi : \overline{\mathbb{R}} \rightarrow \overline{\mathbb{R}}$. En posant $Y = \varphi(X)$, Y est intégrable sur $(\Omega, \mathcal{A}, \mathbb{P})$ si et seulement si φ est intégrable sur $(\overline{\mathbb{R}}, \mathcal{L}(\overline{\mathbb{R}}), \mathbb{P}_X)$ et

$$\mathbb{E}_{\mathbb{P}}(Y) = \int_{\Omega} (\varphi \circ X)(\omega) d\mathbb{P}(\omega) = \sum_{i \in I} p_i \varphi(x_i) = \int_{\mathbb{R}} \varphi(x) d\mathbb{P}_X(x) = \mathbb{E}_{\mathbb{P}_X}(\varphi).$$

Démonstration. Regardons le cas positif, d'après la définition $\mathbb{E}_{\mathbb{P}}(Y) = \sum_{i=0}^{+\infty} \varphi(x_i) p_i$. On note $\mathcal{D} = X(\Omega)$. On va considérer la fonction $\psi : \overline{\mathbb{R}}_+ \rightarrow \overline{\mathbb{R}}_+$ définie par $\psi = \varphi \mathbf{1}_{\overline{\mathbb{R}} \setminus \mathcal{D}}$. On remarque tout d'abord que $\psi = \varphi$ \mathbb{P}_X -p.s. car \mathbb{P}_X "charge uniquement les points de \mathcal{D} ". De plus, si \mathcal{D} est exactement dénombrable, on peut présenter ψ par

$$\psi(x) = \begin{cases} 0 & \text{si } x \notin \mathcal{D} \\ \varphi(x_0) & \text{si } x = x_0 \\ \varphi(x_1) & \text{si } x = x_1 \\ \dots & \dots \end{cases}$$

On peut donc appliquer le lemme de recollement et comprendre que ψ est dénombrablement étagée. Donc φ est aussi mesurable. **C'est ici qu'il est fondamental de travailler avec une tribu et une mesure complétée, la tribu de Lebesgue et la mesure de Lebesgue.** Donc les intégrales ont même valeurs $\int_{\mathbb{R}} \varphi(x) d\mathbb{P}_X(x) = \int_{\mathbb{R}} \psi(x) d\mathbb{P}_X(x)$ et puisque la fonction ψ est dénombrablement étagée, il est possible de calculer explicitement son intégrale

$$\int_{\mathbb{R}} \psi(x) d\mathbb{P}_X(x) = \sum_{i \in \mathbb{N}} \psi(x_i) \mathbb{P}_X(\{x_i\}) + 0 \times \mathbb{P}_X(\mathbb{R}_b \setminus \mathcal{D}) = \sum_{i \in \mathbb{N}} \psi(x_i) p_i + 0 = \sum_{i \in \mathbb{N}} \varphi(x_i) p_i.$$

□

2.4 Cas d'une variable aléatoire réelle générale

2.4.1 Généralités

Il existe des variables aléatoires qui ne sont ni discrètes ni à densité. Par exemple, il suffit de considérer la variable aléatoire $Y = X^+ = (1/2)(X + |X|)$, la partie positive de X , lorsque $X \sim \mathcal{N}(0, 1)$.

Définition 2.4.1. Soit X une variable aléatoire.

- On dit que X est une variable aléatoire diffuse (ou sans atome) si F_X est continue.
- On dit que X est une variable aléatoire continue si \mathbb{P}_X possède une densité par rapport à la mesure de Lebesgue sur \mathbb{R} .

Remarque 2.4.2. Attention au vocabulaire qui est assez dangereux s'il est sorti de son contexte. Considérons \mathbb{R} muni de la loi de probabilité notée \mathbb{P} correspondant à la loi normale centrée réduite ainsi que la fonction $Y : \mathbb{R} \rightarrow \mathbb{R}$

$$Y(t) = \begin{cases} t + 1 & \text{si } t \geq 0; \\ t & \text{sinon.} \end{cases}$$

La fonction Y n'est pas une fonction continue de \mathbb{R} dans \mathbb{R} , mais elle constitue une variable aléatoire continue.

Définition 2.4.3. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \overline{\mathbb{R}}_+$ une variable aléatoire positive. L'intégrale de X par rapport à la mesure \mathbb{P} est appelée espérance de X . Une variable aléatoire est dite intégrable si son espérance est finie.

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int_{\Omega} X d\mathbb{P}$$

Définition 2.4.4. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \overline{\mathbb{R}}$ une variable aléatoire réelle.

Une variable aléatoire est dite intégrable si l'intégrale de $|X|$ par rapport à la mesure \mathbb{P} est finie. L'intégrale de X par rapport à la mesure \mathbb{P} est appelée espérance de X . On note alors

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int_{\Omega} X d\mathbb{P}$$

Lorsqu'elle existe, la quantité $\mathbb{E}(X)$ s'appelle aussi le moment d'ordre 1, et on parle variable aléatoire centrée si $\mathbb{E}(X) = 0$.

Remarque 2.4.5. L'espérance est une intégrale. Réécrivons les propriétés de l'intégrale avec le symbole \mathbb{E} .

- (i) Linéarité : si X et Y sont deux v.a.r. et $a, b \in \mathbb{R}$, $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ (cf. th. 2.2.16).
- (ii) Croissance : si X et Y sont deux v.a.r. telles que $X \leq Y$ (c'est à dire $\forall \omega \in \Omega, X(\omega) \leq Y(\omega)$) alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$ (cf. prop. 2.2.5).
- (iii) Variable aléatoire constante : si X v.a.r. et $a \in \mathbb{R}$ tels que $X(\omega) = a, \forall \omega$, alors $\mathbb{E}(X) = a$ (cf. déf. 2.1.8).
- (iv) Si X et Y v.a.r. telle que $X = Y$ p.p. alors $\mathbb{E}(X) = \mathbb{E}(Y)$ (cf. prop. 2.2.27).
- (v) Si X variable aléatoire à valeurs dans $[0, +\infty]$ telle que $\mathbb{E}(X) < \infty$ alors X est finie p.s. (cf. rem. 2.2.25).

Remarque 2.4.6. Le théorème de transfert sera énoncé dans le paragraphe sur les variables à densité afin de le rendre plus concret. En toute rigueur, il devrait être énoncé ici.

Définition 2.4.7. Si X est une v.a.r. telle que X^2 est intégrable alors la variance de X est la quantité

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 .$$

Lemme 2.4.8. $\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$

Démonstration. Nous allons utiliser les propriétés (i) et (iii) de la remarque 2.4.5.

$$\begin{aligned} \mathbb{E}((X - \mathbb{E}(X))^2) &= \mathbb{E}(X^2 + \mathbb{E}(X)^2 - 2X\mathbb{E}(X)) \\ &= \mathbb{E}(X^2) + \mathbb{E}(\mathbb{E}(X)^2) - 2\mathbb{E}(X\mathbb{E}(X)) \\ &= \mathbb{E}(X^2) + \mathbb{E}(X)^2 - 2\mathbb{E}(X)^2 \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \end{aligned}$$

□

2.4.2 Vecteur aléatoire à densité

2.4.3 Densités

Proposition et définition 2.4.9. Soit $f : \Omega \rightarrow \overline{\mathbb{R}}_+$ et $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. La formule 2.4.1,

$$\text{pour tout } A \in \mathcal{A}, \nu(A) = \int_A f d\mu . \quad (2.4.1)$$

définit une mesure sur (Ω, \mathcal{A}) . On dit alors que ν possède une densité f par rapport à μ .

Remarque 2.4.10. Si f est une densité de ν par rapport à μ , et si $f = g$ p.p. alors, g est aussi une densité de ν par rapport à μ . De plus, si f et g sont des densités de ν par rapport à μ , alors $f = g$ p.p.

Par un abus de langage, on dit néanmoins (même s'il n'y a pas d'unicité) que ν a pour densité f par rapport à μ .

Définition 2.4.11. Soit X une v.a. définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans \mathbb{R}^d et μ une mesure sur \mathbb{R}^d . On dit que X possède une densité $f_X : \mathbb{R}^d \rightarrow \mathbb{R}_+$ par rapport à μ (si on ne précise pas μ , on utilise λ_d la mesure de Lebesgue multidimensionnelle) si \mathbb{P}_X possède f_X comme densité par rapport à ν .

Cela signifie donc que pour tout borélien B de \mathbb{R}^d ,

$$\mathbb{P}(X \in B) = \int_{\mathbb{R}^d} f_X(x) \mathbf{1}_B(x) d\mu(x).$$

La densité de X est la densité de \mathbb{P}_X .

Dans le cas d'une variable aléatoire réelle, si on utilise la mesure $\mu = \lambda$ et si f est (localement-) Riemann-intégrable, on peut écrire

$$F_X(t) = \mathbb{P}(X \leq t) = \int_{]-\infty, t]} f_X(x) d\lambda(x) = \int_{-\infty}^t f_X(x) dx.$$

Proposition et notation 2.4.12. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Si ν a pour densité $f : \Omega \rightarrow \mathbb{R}_+$ alors $\forall \phi$ mesurable positive $\omega \rightarrow \mathbb{R}_+$,

$$\int_{\Omega} \phi d\nu = \int_{\Omega} \phi(\omega) d\nu(\omega) = \int_{\Omega} \phi(\omega) f(\omega) d\mu(\omega) = \int_{\Omega} \phi f d\mu$$

On note alors $d\nu = f d\mu$ ou encore $f = \frac{d\nu}{d\mu}$ (dérivée au sens de Radon-Nikodym).

Remarque 2.4.13. La densité d'une variable aléatoire détermine complètement sa loi.

Par définition, une densité f_X (d'une v.a. X à valeurs dans \mathbb{R}^d) est toujours positive et vérifie $\int_{\mathbb{R}^d} f_X(x) d\lambda_d(x) = 1$.

2.4.4 Le théorème de Lebesgue-Radon-Nikodym

Définition et notation 2.4.14. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Une mesure ν est dite absolument continue par rapport à la mesure μ si, pour tout $A \in \mathcal{A}$:

$$\mu(A) = 0 \Rightarrow \nu(A) = 0,$$

ce qui sera noté : $\nu \ll \mu$.

Exemple 2.4.15. Soit X une v.a.r. suivant une loi discrète alors \mathbb{P}_X n'est pas absolument continue par rapport à la mesure de Lebesgue tandis que si X une v.a.r. suivant une loi $\mathcal{N}(0, 1)$ alors \mathbb{P}_X est absolument continue par rapport à la mesure de Lebesgue.

Remarque 2.4.16. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Si ν a pour densité $f : \Omega \rightarrow \mathbb{R}_+$ par rapport à μ alors $\nu \ll \mu$.

En fait, la condition est également suffisante. On va donner une version simplifiée du théorème de Lebesgue-Radon-Nikodym.

Théorème 2.4.17 (Lebesgue, Radon, Nikodym). Soit un espace mesuré $(\Omega, \mathcal{A}, \mu)$ supposée σ -finie. Si ν est une mesure σ -finie et absolument continue par rapport à μ , alors il existe une fonction mesurable $f : \Omega \rightarrow \overline{\mathbb{R}}_+$, telle que pour tout A dans \mathcal{A} :

$$\nu(A) = \int_A f d\mu.$$

Si X est une v.a.r. définie sur (Ω, \mathcal{A}) telle que \mathbb{P}_X soit absolument continue par rapport à λ (la mesure de Lebesgue), alors elle possède une densité $f_X : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}_+$.

Proposition 2.4.18. *Si X est une v.a. de densité f , sa fonction de répartition F_X vérifie :*

1. $\forall t \in \mathbb{R}, F_X(t) = \int_{]-\infty, t]} f(t) d\lambda_1(t)$.
2. F_X est continue sur \mathbb{R} .
3. Si f est continue au point x_0 , alors F_X est dérivable en x_0 de dérivée $F'_X(x_0) = f(x_0)$.

En effet F_X est dérivable en x_0 de dérivée $F'_X(x_0) = f(x_0)$ si et seulement si

$$\lim_{h \rightarrow 0, h \neq 0} \varphi(h) := \lim_{h \rightarrow 0, h \neq 0} \varphi(h) \frac{F(x_0+h) - F(x_0)}{h} - f(x_0) \rightarrow 0.$$

On peut remarquer si $h > 0$

$$\varphi(h) = \begin{cases} \frac{\int_{](x_0, x_0+h]} (f(u) - f(x_0)) d\lambda(u)}{h} & \text{si } h > 0 \\ \frac{\int_{](x_0+h, x_0]} (-f(u) + f(x_0)) d\lambda(u)}{h} & \text{si } h < 0. \end{cases}$$

Mais par continuité, pour tout $\varepsilon > 0$ il existe $\eta_\varepsilon > 0$ tel que pour tout u dans $]x_0 - \eta_\varepsilon, x_0 + \eta_\varepsilon[$ $|f(u) - f(x_0)| \leq \varepsilon$.

Remarque 2.4.19. *Soit X une v.a.r. et \mathcal{D} un sous-ensemble fini de \mathbb{R} . Si F_X est continue sur \mathbb{R} et de classe \mathcal{C}^1 sur $\mathbb{R} \setminus \mathcal{D}$, alors X a pour densité F'_X (qui existe sauf sur l'ensemble \mathcal{D}).*

2.4.5 Le théorème de transfert

L'énoncé ici dépasse le cadre des variables à densités. Pour des raisons pédagogiques, on l'énonce ici afin de pouvoir l'illustrer par la formulation iii).

Théorème 2.4.20 (Théorème de transfert). *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité.*

i) *Soit une v.a.r. X de $(\Omega, \mathcal{A}, \mathbb{P})$ et $\varphi : \mathbb{R} \rightarrow \overline{\mathbb{R}}_+$ mesurable. En posant $Y = \varphi(X)$,*

$$\mathbb{E}_{\mathbb{P}}(Y) = \mathbb{E}_{\mathbb{P}}(\varphi(X)) = \int_{\Omega} (\varphi \circ X)(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} \varphi(x) d\mathbb{P}_X(x) = \mathbb{E}_{\mathbb{P}_X}(\varphi).$$

ii) *Soit une v.a.r. X de $(\Omega, \mathcal{A}, \mathbb{P})$ et $\varphi : \mathbb{R} \rightarrow \mathbb{R}$. En posant $Y = \varphi(X)$, Y est intégrable sur $(\Omega, \mathcal{A}, \mathbb{P})$ si et seulement si φ est intégrable sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$ et*

$$\mathbb{E}_{\mathbb{P}}(Y) = \int_{\Omega} (\varphi \circ X)(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} \varphi(x) d\mathbb{P}_X(x) = \mathbb{E}_{\mathbb{P}_X}(\varphi).$$

iii) *Soit une v.a.r. X de $(\Omega, \mathcal{A}, \mathbb{P})$ de densité f et $\varphi : \mathbb{R} \rightarrow \overline{\mathbb{R}}_+$ mesurable positive. En posant $Y = \varphi(X)$,*

$$\mathbb{E}_{\mathbb{P}}(Y) = \mathbb{E}_{\mathbb{P}}(\varphi(X)) = \int_{\mathbb{R}} \varphi(x) f(x) d\lambda_1(x).$$

iv) *Soit une v.a.r. X de $(\Omega, \mathcal{A}, \mathbb{P})$ de densité f et $\varphi : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ mesurable. En posant $Y = \varphi(X)$, Y est intégrable sur $(\Omega, \mathcal{A}, \mathbb{P})$ si et seulement si $\varphi \times f$ est intégrable sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda_1)$ et*

$$\mathbb{E}(Y) = \int_{\mathbb{R}} \varphi(x) f(x) d\lambda_1(x).$$

La définition d'une variable aléatoire à densité nécessite de vérifier une propriété pour chaque fonction indicatrice mais en fait on peut se contenter de vérifier sur les fonctions continues positives bornées.

Proposition 2.4.21. *La loi d'une variable aléatoire X à valeurs dans \mathbb{R}^d est uniquement déterminée par le calcul de $\mathbb{E}(\phi(X))$ pour toute fonction $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ continue positive bornée. Autrement dit :*

Soit X variable aléatoire à valeurs dans \mathbb{R}^d . S'il existe $g : \mathbb{R}^d \rightarrow \mathbb{R}$ telle que $\forall \phi : \mathbb{R}^d \rightarrow \mathbb{R}$ continue positive bornée,

$$\mathbb{E}(\phi(X)) = \int_{\mathbb{R}^d} \phi(x)g(x)dx ,$$

alors g est la densité de X .

Notation 2.4.22. *On note $\mathcal{C}_b^+(\mathbb{R}^d)$ l'ensemble des fonctions continues positives bornées $\mathbb{R}^d \rightarrow \mathbb{R}_+$.*

Exemple 2.4.23. *Soit X v.a.r. de densité $x \mapsto \frac{e^{-x^2/2}}{\sqrt{2\pi}}$. Soient $(a, b) \in \mathbb{R}^* \times \mathbb{R}$. Calculons la loi de $aX + b$. Soit $f : \mathbb{R} \rightarrow \mathbb{R}_+$ continue et bornée (on dit que f est une « fonction test »). Par le théorème de transfert, nous avons*

$$\begin{aligned} \mathbb{E}(f(aX + b)) &= \int_{-\infty}^{+\infty} f(ax + b) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\ (\text{changement de variable } y = ax + b) &= \int_{-\infty}^{+\infty} f(y) \frac{e^{-\left(\frac{y-b}{a}\right)^2 \frac{1}{2}}}{\sqrt{2\pi} \times a} dy \end{aligned}$$

Donc, par la proposition 2.4.21, la variable $aX + b$ a une loi de densité $y \mapsto \frac{\exp\left(-\frac{1}{2}\left(\frac{y-b}{a}\right)^2\right)}{\sqrt{2\pi} \times a}$.

2.4.6 Inégalités

Théorème 2.4.24. *Inégalité de Jensen*

Soit $\phi : \mathbb{R} \rightarrow \mathbb{R}$ mesurable convexe. Soit X v.a.r. (définie sur $\Omega, \mathcal{A}, \mathbb{P}$) intégrable telle que $\phi(X)$ est intégrable. Alors

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X)).$$

Exemple 2.4.25. *On retrouve dans un cadre plus général le résultat correspondant pour l'intégrale de Riemann. La fonction $\phi : \mathbb{R} \rightarrow \mathbb{R}$ définie par $\phi(x) = x^2$ est convexe. Donc par le résultat précédent, pour toute fonction $f : [0, 1] \rightarrow \mathbb{R}$ intégrable pensée comme une variable aléatoire X sur l'espace de probabilité $([0, 1], \mathcal{B}([0, 1]), \lambda)$, en utilisant le théorème de transfert,*

$$\phi(\mathbb{E}(X)) = \left(\int_{[0,1]} f(x) d\lambda(x) \right)^2 \leq \int_{[0,1]} f(x)^2 d\lambda(x) = \mathbb{E}(\phi(X)).$$

Démonstration. Regardons la preuve dans le cas régulier ϕ de classe \mathcal{C}^2 (le cas général pourra être abordé ultérieurement dans le cadre du cours d'optimisation).

- Puisque $\phi'' \geq 0$, ϕ' est une fonction croissante (au sens large). On se fixe x_0 réel quelconque alors pour tout $x \in \mathbb{R}$,
si $x \geq x_0$,

$$\phi(x) - \phi(x_0) = \int_{x_0}^x \phi'(t) dt \geq \int_{x_0}^x \phi'(x_0) dt = \phi'(x_0)(x - x_0)$$

si $x \leq x_0$,

$$\phi(x) - \phi(x_0) = \int_{x_0}^x \phi'(t) dt = \int_x^{x_0} -\phi'(t) dt \geq \int_x^{x_0} -\phi'(x_0) dt = \phi'(x_0)(x - x_0)$$

dans les deux cas, l'inégalité est satisfaite pour tout $x \in \mathbb{R}$. Géométriquement, le graphe de ϕ est au dessus de ses tangentes.

- On prend $x_0 = \mathbb{E}(X) = \int_{\Omega} X(t) d\mu(t)$ et $x = X(\omega)$ pour un ω quelconque dans l'inégalité précédente et on a :

$$\phi(X(\omega)) \geq \phi(x_0) + \phi'(x_0)(X(\omega) - x_0) .$$

On intègre ensuite sur Ω en utilisant la mesure de probabilité \mathbb{P} :

$$\begin{aligned} \int_{\Omega} \phi(f(\omega)) d\mathbb{P}(\omega) &\geq \int_{\Omega} \phi(x_0) d\mathbb{P}(\omega) + \int_{\Omega} \phi'(x_0)(X(\omega) - x_0) d\mathbb{P}(\omega) \\ &\geq \phi(x_0)\mathbb{P}(\Omega) + \phi'(x_0) \int_{\Omega} (X(\omega) - x_0) d\mathbb{P}(\omega) \\ &\geq \phi(x_0) + \phi'(x_0) \left(\int_{\Omega} (X(\omega) d\mathbb{P}(\omega) - x_0) \right) \\ &\geq \phi(x_0) + \phi'(x_0)(x_0 - x_0) = \phi(x_0) \end{aligned}$$

□

On va rappeler l'inégalité de Markov (déjà présentée dans le cas des fonctions) qui est l'outil principale de l'inégalité de Bienaymé-Tchebichev.

Théorème 2.4.26. *Inégalité de Markov. Soit X v.a.r. positive, intégrable. Soit $\lambda > 0$. Alors*

$$\mathbb{P}(X \geq \lambda) \leq \frac{1}{\lambda} \mathbb{E}(X) .$$

Théorème 2.4.27. *Inégalité de Bienaymé-Tchebichev*
Si X v.a.r. avec X^2 intégrable et si $\lambda > 0$ alors

$$\mathbb{P}(|X| \geq \lambda) \leq \frac{\mathbb{E}(X^2)}{\lambda^2} .$$

Démonstration.

$$\begin{aligned} \mathbb{P}(|X| \geq \lambda) &= \mathbb{P}(X^2 \geq \lambda^2) \\ (\text{par inégalité de Markov}) &\leq \frac{\mathbb{E}(X^2)}{\lambda^2} . \end{aligned}$$

□

Corollaire 2.4.28. *Soit X v.a.r. telle que X^2 est intégrable. Alors*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{\text{Var}(X)}{\lambda^2} .$$

Démonstration du corollaire 2.4.28. Si on pose $Y = X - \mathbb{E}(X)$ (variable centrée), on a alors $\text{Var}(X) = \text{Var}(Y) = \mathbb{E}(Y^2)$. Donc,

$$\begin{aligned} \mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) &= \mathbb{P}(|Y| \geq \lambda) \\ (\text{par inégalité de Bienaymé-Tchebichev}) &\leq \frac{\mathbb{E}(Y^2)}{\lambda^2} . \end{aligned}$$

□

Chapitre 3

Théorèmes limites et applications

On se donne $(\Omega, \mathcal{A}, \mu)$ un espace mesuré complet. La complétude est essentielle car elle nous garantit la mesurabilité “au sens de la classe égale presque partout”. On supposera à partir de maintenant, pour des raisons techniques, que Ω est σ -fini. C’est bien entendu le cas pour $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda_1)$ ou si si l’espace est un espace de probabilité.

3.1 Convergence presque partout.

Définition 3.1.1. Soit $(f_n)_{n \geq 0}$ une suite de fonctions $\Omega \rightarrow \overline{\mathbb{R}}$. On dit que (f_n) converge partout vers f s’il existe N négligeable tel que $[x \notin N] \Rightarrow [f_n(x) \xrightarrow[n \rightarrow +\infty]{} f(x)]$. On note alors

$$f_n \xrightarrow[n \rightarrow +\infty]{p.p.} f.$$

Si $\mu = \mathbb{P}$ est une mesure de probabilité, en utilisant une terminologie probabiliste, on parle plutôt de convergence presque sûre et on note $f_n \xrightarrow[n \rightarrow +\infty]{p.s.} f$.

Définition 3.1.2. Soit $(f_n)_{n \geq 0}$ une suite de fonctions $\Omega \rightarrow \overline{\mathbb{R}}$. On rappelle que (f_n) convergence simplement vers f si $\forall x, f_n(x) \xrightarrow[n \rightarrow +\infty]{} f(x)$.

Remarque 3.1.3. La convergence presque partout de la suite (f_n) vers f équivaut à la convergence simple de $(f_n \mathbf{1}_{N^c})$ vers $f \mathbf{1}_{N^c}$ pour un ensemble N négligeable. En particulier (pour $N = \emptyset$), la convergence simple de la suite (f_n) vers f implique la convergence presque partout.

Définition 3.1.4. On dit que A est un ensemble de mesure pleine si son complémentaire est de mesure nulle. En particulier, si $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace de probabilité, A est un ensemble de mesure pleine équivaut à $\mathbb{P}(A) = 1$.

Le vocabulaire ci-dessus peut receler une petite difficulté, sur $(\Omega, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda)$, $\mathbb{R} \setminus \mathbb{Q}$ est un ensemble de mesure pleine bien pas \mathbb{R}_+ alors que $\lambda(\mathbb{R}_+)$ soit égal à $\lambda(\mathbb{R})$. Quand à \mathbb{Q} , il est dense dans \mathbb{R} (au sens du cours d’analyse) mais pas de mesure pleine. Pour finir, l’ensemble $\{0\}$ est de mesure pleine dans \mathbb{R} pour la mesure de Dirac en 0.

Proposition 3.1.5. Soit deux suites $(f_n)_n$ et g_n de fonctions $\Omega \rightarrow \overline{\mathbb{R}}$ telle que pour chaque n $f_n = g_n$, μ -p.p. et soit f une fonction alors on a l’équivalence

1. on a l’équivalence
 - i) pour tout n , f_n est mesurable ;
 - ii) pour tout n , g_n est mesurable.
2. On a de plus l’équivalence $f_n \xrightarrow[n \rightarrow +\infty]{p.p.} f \Leftrightarrow g_n \xrightarrow[n \rightarrow +\infty]{p.p.} f$.
3. Si la condition i) est satisfaite et si $f_n \xrightarrow[n \rightarrow +\infty]{p.p.} f$ alors f est mesurable.

Bien entendu, cet énoncé peut se transposer pour des variables aléatoires.

Démonstration. 1. Vu l'hypothèse de complétude, il suffit d'utiliser la proposition 1.5.23.

2. On pose les ensembles négligeables $N_n = \{f_n \neq g_n\}$ et $N = \{\omega \mid f_n \not\rightarrow f(\omega)\}$. Par dénombrabilité, l'ensemble $\tilde{N} := (\cup_{n \in \mathbb{N}} N_n) \cup N$ est aussi négligeable. Puisque sur N_n^c $f_n = g_n$, c'est d'autant plus vrai sur \tilde{N}^c . De plus, D'après la remarque 3.1.3, $(g_n \mathbf{1}_N) = (f_n \mathbf{1}_N)$ converge simplement vers $f \mathbf{1}_N$, donc cela signifie (à nouveau d'après la remarque 3.1.3) que g_n est mesurable.

3. On pose $h_n = f_n \mathbf{1}_{\tilde{N}}$. On a $h_n = f_n$ μ -p.p. donc h_n est mesurable (proposition 1.5.23). Or la suite (h_n) converge simplement vers $f \mathbf{1}_{\tilde{N}}$ donc la fonction $f \mathbf{1}_{\tilde{N}}$ est mesurable. Mais $f = f \mathbf{1}_{\tilde{N}}$ μ -p.p. donc f est également mesurable.

Pour démontrer la mesurabilité de f , il suffit de remarquer que $h_n := f_n \mathbf{1}_{A^c}$ et f_n sont égales p.s. donc h_n est mesurable car l'espace est complet (proposition 1.5.23). La limite $f \mathbf{1}_{A^c}$ est donc mesurable (théorème 1.5.70) mais $f \mathbf{1}_{A^c} = f$ p.p. qui est donc mesurable. □

Remarque 3.1.6. Soit (f_n) une suite presque partout croissante (c'est à dire que μ -p.p. x , $\forall n, f_n(x) \leq f_{n+1}(x)$) de fonctions mesurables à valeurs dans $\overline{\mathbb{R}}$ alors la suite $(f_n)_n$ converge presque partout vers une fonction f .

En effet, par le même type d'arguments, il existe un sous-ensemble $A \subset \Omega$ de mesure pleine tel que $g_n := f_n \mathbf{1}_A$ soit croissant sur Ω . Mais pour chaque ω , la suite $g_n(\omega)$ converge dans $\overline{\mathbb{R}}$. On appelle $f(\omega)$ sa limite. Puisque $g_n \rightarrow f$ sur Ω au sens de la convergence simple, $f_n \xrightarrow[n \rightarrow +\infty]{\text{p.p.}} f$ d'après la proposition 3.1.5.

3.2 Théorèmes de convergence pour les intégrales.

3.2.1 Théorème de convergence monotone.

Théorème 3.2.1. *Théorème de convergence monotone, Beppo Levi (1906)*

Soit (f_n) une suite croissante¹ ($f_n \leq f_{n+1}$) de fonctions mesurables positives $\Omega \rightarrow [0, +\infty[$ convergeant vers une fonction f . Alors

$$\lim_{n \rightarrow +\infty} \int_{\Omega} f_n(x) d\mu(x) = \int_{\Omega} f(x) d\mu(x) .$$

En particulier, si $(X_n)_n$ est une suite croissante ($X_{n+1} \geq X_n$) de variables aléatoires positives, qui converge vers X alors $\mathbb{E}(X_n) \rightarrow \mathbb{E}(X)$.

Démonstration. On commence par remarquer que la limite des intégrales ($\int_{\Omega} f_n d\mu$) existe (théorème de la limite monotone). De plus $\lim_n \int_{\Omega} f_n d\mu \leq \int_{\Omega} f d\mu$ car $f_n \leq f$. Il s'agit de vérifier l'autre inégalité. On va considérer des fonctions f de plus en plus sophistiquées.

Premier cas Si la suite (f_n) converge simplement vers $\beta \mathbf{1}_B := f$ avec B mesurable et β réel.

Deuxième cas Si f est étagée.

Troisième cas Cas général. □

Exemple 3.2.2. Soit l'espace mesuré $(\mathbb{N}^*, \mathcal{P}(\mathbb{N}^*), \mu_C)$ (mesure de comptage). On veut montrer que

$$\sum_{k=1}^{+\infty} \frac{kn}{(kn+1)k^2} \xrightarrow{n \rightarrow +\infty} \sum_{k=1}^{+\infty} \frac{1}{k^2} \quad (3.2.1)$$

1. On pourrait se contenter de l'hypothèse (f_n) presque sûrement croissante, c'est à dire que μ -p.p. x , $\forall n, f_n(x) \leq f_{n+1}(x)$ et d'une convergence presque partout.

3.2.2 Lemme de Fatou

Comme application du théorème de convergence monotone, on va énoncer le lemme de Fatou qui nous donne une inégalité lorsque l'on supprime l'hypothèse de monotonie.

Théorème 3.2.3. *Lemme de Fatou*

Soit $(f_n)_{n \geq 0}$ une suite de fonctions mesurables positives. On note $f = \liminf_{n \rightarrow +\infty} f_n$. Alors f est mesurable positive et

$$\int_{\Omega} f \, d\mu := \int_{\Omega} \liminf_{n \rightarrow +\infty} f_n \, d\mu \leq \liminf_{n \rightarrow +\infty} \int_{\Omega} f_n \, d\mu$$

Démonstration. On pose $i_n = \inf_{k \geq n} f_k(x)$. La suite (i_n) est une suite décroissante de fonctions mesurables qui converge simplement vers f dans $[0, +\infty[$.

On considère $\varepsilon > 0$, puis pour chaque n ,

$$A_n = \{x : i_n(x) \geq f(x) - \varepsilon\} = \{x : \forall p \geq n, f_p(x) \geq (f(x) - \varepsilon)_+\}.$$

Pour tout x , $\exists N \in \mathbb{N}$ tel que $n \geq N \Rightarrow i_n(x) \geq f(x) - \varepsilon$. Nous avons donc $\bigcup_{n \geq 1} A_n = \Omega$. On remarque que pour tout n , $A_n \subset A_{n+1}$. Et donc pour tout x ,

$$(f(x) - \varepsilon)_+ \mathbf{1}_{A_n}(x) \nearrow_{n \rightarrow +\infty} (f(x) - \varepsilon)_+.$$

Donc, par théorème de convergence monotone,

$$\int_{\Omega} (f(x) - \varepsilon)_+ \mathbf{1}_{A_n}(x) \, d\mu(x) \xrightarrow{n \rightarrow +\infty} \int_{\Omega} (f(x) - \varepsilon)_+ \, d\mu(x).$$

Pour tout n , nous avons

$$\int_{\Omega} f_n(x) \, d\mu(x) \geq \int_{\Omega} f_n(x) \mathbf{1}_{A_n}(x) \, d\mu(x) \geq \int_{\Omega} \left(f(x) - \frac{1}{m}\right)_+ \mathbf{1}_{A_n}(x) \, d\mu(x)$$

et donc

$$\liminf_{n \rightarrow +\infty} \int_{\Omega} f_n(x) \, d\mu(x) \geq \int_{\Omega} (f(x) - \varepsilon)_+ \, d\mu(x).$$

On considère une suite $\varepsilon_m = 1/m$ pour $m \geq 1$, puis $g_m = (f - \frac{1}{m})_+$. Nous avons pour tout x , $g_m(x) \nearrow f(x)$. Donc, par théorème de convergence monotone, $\int_{\Omega} (f(x) - \frac{1}{m})_+ \, d\mu(x) \xrightarrow{m \rightarrow \infty} \int_{\Omega} f(x) \, d\mu(x)$. Et donc

$$\liminf_{n \rightarrow +\infty} \int_{\Omega} f_n(x) \, d\mu(x) \geq \int_{\Omega} f(x) \, d\mu(x).$$

□

Remarque 3.2.4. *Attention, on a seulement une inégalité qui peut être stricte. Par exemple, on peut considérer $f_n = \mathbf{1}_{[n, +\infty[}$, on a $f_n \rightarrow 0 = f$ et $\int_{\mathbb{R}} f \, d\lambda < \liminf_{n \rightarrow \infty} \int_{\mathbb{R}} f_n \, d\lambda$.*

Remarque 3.2.5. *Parfois, on peut éviter par translation l'hypothèse de positivité, par exemple si on suppose que $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace de probabilité sur lequel les variables aléatoires (X_n) vérifient, il existe une constante M telle que $|X_n| \leq M$. On ne peut appliquer le lemme de Fatou aux variables aléatoires (X_n) car elles ne sont pas positives mais on peut le faire pour la suite $(Y_n) = (M + X_n)$ et pour la suite $(Z_n) = (M - X_n)$.*

$$\begin{cases} \liminf_{n \rightarrow +\infty} \int_{\Omega} Y_n(x) \, d\mathbb{P}(x) \geq \int_{\Omega} \liminf_{n \rightarrow +\infty} Y_n(x) \, d\mathbb{P}(x). \\ \liminf_{n \rightarrow +\infty} \int_{\Omega} Z_n(x) \, d\mathbb{P}(x) \geq \int_{\Omega} \liminf_{n \rightarrow +\infty} Z_n(x) \, d\mathbb{P}(x). \end{cases}$$

En utilisant le vocabulaire de l'espérance,

$$\begin{cases} \liminf_{n \rightarrow +\infty} \mathbb{E}(Y_n) \geq \mathbb{E}(\liminf_{n \rightarrow +\infty} Y_n). \\ \liminf_{n \rightarrow +\infty} \mathbb{E}(Z_n) \geq \mathbb{E}(\liminf_{n \rightarrow +\infty} Z_n). \end{cases}$$

Si on utilise les propriétés usuelles de l'espérance, on obtient que $\mathbb{E}(Y_n) = M + \mathbb{E}(X_n)$ et que $\mathbb{E}(Z_n) = M - \mathbb{E}(X_n)$. Donc

$$\begin{cases} M + \liminf_{n \rightarrow +\infty} \mathbb{E}(X_n) = \liminf_{n \rightarrow +\infty} \mathbb{E}(M + X_n) \geq \mathbb{E}(\liminf_{n \rightarrow +\infty} (M + X_n)) = M + \mathbb{E}(\liminf_{n \rightarrow +\infty} X_n). \\ M - \limsup_{n \rightarrow +\infty} \mathbb{E}(X_n) = \liminf_{n \rightarrow +\infty} \mathbb{E}(M - X_n) \geq \mathbb{E}(\liminf_{n \rightarrow +\infty} (M - X_n)) = M - \mathbb{E}(\limsup_{n \rightarrow +\infty} X_n). \end{cases}$$

On peut simplifier par M afin d'obtenir que

$$\begin{cases} \liminf_{n \rightarrow +\infty} \mathbb{E}(X_n) \geq \mathbb{E}(\liminf_{n \rightarrow +\infty} X_n). \\ -\limsup_{n \rightarrow +\infty} \mathbb{E}(X_n) \geq -\mathbb{E}(\limsup_{n \rightarrow +\infty} X_n) \Rightarrow \limsup_{n \rightarrow +\infty} \mathbb{E}(X_n) \leq \mathbb{E}(\limsup_{n \rightarrow +\infty} X_n). \end{cases}$$

Si on suppose de plus que X_n est convergente vers X alors on obtient que la suite des intégrales est convergente et que

$$\lim_{n \rightarrow +\infty} \mathbb{E}(X_n) = \mathbb{E}(\lim_{n \rightarrow +\infty} X_n).$$

C'est en fait un cas particulier du théorème de convergence dominée.

3.2.3 Théorème de convergence dominée

Théorème 3.2.6. *Théorème de convergence dominée (appelé aussi théorème de Lebesgue)*

Soit $(f_n)_{n \geq 0}$ une suite de fonctions mesurables sur Ω . Si :

- il existe g positive mesurable et intégrable telle que $\forall n \in \mathbb{N}, \forall x \in \Omega, |f_n(x)| \leq g(x)$
- et $f_n \xrightarrow{p.p.} f$

alors

- $\int_{\Omega} |f(x)| d\mu(x) < \infty$
- $\lim_{n \rightarrow +\infty} \int_{\Omega} |f_n(x) - f(x)| d\mu(x) = 0$.

Ce qui implique en particulier

$$\lim_{n \rightarrow +\infty} \int_{\Omega} f_n(x) d\mu(x) = \int_{\Omega} f(x) d\mu(x).$$

Démonstration. On peut supposer sans perte de généralité que (f_n) converge simplement vers f . Nous avons alors que pour tout x , $|f(x)| \leq g(x)$, donc $\int_{\Omega} |f(x)| d\mu(x) < \infty$, ce qui signifie que f est intégrable.

On commence par remarquer que pour tout x , $|f(x) - f_n(x)| \leq 2g(x)$ ce qui revient à dire que la fonction $h_n := 2g - |f - f_n|$ est positive et $\liminf_{n \rightarrow +\infty} h_n = 2g$ donc par le lemme de Fatou appliqué à la famille (h_n) .

$$\liminf_{n \rightarrow +\infty} \int_{\Omega} (2g(x) - |f(x) - f_n(x)|) d\mu(x) \geq \int_{\Omega} 2g(x) d\mu(x).$$

Mais en fait, il s'agit d'une égalité car le terme de gauche est majoré par $\int_{\Omega} 2g(x) d\mu(x)$. De plus par linéarité de l'intégrale dans l'espace \mathcal{L}^1 ,

$$\liminf_{n \rightarrow +\infty} \int_{\Omega} (2g(x) - |f(x) - f_n(x)|) d\mu(x) = \int_{\Omega} 2g(x) d\mu(x) - \limsup_{n \rightarrow +\infty} \int_{\Omega} |f(x) - f_n(x)| d\mu(x).$$

Donc

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \int_{\Omega} |f(x) - f_n(x)| d\mu(x) &= 0 \\ \lim_{n \rightarrow +\infty} \int_{\Omega} |f(x) - f_n(x)| d\mu(x) &= 0. \end{aligned}$$

Puis

$$\begin{aligned} \left| \int_{\Omega} f_n(x) d\mu(x) - \int_{\Omega} f(x) d\mu(x) \right| &= \left| \int_{\Omega} f(x) - f_n(x) d\mu(x) \right| \\ (\text{par lemme 2.2.14}) &\leq \int_{\Omega} |f(x) - f_n(x)| d\mu(x) \xrightarrow{n \rightarrow +\infty} 0. \end{aligned}$$

□

3.3 Conséquences des théorèmes de convergence déjà connues

3.3.1 Lien avec les intégrales impropres dans le cas positif

Corollaire 3.3.1. *Soit f une fonction positive et mesurable sur $(\mathbb{R}_+, \mathcal{L}(\mathbb{R}), \lambda_1)$ (à valeurs éventuellement infinies). On considère la famille croissante de fonctions $f_n = f \mathbf{1}_{[0,n]}$. Cette suite de fonctions converge simplement vers f et en appliquant le théorème de convergence monotone, on a*

$$\lim_{n \rightarrow +\infty} \int_{[0,n]} f_n(x) d\lambda_1(x) = \int_{\mathbb{R}_+} f(x) d\lambda_1(x).$$

Corollaire 3.3.2. *Soit f une fonction localement Riemann intégrable sur \mathbb{R}_+ de signe constant. Alors on a l'équivalence entre la convergence de l'intégrale généralisée $\int_0^{+\infty} f(x) dx$ et l'intégrabilité de f sur \mathbb{R}_+ . De plus, on peut toujours écrire*

$$\int_0^{+\infty} f(x) dx = \int_{\mathbb{R}_+} f(x) d\lambda_1(x)$$

3.3.2 Retour sur les fonctions dénombrablement étagées

On peut également retrouver le résultat (proposition 2.2.17)

Corollaire 3.3.3. *Pour toute fonction dénombrablement étagée $f : \mathbb{N} \rightarrow \mathbb{R}_+$ positive, en notant $f = \sum_{k=0}^{+\infty} \beta_k \mathbf{1}_{A_k}$ (pour des A_k en partition mesurable dénombrable). Alors f est intégrable si et seulement si la série $\sum_{k=0}^{+\infty} |\beta_k| \mu(A_k)$ est convergente. Dans ce cas*

$$\int_{\Omega} f(x) d\mu(x) = \sum_{k=0}^{+\infty} \beta_k \mu(A_k).$$

Démonstration. On peut supposer sans perte de généralités que f s'écrit $\sum_{k=0}^{+\infty} \beta_k \mathbf{1}_{A_k}$ (pour des A_k en partition mesurable dénombrable) et des β_k tous distincts.

Si on pose $g_n = \sum_{k=0}^n |\beta_k| \mathbf{1}_{A_k}$, la suite g_n est mesurable, positive et croissante, et ce sont des fonctions étagées, de plus elle converge simplement vers $|f|$, donc peut appliquer le théorème de convergence monotone pour déduire que

$$\int_{\Omega} |f|(x) d\mu(x) = \lim_{n \rightarrow +\infty} \int_{\Omega} g_n(x) d\mu(x) = \sum_{k=0}^{+\infty} |\beta_k| \mu(A_k).$$

Pour le calcul de l'intégrale sous l'hypothèse que f est intégrable, on peut dans un second temps utiliser le théorème de convergence dominée pour la suite de fonctions étagées $f_n = \sum_{k=0}^n \beta_k \mathbf{1}_{A_k}$. En effet,

- il existe $g = |f|$ positive mesurable et intégrable telle que $\forall n \in \mathbb{N}, |f_n| \leq g$
- et $f_n \rightarrow f$ (convergence simple)

On a donc

$$\int_{\Omega} f(x) d\mu(x) = \lim_{n \rightarrow +\infty} \int_{\Omega} f_n(x) d\mu(x) = \sum_{k=0}^{+\infty} \beta_k \mu(A_k).$$

□

3.4 Intégrales dépendant d'un paramètre

3.4.1 Position du problème

On se place dans le cadre d'un espace mesuré $(\Omega, \mathcal{A}, \mu)$. Soit $h : \mathbb{R} \times \Omega \rightarrow \mathbb{R}$, on définit une fonction $F(u) = \int_{\Omega} h(u, \omega) d\mu(\omega)$. Cette fonction F s'appelle, suivant les auteurs, une « intégrale à paramètre », « intégrale dépendant d'un paramètre », ... Dans cette partie, nous allons démontrer diverses propriétés des intégrales à paramètre à l'aide du théorème de convergence dominée.

Exemple 3.4.1. On considère un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et X une variable aléatoire intégrable. Pour $u \geq 0$, on considère ;

$$\varphi(F) = \int_{\Omega} \min(X(\omega), F) d\lambda(\omega) = \mathbb{E}(\min(X, F)).$$

Interprétation, si X est une distribution de dommages que l'on assure à l'aide d'une franchise F le montant remboursé est $\min(X, F)$. On peut interpréter $\varphi(F)$ comme le montant moyen du remboursement. Si la variable X est bornée par M , $\varphi(M) = 0$ (en fait pas d'assurance) tandis que $\varphi(0) = \mathbb{E}(X)$ (cas sans franchise). La continuité de φ signifie que la sensibilité par rapport au montant de la franchise est raisonnable. Elle sera une conséquence de cette section.

Exemple 3.4.2. Soit f une fonction mesurable bornée sur l'espace $(\Omega, \mathcal{A}, \mu) = (\mathbb{R}_+, \mathcal{L}(\mathbb{R}_+), \lambda)$. On considère pour $u > 0$, la valeur moyenne de f sur $[0, u]$;

$$\varphi(u) = \frac{1}{u} \int_{[0, u]} f(\omega) d\lambda(\omega) = \int_{\mathbb{R}_+} \frac{f(\omega) \mathbf{1}_{[0, u]}(\omega)}{u} d\lambda(\omega).$$

Donc $h(u, \omega) = \frac{f(\omega) \mathbf{1}_{[0, u]}(\omega)}{u}$.

Exemple 3.4.3. Si X et Y sont des variables aléatoires continues et indépendantes alors $X + Y$ est une variable aléatoire continues dont la densité f_{X+Y} est donnée par une fonction mesurable bornée sur l'espace $(\Omega, \mathcal{A}, \mu) = (\mathbb{R}_+, \mathcal{L}(\mathbb{R}_+), \lambda)$. On considère pour $u > 0$,

$$f_{X+Y}(u) = \int_{\mathbb{R}} f_X(y) f_Y(u - y) d\lambda(y) = \int_{\mathbb{R}} h(\omega, u) d\lambda(t).$$

Donc $h(u, \omega) = f_X(\omega) f_Y(\omega - u)$.

3.4.2 Continuité sous l'intégrale

Théorème 3.4.4. Continuité sous l'intégrale

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Soit $h : \Omega \times \mathbb{R} \rightarrow \mathbb{R}$ telle que

- (i) $\forall u \in \mathbb{R}, \omega \mapsto h(u, \omega)$ est mesurable, i.e. $h(u, \cdot) \in \mathcal{L}^0(\Omega, \mathcal{A}, \mathbb{R})$
- (ii) $\exists \bar{u}$ tel que pour presque tout ω , $u \mapsto h(u, \omega)$ est continue en \bar{u}
- (iii) $\exists g$ positive intégrable telle que $\forall u \in \mathbb{R}, |h(u, \omega)| \leq g(\omega)$, i.e. $g \in \mathcal{L}^1((\Omega, \mathcal{A}, \mu), \mathbb{R})$

Alors la fonction F définie par $F(u) = \int_{\Omega} h(u, \omega) d\mu(\omega)$ est définie en tout point $u \in \mathbb{R}$ et est continue en \bar{u} .

Si de plus, pour tout u de \mathbb{R} , $u \mapsto h(u, \omega)$ est continue, alors F est continue sur \mathbb{R} .

Démonstration. A l'aide de la caractérisation séquentielle de la limite, il suffit de montrer que $F(u_n) \xrightarrow[n \rightarrow +\infty]{} F(\bar{u})$ pour toute suite $(u_n)_{n \geq 0}$ convergeant vers \bar{u} .

On se fixe une telle suite $(u_n)_{n \geq 0}$ et on considère $\forall n, f_n(\omega) = h(u_n, \omega)$. Nous avons $f_n \xrightarrow[n \rightarrow +\infty]{\text{P.P.}} f$ avec $f(\omega) := h(\bar{u}, \omega)$ par (ii). Les fonctions f_n sont mesurables par (i). Par (iii), nous avons $\forall n, \forall x, |f_n(\omega)| \leq g(\omega)$ avec g intégrable. Donc par théorème de convergence dominée,

$$F(u_n) = \int_{\Omega} f_n(\omega) d\mu(\omega) \xrightarrow[n \rightarrow +\infty]{} \int_{\Omega} h(\omega) d\mu(\omega) = F(\bar{u}).$$

□

Exemple 3.4.5. *Convolution*

Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ intégrable et $\phi : \mathbb{R} \rightarrow \mathbb{R}$ bornée et continue. La convolée de f et ϕ est définie par

$$u \mapsto (f \star \phi)(u) := \int_{\mathbb{R}} \phi(u-x)f(x)\lambda(dx)$$

Notons $h(u, x) = \phi(u-x)f(x)$. Pour tout x , $u \mapsto \phi(u-x)f(x)$ est continue. Pour tout u , $|\phi(u-x)f(x)| \leq \|\phi\|_{\infty}|f(x)|$ et $\int_{\Omega} \|\phi\|_{\infty}|f(x)|\lambda(dx) < \infty$ par hypothèse. On rappelle que $\|\phi\|_{\infty} := \sup_{v \in \mathbb{R}} \phi(v)$. Pour tout $u \in \mathbb{R}$, $x \mapsto \phi(u-x)f(x)$ est mesurable comme produit de fonctions mesurables. Donc par le théorème de continuité globale, $f \star \phi$ est continue sur \mathbb{R} .

Remarque 3.4.6. Si f_X et f_Y sont deux densités de variables aléatoires. Si on suppose que f_Y est continue et vérifie $\lim_{x \rightarrow +\infty} f_Y(x) = 0$ et $\lim_{x \rightarrow -\infty} f_Y(x) = 0$ alors le résultat précédent s'applique car f_X est intégrable et f_Y est continue bornée.

3.4.3 Dérivation sous l'intégrale**Théorème 3.4.7.** *Dérivation sous l'intégrale*

Soit I un intervalle ouvert non vide de \mathbb{R} , $\bar{u} \in I$. Soit $h : I \times \Omega \rightarrow \mathbb{R}$ telle que

- (i) $\forall u \in I$, $\omega \mapsto h(u, \omega)$ est intégrable
- (ii) pour presque tout ω , $\frac{\partial h}{\partial u}(\bar{u}, \omega)$ existe
- (iii) $\exists g$ positive intégrable telle que $\forall u \in I$, $\forall \omega \in \Omega$, $|h(u, \omega) - h(\bar{u}, \omega)| \leq g(\omega)|u - \bar{u}|$.

Alors $F(u) := \int_{\Omega} h(u, \omega)\mu(d\omega)$ existe pour tout $u \in I$ et est dérivable en \bar{u} . De plus

$$F'(\bar{u}) = \int_{\Omega} \frac{\partial h}{\partial u}(\bar{u}, \omega) d\mu(\omega).$$

Démonstration. • L'existence de F est assurée par (i).

En ce qui concerne la dérivation, il suffit de montrer que pour toute suite $(u_n)_{n \geq 0}$ convergeant vers \bar{u} avec $\forall n$, $u_n \neq \bar{u}$,

$$\frac{F(u_n) - F(\bar{u})}{u_n - \bar{u}} \xrightarrow{n \rightarrow +\infty} \int_{\Omega} \frac{\partial h}{\partial u}(\bar{u}, \omega) d\mu(\omega).$$

Prenons donc une telle suite $(u_n)_{n \geq 0}$. Posons $\forall n$

$$f_n(\omega) = \frac{h(u_n, \omega) - h(\bar{u}, \omega)}{u_n - \bar{u}}.$$

- Par (ii), $f_n \xrightarrow[n \rightarrow +\infty]{\text{p.p.}} f := \frac{\partial h}{\partial u}(\bar{u}, \cdot)$.
- Par (iii), nous avons pour p.t. ω , $|f_n(\omega)| \leq g(\omega)$. Donc par théorème de convergence dominée,

$$\frac{F(u_n) - F(\bar{u})}{u_n - \bar{u}} = \int_{\Omega} \frac{h(u_n, \omega) - h(\bar{u}, \omega)}{u_n - \bar{u}} d\mu(\omega) \xrightarrow{n \rightarrow +\infty} \int_{\Omega} \frac{\partial h}{\partial u}(\bar{u}, \omega) d\mu(\omega).$$

□

Corollaire 3.4.8. *Dérivation « globale » sous l'intégrale*

Soit I un intervalle ouvert non vide de \mathbb{R} . Soit $h : I \times \Omega \rightarrow \mathbb{R}$ telle que

- (i) $\exists u_0 \in I$, $\omega \mapsto h(u_0, \omega)$ est intégrable
- (ii) pour presque tout ω , $u \mapsto h(u, \omega)$ est dérivable sur I
- (iii) $\forall \omega$, $\forall u$, $|\frac{\partial h}{\partial u}(u, \omega)| \leq g(\omega)$ avec g intégrable.

Alors $F(u) := \int_{\Omega} h(u, \omega) d\mu(\omega)$ existe et est dérivable sur I . De plus

$$F'(u) = \int_{\Omega} \frac{\partial h}{\partial u}(u, \omega) d\mu(\omega).$$

Démonstration. Pour tout $u \in I$,

$$\begin{aligned} |h(u, \omega)| &\leq |h(u_0, \omega)| + |h(u, \omega) - f(u_0, \omega)| \\ &\leq |h(u_0, \omega)| + |u - u_0| \sup_{v \in [u, u_0]} \left| \frac{\partial h}{\partial u}(v, \omega) \right| \\ &\leq |h(u_0, \omega)| + |u - u_0| g(\omega). \end{aligned}$$

Donc, par (i) et (iii), F est bien définie. Pour tous $u, \bar{u} \in I$, pour tout x ,

$$\begin{aligned} |h(u, \omega) - h(\bar{u}, \omega)| &\leq |u - \bar{u}| \sup_{v \in [u, \bar{u}]} \left| \frac{\partial h}{\partial u}(v, \omega) \right| \\ &\leq g(\omega) |u - \bar{u}| \end{aligned}$$

par (iii). Et le théorème précédent finit la démonstration. \square

Remarque 3.4.9. Ces théorèmes restent vrai si on remplace $u \in \mathbb{R}$ par $u \in I$ avec I intervalle de \mathbb{R} , voire un cas multidimensionnel.

3.5 Applications

3.5.1 Fonctions caractéristiques

Définition 3.5.1. Soit X v.a.r, la fonction caractéristique de X est

$$\begin{aligned} \Phi_X &: \mathbb{R} \rightarrow \mathbb{C} \\ \xi &\mapsto \int_{\mathbb{R}} e^{i\xi x} \mathbb{P}_X(dx) = \mathbb{E}(e^{i\xi X}). \end{aligned}$$

Remarque 3.5.2. Pour une fonction $f : \Omega \rightarrow \mathbb{C}$ avec $(\Omega, \mathcal{A}, \mu)$ un espace mesuré quelconque, on note

$$\int_{\Omega} f(x) d\mu(\omega) = \int_{\Omega} \operatorname{Re}(f)(\omega) d\mu(\omega) + i \int_{\Omega} \operatorname{Im}(f)(\omega) d\mu(\omega).$$

et donc dans la définition précédente

$$\Phi_X(\xi) = \int_{\mathbb{R}} \operatorname{Re}(e^{i\xi x}) \mathbb{P}_X(dx) + i \int_{\mathbb{R}} \operatorname{Im}(e^{i\xi x}) \mathbb{P}_X(dx).$$

Exercice 1. Soit X de loi $\mathcal{N}(m, \sigma^2)$. Alors $\Phi_X(\xi) = \exp\left(i\xi m - \frac{\sigma^2 \xi^2}{2}\right)$. En particulier, (loi normale centrée réduite) $\Phi_X(\xi) = \exp\left(-\frac{\xi^2}{2}\right)$.

Démonstration. Nous ne ferons la démonstration que dans le cas $m = 0, \sigma = 1, \xi \in \mathbb{R}$. Nous avons

$$\begin{aligned} \Phi_X(\xi) &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} e^{i\xi x} dx \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \operatorname{Re}(e^{-x^2/2} e^{i\xi x}) dx + i \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \operatorname{Im}(e^{-x^2/2} e^{i\xi x}) dx \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cos(x\xi) dx + i \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \sin(x\xi) dx \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cos(x\xi) dx + 0 \end{aligned}$$

car l'intégrale d'une fonction impaire sur \mathbb{R} est nulle.

Pour tout ξ , $\left| \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cos(x\xi) \right| \leq \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ qui est intégrable sur \mathbb{R} . Pour tout $x \in \mathbb{R}$, $\xi \mapsto \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cos(x\xi)$ est dérivable, de dérivée $\xi \mapsto \frac{1}{\sqrt{2\pi}} e^{-x^2/2} (-x \sin(x\xi))$. Pour tous ξ, x ,

$\left| \frac{1}{\sqrt{2\pi}} e^{-x^2/2} (-x \sin(x\xi)) \right| \leq \frac{1}{\sqrt{2\pi}} e^{-x^2/2} |x|$ qui est intégrable sur \mathbb{R} . En effet, par symétrie,

$$\begin{aligned} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} |x| dx &= 2 \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} x dx \\ &= \left[\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right]_0^{+\infty} \\ &= \frac{1}{\sqrt{2\pi}}. \end{aligned}$$

Donc par théorème de dérivation globale (cf. cor. 3.4.8)

$$\begin{aligned} \Phi'_X(\xi) &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} (-x \sin(x\xi)) dx \\ &= \left[e^{-x^2/2} \sqrt{2\pi} \sin(x\xi) \right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} e^{-x^2/2} \sqrt{2\pi} \xi \cos(x\xi) dx \\ &= 0 - \xi \Phi_X(\xi). \end{aligned}$$

Nous avons donc l'équation

D'où

$$\begin{aligned} \log(\Phi_X(\xi)) - \log(\Phi_X(0)) &= -\frac{\xi^2}{2} \\ \Phi_X(\xi) &= \Phi_X(0) e^{-\xi^2/2}. \end{aligned}$$

Remarquons que $\Phi_X(0) = \mathbb{E}(1) = 1$. Nous avons donc $\Phi_X(\xi) = e^{-\xi^2/2}$. \square

Remarque 3.5.3. Soit X et Y deux variables gaussiennes, si $\Phi_X = \Phi_Y$ alors X et Y ont même loi.

Exercice 2. Soit X une loi discrète finie résumée par $((p_1, x_1), \dots, p_n, x_n)$. Alors

$$\Phi_X(\xi) = p_1 \exp(i\xi x_1) + \dots + p_n \exp(i\xi x_n).$$

Remarque 3.5.4. Soit X et Y deux variables suivant des lois discrète finies, si $\Phi_X = \Phi_Y$ alors X et Y ont même loi.

Proposition 3.5.5. Si X est intégrable alors Φ_X est dérivable et $\Phi'_X(\xi) = \int_{\mathbb{R}} ix(e^{i\xi x}) d\mathbb{P}_X(x)$. En particulier $\Phi'_X(0) = i\mathbb{E}(X)$.

Si de plus X^k est intégrable alors Φ_X est dérivable k fois et $\Phi_X(k)(0) = (i)^k \mathbb{E}(X^k)$.

Démonstration. Si on pose $h(\xi, x) = e^{i\xi x}$, puisque

- pour tout x , l'application $\omega \mapsto h(\xi, x)$ est bornée donc \mathbb{P}_X -intégrable.
- pour tout ω , l'application $\frac{\partial h}{\partial \xi} h(\xi, x)$ existe et vaut $ix e^{i\xi x}$.
- Pour tout ξ , pour tout x , l'application $\left| \frac{\partial h}{\partial \xi} h(\xi, \omega) \right| \leq |x|$ existe (or la fonction identité est \mathbb{P}_X -intégrable). et que l'application

$$\Phi_X(\xi) = \int_{\mathbb{R}} \operatorname{Re}(e^{i\xi x}) \mathbb{P}_X(dx) + i \int_{\Omega} \operatorname{Im}(e^{i\xi x}) \mathbb{P}_X(dx).$$

\square

Théorème 3.5.6. Dans le cas général, la fonction caractéristique d'une v.a.r. caractérise entièrement la loi de cette variable. C'est à dire que si X et Y des v.a.r. ont même fonction caractéristique alors X et Y ont même loi.

3.5.2 Fonctions génératrices (partie non traitée cette année)

Cette partie ne fait pas partie du programme mais est utile pour le prochain semestre.

Définition 3.5.7. Soit X une v.a. à valeurs dans \mathbb{N} . On appelle fonction génératrice de X la fonction

$$\begin{aligned} g_X &: [0, 1] \rightarrow \mathbb{R} \\ r &\mapsto \mathbb{E}(r^X) = \sum_{n=0}^{+\infty} \mathbb{P}(X = n) r^n \end{aligned}$$

A l'aide du cours d'analyse S3 sur les séries entières, on peut énoncer la proposition suivante. En effet, le rayon de convergence est supérieur ou égal à 1. La fonction est \mathcal{C}^∞ sur $] - 1, 1[$ et les coefficients de la série sont liées aux dérivées successives de g_X .

Proposition 3.5.8. *Si X est une v.a. à valeurs dans \mathbb{N} , la fonction génératrice caractérise la loi de X .*

Exemple 3.5.9. *Soit $X \sim \mathcal{P}(\lambda)$ (ce qui veut dire que X est de loi $\mathcal{P}(\lambda)$). Calculons*

$$\begin{aligned} g_X(u) &= \sum_{n=0}^{+\infty} u^n \frac{\lambda^n e^{-\lambda}}{n!} \\ &= e^{\lambda u} e^{-\lambda} = e^{-\lambda(1-u)} . \end{aligned}$$

Chapitre 4

Mesure produit et applications

4.1 Mesure produit et théorèmes de Fubini

4.1.1 Mesure produit

On se donne deux espaces mesurés $(\Omega, \mathcal{A}, \mu)$ et $(\Omega', \mathcal{A}', \mu')$.

Théorème 4.1.1. *Sur l'ensemble $\Omega \times \Omega'$, il existe une « plus petite tribu » \mathcal{C} contenant tous les ensembles de la forme $A \times B$ avec $A \in \mathcal{A}$, $B \in \mathcal{A}'$. On note cette tribu $\mathcal{C} = \mathcal{A} \otimes \mathcal{A}'$.*

De plus, il existe une unique mesure, notée $\mu \otimes \mu'$ sur $\mathcal{A} \otimes \mathcal{A}'$ telle que,

$$\text{si } (A, B) \in \mathcal{A} \times \mathcal{A}', \quad (\mu \otimes \mu')(A \times B) = \mu(A)\mu'(B).$$

Définition 4.1.2. *La mesure $\mu \otimes \mu'$ définie par le théorème ci-dessus s'appelle la mesure produit de μ et μ' . La tribu \mathcal{C} définie par le théorème ci-dessus s'appelle la tribu produit.*

Définition 4.1.3. *On notera $\mathcal{B}(\mathbb{R}^d)$ la tribu $\mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R})^{\otimes d}$ (produit d fois).*

La mesure $\mu_1 \otimes \mu_1$ sur $\mathcal{B}(\mathbb{R}^2)$ mesure les aires, la mesure $\mu_1 \otimes \mu_1 \otimes \mu_1 = \mu_1^{\otimes 3}$ sur $\mathcal{B}(\mathbb{R}^3)$ mesure les volumes, ...

Exemple 4.1.4. *Cas particulier $\Omega = \Omega' = \mathbb{N}$ muni de $\mathcal{P}(\mathbb{N})$ et de la mesure de comptage sur \mathbb{N} . Alors $\mathcal{A} \otimes \mathcal{A}' = \mathcal{P}(\mathbb{N}^2)$ et $\mu_C \otimes \mu_C$ est la mesure de comptage sur \mathbb{N}^2 qui sera également notée μ_C .*

Notation 4.1.5. *Soit $E \subset \Omega \times \Omega'$, on note pour tout x de Ω l'ensemble éventuellement vide $E_{x,\bullet}$, De même si $y \in \Omega'$,*

$$E_{x,\bullet} = \{y \in \Omega' \mid (x, y) \in E\}.$$

$$E_{\bullet,y} = \{x \in \Omega \mid (x, y) \in E\}.$$

Proposition 4.1.6. *Soit $E \subset \Omega \otimes \Omega'$, alors pour tout x de Ω l'ensemble $E_{x,\bullet} \in \mathcal{A}'$, et si $y \in \Omega'$, $E_{\bullet,y} \in \mathcal{A}$.*

Lemme 4.1.7. *Soit $E \subset \Omega \times \Omega'$ un ensemble appartenant à $\mathcal{A} \otimes \mathcal{A}'$. Alors*

- (i) *l'application $x \mapsto \mu_2(E_{x,\bullet}) \in \mathcal{L}^0(\Omega, \mathcal{A}, \mu)$.*
- (ii) *l'application $y \mapsto \mu_1(E_{\bullet,y}) \in \mathcal{L}^0(\Omega', \mathcal{A}', \mu')$*
- (iii) *dans \mathbb{R}_+ , on a*

$$(\mu \otimes \mu')(E) = \int_{\Omega} \mu'(E_{x,\bullet}) d\mu(x) = \int_{\Omega'} \mu(E_{\bullet,y}) d\mu'(y).$$

Exemple 4.1.8. *Si E est un cercle de \mathbb{R}^2 , puisque les ensembles $E_{x,\bullet}$ sont finis donc de mesure nulle, on en déduit que $\mu_2(E) = 0$.*

Remarque 4.1.9. *Dans le cas général $E \subset \Omega \times \Omega'$, on peut remarquer que*

$$\mu'(E_{x,\bullet}) = \int_{E_{x,\bullet}} d\mu'(y) = \int_{\Omega'} \mathbf{1}_{E_{x,\bullet}}(y) d\mu'(y)$$

ce qui permet d'écrire

$$(\mu \otimes \mu')(E) = \int_E d(\mu \otimes \mu')(x, y) \quad (4.1.1)$$

$$= \int_{\Omega \times \Omega'} \mathbf{1}_E(x, y) d(\mu \otimes \mu')(x, y) \quad (4.1.2)$$

$$= \int_{\Omega} \left(\int_{\Omega'} \mathbf{1}_{E_{x, \cdot}}(y) d\mu'(y) \right) d\mu(x) \quad (4.1.3)$$

$$= \int_{\Omega} \left(\int_{\Omega'} \mathbf{1}_E(x, y) d\mu'(y) \right) d\mu(x) \quad (4.1.4)$$

$$(4.1.5)$$

4.1.2 Théorèmes de Fubini

Théorème 4.1.10. *Théorème de Fubini-Tonelli*

Soit $f : \Omega \times \Omega' \rightarrow [0, +\infty]$ mesurable positive. On définit les fonctions ϕ et ψ sur Ω et Ω' respectivement par

$$\phi(x) = \int_{\Omega'} f(x, y) \mu'(dy), \quad \psi(y) = \int_{\Omega} f(x, y) d\mu(x).$$

Ces fonctions sont mesurables positives et vérifient dans $\overline{\mathbb{R}}_+$.

$$\int_{\Omega} \phi(x) d\mu(x) = \int_{\Omega \times \Omega'} f(x, y) \mu \otimes \mu'(dx, dy) = \int_{\Omega'} \psi(y) \mu'(dy)$$

$$\int_{\Omega} \left(\int_{\Omega'} f(x, y) d\mu'(y) \right) d\mu(x) = \int_{\Omega \times \Omega'} f(x, y) d(\mu \otimes \mu')(x, y) = \int_{\Omega'} \left(\int_{\Omega} f(x, y) d\mu(x) \right) d\mu'(y)$$

Exemple 4.1.11. On retient que pour des fonctions positives, on peut intervertir l'ordre des intégrations. Cas particulier $\Omega = \Omega' = \mathbb{N}$ muni de $\mathcal{P}(\mathbb{N})$ et de la mesure de comptage sur \mathbb{N} . Si $(u_{n,k})_{n,k}$ est une suite positive à double indice.

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

Théorème 4.1.12. *Théorème de Fubini (ou Fubini-Lebesgue)*

Soit $f : \Omega \times \Omega' \rightarrow \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction mesurable. On définit les fonctions f_1 et f_2 sur Ω et Ω' respectivement par

$$f_1(x) = \int_{\Omega'} |f(x, y)| \mu'(dy), \quad f_2(y) = \int_{\Omega} |f(x, y)| d\mu(x).$$

(i) Si l'une des fonctions f_1 ou f_2 est intégrable alors l'autre l'est aussi et dans ce cas, f , ϕ et ψ sont intégrables. De plus, nous avons alors

$$\int_{\Omega} \phi(x) d\mu(x) = \int_{\Omega \times \Omega'} f(x, y) \mu \otimes \mu'(dx, dy) = \int_{\Omega'} \psi(y) \mu'(dy).$$

(ii) Si f est intégrable (contre la mesure $\mu \otimes \mu'$) alors f_1 et f_2 sont intégrables et nous avons encore l'égalité ci-dessus.

Exemple 4.1.13. Soit

$$f : [0, 1] \times [0, 1] \rightarrow \mathbb{R}_+ \\ (x, y) \mapsto e^{-(x+y)} \mathbf{1}_{x+y \leq 1}.$$

Cette fonction est mesurable positive. Par Fubini-Tonelli (et puisque les fonctions sont continues, on peut utiliser l'intégrale de Riemann)

$$\begin{aligned}
 \int_{[0,1] \times [0,1]} f(x,y) d(\lambda \otimes \lambda)(x,y) &= \int_0^1 \left(\int_0^1 e^{-(x+y)} \mathbf{1}_{x+y \leq 1} dx \right) dy \\
 &= \int_0^1 e^{-y} \left(\int_0^1 e^{-y} \mathbf{1}_{x+y \leq 1} dx \right) dy \\
 &= \int_0^1 e^{-y} \left(\int_0^{1-y} e^{-x} dx \right) dy \\
 &= \int_0^1 e^{-y} (1 - e^{-(1-y)}) dy \\
 &= \int_0^1 e^{-y} - e^{-1} dy \\
 &= 1 - e^{-1} - e^{-1} = 1 - \frac{2}{e}.
 \end{aligned}$$

Notation 4.1.14. *Intégrale multiple*

Pour toute fonction $f : \mathbb{R}^k \rightarrow \mathbb{R}$ intégrable, on notera indifféremment

$$\begin{aligned}
 \int_{\mathbb{R}^k} f(x_1, \dots, x_k) d\lambda^{\otimes k}(x_1, \dots, x_k) &= \int_{\mathbb{R}^k} f(x_1, \dots, x_k) d\lambda(x_1) \dots d\lambda(x_k) \\
 &= \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f(x_1, \dots, x_k) d\lambda(x_1) \dots d\lambda(x_k) \\
 &= \int_{\mathbb{R}^d} f(u) du
 \end{aligned}$$

(on a remplacé, dans cette écriture, (x_1, \dots, x_d) par u).

Exemple 4.1.15 (non présenté en cours). Soit

$$\begin{aligned}
 f : \mathbb{R}_+ \times [0,1] &\rightarrow \mathbb{R} \\
 (x,y) &\mapsto 2e^{-2xy} - e^{-xy}.
 \end{aligned}$$

Cette fonction est mesurable et n'est pas de signe constant MAIS

$$\int_0^{+\infty} \left(\int_0^1 f(x,y) dx \right) dy \neq \int_0^1 \left(\int_0^{+\infty} f(x,y) dy \right) dx.$$

Exemple 4.1.16. *Interversion de somme et d'intégrale*

Soit $f : \Omega \times \Omega' \rightarrow \mathbb{R}_+$ mesurable positive. Nous supposons dans cet exemple que $(\Omega', \mathcal{A}', \mu') = (\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu_C)$. Comme nous l'avons vu dans l'exemple 3.2.2, pour toute fonction g positive sur Ω' ,

$$\int_{\Omega'} g(y) d\mu'(dy) = \sum_{k \geq 0} g(k).$$

Par Fubini-Tonelli, nous avons alors

$$\int_{\Omega} \left(\sum_{k \geq 0} f(x,k) \right) d\mu(x) = \sum_{k \geq 0} \left(\int_{\Omega} f(x,k) d\mu(x) \right).$$

Remarque 4.1.17. Il importe de comprendre que les théorèmes de Fubini sont un résultat sur les mesures produits, pas seulement sur \mathbb{R}^2 pour $\lambda_2 = \lambda \otimes \lambda$. En particulier, on va utiliser ce résultat sur \mathbb{R}^2 mais pour $\mathbb{P}_X \otimes \mathbb{P}_Y$.

4.2 Changement de variable

4.2.1 Théorème de changement de variable

On va préciser ce que l'on savait de la mesure de Lebesgue sur \mathbb{R}^n .

Remarque 4.2.1. Soit $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ une transformation affine, (si on raisonne matriciellement. i.e. $F(x) = Ax + b$ où A est une matrice carrée). Alors pour toute partie mesurable $D \subset \mathbb{R}^n$, on a $\lambda_n(F(D)) = |\det A| \lambda_n(D)$. Attention à ne pas oublier la valeur absolue dans les calculs. En particulier, elle est invariante par translation, rotation et même par isométrie ($A^{-1} = A^t$).

Définition 4.2.2. Soient U et V deux ouverts de \mathbb{R}^d . Un difféomorphisme ϕ de U dans V est une bijection $\phi : (U \rightarrow V)$ qui est \mathcal{C}^1 telle que ϕ^{-1} est \mathcal{C}^1 aussi.

Rappel : \mathcal{C}^1 veut dire que la fonction est continue et ses dérivées partielles du premier ordre existent et sont continues. De manière plus explicite, la fonction

$$\phi : \begin{array}{ccc} U & \rightarrow & V \\ (u_1, \dots, u_d) & \mapsto & (\phi_1(u_1, \dots, u_d), \dots, \phi_d(u_1, \dots, u_d)) \end{array}$$

est \mathcal{C}^1 si ϕ_1, \dots, ϕ_d sont continues et $\forall i, j, \frac{\partial \phi_i}{\partial u_j}$ existe et est continue.

Définition 4.2.3. Si ϕ est un difféomorphisme de U dans V (deux ouverts de \mathbb{R}^d), on appelle matrice jacobienne la matrice suivante (fonction de (u_1, \dots, u_d))

$$J_\phi = \begin{bmatrix} \frac{\partial \phi_1}{\partial u_1}(u_1, \dots, u_d) & \dots & \frac{\partial \phi_d}{\partial u_1}(u_1, \dots, u_d) \\ \frac{\partial \phi_1}{\partial u_2}(u_1, \dots, u_d) & \dots & \frac{\partial \phi_d}{\partial u_2}(u_1, \dots, u_d) \\ \dots & \dots & \dots \\ \frac{\partial \phi_1}{\partial u_d}(u_1, \dots, u_d) & \dots & \frac{\partial \phi_d}{\partial u_d}(u_1, \dots, u_d) \end{bmatrix}$$

Proposition 4.2.4. Soient U et V deux ouverts de \mathbb{R}^d . Une application ϕ de U dans V est un difféomorphisme si et seulement si elle constitue une bijection qui est \mathcal{C}^1 et telle que J_ϕ est partout inversible.

En particulier, si $n = 1$, soient U et V deux ouverts de \mathbb{R} , une application ϕ de U dans V est un difféomorphisme si et seulement si elle est bijective, \mathcal{C}^1 et dont la dérivée ne s'annule pas.

Lemme 4.2.5. lemme de changement de variable.

Soit $\phi : U \rightarrow U'$ un difféomorphisme entre deux ouverts de \mathbb{R}^n . Alors pour toute partie mesurable $D \subset U$, on a

$$\lambda_n(\phi(D)) = \int_D |\det(J_\phi(x))| d\lambda_n(x)$$

En particulier, si on considère D_ε boule de centre $x_0 \in U$ et de rayon $\varepsilon > 0$. On a

$$|\det(J_\phi(x_0))| = \lim_{\varepsilon \rightarrow 0^+} \frac{\lambda_n(\phi(D_\varepsilon))}{\lambda_n(D_\varepsilon)}$$

Théorème 4.2.6. Théorème de changement de variable.

Soient U, U' deux ouverts de \mathbb{R}^n . Soit $\phi : U \rightarrow U'$ un difféomorphisme. Soit f une fonction borel¹ mesurable et intégrable $U' \rightarrow \mathbb{R}$. Alors la fonction $f \circ \phi : U \rightarrow \mathbb{R}$ est Borel mesurable, de plus et de plus la fonction $|\det(J_\phi)| \times f \circ \phi : U \rightarrow \mathbb{R}$ est intégrable

$$\int_{U'} f(y) d\lambda_n(y) = \int_U (f \circ \phi)(x) \times |\det(J_\phi(x))| d\lambda_n(x)$$

Remarque 4.2.7. Dans le même cadre, si f est à valeurs positives, la formule de changement de variable reste vraie sans hypothèse d'intégralité.

On rappelle le résultat de L2.

Théorème 4.2.8. Soient I et J deux intervalles, f une fonction sur I continue (de primitive F) et g dans $\mathcal{C}^1(J, I)$. Soient $\alpha, \beta \in J$ et $a = g(\alpha)$, $b = g(\beta)$ alors

$$[F]_a^b = \int_a^b f(y) dy = \int_\alpha^\beta f(g(x))g'(x) dx = [F \circ g]_\alpha^\beta.$$

1. Remarque. Dans ce théorème, on suppose que h est borélienne, cela entraîne que $f \circ \phi$ est aussi borélienne. Si on supposait seulement que f est mesurable au sens de Lebesgue, on ne pourrait rien conclure sur $f \circ \phi$.

On peut en première analyse est étonné par la présence de la valeur absolue dans le cas du théorème 4.2.6 qui ne figure pas dans le cas du théorème 4.2.8. Pour comparer les deux versions, il faut d'abord envisager une application g qui soit \mathcal{C}^1 et qui soit un difféomorphisme donc la dérivée ne s'annule pas. On va regarder ici le cas le plus subtil $g' < 0$.

Si $g' < 0$, on pose $U =]\alpha, \beta[$, puisque la fonction g est décroissante, $V =]b, a[$ car $b < a$. La fonction $\Phi = g$ réalise bien un difféomorphisme entre U et V . Attention, lorsque les bornes ne sont pas dans le bon ordre, on ne peut plus identifier l'intégrale au sens de Lebesgue avec l'intégrale au sens de Riemann, il faut modifier le signe.

$$\begin{aligned} [F]_a^b &= \int_a^b f(y) dy = - \int_b^a f(y) dy = - \int_{]b, a[} f(y) dy = - \int_V f(y) dy. \\ \int_\alpha^\beta f(g(x))g'(x) dx &= \int_{] \alpha, \beta [} f(g(x))g'(x) dx = \int_U f(g(x))g'(x) dx \end{aligned}$$

Donc le théorème de L2 nous donne

$$\int_V f(y) dy = \int_U f(g(x))(-g'(x)) dx$$

On retrouve la formule pour l'intégrale de Lebesgue car la matrice jacobienne $J_\Phi(x)$ est ici une matrice 1×1 dont l'unique coefficient vaut $g'(x) < 0$ et donc $-g'(x) = |g'(x)|$.

Donc, en dimension 1, dans le cas d'un difféomorphisme, on peut faire le changement de variable avec le théorème ci-dessus ou directement sur l'intégrale de Riemann, les deux formules coïncident.

4.2.2 Coordonnées polaires

Exemple 4.2.9. *Changement de variables en coordonnées polaires.*

Soit

$$\begin{aligned} \phi :]0, +\infty[\times]0, \frac{\pi}{2}[&\rightarrow \mathbb{R}_+^* \times \mathbb{R}_+^* \\ (\rho, \theta) &\rightarrow (\rho \cos(\theta), \rho \sin(\theta)). \end{aligned}$$

L'application ϕ est un difféomorphisme. Calculons sa matrice jacobienne

$$J_\phi(\rho, \theta) = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\rho \cos(\theta) & \rho \sin(\theta) \end{bmatrix}.$$

Nous avons donc $|\det J_\phi|(\rho, \theta) = |\rho \cos^2(\theta) + \rho \sin^2(\theta)| = |\rho| = \rho$. Donc, par le théorème 4.2.6

$$\begin{aligned} \int_0^{+\infty} \int_0^{+\infty} e^{-(x^2+y^2)} dx dy &= \int_0^{+\infty} \int_0^{\frac{\pi}{2}} e^{-\rho^2} |\rho| d\theta d\rho \\ &= \frac{\pi}{2} \int_0^{+\infty} \rho e^{-\rho^2} d\rho \\ &= \frac{\pi}{2} \left[\frac{1}{2} e^{-\rho^2} \right]_0^{+\infty} = \frac{\pi}{4}. \end{aligned}$$

Or

$$\begin{aligned} \int_0^{+\infty} \int_0^{+\infty} e^{-(x^2+y^2)} dx dy &= \int_0^{+\infty} e^{-x^2} \left(\int_0^{+\infty} e^{-y^2} dy \right) dx \\ &= \left(\int_0^{+\infty} e^{-y^2} dy \right) \times \int_0^{+\infty} e^{-x^2} dx \\ &= \left(\int_0^{+\infty} e^{-y^2} dy \right)^2. \end{aligned}$$

Donc

$$\int_0^{+\infty} e^{-y^2} dy = \frac{\sqrt{\pi}}{2}. \quad (4.2.1)$$

En utilisant la symétrie $\int_{-\infty}^0 e^{-y^2} dy$ vaut la même chose. Pour finir à l'aide d'un changement de variables $t = y\sqrt{2}$, $dt = \sqrt{2} dy$, $dy = \frac{1}{\sqrt{2}} dt$,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} \left(\frac{1}{\sqrt{2}} dt\right) = \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} e^{-y^2} d\lambda(y) = \frac{1}{\sqrt{\pi}} \sqrt{\pi} = 1.$$

4.2.3 Convolution

Exemple 4.2.10. *Convolution*

Soient $f, g : \mathbb{R} \rightarrow \mathbb{R}$ deux fonctions intégrables. On définit la convolée de f et g par

$$(f \star g)(x) = \int_{\mathbb{R}} f(y)g(x-y) dy.$$

Montrons que cette fonction est bien définie (c'est à dire que $f \star g < \infty$ p.p.). Nous avons

$$\begin{aligned} \int_{\mathbb{R}} |(f \star g)(x)| dx &= \int_{\mathbb{R}} \left| \int_{\mathbb{R}} f(y)g(x-y) dy \right| dx \\ &\leq \int_{\mathbb{R}} \int_{\mathbb{R}} |f(y)| \times |g(x-y)| dy dx \\ (\text{Fubini-Tonelli}) &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(y)| \times |g(x-y)| dx \right) dy \\ &= \int_{\mathbb{R}} |f(y)| \int_{\mathbb{R}} |g(x-y)| dx dy \\ &= \int_{\mathbb{R}} |f(y)| \left(\int_{\mathbb{R}} |g(x-y)| dx \right) dy. \end{aligned}$$

Pour y fixé, nous avons par changement de variable en dimension 1 ($u = x - y$, $x = u + y$, $dx = du$)

$$\begin{aligned} \int_{\mathbb{R}} |g(x-y)| dx &= \int_{-\infty}^{+\infty} |g(x-y)| dx \\ &= \int_{-\infty}^{+\infty} |g(u)| du \\ &= \int_{\mathbb{R}} |g(u)| du. \end{aligned}$$

Donc

$$\int_{\mathbb{R}} |(f \star g)(x)| dx \leq \int_{\mathbb{R}} |f(y)| \left(\int_{\mathbb{R}} |g(u)| du \right) dy \quad (4.2.2)$$

$$= \left(\int_{\mathbb{R}} |g(u)| du \right) \times \left(\int_{\mathbb{R}} |f(y)| dy \right) < \infty \quad (4.2.3)$$

car f et g sont intégrables. Par la remarque 2.2.25, $|f \star g|$ est donc finie presque partout, donc $f \star g$ est p.p. finie.

Fixons x et opérons un changement de variable $y = x - u$ dans l'intégrale :

$$\begin{aligned} \int_{\mathbb{R}} f(y)g(x-y) dy &= \int_{-\infty}^{+\infty} f(y)g(x-y) dy \\ &= \int_{+\infty}^{-\infty} f(x-u)g(u)(-du) \\ &= \int_{-\infty}^{+\infty} f(x-u)g(u) du \\ &= \int_{\mathbb{R}} f(x-u)g(u) du. \end{aligned}$$

Donc $f \star g = g \star f$.. Si de plus, f et g sont des densités de probabilité, on peut écrire sans valeur absolue la formule (4.2.3) et on sait que $f \star g$ sera une densité de probabilité.

4.3 Variables indépendantes (vu seulement en cours)

4.3.1 Événements et variables indépendantes

On se donne dans toute la fin du chapitre un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

Définition 4.3.1. On dit que p événements A_1, A_2, \dots, A_p (éléments de \mathcal{A}) sont indépendants et on note $A_1 \perp\!\!\!\perp A_2 \perp\!\!\!\perp \dots \perp\!\!\!\perp A_p$. si

$$\mathbb{P}(A_1 \cap \dots \cap A_p) = \mathbb{P}(A_1) \times \dots \times \mathbb{P}(A_p).$$

On considère une suite d'événements $(A_0, A_1, \dots) \in \mathcal{A}^{\mathbb{N}}$. On dit que les événements A_1, A_2, \dots sont indépendants et on note $A_1 \perp\!\!\!\perp A_2 \perp\!\!\!\perp \dots \perp\!\!\!\perp A$ si $\forall j_1, \dots, j_p$ (indices distincts), les événements $A_{j_1}, A_{j_2}, \dots, A_{j_p}$ sont indépendants.

$$\mathbb{P}(A_{j_1} \cap \dots \cap A_{j_p}) = \mathbb{P}(A_{j_1}) \times \dots \times \mathbb{P}(A_{j_p}).$$

Remarque 4.3.2. Pour que les événements ci-dessus soient indépendants, il ne suffit pas qu'ils soient deux à deux indépendants (c'est à dire $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j), \forall i \neq j$).

Lemme 4.3.3. $A_1, A_2, \dots \in \mathcal{A}$ sont indépendants alors A_1^c, A_2^c, \dots sont indépendants.

Démonstration. Nous ne faisons la démonstration que pour deux événements A_1, A_2 . Nous avons (en utilisant les propriétés des mesures)

$$\begin{aligned} \mathbb{P}(A_1^c \cap A_2^c) &= \mathbb{P}((A_1 \cup A_2)^c) \\ &= 1 - \mathbb{P}(A_1 \cup A_2) \\ &= 1 - \mathbb{P}((A_1 \setminus A_2) \sqcup (A_2 \setminus A_1) \sqcup (A_1 \cap A_2)) \\ &= 1 - \mathbb{P}(A_1 \setminus A_2) - \mathbb{P}(A_2 \setminus A_1) - \mathbb{P}(A_1 \cap A_2) \\ &= 1 - (\mathbb{P}(A_1) - \mathbb{P}(A_1 \cap A_2)) - (\mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)) - \mathbb{P}(A_1 \cap A_2) \\ &= 1 - \mathbb{P}(A_1) - \mathbb{P}(A_2) + \mathbb{P}(A_1)\mathbb{P}(A_2) \quad (\text{car } A_1 \text{ indépendant de } A_2) \\ &= (1 - \mathbb{P}(A_1))(1 - \mathbb{P}(A_2)) = \mathbb{P}(A_1^c)\mathbb{P}(A_2^c) \end{aligned}$$

□

Définition 4.3.4. Soient X_1, \dots, X_n des variables aléatoires à valeurs non nécessairement réelles. On suppose que $X_i \in \mathcal{L}^0((\Omega, \mathcal{A}), (E_i, \mathcal{U}_i))$. On dit que X_1, \dots, X_n sont indépendantes si $\forall (F_1, \dots, F_n) \in \mathcal{U}_1 \times \dots \times \mathcal{U}_n$,

$$\mathbb{P}(\{X_1 \in F_1\} \cap \dots \cap \{X_n \in F_n\}) = \mathbb{P}(\{X_1 \in F_1\}) \times \dots \times \mathbb{P}(\{X_n \in F_n\}).$$

On notera $X_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp X_n$.

Remarque 4.3.5. Pour que X_1, \dots, X_n soient indépendants, il ne suffit pas qu'ils soient deux à deux indépendants.

On rappelle la notion de tribu engendrée par une fonction (définition 1.5.6). Un cas particulier de tribu engendrée est l'exemple introductif du paragraphe 1.2.1.

Définition 4.3.6. Soit Y v.a. à valeurs dans un espace mesurable quelconque (E, \mathcal{E}) . La tribu engendrée par Y est $\sigma(Y) = \{Y^{-1}(A), A \in \mathcal{E}\}$. La famille $\sigma(Y)$ est une tribu et $\sigma(Y) \subset \mathcal{A}$.

Exemple 4.3.7. Soit X une variable aléatoire discrète réelle dont l'ensemble image est $\{a_i \mid i \in I\}$ où I est un sous ensemble fini ou dénombrable et $a_i \neq a_j$. Alors les ensembles $A_i = \{X = a_i\}$ constituent une partition de Ω et Y s'écrit $\sum_{i \in I} a_i \mathbf{1}_{A_i}$ tandis que $\sigma(Y)$ est la tribu engendrée par la partition $(A_i)_{i \in I}$.

Remarque 4.3.8. Si X est une variable aléatoire discrète finie qui a comme valeurs $\{a_1, \dots, a_n\}$ respectivement Y est une variable aléatoire discrète finie qui a comme valeurs $\{b_1, \dots, b_m\}$. Alors X et Y sont indépendants si et seulement si

$$\mathbb{P}(X = a_i, Y = b_j) = \mathbb{P}(X = a_i) \mathbb{P}(Y = b_j).$$

Définition 4.3.9. Soit $\mathcal{A}_1, \dots, \mathcal{A}_n$ n tribus sur Ω . On dit qu'elles sont indépendantes si $\forall A_1 \in \mathcal{A}_1, \dots, \forall A_n \in \mathcal{A}_n, A_1, \dots, A_n$ sont indépendants. (En d'autres termes, des événements appartenant à des tribus indépendantes sont indépendants.)

Remarque 4.3.10. Soient X_1, \dots, X_m des variables indépendantes, alors les tribus associées $\sigma(X_1), \dots, \sigma(X_m)$, sont indépendantes. (En d'autres termes, des événements relatifs à des variables indépendantes sont indépendants.)

Proposition 4.3.11. Soient X_1, \dots, X_n des variables aléatoires réelles indépendantes. Si X_i est à valeurs dans $E_i \in \mathcal{B}(\mathbb{R})$ et si chaque application $f_i : E_i \rightarrow \mathbb{R}$ est mesurable, alors $\sigma(f(X_i)) \subset \sigma(X_i)$, et on en déduit que les variables aléatoires $f_1(X_1), \dots, f_n(X_n)$ sont indépendantes.

4.3.2 Conséquences immédiates de l'indépendance

Théorème 4.3.12. Soient X_1, \dots, X_n des variables aléatoires réelles indépendantes et intégrables. Alors le produit $X_1 \dots X_n$ est intégrable et

$$\mathbb{E}(X_1 \dots X_n) = \mathbb{E}(X_1) \times \dots \times \mathbb{E}(X_n).$$

Plus généralement, si X_i est à valeurs dans $E_i \in \mathcal{B}(\mathbb{R})$ et si chaque application $f_i : E_i \rightarrow \mathbb{R}$ est mesurable, et si $f_i(X_i)$ est intégrable, alors le produit $f_1(X_1) \dots f_n(X_n)$ est intégrable et

$$\mathbb{E}(f_1(X_1) \dots f_n(X_n)) = \mathbb{E}(f_1(X_1)) \times \dots \times \mathbb{E}(f_n(X_n)).$$

Démonstration. Admise en cours, on s'est contenté d'indiquer les grandes lignes.

Compte-tenu de la proposition 4.3.11, il suffit de montrer le premier résultat. Regardons la preuve dans le cas où $n = 2$. On va considérer des situations de plus en plus complexes.

- Si $X_1 = \mathbf{1}_A$ et $X_2 = \mathbf{1}_B$.
- Si X_1 et X_2 sont des variables aléatoires discrètes finies, donc des fonctions étagées.
- Dans le cas où X_1 et X_2 sont mesurables positives et intégrables.
- Dans le cas où X_1 et X_2 sont intégrables.

□

Remarque 4.3.13. On peut généraliser au cas d'une variable possédant une espérance $\mathbb{E}(X_i) \in \mathbb{R} \cup \{+\infty\}$. Par exemple, si $\mathbb{E}(X_1) = +\infty$ et si $\mathbb{E}(X_2) > 0$ alors l'indépendance entraîne que $\mathbb{E}(X_1 X_2) = +\infty$.

Corollaire 4.3.14. Si X et Y sont des v.a.r. indépendantes alors $\Phi_{X+Y} = \Phi_X \Phi_Y$. Si de plus elles sont de carré intégrables alors $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

Démonstration. Après avoir vérifié les hypothèses, on peut écrire :

$$\Phi_{X+Y}(u) = \mathbb{E}(e^{iu(X+Y)}) = \mathbb{E}(e^{iuX} e^{iuY}) = \mathbb{E}(e^{iuX})(e^{iuY}) = \Phi_X(u)\Phi_Y(u)$$

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}((X + Y)^2) - (\mathbb{E}(X + Y))^2 \\ &= \mathbb{E}(X^2 + Y^2 + 2XY) - \mathbb{E}(X)^2 - \mathbb{E}(Y)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) \\ &= \mathbb{E}(X^2) + \mathbb{E}(Y^2) + 2\mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)^2 - \mathbb{E}(Y)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) \\ &= \mathbb{E}(X^2) + \mathbb{E}(Y^2) - \mathbb{E}(X)^2 - \mathbb{E}(Y)^2 \\ &= \text{Var}(X) + \text{Var}(Y). \end{aligned}$$

□

4.3.3 Loi jointes, lois marginales

On énonce dans le cas de $n = 2$ un résultat qui se généralise.

Théorème 4.3.15. Soient X_1, X_2 deux v.a.r.

- (i) X et Y sont indépendants si et seulement si $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$.

- (ii) X et Y sont indépendants si et seulement $F_{(X,Y)}(x,y) = F_X(x)F_Y(y)$, $\forall x,y \in \mathbb{R}$.
- (iii) Dans le cas où X (respectivement Y) a la densité p_X (respectivement p_Y) alors si X et Y sont indépendantes alors (X, X) a pour densité

$$(x, y) \mapsto f_{(X,Y)}(x, y) = f_X(x)f_Y(y).$$

- (iv) Si X et Y sont telles que (X, Y) a une densité de la forme

$$(x, y) \mapsto f_{(X,Y)}(x, y) = g_X(x)g_Y(y)$$

alors il existe une constante C_X (respectivement C_Y), telle que X a pour densité $C_X g_X$ (respectivement $C_Y g_Y$). De plus X et Y sont indépendantes.

Remarque 4.3.16. Quand on se trouve dans le cas (iv) du th. ci-dessus, on détermine la constantes C_X à l'aide de la propriété : $\int_{\mathbb{R}} f_X(x)d\lambda(x) = 1$ (cf. rem 2.4.13). Ce qui donne

$$C_X = \frac{1}{\int_{\mathbb{R}} g_X(x)d\lambda(x)}.$$

Somme de deux variables indépendantes

Exercice 3. Si f et g sont deux densités de probabilité sur \mathbb{R} , alors la convolée $f \star g$ est une mesure de densité. De plus $g \star f = f \star g$.

Proposition 4.3.17. Soient X et Y deux variables indépendantes à valeurs dans \mathbb{R} qui ont des densités respectivement p_X et p_Y , alors $X + Y$ a pour densité $p_X \star p_Y$.

4.3.4 Exemples

Remarque 4.3.18. Si X et Y sont deux variables aléatoires dont on connaît la loi $\mathbb{P}_X, \mathbb{P}_Y$ et si les variables sont indépendantes alors la loi jointe $\mathbb{P}_{(X,Y)}$ est connue car $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$. De plus, les loi marginales de $\mathbb{P}_X \otimes \mathbb{P}_Y$ sont \mathbb{P}_X et \mathbb{P}_Y .

Par contre, si il n'y a pas d'indépendances alors connaître la loi $\mathbb{P}_X, \mathbb{P}_Y$ ne suffit pas à déterminer la loi jointe $\mathbb{P}_{(X,Y)}$.

Exemple 4.3.19. On considère les ensembles A et B puis un couple de variables aléatoires (X_1, X_2) (respectivement (Y_1, Y_2)) qui suit une lois dont la vaut $f_{(X_1, X_2)} = \mathbf{1}_A$, (respectivement $f_{(Y_1, Y_2)} = \mathbf{1}_B$).

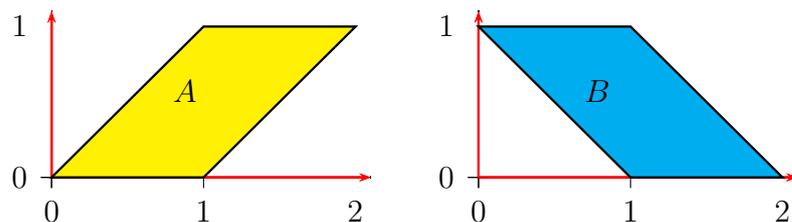


FIGURE 4.1 – Dessins de l'exemple 4.3.19

On vérifie que X_1 et Y_1 ont même loi (donné par la densité $x\mathbf{1}_{]0,1[} + (2-x)\mathbf{1}_{]1,2[}$ (respectivement X_2 et Y_2 ont même loi (loi uniforme sur $]0, 1[$) tandis que la loi du couple (X_1, X_2) diffère de celle de (Y_1, Y_2) .

On peut calculer par exemple (attention $F'_{X_1}(0)$ n'existe pas).

$$F_{X_1}(t) = \begin{cases} 0 & \text{si } t \leq 0 \\ x^2/2 & \text{si } t \in [0, 1] \\ 2x - (x^2/2) - 1 & \text{si } t \in [1, 2] \\ 1 & \text{si } t \geq 2 \end{cases} \quad F'_{X_1}(t) = \begin{cases} 0 & \text{si } t < 0 \\ x & \text{si } t \in]0, 1[\\ 2 - x & \text{si } t \in]1, 2[\\ 0 & \text{si } t \geq 2 \end{cases}$$

Si on compare les covariances $Cov(X_1, X_2) = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)$ tandis que $Cov(Y_1, Y_2) = \mathbb{E}(Y_1 Y_2) - \mathbb{E}(Y_1)\mathbb{E}(Y_2)$. Donc, puisque les lois marginales sont égales, on trouve

$$Cov(X_1, X_2) - Cov(Y_1, Y_2) = \mathbb{E}(X_1 X_2) - \mathbb{E}(Y_1 Y_2)$$

$$\begin{aligned}
 \mathbb{E}(X_1 X_2) &= \int_0^1 \int_0^2 xy \mathbf{1}_A(x, y) dy dx \\
 &= \int_0^1 x \int_0^1 y \mathbf{1}_A(x, y) dy dx + \int_1^2 x \int_0^1 y \mathbf{1}_A(x, y) dy dx \\
 &= \int_0^1 x \int_0^x y dy dx + \int_1^2 x \int_{x-1}^1 y dy dx \\
 &= \frac{1}{2} \int_0^1 x^3 dx + \frac{1}{2} \int_1^2 x(1 - (x-1)^2) dx \\
 &= \frac{1}{8} [x^4]_0^1 + \frac{1}{2} \int_1^2 x(2x - x^2) dx \\
 &= \frac{1}{8} + \frac{1}{2} \left[2 \frac{x^3}{3} - \frac{x^4}{4} \right]_1^2 \\
 &= \frac{1}{8} + \frac{1}{2} \left(\left(2 \frac{2^3}{3} - \frac{2^4}{4} \right) - \left(2 \frac{1^3}{3} - \frac{1^4}{4} \right) \right) \\
 &= \frac{1}{8} + \frac{1}{2} \left(\frac{8 \times 8 - 3 \times 16}{12} - \frac{2 \times 4 - 3}{12} \right) \\
 &= \frac{1}{8} + \frac{1}{2} \left(\frac{16 - 5}{12} \right) \\
 &= \frac{1}{8} + \frac{1}{2} \left(\frac{11}{12} \right) \\
 &= \frac{3}{24} + \frac{11}{24} = \frac{14}{24} = \frac{7}{12}
 \end{aligned}$$

De même,

$$\begin{aligned}
\mathbb{E}(Y_1 Y_2) &= \int_0^1 \int_0^2 xy \mathbf{1}_B(x, y) dy dx \\
&= \int_0^1 x \int_0^1 y \mathbf{1}_B(x, y) dy dx + \int_1^2 x \int_0^1 y \mathbf{1}_B(x, y) dy dx \\
&= \int_0^1 x \int_{1-x}^1 y dy dx + \int_1^2 x \int_0^{2-x} y dy dx \\
&= \frac{1}{2} \int_0^1 x(1 - (1-x)^2) dx + \frac{1}{2} \int_1^2 x(2-x)^2 dx \\
&= \frac{1}{2} \int_0^1 2x^2 - x^3 dx + \frac{1}{2} \int_1^2 4x - 4x^2 + x^3 dx \\
&= \frac{1}{2} \left[2 \frac{x^3}{3} - \frac{x^4}{4} \right]_0^1 + \frac{1}{2} \left[4 \frac{x^2}{2} - 4 \frac{x^3}{3} + \frac{x^4}{4} \right]_1^2 \\
&= \frac{1}{2} \left(\frac{2}{3} - \frac{1}{4} \right) + \frac{1}{2} \left(\left(4 \frac{2^2}{2} - 4 \frac{2^3}{3} + \frac{2^4}{4} \right) - \left(4 \frac{1^2}{2} - 4 \frac{1^3}{3} + \frac{1^4}{4} \right) \right) \\
&= \frac{10}{24} + \frac{1}{2} \left(\left(8 - \frac{32}{3} + 4 \right) - \left(2 - \frac{4}{3} + \frac{1}{4} \right) \right) \\
&= \frac{10}{24} + \frac{1}{2} \left(10 - \frac{28}{3} - \frac{1}{4} \right) \\
&= \frac{10}{24} + \frac{120 - 4 \times 28 - 3}{24} \\
&= \frac{15}{24} = \frac{5}{8}
\end{aligned}$$

Exemple 4.3.20. On considère un couple de variables aléatoires (X, Y) discrètes définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs (pour simplifier) dans $\mathbb{N} \times \mathbb{N}$. On pose $p_{i,j} = \mathbb{P}(X = i, Y = j)$. La donnée de la loi du couple se résume à la donnée de la famille $(p_{i,j})_{(i,j) \in \mathbb{N}^2}$ telle que $p_{i,j} \geq 0$ et $\sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} p_{i,j} = 1$.

Les lois marginales sont des loi discrètes. Pour écrire la loi de X , on note $\mathbb{P}(X = n) := p_{n,\bullet} = \sum_{j \in \mathbb{N}} p_{n,j}$ tandis que $\mathbb{P}(Y = m) := p_{\bullet,m} = \sum_{i \in \mathbb{N}} p_{i,m}$.

Les variables X et Y sont indépendantes si et seulement si pour tout i et pour tout j

$$p_{i,j} = \mathbb{P}(X = i, Y = j) = \mathbb{P}(X = i) \mathbb{P}(Y = j) = p_{i,\bullet} p_{\bullet,j}. \quad (4.3.1)$$

Dans le cas fini où le couple de variables aléatoires (X, Y) définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ est à valeurs dans $\{1, \dots, n\} \times \{1, \dots, m\}$, on peut résumer la loi du couple par la matrice

$$A := \begin{pmatrix} p_{1,1} & \dots & p_{1,m} \\ \vdots & \vdots & \vdots \\ p_{n,1} & \dots & p_{n,m} \end{pmatrix}$$

On montre sans peine que les variables X et Y sont indépendantes si et seulement si la matrice A est de rang 1. Dans ce cas, on a

$$A := \begin{pmatrix} p_{1,1} & \dots & p_{1,m} \\ \vdots & \vdots & \vdots \\ p_{n,1} & \dots & p_{n,m} \end{pmatrix} = \begin{pmatrix} p_{1,\bullet} \\ \vdots \\ p_{n,\bullet} \end{pmatrix} \begin{pmatrix} p_{\bullet,1} & \dots & p_{\bullet,m} \end{pmatrix}$$

Le produit matriciel est compatible et résume la relation (4.3.1).

Exemple 4.3.21. Soit $U_1 \sim \mathcal{E}(1)$ et $U_2 \sim \mathcal{U}([0, 2\pi])$. Les variables U_1 et U_2 sont supposées indépendantes. Soient $X = \sqrt{U_1} \cos(U_2)$, $Y = \sqrt{U_1} \sin(U_2)$.

Comme page 38, on va utiliser une fonction Γ continue bornée sur \mathbb{R}^2 pour déterminer la loi du couple. Calculons à l'aide du théorème de transfert

$$\begin{aligned}\mathbb{E}(\Gamma(X, Y)) &= \mathbb{E}(\Gamma(\sqrt{U_1} \cos(U_2), \sqrt{U_1} \sin(U_2))) \\ &= \int_{\mathbb{R}^2} \Gamma(\sqrt{u_1} \cos(u_2), \sqrt{u_1} \sin(u_2)) \frac{1}{2\pi} \mathbf{1}_{[0, 2\pi[} e^{-u_1} d\lambda_2(u) \\ &= \int_{\mathbb{R}_+ \times [0, 2\pi[} \Gamma(\sqrt{u_1} \cos(u_2), \sqrt{u_1} \sin(u_2)) \frac{1}{2\pi} e^{-u_1} d\lambda_2(u) \\ &= \int_{]0, +\infty[\times]0, 2\pi[} \Gamma(\sqrt{u_1} \cos(u_2), \sqrt{u_1} \sin(u_2)) \frac{1}{2\pi} e^{-u_1} d\lambda_2(u)\end{aligned}$$

Cela ressemble à un changement de variable, on rappelle la formule du changement de variable. Si $\phi : U \rightarrow V$ un difféomorphisme.

$$\int_V f(x) d\lambda_n(x) = \int_U (f \circ \phi)(u) \times |\det(J_\phi(u))| d\lambda_n(u)$$

Changement de variable :

$$\begin{cases} x = \sqrt{u} \cos(v) \\ y = \sqrt{u} \sin(v) \end{cases} \Leftrightarrow \begin{cases} u = x^2 + y^2 \\ v = \arctan(y/x) \end{cases}$$

Difféomorphisme : en posant les ouverts $U =]0, +\infty[\times]0, 2\pi[$ et $V = \mathbb{R}^2 \setminus \{0\}$

$$\begin{aligned}\phi :]0, +\infty[\times]0, 2\pi[&\rightarrow \mathbb{R}^2 \setminus \{0\} \\ (u, v) &\mapsto (\sqrt{u} \cos(v), \sqrt{u} \sin(v)).\end{aligned}$$

Déterminant de la matrice jacobienne de ϕ :

$$\begin{vmatrix} \frac{1}{2\sqrt{u}} \cos v & -\sqrt{u} \sin v \\ \frac{1}{2\sqrt{u}} \sin v & \sqrt{u} \cos v \end{vmatrix} = \frac{1}{2}$$

Donc avec un abus de notation $|\det(J_\phi(u))| d\lambda_2(u) = d\lambda_2(x)$, donc ici $(1/2) d\lambda_2(u) = d\lambda_2(x)$, que l'on va réécrire $d\lambda_2(u) = 2 d\lambda_2(x)$.

Et le terme à intégrer $\Gamma(\sqrt{u_1} \cos(u_2), \sqrt{u_1} \sin(u_2)) \frac{1}{2\pi} e^{-u_1}$ peut s'interpréter $\Gamma(x, y) \frac{1}{2\pi} e^{-(x^2+y^2)}$. Il faut en plus modifier le domaine d'intégration.

$$\mathbb{E}(\Gamma(X, Y)) = \int_U \Gamma(\sqrt{u} \cos(v), \sqrt{u} \sin(v)) \frac{1}{2\pi} e^{-u} d\lambda_2(u, v).$$

Puis par changement de variables :

$$\begin{aligned}\mathbb{E}(\Gamma(X, Y)) &= \int_V \phi(x, y) e^{-x^2-y^2} \frac{1}{2\pi} 2 d\lambda_2(x, y). \\ &= \int_{\mathbb{R}^2} \phi(x, y) e^{-x^2-y^2} \frac{1}{2\pi} 2 d\lambda_2(x, y).\end{aligned}$$

Donc le couple (X, Y) possède une densité et la densité de (X, Y) est $(x, y) \mapsto \frac{1}{\pi} e^{-x^2} e^{-y^2}$ (par (4.2.1), on peut vérifier que c'est bien une fonction dont l'intégrale sur \mathbb{R}^2 vaut 1). La densité du couple est le produit d'une fonction de x et d'une fonction de y donc les variables X et Y sont indépendantes.