

Intégration et probabilités

César Almecija
étudiant à l'école des MINES ParisTech
cesar.almecija@mines-paristech.fr

9 août 2021

Table des matières

I	Théorie de la mesure et intégration	4
1	Introduction à la théorie de la mesure	5
1.1	Ensembles mesurables	5
1.2	Définition de la mesure	7
1.3	Exemples importants	8
1.4	Fonctions mesurables	9
1.5	Approximation des fonctions mesurables	11
2	Construction de l'intégrale de Lebesgue	12
2.1	Pourquoi définir une nouvelle intégrale ?	12
2.2	Définition de l'intégrale de Lebesgue pour des fonctions positives	13
2.3	Propriétés de l'intégrale positive	14
2.4	Définition de l'intégrale de Lebesgue pour des fonctions signées .	14
2.5	Propriétés de l'intégrale	15
2.5.1	Le théorème de convergence dominée	15
2.5.2	Les théorèmes de Fubini-Tonelli	15
2.6	Démontrer en pratique des résultats sur les intégrales	17
2.7	Mesures définies par une intégrale	18
2.8	Généralisation aux fonctions à valeurs dans \mathbb{R}^m	19
3	Cas particulier : l'intégrale de Lebesgue munie de la mesure de Lebesgue	20
3.1	Que deviennent les résultats généraux dans ce cas particulier ? .	20
3.2	L'intégrale indéterminée	20
4	Cas particulier : l'intégrale de Lebesgue munie de la mesure de comptage	21
4.1	Lien avec les séries	21
4.2	Que deviennent les propriétés de l'intégrale dans ce cas particulier ?	23
II	Probabilités	24
5	Généralités sur les probabilités	25

5.1	Premières définitions	25
5.2	Indépendance et conditionnement	25
5.2.1	Trois formules utiles	27
5.3	Fonction de répartition	27
5.4	Deux cas particuliers de mesures de probabilités	28
5.4.1	Mesures de probabilité à densité	28
5.4.2	Mesure de probabilité discrètes	29
6	Variables aléatoires	31
6.1	Définitions	31
6.2	Loi d'une variable aléatoire	32
6.2.1	Loi jointe et lois marginales	34
6.3	Variables aléatoires indépendantes	35
6.3.1	Définitions	35
6.3.2	Deux résultats importants	36
6.4	Moments d'une variable aléatoire réelle ou discrète	37
6.4.1	Formule de transfert	38
6.4.2	Indépendance et espérance	40
6.4.3	Variance et covariance	41
6.5	Moments d'un vecteur aléatoire	42
6.6	Conditionnement par une variable aléatoire	43
6.6.1	Lois conditionnelles	43
6.6.2	Espérance conditionnelle	44
7	Convergence des suites de variables aléatoires	47
7.1	Les différents modes de convergence	47
7.1.1	La convergence en loi	47
7.1.2	La convergence en probabilité	48
7.1.3	La convergence en norme \mathcal{L}^r	49
7.1.4	La convergence presque-sûre	49

Avant-propos

Ce document vise à résumer les notions essentielles d'intégration et de probabilités. Son objectif est de retracer les cheminements de pensée qui mènent aux résultats essentiels, afin de mieux les retenir. Ainsi, il pourra être avantageusement utilisé pour se refamiliariser avec ces notions, au cas-où elles auraient été oubliées. Ce travail peut également servir de préambule à une étude plus approfondie du sujet, pour découvrir les fondements de ce domaine des mathématiques.

Néanmoins, ce document ne saurait se substituer à un ouvrage de référence. En effet, par soucis de clarté et de concision, il ne détaille pas toutes les démonstrations. Celles-ci sont néanmoins nécessaires pour saisir l'entièreté des résultats énoncés ici. De cette manière, **il faut le lire comme un aide-mémoire ou un résumé, et non comme un cours.**

Ce travail se partage en deux grandes parties :

- Une première partie commençant par une introduction à la théorie de la mesure, suivie de la construction rapide de l'intégrale de Lebesgue. Elle explore ensuite les cas particuliers de l'intégrale :
 - munie de la mesure de Lebesgue (l'intégrale *classique*) ;
 - munie de la mesure de comptage (les séries).
- Une deuxième partie qui applique la partie précédente à la théorie des probabilités

Ce document s'est grandement inspiré du cours Calcul Différentiel, Intégral et Stochastique (CDIS) de l'Ecole des MINES ParisTech. L'auteur recommande aux lecteurs intéressés de s'y référer régulièrement.

Première partie

**Théorie de la mesure et
intégration**

Chapitre 1

Introduction à la théorie de la mesure

Pour comprendre cette notion, partons d'un constat évident. On sait mesurer la taille d'un segment : la manière de mesurer cette grandeur est d'utiliser la *longueur*. Par exemple, la longueur du segment $[0, 1]$ vaut 1. Plus généralement, λ , l'application qui à un segment $[a, b]$ associe sa longueur $b - a$, pourrait s'appeler dans le langage courant une *mesure*, c'est-à-dire une fonction permettant d'obtenir la mesure (la taille) d'un objet.

Se posent alors les questions suivantes :

- dans \mathbb{R} , comment mesurer n'importe quel ensemble ?
- plus généralement, comment mesurer dans \mathbb{R}^n ?
- mais est-il possible de *tout* mesurer ? Ou faut-il se restreindre à une catégorie d'ensembles *mesurables* ?

On commencera par répondre à cette dernière question.

1.1 Ensembles mesurables

Il n'est pas possible de tout mesurer ainsi. Il faut introduire une classe d'ensembles, appelée **tribu**, qui va représenter l'ensemble des ensembles **mesurables**.

Une tribu doit cependant garantir certaines propriétés de stabilité par opérations ensemblistes. En effet, si deux ensembles sont mesurables, on aimerait pouvoir donner un sens à la mesure de leur intersection ou de leur union. De même, si un ensemble est mesurable, il est légitime de demander que son complémentaire soit mesurable à son tour. Enfin, demander que l'ensemble vide soit mesurable est raisonnable, et on verra par la suite que la mesure du vide doit valoir 0, conformément à l'intuition.

Ces considérations étant faites, on définit alors une tribu comme suit :

Définition 1.1.1. Une **tribu** (ou σ -**algèbre**) \mathcal{A} d'un ensemble E est une col-

lection d'ensembles $\mathcal{A} \subset \mathcal{P}(E)$ vérifiant les hypothèses suivantes :

1. $\emptyset \in \mathcal{A}$
2. \mathcal{A} est stable par passage au complémentaire
3. \mathcal{A} est stable par union dénombrable

Définition 1.1.2. Un **espace mesurable** (E, \mathcal{A}) est un ensemble E muni d'une tribu \mathcal{A} .

Définition 1.1.3. Soit (E, \mathcal{A}) un espace mesurable. Un ensemble $X \in \mathcal{A}$ est dit **\mathcal{A} -mesurable**.

Remarque 1.1.1. Si le contexte est clair (*ie* il n'y a qu'une seule tribu), on pourra omettre \mathcal{A} et parler simplement d'ensemble mesurable

Remarque 1.1.2. On remarquera qu'il faut distinguer les *espaces* mesurables (définition 1.1.2) des *ensembles* mesurables (définition 1.1.3).

Définition 1.1.4. Soit E un ensemble, et B une collection d'ensembles de E . La **tribu engendrée par B** est la plus petite tribu sur E contenant B . De manière équivalente, c'est l'intersection de toutes les tribus de E contenant B .

Définition 1.1.5. Soit E un espace topologique. La **tribu de Borel** (ou plus simplement, les **boréliens**) est la tribu engendrée par les ouverts de E (ou de manière équivalente, par les fermés de E). Elle est notée $\mathcal{B}(E)$.

Exemple 1.1.1. Considérons \mathbb{N} muni de sa topologie usuelle, la topologie discrète. Par définition, dans la topologie discrète, l'ensemble des ouverts T est $\mathcal{P}(\mathbb{N})$.

Or, une tribu est par définition une partie de $\mathcal{P}(\mathbb{N})$. Ainsi, $\mathcal{P}(\mathbb{N}) = T \subset \mathcal{B}(\mathbb{N}) \subset \mathcal{P}(\mathbb{N})$. Par double inclusion, **les boréliens de \mathbb{N} sont exactement les parties de \mathbb{N}** :

$$\mathcal{B}(\mathbb{N}) = \mathcal{P}(\mathbb{N})$$

On peut généraliser ce résultat à tout espace topologique muni de la topologie discrète.

Soit dit en passant, à titre de petit exercice, on peut vérifier que toute partie d'un espace topologique muni de la topologie discrète est à la fois ouverte et fermée.

Définition 1.1.6. Soit (E, \mathcal{A}) et (F, \mathcal{B}) deux espaces mesurables. On appelle **tribu produit de \mathcal{A} et \mathcal{B}** la tribu sur $E \times F$ engendrée par :

$$\{A \times B, (A, B) \in \mathcal{A} \times \mathcal{B}\}$$

On la note $\mathcal{A} \otimes \mathcal{B}$.

On appelle **espace produit** des espaces mesurables (E, \mathcal{A}) et (F, \mathcal{B}) l'espace mesurable $(E \times F, \mathcal{A} \otimes \mathcal{B})$.

Exemple 1.1.2. On a $\mathcal{B}(\mathbb{R}^n) \otimes \mathcal{B}(\mathbb{R}^m) = \mathcal{B}(\mathbb{R}^{m+n})$.

Notons cependant que ce résultat n'est pas vrai en général. En effet, nous verrons dans la suite de ce chapitre la tribu de Lebesgue $\mathcal{L}(\mathbb{R}^n)$: le résultat est faux pour cette tribu.

Définition 1.1.7. Soit (E, \mathcal{A}) un espace mesurable. Une tribu \mathcal{B} sur E est une **sous-tribu de \mathcal{A}** lorsque $\mathcal{B} \subset \mathcal{A}$.

1.2 Définition de la mesure

Abordons désormais la notion de mesure. Un ensemble mesurable doit intuitivement admettre une mesure positive. De plus, si deux ensembles sont disjoints, il est légitime de demander que la mesure de leur union soit la somme des mesures.

Cela nous amène à définir une mesure ainsi :

Définition 1.2.1. Une **mesure** μ sur un espace mesurable (E, \mathcal{A}) est une application $\mu : \mathcal{A} \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ telle que :

1. $\mu(\emptyset) = 0$
2. Pour une famille au plus dénombrable d'ensembles mesurables $(A_i)_{i \in I}$,

$$\mu\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i)$$

Remarque 1.2.1. Cette deuxième propriété porte le nom de **σ -additivité**.

Remarque 1.2.2. Si $\mu(\emptyset) \neq 0$, le lecteur pourra vérifier que la mesure de n'importe-quel ensemble mesurable (*a fortiori* du vide) vaut $+\infty$. On comprend alors pourquoi on impose que la mesure du vide soit nulle : si ce n'était pas le cas, mesurer n'aurait pour ainsi dire aucun intérêt.

Définition 1.2.2. Un **espace mesuré** (E, \mathcal{A}, μ) est un espace mesurable (E, \mathcal{A}) muni d'une mesure μ .

Définition 1.2.3. Soit (E, \mathcal{A}, μ) un espace mesuré. Un ensemble $X \subset E$ est dit **μ -négligeable** si :

$$\exists Y \in \mathcal{A}, \begin{cases} \mu(Y) = 0 \\ X \subset Y \end{cases}$$

Remarque 1.2.3. On remarque qu'un ensemble négligeable n'est pas forcément mesurable. Il est cependant commode d'imposer qu'un ensemble négligeable soit nécessairement mesurable. Modifier une mesure et la tribu respective pour arriver à ce résultat porte le nom de **complétion d'une mesure**.

Ce processus ne sera pas détaillé ici, mais le lecteur peut néanmoins retenir qu'une mesure complétée vérifie l'équivalence suivante : **un ensemble est négligeable ssi il est mesurable, de mesure nulle**

Remarque 1.2.4. On notera que la notion d'ensemble négligeable dépend de la mesure choisie. Un ensemble négligeable pour une certaine mesure peut très bien ne pas l'être pour une autre mesure.

Définition 1.2.4. Soit (E, \mathcal{A}, μ) un espace mesuré. On dit que μ est une mesure **σ -finie** lorsqu'il existe une collections dénombrables d'ensembles mesurables $(X_n)_{n \in \mathbb{N}}$, tels que :

- $\bigcup_{n \in \mathbb{N}} X_n = E$ (on dit que les X_n recouvrent E) ;
- $\forall n \in \mathbb{N}, \mu(X_n) < +\infty$ (on dit que les X_n sont de mesure finie).

Définition 1.2.5. Soit (E, \mathcal{A}, μ_1) et (F, \mathcal{B}, μ_2) deux espaces mesurés. Une **mesure produit** $\mu_1 \otimes \mu_2$ sur l'espace produit $(E \times F, \mathcal{A} \otimes \mathcal{B})$ est une mesure cet espace produit vérifiant :

$$\forall (A, B) \in (\mathcal{A}, \mathcal{B}), (\mu_1 \otimes \mu_2)(A \times B) = \mu_1(A)\mu_2(B)$$

Une telle mesure existe toujours. Si μ_1 et μ_2 sont σ -finies, alors elle est unique.

1.3 Exemples importants

Il existe trois mesures importantes à connaître.

Définition 1.3.1. La **mesure de Lebesgue** λ est l'unique mesure sur \mathbb{R}^n qui prolonge la notion de volume. Habituellement, sa tribu de définition est la **tribu de Lebesgue** ou la **tribu de Borel**.

Remarque 1.3.1. En définissant la mesure de Lebesgue sur la tribu de Borel, elle ne serait en fait pas complète (voir 1.2.3). La tribu de Lebesgue (notée $\mathcal{L}(\mathbb{R}^n)$) est alors définie comme étant la plus petite tribu de \mathbb{R}^n permettant à cette mesure d'être complète.

On ne retiendra pas plus de détails sur la tribu de Lebesgue, mais il est bon de retenir qu'**un élément Borel-mesurable est Lebesgue-mesurable** (la réciproque est fausse).

Remarque 1.3.2. On remarque immédiatement que :

- lorsque $n = 1$, cette mesure prolonge la notion de longueur vue en introduction du chapitre.
- lorsque $n = 2$, cette mesure prolonge la notion de surface.
- lorsque $n = 3$, cette mesure prolonge la notion de volume.

Par ailleurs, cette mesure sera centrale dans le chapitre 3 concernant les intégrales.

Exemple 1.3.1 (\mathbb{Q} est λ -négligeable).

Définition 1.3.2. Soit (E, \mathcal{A}) un espace mesurable. La **mesure de comptage** c est définie sur cette espace comme suit :

$$c : \begin{cases} \mathcal{A} & \longrightarrow & \mathbb{R}_+ \cup \{+\infty\} \\ A & \longmapsto & \begin{cases} 0 & \text{si } A = \emptyset \\ n & \text{si } A \text{ est fini de cardinal } n \\ +\infty & \text{si } A \text{ est infini} \end{cases} \end{cases}$$

Remarque 1.3.3. Comme son nom l'indique, cette mesure compte les éléments présents dans l'ensemble que l'on mesure.

Remarque 1.3.4. Bien que cette mesure soit définie pour un espace mesurable quelconque, elle sera particulièrement utile dans $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ (notamment pour le chapitre 4 concernant les séries).

Définition 1.3.3. Soit (E, \mathcal{A}) un espace mesurable. La **mesure de Dirac en** $x \in E$ δ_x est définie comme suit :

$$\delta_x : \begin{cases} \mathcal{A} & \longrightarrow & \mathbb{R}_+ \\ A & \longmapsto & \begin{cases} 0 & \text{si } x \in A \\ 1 & \text{si } x \notin A \end{cases} \end{cases}$$

Remarque 1.3.5. La mesure de Dirac et la fonction indicatrice sont reliées de la manière suivante :

$$\forall A \in \mathcal{A}, \forall x \in E, \delta_x(A) = \mathbf{1}_A(x)$$

1.4 Fonctions mesurables

Lorsque deux espaces mesurables sont définis, il peut être utile de définir une classe de fonctions qui font bon ménage avec les tribus des espaces mis en jeu.

Définition 1.4.1. Soit (E, \mathcal{A}) et (F, \mathcal{B}) deux espaces mesurables. Une **fonction \mathcal{A}/\mathcal{B} -mesurable** est une fonction $f : E \longrightarrow F$ telle que :

$$\forall Y \in \mathcal{B}, f^{-1}(Y) \in \mathcal{A}$$

Remarque 1.4.1. Cette caractérisation des fonctions mesurables porte le nom de **critère de l'image réciproque**.

Remarque 1.4.2. Cette notation est lourde, et c'est pourquoi elle est souvent simplifiée.

- dans le cas où $\mathcal{B} = \mathcal{B}(F)$ (la tribu d'arrivée est celle des boréliens), on pourra simplifier la notation et dire simplement que :

f est \mathcal{A} -mesurable

- si, de plus, la tribu de l'espace de départ n'est pas ambiguë, on pourra alors simplement écrire que :

f est mesurable

Remarque 1.4.3. On remarquera que la notion de fonction mesurable est **indépendante des mesures choisies**. Elle ne dépend que des tribus de l'espace de départ et d'arrivée.

Voici un lien bien utile entre ensembles et fonctions mesurables :

Proposition 1.4.1. *Soit (E, \mathcal{A}) un espace mesurable. Soit $A \subset E$. Alors A est mesurable ssi 1_A est mesurable.*

Remarque 1.4.4. On remarquera que l'on a bien utilisé le formalisme de la remarque précédente :

- la tribu de l'espace d'arrivée est la tribu borélienne de \mathbb{R} .
- la tribu de l'espace de départ est \mathcal{A} , non-ambigüe.

La proposition suivante permet de simplifier considérablement la vérification de la mesurabilité d'une fonction, dans le cas où l'espace d'arrivée est muni de la tribu des boréliens.

Elle découle directement du fait que la tribu des boréliens soit la tribu engendrée par les ouverts : pour vérifier que le résultat est vrai pour tout élément de la tribu, il suffit de le vérifier sur les éléments qui engendrent la tribu (en l'occurrence, les ouverts).

Proposition 1.4.2. *Soit (E, \mathcal{A}) et $(F, \mathcal{B}(F))$ deux espaces mesurables. f est \mathcal{A} -mesurable ssi pour tout ouvert U de F , $f^{-1}(U) \in \mathcal{A}$.*

Remarque 1.4.5. En vertu de la définition 1.1.5, on peut remplacer dans la proposition 1.4.2 « ouvert » par « fermé ».

Le résultat suivant est un résultat important concernant la composition des fonctions mesurables :

Proposition 1.4.3. *Soit (E, \mathcal{A}) , (F, \mathcal{B}) et (G, \mathcal{C}) trois espaces mesurables, $f : E \rightarrow F$ une fonction \mathcal{A}/\mathcal{B} -mesurable et $g : F \rightarrow G$ une fonction \mathcal{B}/\mathcal{C} -mesurable. Alors $g \circ f : E \rightarrow G$ est une fonction \mathcal{A}/\mathcal{C} -mesurable*

Voici un résultat concernant les limites de fonctions mesurables :

Proposition 1.4.4. *Soit (E, \mathcal{A}, μ_1) et (F, \mathcal{B}, μ_2) deux espaces mesurés et (f_n) une suite de fonctions mesurables de E dans F qui converge simplement vers f . Alors f est mesurable.*

Enfin, voici une dernière définition qui sera utile dans le chapitre concernant les probabilités :

Définition 1.4.2. Une fonction $f : E \rightarrow F$ est dite **borélienne** lorsqu'elle est $\mathcal{B}(E)/\mathcal{B}(F)$ -mesurable.

Proposition 1.4.5. *Une fonction continue est borélienne.*

Remarque 1.4.6. Cette proposition est une conséquence directe de la caractérisation de la continuité d'une fonction par l'image réciproque des ouverts. Attention, la réciproque de ce résultat est fausse : il existe des fonctions boréliennes qui ne sont pas continues.

Définition 1.4.3. Soit (E, \mathcal{A}) un espace mesurable et f une fonction \mathcal{A} -mesurable à valeurs dans F . La **tribu engendrée par f** est une tribu sur E , définie par :

$$\{f^{-1}(B) | B \in \mathcal{B}(F)\}$$

C'est une sous-tribu de \mathcal{A} .

1.5 Approximation des fonctions mesurables

Les fonctions mesurables, en plus de faire bon ménage avec les ensembles mesurables, peuvent être approximées facilement par des fonctions dites « étagées ». Nous allons commencer par définir cette classe de fonctions, puis nous verrons en quoi consiste cette approximation.

Définition 1.5.1. Une **fonction étagée** est une fonction dont l'image est finie.

La proposition suivante fournit une décomposition pratique des fonctions étagées :

Proposition 1.5.1. Soit f une fonction étagée. Alors :

$$\exists N \in \mathbb{N}, \exists (A_i)_{i \in [1, N]} \subset \mathbb{R}^N, \exists (y_i)_{i \in [1, N]} \in \mathbb{R}^N, f = \sum_{i=1}^N y_i \mathbb{1}_{A_i}$$

Remarque 1.5.1. Lorsque la fonction étagée que l'on manipule est mesurable, il existe toujours une décomposition de la forme précédente où les $(A_i)_{i \in [1, N]}$ ont le bon goût d'être mesurables.

C'est une propriété essentielle des fonctions étagées mesurables, qui sera utile dans prochain chapitre.

Théorème 1.5.1. Toute fonction mesurable positive est limite simple d'une suite croissante de fonctions étagées mesurables positives.

Remarque 1.5.2. Ici, la croissance d'une suite de fonctions $(f_k)_{k \in \mathbb{N}}$ est à comprendre dans le sens suivant :

$$\forall k \in \mathbb{N}, \forall x \in E, f_k(x) \leq f_{k+1}(x)$$

Un corollaire existe pour les fonctions mesurables signées, mais il est assez peu utile. Le voici à titre culturel seulement :

Corollaire 1.5.1. Toute fonction mesurable est limite simple d'une suite absolument croissante de fonctions étagées mesurables.

Remarque 1.5.3. Ici, l'absolue croissance d'une suite de fonctions $(f_k)_{k \in \mathbb{N}}$ est à comprendre dans le sens suivant :

$$\forall k \in \mathbb{N}, \forall x \in E, |f_k(x)| \leq |f_{k+1}(x)|$$

Chapitre 2

Construction de l'intégrale de Lebesgue

Maintenant que nous connaissons les notions de base concernant la théorie de la mesure, nous sommes en mesure de définir l'intégrale de Lebesgue. Cette intégrale est en quelque sorte une généralisation de l'intégrale de Riemann.

2.1 Pourquoi définir une nouvelle intégrale ?

Deux raisons principales peuvent nous amener à définir une nouvelle intégrale.

D'une part, l'intégrale de Riemann ne permet de calculer l'intégrale que de fonctions **continues par morceaux**. Or, cela est très restrictif : par exemple, prenons la fonction $\mathbb{1}_{\mathbb{Q}}$. Elle n'est pas continue par morceaux, donc n'est pas intégrable au sens de Riemann. Pourtant, pour la mesure de Lebesgue, \mathbb{Q} est négligeable (voir l'exemple 1.3.1). Ainsi, $\mathbb{1}_{\mathbb{Q}}$ est Lebesgue-presque-partout égale à 0. On aurait donc envie de dire que « l'influence de \mathbb{Q} est si petite, qu'elle n'a aucun impact dans l'intégrale ». En disant cela, on veut en fait dire que l'intégrale de $\mathbb{1}_{\mathbb{Q}}$ serait égale à l'intégrale de la fonction nulle, qui vaut zéro. Ce faisant, **on pourrait donner un sens à l'intégrale d'une fonction qui n'est pas continue par morceaux !**

Plus généralement, l'intégrale de Lebesgue permettra de définir l'intégrale d'une fonction mesurable, et non plus d'une fonction continue par morceaux.

D'autre part, l'intégrale de Riemann utilise la notion de longueur : par exemple, si f est la fonction constante égale à 1, nous avons

$$\int_0^1 f(x)dx = 1 > 2 = \int_0^2 f(x)dx$$

Mais il est impossible de « mesurer différemment ». En effet, si une mesure μ vérifie $\mu([1, 2]) = 0$, alors on aimerait pouvoir dire que les deux intégrales ci-dessus, calculées avec μ , seraient égales. Plus généralement, **on aimerait pou-**

voir intégrer dans n'importe quel espace par n'importe quelle mesure. L'intégrale de Lebesgue répondra à ce besoin.

2.2 Définition de l'intégrale de Lebesgue pour des fonctions positives

Les fonctions positives ont une place privilégiée dans la théorie de l'intégration de Lebesgue. En effet, leurs intégrales sont toujours définies (quitte à valoir $+\infty$) :

Définition 2.2.1. Soit (E, \mathcal{A}, μ) un espace mesuré et $f : E \rightarrow \mathbb{R}_+$ mesurable.

La μ intégrale de Lebesgue $\int_E f(x) d\mu(x)$ existe toujours et est l'unique élément de $\mathbb{R}_+ \cup \{+\infty\}$ vérifiant :

- **la relation mesure-intégrale** : pour tout ensemble X \mathcal{A} -mesurable,

$$\int_E \mathbf{1}_X(x) d\mu(x) = \mu(X)$$

- **l'hypothèse de linéarité** : pour toute fonction mesurable $g : E \rightarrow \mathbb{R}_+$ et pour tous $(\alpha, \beta) \in (\mathbb{R}_+^*)^2$,

$$\int_E (\alpha f + \beta g)(x) d\mu(x) = \alpha \int_E f(x) d\mu(x) + \beta \int_E g(x) d\mu(x)$$

- **l'hypothèse de convergence monotone** : pour toute suite croissante de fonctions mesurables positives $f_n : E \rightarrow \mathbb{R}_+$ convergeant simplement vers f ,

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

Remarque 2.2.1. L'intégrale d'une fonction mesurable positive vérifie immédiatement, par définition, ces trois hypothèses fondamentales, qui seront centrales dans la suite de ce document. Attention cependant : les deux hypothèses de mesurabilité et de positivité sont absolument nécessaires.

Définition 2.2.2. Soit (E, \mathcal{A}, μ) un espace mesuré et $f : E \rightarrow \mathbb{R}_+$ mesurable.

f est dite **μ -intégrable** lorsque $\int_E f(x) d\mu(x) < +\infty$.

Remarque 2.2.2. On fera donc bien attention au sens précis du mot « intégrable ». Il est plus fort de dire qu'une fonction est intégrable, que de dire que l'intégrale d'une fonction existe. Par exemple, une fonction mesurable positive admet toujours une intégrale, mais elle n'est pas forcément intégrable.

Définition 2.2.3. Soit $f : E \rightarrow \mathbb{R}$, et $F \subset E$. On dit que f est **intégrable sur le sous ensemble F** lorsque $f\mathbf{1}_F$ est intégrable.

Remarque 2.2.3. Pour avoir une chance que f soit intégrable sur F , il est nécessaire que $f\mathbf{1}_F$ soit mesurable.

2.3 Propriétés de l'intégrale positive

2.4 Définition de l'intégrale de Lebesgue pour des fonctions signées

Commençons par des résultats préliminaires qui seront utiles par la suite :

Définition 2.4.1. Soit E un ensemble. Soit $f : E \rightarrow \mathbb{R}$. Les **partie positive** et **partie négative** de f , respectivement f_+ et f_- , sont définies comme suit :

$$\begin{cases} f_+ &:= \max(0, f) \\ f_- &:= \min(0, f) \end{cases}$$

Proposition 2.4.1. Soit E un ensemble et $f : E \rightarrow \mathbb{R}$. Alors :

$$f = f_+ - f_-$$

Le résultat suivant est une conséquence directe de la proposition 1.4.3 :

Proposition 2.4.2. Soit (E, \mathcal{A}) un espace mesurable et $f : E \rightarrow \mathbb{R}$ une fonction mesurable. Alors f_+ et f_- sont mesurables

Désormais, passons à la définition de l'intégrale d'une fonction signée. Celle-ci arrive volontairement dans une nouvelle section, pour souligner la différence avec l'intégrale d'une fonction positive. En effet, comme cela a été souligné au début de la section 2.2, les fonctions positives occupent une place privilégiée dans la théorie de l'intégration de Lebesgue : elles vérifient des résultats qui ne sont pas systématiquement vrais pour les fonctions signées.

La définition de l'intégrale pour une fonction signée est un premier exemple soulignant cette différence :

Définition 2.4.2. Soit (E, \mathcal{A}, μ) un espace mesuré et $f : E \rightarrow \mathbb{R}$ mesurable.

La μ intégrale de Lebesgue $\int_E f(x) d\mu(x)$ existe dès que f_+ ou f_- est μ -intégrable. Dans ce cas, on a :

$$\int_E f(x) d\mu(x) = \int_E f_+ d\mu(x) - \int_E f_- d\mu(x)$$

Remarque 2.4.1. Contrairement au cas positif, **l'intégrale d'une fonction signée n'existe pas systématiquement !**

Définition 2.4.3. Soit (E, \mathcal{A}, μ) un espace mesuré et $f : E \rightarrow \mathbb{R}$ mesurable. f est dite **μ -intégrable** lorsque f_+ et f_- sont μ -intégrables.

Remarque 2.4.2. On notera que, de manière équivalente, **une fonction signée est intégrable ssi l'intégrale de f existe, et est finie.**

Remarque 2.4.3. Comme dans le cas positif (remarque 2.2.2), le mot « intégrable » a un sens bien précis !

Théorème 2.4.1. Soit (E, \mathcal{A}, μ) un espace mesuré et $f : E \rightarrow \mathbb{R}$ mesurable. f est μ -intégrable ssi $|f|$ est μ -intégrable. On dit que **l'intégrale de Lebesgue est absolue**.

Remarque 2.4.4. Pour vérifier qu'une fonction signée est intégrable, on vérifiera en pratique que sa valeur absolue est intégrable.

Le théorème suivant assure que cette nouvelle intégrale vérifie bien les besoins que l'on avait exprimé en début de chapitre.

Théorème 2.4.2. L'intégrale de Lebesgue prolonge l'intégrale de Riemann.

2.5 Propriétés de l'intégrale

Nous énonçons ici quelques résultats importants concernant les intégrales, que les fonctions soient positives ou non.

2.5.1 Le théorème de convergence dominée

Théorème 2.5.1 (de convergence dominée). Soit (E, \mathcal{A}, μ) un espace mesuré. Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables de E dans \mathbb{R} , qui converge simplement vers une fonction f (qui est donc mesurable par la proposition 1.4.4).

Supposons qu'il existe $g : E \rightarrow \mathbb{R}$ μ -intégrable qui **domine** les f_n :

$$\forall n \in \mathbb{N}, \forall x \in E, |f_n(x)| \leq g(x)$$

Alors les f_n et f sont μ -intégrables, et :

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

Remarque 2.5.1. Ce théorème permet donc d'échanger limite et intégrale (ce qui n'est pas vrai dans le cas général).

2.5.2 Les théorèmes de Fubini-Tonelli

Les théorèmes suivants sont connus sous le nom des **théorèmes de Fubini-Tonelli**. Il s'agit en fait de deux théorèmes *a priori* totalement indépendants, mais qui sont souvent utilisés ensemble.

En effet, si l'on se donne (E, \mathcal{A}, μ_1) et (F, \mathcal{B}, μ_2) deux espaces mesurés, et $f : E \times F \rightarrow \mathbb{R}$, on peut résumer grossièrement ces deux théorèmes ainsi :

- le théorème de Fubini permet de calculer l'intégrale de f en calculant d'abord l'intégrale par rapport à E , puis par rapport à F , sous réserve d'intégrabilité ;
- le théorème de Tonelli permet de prouver cette intégrabilité, en calculant d'abord l'intégrale de $|f|$ par rapport à E , puis par rapport à F .

Théorème 2.5.2 (de Fubini pour des fonctions positives). Soit (X, \mathcal{A}, μ_1) et (Y, \mathcal{B}, μ_2) deux espaces mesurés, tels que μ_1 et μ_2 soient σ -finies. On considère l'espace produit $X \times Y$ muni de la mesure produit. Soit $f : X \times Y \rightarrow \mathbb{R}_+$ $\mathcal{A} \otimes \mathcal{B}$ -mesurable. Alors :

- $x \mapsto \int_Y f(x, y) d\mu_2(y)$ existe pour tout $x \in X$ et est \mathcal{A} -mesurable.
- $y \mapsto \int_X f(x, y) d\mu_1(x)$ existe pour tout $y \in Y$ et est \mathcal{B} -mesurable.

De plus, on a :

$$\begin{aligned} \int_{X \times Y} f(x, y) d(\mu_1 \otimes \mu_2)(x, y) &= \int_X \left(\int_Y f(x, y) d\mu_2(y) \right) d\mu_1(x) \\ &= \int_Y \left(\int_X f(x, y) d\mu_1(x) \right) d\mu_2(y) \end{aligned}$$

Remarque 2.5.2. Lorsque la fonction est positive, on peut donc **toujours** calculer l'intégrale variables par variables (quitte à ce que le résultat vaille $+\infty$). Attention, cela n'est pas vrai forcément vrai si f n'est pas positive, comme nous le verrons ci-après.

Voici l'énoncé du théorème dans le cas général :

Théorème 2.5.3 (de Fubini). Soit (X, \mathcal{A}, μ_1) et (Y, \mathcal{B}, μ_2) deux espaces mesurés, tels que μ_1 et μ_2 soient σ -finies. On considère l'espace produit $X \times Y$ muni de la mesure produit. Soit $f : X \times Y \rightarrow \mathbb{R}$ $\mu_1 \otimes \mu_2$ -intégrable. Alors :

- pour μ_1 -presque tout x , $y \mapsto f(x, y)$ est μ_2 -intégrable, et la fonction définie μ_1 -presque partout $x \mapsto \int_Y f(x, y) d\mu_2(y)$ est intégrable.
- pour μ_2 -presque tout y , $x \mapsto f(x, y)$ est μ_1 -intégrable, et la fonction définie μ_2 -presque partout $y \mapsto \int_X f(x, y) d\mu_1(x)$ est intégrable.

De plus, on a :

$$\begin{aligned} \int_{X \times Y} f(x, y) d(\mu_1 \otimes \mu_2)(x, y) &= \int_X \left(\int_Y f(x, y) d\mu_2(y) \right) d\mu_1(x) \\ &= \int_Y \left(\int_X f(x, y) d\mu_1(x) \right) d\mu_2(y) \end{aligned}$$

Remarque 2.5.3. Lorsque la fonction est signée, le théorème n'est vrai que si la fonction est intégrable pour la mesure produit ! De plus, le résultat est un peu plus faible : on ne peut définir les intégrales partielles que presque partout. En pratique, ce n'est pas un problème : il suffit par exemple de faire valoir 0 aux endroits où une telle fonction n'est pas définie (comme c'est uniquement sur un ensemble négligeable, le choix de la valeur n'a pas d'importance car il ne change pas la valeur de l'intégrale).

Voici le théorème de Tonelli, qui permet de prouver l'intégrabilité de telles fonctions :

Théorème 2.5.4 (de Tonelli). Soit (X, \mathcal{A}, μ_1) et (Y, \mathcal{B}, μ_2) deux espaces mesurés, tels que μ_1 et μ_2 soient σ -finies. On considère l'espace produit $X \times Y$ muni de la mesure produit. Soit $f : X \times Y \rightarrow \mathbb{R}_+$ $\mathcal{A} \otimes \mathcal{B}$ -mesurable.

Pour μ_2 -presque tout y , $x \mapsto f(x, y)$ est \mathcal{A} -mesurable.

Puis f est $\mu_1 \otimes \mu_2$ -**intégrable** ssi $x \mapsto f(x, y)$ est μ_1 -intégrable pour μ_2 -presque tout $y \in Y$, et que la fonction alors définie pour μ_2 -presque tout $y \in Y$ (qui est alors \mathcal{B} -mesurable) est μ_2 -intégrable.

Remarque 2.5.4. Il est évident que l'on peut échanger x et y , \mathcal{A} et \mathcal{B} , X et Y et μ_1 et μ_2 dans la dernière phrase du théorème.

Remarque 2.5.5. On remarque que, dans le théorème de Tonelli, la fonction doit être positive. Qu'à cela ne tienne, par le caractère absolu de l'intégrale de Lebesgue, il suffit d'appliquer le théorème à $|f|$ pour avoir l'intégrabilité de $|f|$, ie celle de f .

Remarque 2.5.6. A présent, voyons comment bien utiliser ces deux théorèmes pour intégrer variables par variables. On se donne (X, \mathcal{A}, μ_1) et (Y, \mathcal{B}, μ_2) deux espaces mesurés, tels que μ_1 et μ_2 soient σ -finies.

Soit $f : X \times Y \rightarrow \mathbb{R}$.

- si f est positive, on applique le **théorème de Fubini, cas positif** 2.5.2 : on peut toujours intégrer variable par variable, que f soit intégrable ou non.
- si f n'est pas positive, on doit commencer par prouver l'intégrabilité de f . Pour cela, on applique le **théorème de Tonelli** 2.5.4 à $|f|$, qui nous fait calculer l'intégrale de $|f|$ par rapport à une variable, puis l'intégrale de la fonction obtenue par rapport à l'autre variable. Si le résultat obtenu existe et est fini, alors on peut conclure par ce théorème que f est intégrable.

On peut alors utiliser le **théorème de Fubini, cas général** 2.5.3 qui nous permet de calculer l'intégrale de f variable par variable.

Notons que le calcul effectué pour utiliser le théorème de Tonelli peut servir à nouveau pour intégrer variable par variable (si f et $|f|$ se ressemblent).

2.6 Démontrer en pratique des résultats sur les intégrales

Pour montrer des résultats sur des fonctions mesurables positives, on utilise la décomposition vue au théorème 1.5.1.

- On montre le résultat sur des fonctions indicatrices mesurables.
- Puis on le montre sur une somme de fonctions indicatrices mesurables. On a donc montré le résultat pour les fonctions étagées mesurables.
- En vertu de l'hypothèse de convergence monotone de l'intégrale et de la décomposition fournie par le théorème 1.5.1, on passe à la limite. Le résultat est alors démontré pour n'importe-quelle fonction mesurable positive.

En général, les résultats pour les fonctions signées se déduiront des résultats pour les fonctions positives de la manière suivante :

- On décompose une fonction signée mesurable $f : E \longrightarrow \mathbb{R}$ en sa partie positive f^+ et sa partie négative f_- (voir la définition 2.4.1).
- Ces deux fonctions étant positives (et encore mesurables par la proposition 2.4.2), on applique leur applique un résultat vrai pour les fonctions positives.
- On utilise l'identité $f = f^+ + f_-$ pour obtenir le résultat final.

On rappelle toutefois qu'un certain nombre de résultats vrais pour les fonctions positives ne le sont pas pour les fonctions signées.

2.7 Mesures définies par une intégrale

Ces mesures ont une place très importante dans la théorie de la mesure et en probabilités. Dans toute cette section, on se donne un espace mesuré (E, \mathcal{A}, μ) .

Définition 2.7.1. Soit ν une mesure sur (E, \mathcal{A}) . On dit que ν est définie par une intégrale lorsqu'il existe $f : E \rightarrow \mathbb{R}$, positive et μ -mesurable telle que :

$$\forall A \in \mathcal{A}, \nu(A) = \int_E f(x) d\mu(x)$$

Dans ce cas, on notera $\nu = f\mu$.

On montre le résultat important suivant qui donne toute son importance à ce type de mesures :

Théorème 2.7.1. Soit (E, \mathcal{A}, μ) un espace mesuré et $f : E \rightarrow \mathbb{R}$, positive et μ -mesurable. Soit $\nu := f\mu$. Soit $g : E \rightarrow \mathbb{R}$, ν -mesurable et dont la ν -intégrale existe.

Alors g est ν -intégrable ssi $g \cdot f$ est μ -intégrable, et

$$\int_E g(x) d\nu(x) = \int_E g(x)f(x) d\mu(x)$$

Remarque 2.7.1. En fin de compte, utiliser une mesure définie par une intégrale revient simplement à changer dans l'intégrale $d\nu(x)$ par $f(x)d\mu(x)$. Lorsque la mesure μ est bien connue (par exemple lorsqu'il s'agit de la mesure de Lebesgue ou de comptage), ce théorème permet d'avantageusement remplacer un calcul fastidieux à l'aide de ν avec un calcul plus simple à l'aide de μ et d'une multiplication par f .

Par ailleurs, c'est de là que vient la notation suivante qui est très souvent utilisée :

$$d\nu(x) = f(x)d\mu(x)$$

Dans le cas où μ est la mesure de Lebesgue, on adopte naturellement la notation utilisée dans les intégrales et l'on écrit :

$$d\nu(x) = f(x)dx$$

Cette notation est compatible avec celle du calcul différentiel, mais il convient de se rappeler que ce n'est qu'une notation et qu'il faut en comprendre le sens sous-jacent.

Il existe deux cas particuliers important concernant ce type de mesures, utiles pour les probabilités. Ils seront abordés dans la section 5.4.

2.8 Généralisation aux fonctions à valeurs dans \mathbb{R}^m

Nos fonctions étaient jusqu'ici à valeurs dans \mathbb{R} . Qu'en est-il si elle sont à valeurs dans \mathbb{R}^m ?

Soit $f : E \longrightarrow \mathbb{R}^m$. Décomposons f :

$$f = \sum_{i=1}^m f_i e_i$$

où :

- $(f_i)_{i \in \llbracket 1, m \rrbracket}$ sont les **composantes** de f .
- $(e_i)_{i \in \llbracket 1, m \rrbracket}$ est la base canonique de \mathbb{R}^m .

Posons alors les définitions suivantes :

Définition 2.8.1. L'intégrale de f existe lorsque les m intégrales des f_i existent. Lorsqu'elle existe, l'intégrale de f vaut :

$$\int_E f = \sum_{i=1}^m \int_E f_i e_i$$

(on peut permuter vecteurs et intégrale)

Définition 2.8.2. f est dite **intégrable** lorsque les m fonctions f_i sont intégrables.

Avec ces deux définitions, nous avons très facilement généralisé la notion d'intégrale lorsque les fonctions sont à valeurs dans \mathbb{R}^m . Tous les résultats d'intégrations valables pour des fonctions à valeurs dans \mathbb{R} pourront ainsi se transposer très facilement dans le cas général.

Chapitre 3

Cas particulier : l'intégrale de Lebesgue munie de la mesure de Lebesgue

Dans ce chapitre, nous allons voir les principaux résultats d'intégration, lorsque l'on muni l'intégrale de Lebesgue de la mesure de Lebesgue. Dans la pratique, c'est souvent cette catégorie d'intégrale qui sera utilisée, puisqu'elle a un comportement très proche de l'intégrale de Riemann.

On rappelle que, par définition, la mesure de Lebesgue est définie sur l'espace mesurable $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ ou $(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n))$. Nos fonctions prendront donc leurs valeurs dans \mathbb{R}^n .

Pour faciliter les notations, et souligner cette ressemblance, on notera désormais :

$$\int_A f(x) d\lambda(x) = \int_A f(x) dx$$

3.1 Que deviennent les résultats généraux dans ce cas particulier ?

3.2 L'intégrale indéterminée

Chapitre 4

Cas particulier : l'intégrale de Lebesgue munie de la mesure de comptage

Dans ce chapitre, nous allons voir les principaux résultats d'intégration, lorsque l'on muni l'intégrale de Lebesgue de la mesure de comptage. Nous allons voir que cela nous permet d'explorer la théorie des séries.

Dans tout ce chapitre, on considère l'espace mesuré $(\mathbb{N}^n, \mathcal{P}(\mathbb{N}^n), c)$, où c est la mesure de comptage. Nos fonctions prendront donc leurs valeurs dans \mathbb{N}^n . On énoncera souvent les résultats pour $n = 1$, mais ils se généralisent sans peine dans le cas général.

4.1 Lien avec les séries

Dans toute cette section, on se donne une fonction $f : (\mathbb{N}, \mathcal{P}(\mathbb{N}), c) \rightarrow \mathbb{R}$ (c'est-à-dire une suite numérique).

On commence par énoncer le résultat suivant. Il permet de nous affranchir de vérifier la mesurabilité dans notre cas :

Proposition 4.1.1. $f : (\mathbb{N}, \mathcal{P}(\mathbb{N}), c) \rightarrow \mathbb{R}$ est mesurable.

Démonstration. Pour toute partie A de \mathbb{R} , $f^{-1}(A) \in \mathcal{P}(\mathbb{N})$. A fortiori, cela est vrai pour tout borélien. \square

Remarque 4.1.1. Plus généralement, ce résultat est vrai dès que la tribu de l'espace de départ est l'ensemble des parties.

Théorème 4.1.1. Supposons que l'intégrale de f existe. Alors

$$\int_{\mathbb{N}} f(n) dc(n) = \sum_{n=0}^{+\infty} f(n)$$

Démonstration. Commençons par montrer le résultat dans le cas où f est positive. Dans ce cas, la suite de fonctions suivante $(f_n)_{n \in \mathbb{N}}$ est croissante et converge simplement vers f :

$$\forall n \in \mathbb{N}, \forall k \in \mathbb{N}, f_n(k) := \begin{cases} f(k) & \text{si } k \leq n \\ 0 & \text{sinon} \end{cases}$$

Notons que l'on peut réécrire cette suite sous la forme :

$$\forall n \in \mathbb{N}, \forall k \in \mathbb{N}, f_n(k) = \sum_{i=0}^n f(i) \mathbf{1}_i(k)$$

Et remarquons que, d'après la relation mesure-intégrale, pour $A \subset \mathbb{N}$:

$$\int_{\mathbb{N}} \mathbf{1}_A(k) dc(k) = c(A)$$

Nous pouvons alors écrire que, pour tout $n \in \mathbb{N}$:

$$\begin{aligned} \int_{\mathbb{N}} f_n(k) dc(k) &= \int_{\mathbb{N}} \sum_{i=0}^n f(i) \mathbf{1}_{\{i\}}(k) dc(k) \\ &= \sum_{i=0}^n f(i) \int_{\mathbb{N}} \mathbf{1}_{\{i\}}(k) dc(k) \\ &= \sum_{i=0}^n f(i) c(\{i\}) \\ &= \sum_{i=0}^n f(i) \end{aligned}$$

Or, f_n est une suite croissante de fonctions mesurables positives. On peut donc appliquer l'hypothèse de convergence monotone :

$$\int_{\mathbb{N}} f_n(k) dc(k) \xrightarrow{n \rightarrow +\infty} \int_{\mathbb{N}} f(k) dc(k)$$

De plus, la théorie des séries nous fournit (quitte à ce que la série vaille $+\infty$) :

$$\sum_{i=0}^n f(i) \xrightarrow{n \rightarrow +\infty} \sum_{i=0}^{+\infty} f(i)$$

Ainsi, par unicité de la limite, si f est positive :

$$\int_{\mathbb{N}} f(k) dc(k) = \sum_{i=0}^{+\infty} f(i)$$

Supposons désormais que f soit signée. On décompose f en ses parties positive et négative. Ces deux fonctions étant positives, on peut leur appliquer le résultat que l'on vient d'établir. En faisant la différence des deux, on obtient le résultat désiré. \square

Remarque 4.1.2. Notons que, par hypothèse, l'intégrale de f existe, mais f n'est pas forcément intégrable : on autorise donc la série et l'intégrale à valoir éventuellement $+\infty$ ou $-\infty$. On peut aisément vérifier par ailleurs que :

$$\begin{aligned}
 \int_{\mathbb{N}} f(n) d\mathbb{C}(n) \in \mathbb{R} & \quad \text{ssi} \quad f \text{ est intégrable} \\
 & \quad \text{ssi} \quad \|f\| \text{ est intégrable} \\
 & \quad \text{ssi} \quad \int_{\mathbb{N}} \|f(n)\| d\mathbb{C}(n) < +\infty \\
 & \quad \text{ssi} \quad \sum_{n=0}^{+\infty} \|f(n)\| < +\infty \\
 & \quad \text{ssi} \quad \sum_{n=0}^{+\infty} f(n) \in \mathbb{R}
 \end{aligned}$$

Définition 4.1.1. On dit que la série $\sum f(n)$ **converge** lorsque l'intégrale de la fonction f existe.

On dit que la série $\sum f(n)$ **converge absolument**, ou que la famille $(f(n))_{n \in \mathbb{N}}$ est **sommable**, lorsque f est intégrable, ie lorsque $\sum_{n=0}^{+\infty} |f(n)| < +\infty$

Nous avons vu tout au long de cette section que l'intégrale de Lebesgue est intrinsèquement liée à la théorie des séries. Tous les résultats que nous avons énoncés dans le chapitre ?? sont vrais dans ce cas particulier, et s'expriment assez simplement. L'objet de la section suivante est justement d'explicitier ces propriétés.

4.2 Que deviennent les propriétés de l'intégrale dans ce cas particulier ?

Théorème 4.2.1 (de convergence dominée). Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions de \mathbb{N} dans \mathbb{R} , qui converge simplement vers une fonction f .

Supposons qu'il existe $g : \mathbb{N} \rightarrow \mathbb{R}$, telle que $\sum g(n)$ converge absolument, et qui **domine** les f_n :

$$\forall n \in \mathbb{N}, \forall x \in E, |f_n(x)| \leq g(x)$$

Alors les f_n et f sont μ -intégrables, et :

$$\lim_{n \rightarrow +\infty} \sum_{k=0}^{+\infty} f_n(k) = \sum_{k=0}^{+\infty} f(k)$$

Deuxième partie

Probabilités

Chapitre 5

Généralités sur les probabilités

5.1 Premières définitions

Commençons par définir les notions fondamentales de probabilités, à l'aide du formalisme introduit dans la partie précédente.

Définition 5.1.1. On appelle **espace probabilisable** un espace mesurable, dans le contexte des probabilités.

Définition 5.1.2. Soit (Ω, \mathcal{A}) un espace probabilisable. Une **mesure de probabilité** (ou **loi de probabilité**) \mathbb{P} sur (Ω, \mathcal{A}) est une mesure sur (Ω, \mathcal{A}) , à valeurs dans $[0, 1]$, telle que $\mathbb{P}(\Omega) = 1$.

Définition 5.1.3. On appelle **espace probabilisé** un espace probabilisable muni d'une mesure de probabilité.

Définition 5.1.4. Soit (Ω, \mathcal{A}) un espace probabilisable. Un élément $\omega \in \Omega$ s'appelle une **issue**. Ω s'appelle **l'univers** (l'ensemble des issues possibles). Un élément $A \in \mathcal{A}$ un **évènement** (un ensemble d'issues).

5.2 Indépendance et conditionnement

Dans toute cette section, on se donne un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

Définition 5.2.1. Soit $(A, B) \in \mathcal{A}^2$. On dit que A et B sont des **évènements indépendants** lorsque :

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

Remarque 5.2.1. Attention, l'indépendance de deux évènements dépend de la mesure de probabilité choisie !

Définition 5.2.2. Plus généralement, soit I un ensemble quelconque et $(A_i)_{i \in I} \subset \mathcal{A}^I$. Les $(A_i)_{i \in I}$ sont dits **mutuellement indépendants** (ou tout simplement, indépendants) si pour toute partie finie $J \subset I$,

$$\mathbb{P} \left(\bigcap_{i \in J} A_i \right) = \prod_{i \in J} \mathbb{P}(A_i)$$

Les $(A_i)_{i \in I}$ sont dits **deux à deux indépendants** si pour tous $(i, j) \in I^2$ tels que $i \neq j$,

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \mathbb{P}(A_j)$$

Proposition 5.2.1. *La mutuelle indépendance implique l'indépendance deux à deux.*

Proposition 5.2.2 (Convervation de l'indépendance par passage au complémentaire). *Soit A et B deux évènements indépendants. Alors les couples d'évènements suivants sont toujours indépendants :*

- \bar{A} et B ;
- A et \bar{B} ;
- \bar{A} et \bar{B} .

Plus généralement, si I est un ensemble quelconque et les $(A_i)_{i \in I} \subset \mathcal{A}^I$ des évènements mutuellement indépendants (resp. deux à deux indépendants), alors pour tout $J \subset I$, les évènements de la famille d'évènements suivante :

$$(A'_i)_{i \in I} : \begin{cases} I & \longrightarrow \mathcal{A} \\ i & \longmapsto \begin{cases} \bar{A}_i & \text{si } i \in J \\ A_i & \text{sinon} \end{cases} \end{cases}$$

sont encore mutuellement indépendants (resp. deux à deux indépendants).

Définition 5.2.3. Soit $(A, B) \in \mathcal{A}^2$, tels que $\mathbb{P}(B) > 0$. Alors la **probabilité conditionnelle de A sachant B** , notée $\mathbb{P}(A|B)$ ou $\mathbb{P}_B(A)$ existe et vaut :

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

Proposition 5.2.3. *Soit $B \subset \mathcal{A}$ tel que $\mathbb{P}(B) > 0$. Alors la fonction :*

$$\mathbb{P}(\cdot|B) : \begin{cases} \mathcal{A} & \longrightarrow [0, 1] \\ A & \longmapsto \mathbb{P}(A|B) \end{cases}$$

est une nouvelle loi de probabilité sur l'espace probabilisable (Ω, \mathcal{A}) .

Proposition 5.2.4. *Soit $(A, B) \in \mathcal{A}^2$, tels que $\mathbb{P}(B) > 0$. Alors A et B sont indépendants ssi $\mathbb{P}(A|B) = \mathbb{P}(A)$.*

5.2.1 Trois formules utiles

La **formule de Bayes** permet d'intervertir le conditionnement :

Théorème 5.2.1 (Formule de Bayes). Soit $(A, B) \in \mathcal{A}^2$, tels que $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$. On a :

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

La **formule des probabilités totales** permet de connaître un évènement à l'aide de son conditionnement sur des évènements formant une partition de l'univers :

Théorème 5.2.2 (Formule des probabilités totales). Soit $A \in \mathcal{A}$ et $(B_i)_{i \in I} \subset \mathcal{A}^I$ une partition de Ω au plus dénombrable. Supposons que pour tout $i \in I$, $\mathbb{P}(B_i) > 0$. Alors :

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i \in I} (A \cap B_i)\right) = \sum_{i \in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

Enfin, cette généralisation de la formule de Bayes s'obtient en combinant la formule de Bayes et la formule des probabilités totales :

Théorème 5.2.3 (Formule de Bayes généralisée). Soit $A \in \mathcal{A}$ et $(B_i)_{i \in I} \subset \mathcal{A}^I$ une partition de Ω au plus dénombrable. Supposons que $\mathbb{P}(A) > 0$ et que pour tout $i \in I$, $\mathbb{P}(B_i) > 0$. Alors :

$$\forall i \in I, \mathbb{P}(B_i|A) = \mathbb{P}(A|B_i) \cdot \frac{\mathbb{P}(B_i)}{\sum_{i \in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}$$

5.3 Fonction de répartition

La fonction de répartition est une notion qui ne concerne que les mesures de probabilités sur \mathbb{R} . La notion existe sur \mathbb{R}^n , mais est peu utilisée. Dans cette section, on se place donc sur un espace probabilisé $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$.

Définition 5.3.1. La **fonction de répartition** de \mathbb{P} est :

$$F_{\mathbb{P}} : \begin{cases} \mathbb{R} & \longrightarrow [0, 1] \\ x & \longmapsto \mathbb{P}(]-\infty, x]) \end{cases}$$

Théorème 5.3.1. La fonction de répartition de \mathbb{P} caractérise \mathbb{P} .

Remarque 5.3.1. En particulier, cela implique que si deux mesures de probabilité ont la même fonction de répartition, alors elles sont égales.

Théorème 5.3.2 (Conditions nécessaires et suffisantes pour qu'une fonction soit une fonction de répartition). Soit $F : \mathbb{R} \rightarrow \mathbb{R}$. F est la fonction de répartition d'une mesure de probabilité \mathbb{Q} sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ si et seulement si les quatre conditions suivantes sont remplies :

- F est croissante ;
- F est continue à droite ;
- $\lim_{x \rightarrow -\infty} F(x) = 0$;
- $\lim_{x \rightarrow +\infty} F(x) = 1$.

Proposition 5.3.1 (Propriétés des fonctions de répartition). *La fonction de répartition vérifie les résultats suivants :*

- $F_{\mathbb{P}}$ admet une limite à gauche en tout $x \in \mathbb{R}$, notée $F_{\mathbb{P}}(x^-)$;
- $\mathbb{P}(]x, y]) = F_{\mathbb{P}}(y) - F_{\mathbb{P}}(x)$;
- $\mathbb{P}(]x, y[) = F_{\mathbb{P}}(y^-) - F_{\mathbb{P}}(x)$;
- $\mathbb{P}([x, y]) = F_{\mathbb{P}}(y) - F_{\mathbb{P}}(x^-)$;
- $\mathbb{P}([x, y[) = F_{\mathbb{P}}(y^-) - F_{\mathbb{P}}(x^-)$.

Remarque 5.3.2. En vertu du théorème 5.3.2, on a bien $F_{\mathbb{P}}(x) = F_{\mathbb{P}}(x^+)$, mais nous n'avons aucun résultat équivalent sur la limite à gauche (ie sur la continuité à gauche). Le corollaire suivant permet néanmoins de caractériser la continuité à gauche.

Corollaire 5.3.1 (Caractérisation de la continuité de la fonction de répartition). *Soit $x \in \mathbb{R}$. On a $\mathbb{P}(\{x\}) = F_{\mathbb{P}}(x) - F_{\mathbb{P}}(x^-)$. En particulier, $\mathbb{P}(\{x\}) = 0$ ssi $F_{\mathbb{P}}$ est continue en x .*

5.4 Deux cas particuliers de mesures de probabilités

Il existe un cas particulier où les mesures de probabilité sont simples à exprimer : celui où elles sont définies par une intégrale (cf. section 2.7). Il permet de grandement simplifier les calculs.

Ce cas particulier se décline en deux, selon si la mesure considérée est celle de Lebesgue ou celle de comptage.

5.4.1 Mesures de probabilité à densité

Définition 5.4.1. On se place sur \mathbb{R} (resp. \mathbb{R}^n) muni de la tribu borélienne et de la mesure de Lebesgue λ . Soit f une fonction positive et intégrable d'intégrale 1. Soit $\mathbb{P} := f\lambda$. Alors \mathbb{P} est une mesure de probabilité, et on dit alors que \mathbb{P} admet une **densité**.

Remarque 5.4.1. Notons que la réciproque est fausse : il existe des mesures sur \mathbb{R} (resp. \mathbb{R}^n) qui n'admettent pas de densité.

Par ailleurs, si une mesure de probabilité admet une densité, elle en admet une infinité. En effet, si une fonction f convient, alors une autre fonction g qui égale à f presque partout convient également : il y a **unicité de la densité de probabilité à égalité presque-partout près**.

Enfin, en pratique, pour montrer qu'une mesure de probabilité \mathbb{P} est égale à une autre mesure de probabilité \mathbb{Q} de densité f , il suffira de montrer que \mathbb{P} admet

elle aussi f pour densité (la condition nécessaire et suffisante étant donnée par le fait que \mathbb{P} admette pour densité une fonction g égale λ -presque-partout à f).

Voici justement un théorème qui donne une condition nécessaire et suffisante simple pour qu'une mesure de probabilité admette une certaine fonction de densité. C'est celle qui est utilisée en pratique (plus simple que de passer par la définition).

Théorème 5.4.1 (Condition nécessaire et suffisante pour admettre une densité). *On se place sur \mathbb{R} (resp. \mathbb{R}^n) muni de la tribu borélienne et de la mesure de Lebesgue λ . Soit f une fonction positive et intégrable d'intégrale 1. Une mesure de probabilité \mathbb{P} sur \mathbb{R} admet f comme densité ssi*

$$\forall x \in \mathbb{R}, \mathbb{P}(]-\infty, x]) = \int_{-\infty}^x f(t) dt$$

et resp. une mesure de probabilité \mathbb{P} sur \mathbb{R}^n admet f comme densité ssi

$$\forall (x_1, \dots, x_n) \in \mathbb{R}^n, \mathbb{P}(]-\infty, x_1], \dots,]-\infty, x_n]) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(t) dt$$

Proposition 5.4.1 (Propriétés de la fonction de répartition d'une mesure admettant une densité). *Soit \mathbb{P} une mesure de probabilité sur \mathbb{R} muni de la tribu borélienne admettant une densité f . Alors la fonction de répartition $F_{\mathbb{P}}$ de \mathbb{P} est **continue**, ie $\forall x \in \mathbb{R}, \mathbb{P}(\{x\}) = 0$. De plus, $F_{\mathbb{P}}$ est dérivable en tout point x où f est continue, et le cas échéant $F'_{\mathbb{P}}(x) = f(x)$.*

Proposition 5.4.2 (Une condition suffisante pour admettre une densité (à l'aide de la fonction de répartition)). *Soit \mathbb{P} une mesure de probabilité sur \mathbb{R} muni de la tribu borélienne. Si $F_{\mathbb{P}}$ est dérivable, alors \mathbb{P} admet une densité, donnée par $F'_{\mathbb{P}}$.*

5.4.2 Mesure de probabilité discrètes

Le deuxième cas particulier concerne les espaces discrets (comprendre au plus dénombrables). Sans perte de généralité, on se place uniquement dans le cas de \mathbb{N} .

On commence par énoncer le résultat important suivant :

Théorème 5.4.2. *Soit $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mathbb{P})$ un espace probabilisé. Il existe une unique suite $f : \mathbb{N} \rightarrow \mathbb{R}$ positive et sommable de somme 1 telle que $\mathbb{P} = fc$, où c est la mesure de comptage. De plus, $\forall n \in \mathbb{N}, f(n) = \mathbb{P}(\{n\})$, ie :*

$$\forall A \subset \mathbb{N}, \mathbb{P}(A) = \sum_{a \in A} \mathbb{P}(\{a\})$$

Remarque 5.4.2. Ce résultat nous prouve d'une **mesure de probabilité discrète est caractérisée par ses probabilités élémentaires** (ie par ses valeurs sur les singletons). En particulier, deux mesures de probabilités discrètes sont égales si et seulement si :

- ce sont des mesures sur le même espace discret ;
- elles sont égales sur les probabilités élémentaires.

C'est la caractérisation que l'on utilisera en pratique pour vérifier que deux lois sont égales.

Notons cependant que tout ceci est faux dans le cas général. Par exemple, dans le cas des mesures à densité, toutes les probabilités élémentaires sont nulles (cf. proposition 5.4.1) : il n'est donc pas possible de caractériser les lois à densité de cette manière.

Chapitre 6

Variables aléatoires

Voyons maintenant ce qu'est une variable aléatoire. C'est une appellation étrange, puisqu'une variable aléatoire n'est en fait ni une variable, ni quelque chose d'aléatoire. En réalité, il s'agit simplement d'une fonction qui possède certaines propriétés (notamment la mesurabilité). En général, nous nous limiterons ici à l'étude des variables aléatoires réelles, des vecteurs aléatoires et des variables aléatoires discrètes.

6.1 Définitions

Dans toute cette section, on se donne un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. On rappelle que $\mathcal{B}(E)$ désigne la tribu des boréliens de l'espace topologique E .

Définition 6.1.1. Une **variable aléatoire réelle** est une fonction mesurable

$$X : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

Définition 6.1.2. Un **vecteur aléatoire** est une fonction mesurable

$$X : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$$

Définition 6.1.3. Une **variable aléatoire discrète** est une fonction mesurable

$$X : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (\mathbb{N}, \mathcal{P}(\mathbb{N}))$$

On peut également introduire la définition suivante, qui ne sera pas utile en pratique, mais qui a le mérite d'unifier les trois définitions précédentes :

Définition 6.1.4. Plus généralement, si E est un espace topologique quelconque, une **variable aléatoire** est une fonction mesurable

$$X : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (E, \mathcal{B}(E))$$

Remarque 6.1.1. On remarquera que, exceptés l'espace d'arrivée, ces définitions sont identiques : dans tous les cas, on choisit des fonctions mesurables, et conformément à la simplification expliquée à la remarque 1.4.2, on choisit la tribu borélienne sur l'espace d'arrivée. Cela peut ne pas sembler évident pour la définition 6.1.3 concernant les variables aléatoires discrètes : on rappelle donc que les boréliens de \mathbb{N} (muni de la topologie discrète) sont exactement les parties de \mathbb{N} (pour plus de détails, se référer à l'exemple 1.1.1).

Remarque 6.1.2. On fera attention au fait que :

- une variable aléatoire est une fonction qui prend ses valeurs dans Ω ;
- une mesure de probabilité est une fonction qui prend ses valeurs dans \mathcal{A} .

Ainsi, si $\omega \in \Omega$:

- écrire $X(\omega)$ a du sens.
- écrire $\mathbb{P}(\omega)$ n'a pas de sens (ω n'a aucune chance d'être dans \mathcal{A} , puisque ce n'est même pas une partie de Ω).
- écrire $\mathbb{P}(\{\omega\})$ a du sens dès que $\{\omega\} \in \mathcal{A}$ (ie dès que $\{\omega\}$ est mesurable).
- écrire $X(\{\omega\})$ n'a pas de sens.

Proposition 6.1.1. *Soit E et F deux espaces topologiques. Soit $X : \Omega \longrightarrow E$ une variable aléatoire à valeurs dans E . Soit $g : E \longrightarrow F$ une fonction borélienne. Alors $g \circ X$ est une variable aléatoire à valeurs dans F .*

Remarque 6.1.3. Cette proposition est une conséquence immédiate du résultat de composition de fonctions mesurables.

6.2 Loi d'une variable aléatoire

Dans toute cette section, on se donne

- un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$;
- une variable aléatoire quelconque X à valeurs dans un espace topologique $(E, \mathcal{B}(E))$.

Ces définitions peuvent être données pour des variables aléatoires quelconques, mais seront utiles en pratique pour les trois types de variables aléatoires que nous avons vues.

Définition 6.2.1. La loi de probabilité de X \mathbb{P}_X est définie comme :

$$\mathbb{P}_X = \mathbb{P} \circ X^{-1}$$

C'est la mesure image de \mathbb{P} par X . Il s'agit d'une loi de probabilité sur $(E, \mathcal{B}(E))$

Remarque 6.2.1. Ces notations étant lourdes, on préférera en introduire de nouvelles, plus simples à interpréter. Attention : elles n'ont pas de sens en elles-mêmes, mais ont l'avantage de bien faire comprendre la notion sous-jacente.

- Si $F \in \mathcal{B}(E)$, on notera :

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

d'où :

$$\mathbb{P}(X \in F) = \mathbb{P} \circ X^{-1}(F) = \mathbb{P}(X^{-1}(F)) = \mathbb{P}_X(F)$$

- Si $F = \{e\}$, alors notera plutôt :

$$\{X = e\} := X^{-1}(\{e\}) = \{\omega \in \Omega | X(\omega) = e\} \in \mathcal{A}$$

d'où :

$$\mathbb{P}(X = e) = \mathbb{P} \circ X^{-1}(\{e\}) = \mathbb{P}(X^{-1}(\{e\})) = \mathbb{P}_X(\{e\})$$

- Si $E = \mathbb{R}$, et que $F = [x, y]$, alors on notera plutôt :

$$\{x \leq X \leq y\} := X^{-1}([x, y]) = \{\omega \in \Omega | X(\omega) \in [x, y]\} \in \mathcal{A}$$

d'où :

$$\mathbb{P}(x \leq X \leq y) = \mathbb{P} \circ X^{-1}([x, y]) = \mathbb{P}(X^{-1}([x, y])) = \mathbb{P}_X([x, y])$$

- etc...

Remarque 6.2.2. La fonction de répartition de la loi de X sera abusivement appelée fonction de répartition de X , et ne sera pas notée $F_{\mathbb{P}_X}$ mais plus simplement F_X .

Définition 6.2.2. Soit \mathbb{Q} une loi de probabilité sur $(E, \mathcal{B}(E))$. On dit que X **suit la loi \mathbb{Q}** lorsque $\mathbb{P}_X = \mathbb{Q}$.

Remarque 6.2.3. Conformément à la remarque 5.4.1, et en utilisant le théorème 5.4.1, **si \mathbb{Q} est une loi sur \mathbb{R} qui admet une densité f , pour prouver que X suit la loi \mathbb{Q} , il suffit de vérifier que X est bien à valeurs dans \mathbb{R} , puis de montrer que $\forall x \in \mathbb{R}, \mathbb{P}(X \leq x) = \int_{-\infty}^x f(t)dt$.** En effet,

$$\begin{aligned} \mathbb{P}(X \leq x) &= \mathbb{P}(X \in]-\infty, x]) \\ &= \mathbb{P}(X^{-1}(]-\infty, x])) \\ &= \mathbb{P}_X(]-\infty, x]) \\ &= \int_{-\infty}^x f(t)dt \end{aligned}$$

Evidemment, cette condition est aussi une condition nécessaire.

Dans ce cas, **on dit que X est une variable aléatoire réelle à densité** (*stricto sensu*, on devrait dire que la loi de X est à densité, mais cet abus de langage est toléré).

Remarque 6.2.4. Conformément à la remarque 5.4.2, **si \mathbb{Q} est une loi discrète sur E , pour prouver que X suit la loi \mathbb{Q} , il suffit de vérifier que X est bien à valeurs dans E , puis de montrer que $\forall e \in E, \mathbb{P}(X = e) = \mathbb{Q}(\{e\})$.** En effet,

$$\begin{aligned} \mathbb{P}(X = e) &= \mathbb{P}(X \in \{e\}) \\ &= \mathbb{P}(X^{-1}(\{e\})) \\ &= \mathbb{P}_X(\{e\}) \end{aligned}$$

Evidemment, cette condition est aussi une condition nécessaire.

Remarque 6.2.5. Attention, ce n'est pas parce que deux variables aléatoires ont la même loi qu'elles sont égales ! Voici un contre-exemple. Il utilise cependant des notions qui n'ont pas encore été abordées à ce stade.

Considérons l'espace probabilisé $\Omega := \{\text{beau temps}, \text{mauvais temps}\}$ muni de l'ensemble de ses parties, et de la mesure de probabilité \mathbb{P} uniforme. On a donc :

$$\mathbb{P}(\{\text{beau temps}\}) = \mathbb{P}(\{\text{mauvais temps}\}) = \frac{1}{2}$$

Puis posons la variable aléatoire discrète X à valeurs dans $\{0, 1\}$, telle que $X(\{\text{beau temps}\}) = 1$ et $X(\{\text{mauvais temps}\}) = 0$.

X suit alors la loi uniforme sur l'espace $\{0, 1\}$.

Considérons maintenant l'espace probabilisé $\Omega' := \{\text{temps chaud}, \text{temps froid}\}$ muni de l'ensemble de ses parties, et de la mesure de probabilité \mathbb{P}' uniforme. On a donc :

$$\mathbb{P}'(\{\text{temps chaud}\}) = \mathbb{P}'(\{\text{temps froid}\}) = \frac{1}{2}$$

Puis posons la variable aléatoire discrète Y à valeurs dans $\{0, 1\}$, telle que $Y(\{\text{temps chaud}\}) = 1$ et $Y(\{\text{temps froid}\}) = 0$.

Y suit également la loi uniforme sur l'espace $\{0, 1\}$.

Nous avons donc explicité deux variables aléatoires qui ont même loi, mais qui ne sont pas égales (elles n'ont pas le même espace de départ).

6.2.1 Loi jointe et lois marginales

Dans toute cette sous-section, on se donne

- un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$;
- n variables aléatoires quelconques $(X_i)_{i \in \llbracket 1, n \rrbracket}$ à valeurs dans $(E_i)_{i \in \llbracket 1, n \rrbracket}$.

Commençons par une définition très générale :

Définition 6.2.3. La **loi jointe** des n variables aléatoires $(X_i)_{i \in \llbracket 1, n \rrbracket}$ est la loi de la variable aléatoire (X_1, \dots, X_n) à valeurs dans (E_1, \dots, E_n) . La loi d'une variable aléatoire X_{i_0} est la i_0 e **loi marginale** de X .

Proposition 6.2.1. *La connaissance de la loi jointe permet de connaître la loi marginale. La réciproque est fausse.*

Démonstration. En effet, si l'on souhaite connaître la i_0 e loi marginale, il suffit d'écrire :

$$\begin{aligned} \forall A \in \mathcal{B}(E_{i_0}), \mathbb{P}_{X_{i_0}}(A) &= \mathbb{P}(X_{i_0}^{-1}(A)) \\ &= \mathbb{P}(X_1^{-1}(E_1) \cap \dots \cap X_{i_0}^{-1}(A) \cap \dots \cap X_n^{-1}(E_n)) \\ &= \mathbb{P}((X_1, \dots, X_{i_0}, \dots, X_n)^{-1}(E_1 \times \dots \times A \times \dots \times E_n)) \\ &= \mathbb{P}_{(X_1, \dots, X_{i_0}, \dots, X_n)}(E_1 \times \dots \times A \times \dots \times E_n) \\ &= \int_{E_1} \dots \int_{E_{i_0-1}} \int_A \int_{E_{i_0+1}} \dots \int_{E_n} d\mathbb{P}_{(X_1, \dots, X_n)}(x_1, \dots, x_n) \end{aligned}$$

Cependant, **la réciproque est fausse !** En effet, la connaissance des lois marginales ne donne aucune information sur la *relation* qu'ont les variables aléatoires entre elles. Intuitivement, il est donc impensable de pouvoir connaître la loi du vecteur aléatoire résultant. \square

En pratique cependant, on n'emploie les lois jointes et marginales uniquement lorsque l'on traite avec des variables aléatoires réelles ou discrètes. Dans la suite de cette section, on se donne donc n variables aléatoires réelles ou discrètes $(X_i)_{i \in \llbracket 1, n \rrbracket}$, que l'on peut voir comme un vecteur aléatoire X à n composantes. Remarquons que voir ces variables comme un vecteur aléatoire n'était pas possible jusqu'à maintenant, puisqu'elles n'étaient pas forcément à valeurs réelles.

Ceci étant dit, on énonce de nouveau la définition 6.2.3 dans ce cas particulier :

Définition 6.2.4. La **loi jointe** des n variables aléatoires $(X_i)_{i \in \llbracket 1, n \rrbracket}$ est la loi du vecteur aléatoire X . La loi d'une variable aléatoire X_{i_0} est la i_0 e **loi marginale** de X .

Maintenant, énonçons de nouveau la proposition 6.2.1 dans les deux cas particuliers suivants :

- X est une variable aléatoire réelle à densité (proposition 6.2.2) ;
- X est une variable aléatoire discrète sur \mathbb{N}^n (proposition 6.2.3).

Proposition 6.2.2. *Supposons que X soit un vecteur aléatoire à densité f . Alors les lois marginales sont à densité (ie les $(X_i)_{i \in \llbracket 1, n \rrbracket}$ sont des variables aléatoires à densité), de densité $(f_i)_{i \in \llbracket 1, n \rrbracket}$, vérifiant pour tout $i \in \llbracket 1, n \rrbracket$:*

$$\forall x_i \in \mathbb{R}, f_i(x_i) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(x_1, \dots, x_n) d(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$$

Remarque 6.2.6. Attention, même si toutes les lois marginales sont à densité (ie toutes les variables aléatoires $(X_i)_{i \in \llbracket 1, n \rrbracket}$ sont à densité), la loi jointe peut ne pas admettre de densité. A nouveau, cela est un exemple que la connaissance des lois marginales ne suffit pas pour connaître la loi jointe.

Proposition 6.2.3. *Supposons que X soit une variable aléatoire discrète sur \mathbb{N}^n (ie que toutes les X_1, \dots, X_n soient toutes des variables aléatoires discrètes sur \mathbb{N}). Alors la connaissance de la loi jointe permet de connaître les lois marginales de la manière suivante :*

$$\forall i \in \llbracket 1, n \rrbracket, \forall k_i \in \mathbb{N}, \mathbb{P}(X_i = k_i) = \sum_{k_1, \dots, k_{i-1}, k_{i+1}, \dots, k_n \in \mathbb{N}^{n-1}} \mathbb{P}(X_1 = k_1, \dots, X_n = k_n)$$

6.3 Variables aléatoires indépendantes

Dans toute cette section, on se donne un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

6.3.1 Définitions

Commençons par une définition générale, valable pour n'importe-quel couple de variables aléatoires :

Définition 6.3.1. Soit X et Y deux variables aléatoires quelconques à valeurs dans E et F respectivement. X et Y sont dites **indépendantes** lorsque l'une des deux conditions équivalentes suivantes est remplie :

- la loi jointe de X et Y $\mathbb{P}_{(X,Y)}$ vérifie $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$;
- pour tout $(A, B) \in \mathcal{B}(E) \times \mathcal{B}(F)$, $\{X \in A\}$ et $\{Y \in B\}$ sont deux événements indépendants.

Généralisons maintenant en énonçant la définition pour n variables aléatoires quelconques :

Définition 6.3.2. Soit $(X_n)_{n \in [1,n]}$ n variables aléatoires quelconques, à valeurs dans (E_1, \dots, E_n) . Les $(X_n)_{n \in [1,n]}$ sont dites **mutuellement indépendantes** (ou simplement indépendantes), lorsque l'une des deux conditions équivalentes suivantes est remplie :

- la loi jointe des $(X_n)_{n \in [1,n]}$ $\mathbb{P}_{(X_1, \dots, X_n)}$ vérifie $\mathbb{P}_{(X_1, \dots, X_n)} = \bigotimes_{i=1}^n \mathbb{P}_{X_i}$;
- pour tout $(A_1, \dots, A_n) \in \prod_{i=1}^n \mathcal{B}(E_i)$, $\{X_1 \in A_1\}, \dots, \{X_n \in A_n\}$ sont n événements indépendants.

Elles sont dites **deux à deux indépendantes** lorsque pour tous $\{i, j\} \in [1, n]^2$ tels que $i \neq j$, X_i et X_j sont deux variables aléatoires indépendantes.

Remarque 6.3.1. On vérifie sans peine que la deuxième condition est équivalente à :

$$\forall (A_1, \dots, A_n) \in \prod_{i=1}^n \mathcal{B}(E_i), \mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i)$$

En pratique, pour vérifier l'indépendance de n variables aléatoires, on utilisera cette caractérisation facile. Son défaut est de ne pas faire le parallèle avec l'indépendance des événements (puisqu'on rappelle que pour vérifier l'indépendance d'un ensemble d'événements, il faut vérifier l'égalité pour tout sous-ensemble fini de ces événements).

Enfin, énonçons la définition pour une famille quelconque de variables aléatoires quelconques (en particulier valable pour les suites de variables aléatoires) :

Définition 6.3.3. Soit I un ensemble quelconque. Soit $(X_i)_{i \in I}$ des variables aléatoires quelconques, à valeurs dans $(E_i)_{i \in I}$. Les $(X_i)_{i \in I}$ sont dites **mutuellement indépendantes** (ou simplement indépendantes), lorsque pour toute sous-partie finie $J \subset I$, les $(X_i)_{i \in J}$ sont mutuellement indépendantes.

Elles sont dites **deux à deux indépendantes** lorsque pour tous $\{i, j\} \in I^2$ tels que $i \neq j$, X_i et X_j sont deux variables aléatoires indépendantes.

Proposition 6.3.1. *La mutuelle indépendance implique l'indépendance deux à deux. La réciproque est fausse.*

6.3.2 Deux résultats importants

Théorème 6.3.1 (Transfert d'indépendance). *Soit I un ensemble quelconque. Soit $(X_i)_{i \in I}$ des variables aléatoires quelconques, à valeurs dans $(E_i)_{i \in I}$.*

Supposons que les $(X_i)_{i \in I}$ soient mutuellement indépendantes (resp. deux à deux indépendantes). Soit $(\phi_i)_{i \in I}$ des fonctions de E_i dans des ensembles F_i , $\mathcal{B}(E_i)/\mathcal{B}(F_i)$ -mesurables. Alors les variables aléatoires $(\phi_i \circ X_i)_{i \in I}$, à valeurs dans F_i , sont encore mutuellement indépendantes (resp. deux à deux indépendantes).

Théorème 6.3.2 (Lemme des coalitions).

6.4 Moments d'une variable aléatoire réelle ou discrète

Dans toute cette section, on se donne

- un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$;
- une variable aléatoire réelle ou discrète X .

On ne pourra pas donner de sens à ces notions dans le cas général d'une variable aléatoire quelconque. Nous verrons néanmoins par la suite comment généraliser ces notions aux vecteurs aléatoires.

Définition 6.4.1. Soit $r \in \mathbb{N}$. On dit que X **admet un moment d'ordre r** lorsque $\omega \mapsto X^r(\omega)$ est \mathbb{P} -intégrable.

Remarque 6.4.1. On rappelle que l'intégrale de Lebesgue est absolue. Ainsi, cette définition et les assertions suivantes sont équivalentes :

- $\omega \mapsto |X(\omega)|^r$ est \mathbb{P} -intégrable.
- $\int_{\Omega} |X(\omega)|^r d\mathbb{P}(\omega) < +\infty$

Définition 6.4.2. Soit $r \in \mathbb{N}$. Lorsque X admet un moment d'ordre r , on définit son **moment d'ordre r** comme

$$m_r := \int_{\Omega} X^r(\omega) d\mathbb{P}(\omega)$$

Remarque 6.4.2. On vérifie aisément, par la relation mesure-intégrale, que toute variable aléatoire admet un moment d'ordre 0, et que celui-ci vaut toujours 1.

Proposition 6.4.1. Soit $r \in \mathbb{N}$. Si X admet un moment d'ordre r , alors elle admet un moment d'ordre k pour tout $k \in \llbracket 0, r \rrbracket$. Autrement dit, si X^r est \mathbb{P} -intégrable, alors X^k est \mathbb{P} -intégrable pour tout $k \in \llbracket 0, r \rrbracket$.

Démonstration. Soit $k \in \llbracket 0, r \rrbracket$. Nous avons $|X|^k \leq |X|^r + 1$, donc par croissance et linéarité de l'intégrale :

$$\int_{\Omega} |X(\omega)|^k d\mathbb{P}(\omega) \leq \int_{\Omega} |X(\omega)|^r d\mathbb{P}(\omega) + \int_{\Omega} d\mathbb{P}(\omega)$$

Comme X admet un moment d'ordre r , le premier terme de la somme est fini. Puis, d'après la relation mesure-intégrale, le deuxième terme de la somme vaut 1 (\mathbb{P} est une mesure de probabilité).

Le terme de gauche est donc fini, donc X admet un moment d'ordre k . \square

Remarque 6.4.3. Attention, on remarquera que ceci n'est vrai que parce que \mathbb{P} est une mesure de probabilité ! Dans le cas d'une mesure quelconque (par exemple, la mesure de Lebesgue), cette démonstration ne permet pas de conclure (car $\lambda(\mathbb{R}) = +\infty$)...et heureusement, car ce résultat est faux !

Voici un contre-exemple :

$$x \mapsto \mathbb{1}_{\mathbb{R}_+^*}(x) \frac{1}{x}$$

n'est pas intégrable sur \mathbb{R} , alors que son carré l'est.

Définition 6.4.3. On dit que X **admet une espérance** lorsqu'elle admet un moment d'ordre 1 (*ie* lorsque $\omega \mapsto X(\omega)$ est \mathbb{P} -intégrable). Son **espérance** est alors définie comme son moment d'ordre 1. Elle est notée $\mathbb{E}(X)$.

On note $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ l'ensemble des variables aléatoires qui admettent une espérance (*ie* l'ensemble des variables aléatoires \mathbb{P} -intégrables).

Remarque 6.4.4. Par abus de notation, on pourra écrire $\mathbb{E}(X) = +\infty$ lorsque X n'admet pas d'espérance et qu'elle est positive. Ainsi, on pourra écrire $\mathbb{E}(X)$ dès que X est positive, qu'elle admette une espérance ou non. Vue la place privilégiée qu'occupent les fonctions positives dans la théorie d'intégration de Lebesgue, cette notation ne surprend pas.

Attention cependant, si X n'admet pas d'espérance et n'est pas positive, on s'interdira formellement d'écrire $\mathbb{E}(X)$ (car l'intégrale sous-jacente n'existe peut-être même pas!).

Remarque 6.4.5. Ces considérations étant faites, on remarque que :

- X admet une espérance ssi $\mathbb{E}(|X|) < +\infty$;
- X admet un moment d'ordre r ssi $\mathbb{E}(|X|^r) < +\infty$;
- si X admet un moment d'ordre r , alors $m_r = \mathbb{E}(X^r)$.

6.4.1 Formule de transfert

Dans toute cette section, on se donne

- un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$;
- un espace topologique quelconque E ;
- une variable aléatoire quelconque $X : \Omega \longrightarrow E$.

Théorème 6.4.1 (Formule de transfert). Soit $g : E \longrightarrow \mathbb{R}$ (*resp.* $g : E \longrightarrow \mathbb{N}$) une fonction borélienne. Par la proposition 6.1.1, $g \circ X$ est encore une variable aléatoire ; c'est d'ailleurs une variable aléatoire réelle (*resp.* une variable aléatoire discrète). On sait donc donner un sens à son espérance.

La variable aléatoire réelle (*resp.* discrète) $g \circ X$ sur $(\Omega, \mathcal{A}, \mathbb{P})$ admet une espérance (*ie* $\omega \mapsto (g \circ X)(\omega)$ est \mathbb{P} -intégrable) ssi la variable aléatoire réelle (*resp.* discrète) g sur $(E, \mathcal{B}(E), \mathbb{P}_X)$ admet une espérance (*ie* $x \mapsto g(x)$ est \mathbb{P}_X -intégrable).

Si c'est le cas, alors on a la **formule de transfert** :

$$\mathbb{E}(g \circ X) = \int_E g(x) d\mathbb{P}_X(x)$$

Remarque 6.4.6. On comprend pourquoi cette formule porte ce nom, lorsque l'on remplace l'espérance par sa définition, dans les expressions précédentes :

$$\int_{\Omega} (g \circ X)(\omega) d\mathbb{P}(\omega) = \int_E g(x) d\mathbb{P}_X(x)$$

Cette formule transfère le calcul sur un autre espace, muni d'une autre mesure. Pour peu que la mesure \mathbb{P}_X soit facile à utiliser, on a intérêt à utiliser cette nouvelle formulation.

Voici quelques applications de ce théorème :

Exemple 6.4.1. Soit $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire réelle. La fonction identité $\text{Id} : \mathbb{R} \rightarrow \mathbb{R}$ est continue donc borélienne (voir la proposition 1.4.5).

En appliquant le théorème précédent, X admet une espérance ssi Id est \mathbb{P}_X -intégrable. Dans ce cas, on aura :

$$\mathbb{E}(X) = \mathbb{E}(\text{Id} \circ X) = \int_{\mathbb{R}} x d\mathbb{P}_X(x)$$

Pour peu que l'on sache expliciter \mathbb{P}_X , cette nouvelle l'expression de l'espérance nous simplifie grandement le calcul.

Il est évident que l'on obtient le même résultat avec une variable aléatoire discrète.

Exemple 6.4.2. De même, en choisissant la fonction carrée pour g on obtient (si l'espérance de X^2 existe) :

$$\mathbb{E}(X^2) = \int_{\mathbb{R}} x^2 d\mathbb{P}_X(x)$$

Notons qu'en choisissant l'identité pour g , mais la variable aléatoire X^2 , on obtient (à nouveau sous réserve d'existence) :

$$\mathbb{E}(X^2) = \int_{\mathbb{R}} x d\mathbb{P}_{X^2}(x)$$

Remarque 6.4.7. Voici deux conséquences importantes de la formule de transfert :

- si X est une variable aléatoire réelle de densité f , alors $g \circ X$ admet une espérance ssi

$$\int_{\mathbb{R}} |g(x)| f(x) dx \left(= \int_{\mathbb{R}} |g(x)| d\mathbb{P}_X(x) \right) < +\infty$$

Si c'est le cas, alors :

$$\mathbb{E}(g \circ X) = \int_{\mathbb{R}} g(x) f(x) dx \left(= \int_{\mathbb{R}} g(x) d\mathbb{P}_X(x) \right)$$

La démonstration est immédiate à l'aide du théorème 2.7.1 et de la formule de transfert. Si X était un vecteur aléatoire de densité f , le résultat est le même (remplacer \mathbb{R} par \mathbb{R}^n).

- si X est une variable aléatoire discrète, alors $g \circ X$ admet une espérance ssi

$$\sum_{k=0}^{+\infty} |g(k)| \mathbb{P}(X = k) \left(= \int_{\mathbb{R}} |g(x)| d\mathbb{P}_X(x) \right) < +\infty$$

Si c'est le cas, alors :

$$\mathbb{E}(g \circ X) = \sum_{k=0}^{+\infty} g(k) \mathbb{P}(X = k) \left(= \int_{\mathbb{R}} g(x) d\mathbb{P}_X(x) \right)$$

En effet, dans le cas discret, $\mathbb{P}_X = f c$, où $f(k) = \mathbb{P}_X(\{k\}) = \mathbb{P}(X = k)$ (cf. théorème 5.4.2). Puis le théorème 2.7.1 fournit l'équivalence et l'égalité. Enfin, le théorème 4.1.1 fournit l'expression sous forme de somme. Remarquons que tout a été donné ici dans le cas de \mathbb{N} , mais tout s'énonce de même dans les autres cas.

Dans le cas où $g = \cdot^r$, on obtient les résultats suivants :

- si X est une variable aléatoire réelle de densité f , alors X admet un moment d'ordre r ssi

$$\int_{\mathbb{R}} |x|^r f(x) dx < +\infty$$

Si c'est le cas, alors :

$$\mathbb{E}(X^r) = \int_{\mathbb{R}} x^r f(x) dx$$

- si X est une variable aléatoire discrète, alors X admet un moment d'ordre r ssi

$$\sum_{k=0}^{+\infty} |k|^r \mathbb{P}(X = k) < +\infty$$

Si c'est le cas, alors :

$$\mathbb{E}(X^r) = \sum_{k=0}^{+\infty} k^r \mathbb{P}(X = k)$$

Remarque 6.4.8. La formule de transfert ne doit pas être confondue avec le transfert d'indépendance !

6.4.2 Indépendance et espérance

Cette sous-section vise uniquement à donner un résultat important concernant l'espérance de variables aléatoires, lorsqu'il y a indépendance :

Théorème 6.4.2. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et X et Y deux variables aléatoires quelconques sur cet espace, à valeurs dans deux espaces topologiques E et F respectivement. Soit $g : E \times F \rightarrow \mathbb{R}$ une fonction borélienne, telle que $g \circ (X, Y)$ admette une espérance. Alors :

$$\mathbb{E}(g \circ (X, Y)) = \int_E \left(\int_F g(x, y)(d)\mathbb{P}_Y(y) \right) d\mathbb{P}_X(x)$$

Remarque 6.4.9. Si g est la multiplication, alors on a immédiatement :

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

6.4.3 Variance et covariance

Définition 6.4.4. On dit que X **admet une variance** lorsqu'elle admet un moment d'ordre 2 (ie lorsque $\omega \mapsto X^2(\omega)$ est \mathbb{P} -intégrable). Sa **variance** est alors définie comme :

$$\mathbb{V}(X) := \mathbb{E} \left((X - \mathbb{E}(X))^2 \right) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

On note $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ l'ensemble des variables aléatoires qui admettent une variance (ie l'ensemble des variables aléatoires qui sont de carré \mathbb{P} -intégrables).

Remarque 6.4.10. Cette définition impose de faire trois remarques :

- le théorème qui fournit l'égalité de ces deux expressions s'appelle **le théorème de König-Huygens**. Sa démonstration est immédiate par linéarité de l'intégrale.
- d'après la deuxième expression, l'existence de $\mathbb{V}(X)$ est assurée dès que X admet un moment d'ordre 1 et 2. Cependant, bien que la définition demande que X admette un moment d'ordre 2, elle ne demande pas qu'elle admette un moment d'ordre 1. Nous sommes en fait sauvés par la proposition 6.4.1 !
- d'après la proposition 6.4.1, $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P}) \subset \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$. A nouveau, on rappelle que ceci n'est vrai que parce que \mathbb{P} est une mesure de probabilité (voir la remarque 6.4.3)

Remarque 6.4.11. Dans le cas quelconque, on peut exprimer la variance ainsi (à l'aide de la formule de transfert) :

$$\mathbb{V}(X) = \int_{\mathbb{R}} (x - \mathbb{E}(X))^2 d\mathbb{P}_X(x) = \int_{\mathbb{R}} x^2 d\mathbb{P}_X(x) - (\mathbb{E}(X))^2$$

Cependant, il existe deux cas particuliers où elle s'exprime plus simplement :

- dans le cas où X est à densité, la remarque 6.4.7 simplifie l'expression en :

$$\mathbb{V}(X) = \int_{\mathbb{R}} (x - \mathbb{E}(X))^2 f(x) dx = \int_{\mathbb{R}} x^2 f(x) dx - (\mathbb{E}(X))^2$$

- dans le cas où X est discrète, la même remarque la simplifie en :

$$\mathbb{V}(X) = \sum_{k=0}^{+\infty} (k - \mathbb{E}(X))^2 \mathbb{P}(X = k) = \sum_{k=0}^{+\infty} k^2 \mathbb{P}(X = k) - (\mathbb{E}(X))^2$$

Définition 6.4.5. Soit X et Y deux variables aléatoires réelles (ou discrètes) de $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$. La variable aléatoire $(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))$ admet une espérance et on appelle **covariance de X et Y** le réel :

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

Remarque 6.4.12. Remarquons que $\text{Cov}(X, X) = \mathbb{V}(X)$.

Définition 6.4.6. Soit X et Y deux variables aléatoires réelles (ou discrètes) de $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$. On dit que X et Y sont **non-corrélées** lorsque $\text{Cov}(X, Y) = 0$.

Proposition 6.4.2. Soit X et Y deux variables aléatoires réelles (ou discrètes). Si X et Y sont indépendantes, alors elles sont non-corrélées. La réciproque est fausse.

6.5 Moments d'un vecteur aléatoire

Dans toute cette section, on se donne

- un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$;
- un vecteur aléatoire $X = (X_i)_{i \in \mathbb{N}}$.

On ne peut pas directement adapter les définitions précédentes pour les vecteurs aléatoires, et il est nécessaire de mettre en place de nouvelles définitions. On ne généralisera que les moments d'ordre 1 et 2.

Définition 6.5.1. On dit que X **admet un vecteur espérance** lorsque pour tout $i \in \llbracket 1, n \rrbracket$, X_i admet une espérance. Si c'est le cas, alors le **vecteur espérance de X existe et est défini comme** :

$$\mathbb{E}(X) := (\mathbb{E}(X_i))_{i \in \llbracket 1, n \rrbracket}$$

Définition 6.5.2. On dit que X **admet une matrice de covariance** lorsque pour tout $i \in \llbracket 1, n \rrbracket$, X_i admet une variance. Si c'est le cas, alors la **matrice de covariance de X existe et est définie comme** :

$$\mathbb{V}(X) := (\text{Cov}(X_i, X_j))_{(i,j) \in \llbracket 1, n \rrbracket^2}$$

Remarque 6.5.1. Notons que sur la diagonale, nous retrouvons $(\mathbb{V}(X_i))_{i \in \llbracket 1, n \rrbracket}$.

Proposition 6.5.1. La matrice de covariance de X est une **matrice symétrique positive**.

6.6 Conditionnement par une variable aléatoire

Les lois conditionnelles permettent de répondre à deux problématiques :

- On sait conditionner par un évènement de probabilité non-nulle (cf. section 5.2). Cependant, on ne sait pas conditionner par un évènement de probabilité nulle de la forme $\{X = x\}$. Les lois conditionnelles permettront de donner un sens à une telle notion.
- Lorsque deux variables aléatoires réelles X et Y sont indépendantes, $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$ donc par le théorème de Fubini, on peut calculer les espérances de fonctions de X et Y très facilement en séparant les variables. Lorsqu'elles ne sont pas indépendantes, on énoncera un théorème de Fubini dit « conditionnel », qui permettra de faire une transformation similaire.

De plus, comme on mettra en place de nouvelles lois de probabilité (conditionnelles), on pourra également définir de nouvelles espérances (conditionnelles).

6.6.1 Lois conditionnelles

Dans toute cette section, on se donne :

- un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

Enonçons le théorème d'existence des lois conditionnelles dans le cas de variables aléatoires réelles :

Théorème 6.6.1 (Existence des lois conditionnelles pour des variables aléatoires réelles). *Soit X et Y deux variables aléatoires réelles. Alors il existe une famille de lois de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, notée $(\mathbb{P}_{X|Y=y})_{y \in \mathbb{R}}$, telle que :*

$$\forall (A, B) \in \mathcal{B}(\mathbb{R})^2, \mathbb{P}_{(X,Y)}(A \times B) = \int_B \mathbb{P}_{X|Y=y}(A) d\mathbb{P}_Y(y)$$

$\mathbb{P}_{X|Y=y}$ s'appelle la **loi conditionnelle de X sachant $Y = y$** .

Enonçons-le maintenant de même pour des vecteurs aléatoires :

Théorème 6.6.2 (Existence des lois conditionnelles pour des vecteurs aléatoires). *Soit X et Y deux vecteurs aléatoires à valeurs dans \mathbb{R}^m et \mathbb{R}^n . Alors il existe une famille de lois de probabilité sur $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, notée $(\mathbb{P}_{X|Y=y})_{y \in \mathbb{R}^n}$, telle que :*

$$\forall (A, B) \in \mathcal{B}(\mathbb{R}^m) \times \mathcal{B}(\mathbb{R}^n), \mathbb{P}_{(X,Y)}(A \times B) = \int_B \mathbb{P}_{X|Y=y}(A) d\mathbb{P}_Y(y)$$

$\mathbb{P}_{X|Y=y}$ s'appelle la **loi conditionnelle de X sachant $Y = y$** .

Remarque 6.6.1. Attention, cette notation n'a *a priori* rien à voir avec la théorie des probabilités conditionnées par un évènement. En effet, si $\{Y = y\}$ est de probabilité nulle, la définition habituelle ne peut pas s'appliquer (puisque l'on

doit diviser par 0). C'est d'autant plus embêtant lorsque l'on remarque que si Y admet une densité, alors $\{Y = y\}$ est de probabilité nulle pour tout $y \dots$

Cependant, si $\{Y = y\}$ n'est pas de probabilité nulle, alors on a effectivement (mais c'est un « coup de chance ») :

$$\forall A \in \mathcal{B}(\mathbb{R}), \mathbb{P}_{X|Y=y}(A) = \frac{\mathbb{P}(X \in A, Y = y)}{\mathbb{P}(Y = y)}$$

Ces nouvelles lois conditionnelles vont nous permettre d'énoncer un théorème qui ressemble fortement à celui de Fubini.

Théorème 6.6.3 (Théorème de Fubini conditionnel). *Soit X et Y deux variables aléatoires réelles, et $g : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ une fonction borélienne, telle que la variable aléatoire réelle $g \circ (X, Y)$ admette une espérance. Alors :*

$$\mathbb{E}(g \circ (X, Y)) = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x, y) d\mathbb{P}_{X,Y}(x, y) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g(x, y) d\mathbb{P}_{X|Y=y}(x) \right) d\mathbb{P}_Y(y)$$

Enonçons-le maintenant de même pour des vecteurs aléatoires :

Théorème 6.6.4 (Théorème de Fubini conditionnel). *Soit X et Y deux vecteurs aléatoires à valeurs dans \mathbb{R}^m et \mathbb{R}^n , et $g : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction borélienne, telle que la variable aléatoire réelle $g \circ (X, Y)$ admette une espérance. Alors :*

$$\mathbb{E}(g \circ (X, Y)) = \int_{\mathbb{R}^m} \int_{\mathbb{R}^n} g(x, y) d\mathbb{P}_{X,Y}(x, y) = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} g(x, y) d\mathbb{P}_{X|Y=y}(x) \right) d\mathbb{P}_Y(y)$$

Remarque 6.6.2. Ainsi, dans le cas où X et Y ne sont pas indépendantes, ce théorème permettra de faire facilement les calculs. Si elles le sont, alors par définition de l'indépendance de deux variables aléatoires, le théorème de Fubini « original » s'applique directement (voir le théorème 6.4.2).

6.6.2 Espérance conditionnelle

Cette nouvelle section n'a *a priori* rien à voir avec la section précédente. En réalité, les notions développées ci-dessus rejoignent celles ci-dessous en fin de section.

Dans toute cette section, on se donne :

- un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$;
- une variable aléatoire réelle X ;
- une sous-tribu \mathcal{B} de \mathcal{A} .

Comme fonction \mathcal{A} -mesurable, X n'est pas nécessairement \mathcal{B} -mesurable. Cependant, il est commode « d'approximer » X à l'aide d'une fonction \mathcal{B} -mesurable : c'est l'objet de la proposition suivante, qui effectue cette « approximation » en utilisant un critère basé sur l'espérance.

Proposition 6.6.1. *Il existe une fonction \tilde{X} \mathcal{B} -mesurable telle que pour tout $B \in \mathcal{B}$:*

$$\mathbb{E}(\mathbf{1}_B X) = \mathbb{E}(\mathbf{1}_B \tilde{X})$$

Cette fonction est unique à égalité \mathbb{P} -presque-partout près (ie deux fonctions qui vérifient cette propriété diffèrent au plus sur un ensemble négligeable).

Remarque 6.6.3. La fonction \tilde{X} est donc une variable aléatoire réelle sur l'espace probabilité $(\Omega, \mathcal{B}, \mathbb{P})$ (et évidemment, *a fortiori* une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$).

La variable aléatoire \tilde{X} est justement ce que l'on appelle « espérance conditionnelle ». La définition suivante synthétise donc la proposition précédente :

Définition 6.6.1. L'espérance de X conditionnée par \mathcal{B} $\mathbb{E}(X|\mathcal{B})$ est une variable aléatoire sur $(\Omega, \mathcal{B}, \mathbb{P})$ telle que :

$$\forall B \in \mathcal{B}, \mathbb{E}(\mathbf{1}_B X) = \mathbb{E}(\mathbf{1}_B \mathbb{E}(X|\mathcal{B}))$$

Elle existe toujours, et est unique à égalité \mathbb{P} -presque-partout près.

Remarque 6.6.4. Attention, **l'espérance d'une variable aléatoire réelle conditionnée par une tribu est une nouvelle variable aléatoire !**

Désormais, on peut introduire la notion d'espérance conditionnelle conditionnée par une variable aléatoire. On utilise pour cela la notion de tribu engendrée par une variable aléatoire (cf. définition 1.4.3)

Définition 6.6.2. Soit Y une variable aléatoire quelconque. **L'espérance de X conditionnée par la variable aléatoire Y** $\mathbb{E}(X|Y)$ est une variable aléatoire sur $(\Omega, \sigma(Y), \mathbb{P})$, définie par :

$$\mathbb{E}(X|Y) := \mathbb{E}(X|\sigma(Y))$$

Remarque 6.6.5. X doit être une variable aléatoire **réelle**, puisque l'on souhaite calculer son espérance. Cependant, Y peut être quelconque (on ne fait que conditionner par celle-ci).

Proposition 6.6.2. *Soit Y une variable aléatoire quelconque, à valeurs dans un espace topologique E . Il existe $\psi : E \rightarrow \mathbb{R}$ borélienne, telle que $\mathbb{E}(X|Y) = \psi(Y)$: la nouvelle variable aléatoire $\mathbb{E}(X|Y)$ est une fonction de Y .*

Remarque 6.6.6 (Généralisation aux vecteurs aléatoires). L'espérance conditionnée par une sous-tribu et l'espérance conditionnée par une variable aléatoire peuvent se généraliser si X n'est plus une variable aléatoire réelle mais un vecteur aléatoire. Il suffit simplement de faire de même que pour la définition 6.5.1.

Faisons maintenant le lien avec la section précédente. *A priori*, les lois conditionnelles n'ont rien à voir avec les espérances conditionnelles. Cependant, comme les nouvelles lois conditionnelles sont bien des lois de probabilité, on peut définir une nouvelle espérance en les utilisant :

Définition 6.6.3. Soit X et Y deux variables aléatoires réelles (resp. deux vecteurs aléatoires à valeurs dans \mathbb{R}^m et \mathbb{R}^n respectivement) et $y \in \mathbb{R}$ (resp. $y \in \mathbb{R}^n$). **L'espérance conditionnelle de X sachant $Y = y$** existe toujours et est la quantité réelle (resp. vectorielle) :

$$\mathbb{E}(X|Y = y) := \int_{\mathbb{R}} x d\mathbb{P}_{X|Y=y}(x)$$

Remarque 6.6.7. Pour définir cette espérance conditionnelle, il ne suffit pas que X soit une variable aléatoire réelle (ou un vecteur aléatoire). Il est en effet nécessaire que Y soit également une variable aléatoire réelle (ou un vecteur aléatoire), car le théorème d'existence des lois conditionnelles n'a été énoncé que dans ce cas.

Remarque 6.6.8. Attention à ne pas confondre l'espérance conditionnelle de X sachant $Y = y$ et l'espérance conditionnelle de X sachant Y . La première est une quantité réelle (ou vectorielle), alors que la deuxième est une variable aléatoire réelle (ou un vecteur aléatoire).

On peut cependant faire le lien entre les deux, et c'est l'objet de ce qui suit.

Théorème 6.6.5. Soit X et Y deux variables aléatoires réelles (resp. deux vecteurs aléatoires). Une espérance de X conditionnée par Y est donnée par :

$$\omega \mapsto \mathbb{E}(X|Y = Y(\omega))$$

qui est donc une nouvelle variable aléatoire.

Autrement dit, si $N \subset \Omega$ est négligeable, on a :

$$\forall \omega \in \Omega \setminus N, \mathbb{E}(X|Y)(\omega) = \mathbb{E}(X|Y = Y(\omega))$$

Chapitre 7

Convergence des suites de variables aléatoires

Intéressons-nous désormais à la convergence des suites de variables aléatoires. Il n'existe pas une seule notion de convergence pour les variables aléatoires. Certains modes de convergence sont plus forts que d'autres, et il convient de ne pas les confondre.

7.1 Les différents modes de convergence

7.1.1 La convergence en loi

Dans cette section, on se donne une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ quelconques à valeurs dans E . On se donne également une autre variable aléatoire X à valeurs dans E .

Définition 7.1.1. On dit que $(X_n)_{n \in \mathbb{N}}$ **converge en loi** vers X , et on note $X_n \xrightarrow{\mathcal{L}} X$ lorsque pour toute fonction $f : E \rightarrow \mathbb{R}$ continue et bornée,

$$\lim_{n \rightarrow +\infty} \mathbb{E}(f \circ X_n) = \mathbb{E}(f \circ X)$$

Remarque 7.1.1. Remarquons qu'il n'est pas nécessaire d'imposer une quelconque condition sur E , pourvu que l'on puisse définir la continuité d'une fonction qui prend ses valeurs dans E (par exemple, en le munissant d'une topologie).

Remarque 7.1.2. Pour ce mode de convergence, il n'est pas nécessaire que les $(X_n)_{n \in \mathbb{N}}$ et X soient définies sur le même espace probabilisé. En effet, la définition a encore du sens même si :

- les $(X_n)_{n \in \mathbb{N}}$ sont définies sur une suite d'espaces probabilisés $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)_{n \in \mathbb{N}}$;
- X est défini sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$

En réalité, et comme le suggère la propriété suivante, la convergence en loi ne concerne que la convergence des lois \mathbb{P}_{X_n} vers \mathbb{P}_X .

Pour cette proposition uniquement, on considère que les $(X_n)_{n \in \mathbb{N}}$ et que X sont des variables aléatoires réelles ou discrètes (de manière à ce que la fonction de répartition ait un sens).

Proposition 7.1.1. $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X ssi pour tout $x \in \mathbb{R}$ où la fonction de répartition F_X est continue,

$$\lim_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x)$$

Remarque 7.1.3. Comme nous le verrons par la suite, ce mode de convergence est le plus faible de tous.

Remarque 7.1.4. Enfin, ce mode de convergence sera celui impliqué dans le théorème ?? (théorème centrale limite).

7.1.2 La convergence en probabilité

Dans cette section, on se donne une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ réelles ou discrètes. On se donne également une autre variable aléatoire X réelle ou discrète. On suppose que toutes ces variables aléatoires sont définies sur le même espace probabilisé.

Définition 7.1.2. On dit que $(X_n)_{n \in \mathbb{N}}$ **converge en probabilité** vers X , et on note $X_n \xrightarrow{\mathbb{P}} X$ lorsque :

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

Remarque 7.1.5. Remarquons qu'il est nécessaire que les variables aléatoires soient réelles ou discrètes : on doit en effet pouvoir utiliser la relation d'ordre habituelle sur \mathbb{R} .

Remarque 7.1.6. Pour ce mode de convergence, il est absolument nécessaire que les $(X_n)_{n \in \mathbb{N}}$ et X soient définies sur le même espace probabilisé. En effet, soit $n \in \mathbb{N}$: nous sommes alors amenés à considérer la variable aléatoire $X_n - X$. Pour que cette fonction ait du sens, il faut en particulier que l'espace de départ soit le même. De plus, on doit utiliser la loi de $|X_n - X|$: pour que celle-ci existe et ait du sens, il faut que la tribu et la mesure de probabilité de l'espace de départ soient identiques.

Ce mode de convergence est plus fort que la convergence en loi :

Proposition 7.1.2. Supposons que $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X . Alors $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X .

Remarque 7.1.7. Enfin, ce mode de convergence sera celui impliqué dans le théorème ?? (loi faible des grands nombres)

7.1.3 La convergence en norme \mathcal{L}^r

Dans cette section, on se donne une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ réelles ou discrètes. On se donne également une autre variable aléatoire X réelle ou discrète. On suppose que toutes ces variables aléatoires sont définies sur le même espace probabilisé.

Définition 7.1.3. On dit que $(X_n)_{n \in \mathbb{N}}$ **converge en norme \mathcal{L}^r** vers X , et on note $X_n \xrightarrow{\mathcal{L}^r} X$ lorsque :

$$\lim_{n \rightarrow +\infty} \mathcal{E}(|X_n - X|^r) = 0$$

Remarque 7.1.8. Remarquons qu'il est nécessaire que les variables aléatoires soient réelles ou discrètes : on doit en effet pouvoir donner un sens au moment d'ordre r . Néanmoins, si $r = 1$, on peut également étendre la définition aux vecteurs aléatoires.

Remarque 7.1.9. Lorsque $r = 1$, on parle de **convergence en moyenne**. Lorsque $r = 2$, on parle de **convergence en moyenne quadratique**.

Remarque 7.1.10. Pour ce mode de convergence, tout comme pour la convergence en probabilité, il est absolument nécessaire que les $(X_n)_{n \in \mathbb{N}}$ et X soient définies sur le même espace probabilisé. La justification d'ailleurs sensiblement la même que pour la convergence en probabilité : pour calculer l'espérance, on est amenés à considérer la loi de probabilités de $X_n - X$.

Ce mode de convergence est plus fort que la convergence en probabilité, et *a fortiori* plus fort que la convergence en loi :

Proposition 7.1.3. *Supposons que $(X_n)_{n \in \mathbb{N}}$ converge en norme \mathcal{L}^r vers X . Alors $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X .*

Par ailleurs, on a également l'implication suivante :

Proposition 7.1.4. *Supposons que $(X_n)_{n \in \mathbb{N}}$ converge en norme \mathcal{L}^r vers X . Soit s tel que $1 \leq s \leq r$. Alors $(X_n)_{n \in \mathbb{N}}$ converge en norme \mathcal{L}^s vers X .*

Remarque 7.1.11. Enfin, ce mode de convergence sera un des modes de convergence impliqué dans le théorème ?? (loi forte des grands nombres).

7.1.4 La convergence presque-sûre

Ce mode de convergence est celui qui se rapproche le plus de la convergence simple. Il faut cependant garder à l'esprit que c'est loin d'être le seul mode de convergence existant.

Dans cette section, on se donne une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ quelconques à valeurs dans E . On se donne également une autre variable aléatoire X à valeurs dans E . On suppose que toutes ces variables aléatoires sont définies sur le même espace probabilisé.

Définition 7.1.4. On dit que $(X_n)_{n \in \mathbb{N}}$ **converge presque-sûrement** vers X , et on note $X_n \xrightarrow{\text{p.s.}} X$ lorsque pour \mathbb{P} -presque-tout $\omega \in \Omega$,

$$\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)$$

Remarque 7.1.12. Remarquons qu'il n'est pas nécessaire d'imposer une quelconque condition sur E , pourvu que l'on puisse définir la continuité d'une fonction qui est à valeurs dans E (par exemple, en le munissant d'une topologie). Cependant, dans la plupart des cas, les variables aléatoires seront discrètes, réelles, ou seront des vecteurs aléatoires.

Remarque 7.1.13. Pour ce mode de convergence, il est nécessaire que les $(X_n)_{n \in \mathbb{N}}$ et X soient définies sur le même espace probabilisé, puisqu'on ne considère que la loi \mathbb{P} dans le « \mathbb{P} -presque-partout ».

Ce mode de convergence est plus fort que la convergence en probabilité, et *a fortiori* plus fort que la convergence en loi :

Proposition 7.1.5. *Supposons que $(X_n)_{n \in \mathbb{N}}$ converge presque-sûrement vers X . Alors $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X .*

Remarque 7.1.14. Attention, les convergences en norme \mathcal{L}^r et la convergence presque-sûre ne sont pas équivalentes ! De plus, il n'existe pas sans hypothèse supplémentaire d'implications entre ces deux modes de convergence.

Remarque 7.1.15. Enfin, ce mode de convergence sera l'autre mode de convergence impliqué dans le théorème ?? (loi forte des grands nombres).