Probabilités et statistiques

Paul Viallard
(paul.viallard@inria.fr)

2025

Table des matières

Avant-propos

Bibliographie

Ce cours est basé sur celui de François Schwarzentruber, qui est lui-même basé sur différentes références. Des notations inspirées de mes papiers et de ceux de Pascal Germain, François Laviolette, Mario Marchand, et d'Emilie Morvant.

Références pour la théorie des probabilités

Le cours de François, sur lequel ce cours est basé, s'inspire de deux références pour les probabilités : (Garet et al., 2019) et (Barbe et al., 2021). Cependant, ce cours ajoute des éléments sur la théorie de la mesure qui ne proviennent pas de ces livres. De plus, voici deux livres qui sont intéressants (en tout cas, qui m'intéressent particulièrement...) : (Whittle, 2012 ; Shafer et al., 2019).

Références pour l'apprentissage statistique

Pour une compréhension globale de l'apprentissage statistique, plusieurs références peuvent être intéressantes. Parmi les ouvrages présentant les méthodes d'un point de vue pratique, (Bishop et al., 2006) et (Hastie et al., 2009) sont des références incontournables ; il y a également des références plus récentes comme (Murphy, 2022). D'un point de vue théorique, il y a le livre de (Vapnik, 2013), qui jette les bases théoriques des principes fondamentaux du domaine. Enfin, des introductions plus récentes de ces bases théoriques se trouvent dans (Mohri et al., 2012) et (Bach, 2024).

Partie I

Théorie des probabilités

Chapitre 1

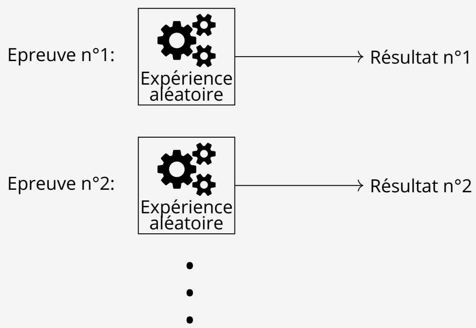
Mesures et espaces probabilisés

1.1 Introduction

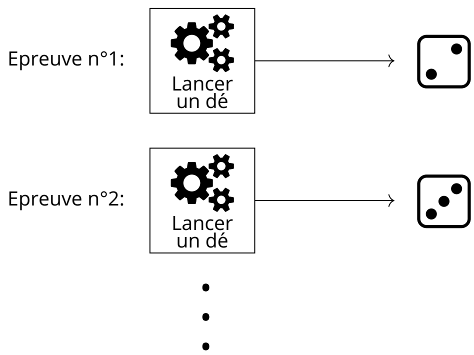
Expérience aléatoire

La théorie des probabilités sert à étudier une expérience aléatoire.

Définition 1.1. Une *expérience aléatoire* est un processus dont le résultat ne peut pas être déterminé à l'avance, même si les conditions initiales sont connues. Autrement dit, le résultat est incertain et dépend du hasard.



Exemple. *Le lancer de dé est une expérience aléatoire :*



Caractérisation d'une expérience aléatoire

Afin de caractériser une expérience aléatoire, nous avons besoin de définir

- de tous les résultats pouvant être obtenus,
- les probabilités associées à chaque résultat pour quantifier la "chance" d'observer un résultat donné.

Exemple. *Pour un lancer de dé, nous avons*

- *l'ensemble des résultats défini par* {⚀, ⚁, ⚂, ⚃, ⚄, ⚅},

- *les probabilités définies par* $\mathbb{P}_{\text{dé}\sim\text{lancer}}\left[\text{dé} = ⚀\right] = \frac{1}{6}$, $\mathbb{P}_{\text{dé}\sim\text{lancer}}\left[\text{dé} = ⚁\right] = \frac{1}{6}$, $\mathbb{P}_{\text{dé}\sim\text{lancer}}\left[\text{dé} = ⚂\right] = \frac{1}{6}$, $\mathbb{P}_{\text{dé}\sim\text{lancer}}\left[\text{dé} = ⚃\right] = \frac{1}{6}$, $\mathbb{P}_{\text{dé}\sim\text{lancer}}\left[\text{dé} = ⚄\right] = \frac{1}{6}$, *et* $\mathbb{P}_{\text{dé}\sim\text{lancer}}\left[\text{dé} = ⚅\right] = \frac{1}{6}$.

Objectif du chapitre

L'objectif de ce chapitre est de définir formellement la notion d'expérience aléatoire. Pour ce faire, nous introduirons les espaces probabilisés (qui formaliseront la notion d'expérience aléatoire) en nous appuyant sur des éléments de théorie de la mesure.

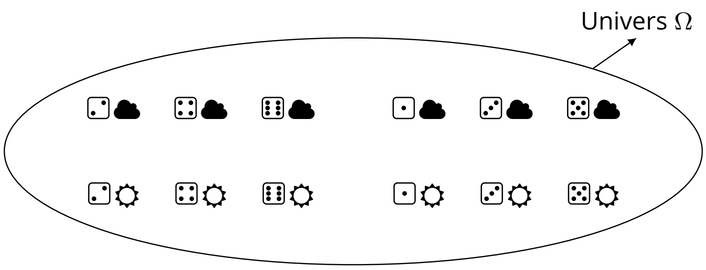
1.2 Elements de la théorie des probabilités

Univers

Afin de caractériser l'ensemble des résultats possibles d'une expérience aléatoire, nous devons définir l'univers.

Définition 1.2. Un *univers* Ω est l'ensemble (non vide) de tous les résultats pouvant être obtenus au cours d'une expérience aléatoire.

Exemple. *Voici un exemple d'univers* Ω *avec 12 résultats possibles :*

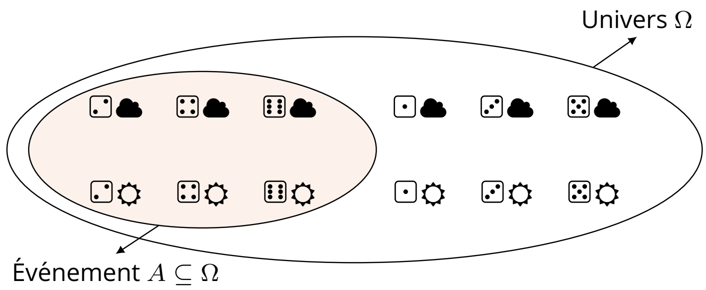


Événement

Parfois, nous pouvons nous intéresser qu'à un sous-ensemble de Ω. Pour cela, nous devons définir un événement.

Définition 1.3. Un *événement* A est un sous-ensemble de Ω désignant un sous-ensemble des résultats possibles d'une expérience aléatoire.

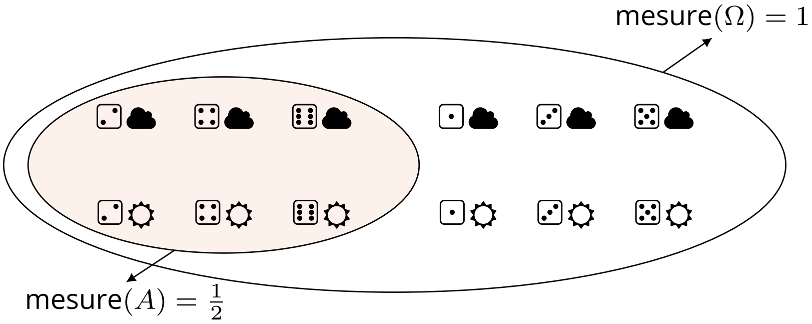
Exemple. *Voici l'événement* $A :=$ *"le dé affiche un nombre pair"*



Pour définir des probabilités...

Pour définir des probabilités, nous allons mesurer la "taille" des événements $A \subseteq \Omega$ à l'aide d'un objet mathématique appelé "mesure". Ainsi, nous verrons que définir une mesure revient à définir des probabilités.

Exemple. *Voici un exemple de mesure sur l'univers* Ω *et l'événement* $A \subseteq \Omega$ *précédent.*



Pour définir une mesure, nous devons également identifier les différents ensembles autorisés à être mesurés.

Exemple. *Nous voulons autoriser que seuls les événements $A \subseteq \Omega$ puissent être mesurés.*

On peut mesurer l'événement $A \subseteq \Omega$ — On ne peut pas mesurer l'ensemble $A \nsubseteq \Omega$

Introduction (rapide) à la théorie de la mesure

Afin de bien définir une mesure, nous allons voir différentes notions issues de la théorie de la mesure.

Tribu (ou σ-algèbre)

Pour définir l'ensemble des ensembles que l'on peut mesurer, nous allons définir une tribu.

Définition 1.4. Une *tribu* Σ_X (ou σ-algèbre) est un ensemble de sous-ensembles de X avec :

1. $\emptyset \in \Sigma_X$;
2. $X \in \Sigma_X$;
3. Σ_X est stable par complémentaire : si $A \in \Sigma_X$, alors $\overline{A} \in \Sigma_X$;
4. Σ_X est stable par union dénombrable : si $\forall n \in \mathbb{N}, A_n \in \Sigma_X$, alors $\cup_{n\in\mathbb{N}} A_n \in \Sigma_X$;
5. Σ_X est stable par intersection dénombrable : si $\forall n \in \mathbb{N}, A_n \in \Sigma_X$, alors $\cap_{n\in\mathbb{N}} A_n \in \Sigma_X$.

Définition 1.5. Si un ensemble X est muni d'une tribu Σ_X, alors (X, Σ_X) est appelé *espace mesurable*.

Mesure

Nous sommes maintenant prêts à définir une mesure : c'est une fonction ν (positive), qui mesure la "taille" d'un ensemble A appartenant à la tribu Σ_X.

Définition 1.6. Une *mesure* est une application ν définie sur la tribu Σ_X et à valeur dans $[0, +\infty]$ qui respecte les deux propriétés suivantes :

- $\nu(\emptyset) = 0$
- Pour toute suite de sous-ensembles $(A_i)_{i\in\mathbb{N}}$ de Σ_X deux à deux disjoints, on a

$$\nu\left(\bigsqcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \nu(A_i).$$

Définition 1.7. Si un ensemble X est muni d'une tribu Σ_X, alors (X, Σ_X, ν) est appelé *espace mesuré*.

La proposition qui suit présente des propriétés naturelles sur un espace mesuré.

Proposition 1.8. Si (X, Σ_X, ν) est un espace mesuré, nous avons les propriétés suivantes :

1. Pour tout $A, B \in \Sigma_X$, $A \cap B = \emptyset$ implique $\nu(A \sqcup B) = \nu(A) + \nu(B)$;
2. Pour tout $A, B \in \Sigma_X$, $\nu(A \cup B) = \nu(A) + \nu(B) - \nu(A \cap B)$;
3. Pour tout $A, B \in \Sigma_X$, $\nu(A \cup B) \leq \nu(A) + \nu(B)$;
4. Pour tout $A, B \in \Sigma_X$, $\nu(A \cap B) \leq \min(\nu(A), \nu(B))$;
5. Pour tout $A, B \in \Sigma_X$, $\nu(A \cup B) \geq \max(\nu(A), \nu(B))$.

1.3 Théorie de la mesure appliquée à la théorie des probabilités

Tribu sur un univers

Pour définir des probabilités, un univers Ω doit être muni d'une tribu.

Définition 1.9. Si un univers Ω est muni d'une tribu Σ_Ω, alors (Ω, Σ_Ω) est appelé *espace probabilisable*.

Définition 1.10. Un événement est un sous-ensemble A qui est contenu dans Σ_Ω.

Mesure de probabilité

Ensuite, une mesure de probabilité est une fonction μ, qui, à tout événement A, associe un nombre $\mu(A)$ entre 0 et 1, qui mesure de combien l'événement est probable.

Définition 1.11. Une *mesure de probabilité* est une application de $\mu : \Sigma_\Omega \to [0, 1]$ vérifiant :

- $\mu(\Omega) = 1$;
- $\mu(\emptyset) = 0$;
- Pour toute suite de sous-ensembles $(A_i)_{i \in \mathbb{N}}$ de Σ_Ω deux à deux disjoints, on a

$$\mu\left(\bigsqcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} \mu(A_i).$$

Plus précisément, $\mu(\emptyset) = 0$ car nous mesurons un événement sans aucun résultat possible. De même, comme Ω est l'ensemble de tous les résultats possibles, nous avons $\mu(\Omega) = 1$. On peut aussi imaginer que Ω est une surface d'aire 1. Alors $\mu(A)$ est l'aire de A.
Nous verrons dans le prochain chapitre que la mesure de probabilité nous permet de définir directement des probabilités. Autrement dit, le nombre $\mu(A)$ est la probabilité que A soit vraie.

Proposition 1.12. Soit $(\Omega, \Sigma_\Omega, \mu)$ un espace probabilisé. Nous avons

$$\mu(A) = \mathop{\mathbb{P}}_{\omega \sim \mu}\left[\omega \in A\right].$$

Espace probabilisé

Un espace probabilisé est constitué d'un univers Ω, d'une tribu définie sur cet univers, et d'une mesure de probabilité.

Définition 1.13. Un *espace probabilisé* est $(\Omega, \Sigma_\Omega, \mu)$ est la donnée :

- d'un espace Ω ;
- d'une tribu Σ_Ω sur Ω ;
- d'une mesure de probabilité $\mu : \Sigma_\Omega \to [0, 1]$.

En plus des propriétés de Proposition 1.8, nous avons la propriété suivante pour un espace probabilisé.

Proposition 1.14. Si $(\Omega, \Sigma_\Omega, \mu)$ est un espace probabilisé, nous avons :

1. Pour tout $A \in \Sigma_\Omega$, $\mu(\overline{A}) = 1 - \mu(A)$.

Voici deux exemples simples d'espaces probabilisés où Ω est un ensemble fini.

Exemple (Lancer d'un dé)**.**

- $\Omega = \{1, 2, 3, 4, 5, 6\}$;
- $\Sigma_\Omega = 2^{\{1,2,3,4,5,6\}}$;
- *Pour tout* $A \subseteq \Omega$, *nous avons* $\mu(A) = \frac{|A|}{6}$.

Exemple (Météo)**.**

- $\Omega = \{☼, ☁\}$;
- $\Sigma_\Omega = 2^{\{☼, ☁\}}$;
- $\mu(\emptyset) = 0$, $\mu(\Omega) = 1$, $\mu(\{☼\}) = 0.2$, $\mu(\{☁\}) = 0.8$.

Par contre, lorsque l'on veut définir une mesure de probabilité sur un intervalle, tout se complique ... En effet, nous pouvons prouver le théorème suivant.

Théorème 1.15. Soit $\Omega = [0, 1]$ et $\Sigma_\Omega = 2^\Omega$, alors il n'existe pas d'espace probabilisé $(\Omega, \Sigma_\Omega, \mu)$ tel que $\mu([a, b]) = b - a$ avec $0 \leq a \leq b \leq 1$.

1.4 Tribu borélienne et mesure de Lebesgue

Autrement dit, nous ne pouvons pas définir de mesure de probabilité μ naturelle sur un intervalle (où sa mesure correspond à la longueur de l'intervalle) si la tribu est l'ensemble des parties de Ω.

Tribu engendrée et tribu borélienne

Pour réussir à définir une mesure μ sur $[0, 1]$, il faut utiliser une autre tribu : la tribu borélienne.
Pour la construire, il faut supposer que

- $[a, b]$ avec $a, b \in [0, 1]$ est dans *Borel*$([a, b])$.

Puis, *Borel*$([a, b])$ est stable par complémentaire, union, et intersection dénombrables. Nous avons, par exemple

- $[0, 1] \setminus [a, b]$ est dans *Borel*$([0, 1])$;
- $\bigcup_{i \in \mathbb{N}} [a_i, b_i]$ est dans *Borel*$([0, 1])$ avec $0 \leq a_i \leq b_i \leq 1$;
- $\bigcap_{i \in \mathbb{N}} [a_i, b_i]$ est dans *Borel*$([0, 1])$ avec $0 \leq a_i \leq b_i \leq 1$;
- $\{a\} = [a, a]$ est dans *Borel*$([0, 1])$;
- Tout ensemble dénombrable $X = \bigcup_{x \in X} \{x\}$ est dans *Borel*$([0, 1])$.

Les deux définitions suivantes formalisent exactement cette construction.

Définition 1.16. Soit A un ensemble de sous-ensembles de Ω. La *tribu engendrée* par A est la plus petite tribu contenant A.

Définition 1.17. La *tribu borélienne Borel*$([0, 1])$ est la tribu engendrée par l'ensemble $\{[a, b] \mid a, b \in [0, 1]\}$.

Tribu borélienne sur un ensemble $F \subseteq \mathbb{R}^d$

Nous allons voir que nous pouvons étendre la notions de tribu borélienne à un ensemble $F \subseteq \mathbb{R}^d$, avec $d \in \mathbb{N}^*$. Ceci nous servira plus tard dans le cours.

Définition 1.18. La *tribu borélienne Borel*(F) où $F \subseteq \mathbb{R}^d$, avec $d \in \mathbb{N}^*$, est la tribu engendrée par l'ensemble

$$\left\{ \prod_{i=1}^{d} [a_i, b_i] \;\middle|\; \prod_{i=1}^{d} [a_i, b_i] \subseteq F \right\}$$

Mesure de Lebesgue

Nous pouvons définir la mesure de Lebesgue, qui est une mesure sur un produit cartésien d'intervalles provenant de *Borel*(F).

Définition 1.19. Soit I_i un intervalle de la forme $[a_i, b_i]$, $]a_i, b_i]$, $[a_i, b_i[$, ou $]a_i, b_i[$. La *mesure de Lebesgue* ν d'un produit cartésien d'intervalles est donnée par

$$\nu(I_1 \times \cdots \times I_d) := \prod_{i=1}^{d}(b_i - a_i).$$

Espace probabilisé avec une tribu borélienne et une mesure de Lebesgue

Grâce à la tribu borélienne *Borel*$([0, 1])$ et sa mesure de Lebesgue associée, nous pouvons définir un espace probabilisé sur l'interval $[0, 1]$.

Exemple (Loi uniforme sur un intervalle)**.**

- $\Omega = [0, 1]$ *;*
- $\Sigma_\Omega =$ *Borel*$([0, 1])$ *;*
- $\mu([a, b]) = \mu(]a, b]) = \mu([a, b[) = \mu(]a, b[) = b - a$ *avec* $0 \leq a \leq b \leq 1.$

Chapitre 2

De l'intégration aux probabilités

Introduction

Dans le chapitre précédent, nous avons vu avec Proposition 1.12 l'égalité suivante.

Soit $(\Omega, \Sigma_\Omega, \mu)$ un espace probabilisé, alors nous avons

$$\mu(A) = \underset{\omega\sim\mu}{\mathbb{P}}\left[\omega \in A\right].$$

Dans ce chapitre, nous allons démontrer cette proposition.
De plus, nous allons définir des notions essentielles pour caractériser une expérience aléatoire. En effet, étant donné un espace probabilisé $(\Omega, \Sigma_\Omega, \mu)$, nous allons définir une probabilité $\mathbb{P}_{\omega\sim\mu}\left[\cdot\right]$, une espérance $\mathbb{E}_{\omega\sim\mu}\left[\cdot\right]$, et voir certaines de leurs propriétés.
Afin de formaliser ces notions fondamentales en théorie des probabilités, il est indispensable de définir l'intégrale de Lebesgue. Ainsi, nous verrons : *(i)* les fonctions simples, qui servent d'éléments de base pour la définition de l'intégrale, et *(ii)* la notion de fonction mesurable, c'est-à-dire une fonction pour laquelle cette intégrale peut être correctement définie.

2.1 Fonction simple

Fonction simple

Pour définir l'intégrale de Lebesgue, nous avons besoin de définir un type de fonction particulier.

Définition 2.1. Soit (X, Σ_X) un espace mesurable. Soient $a_1, \dots, a_n \in \mathbb{R}$ et $A_1, \dots, A_n \in \Sigma_X$ des ensembles mesurables deux à deux disjoints. Une *fonction simple* $f : X \to \mathbb{R}$ est définie par

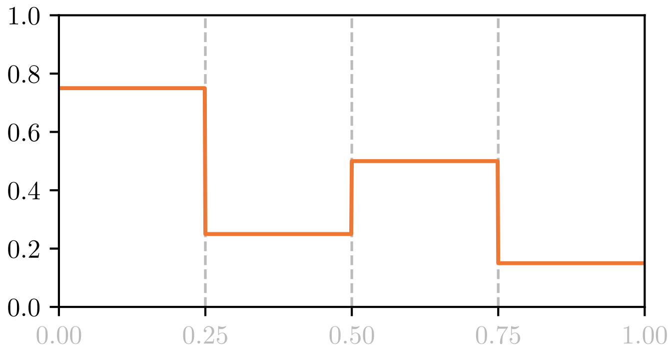
$$f(x) = \sum_{k=1}^{n} a_k \mathbb{1}[x \in A_k],$$

où $\mathbb{1}[a] = 1$ si la proposition a est vraie et 0 sinon.

En d'autres termes, une fonction simple prend un nombre fini de valeurs réelles. De plus, pour tout élément $x \in A_k$, la fonction simple donne la valeur associée à l'ensemble A_k, c'est-à-dire que nous avons $f(x) = a_k$.
Voici un exemple de fonction simple, ainsi que sa représentation graphique.

Exemple.

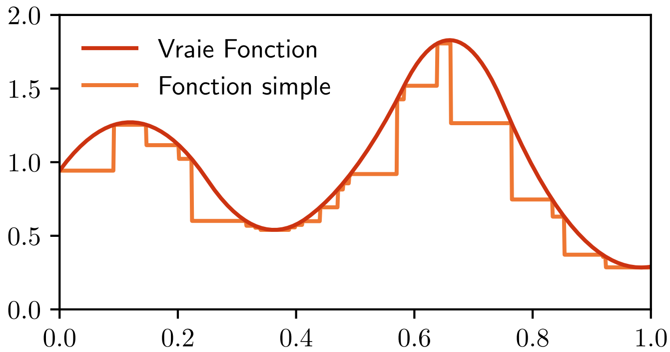
$$\begin{aligned} f(x) = {} & 0.75 \cdot \mathbb{1}\Big[x \in [0.0, 0.25[\Big] + 0.25 \cdot \mathbb{1}\Big[x \in [0.25, 0.5[\Big] \\ & + 0.5 \cdot \mathbb{1}\Big[x \in [0.5, 0.75[\Big] + 0.15 \cdot \mathbb{1}\Big[x \in [0.75, 1.0]\Big] \end{aligned}$$



Approximation par fonctions simples

En pratique, une fonction simple permet d'approcher une fonction (plus compliquée).

Exemple.



Cette approximation est justifiée par un théorème que nous allons voir plus tard dans ce chapitre. Ainsi, le théorème suivant n'est que sa version informelle.

Théorème (informel !)**.** Pour un certain type de fonction $f : X \to \mathbb{R}$, il existe une séquence de fonctions simples $\{f_n\}_{n\in\mathbb{N}}$ telle que

$$\text{pour tout } x \in X, \quad \text{nous avons} \quad \lim_{n\to\infty} f_n(x) = f(x).$$

En réalité, il permet d'approcher des fonctions mesurables (que nous allons voir ensuite).

2.2 Fonction mesurable

Fonction mesurable

L'intégrale de Lebesgue permet d'intégrer un type de fonction, appelé fonction mesurable, que nous allons voir dans la définition suivante.

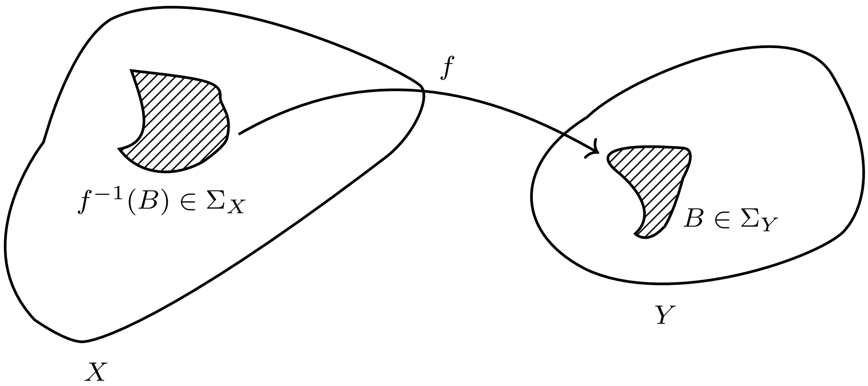
Définition 2.2. Soient deux espaces (X, Σ_X) et (Y, Σ_Y) mesurables. Une fonction $f : X \to Y$ est dite *mesurable* si nous avons

$$\text{pour tout } B \in \Sigma_Y, \quad \text{nous avons} \quad f^{-1}(B) \in \Sigma_X,$$

$$\text{où} \quad f^{-1}(B) = \{x \in X \mid f(x) \in B\}.$$

Voici une illustration pour clarifier cette définition.

Exemple.



Cette définition garantit que la fonction f "transporte" les ensembles mesurables de Y (c'est-à-dire $B \in \Sigma_Y$) vers des ensembles mesurables de X (c'est-à-dire $f^{-1}(B) \in \Sigma_X$). Autrement dit, elle assure que f préserve la structure mesurable des espaces X et Y, ce qui permet d'étudier ses propriétés à travers l'intégration.

Propriétés des fonctions mesurables

Si l'espace Y est restreint à l'ensemble des réels $\mathbb{R}$, nous pouvons trouver une caractérisation (plus naturelle) des fonctions mesurables.

Théorème 2.3. Soient (X, Σ_X) un espace mesurable et une fonction $f : X \to \mathbb{R}$. Les propositions suivantes sont équivalentes

1. f est une fonction mesurable,
2. pour tout $\alpha \in \mathbb{R}$, nous avons $\{x \in X \mid f(x) \leq \alpha\} \in \Sigma_X$,
3. pour tout $\alpha \in \mathbb{R}$, nous avons $\{x \in X \mid f(x) < \alpha\} \in \Sigma_X$,
4. pour tout $\alpha \in \mathbb{R}$, nous avons $\{x \in X \mid f(x) \geq \alpha\} \in \Sigma_X$,
5. pour tout $\alpha \in \mathbb{R}$, nous avons $\{x \in X \mid f(x) > \alpha\} \in \Sigma_X$.

En effet, pour une fonction $f : X \to \mathbb{R}$ qui est mesurable, il suffit de vérifier que les ensembles où la fonction est supérieure ou inférieure à un certain seuil α sont mesurable, c'est-à-dire qu'ils appartiennent à la tribu Σ_X. Autrement dit, cette propriété reflète le fait que f "transporte" les ensembles mesurables de $\mathbb{R}$ (comme les intervalles $]-\infty, \alpha]$ ou $[\alpha, +\infty[$) vers des ensembles mesurables de X.
Grâce à cette propriété, nous pouvons prouver que les fonctions simples sont des fonctions mesurables.

Théorème 2.4. Soit f une fonction simple, alors f est une fonction mesurable.

De plus, une caractérisation importante des fonctions mesurables est qu'elles peuvent être approchées par des fonctions simples. En effet, nous avons le théorème suivant.

Théorème 2.5. Soit une fonction mesurable $f : X \to \mathbb{R}$, alors il existe une séquence de fonctions simples $\{f_n\}_{n \in \mathbb{N}}$ telle que

$$\text{pour tout } x \in X, \quad \text{nous avons} \quad \lim_{n \to \infty} f_n(x) = f(x).$$

Autrement dit, nous pouvons approcher n'importe quelle fonction mesurable f à l'aide d'une séquence $\{f_n\}_{n \in \mathbb{N}}$ de fonctions simples.

2.3 Intégrale de Lebesgue

Maintenant, nous pouvons définir l'intégrale de Lebesgue pour les fonctions mesurables. Nous allons tout d'abord nous concentrer sur les fonctions simples (qui sont des fonctions mesurables).

Définition 2.6. Soit un espace mesuré (X, Σ_X, ν), soit une fonction simple $f : X \to \mathbb{R}$ où

$$\text{pour tout} \quad x \in X, \quad \text{nous avons} \quad f(x) = \sum_{k=1}^{n} a_k \mathbb{1}[x \in A_k].$$

L'*intégrale de Lebesgue pour une fonction simple* f est définie par

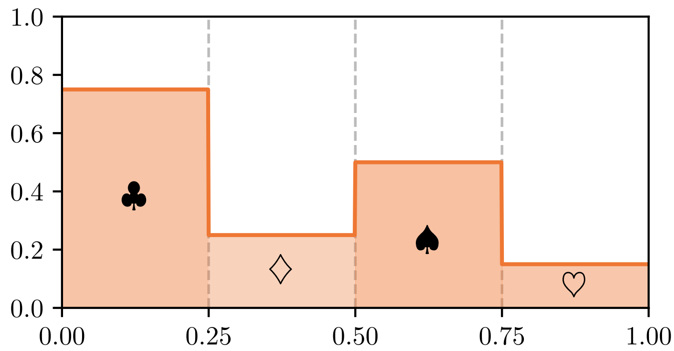
$$\int_{\mathcal{X}} f(x) d\mu(x) := \sum_{k=1}^{n} a_k \nu(A_k).$$

De façon non formelle, lorsque nous avons une fonction simple $f : X \to \mathbb{R}$, l'intégrale de Lebesgue revient à sommer les aires des rectangles, avec pour "longueur" a_k et "largeur" $\nu(A_k)$.
Voici un exemple d'intégration avec la fonction simple précédente.

Exemple.

$$\int_{\mathcal{X}} f(x) d\mu(x) = \underbrace{0.75 \cdot \nu([0.0, 0.25[)}_{\clubsuit} + \underbrace{0.25 \cdot \nu([0.25, 0.5[)}_{\diamondsuit} + \underbrace{0.5 \cdot \nu([0.5, 0.75[)}_{\spadesuit} + \underbrace{0.15 \cdot \nu([0.75, 1.0])}_{\heartsuit}$$

(où ν est la mesure de Lebesgue)



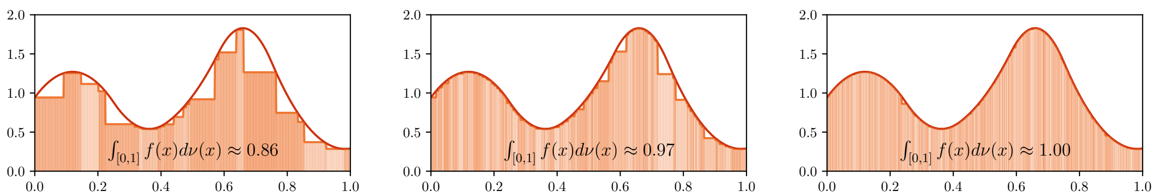
Maintenant que nous avons une intégrale pour les fonctions simples, nous pouvons les définir pour des fonctions plus générales qui sont les fonctions mesurables positives.

Définition 2.7. Soit un espace mesuré (X, Σ_X, ν), l'*intégrale de Lebesgue pour une fonction* $f : X \to \mathbb{R}^+$ *mesurable (positive)* est définie par

$$\int_X f(x) d\nu(x) := \sup \left\{ \int_X s(x) d\nu(x) \;\middle|\; s \text{ fonction simple avec } 0 \leq s(x) \leq f(x) \text{ pour tout } x \in X \right\}.$$

En d'autres termes, l'intégrale de Lebesgue d'une fonction mesurable positive f est obtenue en approximant f par des fonctions simples s telles que $0 \leq s(x) \leq f(x)$ pour tout $x \in X$. L'intégrale de f est alors définie comme la borne supérieure des intégrales de ces fonctions simples s.
Voici une illustration de l'intégration d'une fonction mesurable utilisée dans un exemple précédent.

Exemple.



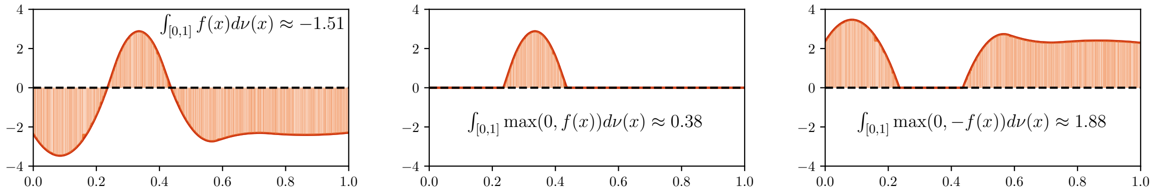
Malheureusement, l'intégrale est définie pour des fonctions mesurables positives. Ainsi, nous pouvons étendre la définition de l'intégrale de Lebesgue aux fonctions mesurables.

Définition 2.8. Soit un espace mesuré (X, Σ_X, ν), l'*intégrale de Lebesgue pour une fonction $f : X \to \mathbb{R}$ mesurable* est définie par

$$\int_X f(x) d\nu(x) := \int_X \max(0, f(x)) d\nu(x) - \int_X \max(0, -f(x)) d\nu(x).$$

L'idée est d'utiliser deux intégrales de Lebesgue sur deux fonctions mesurables positives $x \mapsto \max(0, f(x))$ et $x \mapsto \max(0, -f(x))$ et de soustraire la valeur de l'intégrale associée à $x \mapsto \max(0, -f(x))$.
Voici un exemple d'intégration pour une fonction mesurable.

Exemple.



Après avoir défini l'intégrale de Lebesgue pour une fonction mesurable, il est naturel de préciser la condition sous laquelle l'intégration se déroule correctement, c'est-à-dire lorsque l'intégrale prend une valeur finie.

Définition 2.9. Soit un espace mesuré (X, Σ_X, ν) et une fonction $f : X \to \mathbb{R}$ mesurable. Si

$$\int_X |f(x)| d\nu(x) < +\infty$$

alors la fonction f est dite *intégrable* (au sens de Lebesgue).

Nous pouvons également dériver les propriétés suivantes pour l'intégrale de Lebesgue.

Théorème 2.10. Soient un espace mesuré (X, Σ_X, ν) et $f : X \to \mathbb{R}$ et $g : X \to \mathbb{R}$ des fonctions intégrables, alors

1. pour tout $c \in \mathbb{R}$,

$$\int_X (c \cdot f(x))\, d\nu(x) = c \cdot \int_X f(x) d\nu(x),$$

2. nous avons

$$\int_X (f(x) + g(x))\, d\nu(x) = \int_X f(x) d\nu(x) + \int_X g(x) d\nu(x),$$

3. nous avons

$$\int_X \mathbb{1}[x \in A] f(x) d\nu(x) = \int_A f(x) d\nu(x).$$

2.4 Espérance et probabilité

Espérance

Grâce à l'intégrale de Lebesgue, nous pouvons maintenant définir l'espérance.

Définition 2.11. Soit un espace probabilisé $(\Omega, \Sigma_\Omega, \mu)$, l'espérance de $\omega \mapsto f(\omega)$ est définie par

$$\mathop{\mathbb{E}}_{\omega \sim \mu} \left[f(\omega) \right] := \int_\Omega f(\omega) d\mu(\omega).$$

Grâce à la définition de l'espérance, nous pouvons dériver les propriétés suivantes qui sont directement issues de l'intégrale de Lebesgue.

Corollaire 2.12. Soient un espace probabilisé $(\Omega, \Sigma_\Omega, \mu)$ et $f, g : \Omega \to \mathbb{R}$ des fonctions intégrables, alors

1. pour tout $c \in \mathbb{R}$, nous avons

$$\mathop{\mathbb{E}}_{\omega \sim \mu} [c \cdot f(\omega)] = c \cdot \mathop{\mathbb{E}}_{\omega \sim \mu} [f(\omega)]$$

2. pour tout $c \in \mathbb{R}$, nous avons

$$\mathop{\mathbb{E}}_{\omega \sim \mu} [f(\omega) + g(\omega)] = \mathop{\mathbb{E}}_{\omega \sim \mu} [f(\omega)] + \mathop{\mathbb{E}}_{\omega \sim \mu} [g(\omega)]$$

Probabilité

De plus, grâce à l'espérance, nous pouvons également définir formellement une probabilité.

Définition 2.13. Soit un espace probabilisé $(\Omega, \Sigma_\Omega, \mu)$, la probabilité que $\omega \in A$ où $\omega \sim \mu$ est définie par

$$\mathop{\mathbb{P}}_{\omega \sim \mu} [\omega \in A] := \mathop{\mathbb{E}}_{\omega \sim \mu} \mathbb{1} [\omega \in A]$$

Enfin, nous pouvons relier une probabilité à sa mesure (grâce à la définition d'une probabilité et de l'espérance). Autrement dit, nous sommes maintenant prêts à prouver Proposition 1.12.

Soit $(\Omega, \Sigma_\Omega, \mu)$ un espace probabilisé. Nous avons

$$\mu(A) = \mathop{\mathbb{P}}_{\omega \sim \mu} [\omega \in A] .$$

Démonstration. Grâce aux Définitions 2.13 et 2.11, nous avons

$$\begin{aligned}
\mathop{\mathbb{P}}_{\omega \sim \mu} [\omega \in A] &= \mathop{\mathbb{E}}_{\omega \sim \mu} \mathbb{1} [\omega \in A] \\
&= \int_\Omega \mathbb{1} [\omega \in A] \, d\mu(\omega).
\end{aligned}$$

De plus grâce à Théorème 2.10 (propriété 3) et Définition 2.6, nous avons

$$\begin{aligned}
\mathop{\mathbb{P}}_{\omega \sim \mu} [\omega \in A] &= \int_\Omega \mathbb{1} [\omega \in A] \, d\mu(\omega) \\
&= \int_A 1 d\mu(\omega) \\
&= \mu(A).
\end{aligned}$$

$\square$

Références

Christopher M. Bishop et Nasser M. Nasrabadi. Pattern recognition and machine learning. *Springer*. (2006).

Trevor Hastie, Robert Tibshirani, Jerome H Friedman et Jerome H Friedman. The elements of statistical learning : data mining, inference, and prediction. *Springer*. (2009).

Mehryar Mohri, Afshin Rostamizadeh et Ameet Talwalkar. Foundations of Machine Learning. *MIT Press*. (2012).

Peter Whittle. Probability via expectation. *Springer Science & Business Media*. (2012).

Vladimir Vapnik. The nature of statistical learning theory. *Springer science & business media*. (2013).

Olivier Garet et Aline Kurtzmann. De l'intégration aux probabilités-2e édition augmentée. *Editions Ellipses*. (2019).

Glenn Shafer et Vladimir Vovk. Game-theoretic foundations for probability and finance. *John Wiley & Sons*. (2019).

Philippe Barbe et Michel Ledoux. Probabilité. *Probabilité. EDP Sciences*. (2021).

Kevin P. Murphy. Probabilistic machine learning : an introduction. *MIT press*. (2022).

Francis Bach. Learning theory from first principles. *MIT press*. (2024).