

# Probabilités et statistiques

Paul Viallard  
([paul.viallard@inria.fr](mailto:paul.viallard@inria.fr))

2025



# Table des matières

<b>I</b>	<b>Théorie des probabilités</b>	<b>7</b>
<b>1</b>	<b>Mesures et espaces probabilisés</b>	<b>9</b>
1.1	Introduction . . . . .	9
1.2	Elements de la théorie des probabilités . . . . .	10
1.3	Théorie de la mesure appliquée à la théorie des probabilités . . . . .	12
1.4	Tribu borélienne et mesure de Lebesgue . . . . .	13



# Avant-propos

## Bibliographie

Ce cours est basé sur celui de [François Schwarzenrubler](#), 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

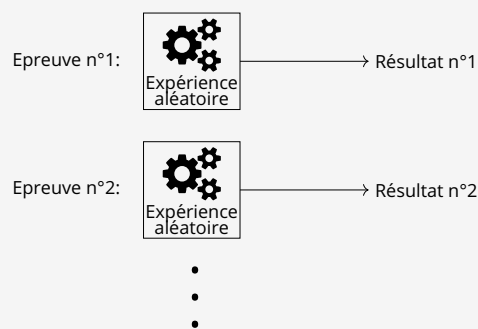
## Mesures et espaces probabilisés

### 1.1 Introduction

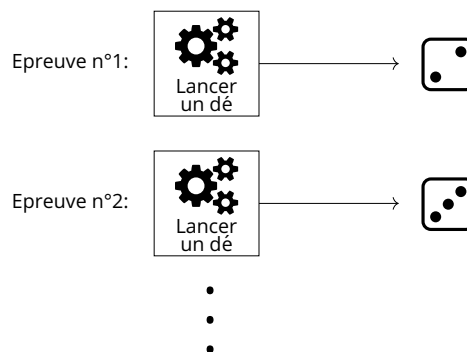
#### Expérience aléatoire

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  $\{\square, \begin{smallmatrix} \square \\ \square \end{smallmatrix}, \begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix}, \begin{smallmatrix} \square & \square & \square \\ \square & \square & \square \end{smallmatrix}, \begin{smallmatrix} \square & \square & \square & \square \\ \square & \square & \square & \square \end{smallmatrix}\}$ ,

- les probabilités définies par  $\mathbb{P}_{\text{dé} \sim \text{lancer}} [\text{dé} = \square] = \frac{1}{6}$ ,  $\mathbb{P}_{\text{dé} \sim \text{lancer}} [\text{dé} = \square] = \frac{1}{6}$ ,  $\mathbb{P}_{\text{dé} \sim \text{lancer}} [\text{dé} = \square] = \frac{1}{6}$ ,  
 $\mathbb{P}_{\text{dé} \sim \text{lancer}} [\text{dé} = \boxtimes] = \frac{1}{6}$ ,  $\mathbb{P}_{\text{dé} \sim \text{lancer}} [\text{dé} = \boxtimes] = \frac{1}{6}$ , et  $\mathbb{P}_{\text{dé} \sim \text{lancer}} [\text{dé} = \boxplus] = \frac{1}{6}$ .

## 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.

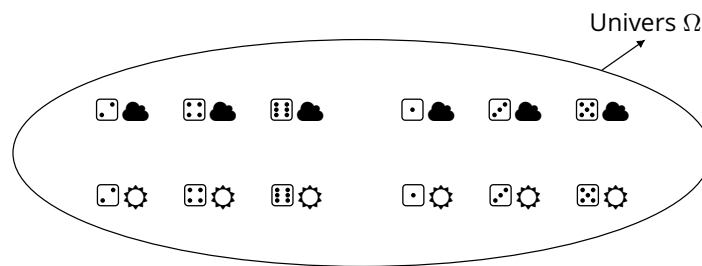
## 1.2 Elements de la théorie des probabilités

### Univers

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

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

**Exemple.** Voici un exemple d'univers  $\Omega$  avec 12 résultats possibles :

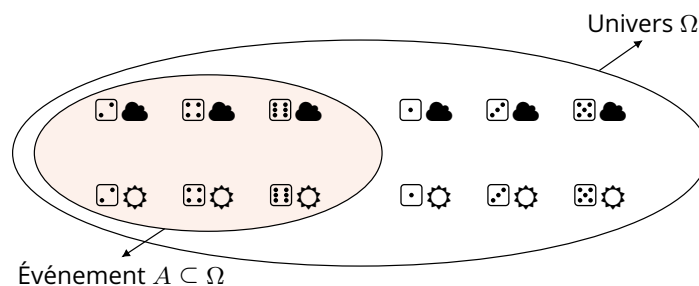


### Événement

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

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

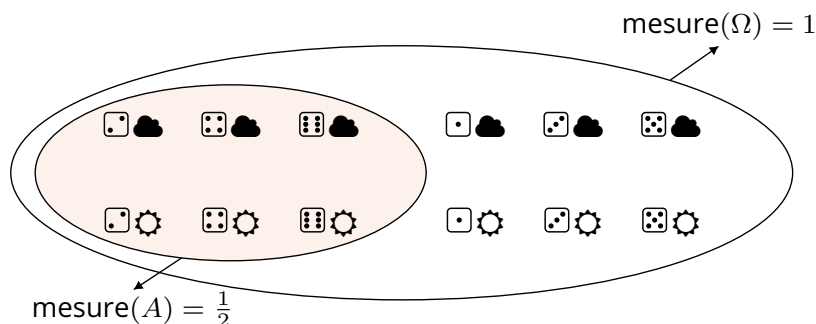
**Exemple.** Voici l'événement  $A := \text{"le dé affiche un nombre pair"}$



### 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.

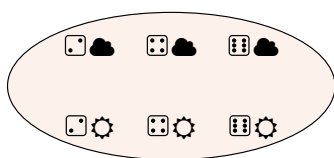
**Exemple.** Voici un exemple de mesure sur l'univers  $\Omega$  et l'événement  $A \subseteq \Omega$  précédent.



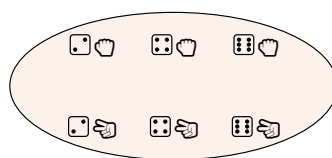
Pour définir une mesure, nous devons également identifier les différents ensembles autorisés à être mesurés.

**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  $A \not\subseteq \Omega$



## 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 $\sigma$ -algèbre)

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

**Définition 1.4.** Une *tribu*  $\Sigma_X$  (ou  $\sigma$ -algèbre) est un ensemble de sous-ensembles de  $X$  avec :

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

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

### Mesure

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

**Définition 1.6.** Une *mesure* est une application  $\nu$  définie sur la tribu  $\Sigma_X$  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  $\Sigma_X$  deux à deux disjoints, on a

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

**Définition 1.7.** Si un ensemble  $X$  est muni d'une tribu  $\Sigma_X$ , alors  $(X, \Sigma_X, \nu)$  est appelé *espace mesuré*.

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

**Proposition 1.8.** Si  $(X, \Sigma_X, \nu)$  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  $\Omega$  doit être muni d'une tribu.

**Définition 1.9.** Si un univers  $\Omega$  est muni d'une tribu  $\Sigma_\Omega$ , alors  $(\Omega, \Sigma_\Omega)$  est appelé *espace probabilisable*.

**Définition 1.10.** Un événement est un sous-ensemble  $A$  qui est contenu dans  $\Sigma_\Omega$ .

### Mesure de probabilité

Ensuite, une mesure de probabilité est une fonction  $\mu$ , 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  $\mu : \Sigma_\Omega \rightarrow [0, 1]$  vérifiant :

- $\mu(\Omega) = 1$  ;
- $\mu(\emptyset) = 0$  ;
- Pour toute suite de sous-ensembles  $(A_i)_{i \in \mathbb{N}}$  de  $\Sigma_\Omega$  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  $\Omega$  est l'ensemble de tous les résultats possibles, nous avons  $\mu(\Omega) = 1$ . On peut aussi imaginer que  $\Omega$  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, alors nous avons

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

### Espace probabilisé

Un espace probabilisé est constitué d'un univers  $\Omega$ , 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  $\Omega$  ;
- d'une tribu  $\Sigma_\Omega$  sur  $\Omega$  ;
- d'une mesure de probabilité  $\mu : \Sigma_\Omega \rightarrow [0, 1]$ .

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

Voici deux exemples simples d'espaces probabilisés où  $\Omega$  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).**

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

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  $\mu$  naturelle sur un intervalle (où sa mesure correspond à la longueur de l'intervalle) si la tribu est l'ensemble des parties de  $\Omega$ .

**Tribu engendrée et tribu borélienne**

Pour réussir à définir une mesure  $\mu$  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  $\Omega$ . 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

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

### 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*  $\nu$  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'intervalle  $[0, 1]$ .

**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  $0 \leq a \leq b \leq 1$ .

## 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).