

# M61B : Probabilités et intégration, *parcours classique* – 18h

Clément Erignoux \*

*Pour l'écriture de ces notes de cours, je me suis particulièrement inspiré du livre **Mesures, Intégration et Probabilités**, de Thierry Gallouët et Raphaële Herbin, disponible sur HAL, <https://hal.archives-ouvertes.fr/hal-01283567>, dont je remercie les auteurs. Pour plus d'exercices, d'exemples, et d'approfondissements, je vous invite donc à consulter ce dernier. Les résultats et notions clés de ce cours sont indiqués par une ★.*

## Table des matières

<b>1</b>	<b>Théorie de la mesure</b>	<b>5</b>
1.1	Tribus . . . . .	5
1.2	Fonctions mesurables . . . . .	8
1.3	Mesures . . . . .	11
<b>2</b>	<b>Variables aléatoires</b>	<b>16</b>
2.1	Loi d'une variable aléatoire . . . . .	16
2.2	Rappels : variables aléatoires discrètes . . . . .	17
2.3	Rappels : quelques lois discrètes classiques . . . . .	21
2.4	Intégrale et intégrabilité sur $\mathbb{R}$ . . . . .	22
2.5	Variables aléatoires réelles à densité . . . . .	24
2.6	Fonction de répartition . . . . .	26
2.7	Quelques lois classiques à densité continue par morceaux . . . . .	29
<b>3</b>	<b>Intégrale par rapport à une mesure <math>\sigma</math>-finie.</b>	<b>34</b>
3.1	Définition de l'intégrale d'une fonction mesurable . . . . .	34
3.2	Convergence d'intégrales . . . . .	38

---

\*Pour toute typo/question, me contacter à [clement.erignoux@inria.fr](mailto:clement.erignoux@inria.fr). Les notes de cours sont disponibles sur ma page web, <http://chercheurs.lille.inria.fr/cerignou/homepage.html>

<b>4</b>	<b>Variables aléatoires réelles</b>	<b>43</b>
4.1	Préambule : caractérisation de mesures . . . . .	43
4.2	Variable aléatoire réelles et fonction de répartition . . . . .	44
4.3	Moments d'une variable aléatoire réelle . . . . .	46
4.4	Bonus : les variables discrètes, variables réelles comme les autres . . . . .	50
4.5	Propriétés et inégalités classiques . . . . .	51
<b>5</b>	<b>Indépendance, loi des grands nombres et TCL</b>	<b>54</b>
5.1	Indépendance de variables aléatoires . . . . .	54
5.2	Convergence des variables aléatoires réelles . . . . .	55
5.3	Loi des grands nombres . . . . .	57
5.4	Théorème central limite et statistiques descriptives . . . . .	58
<b>6</b>	<b>Bonus : vecteurs aléatoires et suites de variables aléatoires</b>	<b>62</b>
6.1	Espaces mesurés produits et Théorème de Fubini . . . . .	62
6.2	Vecteurs aléatoires . . . . .	64

## Introduction : pourquoi une nouvelle intégrale ?

La construction de la théorie des probabilités sur des ensembles continus (ou, en tout cas, non dénombrables) nécessite l'introduction d'une nouvelle notion d'intégrale. Cette intégrale est l'*intégrale de Lebesgue*, qui repose elle-même sur la *théorie de la mesure*, dont nous allons voir ensemble quelques éléments. Cette théorie va bien au-delà de ce que nous allons voir en cours, puisque nous ne traiterons que les éléments nécessaires à l'introduction des probabilités continues. Avant tout, toutefois, commençons par nous poser la question : pourquoi une nouvelle intégrale pour compléter la notion, plus élémentaire, d'intégrale de Riemann ?

Commençons par quelques rappels sur l'intégrale de Riemann. Étant donnée une fonction sur un segment, par exemple  $[0, 1]$ , on définit pour toute suite finie  $0 \leq x_1 < x_2 < \dots < x_n \leq 1$  la somme de Riemann correspondante

$$\sum_{i=1}^{n-1} f(x_i)(x_{i+1} - x_i).$$

En pratique, on pourra par exemple considérer la suite  $x_i = i/n$ , dont la somme de Riemann est donnée par  $\frac{1}{n} \sum_{i=1}^n f(x_i)$ . Une fonction est Riemann intégrable si, en faisant tendre le *pas*  $\max\{x_{i+1} - x_i\}$  vers 0, la somme de Riemann converge, vers une quantité qui est alors l'intégrale (de Riemann) de la fonction sur le segment  $[0, 1]$ .

Alors, pourquoi y a-t-il besoin d'une autre construction de l'intégrale ? Un premier élément de réponse vient de l'espace des fonctions considéré : L'intégrale de Riemann est très limitée, puisque sa construction ne la rend exploitable que pour les fonctions *continues* (ou plus rigoureusement presque-partout continues) définies sur un segment. La théorie de la mesure substitue aux fonctions continues les fonctions *mesurables*, qui est une notion beaucoup plus générale.

Prenons par exemple l'ensemble des nombres rationnels  $\mathbb{Q}$ , et considérons la fonction indicatrice  $\mathbf{1}_{\mathbb{Q}}$ , définie sur  $\mathbb{R}$  par

$$\mathbf{1}_{\mathbb{Q}}(x) = \begin{cases} 1 & \text{si } x \in \mathbb{Q} \\ 0 & \text{si } x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$$

Cette fonction n'est *continue* en aucun point de  $\mathbb{R}$  puisque tout réel peut être approché par une suite de nombres rationnels, et n'est donc pas *intégrable* sous la théorie de l'intégrale de Riemann. Toutefois, cette fonction est *mesurable*, et donc *intégrable*, dans le contexte de l'intégrale de Lebesgue.

La notion de *convergence* est une autre raison naturelle pour l'emploi de l'intégrale de Lebesgue. Considérons l'ensemble  $E = C([0, 1], \mathbb{R})$  des fonctions continues sur  $[0, 1]$  à valeurs dans  $\mathbb{R}$ . Étant donnée une suite  $(f_n)_{n \in \mathbb{N}} \in E^{\mathbb{N}}$  de fonctions continues, on veut pouvoir formaliser mathématiquement la notion de convergence sous l'intégrale, c'est-à-dire trouver des conditions sur la suite  $(f_n)_n$  permettant d'affirmer que

$$\left[ f_n \xrightarrow[n \rightarrow \infty]{} f \right] \implies \left[ \int_{[0,1]} f_n(x) dx \xrightarrow[n \rightarrow \infty]{} \int_{[0,1]} f(x) dx \right].$$

L'intégrale de Riemann n'offre qu'un résultat faible de convergence, qui repose sur la *convergence uniforme* :

### Théorème 1 : Théorème de convergence uniforme

Supposons qu'une suite  $(f_n)_{n \in \mathbb{N}} \in E^{\mathbb{N}}$  converge uniformément vers une fonction  $f \in E$ , c'est-à-dire que  $\max_{x \in [0,1]} |f_n(x) - f(x)| \xrightarrow[n \rightarrow \infty]{} 0$ . Alors,

$$\int_{[0,1]} f_n(x) dx \xrightarrow[n \rightarrow \infty]{} \int_{[0,1]} f(x) dx.$$

En pratique, ce théorème est très faible car l'hypothèse de convergence uniforme est extrêmement forte, et ne peut donc souvent pas être obtenue. L'intégrale de Lebesgue donne par exemple accès au *théorème de convergence dominée*, beaucoup plus utile, car il repose sur des hypothèses plus faibles :

### Théorème 2 : Théorème de convergence dominée

Supposons qu'une suite  $(f_n)_{n \in \mathbb{N}} \in E^{\mathbb{N}}$  converge simplement vers  $f \in E$ , c'est-à-dire que pour tout  $x \in [0, 1]$ ,  $f_n(x) \xrightarrow[n \rightarrow \infty]{} f(x)$ . Supposons également que les  $f_n$  sont uniformément bornées, i.e. qu'il existe une constante  $C$  telle que pour tout  $x \in [0, 1]$ ,  $n \in \mathbb{N}$ , on ait  $|f_n(x)| \leq C$ . Alors,

$$\int_{[0,1]} f_n(x) dx \xrightarrow[n \rightarrow \infty]{} \int_{[0,1]} f(x) dx.$$

Pour illustrer la différence entre ces deux résultats, considérons pour  $n \geq 1$  la suite de fonctions  $f_n$ , représentées en Figure 1,

$$f_n(x) = \begin{cases} 2nx & \text{pour } x \in [0, 1/2n] \\ 2 - 2nx & \text{pour } x \in [1/2n, 1/n] \\ 0 & \text{pour } x \in [1/n, 1]. \end{cases}$$

On remarque facilement que pour tout  $x \in [0, 1]$ ,  $f_n(x) \rightarrow f(x) := 0$  quand  $n \rightarrow \infty$ . Pour cette suite de fonctions, le *théorème de convergence uniforme* ne s'applique pas, car  $\max_{x \in [0,1]} |f_n(x) - f(x)| = 1$  pour tout  $n \geq 1$ .

En revanche, on peut appliquer le *théorème de convergence dominée* avec  $C = 1$ , et on obtient alors  $\int_{[0,1]} f_n(x) dx \xrightarrow{n \rightarrow \infty} \int_{[0,1]} f(x) dx = 0$ , ce qui est vrai puisque  $\int_{[0,1]} f_n(x) dx = 1/2n$ .

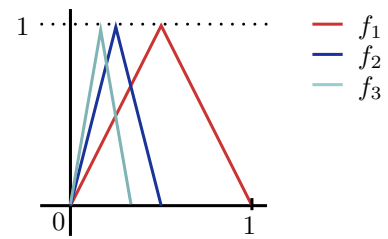


FIGURE 1

# 1 Théorie de la mesure

## 1.1 Tribus

Afin de construire la mesure de Lebesgue, puis l'intégrale de Lebesgue mentionnée ci-dessus, et plus généralement les probabilités continues, nous allons définir la notion d'ensembles mesurables. En effet, en prenant par exemple  $\mathbb{R}^n$  comme espace de travail, on veut choisir un sous-ensemble  $\mathcal{T} \subset \mathcal{P}(\mathbb{R}^n)$  de parties de  $\mathbb{R}^n$  dont la mesure sera bien définie.

Or, il s'avère que, pour que notre mesure respecte bien les propriétés de  $\sigma$ -additivité que l'on attend (cf. Définition 5 ci-dessous), on ne peut pas choisir  $\mathcal{T} = \mathcal{P}(\mathbb{R}^n)$  (c'est-à-dire que toute partie de  $\mathbb{R}^n$  aurait une mesure), puisque cet espace est trop gros et fait apparaître des contradictions dans la construction de notre mesure. Les curieux pourront par exemple se renseigner sur le paradoxe de Banach-Tarski : sous ce paradoxe, et en admettant que l'on arrive à construire la mesure de Lebesgue sur toutes les parties de  $\mathbb{R}^n$ , on montrerait alors que  $1 = 2$ . Ce qui est un peu problématique.

### Definition 1 : Tribu, espace mesurable



Soit  $E$  un ensemble (on peut penser par exemple à  $E = \mathbb{R}^n$ ) et  $\mathcal{T} \subset \mathcal{P}(E)$  un ensemble de parties de  $E$ . La famille  $\mathcal{T}$  est une *tribu*, ou  $\sigma$ -algèbre, sur  $E$  si  $\mathcal{T}$  vérifie

- $\emptyset, E \in \mathcal{T}$ ,
- $\mathcal{T}$  est stable par union et intersection dénombrable, c'est-à-dire que si  $A_n \in \mathcal{T}$  pour tout  $n \in \mathbb{N}$ , alors  $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{T}$  et  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{T}$ ,
- $\mathcal{T}$  est stable par passage au complémentaire, c'est-à-dire que si  $A \in \mathcal{T}$ , alors  $A^c := E \setminus A \in \mathcal{T}$ .

Un couple  $(E, \mathcal{T})$ , où  $\mathcal{T}$  est une tribu sur  $E$ , est appelé un *espace mesurable*.

Pourquoi cette définition, et pourquoi de telles propriétés? Comme on le verra plus tard dans le cours, une tribu sera l'espace de départ d'une mesure de probabilité, c'est à dire que les éléments de la tribu seront des ensemble dont la probabilités est bien définie. Or si la probabilité d'un "événement" (un élément de la tribu) est bien définie, on veut que la probabilité que cet événement *n'arrive pas* soit bien définie également, d'où la troisième propriété. De la même manière, si une famille (dénombrable) d'événements ont tous des probabilités bien définies, on veut que la probabilité que tous ces événements arrivent simultanément (l'intersection, donc) soit également bien définie. Par passage au complémentaire, la propriété de stabilité par union en découle également naturellement.

**EXEMPLE** : La tribu dite *grossière*  $\mathcal{T} = \{\emptyset, E\}$  et la tribu *discrète*  $\mathcal{T} = \mathcal{P}(E)$  sont toujours des tribus (pas très intéressantes).

**REMARQUE 1** : En pratique, on n'aura pas à vérifier à la fois la stabilité par union et celle par intersection, une seule des deux suffit, puisque l'on peut passer d'une

union à une intersection en passant au complémentaire :  $\cup_{n \in \mathbb{N}} A_n = (\cap_{n \in \mathbb{N}} A_n^c)^c$ .

### Exercice 1 : tribu image réciproque

Soit  $E$  et  $F$  deux ensembles,  $\mathcal{T}$  une tribu sur  $F$  et  $\phi : E \rightarrow F$  une application. Montrer que  $\mathcal{T}' = \{\phi^{-1}(A) \mid A \in \mathcal{T}\}$  est une tribu sur  $E$ .

**SOLUTION :** On commence par vérifier que  $\emptyset$  et  $E$  sont dans  $\mathcal{T}'$ . C'est vrai car  $\emptyset$  et  $F$  sont dans  $\mathcal{T}$  (car c'est une tribu), et par ailleurs  $E = \phi^{-1}(F)$ , et  $\emptyset = \phi^{-1}(\emptyset)$ . Soit maintenant  $(A_n)_{n \in \mathbb{N}}$  une suite d'éléments de  $\mathcal{T}'$ , par définition il doit exister pour tout  $n$  un élément  $B_n \in \mathcal{T}$  tel que  $A_n = \phi^{-1}(B_n)$ . Or  $\mathcal{T}$  est une tribu, par conséquent on doit avoir  $\cup_{n \in \mathbb{N}} B_n \in \mathcal{T}$ . Or

$$\cup_{n \in \mathbb{N}} A_n = \cup_{n \in \mathbb{N}} \phi^{-1}(B_n) = \phi^{-1}(\cup_{n \in \mathbb{N}} B_n) \in \mathcal{T}'.$$

Par conséquent,  $\mathcal{T}'$  est stable par union dénombrable. Pour le complémentaire, on procède de même, puisque  $E \setminus \phi^{-1}(A) = \phi^{-1}(F \setminus A)$ .  $\square$

### Proposition 3 : Intersection de tribus

Soient  $I$  un ensemble non vide (pas nécessairement dénombrable), et une famille  $(\mathcal{T}_i)_{i \in I}$  de tribus. Alors,

$$\bigcap_{i \in I} \mathcal{T}_i = \{A \in \mathcal{P}(E) ; A \in \mathcal{T}_i \forall i \in I\}$$

est aussi une tribu.

**PREUVE :** Il s'agit simplement de vérifier les trois propriétés. Comme  $\emptyset$  et  $E$  appartiennent à chacun des  $\mathcal{T}_i$ , ils appartiennent également à l'intersection. Par ailleurs, si  $A_j \in \mathcal{T}_i \forall j \in \mathbb{N}$ ,  $i \in I$ , alors, comme les  $\mathcal{T}_i$  sont des tribus,  $\cup_j A_j \in \mathcal{T}_i$  pour tout  $i$  et donc  $\cup_j A_j \in \bigcap_{i \in I} \mathcal{T}_i$ . Il en va de même pour les intersections dénombrables et pour le passage au complémentaire.  $\square$

Ce résultat permet de définir la tribu engendrée par un ensemble de parties de  $E$ .

### Definition 2 : Tribu engendrée par un ensemble de parties

Soit  $C \subset \mathcal{P}(E)$  un ensemble de parties de  $E$ . On appelle *tribu engendrée par  $C$*  la plus petite tribu contenant  $C$ , notée  $\sigma(C)$  définie par

$$\sigma(C) = \bigcap \{ \mathcal{T} \subset \mathcal{P}(E) ; \mathcal{T} \text{ tribu contenant } C \}.$$

À noter que le membre de droite est bien défini, puisque la tribu discrète  $\mathcal{T} = \mathcal{P}(E)$  contient  $C$ , et donc l'intersection est non vide.

Pour  $E = \mathbb{R}$ , on définit une tribu particulière sur laquelle va être construite la mesure de Lebesgue.

**Definition 3 : tribu borélienne sur  $\mathbb{R}$**  ★

On appelle *tribu borélienne* sur  $\mathbb{R}$  la tribu, notée  $\mathcal{B}(\mathbb{R})$ , engendrée par les intervalles ouverts de  $\mathbb{R}$ . Autrement dit, étant donné

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

on définit

$$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{I}) = \bigcap \{ \mathcal{T} \subset \mathcal{P}(\mathbb{R}); \mathcal{T} \text{ tribu contenant } \mathcal{I} \}.$$

On appelle *boréliens* les éléments de la tribu borélienne.

**REMARQUE 2 :** Définissons

$$\mathcal{I}_1 = \{]a, b[, a \in \mathbb{R}, b \in \mathbb{R}, a \leq b\},$$

$$\mathcal{I}_2 = \{[a, b], a \in \mathbb{R}, b \in \mathbb{R}, a \leq b\},$$

$$\mathcal{I}_3 = \{]a, b[, a \in \mathbb{Q}, b \in \mathbb{Q}, a \leq b\}.$$

$$\mathcal{I}_4 = \{]-\infty, a[, a \in \mathbb{R}\}.$$

On peut alors facilement montrer que

$$\mathcal{B}(\mathbb{R}) := \sigma(\mathcal{I}) = \sigma(\mathcal{I}_1) = \sigma(\mathcal{I}_2) = \sigma(\mathcal{I}_3) = \sigma(\mathcal{I}_4).$$

En effet, Il est immédiat que  $\mathcal{B}(\mathbb{R})$  contient  $\sigma(\mathcal{I}_1)$  et  $\sigma(\mathcal{I}_3)$ , car  $\mathcal{B}(\mathbb{R})$  est une tribu contenant les intervalles ouverts, et donc en particulier elle contient les intervalles finis ouverts et les intervalles dont les extrémités sont rationnelles. Par ailleurs, pour tout intervalle  $]a, b[ \in \mathcal{I}$ , il existe une suite monotone de rationnels  $(a_n)$  (resp.  $(b_n)$ ) qui décroît vers  $a$  (resp. croît vers  $b$ ), et on peut donc écrire

$$]a, b[ = \bigcup_{n \in \mathbb{N}} ]a_n, b_n[.$$

Or pour tout  $n$ ,  $]a_n, b_n[ \in \sigma(\mathcal{I}_3)$ . donc comme  $\sigma(\mathcal{I}_3)$  est une tribu,  $]a, b[ \in \sigma(\mathcal{I}_3)$ . On en déduit que  $\sigma(\mathcal{I}_3)$  est une tribu contenant tous les intervalles ouverts, et par conséquent qu'elle contient la tribu borélienne, ce qui prouve l'égalité. On peut montrer par le même type de raisonnement les autres égalités (À faire en exercice).

Pour  $\mathcal{I}_4$ , il suffit de remarquer que  $]a, b[ = ]-\infty, b[ \cap ]-\infty, a[^c$ , ce qui permet ensuite d'écrire

$$]a, b[ = \bigcup_{n \in \mathbb{N}} ]a + 1/n, b[.$$

On en déduit facilement que  $\sigma(\mathcal{I}_4) = \sigma(\mathcal{I})$ .

**REMARQUE 3 :** La tribu borélienne peut en réalité être définie pour tout espace topologique  $E$ . Elle est alors définie comme la plus petite tribu contenant tous

les ouverts de la topologie de  $E$  (cf. Exercice 4). En particulier, on retiendra les exemples fondamentaux suivants.

1. Étant donné un intervalle  $\Gamma \subset \mathbb{R}$ , on peut définir **la tribu borélienne sur  $\Gamma$** , notée  $\mathcal{B}(\Gamma)$ , qui est engendrée par les intervalles de  $\Gamma$ . Cette dernière est également donnée par

$$\mathcal{B}(\Gamma) = \{A \cap \Gamma, A \in \mathcal{B}(\mathbb{R})\},$$

ce qui est complètement équivalent.

2. Pour  $d \geq 1$ , **la tribu des boréliens sur  $\mathbb{R}^d$** , qui est par définition engendrée par les ouverts, mais qui est également engendrée, plus simplement, par les boules (Euclidiennes) ouvertes

$$B_r(x) := \{y; \|x - y\|_2 < r\},$$

et par les cubes ouverts,

$$C_r(x) := \{y; \|x - y\|_\infty < r\},$$

c'est-à-dire que

$$\mathcal{B}(\mathbb{R}^d) = \sigma(\{B_r(x), x \in \mathbb{R}^d, r > 0\}) = \sigma(\{C_r(x), x \in \mathbb{R}^d, r > 0\}).$$

3. Enfin, on rappelle que  $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$  est la fermeture de  $\mathbb{R}$ , on peut définir **la tribu des boréliens sur  $\overline{\mathbb{R}}$**  comme la tribu engendrée par les intervalles  $[a, b]$ , avec  $a, b \in \overline{\mathbb{R}}$ .

### Exercice 2

Démontrer que  $A = \{x \in \mathbb{R} / \exists n \in \mathbb{N} \setminus \{0\}, |x - n| < \frac{1}{n}\}$  est un borélien de  $\mathbb{R}$ .  
Même question si l'inégalité est large.

**SOLUTION :** On définit

$$A_n = \left]n - \frac{1}{n}, n + \frac{1}{n}\right[.$$

$A_n$  est un intervalle ouvert et donc un Borélien. On a alors  $A = \cup_{n \in \mathbb{N}} A_n$ , et comme la tribu des boréliens est stable par union dénombrable,  $A$  est un borélien.  $\square$

————— Fin du cours 1 —————

## 1.2 Fonctions mesurables

Maintenant munis de la notion de *tribus*, nous allons pouvoir définir une classe de fonctions beaucoup plus générales que les fonctions *continues*, sur laquelle l'intégrale de Lebesgue sera construite.



#### Definition 4 : Fonction mesurable



Soient deux espaces mesurables  $(E, \mathcal{T})$  et  $(E', \mathcal{T}')$ . Une fonction  $f : E \rightarrow E'$  est dite *mesurable* (ou, plus précisément,  $(\mathcal{T}, \mathcal{T}')$ -mesurable) si pour tout  $A \in \mathcal{T}'$ , on a

$$f^{-1}(A) := \{x \in E \mid f(x) \in A\} \in \mathcal{T}.$$

Si la tribu d'arrivée  $\mathcal{T}'$  est la tribu borélienne, on dira que la fonction est *borel-mesurable*, ou *borélienne*.

**REMARQUE 4 :** La mesurabilité, tant pour les fonctions que pour les ensembles, est une propriété extrêmement stable : en pratique, il est très difficile de produire des fonctions réelles non-mesurables, parce qu'il est très difficile de produire des ensembles non-mesurables. Par ailleurs, comme vous le verrez en exercices, la plupart des opérations auxquelles on pourrait penser sur les fonctions mesurables, somme, produit, composition, limites, liminf, limsup, etc. préservent toutes la mesurabilité.

**REMARQUE 5 :** Attention, le fait qu'une fonction soit ou non mesurable dépend profondément des tribus dont sont munis son espace de départ et son espace d'arrivée. Par exemple, on peut facilement montrer que si l'espace d'arrivée  $F$  est muni de la tribu grossière  $\mathcal{T}' = \{\emptyset, F\}$ , alors toutes les fonctions sont mesurables (cf Exercice 3 ci-dessous). En particulier, on peut facilement se convaincre qu'étant donnée une application  $\phi : E \rightarrow (F, \mathcal{T}')$ , la tribu image réciproque définie à l'Exercice 1 est la plus petite tribu sur  $E$  rendant  $\phi$  mesurable.

#### Exercice 3 : Tribus de départ et d'arrivées

Soient  $(E, \mathcal{T})$  et  $(F, \mathcal{T}')$  deux espaces mesurables.

- 1) Montrer que si  $\mathcal{T}' = \{\emptyset, F\}$  est la tribu grossière, alors toute application  $f : E \rightarrow F$  est mesurable.
- 2) Montrer que si  $\mathcal{T} = \mathcal{P}(E)$  est la tribu discrète, alors toute application  $f : E \rightarrow F$  est mesurable.
- 3) Supposons que  $\mathcal{T}'$  contient les singletons, i.e.  $\forall x \in F, \{x\} \in \mathcal{T}'$ . Montrer que si  $\mathcal{T} = \{\emptyset, E\}$ , les seules applications mesurables sont les applications constantes,  $f \equiv x$ .

#### SOLUTION :

1) Il suffit de vérifier que pour tout ensemble mesurable de la tribu d'arrivée, leur image réciproque est mesurable. Comme  $\mathcal{T}'$  est la tribu grossière, il suffit donc de vérifier que  $f^{-1}(\emptyset) = \emptyset$  et  $f^{-1}(F) = E$  sont des ensembles mesurables, ce qui est le cas puisque  $\mathcal{T}$  est une tribu sur  $E$  et contient en particulier  $\emptyset$  et  $E$ . Toutes les applications de  $(E, \mathcal{T})$  dans  $(F, \mathcal{T}')$  sont donc mesurables.

2) Là encore, il suffit de vérifier que pour tout ensemble mesurable de la tribu d'arrivée, leur image réciproque est mesurable. Or pour tout  $A \in \mathcal{T}'$ ,  $f^{-1}(A) \in \mathcal{P}(E) = \mathcal{T}$ ,  $f$  est donc mesurable.

3) Soit  $x \in f(E)$ , il existe  $e \in E$  tel que  $f(e) = x$ . Supposons que  $f$  est mesurable, en

particulier  $f^{-1}(\{x\})$  est mesurable et contient  $e$ . Or les seuls ensembles mesurables sont  $\emptyset$  et  $E$ , par conséquent  $f^{-1}(\{x\}) = E$ , et  $f$  est donc constante égale à  $x$ .  $\square$

#### Exercice 4 : Mesurabilité des fonctions continues

Soit  $f : \mathbb{R} \rightarrow \mathbb{R}$  une fonction continue. On munit  $\mathbb{R}$ , à l'arrivée comme au départ, de la tribu borélienne. On définit

$$\mathcal{G} = \{B \in \mathcal{B}(\mathbb{R}), f^{-1}(B) \in \mathcal{B}(\mathbb{R})\}.$$

À noter que  $\mathcal{G}$  n'est pas la tribu image réciproque de  $f$ , puisque  $\mathcal{G}$  est un ensemble de partie de l'espace d'arrivée de  $f$ , alors que la tribu image réciproque est une tribu sur son espace de départ.

- 1) Montrer que  $\mathcal{G}$  est une tribu.
- 2) Montrer que tout ouvert  $O \subset \mathbb{R}$  est un borélien, i.e.  $O \in \mathcal{B}(\mathbb{R})$ . Indice : on pourra considérer les  $I_{x,r} = ]x - r, x + r[$  pour  $x \in \mathbb{Q}$  et  $r \in \mathbb{Q}_+$ .
- 3) Montrer que  $\mathcal{G}$  contient les intervalles.
- 4) En déduire que  $\mathcal{G} = \mathcal{B}(\mathbb{R})$ , et en conclure que toute fonction continue est mesurable.

#### SOLUTION :

1) On commence par remarquer que  $f^{-1}(\emptyset) = \emptyset$  et  $f^{-1}(\mathbb{R}) = \mathbb{R}$ , donc  $\emptyset, \mathbb{R} \in \mathcal{G}$ . Par ailleurs, si  $A_n \in \mathcal{G} \forall n \in \mathbb{N}$ , alors  $f^{-1}(\cup A_n) = \cup f^{-1}(A_n) \in \mathcal{B}(\mathbb{R})$ . Enfin,  $f^{-1}(A)^c = f^{-1}(A^c)$  donc si  $A \in \mathcal{G}$ , on a également  $A^c \in \mathcal{G}$ .  $\mathcal{G}$  est donc une tribu.

2) Soit  $O$  un ouvert de  $\mathbb{R}$ , pour tout  $x \in \mathbb{Q}$  on définit  $\mathcal{Q}_x = \{r \in \mathbb{Q}_+, ]x - r, x + r[ \subset O\}$ . On peut alors écrire

$$O = \bigcup_{x \in \mathbb{Q} \cap O} \bigcup_{r \in \mathcal{Q}_x} ]x - r, x + r[,$$

qui est une union dénombrable d'intervalles, et est donc un borélien.

3) On rappelle que par définition (topologique) une fonction est continue si et seulement si son image réciproque d'un ouvert est un ouvert. Soit  $I$  un intervalle (ouvert),  $f$  étant continue,  $f^{-1}(I)$  est un ouvert de  $\mathbb{R}$ , et est donc borélien d'après la question précédente. On en déduit que  $I \in \mathcal{G}$ .

4)  $\mathcal{G}$  étant une tribu qui contient les intervalles, par définition elle contient la tribu Borélienne, qui est engendrée par les intervalles. Or par construction,  $\mathcal{G} \subset \mathcal{B}(\mathbb{R})$ , donc  $\mathcal{G} = \mathcal{B}(\mathbb{R})$ . Cela signifie que pour tout  $B \in \mathcal{B}(\mathbb{R})$ ,  $f^{-1}(B) \in \mathcal{B}(\mathbb{R})$ , donc  $f$  est une fonction mesurable.  $\square$

#### Proposition 4 : Mesurabilité et tribu engendrée

Soient  $(E, \mathcal{T})$  un espace mesurable,  $C$  un ensemble de parties de  $F$  et  $\sigma(C) \subset \mathcal{P}(F)$  sa tribu engendrée. Soit  $f : (E, \mathcal{T}) \rightarrow (F, \sigma(C))$  une application telle que  $f^{-1}(C) \subset \mathcal{T}$ . Alors,  $f$  est mesurable.

Concrètement, cette proposition donne une stratégie pour montrer qu'une application est mesurable : lorsque la tribu d'arrivée est la tribu borélienne, par exemple, cette proposition montre que pour que  $f$  soit borélienne, il suffit que  $f^{-1}(]-\infty[, a) := \{e \in E, f(e) < a\}$  soit mesurable pour tout  $a \in \mathbb{R}$ . En effet, d'après la Remarque 2, la tribu borélienne est engendrée par la classe  $\mathcal{I}_4 := \{]-\infty, a[, a \in \mathbb{R}\}$ .

**PREUVE** : La stratégie est similaire à celle de l'exercice 4. Soit  $\mathcal{T}'$  la tribu

$$\mathcal{T}' := \{A \in \sigma(C), f^{-1}(A) \in \mathcal{T}\}.$$

En d'autres termes,  $\mathcal{T}'$  est l'ensemble des éléments de  $\sigma(C)$  dont l'image réciproque par  $f$  est mesurable. Notre objectif est de montrer que  $\mathcal{T}' = \sigma(C)$  tout entier, ce qui prouvera que  $f : (E, \mathcal{T}) \rightarrow (F, \sigma(C))$  est mesurable. Il suffit de montrer que

1.  $\mathcal{T}'$  est une tribu, ce que l'on montre facilement en suivant le même raisonnement que dans l'exercice 4.
2.  $\mathcal{T}'$  contient  $C$ , ce qui est vrai par hypothèse.

Il en découle que  $\mathcal{T}'$  contient  $\sigma(C)$ , et comme l'inclusion réciproque est immédiate, que  $\mathcal{T}' = \sigma(C)$ . □

### Proposition 5 : Mesurabilité et opérations classiques

La somme, le produit, la composition, la limite, le supremum, etc. de fonction mesurable est mesurable. Par exemple, si  $f$  et  $g$  sont deux fonctions mesurables  $(\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , alors  $h := f + g$  est également mesurable.

**PREUVE** : On ne montre que le dernier exemple, vous pouvez vous attaquer aux autres affirmations en exercice. Pour montrer que  $f + g$  est mesurable, on va utiliser la proposition précédente, et montrer simplement que les ensembles

$$\{e \in E, f(e) + g(e) < a\}$$

sont mesurables pour tout  $a \in F$ . Pour cela, il suffit de se convaincre que

$$\{e \in E, f(e) + g(e) < a\} = \bigcup_{q \in \mathbb{Q}} (\{e \in E, f(e) < q\} \cap \{e \in E, g(e) < a - q\}).$$

Le membre de droite est un ensemble mesurable, car  $f$  et  $g$  sont deux applications mesurables, ce qui prouve que  $f + g$  l'est aussi. □

## 1.3 Mesures

On s'intéresse maintenant à la notion de mesure positive.

### Definition 5 : Mesure (positive) ★

Soit  $(E, \mathcal{T})$  un espace mesurable, une application  $\mu : \mathcal{T} \rightarrow \mathbb{R}_+ \cup \{\infty\}$  est une *mesure* si les deux conditions suivantes sont satisfaites :

- l'ensemble vide a mesure nulle,  $\mu(\emptyset) = 0$ ,
- $\mu$  est  $\sigma$ -additive, c'est-à-dire que pour toute famille dénombrable  $(A_n)_{n \in \mathbb{N}}$  d'ensembles disjoints deux-à-deux (i.e. tels que  $A_n \cap A_m = \emptyset$  si  $n \neq m$ ), on a

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n).$$

Le triplet  $(E, \mathcal{T}, \mu)$  est alors appelé *espace mesuré*.

**NOTATION** : On notera dans toute la suite du cours par " $\sqcup$ " les unions disjointes, c'est-à-dire que

$$\left[ A = B \sqcup C \right] \iff \left[ A = B \cup C \text{ et } B \cap C = \emptyset \right].$$



**ATTENTION** : il ne faut pas confondre fonction mesurable et mesure ! Les deux sont des fonctions, mais une fonction mesurable prend pour variable un élément  $e \in E$ , alors qu'une mesure prend pour variable un événement  $A \in \mathcal{T}$ .

### Definition 6 : Mesures finies, $\sigma$ -finies, de probabilité

Soit  $(E, \mathcal{T}, \mu)$  un espace mesuré. On dit que la mesure  $\mu$  est

- *finie* si  $\mu(E) < \infty$ .
- $\sigma$ -*finie* si il existe une *partition* dénombrable  $E = \sqcup_{i \geq 1} E_i$ , telle que pour tout  $i \in \mathbb{N}$ ,  $E_i \in \mathcal{T}$  et  $\mu(E_i) < \infty$ .
- une *mesure de probabilité* si  $\mu(E) = 1$ .

**EXEMPLE** : la *mesure de Lebesgue* sur  $\mathbb{R}$ , notée  $\lambda$  est l'unique mesure sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  telle que pour tout  $a \leq b$ , on ait

$$\lambda([a, b[) = b - a.$$

en d'autres termes, la mesure de Lebesgue d'un intervalle est la *longueur* de cet intervalle. Cette mesure est un objet fondamental à la fois en théorie de la mesure et en théorie des probabilités : comme on le verra plus tard dans le cours, elle permet d'étendre la notion d'intégrale au sens de Riemann à toute fonction mesurable (borélienne). La mesure de Lebesgue sur  $\mathbb{R}$  est  $\sigma$ -finie, puisque  $\mathbb{R} = \sqcup [n, n + 1[$ , et la mesure de Lebesgue de chaque intervalle  $[n, n + 1[$  vaut 1.

De manière analogue, la mesure de Lebesgue sur un intervalle fini  $I$  (cf. Remarque 3), est définie sur sa tribu borélienne  $\mathcal{B}(I)$  par  $\lambda_I(A) = \lambda(A \cap I)$ . Enfin, la

mesure de Lebesgue  $\lambda_{\mathbb{R}^d}$  sur  $\mathbb{R}^d$  est caractérisée, pour  $d$  boréliens  $B_1, \dots, B_d \in \mathcal{B}(\mathbb{R})$ , par

$$\lambda_{\mathbb{R}^d}(B_1 \times B_2 \cdots \times B_d) = \prod_{i=1}^d \lambda(B_i).$$

Naturellement, il existe des boréliens de  $\mathbb{R}^d$  qui ne sont pas factorisables sous la forme  $B_1 \times B_2 \cdots \times B_d$ , mais de la même manière qu'il suffit de définir la mesure de Lebesgue sur les intervalles pour la définir sur  $\mathcal{B}(\mathbb{R})$ , il suffit de la définir sur les ensembles factorisables  $B_1 \times B_2 \cdots \times B_d$ , qui engendrent  $\mathcal{B}(\mathbb{R}^d)$ , pour la définir sur  $\mathcal{B}(\mathbb{R}^d)$  tout entière.

**REMARQUE 6 :** L'existence et l'unicité de la mesure de Lebesgue telle que définie ci-dessus n'est *a priori* pas garantie. La démonstration n'est pas à notre programme, et donc nous admettrons donc la "définition" ci-dessus comme la définition de la mesure de Lebesgue (cf paragraphe 4.1).

**EXEMPLE :** Soit  $(E, \mathcal{T})$  un espace mesurable, la *mesure de Dirac* en  $x$ , notée  $\delta_x$ , est la mesure définie par

$$\delta_x(A) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon.} \end{cases}$$

En pratique, la quantité ci-dessus peut être vue soit comme une mesure, soit comme une fonction, selon la variable choisie : soit on fixe  $x$ , et l'application  $\delta_x(A)$  est une mesure puisqu'elle prend en argument un ensemble mesurable, soit on fixe l'ensemble mesurable  $A$ , et  $\delta_x(A)$  s'écrit alors  $\mathbf{1}_A(x)$  qui est la fonction indicatrice de l'ensemble  $A$ , que nous reverrons plus tard (cf. Définition 27).

On finit cette section par quelques notions et résultats utiles sur les mesures et les suites d'événements.

### Proposition 6 : Propriétés générales des mesures

Soit  $(E, \mathcal{T}, \mu)$  un espace mesuré.

1. (Monotonie ★) Si  $A$  et  $B$  sont mesurables et  $A \subset B$ , alors  $\mu(A) \leq \mu(B)$ .
2. (Sous-additivité ★) Pour toute suite d'événements  $A_n \in \mathcal{T}^{\mathbb{N}}$ , on a

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} \mu(A_n).$$

3. (Continuité par rapport aux unions croissantes) Soit  $(A_n)$  une suite croissante (pour l'inclusion) d'ensembles mesurables, i.e.  $A_n \subset A_{n+1} \forall n$ . Alors, on a

$$\lim_{n \rightarrow \infty} \mu(A_n) = \mu\left(\bigcup_{n \in \mathbb{N}} A_n\right).$$

4. (Continuité par rapport aux intersections décroissantes) Par ailleurs, soit  $(A_n)$  une suite décroissante (pour l'inclusion) d'ensembles mesurables, i.e.  $A_{n+1} \subset A_n \forall n$ . Alors, **en supposant qu'il existe  $n_0 \in \mathbb{N}$  tel que  $\mu(A_{n_0}) < \infty$** , on a

$$\lim_{n \rightarrow \infty} \mu(A_n) = \mu\left(\bigcap_{n \in \mathbb{N}} A_n\right).$$

**PREUVE :**

1. On écrit  $\mu(B) = \mu(A) + \mu(B \setminus A)$ , le dernier terme étant positif, l'inégalité suit.
2. On définit  $B_n = A_n \setminus \cup_{k < n} A_k$ . Les  $B_n$  sont disjoints, et  $\sqcup_{n \in \mathbb{N}} B_n = \cup_{n \in \mathbb{N}} A_n$ . Par ailleurs, pour tout  $n$ ,  $B_n \subset A_n$ , et donc par monotonie  $\mu(B_n) \leq \mu(A_n)$ . On obtient

$$\mu(\cup_{n \in \mathbb{N}} A_n) = \mu(\cup_{n \in \mathbb{N}} B_n) = \sum_{n \in \mathbb{N}} \mu(B_n) \leq \sum_{n \in \mathbb{N}} \mu(A_n).$$

3. Soit  $(A_n)$  une suite croissante, on définit la suite d'événements incompatibles  $B_0 = A_0$  et  $B_n = A_n \setminus A_{n-1}$  pour tout  $n \geq 1$ . On a alors, par  $\sigma$ -additivité de la mesure  $\mu$

$$\begin{aligned} \mu(\cup_{n \in \mathbb{N}} A_n) &= \mu(\cup_{n \in \mathbb{N}} B_n) \\ &= \sum_{n \in \mathbb{N}} \mu(B_n) = \lim_{N \rightarrow \infty} \sum_{n \leq N} \mu(B_n) = \lim_{N \rightarrow \infty} \mu(\cup_{n \leq N} B_n) = \lim_{N \rightarrow \infty} \mu(A_N). \end{aligned}$$

4. Soit  $(A_n)$  une suite décroissante, on définit la suite d'événements croissante  $B_n = A_{n_0} \setminus A_{n_0+n}$ , à laquelle on peut appliquer l'identité précédente.

$$\begin{aligned} \mu(A_{n_0}) - \lim_{n \rightarrow \infty} \mu(A_n) &= \lim_{n \rightarrow \infty} \mu(B_n) \\ &= \mu(\cup_{n \in \mathbb{N}} B_n) = \mu(A_{n_0} \setminus \cap_{n \geq n_0} A_n) = \mu(A_{n_0}) - \mu(\cap_{n \geq n_0} A_n) = \mu(A_{n_0}) - \mu(\cap_{n \geq 0} A_n). \end{aligned}$$

□

————— Fin du cours 2 —————

### Exercice 5

Trouver un contre-exemple à la seconde affirmation si l'on ne suppose plus qu'un des  $A_n$  est de mesure finie.

**SOLUTION :** On peut considérer la mesure de Lebesgue, et  $A_n = ]n, +\infty[$  : chacun des  $A_n$  a une mesure  $\lambda(A_n)$  infinie, pourtant l'intersection des  $A_n$  est vide, et a donc une mesure nulle. □

### Definition 7 : Limsup et liminf d'ensembles

Soit  $(E, \mathcal{F})$  un espace mesurable. Soit  $(A_n)_{n \in \mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$  une suite d'ensembles mesurables. On définit les ensembles

$$\limsup A_n = \bigcap_{n \geq 0} \bigcup_{p \geq n} A_p \quad \text{et} \quad \liminf A_n = \bigcup_{n \geq 0} \bigcap_{p \geq n} A_p.$$

**REMARQUE 7 :** Comme vous le montrerez en TD, la  $\limsup A_n$  est l'ensemble des éléments de  $\Omega$  qui appartiennent à une infinité de  $A_n$ . La  $\liminf A_n$  est l'ensemble des éléments de  $\Omega$  qui appartiennent à tous les  $A_n$  à partir d'un certain rang. Ces

notions sont analogues aux notions de  $\limsup$  et  $\liminf$  de suites de réels. De la même manière que pour les réels, ces deux limites sont toujours bien définies, même si la suite  $(A_n)$  n'est pas monotone (au sens de l'inclusion)

### Proposition 7 : Lemme de Borel-Cantelli

Soit  $(E, \mathcal{F}, \mu)$  un espace mesuré, alors pour toute suite d'ensembles mesurables  $(A_n)_{n \in \mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$  telle que  $\sum_{n=0}^{\infty} \mu(A_n) < \infty$ , on a

$$\mu(\limsup A_n) = 0.$$

En d'autres termes, si une suite d'événements est de somme de probabilité finie, un nombre fini de ces événements se réalisent avec probabilité 1.

**PREUVE** : On définit  $B_n = \bigcup_{p \geq n} A_p$ . C'est une suite décroissante d'événements, et par hypothèse ils sont tous de mesure finie puisque  $\mu(\bigcup_{n \in \mathbb{N}} A_n) \leq \sum_{n=1}^{\infty} \mu(A_n) < \infty$ . On peut donc appliquer la Propriété 6, on obtient par sous-additivité

$$\mu(\limsup A_n) = \mu(\bigcap_{n \in \mathbb{N}} B_n) = \lim_{n \rightarrow \infty} \mu(B_n) \leq \lim_{n \rightarrow \infty} \sum_{p \geq n} \mu(A_p) = 0,$$

puisque la série  $\sum \mu(A_p)$  est absolument convergente. □

Il existe en réalité un second Lemme de Borel-Cantelli dans le contexte des probabilités, que l'on donnera un peu plus bas (cf. Proposition 10 ci-dessous) pour culture, et dont la preuve sera admise. Le second Lemme, toutefois, requiert la notion *d'indépendance*.

## 2 Variables aléatoires

### 2.1 Loi d'une variable aléatoire

Avec ces éléments de théorie de la mesure, nous pouvons maintenant introduire plus généralement la notion de *variable aléatoire (v.a.)*.

#### Definition 8 : Espace de probabilités, variables aléatoires

★

Un *espace de probabilités* est un espace mesuré  $(\Omega, \mathcal{F}, \mathbb{P})$ , formé d'un ensemble  $\Omega$  appelé *univers*, une tribu  $\mathcal{F}$  sur  $\Omega$  et  $\mathbb{P}$  une mesure de probabilité sur  $(\Omega, \mathcal{F})$ . Une *variable aléatoire* est une fonction *mesurable*  $X : \Omega \rightarrow E$ , où  $(E, \mathcal{T})$  est un *espace mesurable donné*. Dans le contexte de la théorie des probabilités, les éléments de  $\mathcal{F}$  sont appelés *événements*. L'espace  $E$  est alors appelé *espace d'états* de la variable aléatoire  $X$ .

La fonction  $X$  étant mesurable, on peut alors définir sa *distribution*, ou *loi*,  $\mathbb{P}_X$ , comme la mesure sur  $(E, \mathcal{T})$  définie pour tout ensemble mesurable  $A \in \mathcal{T}$  par

$$\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)).$$

Dans toute la suite, on fixe un espace de probabilités  $(\Omega, \mathcal{F}, \mathbb{P})$  donné.

#### Exercice 6

Montrer que, étant donnée une variable aléatoire  $X$ , l'application  $\mathbb{P}_X : A \in \mathcal{T} \mapsto \mathbb{P}(X^{-1}(A))$  est bien une mesure, et qui plus est une mesure de probabilité, sur  $(E, \mathcal{T})$ .

**SOLUTION** : On commence par remarquer  $X^{-1}(\emptyset) = \emptyset$  et donc  $\mathbb{P}_X(\emptyset) = \mathbb{P}(\emptyset) = 0$  car  $\mathbb{P}$  est une mesure. Par ailleurs, on a pour toute famille disjointe d'événements  $X^{-1}(\sqcup_k A_k) = \sqcup_k X^{-1}(A_k)$ , et donc la  $\sigma$ -additivité de  $\mathbb{P}_X$  découle de celle de  $\mathbb{P}$  :

$$\mathbb{P}_X(\sqcup_k A_k) = \mathbb{P}(\sqcup_k X^{-1}(A_k)) = \sum_k \mathbb{P}_X(A_k).$$

L'application  $\mathbb{P}_X$  est donc une mesure. C'est une mesure de probabilité car  $\mathbb{P}$  en est une, et  $X^{-1}(E) = \Omega$ .  $\square$

**NOTATION** : On notera  $\{X \in A\}$  l'événement  $\{X \in A\} := X^{-1}(A) \in \mathcal{F}$ , et par conséquent

$$\mathbb{P}(X \in A) = \mathbb{P}_X(A).$$

Dans le contexte de deux variables aléatoires  $X$  et  $Y$ , on notera également

$$\{X \in A, Y \in B\} := \{X \in A\} \cap \{Y \in B\}.$$

En pratique, l'espace  $\Omega$  n'est pas accessible, car il est gigantesque : on n'y a accès que par des observations, qui sont les *variables aléatoires*. Prenons l'exemple



d'un lancer de dé : chaque élément  $\omega$  de l'univers  $\Omega$  contient l'ensemble des paramètres du lancer, toutes les positions et vitesses de la main qui lance le dé, la position et orientation du dé, toutes les caractéristiques de la surface sur lequel le dé est lancé, etc. En somme, tous les paramètres physiques de l'expérience qui peuvent avoir une influence sur son résultat. Comme on n'a pas accès à toutes ces données, on va donc *encoder* le résultat de l'expérience, à travers la *variable aléatoire*  $X$ ="résultat du dé", qui prend ses valeurs dans  $\{1, 2, \dots, 6\}$ . Si le dé est bien équilibré, on estime que la loi de cette variable aléatoire est uniforme sur  $\{1, 2, \dots, 6\}$ .

On regroupe ci-dessous les différences de vocabulaire entre théorie de la mesure et théorie des probabilités.

THÉORIE DE LA MESURE	PROBABILITÉS
Espace mesuré $(E, \mathcal{F}, \mu)$	Espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$
ensemble mesurable	événement
mesure	loi, distribution
fonction mesurable $f : (E, \mathcal{F}) \rightarrow (F, \mathcal{F}')$	variable aléatoire $X : \Omega \rightarrow F$
Espace d'arrivée $F$	Espace d'états $F$
$f^{-1}(A) \in \mathcal{F}$	$\{X \in A\} = X^{-1}(A) \in \mathcal{F}$

## 2.2 Rappels : variables aléatoires discrètes

Pour comprendre une variable aléatoire  $X$ , il nous faut maintenant comprendre sa distribution  $\mathbb{P}_X$ . Commençons par quelques rappels sur les lois discrètes : dans ce contexte, on se donne un espace  $E$  fini ou dénombrable, identifié, quitte à numéroter ses éléments, à  $E = \{1, \dots, N\}$  (cas fini) ou  $E = \mathbb{N}$  (cas infini dénombrable). On munit  $E$  de la *tribu discrète*  $\mathcal{T} = \mathcal{P}(E)$ .

Pour caractériser la distribution  $\mathbb{P}_X$  d'une v.a. discrète  $X$ , il suffit alors d'estimer  $\mathbb{P}_X(\{e\}) = \mathbb{P}(X = e)$  pour chaque élément  $e \in E$ . Cela permet alors de caractériser  $\mathbb{P}_X$  sur toute la tribu discrète  $\mathcal{T}$ . Pour cette raison, la notion de tribu n'est pas nécessaire pour les variables *discrètes*. On verra toutefois qu'elle est fondamentale pour les variables *continues*.

### Definition 9 : Variables aléatoires discrètes ★

Une *variable aléatoire discrète* est une fonction mesurable  $X : \Omega \rightarrow E$  d'un *espace de probabilités*  $(\Omega, \mathcal{F}, \mathbb{P})$  dans un espace mesurable  $(E, \mathcal{T})$  fini ou dénombrable. La tribu  $\mathcal{T}$  étant discrète, la *loi*, ou *distribution*, de  $X$  est entièrement caractérisée par la famille  $(\mu_k)_{k \in E}$ , où  $\mu_k := \mathbb{P}(X = k)$ , que l'on représentera par un vecteur ligne  $\mu = (\mu_1, \mu_2, \dots)$ .

**REMARQUE 8** : La définition précédente des variables aléatoires discrètes est ambiguë : en effet, toute variable aléatoire à valeurs dans  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ , par exemple, peut également être interprétée comme une variable aléatoire à valeurs dans  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ,

puisque pour tout borélien  $B \in \mathcal{B}(\mathbb{R})$

$$X^{-1}(B) = \bigcup_{x \in B \cap \mathbb{N}} X^{-1}(\{x\}) \in \mathcal{F},$$

puisque l'union est dénombrable, et chacun de ses éléments est dans  $\mathcal{F}$  (car  $X : \Omega \rightarrow \mathbb{N}$  est une variable aléatoire).

Par conséquent, caractériser l'espace d'état n'est pas suffisant puisque cet espace d'état n'est pas a priori défini de manière unique. Pour définir les variables aléatoires discrètes de manière non ambiguë, on peut donc donner la définition suivante : une variable aléatoire  $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{T})$  est *discrète* s'il existe un sous ensemble fini ou dénombrable  $F \subset E$ , dont tous les singletons sont mesurables, et tel que  $\mathbb{P}(X \in F) = 1$ . On peut alors voir  $F$  comme l'espace d'état "réel" de  $X$ .

### Definition 10 : Espérance et variance d'une v.a. discrète ★

Rappelons nous que  $E$  est vu comme un sous-ensemble de  $\mathbb{N}$ . L'espérance, ou moyenne d'une v.a. discrète  $X : \Omega \rightarrow E$  de distribution  $\mu$  est donnée par

$$\mathbb{E}(X) = \sum_{k \in E} k \mu_k.$$

Sa variance est donnée par

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

Ces deux quantités sont bien définies si  $E$  est un ensemble fini, et sont définies sous réserve d'absolue convergence de la série si  $E = \mathbb{N}$ .

Une variable est dite *centrée* si son espérance est nulle,  $\mathbb{E}(X) = 0$ . Une variable est dite *réduite* si sa variance vaut  $\text{Var}(X) = 1$ .

### Definition 11 : Probabilités conditionnelles et indépendance ★

Étant donnés deux événements  $A$  et  $B$ , en supposant que  $\mathbb{P}(B) \neq 0$ , on définit la *probabilité conditionnelle de  $A$  sachant  $B$* , notée  $\mathbb{P}(A | B)$  ou parfois  $\mathbb{P}_B(A)$ , par

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

Les événements  $A$  et  $B$  sont *indépendants* si  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ .

### Exercice 7

Montrer que pour tout  $B \in \mathcal{F}$ ,  $\mathbb{P}_B : A \mapsto \mathbb{P}_B(A)$  est une mesure de probabilité.

**SOLUTION :** On remarque que  $\emptyset \cap B = \emptyset$ , et donc  $\mathbb{P}_B(\emptyset) = 0$ . Par ailleurs, si les  $(A_k)$  sont une famille d'événements disjoints, les  $(B \cap A_k)_k$  sont aussi des événements

disjoints, et donc  $\mathbb{P}_B$  est  $\sigma$ -additive car  $\mathbb{P}$  l'est. Finalement,  $\mathbb{P}_B(\Omega) = \mathbb{P}(B)/\mathbb{P}(B) = 1$ ,  $\mathbb{P}_B$  est donc une mesure de probabilité.  $\square$

### Proposition 8 : Formule des probabilités totales ★

Étant donnés des événements  $A \in \mathcal{F}$  et une partition  $\Omega = \sqcup_{k=1}^n B_k$ , où  $B_1, \dots, B_n \in \mathcal{F}$  (on dira que  $(B_k)_{k \leq n}$  est un *système complet d'événements*), on a

$$\mathbb{P}(A) = \sum_{k=1}^n \mathbb{P}(A | B_k) \mathbb{P}(B_k).$$

**PREUVE :** Les événements  $A \cap B_k, A \cap B_{k'}$  étant disjoints deux à deux, par  $\sigma$ -additivité de la mesure  $\mathbb{P}$ , on peut écrire

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap (\sqcup_{k=1}^n B_k)) = \sum_{k=1}^n \mathbb{P}(A \cap B_k) \\ &= \sum_{k=1}^n \mathbb{P}(A | B_k) \mathbb{P}(B_k). \end{aligned}$$

$\square$

### Proposition 9 : Théorème de transfert pour les v.a. discrètes ★

Soit une fonction  $\varphi : E \rightarrow \mathbb{N}$ , alors pour toute v.a. discrète  $X : \Omega \rightarrow E$ , la fonction  $\varphi(X) : \Omega \rightarrow \mathbb{N}$  est également une variable aléatoire discrète, et son espérance s'exprime en fonction de la distribution  $\mu$  de  $X$  par

$$\mathbb{E}(\varphi(X)) = \sum_{k \in E} \varphi(k) \mu_k.$$

**PREUVE :** On commence par montrer que  $\varphi(X) : \omega \mapsto \varphi(X(\omega))$  est bien une variable aléatoire, c'est-à-dire qu'elle est mesurable. Pour cela, il faut montrer que  $\varphi(X)^{-1}(k) \in \mathcal{F}$  pour tout  $k \in \mathbb{N}$ . Or

$$\varphi(X)^{-1}(k) = \{\omega \in \Omega \mid \varphi(X(\omega)) = k\} = \bigcup_{\substack{k' \in E \\ \varphi(k')=k}} X^{-1}(k'),$$

qui est une union dénombrable d'ensembles mesurables, et est donc mesurable.

La formule de transfert en découle immédiatement : par définition de  $\mathbb{E}(\varphi(X))$ , et en utilisant que la famille d'événements  $(\{X = k'\})_{k' \in E}$  est un système complet d'événements,

$$\mathbb{E}(\varphi(X)) = \sum_{k \in \mathbb{N}} k \mathbb{P}(\varphi(X) = k) = \sum_{\substack{k \in \mathbb{N}, \\ k' \in E}} k \mathbb{P}(\varphi(X) = k, X = k')$$

$$= \sum_{\substack{k \in \mathbb{N}, \\ k' \in E}} \varphi(k') \mathbb{P}(\varphi(X) = k, X = k') = \sum_{k' \in E} \varphi(k') \mathbb{P}(X = k').$$

□

### Definition 12 : Suite d'événements indépendants ★

Une suite d'événements  $(A_n)_{n \in \mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$  est dite (*mutuellement*) *indépendante*, si pour tout entier  $k \geq 1$ , et pour toute famille d'indices  $n_1 \neq n_2 \neq \dots \neq n_k$ , on a

$$\mathbb{P}(A_{n_1} \cap A_{n_2} \cap \dots \cap A_{n_k}) = \mathbb{P}(A_{n_1}) \mathbb{P}(A_{n_2}) \dots \mathbb{P}(A_{n_k}).$$

La suite est dite *indépendante deux-à-deux* si la propriété ci dessus est vraie pour  $k = 2$ , c'est à dire que pour  $n, m$

$$\mathbb{P}(A_n \cap A_m) = \mathbb{P}(A_n) \mathbb{P}(A_m).$$

A noter naturellement que si une suite d'événements est *indépendante*, alors elle est en particulier *indépendante deux-à-deux*. On peut maintenant énoncer, pour culture, le second Lemme de Borel-Cantelli :

### Proposition 10 : Second Lemme de Borel-Cantelli

Soit  $(A_n)_{n \in \mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$  une suite *indépendante* d'événements avec  $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ , alors

$$\mathbb{P}(\limsup A_n) = 1.$$

En d'autres termes, si une suite d'événements est de somme de probabilité infinie, un nombre infini de ces événements se réalisent avec probabilité 1.

### Exercice 8

Trouver un contre-exemple au second Lemme de Borel-Cantelli si l'on ne suppose pas la suite d'événements indépendante.

**SOLUTION** : On sait que l'on doit choisir des événements qui ne sont pas indépendants. Un exemple évident de telle suite est de fixer un événement  $A$ , de probabilité  $\mathbb{P}(A) = 1/2$  par exemple, et de choisir  $A_n = A$  pour tout  $n \in \mathbb{N}$ . On a alors  $\limsup A_n = A$ , et donc en particulier  $\mathbb{P}(\limsup A_n) = 1/2$  et pourtant  $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} 1/2 = \infty$ . □

On aborde maintenant l'indépendance de variables aléatoires discrètes.

### Definition 13 : Indépendance de v.a. discrètes



Deux variables aléatoires  $X$  et  $Y$  sur le même espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ , et à valeurs dans deux espaces discrets (au plus dénombrables)  $E, F$  sont *indépendantes* si pour tout  $e \in E, f \in F$ ,

$$\mathbb{P}(X = e, Y = f) = \mathbb{P}(X = e)\mathbb{P}(Y = f).$$

Une suite de variable aléatoires discrètes  $(X_i)_{i \in \mathbb{N}}$  à valeur dans un espace  $E$  est dite (*mutuellement*) *indépendante* (resp. indépendante deux-à-deux) si pour toute suite  $(e_n)_{n \in \mathbb{N}} \in E^{\mathbb{N}}$ , la suite d'événements de terme général  $A_n = \{X_n = e_n\}$  est indépendante (resp. deux-à-deux).

————— Fin du cours 3 —————

## 2.3 Rappels : quelques lois discrètes classiques

Voici pour rappel les lois discrètes les plus importantes. À noter que toutes ces variables sont à valeurs dans  $\mathbb{N}$ , mais toute variable aléatoire à valeurs dans un ensemble fini ou dénombrable peut être considérée comme une variable aléatoire discrète.

### Definition 14 : Loi de Bernoulli

La loi Bernoulli( $p$ ) modélise une expérience avec deux résultats possibles (succès  $X = 1$ , échec,  $X = 0$ ), avec probabilité  $p$  de succès.

- Paramètre  $p \in [0, 1]$ .
- Espace d'états  $E = \{0, 1\}$ .
- Distribution  $\mathbb{P}(X = 1) = p, \mathbb{P}(X = 0) = 1 - p$ .
- Espérance  $\mathbb{E}(X) = p$ , variance  $Var(X) = p(1 - p)$ .

### Definition 15 : Loi binomiale



La loi Binomiale( $n, p$ ) modélise le nombre de succès parmi  $n$  tentatives, d'une expérience qui a une probabilité  $p$  de succès à chaque tentative. C'est la somme de  $n$  variables de Bernoulli( $p$ ) indépendantes.

- Paramètres  $n \in \mathbb{N}, p \in [0, 1]$ .
- Espace d'états  $E = \{0, 1, \dots, n\}$ .
- Distribution  $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ .
- Espérance  $\mathbb{E}(X) = np$ , variance  $Var(X) = np(1 - p)$ .

### Definition 16 : Loi uniforme

La loi uniforme( $n$ ) modélise le choix au hasard d'un objet parmi  $n$  objets identiques. (exemple, lancer de dé)

- Paramètres  $n \in \mathbb{N} \setminus \{0\}$ .
- Espace d'états  $E = \{1, \dots, n\}$ .
- Distribution  $\mathbb{P}(X = k) = 1/n$ .
- Espérance  $\mathbb{E}(X) = (n + 1)/2$ , variance  $\text{Var}(X) = (n^2 - 1)/12$ .

### Definition 17 : Loi géométrique



La loi Géométrique( $p$ ) modélise le nombre de tentatives pour avoir son premier succès, pour une expérience qui a une probabilité  $p$  de succès à chaque tentative.

- Paramètre  $p \in [0, 1]$ .
- Espace d'états  $E = \mathbb{N} \setminus \{0\}$ .
- Distribution  $\mathbb{P}(X = k) = p(1 - p)^{k-1}$ .
- Espérance  $\mathbb{E}(X) = 1/p$ , variance  $\text{Var}(X) = (1 - p)/p^2$ .

### Definition 18 : Loi de Poisson

La loi de Poisson( $\lambda$ ) modélise le nombre d'événements arrivant dans un intervalle de temps donné, si ces dernières arrivent avec une fréquence moyenne,  $1/\lambda$ , connue.

- Paramètre  $\lambda \in \mathbb{R}_+^*$ .
- Espace d'états  $E = \mathbb{N}$ .
- Distribution  $\mathbb{P}(X = k) = \lambda^k e^{-\lambda} / k!$ .
- Espérance  $\mathbb{E}(X) = \lambda$ , variance  $\text{Var}(X) = \lambda$ .

## 2.4 Intégrale et intégrabilité sur $\mathbb{R}$

Après ces rappels sur les lois discrètes, intéressons nous maintenant aux lois continues. Dans tout ce qui suit, et sauf mention contraire, **nous nous placerons sur la droite réelle  $E = \mathbb{R}$ , munie de la tribu borélienne  $\mathcal{B}(\mathbb{R})$** . Comme nous n'avons pas encore introduit l'intégrale de Lebesgue, nous commençons par définir les lois continues dont la densité est continue par morceaux, que l'on peut introduire via l'intégrale de Riemann. L'intégrale de Riemann d'une fonction n'est définie *a priori* que sur un segment. Pour introduire des lois classiques sur  $\mathbb{R}$ , toutefois, on a besoin de pouvoir définir l'intégrale sur  $\mathbb{R}$ , que nous allons définir pour les fonctions continue par morceaux comme l'extension de l'intégrale de Riemann sur un segment.

Rappelons qu'une fonction est continue par morceaux sur  $\mathbb{R}$  si pour tous  $a < b \in \mathbb{R}$ , il existe  $x_0 = a < x_1 < \dots < x_{n-1} < x_n = b$  tels que  $f$  est continue sur chaque intervalle  $]x_{k-1}, x_k[$ , et prolongeable en une fonction continue sur  $[x_{k-1}, x_k]$ .

### Definition 19 : Intégrale sur $\mathbb{R}$

Soit une fonction  $f$  positive et continue par morceaux sur  $\mathbb{R}$ , Riemann-intégrable sur tout segment  $[-A, A]$ , telle que la limite

$$\lim_{A \rightarrow \infty} \int_{-A}^A f(x) dx$$

existe et soit finie. On dit alors que la fonction  $f$  est *intégrable* sur  $\mathbb{R}$ , et on définit

$$\int_{\mathbb{R}} f(x) dx = \lim_{A \rightarrow \infty} \int_{-A}^A f(x) dx.$$

Une fonction  $f$  continue sur  $\mathbb{R}$  est *intégrable* si la fonction positive  $|f|$  (qui est continue par morceaux) est *intégrable*. On admettra alors que l'intégrale  $\int_{\mathbb{R}} f(x) dx = \lim_{A \rightarrow \infty} \int_{-A}^A f(x) dx$  existe et est finie.

A noter que le choix de faire tendre les deux bornes de l'intégrale ensemble pour définir l'intégrabilité sur  $\mathbb{R}$  est arbitraire : Si l'intégrale converge, ça veut dire que les "restes" sur  $[A, +\infty[$  et  $] -\infty, -A]$  tendent tous les deux vers 0, peu importe que l'on prenne la limite simultanée en  $A$  et  $-A$  ou non.

Plus généralement, on peut se poser la question de l'intégrabilité d'une fonction, cette dernière pouvant avoir des discontinuités. Si la fonction est continue par morceaux, alors la définition précédente fonctionne. On peut par contre considérer des fonctions simples comme la fonction  $x \mapsto 1/x$  sur  $\mathbb{R}$ , qui ne sont pas continue par morceaux sur  $\mathbb{R}$  : cette dernière, en particulier, n'est prolongeable ni à gauche ni à droite en une fonction continue en 0. D'une manière générale, une fonction  $f$  peut ne pas être intégrable

- À l'infini.
- Localement, en un point de discontinuité  $y_0$  de  $f$  où elle diverge.

En effet, l'intégrale est une aire, et est grossièrement le produit entre la taille de la base (sur  $\mathbb{R}$ , la largeur d'un intervalle par exemple) et la hauteur (la valeur locale de la fonction). Pour obtenir une aire divergente, il faut donc soit que la taille de la base diverge (intégrabilité à l'infini), soit que la hauteur diverge, c'est à dire que localement la fonction diverge. Dans les deux cas, une fonction est intégrable si l'intégrale de sa valeur absolue reste finie. En pratique, trois cas sont possibles :

- soit on sait primitiver la fonction  $|f|$ , auquel cas on peut calculer directement si l'intégrale reste finie lorsque les bornes tendent vers  $y_0$  et  $\infty$ .
- soit on ne sait pas calculer la primitive de  $|f|$ , et on peut minorer  $|f|$  par une fonction  $g \leq f$  positive, dont on sait calculer l'intégrale, et d'intégrale divergente, auquel cas  $f$  n'est pas intégrable par monotonie.

- soit on ne sait pas calculer la primitive de  $|f|$ , et on peut majorer  $|f|$  par une fonction  $g \geq f$ , dont on sait calculer l'intégrale, et d'intégrale convergente, auquel cas  $f$  est intégrable par monotonie.

À titre d'exemple, on peut considérer la fonction  $x \mapsto 1/x$  qui n'est intégrable ni en  $\pm\infty$ , ni en  $0^\pm$ . En effet, Pour tout  $0 < \varepsilon < A < +\infty$

$$\int_{\varepsilon}^{+\infty} \frac{1}{x} dx = \log A - \log \varepsilon,$$

qui diverge quand  $\varepsilon \rightarrow 0$  ( $1/x$  n'est donc pas intégrable en  $0^+$ ) et quand  $A \rightarrow \infty$  ( $1/x$  n'est pas intégrable en  $+\infty$ ). A titre de pense-bête, reprenez que pour tout  $\varepsilon > 0$  toute fonction qui décroît "plus vite" que  $1/x^{1+\varepsilon}$  va être intégrable en  $\infty$  ( $1/x$ ,  $e^{-|x|}$  par exemple), et toute fonction qui explose moins vite que  $1/x^{1-\varepsilon}$  ( $1/\sqrt{x}$  par exemple) va être intégrable en 0.

## 2.5 Variables aléatoires réelles à densité

On se place sur un espace de probabilités  $(\Omega, \mathcal{F}, \mathbb{P})$ , et dans ce qui suit, on munit  $\mathbb{R}$  de la tribu des boréliens  $\mathcal{B}(\mathbb{R})$  (cf. Définition 3).

### Definition 20 : Lois à densité continue par morceaux ★

Une fonction  $f : \mathbb{R} \rightarrow \mathbb{R}$  est une *densité de probabilité* si elle est

- positive, i.e.  $f(x) \in \mathbb{R}_+ \quad \forall x \in \mathbb{R}$ .
- borélienne et intégrable sur  $\mathbb{R}$ , et satisfait  $\int_{\mathbb{R}} f(x) dx = 1$ .

Dans le cadre de ce cours, on considérera des densités de probabilités continue par morceaux.

Une *variable aléatoire réelle* est une fonction mesurable  $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Une *variable aléatoire réelle*  $X : \Omega \rightarrow \mathbb{R}$  est dite à *densité*  $f_X$  si pour tout  $a \leq b \in \mathbb{R}$ ,

$$\mathbb{P}_X([a, b]) = \mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx.$$



**ATTENTION :** On remarque que pour toute variable à densité  $X$ , et pour tout  $x \in \mathbb{R}$ , on a nécessairement  $\mathbb{P}(X = x) = \int_x^x f_X(x) dx = 0$ . En particulier, pour les variables à densité, les inégalités strictes et larges peuvent être interverties librement. Attention, ce n'est pas du tout vrai en général! Pour une variable de Bernoulli  $X \sim \text{Ber}(1/2)$ , par exemple, on a  $\mathbb{P}(X < 1) = 1/2 \neq \mathbb{P}(X \leq 1) = 1$ .

### Definition 21 : Espérance et Variance de lois à densité ★

Soit  $X$  une variable à densité  $f_X$ , continue par morceaux. On suppose que la fonction  $|x|f_X(x)$  est intégrable sur  $\mathbb{R}$ , on peut alors définir l'*espérance*  $\mathbb{E}(X)$



de  $X$  par

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f_X(x) dx,$$

et on dit alors que la variable aléatoire  $X$  est intégrable. Si  $x^2 f_X(x)$  est intégrable, on dit que  $X$  est de carré intégrable, et on définit la *variance*  $Var(X)$  de  $X$  par

$$Var(X) = \int_{\mathbb{R}} x^2 f_X(x) dx - \mathbb{E}(X)^2.$$

On rappelle qu'une variable est dite *centrée* si son espérance est nulle,  $\mathbb{E}(X) = 0$ . Une variable est dite *réduite* si sa variance vaut  $Var(X) = 1$ .

### Exercice 9 : Une v.a. de carré intégrable est intégrable

Montrer qu'une variable à densité de carré intégrable  $X$  est intégrable.

**SOLUTION :** Supposons  $X$  de carré intégrable, c'est à dire que  $x^2 f_X(x)$  est intégrable. Cette dernière est une fonction positive, par conséquent, on en déduit que

$$\int_{\mathbb{R}} x^2 f_X(x) dx < \infty.$$

On cherche à montrer que  $\int_{\mathbb{R}} |x| f_X(x) dx < \infty$ . Pour le montrer, on écrit

$$\int_{\mathbb{R}} |x| f_X(x) dx = \int_{[-1,1]} |x| f_X(x) dx + \int_{\mathbb{R} \setminus [-1,1]} |x| f_X(x) dx.$$

Or, si  $x \in [-1, 1]$ ,  $|x| \leq 1$ , et si  $x \notin [-1, 1]$ ,  $|x| \leq x^2$ , par conséquent

$$\begin{aligned} \int_{\mathbb{R}} |x| f_X(x) dx &\leq \int_{[-1,1]} f_X(x) dx + \int_{\mathbb{R} \setminus [-1,1]} x^2 f_X(x) dx \\ &\leq \int_{\mathbb{R}} f_X(x) dx + \int_{\mathbb{R}} x^2 f_X(x) dx \\ &= 1 + \int_{\mathbb{R}} x^2 f_X(x) dx < \infty, \end{aligned}$$

ce qui prouve que  $X$  est intégrable. □

### Definition 22 : indépendance de lois à densité

Soit  $X$  et  $Y$  deux variable à densité  $f_X, f_Y$  continues par morceaux. On dit que  $X$  et  $Y$  sont indépendantes si pour tous réels  $a, b, c, d \in \mathbb{R}$

$$\mathbb{P}(X \in [a, b], Y \in [c, d]) = \int_a^b f_X(x) dx \int_c^d f_Y(y) dy.$$

## 2.6 Fonction de répartition

On introduit maintenant une notion fondamentale des variables aléatoires réelles, la fonction de répartition.

### Definition 23 : Fonction de répartition ★

Soit  $X$  une variable aléatoire, on appelle *fonction de répartition* de  $X$  la fonction  $F_X : \mathbb{R} \rightarrow [0, 1]$  définie par

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}_X(]-\infty, x[).$$

Si de plus  $X$  est une variable à densité, on peut écrire la fonction de répartition sous la forme

$$F_X(x) = \int_{-\infty}^x f_X(y) dy.$$

**REMARQUE 9** : La première identité,  $F_X(x) = \mathbb{P}(X \leq x)$  est valide pour toutes variables aléatoires à valeurs dans  $\mathbb{R}$ , y compris les variables discrètes à valeurs dans  $\mathbb{N}$ . Toutefois, on utilise rarement la fonction de répartition pour les variables discrètes, pour lesquelles sa valeur est rarement explicite.

**REMARQUE 10** : Par sa définition, on remarque immédiatement que la fonction de répartition d'une variable à densité est la primitive de la densité. En particulier, la fonction de répartition d'une variable à densité est continue sur  $\mathbb{R}$  et elle est non-dérivable au plus en un nombre dénombrable de points. Pour obtenir la densité à partir de la fonction de répartition, il suffit donc de dériver. La réciproque est vraie : si  $F_X$  est continue sur  $\mathbb{R}$  et dérivable sauf en un nombre fini de points (c'est à dire continue sur  $\mathbb{R}$  et  $C^1$  par morceaux), alors  $X$  est à densité, et sa densité est donnée par l'identité

$$F'_X(x) = f_X(x).$$

### Proposition 11 : Propriétés de la fonction de répartition, admis

La fonction de répartition  $F_X$  d'une v.a. réelle est croissante, continue à droite, et satisfait

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow +\infty} F_X(x) = 1.$$

Par ailleurs, elle caractérise entièrement la loi de la variable  $X$ , c'est-à-dire que si deux variables aléatoires  $X$  et  $Y$  ont même fonction de répartition, alors leurs distributions  $\mathbb{P}_X$  et  $\mathbb{P}_Y$  sont identiques.

**REMARQUE 11** : Nous reverrons plus tard ces propriétés de la fonction de répartition (cf. Proposition 20 et Exercice 20). En pratique, il est important de savoir reconnaître, éventuellement graphiquement, les fonctions de répartition classiques, qui peuvent permettre d'un coup d'œil d'identifier les distributions correspondantes.

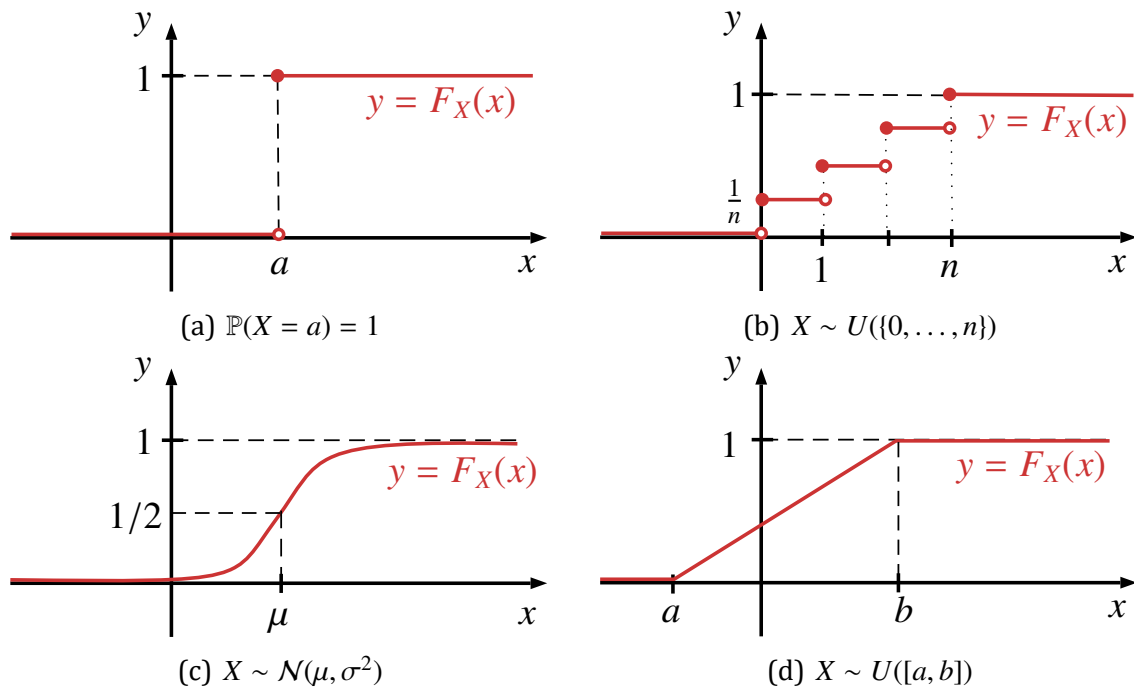


FIGURE 2 – Representation de la fonction de répartition de quelques lois classiques.

### Exercice 10

Soit  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  une fonction *positive*, on admettra le théorème de Fubini, qui garantit que pour tous intervalles  $I, I' \subset \mathbb{R}$ , l'ordre d'intégration sur  $I \times I'$  n'a pas d'importance, c'est-à-dire que

$$\int_I \int_{I'} f(x, y) dx dy = \int_{I'} \int_I f(x, y) dy dx$$

Montrer que si  $X$  est une v.a. réelle, positive, à densité  $f_X$  dont l'espérance est bien définie, alors

$$\mathbb{E}(X) = \int_0^{\infty} (1 - F_X(y)) dy.$$

*Indice* : on pourra commencer par remarquer que  $x = \int_0^x dy$ , et utiliser la fonction

$$\mathbf{1}_{\{y \leq x\}} = \begin{cases} 1 & \text{si } y \leq x \\ 0 & \text{sinon.} \end{cases}.$$

**SOLUTION** : Comme  $X$  est une variable positive, on en déduit que  $f_X = 0$  sur  $]-\infty, 0]$ . On en déduit

$$\mathbb{E}(X) = \int_0^{+\infty} x f_X(x) dx.$$

On écrit alors  $x = \int_0^x dy = \int_0^{+\infty} \mathbf{1}_{\{0 \leq y \leq x\}} dy$ , que l'on introduit dans l'intégrale, puis on

intervertit les deux intégrales par théorème de Fubini, pour obtenir

$$\mathbb{E}(X) = \int_0^{+\infty} \left[ \int_0^{+\infty} \mathbf{1}_{\{0 \leq y \leq x\}} f_X(x) dx \right] dy = \int_0^{+\infty} \left[ \int_y^{+\infty} f_X(x) dx \right] dy = \int_0^{+\infty} \mathbb{P}(X > y) dy,$$

which proves the result. □

On regarde maintenant ce qu'il se passe lorsque la fonction de répartition n'est pas continue à gauche.

### Proposition 12 : Fonction de répartition et atomes

Soit  $X$  une variable aléatoire réelle et  $\mathbb{P}_X$  sa distribution. On dit que  $\mathbb{P}_X$  admet un atome  $x$  si

$$\mathbb{P}(X = x) = \mathbb{P}(\{x\}) > 0.$$

Alors,  $\mathbb{P}_X$  admet un atome en  $X$  si et seulement si  $F_X$  est discontinue (à gauche, forcément, puisqu'elle est continue à droite) en  $x$ , et par ailleurs

$$\mathbb{P}(X = x) = F_X(x) - \lim_{y \nearrow x} F_X(y)$$

est la taille du saut en  $x$ .

**PREUVE :** On commence par remarquer que

$$\{x\} = \bigcap_{n \geq 1} ]x - 1/n, x].$$

Ces intervalles sont décroissants,  $\mathbb{P}_X$  est une mesure de probabilités, elle est donc continue par rapport aux intersections décroissantes (cf. Proposition 6-4)), et par conséquent

$$\begin{aligned} \mathbb{P}_X(\{x\}) &= \mathbb{P}_X(\bigcap_{n \geq 1} ]x - 1/n, x]) = \lim_{n \rightarrow \infty} \mathbb{P}_X(]x - 1/n, x]) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}_X(]-\infty, x]) - \mathbb{P}_X(]-\infty, x - 1/n]) \\ &= F_X(x) - \lim_{n \rightarrow \infty} F_X(x - 1/n) \\ &= F_X(x) - \lim_{y \nearrow x} F_X(y) \end{aligned}$$

□

**REMARQUE :** A noter en particulier que pour une variable discrète  $X$  prenant par exemple des valeurs  $x_1 < x_2 \dots$  avec probabilités  $p_1, p_2, \dots$ , la fonction de répartition fait à chaque  $x_k$  un saut de taille  $p_k$ , puis reste constante jusqu'à  $x_{k+1}$ . Pour les variables à densité, qui n'ont pas d'atomes, on retrouve bien que la fonction de répartition est continue.

## 2.7 Quelques lois classiques à densité continue par morceaux

### Definition 24 : Loi uniforme sur un segment

La loi uniforme  $U([a, b])$  modélise le tirage d'un nombre réel entre  $a$  et  $b$ .

- Paramètres  $a < b \in \mathbb{R}$ .
- Densité

$$f_{a,b}(x) = \begin{cases} 1/(b-a) & \text{si } x \in [a, b] \\ 0 & \text{sinon.} \end{cases}$$

- Espérance  $\mathbb{E}(X) = \frac{a+b}{2}$ , variance  $\text{Var}(X) = \frac{(b-a)^2}{12}$ .



*ATTENTION : il ne faut pas confondre les variables uniformes sur un segment ( $X \sim U([0, 1])$ ), et les variables uniformes sur un ensemble discret (par exemple  $X \sim U(\{1, \dots, n\})$ ). La première est une variable à densité, dont l'espace d'états est infini non dénombrable, alors que la seconde est une variable discrète dont l'espace d'état est fini (et donc en particulier dénombrable).*

**REMARQUE :** À noter qu'il n'existe pas de loi  $\mathbb{P}_X$  (de probabilité) uniforme sur  $\mathbb{N}$  tout entier, ou sur  $\mathbb{Q}$  tout entier : en effet, en supposant qu'il en existe une, chaque élément de  $\mathbb{N}$ , par exemple, devrait avoir la même probabilité  $p \geq 0$ , mais par  $\sigma$ -additivité, on aurait alors  $1 = \mathbb{P}_X(\mathbb{N}) = \sum_{k \in \mathbb{N}} p = \infty \times p$ , ce qui est impossible, puisque la série vaut 0 si  $p = 0$  et  $\infty$  sinon. De la même manière, on ne peut pas construire de variable uniforme sur un segment non-borné ( $\mathbb{R}$  par exemple), on se heurterait au même problème.

### Exercice 11

Soit  $X$  une variable aléatoire uniforme sur  $[0, 1]$ . Montrer que  $Y = 1 - X$  :  $\omega \mapsto 1 - X(\omega)$  est une variable uniforme sur  $[0, 1]$ .

**SOLUTION :** Étant donné que  $X$  est une variable aléatoire uniforme, on calcule/connait sa fonction de répartition :

$$F_X(x) = \int_{-\infty}^x f_X(u) du = \begin{cases} 0 & \text{for } x \leq 0 \\ x & \text{for } 0 \leq x \leq 1 \\ 1 & \text{for } x \geq 1. \end{cases}$$

On calcule maintenant la fonction de répartition de  $Y$ ,

$$F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}(1 - X \leq x) = \mathbb{P}(X \geq 1 - x) = 1 - \mathbb{P}(X < 1 - x) = 1 - F_X(1 - x).$$

car  $x$  est une variable à densité et donc  $\mathbb{P}(x = \alpha) = 0$  pour tout réel  $\alpha$ . un rapide

calcul nous donne alors

$$F_Y(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ x & \text{for } 0 \leq x \leq 1 \\ 1 & \text{for } x \geq 1 \end{cases} = F_X(x).$$

Les deux variables  $X$  et  $Y$  ont même fonction de répartition, elles ont donc également la même loi.  $\square$

### Exercice 12 : Minimums de lois uniformes

On considère une famille i.i.d.  $(X_k)_{k=1,\dots,n}$  de variables aléatoires de loi uniforme sur  $[0, 1]$ .

1) On pose

$$Y = \min_{k=1,\dots,n} X_k,$$

calculer la fonction de répartition de  $Y$ . Montrer que  $Y$  est à densité et calculer sa densité.

2) Soit  $N \sim \text{Bin}(n, 1/2)$  indépendante des  $X_k$ , on pose  $Y'(\omega) = \min_{k=1,\dots,N(\omega)} X_k(\omega)$ . Montrer que  $Y'$  est à densité et calculer sa densité.

#### SOLUTION :

1) On calcule la fonction de répartition de  $Y$  :

$$F_Y(x) = \mathbb{P}(Y \leq x) = 1 - \mathbb{P}(X_k \geq x, \forall k \in \{1, \dots, n\}) = 1 - \mathbb{P}(X_1 \geq x)^n$$

ou la dernière identité est obtenue par indépendance. les  $X_i$  étant à densité,  $\mathbb{P}(X_1 \geq x) = \mathbb{P}(X_1 > x) = 1 - F_{X_1}(x)$ , ce dont on déduit que

$$F_Y(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - (1 - x)^n & \text{si } x \in (0, 1) \\ 1 & \text{si } x \geq 1 \end{cases}$$

On remarque que  $F_Y$  est de classe  $C^1$  sur  $\mathbb{R} \setminus \{0, 1\}$ , et continue sur  $\mathbb{R}$ , sa dérivée est donc définie partout sauf en 0 et 1, et on a donc pour tout  $x \in \mathbb{R}$   $F_Y(x) = \int_0^x F'_Y(t) dt$ ,  $Y$  est donc à densité, et  $f_Y = n(1 - x)^{n-1} \mathbf{1}_{[0,1]}(x)$ .

2) Pour la seconde question, on projette selon les valeurs possible de  $N$ , on note  $F_n$  la fonction de répartition obtenue à la question précédente. Grâce à la formule des probabilités totales, on a  $F_{Y'}(x) = \sum_{k=0}^n \binom{n}{k} 2^{-n} F_n(x)$ ,  $F_{Y'}(x)$  est donc elle même dérivable sur  $\mathbb{R} \setminus \{0, 1\}$  et continue sur  $\mathbb{R}$ , et on obtient la densité de  $Y'$  en dérivant  $F_{Y'}$ , ce qui après un petit calcul donne  $f_{Y'} = n2^{-n}(2 - x)^{n-1} \mathbf{1}_{[0,1]}(x)$ .  $\square$

### Definition 25 : Loi gaussienne/normale



La loi normale  $\mathcal{N}(\mu, \sigma^2)$  est une loi fondamentale en théorie des probabilités parce qu'elle modélise, entre autres, les fluctuations de la moyenne empirique (cf. équation (3) ci-dessous) d'un grand nombre de tirages indépendants autour de leur espérance. C'est le *théorème central limite*, que nous verrons en fin de cours (cf. Théorème 29).

— Paramètres  $\mu \in \mathbb{R}$  et  $\sigma^2 > 0$ .

— Densité

$$f_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

— Espérance  $\mathbb{E}(X) = \mu$ , variance  $\text{Var}(X) = \sigma^2$ .

La *fonction de répartition* de la loi normale n'est pas une fonction explicite. Lorsque  $\mu = 0$  et  $\sigma = 1$ , on parlera de loi *normale centrée réduite*, notée  $\mathcal{N}(0, 1)$ .

Fin du cours 4

### Definition 26 : Loi exponentielle



La loi exponentielle représente le temps d'attente pour des événements indépendants intervenant à fréquence moyenne donnée. Elle est très utile pour modéliser des files d'attente.

— Paramètre  $\lambda > 0$ .

— Densité  $f_\lambda(x) = \lambda e^{-\lambda x} \mathbf{1}_{\{x \geq 0\}}$ , fonction de répartition

$$F_\lambda(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{sinon.} \end{cases}$$

— Espérance  $\mathbb{E}(X) = 1/\lambda$ , variance  $\text{Var}(X) = 1/\lambda^2$ .



*ATTENTION : il pourra nous arriver d'écrire "Exp( $\lambda$ )" au lieu de "Y, où  $Y \sim \text{Exp}(\lambda)$ " (cf par exemple la solution de l'exercice suivant). Cela revient à assimiler une distribution  $\mathbb{P}_X$  et une variable aléatoire  $X$  ayant cette distribution. Il ne faut donc, en particulier, pas se méprendre sur la signification de  $\mathbb{P}(\text{Exp}(1) \leq x)$ , dans la quelle  $\text{Exp}(1)$  désigne une variable aléatoire  $X \sim \text{Exp}(1)$ , et non pas le nombre réel  $\exp(1) = e^1 = e$ .*

### Exercice 13 : absence de mémoire de la loi exponentielle

Barnabé attend le bus. Ayant étudié en détail les transports publics, il sait que le temps, exprimé en heures, qui sépare le passage de deux bus successifs suit une loi exponentielle de paramètre  $\lambda = 1$ .

1) *Ilevia* informe Barnabé qu'un bus est passé à 16h. Quelle est la probabilité

que le suivant arrive avant 16h30?

2) Un bus étant passé à 16h, Barnabé arrive à l'arrêt à 16h30, et remarque que le bus suivant n'est pas encore passé. Quelle est la probabilité que Barnabé attende moins de 30min le bus suivant?

3) En déduire la propriété, dite *d'absence de mémoire*, de la loi exponentielle : soit  $X$  une v.a. de loi exponentielle, pour tous  $a, b \geq 0$

$$\mathbb{P}(X \geq a + b \mid X \geq a) = \mathbb{P}(X \geq b).$$

**SOLUTION :**

1) On note  $X_k$  l'heure à laquelle passe le  $k$ -ème bus, et  $Y_k = X_k - X_{k-1}$  le temps d'attente du  $k$ -ème bus. On suppose arbitrairement que le bus 0 passe à 16h, i.e.  $X_0 = 16$ . Comme une loi exponentielle ne peut pas être négative, on cherche

$$\mathbb{P}(X_1 \leq 16.5) = \mathbb{P}(Y_1 \leq 0.5) = \mathbb{P}(\text{Exp}(1) \leq 0.5) = F(0.5),$$

où  $F(x) = 1 - e^{-x}$  représente la fonction de répartition d'une variable exponentielle de paramètre  $\lambda = 1$ . Par conséquent,  $\mathbb{P}(X_1 \leq 16.5) = 1 - e^{-0.5}$ .

2) On pose  $X_0 = 0$ , on cherche

$$\begin{aligned} \mathbb{P}(Y_1 \leq 1 \mid X_1 \geq 16.5) &= \frac{\mathbb{P}(Y_1 \leq 1, X_1 \geq 16.5)}{\mathbb{P}(X_1 \geq 16.5)} = \frac{\mathbb{P}(0.5 \leq Y_1 \leq 1)}{\mathbb{P}(Y_1 \geq 0.5)} \\ &= \frac{F(1) - F(0.5)}{1 - F(0.5)} = \frac{e^{-0.5} - e^{-1}}{e^{-0.5}} = F(0.5). \end{aligned}$$

3) Pour une loi exponentielle de paramètre  $\lambda$ , on a

$$\mathbb{P}(X \geq a + b \mid X \geq a) = \frac{\mathbb{P}(X \geq a + b)}{\mathbb{P}(X \geq a)} = \frac{1 - F(a + b)}{1 - F(a)} = e^{-\lambda(a+b)+\lambda a} = e^{-\lambda b} = \mathbb{P}(X \geq b).$$

Autrement dit, étant donné une loi exponentielle  $X$  de paramètre  $\lambda$ , conditionnellement à  $X \geq a$ , la variable aléatoire  $Y - a$  est suit également une loi exponentielle de paramètre  $\lambda$ . □

**Exercice 14 : lois ni discrètes ni à densité**

Barnabé arrive à un guichet de poste. S'il n'y a pas de queue, ce qui arrive avec probabilité  $p = 1/2$ , on s'occupe de lui immédiatement et il n'attend pas. Sinon, il doit attendre un temps exponentiel (en heures) de paramètre  $\lambda = 1$  avant que l'on s'occupe de lui.

1) On note  $X$  le temps d'attente (en heures) de Barnabé. La loi de  $X$  est-elle discrète? À densité?

2) Quelle est la probabilité que Barnabé attende une heure ou moins?

**SOLUTION :**



1) On note  $V$  l'événement "Le guichet est vide". La loi de  $X$  n'est pas à densité, puisque

$$\mathbb{P}(X = 0) = \mathbb{P}(X = 0 | V)\mathbb{P}(V) + \mathbb{P}(X = 0 | V^c)\mathbb{P}(V^c) = 1 * 1/2 + 1/2\mathbb{P}(Exp(1) = 0) = 1/2.$$

Si  $X$  était à densité, on aurait  $\mathbb{P}(X = 0) = \int_0^0 f_X(x)dx = 0$ .

On montre maintenant que  $X$  n'est pas discrète. Soit  $(a_k)_{k \in \mathbb{N}}$  une suite de réels non nuls. On calcule

$$\mathbb{P}(X \in \{a_k, k \in \mathbb{N}\}) = \sum_k \mathbb{P}(X = a_k | V)\mathbb{P}(V) = \frac{1}{2} \sum_k \mathbb{P}(Exp(1) = a_k) = 0,$$

En particulier, pour tout ensemble dénombrable  $\Gamma$ ,  $\mathbb{P}(X \in \Gamma) = \frac{1}{2} \mathbf{1}_{\{0 \in \Gamma\}} < 1$ , la variable  $X$  n'est donc pas discrète.

2) On écrit

$$\mathbb{P}(X \leq 1) = \mathbb{P}(X \leq 1 | V)\mathbb{P}(V) + \mathbb{P}(X \leq 1 | V^c)\mathbb{P}(V^c) = 1 * \frac{1}{2} + \frac{1}{2}F(1) = 1 - \frac{1}{2}e^{-1},$$

où  $F$  est la fonction de répartition d'une loi exponentielle de paramètre 1. □

### 3 Intégrale par rapport à une mesure $\sigma$ -finie.

Les variables aléatoires discrètes et les variables aléatoires à densité ne sont que des exemples de variables aléatoires réelles. Comme le montre l'exercice précédent, on peut facilement construire des variables aléatoires qui ne sont ni l'un, ni l'autre. Nous allons maintenant construire l'intégrale d'une manière beaucoup plus générale que celle de Riemann, ce qui nous donnera accès à de nombreux résultats fort utiles. On se place pour l'instant dans un cadre aussi général que possible, en fixant un espace mesuré  $(E, \mathcal{T}, \mu)$ , muni d'une mesure  $\sigma$ -finie  $\mu$  (cf. Définition 6). On considérera dans la suite uniquement des fonctions  $f : (E, \mathcal{T}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  mesurables.

#### 3.1 Définition de l'intégrale d'une fonction mesurable



Lorsque  $(E, \mathcal{T}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ , où  $\lambda$  est la *mesure de Lebesgue* déjà évoquée à la fin de la Section 1, cette nouvelle intégrale sera l'*intégrale de Lebesgue*, beaucoup plus puissante que l'intégrale de Riemann. La construction de cette nouvelle intégrale se fait par étapes, commençons par définir des fonctions essentielles, les fonctions *indicatrices* d'ensembles mesurables.

##### Definition 27 : Fonction indicatrice d'un ensemble mesurable



Soit  $A \in \mathcal{T}$  un ensemble mesurable, on définit la fonction  $\mathbf{1}_A : E \rightarrow \mathbb{R}$ , appelée *fonction indicatrice de A*, par

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \in A^c. \end{cases}$$

À noter que  $\mathbf{1}_A(x)\mathbf{1}_B(x) = \mathbf{1}_{A \cap B}(x)$ .

Les fonctions indicatrices forment la base de la construction de l'intégrale au sens de Lebesgue. Définissons maintenant l'intégrale d'une combinaison linéaire de fonctions indicatrices.

##### Definition 28 : Fonctions étagées

Une fonction  $f : E \rightarrow \mathbb{R}$  est *étagée* s'il existe  $n \in \mathbb{N}$ , une famille d'ensembles mesurables  $A_1, \dots, A_n \in \mathcal{T}$ , et  $a_1, \dots, a_n \in \mathbb{R}$  tels que

$$f = \sum_{i=1}^n a_i \mathbf{1}_{A_i}.$$

On note  $\mathcal{E}$  l'ensemble des fonctions étagées sur  $E$ . On note  $\mathcal{E}_+$  l'ensemble des fonctions étagées positives, c'est-à-dire telles que  $a_i \in \mathbb{R}_+ \forall i \leq n$ .

**Definition 29 : Intégrale d'une fonction étagée positive**

Soit  $f = \sum_{i=1}^n a_i \mathbf{1}_{A_i} \in \mathcal{E}_+$  une fonction étagée positive, on définit l'intégrale de  $f$  (contre  $\mu$ ) par

$$\int f d\mu = \sum_{i=1}^n a_i \mu(A_i).$$

Cette définition reste valable si  $f$  n'est pas supposée positive, mais que la mesure  $\mu$  est finie.

**Exercice 15**

Montrer que la fonction  $x \mapsto \mathbf{1}_{\mathbb{Q}}(x)$  est intégrable sur  $\mathbb{R}$  et calculer son intégrale  $\int_{\mathbb{R}} \mathbf{1}_{\mathbb{Q}} d\lambda$  contre la mesure de Lebesgue sur  $\mathbb{R}$ .

**SOLUTION :** L'ensemble  $\mathbb{Q}$  est une union dénombrable de singletons (qui sont chacun de mesure de Lebesgue nulle), et donc  $\lambda(\mathbb{Q}) = 0$ . La fonction  $\mathbf{1}_{\mathbb{Q}}$  est une fonction indicatrice, et donc étagée, son intégrale est donc donnée par

$$\int_{\mathbb{R}} \mathbf{1}_{\mathbb{Q}} d\lambda = 1 \times \lambda(\mathbb{Q}) = 0.$$

□

**Exercice 16**

Montrer qu'une fonction est étagée si et seulement si il existe une famille  $B_1, \dots, B_k \in \mathcal{T}$  d'ensembles mesurables *disjoints*, et  $b_1, \dots, b_k \in \mathbb{R}$  tels que

$$f = \sum_{i=1}^k b_i \mathbf{1}_{B_i}.$$

**SOLUTION :** Soit  $f = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$  une fonction étagée. Pour  $\sigma \subset \{1, \dots, n\}$ , on définit  $\sigma^c = \{1, \dots, n\} \setminus \sigma$  et l'ensemble mesurable

$$B_\sigma = \left( \bigcap_{k \in \sigma} A_k \right) \cap \left( \bigcap_{\ell \in \sigma^c} A_\ell^c \right).$$

On vérifie facilement que pour  $\sigma \neq \sigma'$ ,  $B_\sigma$  et  $B_{\sigma'}$  sont disjoints. On définit également  $b_\sigma = \sum_{k \in \sigma} a_k$ , et on peut alors vérifier (par récurrence sur  $n$  par exemple) que

$$f = \sum_{\sigma \subset \{1, \dots, n\}} b_\sigma \mathbf{1}_{B_\sigma}$$

□

**Definition 30 : Intégrale d'une fonction positive**

Soit  $f : E \rightarrow \mathbb{R}_+$  une fonction *mesurable positive*. On définit l'intégrale de  $f$ , notée  $\int f d\mu \in \mathbb{R}_+ \cup \{+\infty\}$ , par

$$\int f d\mu = \sup \left\{ \int g d\mu; g \in \mathcal{E}_+, g \leq f \right\}.$$

La fonction  $f$  est dite *intégrable* si  $\int f d\mu < \infty$ .

**Definition 31 : Ensembles négligeables**

On dit qu'un ensemble  $A \subset E$  est *négligeable* s'il existe un ensemble mesurable  $N \supset A$  satisfaisant  $\mu(N) = 0$ . On dira qu'une propriété est vraie *presque-partout* si elle est vraie partout sauf sur un ensemble négligeable.

**Exercice 17**

Soient  $f, g$  deux fonctions positives intégrables. On suppose que l'ensemble mesurable  $\{x \in E : f(x) \neq g(x)\}$  est négligeable (Autrement dit, on suppose que  $f = g$  presque-partout). Montrer que l'on a

$$\int f d\mu = \int g d\mu.$$

**SOLUTION** : On va montrer que  $\int f d\mu \leq \int g d\mu$ , ce qui, comme  $f$  et  $g$  jouent des rôles symétriques, montre le résultat. Pour cela, il suffit de montrer que pour toute fonction étagée positive  $h \in \mathcal{E}_+$ , satisfaisant  $h \leq f$ , on a

$$\int h d\mu \leq \int g d\mu.$$

Par hypothèse, il existe un ensemble mesurable  $N \subset E$  tel que  $\mu(N) = 0$  et  $f = g$  sur  $E \setminus N$ . On définit alors

$$\tilde{h}(x) = h(x)\mathbf{1}_{E \setminus N}(x).$$

On remarque facilement que  $\tilde{h} = \sum_{i=1}^n a_i \mathbf{1}_{A_i \setminus N}$  est une fonction étagée positive car  $h$  l'est, par ailleurs comme on ne l'a modifiée que sur un ensemble négligeable, on a également

$$\int h d\mu := \sum_{i=1}^n a_i \mu(A_i) = \int \tilde{h} d\mu := \sum_{i=1}^n a_i \mu(A_i \setminus N).$$

Par ailleurs, comme  $h \leq f$ ,  $g \geq 0$  et  $f = g$  sur  $E \setminus N$ , on a  $\tilde{h} \leq g$ . On en déduit par définition que

$$\int \tilde{h} d\mu \leq \int g d\mu,$$

et par conséquent

$$\int h d\mu \leq \int g d\mu.$$

Cela montre que  $\int f d\mu \leq \int g d\mu$ . □

### Definition 32 : Intégrale d'une fonction mesurable, espace $\mathcal{L}^1$ ★

Soit  $f$  une fonction mesurable, on définit les fonctions  $f_+(x) = \max\{f(x), 0\}$  et  $f_-(x) = \max\{-f(x), 0\}$ . Ces fonctions sont positives et mesurables, et on a les identités

$$f = f_+ - f_- \quad \text{et} \quad |f| = f_+ + f_-.$$

On dit que la fonction  $f$  est *intégrable* si  $f_+$  et  $f_-$  sont toutes les deux intégrables. L'intégrale de  $f$  est alors définie par

$$\int f d\mu = \int f_+ d\mu - \int f_- d\mu.$$

On note  $\mathcal{L}^1(E, \mathcal{T}, \mu)$ , ou plus simplement  $\mathcal{L}^1$  ou  $\mathcal{L}^1(\mu)$ , l'ensemble des fonctions intégrables sur  $E$ .

**NOTATION** : Étant données une fonction  $f$  et une mesure  $\mu$ , on notera indifféremment

$$\int f d\mu, \quad \int f(x) d\mu(x) \quad \text{ou} \quad \int f(x) \mu(dx)$$

pour l'intégrale de  $f$  contre  $\mu$ .

Lorsque  $\mu = \lambda$  est la mesure de Lebesgue sur  $\mathbb{R}$ , on pourra également adopter la notation  $\int f(x) dx = \int f d\lambda$  pour son intégrale contre la mesure de Lebesgue.

**REMARQUE 12** : On voit facilement à l'aide de l'exercice 17 que si deux fonctions mesurables (pas nécessairement positives) sont égales presque-partout, et que l'une est intégrable, alors l'autre est aussi intégrable et leurs intégrales sont égales.

### Definition 33 : Intégrale sur un sous-ensemble mesurable

Soit  $f$  une fonction mesurable, et  $A \in \mathcal{T}$  un ensemble mesurable, on définit, si elle existe,

$$\int_A f d\mu = \int_E f \mathbf{1}_A d\mu.$$

L'intégrale de Lebesgue étant plus générale que l'intégrale de Riemann, elle en partage deux propriétés fondamentales.

### Proposition 13 : Propriétés de l'intégrale, admis



Linéarité : Pour deux fonctions mesurables  $f$  et  $g$ , et réels  $\alpha$  et  $\beta$ , on admettra que la fonction  $\alpha f + \beta g$  est mesurable. L'intégrale est linéaire, c'est-à-dire qu'on a

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu.$$

Monotonie : l'intégrale est monotone, c'est-à-dire que si  $f \leq g$  sont deux fonctions mesurables, on a

$$\int f d\mu \leq \int g d\mu.$$

Fin du cours 6

## 3.2 Convergence d'intégrales

Grâce à cette nouvelle intégrale, nous allons pouvoir obtenir des résultats de convergence généraux, et notamment des résultats forts d'interversion de limites et d'intégrale.

Commençons par la notion même de convergence. Pour une suite de réels, pour lesquelles il existe une notion de convergence canonique

$$\lim_{n \rightarrow \infty} x_n = x \quad \Longleftrightarrow \quad \forall \varepsilon > 0, \text{ pour } n \text{ assez grand, } |x_n - x| \leq \varepsilon.$$

Au contraire, les suites de fonctions et de mesures sur lesquelles portent le reste du cours peuvent converger de façons diverses. En particulier, il est fondamental, en parlant de convergence de suite de fonctions et de variables aléatoires, de toujours préciser quel sens est donné à la convergence. Commençons par la forme de convergence de fonctions la plus intuitive, la convergence simple.

### Definition 34 : Convergence simple



Une suite  $(f_n)_{n \in \mathbb{N}}$  de fonctions de  $E$  dans  $\mathbb{R}$  converge simplement vers une fonction  $f$  si pour tout  $x \in E$ ,  $\exists \lim_{n \rightarrow \infty} f_n(x) = f(x)$ .

### Definition 35 : Convergence dans $\mathcal{L}^1$



On dira qu'une suite  $(f_n)_{n \in \mathbb{N}}$  de fonctions mesurables converge dans  $\mathcal{L}^1$  vers une fonction  $f$  si les  $f_n$  et  $f$  sont intégrables (i.e.  $f_n, f \in \mathcal{L}^1$ ), et si

$$\lim_{n \rightarrow \infty} \int |f_n - f| d\mu$$

existe et vaut 0.



**ATTENTION** : aucune de ces deux convergences n'implique l'autre! En fait, dans le contexte de convergence en intégrale, il est généralement préférable de parler de convergence presque-partout, puisque une fonction peut être modifiée arbitrairement sur un ensemble de mesure nulle sans changer son intégrale! (cf. Exercice 17). Toutefois, même la convergence presque-partout n'est pas nécessaire pour converger dans  $\mathcal{L}^1$  (cf. Remarque 24 ci-dessous)

**REMARQUE 13** : On peut évoquer une autre notion de convergence de fonctions, la convergence uniforme : une suite  $(f_n)_{n \in \mathbb{N}}$  de fonctions de  $E$  dans  $\mathbb{R}$  converge uniformément vers une fonction  $f$  si

$$\lim_{n \rightarrow \infty} \left[ \sup_{x \in E} |f_n(x) - f(x)| \right] = 0.$$

### Exercice 18 : densité des fonctions étagées

Soit  $f : E \rightarrow \bar{\mathbb{R}}_+$  une fonction mesurable positive réelle.

- 1) Montrer que  $f$  est la limite simple d'une suite de fonctions étagées positives.
- 2) On ne suppose plus  $f$  positive, montrer que  $f$  est limite simple d'une suite de fonctions étagées.
- 3) Soit  $f$  une fonction mesurable bornée. Montrer qu'il existe une suite de fonctions étagées qui converge uniformément vers  $f$ .

**SOLUTION** :

1) On définit pour tout  $n \geq 0$  et pour tout  $1 \leq k \leq n^2$  l'intervalle  $I_{n,k} = [(k-1)/n, k/n[$ . Posons  $A_{n,k} = f^{-1}(I_{n,k})$  et  $A_{n,\star} = f^{-1}([n, +\infty[)$ . On construit alors

$$\hat{f}_n = n \mathbf{1}_{A_{n,\star}} + \sum_{k=1}^{n^2} \frac{k-1}{n} \mathbf{1}_{A_{n,k}}.$$

C'est une suite de fonctions étagées, dont la limite simple est  $f$ .

- 2) La partie positive et la partie négative de  $f$ ,  $f_+$  et  $f_-$  sont mesurables positives, on peut donc leur appliquer le 1).
- 3) Notons avec  $\pm$  les quantités et ensembles relatives à  $f_{\pm}$ . Si  $f$  est bornée, il existe  $n$  tel que  $A_{n,\star}^{\pm} = \emptyset$ . Or pour tout  $k \leq n^2$ , pour tout  $x \in A_{n,k}$ ,

$$|\hat{f}_{\pm,n}(x) - f_{\pm}(x)| \leq 1/n.$$

La convergence de  $\hat{f}_n$  vers  $f$  est donc uniforme. □

**REMARQUE 14** : Comme c'est le cas de l'intégrale (voir Exercice 17), la limite dans  $\mathcal{L}^1$  d'une suite de fonctions  $(f_n)$  est toujours définie à un ensemble négligeable près : en effet, si  $f = g$  presque-partout, et si  $(f_n)$  converge vers  $f$  dans  $\mathcal{L}^1$ , alors  $(f_n)$  converge aussi vers  $g$  dans  $\mathcal{L}^1$ .

### Definition 36 : lim inf et lim sup d'une suite de réels

Soit  $(a_n)_{n \in \mathbb{N}}$  une suite réelle. Remarquons que les suites de termes généraux  $v_n = \inf_{p \geq n} a_p$  et  $w_n = \sup_{p \geq n} a_p$  sont respectivement croissante et décroissante, et par conséquent admettent une limite. On définit alors

$$\liminf_{n \rightarrow \infty} a_n := \lim_{n \rightarrow \infty} \inf_{p \geq n} a_p \quad \text{et} \quad \limsup_{n \rightarrow \infty} a_n := \lim_{n \rightarrow \infty} \sup_{p \geq n} a_p.$$

Ces deux quantités sont respectivement la plus petite et la plus grande *valeur d'adhérence* de la suite  $a_n$ .

Étant donnée une suite de fonctions réelles  $(f_n)_{n \in \mathbb{N}}$ ,  $f_n : E \rightarrow \mathbb{R}$ , on définit de manière analogue les fonctions  $\liminf f_n$  et  $\limsup f_n$  en posant pour tout  $x \in E$

$$\liminf f_n : x \mapsto \liminf_{n \rightarrow \infty} f_n(x) \quad \text{et} \quad \limsup f_n : x \mapsto \limsup_{n \rightarrow \infty} f_n(x).$$

Grâce à cette définition, on donne un premier résultat très pratique de convergence d'intégrales, le Lemme de Fatou.

### Proposition 14 : Lemme de Fatou, admis

Soit  $(f_n)_{n \in \mathbb{N}}$  une suite de fonctions *mesurables positives*  $E \rightarrow \mathbb{R}$ . La fonction  $f = \liminf f_n$  est mesurable et

$$\int_E \liminf_{n \rightarrow \infty} f_n d\mu \leq \liminf_{n \rightarrow \infty} \int_E f_n d\mu.$$

Donnons deux exemples où l'inégalité est stricte. Définissons sur  $\mathbb{R}$

$$f_n(x) = n \mathbf{1}_{]0, 1/n[}(x) \quad \text{et} \quad g_n(x) = (1/n) \mathbf{1}_{[0, n]}(x),$$

les  $(f_n)_{n \in \mathbb{N}}$  et les  $(g_n)_{n \in \mathbb{N}}$  sont toutes d'intégrale 1, mais  $\liminf f_n = \liminf g_n \equiv 0$ , et donc  $\int_E \liminf_{n \rightarrow \infty} f_n d\mu = 0$ .

On énonce maintenant deux résultats fondamentaux, les théorèmes de convergence monotone et dominée.

### Théorème 15 : Théorème de convergence monotone, admis

★

Soit  $(f_n)_{n \in \mathbb{N}}$  une suite *croissante* de fonctions *mesurables positives* de  $E$  dans  $\mathbb{R}$ , c'est-à-dire que  $\forall x \in E, \forall n \in \mathbb{N}, f_{n+1}(x) \geq f_n(x)$ . Alors, il existe une fonction *mesurable*  $f$  qui est limite simple de  $(f_n)_{n \in \mathbb{N}}$ , et on a

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu.$$

À noter que l'on ne fait aucune hypothèse d'intégrabilité, les deux quantités ci-dessus peuvent être infinies. De la même manière,  $f$  est à valeurs dans



$\mathbb{R} \cup \{+\infty\}$ .

**REMARQUE 15** : En appliquant ce théorème à la suite  $(-f_n)_{n \in \mathbb{N}}$ , on voit immédiatement que le théorème de convergence monotone s'applique également à une suite décroissante de fonctions négatives.

### Exercice 19

Trouver un contre-exemple au théorème de convergence monotone si on ne suppose pas les fonctions positives.

**SOLUTION** : Il suffit de considérer la suite de fonctions  $f_n = -\mathbf{1}_{[n, +\infty[}$ , qui est croissante et converge simplement vers  $f \equiv 0$ , et pourtant dont toutes les intégrales sont égales à  $-\infty$ .  $\square$

**REMARQUE 16** : Nous ne verrons pas la preuve de ce théorème, toutefois grâce au Lemme de Fatou, la convergence est très facile à obtenir. Notons  $I_n = \int f_n d\mu$  et  $I = \int f d\mu$ . La suite  $(f_n)_{n \in \mathbb{N}}$  étant croissante, on a  $f_n \leq f$ , et donc par monotonie  $I_n \leq I$ . L'inégalité  $I \leq I_n$  est une conséquence directe du Lemme de Fatou, puisque pour toute suite réelle  $(u_n)_{n \rightarrow \infty}$  admettant une limite, on a

$$\liminf_{n \rightarrow \infty} u_n = \limsup_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} u_n.$$

On énonce maintenant un des résultats les plus importants de convergence sous le signe intégral. Ce résultat, très utile, énonce que si une suite de fonction et sa limite simple peuvent toutes majorées par une fonction intégrable, alors la convergence a également lieu dans  $\mathcal{L}^1$ .

### Théorème 16 : Théorème de convergence dominée, admis ★

Soit une suite  $(f_n)_{n \in \mathbb{N}}$  de fonctions de  $E$  dans  $\mathbb{R}$  convergeant simplement vers une fonction  $f$ . On suppose qu'il existe une fonction intégrable  $g \in \mathcal{L}^1$  telle que

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

Alors,  $f_n$  converge vers  $f$  dans  $\mathcal{L}^1$ . En particulier,

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu.$$

**REMARQUE 17** : Reprenons les exemples d'inégalité stricte du Lemme de Fatou,

$$f_n(x) = n\mathbf{1}_{]0, 1/n[}(x) \quad \text{et} \quad g_n(x) = (1/n)\mathbf{1}_{[0, n]}(x),$$

qui convergent simplement vers  $f \equiv g \equiv 0$  sur  $]0, +\infty[$ . Comme l'inégalité est stricte pour le Lemme de Fatou, en particulier le théorème de convergence dominée de

doit pas pouvoir s'appliquer; en réalité, une domination naturelle de ces deux suites de fonctions serait par la fonction  $h : x \mapsto 1/x$  sur l'intervalle  $]0, +\infty[$ . On peut rapidement vérifier que  $|f_n|, |g_n| \leq h$ , mais la fonction  $h$  n'est *pas* intégrable, que ce soit en 0 (suite  $(f_n)_n$ ) ou en  $+\infty$  (suite  $(g_n)_n$ ).

————— *Fin du cours 7* —————

## 4 Variables aléatoires réelles

### 4.1 Préambule : caractérisation de mesures

Nous n'allons pas démontrer le théorème de Carathéodory. Toutefois, il est important de savoir qualifier les mesures, en particulier les mesures sur la tribu borélienne. Les boréliens généraux étant extrêmement difficiles à caractériser, on va avoir besoin d'outils pour montrer des égalités entre mesures, pour pouvoir dire que si deux mesures coïncident sur une classe d'événements  $C$ , elles sont égales. Le seul résultat à retenir de ce paragraphe est la conséquence du Lemme des classes monotones donnée dans le Corollaire 18, et en particulier ses conséquences sur les mesures sur la tribu borélienne et sur la fonction de répartition des variables aléatoires réelles (cf. Proposition 20).

#### Definition 37 : $\pi$ -système

Une classe  $C$  de parties de  $E$  est un  $\pi$ -système si elle est stable par intersection finie :

$$\{A \in C \text{ et } B \in C\} \Rightarrow \{A \cap B \in C\}.$$

**EXEMPLE** : L'ensemble des intervalles ouverts est un  $\pi$ -système, de même que l'ensemble des singletons.

#### Definition 38 : Classe monotone

Une classe  $\mathcal{M}$  de parties de  $E$  est un  $\lambda$ -système, ou *classe monotone* si elle contient  $E$ , et est stable par différence et par union croissante :

$$\{A \in \mathcal{M} \text{ et } B \in \mathcal{M} \text{ et } B \subset A\} \Rightarrow \{A \setminus B \in \mathcal{M}\}.$$

$$\{\forall n \geq 0, A_n \in \mathcal{M} \text{ et } A_n \subset A_{n+1}\} \Rightarrow \{\cup_{n \geq 0} A_n \in \mathcal{M}\}.$$

**EXEMPLE** : Étant données deux mesures de probabilité  $\mathbb{P}$  et  $\mathbb{Q}$  sur  $(\Omega, \mathcal{F})$ , la classe  $\mathcal{M} = \{A \in \mathcal{F} \mid \mathbb{P}(A) = \mathbb{Q}(A)\}$  est une classe montone.

#### Lemme 17 : Lemme des classes monotones (admis)

La plus petite classe monotone contenant le  $\pi$ -système  $C$  est la tribu  $\sigma(C)$  engendrée par  $C$ .

#### Corollaire 18 : Unicité des mesures de probabilités



Deux mesures de probabilité  $\mathbb{P}$  et  $\mathbb{Q}$  sur  $(\Omega, \mathcal{F})$  qui coïncident sur un  $\pi$ -système  $C \subset \mathcal{F}$  coïncident également sur la tribu  $\sigma(C)$  engendrée par  $C$ .

**PREUVE :** Considérons deux mesures de probabilité  $\mathbb{P}$  et  $\mathbb{Q}$  qui coïncident sur un  $\pi$ -système  $C \subset \mathcal{F}$ . La classe  $\mathcal{M} = \{A \in \mathcal{F} \mid \mathbb{P}(A) = \mathbb{Q}(A)\}$  est une classe monotone contenant  $C$ , d'après le Lemme des classes monotones,  $\mathcal{M} = \sigma(C)$ .  $\square$

En pratique, ce corollaire est très utile! Il signifie que pour caractériser une mesure (par exemple la loi d'une variable aléatoire), il n'y a pas besoin de la connaître sur la tribu tout entière, mais il suffit de la caractériser sur un  $\pi$ -système engendrant cette tribu. La tribu borélienne, par exemple, est engendrée par le  $\pi$ -système

$$C = \{ ] - \infty, x], x \in \mathbb{R} \},$$

on en déduira plus loin dans le cours que la fonction de répartition caractérise entièrement la distribution d'une variable aléatoire! Pour l'instant, on peut toutefois en déduire l'unicité de la mesure de Lebesgue, qui est une des conclusions du théorème de Carathéodory :

### Théorème 19 : Théorème de Carathéodory, admis ★

Il existe une unique mesure  $\lambda$  sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  telle que pour tout  $a < b \in \mathbb{R} \cup \{-\infty, +\infty\}$ ,

$$\lambda(]a, b[) = b - a$$

Cette mesure est appelée *mesure de Lebesgue*.

Comme on l'a vu précédemment, L'intégrale contre la mesure de Lebesgue coïncide avec l'intégrale de Riemann pour toute fonction telle que les deux intégrales sont bien définies (typiquement, les fonctions continues par morceaux sur un segment). On pourra donc écrire pour la mesure de Lebesgue  $dx = d\lambda(x) = \lambda(dx)$  pour l'élément infinitésimal d'intégration. En l'absence d'ambiguïté, lorsque l'on intègre contre la mesure de Lebesgue, on utilisera la notation la plus simple, c'est-à-dire la première, i.e.

$$\int f d\lambda = \int f(x) dx.$$

En dehors de la construction différentes des deux intégrales, comme on l'a vu, l'intérêt fondamental de la mesure de Lebesgue est que l'intégrale par rapport à  $\lambda$  est définie pour la classe des fonctions *mesurables*, qui est une classe bien plus générale que celle des fonctions continues par morceau sur un segment! Le Lemme des classes monotones garantit que la mesure de Lebesgue est unique, puisque l'ensemble des intervalles ouverts est un  $\pi$ -système qui engendre la tribu borélienne. Caractériser une mesure sur l'ensemble des intervalles ouverts est donc suffisant pour la caractériser sur la tribu borélienne tout entière.

————— Fin du cours 8 + Retour DS1 —————

## 4.2 Variable aléatoire réelles et fonction de répartition

On a maintenant tous les outils nécessaires pour introduire les probabilités continues. Dans tout ce qui suit, on fixe un *espace de probabilités*  $(\Omega, \mathcal{F}, \mathbb{P})$ .

On dira qu'une propriété  $\mathcal{P}$  est vérifiée *presque-sûrement* s'il existe  $F \in \mathcal{F}$  tel que  $\mathbb{P}(F) = 1$  et tel que  $\mathcal{P}$  est vérifiée pour tout  $\omega \in F$ .

### Definition 39 : variable aléatoire réelle, fonction de répartition ★

Comme nous l'avons déjà vu dans la Déf. 8, une variable aléatoire réelle est une fonction mesurable  $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . On note alors  $\mathbb{P}_X(\cdot) = \mathbb{P}(X \in \cdot)$  sa distribution, définie sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

Les *variables à densité* introduites dans le paragraphe 2.5 sont un cas particulier de variables aléatoires réelles  $X$  pour lesquelles  $\mathbb{P}_X(A) = \int_A f_X(x)dx$ , pour une fonction  $f_X$  borélienne. La fonction  $f_X$  est alors appelée la *densité par rapport à la mesure de Lebesgue* de la loi  $\mathbb{P}_X$ , et on écrit formellement  $\mathbb{P}_X(dx) = f_X(x)dx$ .

La *fonction de répartition* de la variable aléatoire  $X$  est la fonction  $F_X : \mathbb{R} \rightarrow [0, 1]$ , définie par

$$F_X(x) = \mathbb{P}(X \leq x).$$

### Proposition 20 : Propriétés de la fonction de répartition ★

La fonction de répartition  $F_X$  d'une v.a. réelle est croissante, continue à droite, et satisfait

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow +\infty} F_X(x) = 1.$$

Enfin, la fonction de répartition caractérise entièrement la loi d'une v.a. réelle, c'est-à-dire que si  $X$  et  $Y$  ont même fonction de répartition,  $F_X = F_Y$ , alors  $\mathbb{P}_X = \mathbb{P}_Y$ .

**PREUVE :** La croissance est immédiate, et cette dernière garantit l'existence des deux limites, ainsi que l'identité

$$\lim_{x \in \mathbb{R} \rightarrow -\infty} F_X(x) = \lim_{n \in \mathbb{N} \rightarrow -\infty} F_X(n).$$

La continuité à droite est une conséquence directe de la continuité des mesures par rapport aux intersections décroissantes : pour toute suite positive  $\varepsilon_n \rightarrow 0$ , pour tout  $x \in \mathbb{R}$ ,

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\cap_{n \in \mathbb{N}} \{X \leq x + \varepsilon_n\}) = \lim_{n \rightarrow \infty} \mathbb{P}(X \leq x + \varepsilon_n) = \lim_{n \rightarrow \infty} F(x + \varepsilon_n).$$

On applique ensuite la Proposition 6, puisque  $\mathbb{P}_X$  est une mesure de probabilité, et tout événement a donc probabilité finie,

$$\lim_{x \rightarrow -\infty} F_X(x) = \lim_{n \rightarrow -\infty} \mathbb{P}_X(] - \infty, n]) = \lim_{n \rightarrow \infty} \mathbb{P}_X \left( \bigcap_{\substack{m \in \mathbb{N}, \\ m \leq n}} ] - \infty, m[ \right) = \mathbb{P}_X(\emptyset) = 0.$$

La seconde limite se montre de manière analogue.

Enfin, la dernière affirmation est une conséquence du Corollaire 18 du Lemme des classes monotones : le  $\pi$ -système

$$C = \{ ] - \infty, x], x \in \mathbb{R} \}$$

engendre la tribu des boréliens (c'est-à-dire  $\sigma(C) = \mathcal{B}(\mathbb{R})$ ), et si  $F_X = F_Y$ , on a par définition que  $\mathbb{P}_X = \mathbb{P}_Y$  sur  $C$ . Le Corollaire 18 garantit alors que  $\mathbb{P}_X = \mathbb{P}_Y$  sur  $\mathcal{B}(\mathbb{R})$  tout entier.  $\square$

### 4.3 Moments d'une variable aléatoire réelle

La définition de l'intégrale de Lebesgue dans la section précédente nous permet désormais de définir l'espérance pour des variables aléatoires générales. Nous nous concentrons sur des variables aléatoires réelles, mais la notion d'espérance est beaucoup plus générale.

#### Definition 40 : Espérance d'une v.a. réelle ★

Rappelons (cf. Définition 8) que la loi  $\mathbb{P}_X$  d'une variable aléatoire  $X : \Omega \rightarrow E$  est définie comme la mesure  $\mathbb{P}_X(A) = \mathbb{P}(X \in A)$ , pour  $A \in \mathcal{T}$ . On dit qu'une variable aléatoire *réelle*  $X : \Omega \rightarrow \mathbb{R}$  est *intégrable*, ou *dans*  $\mathcal{L}^1$ , si  $\int_{\mathbb{R}} |x| \mathbb{P}_X(dx) < \infty$ . On définit alors l'*espérance*, ou *moyenne*, de  $X$  par

$$\mathbb{E}(X) = \int_{\mathbb{R}} x \mathbb{P}_X(dx). \quad (1)$$

On dira que la variable  $X$  est *centrée* si  $\mathbb{E}(X) = 0$ . On admettra que l'espérance a les mêmes propriétés que l'intégrale :

$$\textit{monotonie} : \quad \forall X, Y \text{ v.a.r. } X \leq Y \text{ p.s.} \implies \mathbb{E}(X) \leq \mathbb{E}(Y),$$

$$\textit{linéarité} : \quad \forall a, b \in \mathbb{R} \text{ et } X, Y \text{ v.a.r. } \mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

En particulier, on remarque qu'une variable aléatoire *bornée* est *intégrable*.

#### Exercice 20

Dans cet exercice, on suppose à nouveau le théorème de Fubini (cf. Exercice 10) : deux intégrales successives d'une fonction positive peuvent être interverties.

- 1) Montrer que si  $X$  est intégrable,  $\mathbb{E}(X) = \int_0^{+\infty} (1 - F_X(x))dx - \int_{-\infty}^0 F_X(x)dx$ .
- 2) En commençant par une variable aléatoire  $X$  positive, montrer la formule

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) \mathbb{P}(d\omega). \quad (2)$$

- 3) En déduire que l'espérance a les mêmes propriétés que l'intégrale : elle est *linéaire*

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y),$$

et monotone

$$X \leq Y \implies \mathbb{E}(X) \leq \mathbb{E}(Y).$$

**SOLUTION** : Pour traiter cet exercice, il faut sans cesse se ramener à des indicatrices d'événements, dont on connaît explicitement les intégrales.

1) On commence par montrer le résultat pour  $X : \Omega \rightarrow \mathbb{R}^+$  positive. On remarque que  $x = \int_0^x dy = \int_0^{+\infty} \mathbf{1}_{[0,x]}(y) dy$ , et on peut alors écrire grâce au théorème de Fubini

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{+\infty} x \mathbb{P}_X(dx) = \int_0^{+\infty} \left[ \int_0^{+\infty} \mathbf{1}_{[0,x]}(y) dy \right] \mathbb{P}_X(dx) \\ &= \int_0^{+\infty} \left[ \int_0^{+\infty} \mathbf{1}_{[0,x]}(y) \mathbb{P}_X(dx) \right] dy. \end{aligned}$$

Or pour une variable positive, on peut écrire

$$\int_0^{+\infty} \mathbf{1}_{[0,x]}(y) \mathbb{P}_X(dx) = \int_0^{+\infty} \mathbf{1}_{\{y < x\}} \mathbb{P}_X(dx) = \mathbb{P}(X > y) = 1 - F_X(y).$$

Pour une variable  $X$  négative, on obtient de manière analogue

$$\mathbb{E}(X) = - \int_{-\infty}^0 F_X(y) dy$$

On sépare alors  $X$  en partie positive et partie négative,  $X_+ = X \mathbf{1}_{X \geq 0}$ ,  $X_- = X \mathbf{1}_{X < 0}$ , et on vérifie que pour tout  $t \geq 0$ ,  $F_{X_+}(t) = F_X(t)$ , et pour  $t \leq 0$ ,  $F_{X_-}(t) = F_X(t)$ , ce qui conclut la preuve.

2) Pour une variable  $X$  positive, on écrit

$$\int_{\Omega} X(\omega) \mathbb{P}(d\omega) = \int_{\Omega} \left[ \int_0^{+\infty} \mathbf{1}_{[0,X(\omega)]}(y) dy \right] \mathbb{P}(d\omega),$$

et on inverse encore les deux intégrales par le théorème de Fubini, pour obtenir

$$\int_{\Omega} X(\omega) \mathbb{P}(d\omega) = \int_0^{+\infty} \left[ \int_{\Omega} \mathbf{1}_{[0,X(\omega)]}(y) \mathbb{P}(d\omega) \right] dy = \int_0^{+\infty} \mathbb{P}(X > y) dy = \int_0^{+\infty} (1 - F_X(y)) dy,$$

on retrouve donc le résultat de la question 1. Le cas où  $X$  est négatif se montre de manière analogue, ce qui prouve la formule alternative

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

3) Avec cette nouvelle formule, la linéarité et la monotonie découlent immédiatement de la linéarité et la monotonie de l'intégrale.  $\square$

On peut maintenant énoncer le théorème de transfert pour des variables aléatoires générales, déjà énoncé pour les variables discrètes.

### Théorème 21 : Théorème de transfert (admis)



Soit  $X : \Omega \rightarrow \mathbb{R}$  une variable aléatoire réelle et  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  mesurable. Alors, la variable aléatoire  $\varphi(X)$  est intégrable (par rapport à  $\mathbb{P}$ ) si et seulement si  $\varphi$  est intégrable par rapport à  $\mathbb{P}_X$ , et on a alors

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

**EXEMPLE** : soit  $X : \Omega \rightarrow \mathbb{R}$  une variable aléatoire réelle, pour tout ensemble borélien  $B$ , on peut définir la variable aléatoire

$$\mathbf{1}_{\{X \in B\}} : \omega \mapsto \mathbf{1}_B(X(\omega)).$$

Par le théorème de transfert, son espérance est donnée par

$$\mathbb{E}(\mathbf{1}_{\{X \in B\}}) = \int_{\mathbb{R}} \mathbf{1}_B(x)\mathbb{P}_X(dx) = \int_B \mathbb{P}_X(dx) = \mathbb{P}(X \in B).$$

On vient de donner la définition, grâce à l'intégrale de Lebesgue, de l'espérance d'une variable aléatoire réelle. Voyons maintenant comment cette définition rejoint celle que nous avons déjà vu pour les variables à densité, et les variables discrètes. Commençons par une variable à densité  $X$ , dont la mesure  $\mathbb{P}_X$  est définie pour tout borélien  $B$  par

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \int_B \mathbb{P}_X(dx) = \int_B f_X(x)dx.$$

Cela nous permet d'écrire formellement  $\mathbb{P}_X(dx) = f_X(x)dx$ . En transposant cette identité dans notre définition générale de l'espérance, (1), on retrouve bien la Définition 21 de l'espérance d'une variable à densité :

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f_X(x)dx.$$

Considérons maintenant les variables discrètes. Bien qu'on puisse les définir sur la tribu borélienne, cette dernière n'est pas un choix naturel pour le cas discret, comme on l'a déjà vu. On considère donc une variable aléatoire discrète  $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{N}, \mathcal{P}(\mathbb{N}))$ . Notez que l'on a muni  $\mathbb{N}$  de la tribu *discrète*  $\mathcal{P}(\mathbb{N})$ , ce qui pour des variables discrètes n'est pas un souci. On suppose dans un premier temps que  $X$  ne prend qu'un nombre fini de valeurs,  $X \in \{0, \dots, n\}$ . Alors, la variable aléatoire  $X$  est une *fonction étagée* sur  $(\Omega, \mathcal{F}, \mathbb{P})$ , puisqu'elle peut s'écrire comme

$$X(\omega) = \sum_{k=0}^n k \mathbf{1}_{\{X=k\}}(\omega),$$

et les événements  $\{X = k\} = X^{-1}(\{k\})$  sont mesurables puisque  $X$  est une variable aléatoire. Or on connaît l'intégrale d'une fonction étagée contre une mesure finie  $\mathbb{P}$  (voir Définition 29), et on peut donc écrire

$$\mathbb{E}(X) = \int_{\{1, \dots, n\}} x \mathbb{P}_X(dx) = \int_{\Omega} X(\omega) \mathbb{P}(d\omega) = \sum_{k=0}^n k \mathbb{P}(X = k).$$



Si maintenant  $X$  est intégrable mais n'est pas bornée, on approche  $X$  par les fonctions étagées  $X^{(n)} = X\mathbf{1}_{X \leq n}$ . Alors, grâce au théorème de convergence dominée, qui s'applique car  $|X|$  est intégrable, et on a à la fois  $|X| \leq |X|$  et  $|X^{(n)}| \leq |X|$ , on peut écrire

$$\mathbb{E}(X) = \lim_{n \rightarrow \infty} \mathbb{E}(X^{(n)}) = \lim_{n \rightarrow \infty} \sum_{k=0}^n k\mathbb{P}(X = k) = \sum_{k=0}^{\infty} k\mathbb{P}(X = k).$$

On voit bien dans ces deux exemples que l'espérance d'une variable aléatoire peut être vue comme une intégrale sur l'univers  $\Omega$  (ce qui nous a permis de retrouver l'expression de l'espérance d'une variable aléatoire discrète), ou comme une intégrale sur l'espace d'arrivée de  $X$ , ce qui nous permet de retrouver l'expression de l'espérance des variables à densité.

————— Fin du cours 9 —————

### Definition 41 : Moments d'une v.a. réelle ★

Étant donné  $k \in \mathbb{N}$ , le  $k$ -ème moment d'une variable aléatoire  $X : \Omega \rightarrow \mathbb{R}$  est donné, s'il existe, par

$$\mathbb{E}(X^k) = \int_{\mathbb{R}} x^k \mathbb{P}_X(dx).$$

On dira que  $X$  admet un  $k$ -ème moment fini si  $X^k$  est intégrable. En particulier, une variable a son premier moment fini si et seulement si elle est *intégrable*. Pour une variable admettant un second moment fini, on dira aussi qu'elle est *de carré intégrable*.

Comme nous le verrons plus tard, (cf. Théorème 22 et Exercice 22 ci-dessous) une variable aléatoire *de carré intégrable* est *intégrable*. En particulier, supposer que  $X$  est de carré intégrable nous permet de définir sa variance :

### Definition 42 : Variance d'une v.a. réelle ★

Soit  $X$  une variable aléatoire réelle de second moment fini, sa variance est définie par

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \int_{\mathbb{R}} x^2 \mathbb{P}_X(dx) - \mathbb{E}(X)^2.$$

Son écart-type est défini comme  $\sigma(X) = \sqrt{\text{Var}(X)}$ . On dira parfois que la variable  $X$  est *réduite* si  $\sigma(X) = 1$ .

### Exercice 21

Soit  $Y$  une variable aléatoire presque-sûrement positive, montrer que  $\mathbb{E}(Y) = 0 \Rightarrow \mathbb{P}(Y = 0) = 1$ . Soit  $X$  une variable aléatoire réelle de carré intégrable, on suppose que  $\text{Var}(X) = 0$ . En déduire qu'il existe une constante  $c \in \mathbb{R}$  telle que  $\mathbb{P}(X = c) = 1$ . Autrement dit, une variable aléatoire de variance nulle est

presque sûrement constante.

**SOLUTION** : Commençons par montrer qu'une variable positive d'espérance nulle est nulle presque sûrement. Soit  $Y$  une variable aléatoire positive, supposons que  $\mathbb{E}(Y) = 0$ . Soit  $F_Y$  sa fonction de répartition, on va montrer que  $F_Y(0) = 1$ . On rappelle que  $F_Y$  est croissante et continue à droite, on en déduit que  $F_Y(0) = \lim_{x \downarrow 0} F_Y(x) \leq 1$ . Supposons que  $F_Y(0) < 1$ , il doit alors exister  $\varepsilon > 0$  tel que  $F_Y(\varepsilon) < 1$ . En particulier, cela signifie que  $\mathbb{P}(Y > \varepsilon) > 0$ , et donc  $\mathbb{E}(Y) \geq \varepsilon \mathbb{P}(Y > \varepsilon) > 0$ , ce qui est impossible par hypothèse. Par conséquent, on doit avoir  $F_Y(0) = 1$ , donc  $\mathbb{P}(Y \leq 0) = 1$ , ce qui, comme  $Y$  est positive presque-sûrement, implique  $\mathbb{P}(Y = 0) = 1$ .

Pour montrer la seconde affirmation, on applique le résultat précédent à la variable  $Y = (X - \mathbb{E}(X))^2$ , qui est positive et donc doit être nulle presque-sûrement. Par conséquent, si  $\text{Var}(X) = 0$ , on peut en déduire que  $X$  est presque-sûrement égale à sa variance.  $\square$

#### 4.4 Bonus : les variables discrètes, variables réelles comme les autres

Nous avons maintenant défini les variables aléatoires réelles dans un cadre général. On a déjà vu (cf. Définition 39) que les variables à densité en sont des cas particuliers, dont la loi prend la forme  $\mathbb{P}_X(A) = \int_A f_X(x) dx$ . Nous allons maintenant donner une nouvelle perspectives sur les variables aléatoires discrètes. Pour cela, on revient sur la notion de mesure de Dirac sur  $\mathbb{R}$ .

##### Definition 43 : Mesure de Dirac sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

Soit  $x \in \mathbb{R}$ , on définit la mesure de Dirac  $\delta_x : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  comme la mesure de probabilités donnée par

$$\delta_x(B) = \begin{cases} 1 & \text{si } x \in B \\ 0 & \text{sinon.} \end{cases}$$

**REMARQUE 18** : Soit  $x \in \mathbb{R}$ , et soit  $X$  une variable aléatoire réelle, on suppose que la distribution de  $X$  est donnée par  $\delta_x$ . On voit alors facilement

$$\mathbb{P}(X = x) = \mathbb{P}_X(\{x\}) = \delta_x(\{x\}) = 1.$$

Autrement dit, une variable aléatoire dont la distribution est une mesure de Dirac en  $x$  est presque-sûrement égale à  $x$ .

Par ailleurs, on peut également facilement remarquer que  $X$  est intégrable, et que  $\mathbb{E}(X) = x$ . En effet, pour une mesure de Dirac en  $x$ ,  $\mathbb{R} \setminus \{x\}$  est négligeable, par conséquent, d'après la Remarque 12, on a

$$\mathbb{E}(X) := \int_{\mathbb{R}} y \delta_x(dy) = \int_{\{x\}} y \delta_x(dy) = x \delta_x(\{x\}) = x,$$

puisque sur  $\{x\}$  la fonction  $y \mapsto y$  est trivialement étagée (c'est le cas de toute fonction prenant ses valeurs dans un ensemble fini). On remarque facilement en suivant le même raisonnement que  $\text{Var}(X) = 0$ .

#### Definition 44 : Variables aléatoires discrètes à valeurs dans $\mathbb{N}$

Comme nous l'avons vu plus tôt dans le cours (cf. Définition 9), une variable aléatoire discrète  $X$  à valeurs entières est une fonction mesurable  $X : \Omega \rightarrow \mathbb{N}$ , où  $\mathbb{N}$  est muni de la tribu discrète. On note alors  $\mu_k = \mathbb{P}(X = k)$ , qui caractérise entièrement sa loi.

Toutefois, si  $X : \Omega \rightarrow \mathbb{N}$  est mesurable, elle peut être également vue comme une fonction mesurable  $X : \Omega \rightarrow \mathbb{R}$ , puisque  $\mathbb{N} \subset \mathbb{R}$  et que les singletons  $\{k\}$  sont des ensembles mesurables de la tribu borélienne. La variable aléatoire  $X$  est alors à valeurs dans  $\mathbb{R}$  muni de la tribu borélienne  $\mathcal{B}(\mathbb{R})$ , et sa distribution  $\mathbb{P}_X$  sur  $\mathcal{B}(\mathbb{R})$  est donnée en fonction du vecteur  $\mu_k$  par la mesure sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

$$\mathbb{P}_X = \sum_{k \in \mathbb{N}} \mu_k \delta_k.$$

Naturellement, cette définition est généralisable aux variables discrètes réelles  $X$ , prenant leurs valeurs dans un sous ensemble dénombrable  $E \subset \mathbb{R}$  quelconque de  $\mathbb{R}$ . On peut alors écrire la loi de  $X$  comme

$$\mathbb{P}_X = \sum_{e \in E} \mathbb{P}(X = e) \delta_e.$$

#### 4.5 Propriétés et inégalités classiques

On commence par une inégalité de convexité. Rappelons qu'une fonction  $\varphi$  sur  $\mathbb{R}$  est *convexe* si elle est au-dessus de toutes ses tangentes, c'est à dire si

$$\varphi(\lambda x + (1 - \lambda)y) \leq \lambda\varphi(x) + (1 - \lambda)\varphi(y) \quad \forall x, y \in \mathbb{R}, \lambda \in [0, 1].$$

Une fonction  $\varphi$  de classe  $C^2$  est convexe si et seulement si  $\varphi'' \geq 0$ .

#### Théorème 22 : Inégalité de Jensen, admis

Soit  $X$  une variable aléatoire, et  $\varphi$  une fonction convexe. Alors,

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

Cette inégalité reste valable si  $X$  n'est pas intégrable.

**REMARQUE 19** : On voit naturellement que l'inégalité est inversée si  $\varphi$  est concave, puisque

$$\varphi \text{ convexe} \iff -\varphi \text{ concave.}$$

### Exercice 22

Utiliser l'inégalité de Jensen pour montrer que si  $X$  admet un  $k$ -ème moment, alors elle admet un  $j$ -ème moment pour tout  $j \leq k$ .

**SOLUTION** : Il suffit de remarquer que pour tout entiers positifs  $1 \leq j \leq k$ , la fonction  $\varphi : x \mapsto x^{j/k}$  est concave sur  $\mathbb{R}_+$  car  $j/k \leq 1$ . On en déduit que

$$\mathbb{E}(|X|^j) = \mathbb{E}\left[ (|X|^k)^{j/k} \right] \leq \mathbb{E}\left[ |X|^k \right]^{j/k} < \infty,$$

car  $X$  admet un  $k$ -ème moment par hypothèse. □

### Proposition 23 : Inégalité de Markov ★

Soit  $X : \Omega \rightarrow \mathbb{R}_+$  une variable aléatoire *positive*, pour tout  $a > 0$ , on a

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

**REMARQUE 20** : Cette inégalité reste valide lorsque  $X$  n'est pas intégrable, mais elle devient triviale, puisque le membre de droite est alors infini!

**PREUVE** : Fixons  $a > 0$ , comme  $X$  est une v.a. *positive*, on commence par écrire l'inégalité

$$a\mathbf{1}_{\{X \geq a\}} \leq X,$$

valide pour tout  $\omega$ . L'espérance étant monotone, on peut prendre l'espérance de part et d'autre :

$$\mathbb{E}(a\mathbf{1}_{\{X \geq a\}}) = a\mathbb{P}(X \geq a) \leq \mathbb{E}(X).$$

On divise ensuite de part et d'autre par  $a$  pour conclure la preuve. □

### Exercice 23

Trouver un contre-exemple à l'inégalité de Markov si  $X$  n'est pas positive.

**SOLUTION** : Il suffit de prendre une variable centrée, comme une loi normale centrée réduite, pour laquelle  $\mathbb{E}(X) = 0$  et pourtant  $\mathbb{P}(X \geq a) > 0$  pour tout  $a > 0$ . □

L'inégalité de Markov est extrêmement utile, c'est l'outil standard pour contrôler la probabilité qu'une variable aléatoire intégrable soit grande. Une de ses conséquences est l'inégalité de Bienaymé-Chebychev, qui permet de contrôler la probabilité qu'une variable aléatoire soit loin de sa moyenne.

### Proposition 24 : Inégalité de Bienaymé-Chebychev

Soit  $X : \Omega \rightarrow \mathbb{R}$  une variable aléatoire de second moment fini, et  $a > 0$ , on a

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

**REMARQUE 21** : Là encore, l'inégalité reste valide, mais inintéressante si  $X$  n'admet pas de second moment. Par ailleurs, l'inégalité de Bienaymé-Chebychev peut se réécrire de la manière suivante : pour tout  $k > 0$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq k\sigma) \leq \frac{1}{k^2},$$

où  $\sigma < \infty$  est l'écart type de  $X$ .

**PREUVE** : On commence par remarquer que

$$|X - \mathbb{E}(X)| \geq a \iff (X - \mathbb{E}(X))^2 \geq a^2.$$

On peut donc appliquer l'inégalité de Markov à la v.a. positive  $Y := (X - \mathbb{E}(X))^2$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) = \mathbb{P}(Y \geq a^2) \underset{\text{inég. Markov}}{\leq} \frac{\mathbb{E}(Y)}{a^2} = \frac{\text{Var}(X)}{a^2}.$$

□

————— Fin du cours 10 —————

## 5 Indépendance, loi des grands nombres et TCL

### 5.1 Indépendance de variables aléatoires

On revient maintenant sur la notion d'indépendance, que nous avons déjà vue pour les événements et pour les variables aléatoires discrètes.

#### Definition 45 : Variables aléatoires indépendantes ★

Deux variables aléatoires réelles  $X$  et  $Y$  sont *indépendantes* si pour tous boréliens  $A, B \in \mathcal{B}(\mathbb{R})$ ,

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

Une famille  $(X_n)_{n \in \mathbb{N}}$  de variables aléatoires est dite (*mutuellement*) *indépendante* si pour tout  $k \in \mathbb{N}$ , pour tous  $n_1 \neq n_2 \neq \dots \neq n_k \in \mathbb{N}$ , et pour tous boréliens  $A_1, \dots, A_k \in \mathcal{B}(\mathbb{R})$

$$\mathbb{P}(X_{n_i} \in A_i \forall i = 1 \dots, k) = \prod_{i=1}^k \mathbb{P}(X_{n_i} \in A_i).$$

Si la famille  $(X_n)_{n \in \mathbb{N}}$  est (*mutuellement*) *indépendante* et de plus les  $X_n$  ont tous la même loi  $\mathbb{P}_X$ , on dira que la famille  $(X_n)_{n \in \mathbb{N}}$  est *indépendante et identiquement distribuée*, ou *i.i.d.*.

**NOTATION** : On notera  $X \perp\!\!\!\perp Y$  pour "X et Y sont indépendantes".



**ATTENTION** : Pour que la famille  $(X_n)_{n \in \mathbb{N}}$  soit indépendante, il ne suffit pas que les  $(X_n)_{n \in \mathbb{N}}$  soient indépendantes deux à deux, i.e.  $X_n \perp\!\!\!\perp X_m \forall n \neq m$ . Considérons deux variables indépendantes  $X$  et  $Y$  de même loi, où  $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$ . Alors, en posant  $Z = XY$ , la famille  $(X, Y, Z)$  est indépendante deux à deux, mais pas indépendante.

**REMARQUE 22** : En pratique, pour vérifier l'indépendance, il suffit de considérer des boréliens de type  $A = ]a, +\infty[$ . Par exemple,

$$X \perp\!\!\!\perp Y \iff \mathbb{P}(X > a, Y > b) = \mathbb{P}(X > a)\mathbb{P}(Y > b) \quad \forall a, b \in \mathbb{R}.$$

#### Proposition 25 : Espérance et indépendance, admis ★

Soient  $X$  et  $Y$  deux variables aléatoires indépendantes à valeurs dans  $\mathbb{R}$ , et  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  deux fonctions boréliennes. Alors,  $f(X)$  et  $g(Y)$  sont également indépendantes. Par ailleurs,

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y)).$$

De manière analogue, si la famille  $(X_n)_{n \in \mathbb{N}}$  est indépendante, alors  $\forall k \in \mathbb{N}$ ,

$\forall n_1 \neq n_2 \neq \dots \neq n_k \in \mathbb{N}$ , et pour toutes fonctions boréliennes  $f_1, \dots, f_n$ , on a

$$\mathbb{E} \left( \prod_{i=1}^k f_i(X_{n_i}) \right) = \prod_{i=1}^k \mathbb{E}(f_i(X_{n_i})).$$

### Corollaire 26 : Variance et somme de variables indépendantes ★

Soient  $X$  et  $Y$  deux variables aléatoires indépendantes de carré intégrable, on a

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Plus généralement, si  $(X_n)_{n \leq N}$  est une famille de variables aléatoires deux-à-deux indépendantes toutes de carré intégrable, on a

$$\text{Var} \left( \sum_{n=1}^N X_n \right) = \sum_{n=1}^N \text{Var}(X_n).$$

**PREUVE** : Par linéarité de l'espérance,

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}((X + Y)^2) - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \\ &= \mathbb{E}(X^2) + \mathbb{E}(Y^2) - 2\mathbb{E}(XY) - \mathbb{E}(X)^2 - \mathbb{E}(Y)^2 + 2\mathbb{E}(X)\mathbb{E}(Y). \end{aligned}$$

Par indépendance,  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ , ce qui montre l'identité. La seconde identité se montre facilement par récurrence. □

## 5.2 Convergence des variables aléatoires réelles

On définit pour les variables aléatoires des notions de convergence analogues au convergences d'intégrales mais avec un vocabulaire légèrement différent de celui introduit à la Définition 35.

### Définition 46 : Convergence p.s., $\mathcal{L}^1$ , en loi ★

Une suite  $(X_n)_{n \in \mathbb{N}}$  de variables aléatoires réelles *converge presque-sûrement*, ou *p.s.*, vers  $X$ , noté  $X_n \xrightarrow{p.s.} X$ , si il existe  $A \in \mathcal{F}$  tel que  $\mathbb{P}(A) = 1$  et  $X_n(\omega)$  converge vers  $X(\omega)$  pour tout  $\omega \in A$ . autrement dit

$$X_n \xrightarrow{p.s.} X \iff \mathbb{P} \left( \omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right) = 1.$$

Une suite  $(X_n)_{n \in \mathbb{N}}$  de variables aléatoires *converge dans  $\mathcal{L}^1$*  vers une variable aléatoire  $X$ , noté  $X_n \xrightarrow{\mathcal{L}^1} X$  si les  $X_n$  et  $X$  sont intégrables et si

$$\mathbb{E}(|X_n - X|) \xrightarrow{n \rightarrow \infty} 0.$$

Une suite  $(X_n)_{n \in \mathbb{N}}$  de variables aléatoires *converge en probabilité* vers une variable aléatoire  $X$ , noté  $X_n \xrightarrow{\mathbb{P}} X$  si pour tout  $\varepsilon > 0$

$$\mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0.$$

Une suite  $(X_n)_{n \in \mathbb{N}}$  de variables aléatoires *converge en loi, ou en distribution*, vers une variable aléatoire  $X$ , noté  $X_n \xrightarrow{(d)} X$  si pour toute fonction  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  continue bornée,

$$\mathbb{E}(\varphi(X_n)) \xrightarrow{n \rightarrow \infty} \mathbb{E}(\varphi(X)).$$

**REMARQUE 23 :** La convergence en loi est différente des convergences en probas, p.s. et  $\mathcal{L}^1$ , car elle porte sur les mesures  $\mathbb{P}_{X_i}$  et  $\mathbb{P}_X$  et non pas directement sur les variables aléatoires  $X_i$  et  $X$ . En particulier, la convergence en loi ne requiert pas que les v.a.  $X_i$  et  $X$  soient définies sur le même espace de probabilité.

**NOTATION :** Dans le cadre d'une convergence en loi, on identifiera parfois une variable aléatoire et sa distribution. Dans le contexte du *théorème central limite*, par exemple, introduit ci-dessous, on écrira " $Y_n \xrightarrow{(d)} \mathcal{N}(0, 1)$ " plutôt que " $Y_n \xrightarrow{(d)} N$ ", où  $N$  est une variable aléatoire de loi normale centrée réduite  $\mathcal{N}(0, 1)$ ".

**REMARQUE 24 :** Ces convergences ne sont pas toutes aussi fortes les unes que les autres! En fait, on peut montrer que

$$\left[ X_n \xrightarrow{L^1 \text{ ou p.s.}} X \right] \implies \left[ X_n \xrightarrow{\mathbb{P}} X \right] \implies \left[ X_n \xrightarrow{(d)} X \right].$$

Attention toutefois, on n'a ni  $X_n \xrightarrow{p.s.} X \implies X_n \xrightarrow{L^1} X$ , ni l'implication inverse! Pour la première, il faut le théorème de convergence dominée (Théorème 16). Pour la seconde, il n'y a pas de théorème général, mais le contre exemple suivant montre que l'implication n'est pas vraie : on considère des Bernoullis indépendantes  $X_n$  de paramètre  $1/n$ . On calcule facilement que  $\mathbb{E}(|X_n|) = 1/n \rightarrow 0$ , et donc  $X_n \xrightarrow{L^1} 0$ . Pourtant, par le second lemme de Borel-Cantelli (hors-programme), une infinité de  $X_n$  valent 1 avec probabilité 1 puisque  $\sum_n 1/n = \infty$  et que les événements  $\{X_n = 1\}$  sont indépendants. Il en est de même pour les événements  $\{X_n = 0\}$ , on a donc en particulier, presque-sûrement, que la suite  $(X_n(\omega))_{n \in \mathbb{N}}$  ne converge pas!

### Proposition 27 : Caractérisation de la convergence en loi

Soit  $F_n = F_{X_n}$  la fonction de répartition de la v.a.  $X_n$  et  $F$  celle de  $X$ . Alors,

$$X_n \xrightarrow{(d)} X \iff (F_n(x) \rightarrow F(x), \forall x \text{ où } F \text{ est continue}).$$

En d'autres termes, pour des variables aléatoires réelles, la convergence en loi est équivalente à la convergence simple de la fonction de répartition.



### 5.3 Loi des grands nombres

Nous avons maintenant tous les outils nécessaires pour énoncer les deux principaux résultats de ce cours, la loi des grands nombres, et le théorème central limite. Pour cela, étant donné une suite i.i.d. de v.a.  $(X_i)_{i \in \mathbb{N}}$ , on définit sa *moyenne empirique*

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}. \quad (3)$$

#### Théorème 28 : Loi forte des grands nombres (LGN) ★

Soit  $(X_i)_{i \in \mathbb{N}}$  une famille i.i.d. de variables aléatoires, on suppose  $X_1$  *intégrable*. Alors,  $\bar{X}_n$  converge p.s. quand  $n \rightarrow \infty$ , et

$$\lim_{n \rightarrow \infty} \bar{X}_n = \mathbb{E}(X_1).$$



*ATTENTION : il est fondamental que la suite soit indépendante pour pouvoir appliquer la loi des grands nombres! (de même que le théorème central limite ci-dessous) Prenons par exemple  $X_1 \sim \text{Bernoulli}(1/2)$ , et  $X_i = X_1$  pour tout  $i \in \mathbb{N}$  (c'est à dire que pour tout  $\omega$ ,  $X_i(\omega) = X_1(\omega)$ ). Alors, la suite  $(X_i)$  est bien identiquement distribuée, puisque tous les  $X_i$  ont également pour loi une Bernoulli(1/2), mais elle n'est pas indépendante. On remarque facilement que  $\bar{X}_n = X_1$  p.s., et reste donc aléatoire, et donc  $\bar{X}_n$  ne converge pas p.s. vers  $\mathbb{E}(X_1) = 1/2$ .*

Malheureusement, la loi forte des grands nombres ne dit rien sur la vitesse à laquelle  $\bar{X}_n$  converge vers  $X$ . Et pour cause, la vitesse de cette convergence dépend des propriétés d'intégrabilité de  $X_1$ ! Par exemple, on peut montrer le résultat suivant.

#### Exercice 24 : vitesse de convergence de la LFGN

On suppose que  $(X_i)_{i \in \mathbb{N}}$  une famille i.i.d. de variables aléatoires, on suppose que  $X_1$  admet un moment d'ordre deux. Montrer que

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}(X_1)| \geq \varepsilon) \leq \frac{\text{Var}(X_1)}{\varepsilon^2 n}.$$

**SOLUTION :** On applique l'inégalité de Bienaymé-Chebychev :

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}(X_1)| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2},$$

il suffit donc maintenant de calculer la variance de  $\bar{X}_n$ . Comme l'espérance est linéaire,  $\text{Var}(aX) = a^2 \text{Var}(X)$ . Grâce à cette affirmation, et en utilisant le Corollaire 26, on obtient donc que

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) \stackrel{(X_k) \text{ i.i.d.}}{=} \frac{n}{n^2} \text{Var}(X_1) = \frac{\text{Var}(X_1)}{n}.$$

□

**REMARQUE 25 :** En réalité, si l'on suppose de plus fortes conditions d'intégrabilité sur  $X_1$ , on peut montrer des convergences beaucoup plus rapides vers la moyenne. Par exemple, si  $\mathbb{E}(e^{\lambda X_1}) < +\infty$  pour tout  $\lambda \in \mathbb{R}$ , le *théorème de Cramér* (complètement hors programme!) affirme qu'il existe une fonction  $\Lambda(\varepsilon)$  telle que

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}(X_1)| \geq \varepsilon) \leq e^{-\Lambda(\varepsilon)n},$$

qui est une estimation beaucoup plus forte que celle que nous avons obtenue dans l'exercice précédent!

## 5.4 Théorème central limite et statistiques descriptives

### Théorème 29 : Théorème central limite

★

On suppose que  $X_1$  admet un moment d'ordre deux (c'est-à-dire que sa variance est finie). On note  $\mu = \mathbb{E}(X)$  et  $\sigma^2 = \text{Var}(X)$ . Alors,

$$Y_n := \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{(d)} \mathcal{N}(0, 1),$$

où  $\mathcal{N}(0, 1)$  est une loi normale centrée réduite introduite dans la Définition 25.

Le théorème central limite que nous venons de voir est le pilier des statistiques descriptives, et a de nombreuses applications. Nous allons illustrer le principe général des intervalles de confiance statistiques avec l'exercice suivant.

————— Fin du cours 11 —————

On cherche à déterminer la proportion de personnes en France qui sont actuellement infectées par le Covid. Pour cela, on va tester un certain nombre de personnes, que l'on suppose choisies uniformément dans la population. Évidemment, comme on n'a accès qu'à une partie de la population, le taux d'infections estimé ne sera pas exactement égal au taux d'infections général de la population. On veut donc estimer la probabilité que le taux d'infection en France soit dans un intervalle donné, dit *intervalle de confiance*, à l'aide de tests de coronavirus dans un échantillon de taille  $n$ .

### Exercice 25

On associe à chaque individu  $1 \leq i \leq n$  de l'échantillon une variable de Bernoulli  $B_i$  de paramètre  $p$ , valant 1 si l'individu  $i$  est infecté, et 0 sinon. Construire à l'aide de l'échantillon observé un intervalle  $I_n$  tel que

$$\mathbb{P}(p \in I_n) \geq 0,95.$$

Attention, contrairement aux exemples que l'on a vu précédemment, la quan-

tité  $p$  est un *paramètre déterministe*, alors que l'intervalle  $I_n = [a_n, b_n]$  est *aléatoire*!

**SOLUTION** : On suppose que la population française est assez grande pour considérer qu'elle est infinie (si, si!). Une proportion  $p$  de la population est infectée par le coronavirus. Le paramètre  $p$  est inconnu, notre objectif de l'estimer. On va donc tester un échantillon de la population, composé de  $n$  individus. On définit la proportion *observée* d'infections dans l'échantillon  $p_n = \frac{1}{n} \sum_{i=1}^n B_i$ .

Par le théorème central limite,

$$\mathbb{P}\left(\frac{\sqrt{n}|p_n - p|}{\sigma_p} \leq \varepsilon\right) \underset{n \rightarrow \infty}{\simeq} \mathbb{P}(|\mathcal{N}(0, 1)| \leq \varepsilon),$$

où  $\sigma_p = \sqrt{p(1-p)}$  est l'écart-type de la loi de Bernoulli. Typiquement, on cherche à construire un intervalle de confiance à 95%, c'est à dire un intervalle dans lequel le paramètre  $p$  à une probabilité de 0.95 de se trouver. On va donc choisir  $\varepsilon$  aussi petit que possible tel que  $\mathbb{P}(|\mathcal{N}(0, 1)| \leq \varepsilon) \geq 0.95$ . En consultant des tables de lois normales, on obtient

$$\varepsilon = 1.96.$$

On déduit donc du théorème central limite

$$\mathbb{P}\left(p_n - \frac{1.96 \sigma_p}{\sqrt{n}} \leq p \leq p_n + \frac{1.96 \sigma_p}{\sqrt{n}}\right) \geq 0.95. \quad (4)$$

Malheureusement, dans les bornes ci-dessus, l'écart-type  $\sigma_p$  dépend encore de  $p$ . On va donc également devoir l'estimer, en écrivant à l'aide de la loi des grands nombres

$$s_n = \frac{1}{n} \sum_{i=1}^n B_i^2 - \left(\frac{1}{n} \sum_{i=1}^n B_i\right)^2 = p_n(1 - p_n) \xrightarrow{n \rightarrow \infty} p(1 - p).$$

On va donc remplacer dans (4) l'écart type  $\sigma_p$  par l'estimateur de l'écart type  $\sqrt{s_n}$ . On définit donc

$$I_n = \left[ p_n - \frac{1.96 \sqrt{s_n}}{\sqrt{n}}, p_n + \frac{1.96 \sqrt{s_n}}{\sqrt{n}} \right].$$

À noter que  $s_n$  est construit uniquement à partir des observations  $(B_k)$ , et non pas des paramètres théoriques du problème. Par ailleurs, on a comme on le voulait

$$\mathbb{P}(p \in I_n) \geq 0.95.$$

Quelques remarques sur ce raisonnement.

- Dans la dernière identité, on a triché en remplaçant  $\sigma_p$  par son estimateur  $\sqrt{s_n}$ . Ce faisant, on fait, d'après le théorème central limite appliqué à la suite  $(B_k^2)_{k \in \mathbb{N}}$ , une erreur de l'ordre de  $1/\sqrt{(n)}$  sur  $\sigma_p$ . Ce n'est pas un problème en réalité, puisque cela entraîne une correction d'ordre  $1/\sqrt{n^2} = 1/n$  dans l'intervalle  $I_n$ , qui est de taille  $O(1/\sqrt{n})$ .

- Cet intervalle de confiance permet donc de déterminer la taille d'échantillon nécessaire pour avoir un niveau de précision donné. Imaginons par exemple que le gouvernement souhaite pouvoir estimer, avec un niveau de confiance de 95%, et avec une précision donnée de  $\delta$ , le taux d'infection dans la population générale. Il suffit pour cela de choisir  $n$  assez grand pour que  $1.96 \sqrt{s_n} / \sqrt{n} \leq \delta$ .
- Tout ce raisonnement repose, comme la loi des grands nombres et le théorème central limite, sur l'indépendance des échantillons  $B_k$ . Pour cette raison, il est important, pour que les résultats puissent être pertinents, de tester des personnes qui viennent d'environnements aussi divers et distants que possibles! Imaginez par exemple que tous les tests viennent de patients présentant des symptômes du coronavirus. On surestimerait alors énormément la prévalence  $p$  du coronavirus dans la population!

□

On cherche maintenant à estimer si, parmi les patients hospitalisés du covid, la létalité du coronavirus dépend de l'âge du patient.

### Exercice 26

On a accès à deux échantillons de patients ayant été hospitalisés, chaque échantillon contenant 100 patients supposés indépendants. Dans le premier groupe, constitué de patients de plus de 50 ans, 50 patients sont décédés. Dans le second groupe, constitué de patients de 15 à 49 ans, 25 personnes sont décédées du covid.

On pose deux hypothèses de travail. Sous

$H_0$  : le taux de décès du covid est indépendant de l'âge,

le taux théorique de décès suite à l'hospitalisation d'un patient est donné par un paramètre  $p$ , commun à tous les patients indépendamment de leur âge. Sous la seconde hypothèse,

$H_1$  : le taux de décès du covid dépend de l'âge,

les deux groupes de patients ont deux taux théoriques de décès  $p_1$  et  $p_2$ , respectivement  $p_1$  pour les patients de plus de 50 ans et  $p_2$  pour les 15 – 49 ans. Avec quel degré de confiance peut-on justifier l'hypothèse  $H_0$ ?

*On admettra que la somme de deux gaussiennes indépendantes suit également une distribution gaussienne.*

**SOLUTION** : Posons  $n = 100$ , on va appliquer le même type de raisonnement que dans l'exercice précédent. Supposons que  $H_0$  est vraie. Alors, en définissant  $p_n$  la proportion observée de décès dans le premier échantillon et  $q_n$  celle dans le second échantillon, on a

$$p_n = 0.5 \quad \text{et} \quad q_n = 0.25.$$

D'autre part, par le théorème central limite,

$$X := \frac{\sqrt{n}(p_n - p)}{\sqrt{p(1-p)}} \simeq \mathcal{N}(0, 1) \quad \text{et} \quad Y := \frac{\sqrt{n}(q_n - p)}{\sqrt{p(1-p)}} \simeq \mathcal{N}(0, 1).$$

On va maintenant supposer que ces deux approximations sont en fait des égalités, et que  $X, Y \sim \mathcal{N}(0, 1)$ . Par symétrie de la distribution gaussienne,  $-Y \sim \mathcal{N}(0, 1)$ . Par ailleurs, comme  $X$  et  $Y$  reposent sur des individus tous indépendants entre eux, on a  $X \perp -Y$ , et donc  $X - Y \sim \mathcal{N}(m, \sigma^2)$ , puisque une somme de gaussiennes est encore une gaussienne. Or  $m = \mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = 0$ . et pour des variables indépendantes

$$\sigma^2 = \text{Var}(X - Y) = \text{Var}(X) + \text{Var}(-Y) = \text{Var}(X) + \text{Var}(Y) = 2,$$

et donc  $(X - Y)/\sqrt{2} \sim \mathcal{N}(0, 1)$ . On reprend maintenant les résultats numériques de l'expérience, et on obtient

$$\frac{X - Y}{\sqrt{2}} = \frac{\sqrt{n}(p_n - q_n)}{\sqrt{p(1-p)}}.$$

Comme on ne connaît pas  $p$ , on va à nouveau estimer l'écart type  $\sqrt{p(1-p)}$  grâce à l'écart type empirique dans l'échantillon total,

$$s_n = \sqrt{\frac{p_n + q_n}{2} \left(1 - \frac{p_n + q_n}{2}\right)} = \sqrt{15}/8,$$

et on obtient la valeur numérique  $(X - Y)/\sqrt{2} \simeq 40/\sqrt{15} \simeq 10,3$ .

Sous l'hypothèse  $H_0$ , notre variable observée  $(X - Y)/\sqrt{2} \sim \mathcal{N}(0, 1)$  prend une valeur supérieure à 10. En réalité, cette probabilité est infinitésimale, puisque on a déjà que

$$\mathbb{P}(\mathcal{N}(0, 1) > 5) \simeq 3 \times 10^{-7}.$$

On peut donc rejeter  $H_0$ , et conclure que le taux de décès dû au covid dépend bien de l'âge du patient. □

## 6 Bonus : vecteurs aléatoires et suites de variables aléatoires

### 6.1 Espaces mesurés produits et Théorème de Fubini

On revient très brièvement sur des éléments de théorie de la mesure pour pouvoir introduire les variables aléatoires sur  $\mathbb{R}^N$ .

#### Definition 47 : Tribu produit et boréliens sur $\mathbb{R}^N$

Soient  $(E_1, \mathcal{T}_1)$  et  $(E_2, \mathcal{T}_2)$  deux espaces mesurables, On définit la tribu produit  $\mathcal{T} = \mathcal{T}_1 \otimes \mathcal{T}_2$  sur  $E_1 \times E_1$  comme la tribu engendrée par les produits d'ensembles mesurables,

$$\mathcal{T} = \sigma(\{A \times B, A \in \mathcal{T}_1, B \in \mathcal{T}_2\}).$$

En particulier, la tribu borélienne  $\mathcal{B}(\mathbb{R}^N)$  sur  $\mathbb{R}^N$  est définie comme la tribu produit de la tribu borélienne sur  $\mathbb{R}$ , mais aussi comme la tribu engendrée par les hypercubes de  $\mathbb{R}^N$

$$\mathcal{B}(\mathbb{R}^N) = \underbrace{\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R})}_{N \text{ fois}} = \sigma\left(\left\{\prod_{i=1}^N ]a_i, b_i[, -\infty \leq a_i \leq b_i \leq +\infty \forall i \leq N\right\}\right).$$

L'introduction de la tribu produit permet maintenant d'introduire la notion de mesure produit, sous la forme d'une proposition que l'on admettra.

#### Proposition 30 : Mesure produit

Soient  $(E_1, \mathcal{T}_1, \mu_1)$  et  $(E_2, \mathcal{T}_2, \mu_2)$  deux espaces mesurés  $\sigma$ -finis. Il existe une unique mesure, notée  $\mu_1 \otimes \mu_2$  et appelée *mesure produit de  $\mu_1$  et  $\mu_2$*  sur l'espace mesurable  $(E_1 \times E_2, \mathcal{T}_1 \otimes \mathcal{T}_2)$ , et satisfaisant pour tous  $A \in \mathcal{T}_1, B \in \mathcal{T}_2$  de mesure finie

$$\mu \otimes \nu(A \times B) = \mu(A)\nu(B).$$

**REMARQUE 26 :** On montre facilement que  $\mu \otimes \nu$  est également  $\sigma$ -finie.

**REMARQUE 27 :** Cette proposition permet en particulier de définir la mesure de Lebesgue  $\lambda_N$  sur  $\mathbb{R}^N$ , comme la mesure produit ( $N$  fois) de la mesure de Lebesgue sur  $\mathbb{R}$ . C'est en particulier l'unique mesure sur  $\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N)$ , satisfaisant

$$\lambda_N\left(\prod_{i=1}^N ]a_i, b_i[ \right) = \prod_{i=1}^N (b_i - a_i),$$

c'est à dire que la mesure de Lebesgue d'un hypercube est son volume.

On en arrive maintenant à un résultat fondamental, que nous avons déjà abordé en exercices : le théorème de Fubini.

### Théorème 31 : Théorème de Fubini Tonelli

Soient  $(E_1, \mathcal{T}_1, \mu_1)$  et  $(E_2, \mathcal{T}_2, \mu_2)$  deux espaces mesurés  $\sigma$ -finis, on note

$$(E, \mathcal{T}, \mu) := (E_1 \times E_2, \mathcal{T}_1 \otimes \mathcal{T}_2, \mu_1 \otimes \mu_2)$$

leur espace mesuré produit. Soit  $f : E \rightarrow \mathbb{R}_+ \cup \{+\infty\}$  une fonction mesurable **positive** sur l'espace produit. Alors,

- i) Pour tout  $x_1 \in E_1$ , la fonction  $f(x_1, \cdot)$  est mesurable positive sur  $(E_2, \mathcal{T}_2)$ , et la fonction  $\varphi_f : E_1 \rightarrow \mathbb{R}_+ \cup \{+\infty\}$  définie par

$$\varphi_f(x_1) = \int_{E_2} f(x_1, x_2) d\mu_2(x_2)$$

est mesurable positive sur  $(E_1, \mathcal{T}_1)$ .

- ii) On a par ailleurs l'identité

$$\int f(x_1, x_2) d\mu(x_1, x_2) = \int_{E_1} \varphi_f(x_1) d\mu_1(x_1) = \int_{E_1} \left[ \int_{E_2} f(x_1, x_2) d\mu_2(x_2) \right] d\mu_1(x_1).$$

Le même résultat est vrai en inversant les rôles de  $E_1$  et  $E_2$ , et donc en particulier

$$\int_{E_1} \left[ \int_{E_2} f(x_1, x_2) d\mu_2(x_2) \right] d\mu_1(x_1) = \int_{E_2} \left[ \int_{E_1} f(x_1, x_2) d\mu_1(x_1) \right] d\mu_2(x_2).$$

En particulier, une fonction  $f$  est intégrable contre la mesure produit  $\mu$  si et seulement si

$$\int_{E_1} \left[ \int_{E_2} |f(x_1, x_2)| d\mu_2(x_2) \right] d\mu_1(x_1) < \infty$$

ou

$$\int_{E_2} \left[ \int_{E_1} |f(x_1, x_2)| d\mu_1(x_1) \right] d\mu_2(x_2) < \infty.$$

Si l'une de ces identités est vraie, alors l'application  $f(x_1, \cdot)$  (resp.  $f(\cdot, x_2)$ ) est intégrable sur  $E_2$  pour  $\mu_1$ -presque-tout  $x_1$ , resp. sur  $E_1$   $\mu_2$ -presque-tout  $x_2$ .

Par récurrence, toutes ces notions (tribus produit, mesures produit, intégrales multiples) sont généralisables à un nombre fini d'espaces/ de variables. La définition des espaces produits permet d'introduire la notion de convolution de mesure, très utile en pratique.

### Definition 48 : Convolution de mesures sur $\mathbb{R}$

Soient  $\mu, \nu$  deux mesures sur  $\mathcal{B}(\mathbb{R})$ . On définit la mesure de convolution  $\mu * \nu$  sur  $\mathcal{B}(\mathbb{R})$  par

$$\mu * \nu(B) = \int_{\mathbb{R} \times \mathbb{R}} \mathbf{1}_{\{x+y \in B\}} d\mu \otimes \nu(x, y)$$

En particulier, on admettra que si  $\mu(dx) = f(x)dx$  et  $\nu(dx) = g(x)dx$  sont toutes les deux à densités par rapport à la mesure de Lebesgue, la convolution  $\mu * \nu$  est également à densité par rapport à la mesure de Lebesgue, et sa densité est donnée par le produit de convolution de  $f$  et  $g$ ,

$$\mu * \nu(B) = \int_B f * g(x)dx, \quad \text{avec} \quad f * g(x) := \int_{\mathbb{R}} f(t)g(x-t)dt.$$

Le produit de convolution permet de caractériser la loi d'une somme de variables aléatoires indépendantes.

### Proposition 32 : Somme de variables aléatoires

Soient  $X, Y$  deux variables aléatoires réelles indépendantes, de lois respectives  $\mathbb{P}_X$  et  $\mathbb{P}_Y$ , alors la loi  $\mathbb{P}_{X+Y}$  de  $X + Y$  est donnée par  $\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y$ . En particulier, si  $X$  et  $Y$  sont toutes deux à densité, de densités respectives  $f$  et  $g$ , alors la variable aléatoire  $X + Y$  est également à densité, et sa densité est donnée par

$$f * g(x) := \int_{\mathbb{R}} f(t)g(x-t)dt.$$

A noter que l'on n'a pas montré a priori que  $f * g$  est bien définie, parce qu'un produit de fonctions intégrables n'est pas nécessairement intégrable. Nous passons sous silence ce point technique, qui nous forcerait à aborder trop de nouvelles difficultés.

## 6.2 Vecteurs aléatoires

Dans toute cette section, on fixe  $N \in \mathbb{N}$  et  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace de probabilités. on notera  $\underline{x} = (x_1, \dots, x_N)$  les éléments de  $\mathbb{R}^N$ .

### Definition 49 : Vecteur aléatoire

Un vecteur aléatoire (de dimension  $N$ , à valeurs réelles)  $X$  est une variable aléatoire à valeurs dans  $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$ .

La loi d'un vecteur aléatoire  $X$  est la mesure  $\mathbb{P}_X$  sur  $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$  définie par

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A), \quad \forall A \in \mathcal{B}(\mathbb{R}^N).$$

Un vecteur aléatoire est dit à densité  $f_X$  (par rapport à la mesure de Lebesgue sur  $\mathbb{R}^N$ ) s'il existe une fonction positive mesurable :  $\mathbb{R}^N \rightarrow \mathbb{R}_+$  telle que  $\int_{\mathbb{R}^N} f(\underline{x})d\underline{x} = 1$  et satisfaisant

$$\mathbb{P}(X \in A) = \int_A f(\underline{x})d\underline{x}.$$



**NOTATION** : Pour tout vecteur  $\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^N$ , on note, pour tout  $n \leq N$ ,  $\pi_n(\underline{x}) = x_n$  le projecteur sur la  $n$ -ème coordonnée de  $\underline{x}$ .

### Definition 50 : Probabilité marginale

Soit  $\mu$  une mesure de probabilités sur  $\mathbb{R}^N$ , on appelle  $n$ -ème probabilité marginale de  $\mu$  la mesure sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

$$\mu_n(B) = \{\mu(\pi_n^{-1}(B))\}, \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

Autrement dit, on en particulier  $\mu_1(B) = \mu(B \times \mathbb{R}^{N-1})$ .

Soit  $X$  un vecteur aléatoire dans  $\mathbb{R}^N$ , on appelle  $n$ -ème loi marginale de  $X$  la  $n$ -ème loi marginale de  $\mathbb{P}_X$ . On admettra que la  $n$ -ème loi marginale de  $X$  est la loi de  $X_n$ .

### Exercice 27 : Variables à densité indépendantes

Soit  $X = (X_1, X_2)$  un vecteur aléatoire, on suppose que les deux lois marginales de  $X$  sont à densités  $f_1$  et  $f_2$ . Montrer que  $X_1$  et  $X_2$  sont indépendantes si et seulement si  $X$  est un vecteur aléatoire de densité  $f_X(x_1, x_2) = f_1(x_1)f_2(x_2)$ .

**SOLUTION** : On rappelle que la tribu des boréliens sur  $\mathbb{R}^2$  est engendrée par les rectangles  $]a, b[ \times ]c, d[$ , avec  $a \leq b, c \leq d$ , qui forment un  $\pi$ -système. Pour caractériser une mesure sur  $\mathbb{R}^2$ , d'après le Lemme des classes monotones (cf. Lemme 17), il suffit donc de la caractériser sur les rectangles. On peut alors écrire

$X_1$  et  $X_2$  sont indépendantes.

$$\Leftrightarrow \mathbb{P}(X \in ]a, b[ \times ]c, d[) = \mathbb{P}(X_1 \in ]a, b[) \mathbb{P}(X_2 \in ]c, d[) \quad \forall a \leq b, c \leq d \in \mathbb{R}$$

$$\Leftrightarrow \mathbb{P}(X \in ]a, b[ \times ]c, d[) = \int_a^b \int_c^d f_1(x_1) f_2(x_2) dx_2 dx_1 \quad \forall a \leq b, c \leq d \in \mathbb{R}$$

$$\Leftrightarrow X \text{ est à densité, donnée par } f_X(x_1, x_2) := f_1(x_1) f_2(x_2).$$

□

### Definition 51 : Espérance d'un vecteur aléatoire

Soit  $X$  un vecteur aléatoire, on suppose que toutes les lois marginales de  $X$  sont intégrables. On définit l'espérance de  $X$  comme le vecteur

$$\mathbb{E}(X) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_N)),$$

qui est bien défini puisque chaque coordonnée l'est.

Comme les coordonnées d'un vecteur aléatoires ne sont pas nécessairement indépendantes, la notion de Variance, pour les variables réelles, est à remplacer pour les vecteurs aléatoires par la notion de matrice de covariance.

### Definition 52 : Matrice de covariance

Soit  $X$  un vecteur aléatoire, on suppose que tout  $n \leq N$ ,  $X_n$  est de carré intégrable. Alors, pour tout  $n, m \leq N$ ,  $X_n X_m$  est également intégrable, et on définit la matrice de covariance  $Cov(X)$  des  $(X_n)_{n \leq N}$  comme la matrice  $N \times N$  dont les entrées sont données par

$$Cov(X)_{n,m} = Cov(X_n, X_m) = \mathbb{E}[(X_n - \mathbb{E}(X_n))(X_m - \mathbb{E}(X_m))].$$

**REMARQUE 28 :** Si les coordonnées d'un vecteur aléatoire sont indépendantes deux à deux, sa matrice de covariance est diagonale.

Un exemple fondamental de vecteur aléatoire est la loi normale multidimensionnelle, aussi appelée vecteur gaussien.

### Definition 53 : Vecteur gaussien

Soit  $\Sigma$  une matrice  $N \times N$  symétrique définie positive,  $M \in \mathbb{R}^N$  un vecteur réel. Un vecteur aléatoire  $X$  est dit *gaussien* de matrice de covariance  $\Sigma$  et de moyenne  $M$  si  $X$  a pour densité

$$f(\underline{x}) = \frac{1}{\sqrt{(2\pi)^N Det(\Sigma)}} \exp\left(-(\underline{x} - M)' \Sigma^{-1} (\underline{x} - M)/2\right),$$

où  $\underline{y}'$  désigne le vecteur ligne transposé de  $\underline{y}$ .

### Exercice 28

Soient  $X_1, \dots, X_N$  des variables aléatoires gaussiennes réelles, indépendantes, de moyennes et variances respectives  $m_n, \sigma_n^2$ , pour  $1 \leq n \leq N$ . Quelle est la loi du vecteur aléatoire  $X = (X_1, \dots, X_N)$ ?

**SOLUTION :** On vérifie facilement qu'alors, le vecteur  $(X_1, \dots, X_N)$  est un vecteur gaussien (voir Définition 53) de moyenne  $M = (m_1, \dots, m_N)$  et de matrice de covariance  $\Sigma = \text{Diag}(\sigma_1^2, \dots, \sigma_N^2)$ .  $\square$



*ATTENTION : Cet exercice montre qu'un vecteur aléatoire dont toutes les lois marginales sont des lois normales indépendantes est un vecteur gaussien. Ce n'est plus vrai si les coordonnées ne sont pas indépendantes, toutefois. Par exemple, soit  $X$  une loi normale centrée réduite, le vecteur  $(X, X, \dots, X)$  prend ses valeurs dans l'ensemble  $\{\underline{x} = (x_1, \dots, x_N) \in \mathbb{R}^N, x_1 = x_2 = \dots = x_N\}$ , qui est de mesure de Lebesgue dans  $\mathbb{R}^N$  nulle, et ne peut donc pas être à densité. La réciproque est également fautive : les lois marginales d'un vecteur gaussien sont bien des lois gaussiennes, par contre elles ne sont pas nécessairement indépendantes.*