Intégration et probabilités

Paul Fournier et Constantin Vaillant-Tenzer, d'après le cours de Patrick Bernard

Premier semestre 2020

Table des matières

Chapitre 1

Préliminaires

1.1 Introduction

Références :
- Billingsley, *Probabiliy and measure*
- Kolmogorov & Fomin, tome 2

Motivations :
- Définir la longueur d'une partie de $\mathbb{R}$
- Définir l'aire d'une partie de $\mathbb{R}^2$
- Définir $\displaystyle\int f \mathrm{d}x$ pour $f : \mathbb{R}^d \to \mathbb{R}$
- Définir, préciser la notion mathématique décrivant une suite infinie de jets de dés

Par exemple :
- Si $f : \mathbb{R} \to \mathbb{R}$, on peut définir $\displaystyle\int f$ comme l'aire algébrique définie par le graphe de f. Ainsi, définir une aire permet de définir une intégrale
- De même, $\lambda(A) = \mathbb{1}_A$ avec $\mathbb{1}_A(x) = 1$ ssi $x \in A$. Donc définir une intégrale revient à définir une mesure.
- Tirer un nombre au hasard dans $[0,1]$, cela revient à tirer au hasard la suite de ses décimales au D10, car on mesure une partie de $\{0,1,\dots 9\}^{\mathbb{N}}$

On se demande alors comment définir la surface d'une partie du plan.

Méthode 1 : à la Riemann. On approxime avec un quadrillage. On compte le nombre de carrés qui intersectent l'ensemble considéré, puis on conclut en passant à la limite quand le côté du quadrillage tend vers 0.

Méthode 2 : on pose $\lambda(A) := \inf\limits_{(R_i)} \displaystyle\sum_{i=1}^{\infty} \lambda(R_i)$ où R_i est une suite de rectangles recouvrant A.

À noter : les deux méthodes ont des cas pathologiques différents.

1.2 Ensembles dénombrables

Définition : Un ensemble est dénombrable ssi il est en bijection avec $\mathbb{N}$

Propriété : Toute partie d'un ensemble dénombrable est au plus dénombrable

Démonstration : On pose $x : \mathbb{N} \to X, Y \subset X$. Si Y n'est pas fini :

$$\begin{aligned} &i_1 = \min\{i \in \mathbb{N}, x_i \in Y\} \\ &\dots \\ &i_n = \min\{i \in \mathbb{N}, x_i \in Y \setminus \{x_1, \dots, x_{n-1}\}\} \end{aligned}$$

Ainsi, $k \mapsto x_{n_k}$ est une bijection de $\mathbb{N}$ vers Y.

□

Propriété : L'image d'une suite est au plus dénombrable.

Démonstration : On note $x : \mathbb{N} \to X$ une suite. On crée de manière analogue une sous-suite injective de x de même image que x (sauf si $f(x(\mathbb{N}))$ est fini).

□

Propriété : $\mathbb{N} \times \mathbb{N}$ est dénombrable.

Démonstration : $(n_1, n_2) \mapsto 2^{n_1}(2n_2+1) - 1$ est une bijection $\mathbb{N}^2 \to \mathbb{N}$.

$\square$

Propriété : Une réunion au plus dénombrables d'ensembles au plus dénombrable est au plus dénombrable.

Démonstration : On traite le cas "union dénombrable d'ensembles dénombrables".
Soit A_i des parties dénombrables d'un ensemble X. Pour tout i, il existe $b_i : \mathbb{N} \to A_i$ bijection. (nb : ceci requiert en fait l'axiome du choix dénombrable) Alors $\begin{array}{c}(i,j) \mapsto b_i(j) \\ \mathbb{N}^2 \to \bigcup\limits_i A_i\end{array}$ est surjective.

Donc $\bigcup\limits_i A_i$ est au plus dénombrable.

Or $\bigcup\limits_i A_i \supset A_i$.

Donc $\bigcup\limits_i A_i$ est dénombrable.

$\square$

Propriété : Si X est dénombrable, $\mathcal{P}(X)$ ne l'est pas.
Plus généralement, quel que soit X, X et $\mathcal{P}(X)$ ne sont jamais en bijection (théorème de Cantor).

Démonstration : Supposons qu'il existe $x : \begin{array}{c} X \to \mathcal{P}(X) \\ x \mapsto A_x \end{array}$ une bijection.
Considérons $B := \{x, x \notin A_x\}$. Comme x est une bijection, il existe $y \in X$ tel que $B = A_y$.
Question : a-t-on $y \in B$. On arrive à un paradoxe type Russel.

$\square$

Exercice :
— $\{0,1\}^{\mathbb{N}}$ est non dénombrable.
— $\mathbb{R}$ est non dénombrable.

1.3 $\limsup$ et $\liminf$

Définition :
Soit $(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ (plus généralement $\in \bar{\mathbb{R}}^{\mathbb{N}}$). Alors $s_n := \sup\limits_{k\geq n} x_k$.
s_n est décroissante (donc a une limite dans $\bar{\mathbb{R}}$).
Alors $\lim s_n =: \limsup x_n = \inf s_n$.
De même pour $\liminf x_n$.

Propriété : $\lim x_n$ existe ssi $\liminf x_n = \limsup x_n$. Dans ce cas, $\lim x_n = \limsup x_n = \liminf x_n$.

Démonstration : $\Leftarrow$: $i_n \leq x_n \leq s_n$. On conclut par théorème d'encadrement.
$\Rightarrow$: Si $x_n \to l$ alors : $\forall \varepsilon > 0, \exists N \in \mathbb{N}, \forall n \geq N, l - \varepsilon \leq i_n \leq l \leq s_n \leq l + \varepsilon$. Donc $s_n \to l$ et $i_n \to l$.

$\square$

Propriété : Si y_n est une sous-suite de x_n, alors $\liminf x_n \leq \liminf y_n \leq \limsup y_n \leq \limsup x_n$

Ainsi, si l est valeur d'adhérence de x_n, alors $\liminf x_n \leq l \leq \limsup x_n$.

Propriété : $\limsup x_n = -\liminf(-x_n)$

Propriété : Il existe une sous-suite de x_n qui converge vers $\limsup x_n$. Idem pour $\liminf x_n$.

Démonstration : On choisit $k_n \geq n$ tel que $s_n - \dfrac{1}{n} \leq x_{k_n} \leq s_n$. $n \mapsto x_{k_n}$ converge vers $\limsup x_n$.

$\square$

1.4 Familles sommables

On pose $(a_i)_{i\in I}$ famille de nombres positifs.
Définition : $\sum_{i\in I} a_i := \sup_{F\subset I \text{fini}} \sum_{i\in F} a_i$

Propriété : Si $\sum_{i\in I} a_i$ est fini, alors $\{i \in I, a_i \neq 0\}$ est au plus dénombrable.

Démonstration : $\{i \in I, a_i \in \mathbb{R}\setminus\{0\}\} \subset \bigcup_{k\in\mathbb{N}} \underbrace{\{i \in I, a_i \geq \frac{1}{k}\}}_{\#\leq k\sum_{i\in I} a_i}$

□

À partir de maintenant, on considérera I dénombrable.

Propriété : Si $\sigma : \mathbb{N} \to I$ est une bijection, alors $\sum_{i\in I} a_i = \lim_{n\to+\infty} \sum_{k=1}^{n} a_{\sigma(k)} =: \sum_{k=1}^{+\infty} a_{\sigma(k)}$

Démonstration : $\forall F \subset I$ fini, $\sigma^{-1}(F)$ est fini donc majoré par un entier N.
$\sum_{i\in F} a_i = \sum_{k\in\sigma^{-1}(F)} a_{\sigma(k)} \leq \sum_{k=1}^{N} a_{\sigma(k)} \leq \sum_{k=1}^{+\infty} a_{\sigma(k)}$

Donc par passage au sup : $\sum_{i\in I} a_i \leq \sum_{k=1}^{+\infty} a_{\sigma(k)}$.

Réciproquement, $\sum_{k=1}^{N} a_{\sigma(k)} = \sum_{i\in\sigma([\![1,N]\!])} a_i \leq \sum_{i\in I} a_i$. On conclut par passage à la limite.

□

Corollaire : Si $(a_k) \in \mathbb{R}_+^{\mathbb{N}}$, $\sum_{k=1}^{+\infty} a_k = \sum_{k=1}^{+\infty} a_{\sigma(k)}$ et ce quel que soit $\sigma : \mathbb{N} \to \mathbb{N}$ bijection.

En particulier dans le cas $I = \mathbb{N}^2$, $(a_{i,j})_{(i,j)\in I} \in \mathbb{R}_+^I$:

Propriété : $\sum_{(i,j)\in I} a_{i,j} = \sum_{i=1}^{+\infty}\left(\sum_{j=1}^{+\infty} a_{i,j}\right) = \sum_{j=1}^{+\infty}\left(\sum_{i=1}^{+\infty} a_{i,j}\right)$

Démonstration : $F \subset I$ fini. Il existe $N \in \mathbb{N}$ tel que $F \subset [\![1,N]\!]^2$. Donc $\sum_{(i,j)\in F} a_{i,j} \leq \sum_{i=1}^{N}\sum_{j=1}^{N} a_{i,j} \leq \sum_{i=1}^{N}\sum_{j=1}^{+\infty} a_{i,j} \leq \sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty} a_{i,j}$.

Réciproquement, $\forall N \in \mathbb{N}, \forall M \in \mathbb{N}$, $\sum_{i=1}^{N}\sum_{j=1}^{M} a_{i,j} \leq \sum_{(i,j)\in\mathbb{N}^2} a_{i,j}$.

Donc $(M \to +\infty)$, $\sum_{i=1}^{N}\sum_{j=1}^{+\infty} a_{i,j} \leq \sum_{(i,j)\in\mathbb{N}^2} a_{i,j}$.

Donc $(N \to +\infty)$, $\sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty} a_{i,j} \leq \sum_{(i,j)\in\mathbb{N}^2} a_{i,j}$.

□

1.5 Séries absolument convergentes

Soit $(a_i)_{i\in I}$ une famille de réels tels que $\sum\limits_{i\in I} |a_i|$ soit finie.
On définit $a_i^+ := \max(a_i, 0)$, $a_i^- := \max(-a_i, 0)$.
Donc $a_i^+ - a_i^- = a_i$ et $a_i^+ + a_i^- = |a_i|$.

Propriété : $\sum\limits_{i\in I} a_i^+ - \sum\limits_{i\in I} a_i^- = \sum\limits_{k=1}^{+\infty} a_{\sigma(k)}$ et ce quel que soit $\sigma : \mathbb{N} \to I$ bijection.

Démonstration : $\sum\limits_{i\in I} a_i^+ \leq \sum\limits_{i\in I} |a_i|$ donc la somme est finie. Idem pour $\sum\limits_{i\in I} a_i^-$.

$\sum\limits_{k=1}^{n} a_{\sigma(k)} = \sum\limits_{k=1}^{n} a_{\sigma(k)}^+ - \sum\limits_{k=1}^{n} a_{\sigma(k)}^- \underset{n\to+\infty}{\to} \sum\limits_{k=1}^{+\infty} a_{\sigma(k)}^+ - \sum\limits_{k=1}^{+\infty} a_{\sigma(k)}^-$

□

Corollaire : Sous réserve de convergence absolue, on a :

$$\sum_{k=1}^{+\infty} a_k = \sum_{k=1}^{+\infty} a_{\sigma(k)}$$

$$\sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty} a_{i,j} = \sum_{j=1}^{+\infty}\sum_{i=1}^{+\infty} a_{i,j}$$

Chapitre 2

Probabilités

2.1 Vocabulaire

Définition : Soit X un ensemble. On dit que $\mathcal{A} \subset \mathcal{P}(X)$ est :
- une algèbre (d'ensembles) si elle est stable par union finie, intersection finie et passage au complémentaire, contient $\emptyset$ et X.
- une tribu (ou σ-algèbre) si c'est une algèbre stable par réunion/intersection dénombrable.

Exemple :
- $\mathcal{P}(X)$ est une tribu.
- $\{\emptyset, X\}$ est une tribu.

Si on se donne une partition finie de X : $X = X_1 \sqcup X_2 \cdots \sqcup X_k$, alors l'ensemble des $A \subset X$ de la forme $A = \bigcup_{n \in I \subset [\![1,k]\!]} X_n$ est une tribu finie.

> **Lemme :** Toute algèbre finie est associée à une partition finie.

Démonstration : Soit $\mathcal{A}$ une algèbre finie.
$\forall x \in X, A(x) := \bigcap_{\substack{A \in \mathcal{A} \\ x \in A}} A.$
Pour x et y donnés, soit $A(x) = A(y)$, soit $A(x) \cap A(y) = \emptyset$.
Fixons $x \in X, B \in \mathcal{A}$.
- Soit $x \in B$ et alors $A(x) \subset B$.
- Soit $x \in B^{\complement}$ et alors $A(x) \subset B^{\complement}$ i.e. $A(x) \cap B = \emptyset$

On conclut avec $B = A(y)$.

□

Définition : Si $\mathcal{A}$ est une algèbre de X et $m : \mathcal{A} \to [0, +\infty]$ une fonction.
On dit que m est une *mesure additive* si :
- $m(\emptyset) = 0$
- $m(A \sqcup B) = m(A) + m(B) \qquad (A \cap B = \emptyset)$

Définition : Si $\mathcal{T} \subset \mathcal{P}(X)$ est une tribu, $m : \mathcal{T} \to [0, +\infty]$ est une *mesure* si :
- $m(\emptyset) = 0$
- $m(\bigsqcup_{i \in I} A_i) = \sum_{i \in I} m(A_i)$ pour $(A_i)_{i \in I}$ famille dénombrable disjointe.

Remarque : Toute mesure est une mesure additive.
Remarque : On appelle parfois les mesures "mesures σ-additives".
Remarque : Lorsque $m : \mathcal{A} \to [0, +\infty]$ est une mesure additive sur une algèbre, les propriétés suivantes sont équivalentes :

1. Si $A_i \in \mathcal{A}$ sont disjoints, (A_i) dénombrable, $\bigsqcup_{i \in I} A_i \in \mathcal{A}$, alors $m(\bigsqcup_{i \in I} A_i) = \sum_{i \in I} m(A_i)$
2. Si $A, A_i \in \mathcal{A}$, $A \subset \bigcup_{i \in I} A_i$, alors $m(A) \leq \sum_{i \in I} m(A_i)$.

Dans ce cas, on dit que m est σ-additive.
Démonstration : (1) $\Rightarrow$ (2) :
Soit $A_i \in \mathcal{A}$. On définit $\tilde{A}_i$ par : $\tilde{A}_1 = A_1, \ldots \tilde{A}_n = A_n \setminus \tilde{A}_{n-1} \quad \forall n \geq 1$
Alors $\bigcup A_i = \bigsqcup \tilde{A}_i$.

Si $A \subset \bigcup A_i$, alors $A \subset \bigsqcup \tilde{A}_i$. Alors $A = \bigsqcup (A \cap \tilde{A}_i)$.
Donc $m(A) = m(\bigsqcup (\tilde{A}_i \cap A)) \leq \sum m(A_i)$.
$(2) \Rightarrow (1)$:
Si $A = \bigsqcup A_i \overset{(2)}{\Rightarrow} m(A) \leq \sum m(A_i)$.
$A \supset \bigsqcup_{i=1}^{n} A_i$ quel que soit n.
Donc $m(A) \geq \sum_{i=1}^{n} m(A_i)$. Donc $(n \to +\infty)$, $m(A) \geq \sum_{i=1}^{+\infty} m(A_i)$.

□

Définition : Soit $f : \Omega \to X$ une application. Si $\mathcal{A}$ est une algèbre (ou une tribu) sur Ω, alors on définit l'algèbre (tribu) image par :

$$f_* \mathcal{A} = \{A \subset X, f^{-1}(A) \in \mathcal{A}\}$$

Si $\mathcal{A}$ est une algèbre (tribu) sur X, alors

$$f^* \mathcal{A} = \{f^{-1}(A), A \in \mathcal{A}\}$$

est une algèbre (tribu) sur Ω.

La vérification du fait que $f^* \mathcal{A}$ et $f_* \mathcal{A}$ est une algèbre (tribu) découle des propriétés des préimages :

$$\begin{aligned} f^{-1}(A \cap B) &= f^{-1}(A) \cap f^{-1}(B) \\ f^{-1}(A \cup B) &= f^{-1}(A) \cup f^{-1}(B) \\ f^{-1}(A \setminus B) &= f^{-1}(A) \setminus f^{-1}(B) \end{aligned}$$

Définition : Si $f : (\Omega, \mathcal{A}, m) \to X$ est une application, on définit la mesure image (ou la loi) comme la mesure :

$$(f_* m)(Y) := m(f^{-1}(Y))$$

définie sur $f_* \mathcal{A}$.

Définition : m est dite finie ssi $m(X) < +\infty$

Définition : m est dite de probabilité ssi $m(X) = 1$

Définition : $f : (\Omega, \tau) \to (X, \mathcal{T})$ est dite mesurable si :

$$\forall Y \in \mathcal{T}, f^{-1}(Y) \in \tau$$

i.e.

$$\begin{aligned} f_* \tau &\supset \mathcal{T} \\ f^* \mathcal{T} &\subset \tau \end{aligned}$$

Exercice : Soit Ω, X des ensembles, $\mathcal{T}$ une tribu sur X. Soit $f : \Omega \to X$ une application, $g : \Omega \to X$ une application à valeurs dans un ensemble fini Y. Alors g est $f^* \mathcal{T}$ mesurable ssi $\exists h : (X, \mathcal{T}) \to (Y, \mathcal{P}(Y))$ mesurable telle que $g = h \circ f$. i.e. "g est f-mesurable ssi g ne dépend que de f".

2.2 Modélisation d'une expérience aléatoire finie (ex : jets de dés)

Soit Y un ensemble fini représentant les issues possibles. Il y a 2 manières de représenter un tirage aléatoire sur Y.

1. On se donne une mesure de probabilité sur $(Y, \mathcal{P}(Y))$. Pour ceci, il suffit de donner $p : Y \to [0, 1]$ tel que $\sum_{y \in Y} p(y) = 1$. On note P la mesure de probabilité ainsi créée.
2. On se donne un espace de probabilité abstrait $(\Omega, \mathcal{T}, \mathbb{P})$ et une application mesurable $f : \Omega \to Y$ telle que $f_* \mathbb{P} = P$.

Pour passer de 1. à 2., il suffit de prendre $\Omega = Y$, $\mathcal{T} = \mathcal{P}(Y)$, $\mathbb{P} = P$, $f = \mathrm{id}$.

L'expérience aléatoire consistant à jeter un nombre fini k de dés de valeurs possibles $Y_1, \dots, Y_k$ est simplement une expérience aléatoire à valeurs dans le produit $Y = Y_1 \times Y_2 \times \cdots \times Y_k$.

La description en termes de variables aléatoires consiste donc à se donner une application mesurable $f : (\Omega, \mathcal{T}, P) \to Y$, c'est à dire k applications mesurables $f_i : (\Omega, \mathcal{T}, P) \to Y_i$, définies sur *le même espace de probabilités*.

Définition : La loi de f (qui est une probabilité sur Y) est dite *loi jointe*. Les lois des f_i (qui sont des probabilités sur Y_i) sont dites *lois marginales*.

Remarque : La loi jointe détermine les lois marginales, qui peuvent se décrire explicitement par $m_i(y_i) = \sum_{y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_k} m(y_1, \dots, y_k)$.

Plus abstraitement, ce soint les mesures images $m_i = (\Pi_i)_* m_i$ où $\Pi_i : Y \to Y_i$ est la projection.

Remarque : La loi jointe est déterminée par $|Y_1| \times \cdots \times |Y_k| - 1$ nombres réels (-1 à cause de la contrainte $\sum p = 1$).

Les lois maginales sont déterminées par $|Y_1| + \cdots + |Y_k| - k$ nombres réels, ce qui est beaucoup moins.

Si on se donnes les marginales $m_1, \dots, m_k$, ilm existe de nombreuses lois jointes qui engendrent ces marginales. L'une d'entre elles est particulièrement intéressante : la loi produit $m((y1, \dots, y_k)) = m_1(y_1) \cdot \cdots \cdot m_k(y_k)$, qui correspond (par définition) à des expériences indépendantes.

Définition :

- Les événement A, B dans un espace de probabilité $(\Omega, \mathcal{T}, P)$ sont dits indépendants si $P(A \cap B) = P(A)P(B)$.
- Si $(X_i, \mathcal{T}_i)_{1 \le i \le k}$ sont des espaces mesurables (c'est à dire munis de tribus $\mathcal{T}_i$), les variables aléatoires (applications mesurables) $f_i : (\Omega, \mathcal{T}, P) \to (X_i, \mathcal{T}_i)$ sont dites *indépendantes* si $\forall Z_i \in \mathcal{T}_i, P(f_1 \in Z_1, \dots, P_k \in Z_k) = P(f_1 \in Z_i) \cdot \cdots \cdot P(f_k \in Z_k)$

Propriété : Les événements A et B sont indépendants ssi les variables aléatoires $\mathbb{1}_A, \mathbb{1}_B : (\Omega, \mathcal{T}, P) \to \{0, 1\}$ le sont.

Démonstration : Il suffit de montrer que $A^{\complement}$ et B sont indépendants (le reste est évident ou vient par symétrie).

$$\begin{aligned} P(A^{\complement} \cap B) &= P((\Omega \setminus A) \cap B) \\ &= P(B \setminus A \cap B) \\ &= P(B) - P(A \cap B) \\ &= P(B) - P(A)P(B) \\ &= (1 - P(A))P(B) \\ &= P(A^{\complement})P(B) \end{aligned}$$

□

Définition : Les événements $A_1, \dots, A_k$ sont dits indépendants si $\mathbb{1}_{A_1} \dots \mathbb{1}_{A_k} : \Omega \to \{0, 1\}$ le sont.

Remarque : Il ne suffit pas d'avoir l'indépendance deux à deux ou $P(A_1 \cap \cdots \cap A_k) = P(A_1) \cdot \cdots \cdot P(A_k)$.

Propriété : Il suffit d'avoir $P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdot \cdots \cdot P(A_{i_k})$, et ce $\forall \{i_1, \dots, i_k\} \subset [\![1, k]\!]$.

Démonstration : Il faut montrer que

$$(*) \quad P(B_1 \cap \cdots \cap B_k) = P(B_1) \cdot \cdots \cdot P(B_k) \forall B_i \in \{\emptyset, A_i, A_i^{\complement}, \Omega\}$$

Il découle de l'hypothèse que c'est vrai pour $B_i \in \{\emptyset, A_i, \Omega\}$.

Il suffit donc de constater que $(*)$ implique $P(B_1^{\complement} \cap B_2 \cap \cdots \cap B_k) = P(B_1^{\complement})P(B_2) \cdot \cdots \cdot P(B_k)$, ce qui se montre comme ce-dessus. On conclut par récurrence finie.

□

Exemple : Tirage non indépendant :

On tire $-$ chiffres dans $[\![1, 6]\!]$, en leur imposant d'être distincts. La loi jointe est donc : $P(y_1, \dots, y_6) = \begin{cases} 0 & \text{si non distincts} \\ \frac{1}{6!} & \text{si distincts} \end{cases}$.

Les lois marginales sont : $P_1(y_1) := \sum_{y_2, \dots, y_k} P(y_1, \dots, y_k) = \frac{5!}{6!} = \frac{1}{6}$. Les lois marginales sont donc les mêmes que pour un tirage indépendant !

Définition : On dit que $f_i,\ i \in I$ sont indépendantes si $f_i,\ i \in F$ le sont pour tout $F \subset I$ fini.

2.3 Modélisation d'une suite infinie de jets dés indépendants

Donnons-nous une suite infinie d'espaces de probabilités finis (Y_i, P_i) (la tribu est $\mathcal{P}(Y_i)$).

Pour chaque n, on a vu que l'on peut trouver des variables aléatoires indépendantes $f_i : (\Omega_n, \mathcal{T}_n, P_n) \to Y_i$ de loi P_i.

Question : peut-on prendre $(\Omega_n, \mathcal{T}_n, P_n)$ indépendant de n ?

Théorème 1
Il existe un espace de probabilité $(\Omega, \mathcal{T}, P)$ et une suite de variables aléatoires $f_i : \Omega \to Y_i$ qui sont indépendantes et de loi P_i.

Remarque : Les variables aléatoires f_i, $i \in \mathbb{N}$ sont indépendantes ssi $f_1, \dots, f_n$ le sont pour tout n.

L'hypothèse d'indépendance consiste donc à dire que, pour tout n et pour tout $(y_1, \dots, y_n) \in Y_1 \times \cdots \times Y_n$, l'événement $\{f_1 = y_1, \dots, f_n = y_n\}$ est mesurable $() \in \mathcal{T})$ et de mesure $P(f_1 = y_1, \dots f_n = y_n) = P_1(y_1) \cdot \cdots \cdot P_n(y_n)$.

En termes de loi, ceci implique que $\{y_1\} \times \dots \{y_n\} \times Y_{n+1} \times \dots$ est mesurable sur $X := \prod Y_i$ et que sa mesure est $m(\{y_1\} \times \cdots \times Y_{n+1}) = P_1(y_1) \cdot \cdots \cdot P_n(y_n)$.

2.4 Introduction de l'algèbre $\mathcal{A}_\infty$ engendrée par les cylindres finis

Sur le produit $X = \prod Y_i$, pour n fixé, les ensembles de la forme $\{y_1\} \times \cdots \times \{y_n\} \times Y_{n+1} \times \dots$ forment une partition finie (ce sont les cylindres finis), qui engendre une algèbre finie $\mathcal{A}_n$ (qui est donc aussi une tribu).

C'est l'algèbre engendrée par les n premières coordonnées. En effet si $\Pi : X \to Y_1 \times \cdots \times Y_n$ est la projection, alors $\mathcal{A}_n = \Pi^*(\mathcal{P}(Y_1 \times \cdots \times Y_n))$.

Cette algèbre décrit les parties de X qui peuvent être décrites en termes des n premières coordonnées.

On a $\mathcal{A}_n \subset \mathcal{A}_{n+1}$. On note $\mathcal{A}_\infty = \bigcup_{n \geq 1} \mathcal{A}_n$.

$\mathcal{A}_\infty$ est donc l'algèbre des parties de X qui dépendent d'un nombre fini de coordonnées. C'est l'algèbre engendrée par les cylindres finis.

Contrairement aux $\mathcal{A}_n$, $\mathcal{A}_\infty$ est infinie et ce n'est pas une tribu !

L'hypothèse d'indépendance des f_i implique que la loi m doit être définie sur $\mathcal{A}_\infty$, et qu'elle y est déterminée par la relation

$$(*) \quad m(\{y_1\} \times \cdots \times \{y_n\} \times Y_{n+1} \times \dots) = P_1(y_1) \cdot \cdots \cdot P(y_n)$$

Théorème 2
Il existe sur $X = \prod Y_i$ une tribu τ, qui contient $\mathcal{A}_\infty$, et une mesure m sur $\mathcal{T}$ qui vérifie $(*)$.

On vient en fait de voir que le théorème 1. implique le théorème 2. Réciproquement, il suffit de prendre $\Omega = X, \mathcal{T} = \tau, P = m, f =$ projection.

Pour démontrer l'utilité du théorème 2., donnons des exemples d'ensembles qu'il est naturel de considérer et qui sont dans τ mais pas dans $\mathcal{A}_\infty$. On suppose $Y_i \subset \mathbb{R}$

Exemple : L'ensemble $\{(y_i) \in X, \dfrac{y_1 + \dots y_n}{n} \to l\}$ est mesurable. En effet, il s'écrit : $\bigcap_{k \geq 1} \bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} \{\left|\dfrac{y_1 + \dots y_n}{n} - l\right| \leq \dfrac{1}{k}\}$, i.e. $\forall k \geq 1, \exists n \in \mathbb{N}, \forall m \geq n, \dots$. Chacun des ensembles est dans $\mathcal{A}_\infty$ donc l'ensemble considéré est dans τ.

2.5 Quelques résultats d'extension des mesures

Si A_n est une suite d'ensembles, on note :

$$\liminf A_n = \bigcup_n \bigcap_{m \geq n} A_m = \{A_i \text{ APCR}\}$$

$$\limsup A_n = \bigcap_n \bigcup_{m \geq n} A_m = \{A_i \text{ infinitely often (i.o.)}\}$$

Si τ est une tribu, que les $A_n \in \tau$, alors $\limsup A_n \in \tau$ et $\liminf A_n \in \tau$.

Propriété :
- $\liminf A_n \subset \limsup A_n$
- $\liminf A_n^\complement = (\limsup A_n)^\complement$

Démonstration : $\forall m, M, \quad \bigcap_{n\geq m} A_n \subset \underset{n\geq M}{A}_n$. Donc $\bigcap_{n\geq m} A_n \subset \limsup A_n$, donc $\liminf A_n \subset limsupA_n$

□

Exercice : $\limsup \mathbb{1}_{A_n} = \mathbb{1}_{\limsup A_n}$

Exemple : On considère un tirage aléatoire indépendant $f_n \in -1, 1^{\mathbb{N}}$, ce que l'on voit comme un jeu de hasard (le joueur gagne ou perd 1 à chaque étape). Étant donnée la richesse initiale r_0 et un objectif R, on considère l'événement {le joueur atteint la richesse R avant de se ruiner}.

Il s'écrit $\bigcup_{n\geq 1} \{y_1 + \dots y_n \geq -r_0 \quad \forall k < n \text{ et } y_1 + \dots + y_k = R - r_0\}$.

C'est une réunion dénombrable d'éléments de $\mathcal{A}_\infty$

Le théorème 2 sera déductible du théorème suivant :

> **Théorème 3 Hahn-Kolmogorov**
> Soit $\mathcal{A}$ une algèbre d'ensembles sur X. Soit $\underline{m}$ une mesure de probabilité additive sur $\mathcal{A}$, qui vérifie la propriété de σ-additivité.
> Alors il existe une tribu τ contenant $\mathcal{A}$, et une mesure de proba m sur τ qui prolonge $\underline{m}$. De plus, on peut prendre : $m(B) = \inf_{B\subset\bigcup A_i} \sum_{i\in\mathbb{N}} \underline{m}(A_i)$, où le inf est pris sur les recouvrements dénombrables de B par des éléments de $\mathcal{A}$.

Pour démontrer le théorème 2, on va appliquer le théorème 3 avec $\mathcal{A} = \mathcal{A}_\infty$, et $\underline{m}$ la mesure additive déterminée par $\underline{m}(\{y_1\} \times \dots \times \{y_n\} \times Y_{n+1} \times \dots) = P_1(y_1) \dots P_n(y_n)$.

Il nous suffit donc de vérifier que cette mesure additive a la propriété de σ-additivité.

> **Propriété :** Toute mesure additive sur $\mathcal{A}_\infty$ est σ-additive.

Démonstration : Soient $A \in \mathcal{A}_\infty$ et $A_i \in \mathcal{A}_i nfty$ tel que $A \subset \bigcup_i A_i$, alors $\exists n, A \subset \bigcup_{i=1}^{n} A_i$.

— méthode savante : c'est la compacité de A dans X muni de la topologie produit (les A_i sont ouverts et compacts)

— à la main : On pose $B_n = A \setminus \bigcup_{i=1}^{n} A_n$. On veut montrer que $\exists n, B_n = \emptyset$, sachant que $\bigcap_{n\geq 0} B_n = \emptyset$.
  On suppose que $B_n \neq \emptyset, \forall n$. On note $B_n(y_1) := \Pi_1^{-1}(y_1) \cup B_n$, ce sont les éléments de B_n qi commencent par y_1.
  Pour chaque y_1, $n \mapsto B_n(y_1)$ est décroissante. Comme $B_n = \bigcup_{y_1\in Y_1} B_n(y_1)$ (union finie) (et $B_n \neq \emptyset$), il existe y_1 tel que les $B_n(y_1)$ sont tous non vides.
  On fixe maintenant un tel y_1 et on reprend le même raisonnement sur y_2, puis... On obtient de la sorte une suite y.
  Ainsi, il existe une suite $(y_1, \dots) \in B_n \forall n$ car $\forall n, \exists k_n, B_n \in \mathcal{A}_{k_n}$.
  Ainsi, $\forall n, B_n \ni y$ donc $\bigcap B_n \neq \emptyset$. Absurde.

□

> **Propriété :** Dans le contexte du théorème d'Hahn-Kolmogorov, $m^* : \mathcal{P}(X) \to [0, \infty]$ est une *mesure extérieure*, c'est à dire que $m^*(\emptyset) = 0$, m^* est croissante, et $m^*\left(\bigcup_{i\in\mathbb{N}} Z_i\right) \leq \sum_{i\in\mathbb{N}} m^*(Z_i), \forall Z_i$.

Démonstration : Démontrons la dernière propriété. Fixons $\varepsilon > 0$. Pour tout i, il existe un recouvrement $A_{i,j}, j \in \mathbb{N}$ de Z_i tel que $\sum_j \underline{m}(A_{i,j}) \geq m^*(Z_i) \geq \sum_j \underline{m}(A_{i,j}) - \varepsilon 2^{-i}$, alors $A_{i,j}, i \in \mathbb{N}, j \in \mathbb{N}$ est un recouvrement de $\bigcup Z_i$, et $m^*(\bigcup Z_i) \leq \sum_{i,j} \underline{m}(A_{i,j}) \leq \sum_{i\geq 1}(m^*(Z_i) + \varepsilon 2^{-1}) \leq \varepsilon + \sum_{i\geq 1} m^*(Z_i)$.

□

Démonstration Démonstration du théorème d'Hahn-Kolmogorov : Deux étapes :

1. $m^*|_{\mathcal{A}} = \underline{m}$ Si $A \subset \bigcup_i A_i$, alors $\underline{m}(A) \leq \sum \underline{m}(A_i)$ par σ-additivité de $\underline{m}$. En prenant l'inf, on obtient $\underline{m}(A) \leq m^*(A)$. L'inégalité réciproque s'obtient en considérant le recouvrement trivial $A_1 = A, A_2 = A_3 = \dots = \emptyset$.

2. On dit que $Y \subset X$ est mesurable si, pour tout $\varepsilon > 0, \exists A \in \mathcal{A}$ tel que $m^*(Y \Delta A) \leq \varepsilon$. Alors l'ensembre $\mathcal{T}$ des parties mesurables est une algèbre.
   Démonstration :
   — si $m^*(Y \Delta A) \leq \varepsilon$, alors $m^*(Y^{\complement} \cap A^{\complement}) \leq \varepsilon$, donc $\mathcal{T}$ est stable par complément.
   — Soient Y, Z mesurables et A, B tels que $m^*(Y \Delta A) \leq \varepsilon, m^*(Z \Delta B) \leq \varepsilon$ alors $m^*((Y \cup Z)\Delta(A \cup B)) \leq 2\varepsilon$ car $(Y \cup Z)\Delta(A \cup B) \subset (Y \Delta A) \cup (Z \Delta B)$.

   □

3. m^* est une mesure additive sur $\mathcal{T}$.
   Démonstration : Y, Z disjoints, A, B comme ci-dessus.
   $(A \cap B) = (Y \cup (A \setminus Y)) \cap (Z \cup (B \setminus Z)) \subset Y \cap Z \cup (B \setminus Z) \cup (A \setminus Y)$
   donc $\underline{m}(A \cap B) \leq 2\varepsilon$
   $A \cup B = (Y \cup (A \setminus Y)) \cup (Z \cup (B \setminus Z)) \subset Y \cup Z \cup (A \setminus Y) \cup (B \setminus Z)$
   $\underline{m}(A \cup B) \leq m^*(Y \cup Z) + 2\varepsilon$
   et $\underline{m}(A \cup B) = \underline{m}(A) + \underline{m}(B) - \underline{m}(A \cap B) \geq \underline{m}(A) + \underline{m}(B) - 2\varepsilon \geq m^*(Y) - \varepsilon + m^*(Z) - \varepsilon - 2\varepsilon$.
   Finalement, $m^*(Y) + m^*(Z) \leq m^*(Y \cup Z) + 6\varepsilon$

   □

   Comme m^* est une mesure extérieure et une mesure additive sur l'algèbre $\mathcal{T}$, elle a la propriété de σ-additivité.

4. $\mathcal{T}$ est une tribu.
   Démonstration : $Y_i \in \mathcal{T}$. On veut montrer que $Y_\infty := \bigcup_i Y_i \in \mathcal{T}$. On peut supposer que les Y_i sont disjoints. Alors $\forall n, m^*(\bigsqcup_{i=1}^n Y_i) = \sum_{i=1}^n m^*(Y_i) \leq m^*(X) = 1$. Donc la série $\sum m^*(Y_i)$ converge, donc $\forall \varepsilon, \exists n, \sum_{i=n+1}^{+\infty} m^*(Y_i) \leq \varepsilon$.
   Alors en posant $Z = \bigcup_{i=1}^n Y_i$, on a $m^*(Y_\infty \setminus Z) \leq \varepsilon$, $Z \subset Y_\infty$. Ensuite, on prend $A \in \mathcal{A}$ tel que $m^*(A \setminus Z) \leq \varepsilon, m^*(Z \setminus A) \leq \varepsilon$. On obtient $A \setminus Y_\infty \subset A \setminus Z, Y_\infty \setminus A \subset (Z \setminus A) \cup (Y_\infty \setminus Z)$.

   □

□

Complément : on aurait pu donner une autre preuve du théorème 3 basée sur un résultat général sur les mesures extérieures. Lorsque m^* est une mesure extérieure, on dit que $Y \subset X$ est m^*-mesurable si
$\forall Z \subset X, m^*(Z) = m^*(Z \cap Y) + m^*(Z \cap Y^{\complement})$.

Théorème 4 Carathéodory
Si m^* est une mesure extérieure, l'ensemble $\mathcal{T}$ des parties m^*-mesurables est une tribu, et $m^*|_{\mathcal{T}}$ est une mesure.

Remarque : Dans le cas du théorème de Hahn, la tribu $\mathcal{T}$ est la même que celle introduite dans la démonstration précédente.
Démonstration Carathéodory ⇒ Hahn-Kolmogorov :
Il suffit de montrer que les éléments de $\mathcal{A}$ sont m^*-mesurables, et que $m^*|_{\mathcal{A}} = \underline{m}$.
— $m^*(A) \leq \underline{m}(A) \forall A \in \mathcal{A}$
— $m^*(A) \geq \underline{m}(A) \forall A \in \mathcal{A}$. En effet, si $A \subset \bigcup_i A_i$, on peut supposer les A_i disjoints. Alors par σ-additivité de $\underline{m}$ sur $\mathcal{A}$: $\underline{m}(A) = \sum_i \underline{m}(A_i) \geq m^*(A)$.
— Soit $A \in \mathcal{A}$ et $Z in \mathcal{P}(X)$. On considère un recouvrement A_i de Z.
  $\sum_i \underline{m}(A_i) = \sum_i \underline{m}(A_i \cap A) + \underline{m}(A_i \cap A^{\complement}) \geq m^*(Z \cap A) + m^*(Z \cap A^{\complement})$.
  On prend l'inf : $m^*(Z) \geq m^*(Z \cap A) + m^*(Z \cap A^{\complement})$. L'autre inégalité découle de la sous-additivité.

□

Démonstration Carathéodory :

1. $\mathcal{T}$ est une algèbre.
   Démonstration : On a $\emptyset \in \mathcal{T}, X \in \mathcal{T}$, et stabilité par complément de manière triviale.
   $A, B \in \mathcal{T} \Rightarrow \forall Y, m^*(Y) = m^*(Y \cap A) + m^*(Y \cap A^{\complement}) = m^*(Y \cap A \cap B) + m^*(Y \cap A \cap B^{\complement}) + m^*(Y \cap A^{\complement} \cap B^{\complement}) + m^*(Y \cap A^{\complement} \cap B)$.
   Remarque : $(A \cap B)^{\complement} = (B^{\complement} \cap A) \cup (B \cap A^{\complement}) \cup (A^{\complement} \cap B^{\complement})$ Donc $m^*(Y) \geq m^*(Y \cap (B \cup A)) + m^*(Y \cap (B \cap A)^{\complement})$

   □

2. m^* est additive sur $\mathcal{T}$
   Démonstration : $A, B \in \mathcal{T}$, $A \cap B = \emptyset$.
   $m^*(A \cup B) = m^*((A \cup B) \cap A) + m^*((A \cup B) \cap A^{\complement}) = m^*(A) + m^*(B)$

   □

3. $\mathcal{T}$ est une tribu.
   Démonstration : soit A_n une suite d'éléments deux à deux disjoints de $\mathcal{T}$. Posons $B_n = \bigcup_{k=1}^{n} A_n$ et $B_\infty = \bigcup_{k=1}^{\infty} A_n$.
   $\forall Y \subset X, m^*(Y \cap B_n) = m^*(Y \cap B_n \cap A_n) + m^*(Y \cap B_n \cap A_n^{\complement}) = m^*(Y \cap A_n) + m^*(Y \cap B_{n-1})$.
   Donc $m^*(Y \cap B_n) = \sum_{k=1}^{n} m^*(Y \cap A_k)$.
   Alors $m^*(Y) = m^*(Y) = m^*(Y \cap B_n) + m^*(Y \cap B_n^{\complement}) \geq \sum_{k=1}^{n} m^*(Y \cap A_k) + m^*(Y \cap B_{infty}^{\complement})$.
   À la limite : $m^*(Y) \geq \sum_{n=1}^{\infty} m^*(Y \cap A_n) + m^*(Y \cap B_\infty^{\complement})$

   □

□

On peut cependant se poser la question de l'unicité de m^* dans Hahn-Kolmogorov.

Théorème 5
Si $\mu : B \to [0,1]$ est une mesure sur une tribu $B \subset \mathcal{A}$, $\mu|_{\mathcal{A}} = \underline{m}$, alors $\mu = m^*$ sur $\mathcal{T} \cap B$.

Remarque : Il existe une plus petite tribu contenant $\mathcal{A}$ ($\bigcap_{\mathcal{T} \subset \mathcal{A},\ \mathcal{T} \text{ tribu}} \mathcal{T}$). Sur cette tribu, il existe une unique mesure prolongeant $\underline{m}$.
Démonstration :

1. Si $B \subset \bigcup_i A_i, A_i \in \mathcal{A}_i, B \in \mathcal{B}$, alors $\mu(B) \leq \sum_i \mu(A_i) = \sum_i \underline{m}(A_i)$. En prenant l'inf sur les familles A_i, on conclut $\mu \leq m^*|_{\mathcal{B}}$
2. Comme $\mu(B) \leq 1 - \mu(B^{\complement})$, si $B \in \mathcal{T}$, on a $A - m^*(B^{\complement}) = m^*(B)$ donc $\mu(b) \geq m^*(B)$ et donc $\mu(B) = m^*(B) si B \in \mathcal{T}$

□

2.6 Loi des grands nombres

On se donne $Y \subset \mathbb{R}$ fini, une mesure de probabilité p sur Y, et une suite finie $f_{i, i\in\mathbb{N}} : \Omega \to Y$ de variables aléatoires iid suivant la loi p. L'existence d'une telle suite découle des théorèmes de la section précédente.

Définition : Si $f : \Omega \to \mathbb{R}$ est une variable aléatoire prenant un nombre fini de valeurs, on note $E(f) = \sum_{y \in f(\Omega)} yP(f = y)$ l'espérance de f.

Dans notre contexte on note $e := E(f)$.

On définit $S_n = \frac{f_1 + \cdots + f_n}{n} : \Omega \to \mathbb{R}$. Chacune des variables aléatoires S_n prend un nombre fini de valeurs, mais les variables S_n ne sont pas indépendantes.

On veut montrer les trois énoncés suivants :

Théorème 6 Loi faible des grands nombres

$$P\left(\left|\frac{S_n}{n} - e\right| \geq \varepsilon\right) \to 0$$

Théorème 7 Loi forte des grands nombres

$$P\left(\frac{S_n}{n} \to e\right) = 1$$

Théorème 8

$$\forall \alpha > \frac{1}{2}, \quad P\left(\frac{S_n - ne}{n^\alpha} \to 0\right) = 1$$

2.6.1 Quelques outils de théorie de la probabilité

Pour démontrer ces résultats, on va avoir besoin d'un certain nombre d'autres outils.

Théorème 9 Inégalité de Markov
Si f est une variable aléatoire positive,

$$\forall a \in \mathbb{R}_*^+, \quad P(f > a) \leq \frac{E(f)}{a}$$

Démonstration : On écrit la définition de $E(f)$, on coupe la somme en deux selon $y > a$ ou $y \leq a$, on majore brutalement et on conclut.

□

Théorème 10 Inégalité de Bienaymé-Tchebychev

$$\forall a \in \mathbb{R}_*^+, \quad P(|f - E(f))| > a) \leq \frac{\mathrm{Var}(f)}{a^2}$$

Démonstration : On élève l'événement au carré, on conclut par Markov.

□

Propriété : $E(XY) := E(X)E(Y) + \mathrm{Cov}(X, Y)$,
$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$

Démonstration : Il suffit de l'écrire.

□

Lemme : Si X et Y sont deux variables aléatoires indépendantes, $\mathrm{Cov}(X, Y) = 0$.

Démonstration : Trivial.

□

Propriété Convergence monotone : Soit $(\Omega, \mathcal{T}, m)$ un espace mesuré.
Si (A_n) est une suite décroissante et que $m(A_1)$ est fini, alors $m(\bigcap_{n\in\mathbb{N}} A_n) = \lim m(A_n)$.
Si (A_n) est une suite croissante, alors $m(\bigcup_{n\in\mathbb{N}} A_n) = \lim m(A_n)$.

Démonstration : On pose $B_n = A_n \setminus A_{n-1}$, et par double passage à la limite, la propriété sur les suites croissantes est immédiate. Le résultat sur les suites décroissantes vient du passage au complémentaire.

□

Lemme Fatou ensembliste :

— $m(\liminf A_n) \leq \liminf m(A_n)$
— Si m est finie, $m(\limsup A_n) \geq \limsup m(A_n)$
— Si m est finie et $\limsup A_n = \liminf A_n = A$, alors $m(A_n) \to A$

Démonstration : On pose $B_n = \bigcap_{m \geq n} A_m$. C'est une suite croissante.

$m(B_n) \to m(\bigcup_n B_n) = m(\liminf A_n)$

$m(B_n) \leq m(A_n) \Longrightarrow \liminf m(A_n) \leq m(\liminf A_n)$

□

Lemme Premier lemme de Borel-Cantelli : Si $\sum_{n \geq 0} m(A_n)$ est finie, alors $m(\limsup A_n) = 0$.

Démonstration : $B_n := \underset{m \geq n_m}{A}$.

$m(B_n) \leq \sum_{m \geq n}^{\infty} m(A_n) \to 0$ (reste de série convergente)

Or, $\lim m(B_n) = m(\limsup A_n) = 0$.

□

Théorème 11 Inégalité de Kolmogorov

$P(\max_{A \leq k\ len} |\tilde{S}_k| \geq a) \leq \frac{n\mathrm{Var}(f)}{a^2}$

Démonstration : $T(\omega) :=$ le premier temps pour lequel $|\tilde{S}_n \geq a$. $T(\omega) \in \mathbb{N} \cup \{+\infty\}$
$(T = k) = \{|\tilde{S}_1| < a\} \cap \cdots \cap \{|\tilde{S}_{n-1}| < a\} \cap \{|\tilde{S}_n| \geq a\} \in \mathcal{A}_k$.
T est ainsi un *temps d'arrêt*.

$$\begin{aligned}
\mathrm{Var}\tilde{S}_n = E(\tilde{S}_n^2) &\geq \sum_{k=1}^{n} E(S_n^2 \mathbb{1}_{\{T=k\}}) \\
&= \sum_{k=1}^{n} E((\tilde{S}_n + \tilde{S}_k - \tilde{S}_k)\mathbb{1}_{\{T=k\}}) \\
&= \sum_{k=1}^{n} E(\tilde{S}_k^2 \mathbb{1}_{\{T=k\}}) + \sum_{k=1}^{n} E((\tilde{S}_n - \tilde{S}_k)\tilde{S}_k \mathbb{1}_{\{T=k\}}) \\
&\geq \sum_{k=1}^{n} a^2 P(T = k) + 0 + 0(*) \\
&\geq a^2 P(M_n \geq a)
\end{aligned}$$

et $(*)$: $\tilde{S}_n - \tilde{S}_k = \tilde{f}_{k+1} + \cdots + \tilde{f}_n$. Or $\tilde{S}_k \mathbb{1}_{\{T=k\}}$ ne dépend que des k premières valeurs (indépendance).

□

Remarque : Illustration de la notion de temps d'arrêt :

On considère un jeu de hasard : une suite f_i de v.a. i.i.d. à valeurs dans $\{-1, 1\}$, avec $P(1) = p$.

Supposons que le joueur choisit un temps $T(\omega)$ pour miser. Peut-il optimiser sa probabilité de gain $P(f_{T(\omega)}(\omega) = 1)$?

On peut choisir $T(\omega)$ le premier temps tel que $f_{T(\omega)} = 1$, mais cela nécessite de connaître tous les tirages.

En réalité, on ne dispose pas de l'almanach des sports, on n'a que l'information des $k - 1$ premiers tirages, i.e. $\{T = k\} \in \mathcal{A}_{k-1}$, c'est un temps d'arrêt.

Propriété : Si T vérifie cette condition, $P(f_{T(\omega)}(\omega) = 1) = p$.

Démonstration : $P(f_{T(\omega)}(\omega) = 1) = \sum_{k=1}^{\infty} P((T = k) \cap (f_k = 1))$. On conclut par indépendance.

□

Théorème 12 Inégalité de Hoeffding

$$P\left(\left|\tilde{S}_n\right| \geq a\right) \leq 2\exp(-\frac{2a^2}{Cn}), \qquad C = (\max f - \min f)^2$$

Lemme : Pour $\tilde{f}$ une v.a. centrée prenant un nombre fini de valeurs, on a :

$$E\left(e^{\theta\tilde{f}}\right) \leq e^{C\theta^2/8}, \qquad C = (\max\tilde{f} - \min\tilde{f})^2$$

Démonstration lemme : On pose $g(\theta) := \ln(E(e^{\theta\tilde{f}}))$. g est $\mathcal{C}^\infty$.

On a : $g(0) = 0$, $g'(\theta) = \dfrac{E(\tilde{f}e^{\theta\tilde{f}})}{E(e^{\theta\tilde{f}})}$ donc $g'(0) = E(\tilde{f}) = 0$.

Au voisinage de $\theta = 0$, il existe une constance c telle que $g(\theta) \leq c\theta^2$.

$g''(\theta) = \dfrac{E(\tilde{f}^2 e^{\theta\tilde{f}})E(e^{\theta\tilde{f}}) - E(\tilde{f}e^{\theta\tilde{f}})^2}{E(\theta e^{\theta\tilde{f}})^2}$

Ceci est la variance de la loi de proba sur $\tilde{Y}$ donnée par $P_\theta(\tilde{y}) = \dfrac{P(\tilde{y})e^{\theta\tilde{f}}}{E(e^{\theta\tilde{f}})}$.

En effet, $g''(\theta) = E\left(\dfrac{\tilde{f}^2 e^{\theta\tilde{f}}}{E(e^{\theta\tilde{f}})}\right) - E\left(\tilde{f}\dfrac{e^{\theta\tilde{f}}}{E(e^{\theta\tilde{f}})}\right)^2$.

Si g est une v.a. prenant un nombre fini de valeurs, alors $\mathrm{Var}(g) \leq (\max g - \min g)^2/4$. On remarque que la variance est invariante à translation de g près, donc on peut supposer $\max g = -\min g$. Or, $\mathrm{Var}(g) = E(g^2) - E(g)^2 \leq E(g^2) \leq (\max g)^2 \leq (\max g - \min g)^2/4$.

Ainsi, par l'inégalité des accroissements finis, comme $g''(\theta) \leq \dfrac{C}{4}$, on a $g(\theta) \leq \dfrac{C}{8}\theta^2$

Donc $E(e^{\theta\tilde{f}}) \leq \exp\left(\dfrac{C\theta^2}{8}\right)$

□

Démonstration Hoeffding : $P(\tilde{S}_n \geq a) = P(e^{theta\tilde{S}_n} \geq e^{\theta a}) \leq e^{-a\theta}E(e^{\theta\tilde{S}_n})$ par Markov.

$E(e^{\theta\tilde{S}_n}) = E(\prod e^{\theta\tilde{f}_i}) = \prod E(e^{\theta\tilde{f}_i}) = E(e^{\theta\tilde{f}})^n$.

$P(\tilde{S}_n \geq a) \leq e^{-a\theta}E(e^{\theta\tilde{f}})^n \leq \exp\left(\dfrac{nC\theta^2}{8} - \theta a\right)$.

On optimise par rapport à θ ($\theta = \dfrac{4a}{nC}$), et on conclut.

□

Remarque : Intérêt de ces inégalités en statistiques

Bienaymé-Tchebychev : $P(|\tilde{S}_n| \geq a) \leq \dfrac{n\mathrm{Var}(f)}{a^2}$

Hoeffding : $P(|\tilde{S}_n| \geq a) \leq 2\exp(-\dfrac{2a^2}{nC})$

Dans les deux cas, on note une décroissance en $\dfrac{n}{a^2}$.

Exemple d'application : sondage. La population peut avoir deux avis : $Y = \{0, 1\}$. On cherche à estimer par un sondage quelle est la proportion p de la population qui a l'avis 1 ($p = P(1)$).

On interprète un sondage comme étant une suite finie de variable aléatoires iid $f_1, \ldots, f_n$ tirées suivant la loi ci-dessus.

On s'attend à ce que $\dfrac{S_n}{n} \approx p$. Les inégalités rappelées ci-dessus nous donnent des majorations de $P\left(\left|\dfrac{S_n}{n} - p\right| \geq \varepsilon\right)$. Dans ce contexte, Hoeffding donne de meilleures estimations.

Exemple de valeurs numériques :

n	ε	Résultat
1000	5%	$p = 1,3\%$
1000	1%	$n\varepsilon^2 = 1$, on ne peut rien conclure

La première ligne indique que la probabilité d'être à plus de 5% d'erreur est d'au plus 1,3%.

Intervalle de confiance :

Ici, on fixe p la probabilité d'erreur, et on cherche ε, c'est à dire, par Hoeffding, $\varepsilon \leq \sqrt{\frac{1}{2n} \ln\left(\frac{2}{p}\right)}$. Par exemple, pour $n = 1000$, $p = 5\%$, on obtient $\varepsilon = 4,"\%$.

En général, avec ces données, on donne $\varepsilon = 3\%$. Cette disparité vient de la non-optimalité de l'inégalité de Hoeffding, et par le fait qu'il existe des modèles plus précis (approcher cette binomiale par une gaussienne grâce au théorème central limite par exemple).

2.6.2 Démonstrations des lois des grands nombres

Démonstration Loi faible des grands nombres :
$P(\left|\frac{S_n}{n} - e\right| \geq \alpha) = P(|S_n - E(S_n)| \geq n\alpha) \leq \frac{\mathrm{Var}(S_n)}{n^2\alpha^2} = \frac{n\mathrm{Var}(f)}{n^2\alpha^2} = \frac{\mathrm{Var}(f)}{n\alpha^2}$

□

Démonstration Loi forte des grands nombres : $\sum_n P(A_{n^2}(\varepsilon))$ converge.

Donc $m(\limsup(A_{n^2}(\varepsilon))) = 0$ d'après Borel-Cantelli. Donc $\frac{S_{n^2}}{n^2} \to E(f)$ p.p.

Montrons alors que si $\frac{S_{n^2}}{n^2} \to E(f)$ alors $\frac{S_n}{n} \to E(f)$.

On note $M = \max |f|$. Soit $k(n)$ tel que $k(n)^2 \leq n < (k(n)+1)^2$.

$$\begin{aligned}\left|\frac{S_n - nE(f)}{n}\right| &\leq \frac{|S_{k(n)^2} - k(n)^2E(f)| + (n - k(n)^2)(M + E(f))}{k(n)^2} \\ &\leq \left|\frac{S_{k(n)^2} - k(n)^2E(f)}{k(n)^2}\right| + \frac{(k(n)^2+1) - k(n)^2}{k(n)^2}(M + E(f))\end{aligned}$$

Chacun des termes tend vers 0, ce qui achève la preuve.

□

Démonstration Inégalité de Kolmogorov ⇒ théorème 3 :
$P(\frac{M_n}{n^\alpha} \geq \varepsilon) \leq \frac{\mathrm{Var}(f)}{n^{2\alpha-1}\varepsilon^2}$ où $M_n = \max_{1\leq k\leq n} |\tilde{S}_k|$

On fixe $R \in \mathbb{N}$ tel que $(2\alpha - 1)r > 1$.

$$P(\frac{M_{n^r}}{n^{r\alpha}} \geq \varepsilon) \leq \frac{\mathrm{Var}(f)}{\varepsilon^2 n^{(2\alpha-1)r}}$$

C'est le terme général d'une série convergente, donc par le premier lemme de Borel-Cantelli, on en déduit que $P(\frac{M_{n^r}}{n^{r\alpha}} \geq \varepsilon \text{ i.o.}) = 0$, et donc que p.p., $\frac{M_{n^r}}{n^{r\alpha}} \to 0$.

Mais $\frac{M_{n^r r}}{n^{r\alpha}} \to 0 \Rightarrow \frac{\tilde{S}_n}{n^\alpha} \to 0$.

En effet, soit $k(n)$ tel que $(k(n)-1)^r \leq n < k(n)^r$. Alors,

$$\begin{aligned}\frac{\tilde{S}_n}{n^\alpha} &\leq \frac{M_{k(n)^r}}{n^\alpha} \\ &= \frac{M_{k(n)^r}}{k(n)^{\alpha r}} \cdot \frac{k(n)^{\alpha r}}{n^\alpha} \\ &\leq \frac{M_{k(n)^r}}{k(n)^{\alpha r}} \cdot \frac{k(n)^{\alpha r}}{(k(n)-1)^{\alpha r}}\end{aligned}$$

Le premier terme tend vers 0, le second vers 1, ce qui conclut la preuve.

□

Démonstration Inégalité de Hoeffding ⇒ théorème 3 : $P(|\tilde{S}_n| \geq \varepsilon n^\alpha) \leq 2\exp\left(\frac{-2\varepsilon^2 n^{2\alpha-1}}{C}\right)$

C'est le terme général d'une série convergente, donc d'après le 1er lemme de Borel-Cantelli :
$P(|\tilde{S}_n| \geq \varepsilon n^\alpha \text{ i.o.})$ ie $\frac{\tilde{S}_n}{n^\alpha} \to 0$ p.p..

□

2.7 Estimations inférieures

On suppose que les variables aléatoires prennent au moins deux valeurs avec probabilité non nulle.

2.7.1 Quelques estimations inférieures

Théorème 13

$$P((S_n) \text{ bornée}) = 0$$

Théorème 14

$$P(\limsup \frac{\tilde{S}_n}{\sqrt{n}} = +\infty) = 1$$

Théorème 15 Loi du logarithme itéré
Presque partout, on a :

$$\limsup \frac{\tilde{S}_n}{\sqrt{2(\mathrm{Var}(f))n \ln\ln n}} = 1$$

2.7.2 Quelques résultats utiles

Lemme 2nd lemme de Borel-Cantelli : Si (A_n) est une suite d'événements indépendants, telle que $\sum_n P(A_n) = +\infty$, alors $P(\limsup A_n) = 1$.

Remarque : Si (A_n) est une suite d'événements indépendants, alors d'après les deux lemmes de Borel-Cantelli : $P(\limsup A_n) \in \{0, 1\}$.

Démonstration Second lemme de Borel-Cantelli :

$$B_n := \{f_n = a, \dots, f_{n+k} = a\}$$

$$P(B_n) = \underbrace{P(f = a)^k}_{\sum = +\infty}$$

Par BC2 ((B_n) indépendants), $P(B_n \text{ i.o.}) = 1$

Donc $P(B_n \text{ i.o.}) = 1$

□

Lemme : $\forall p, \exists C, \forall n, P_p(S_n = k) \geq \frac{1}{C\sqrt{n}} \exp\left(-nC \left|\frac{n}{k}n - p\right|^2\right)$

Démonstration : On cherche en fait à faire une estimation de la loi binomiale $P_p(S_n = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$

$\frac{P_p(S_n = k)}{P_{\frac{n}{k}}(S_n = k)} = \frac{p^k (1-p)^{n-k}}{\left(\frac{n}{k}\right)^k \left(1 - \frac{n}{k}\right)^{n-k}} = \exp\left(-nh\left(\frac{k}{n}, p\right)\right)$ où $h(s, p) = s \ln\left(\frac{s}{p}\right) + (1-s) \ln\left(\frac{1-s}{1-p}\right)$.

En particulier, comme h est $\mathcal{C}^2$ sur un compact, $h(s, p) \leq C|s - p|^2$ (à p fixé).

$P_p(S_n = k) \geq P_{\frac{k}{n}}(S_n = k) \exp\left(-n\left|\frac{k}{n} - p\right|^2\right)$

Il reste à estimer $P_{\frac{k}{n}}(S_n = k) = \frac{n!}{k!(n-k)!} \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k}$.

On conclut par Striling.

□

Propriété : $\forall A>0, \exists \delta>0, \forall n\in\mathbb{N}, P(\tilde{S}_n \geq A\sqrt{n}) \geq \delta$

Démonstration :

$$\begin{aligned}
P(\tilde{S}_n \geq A\sqrt{n}) &\overset{\tilde{S}_n=S_n-np}{=} \sum_{k\geq np+A\sqrt{n}} P(S_n=k)\\
&\geq \sum_{np+A\sqrt{n}\leq k\leq np+(A+\frac{1}{2})\sqrt{n}} P(S_n=k)\\
&\geq \sqrt{n}\frac{1}{C\sqrt{n}}\exp\left(-nC\frac{(A+\frac{1}{2})^2}{n}\right)\\
&= \underbrace{\frac{1}{C}\exp\left(-C(A+\frac{1}{2})^2\right)}_{\delta}
\end{aligned}$$

□

2.7.3 Démonstration des théorèmes 13-15

Démonstration Théorème 13 : On se donne $a \neq 0$, $P(a)>0$.

Affirmation : Pour tout $k\in\mathbb{N}$, il existe une infinité de $n\in\mathbb{N}$ tels que $f_n=a,\dots,f_{n+k}=a$.

Démonstration : On démontre cette affirmation en passant au complémentaire.
$B_n := A_n^{\complement}$. Les (B_n) sont indépendants.
On veut montrer que $P(\liminf B_n)=0$ i.e. $P(B_n \text{ APCR})=0$.
Or, $\liminf B_n = \bigcup_{n\in\mathbb{N}}\bigcap_{m\geq n} B_m$. Il suffit alors de montrer que $P(\bigcup_{m\geq n} B_m)=0$.

$$P(\bigcap_{m=n}^{l} B_m) \overset{\perp\!\!\!\perp}{=} \prod_{m=n}^{l} P(B_m) = \prod_{m=n}^{l}(1-P(A_m)) \overset{1-x\leq e^{-x}}{\leq} e^{-\sum_{m=n}^{l} P(A_m)} \underset{l\to+\infty}{\longrightarrow} 0.$$

Donc $P(\bigcap_{m\geq n} B_m)=0$.

□

On en déduit immédiatement le caractère non borné de la suite (S_n)

□

Démonstration Théorème 14 : On suppose A "assez grand" (dans un sens qui se précisera dans le corps de la preuve).

On considère la sous-suite n_k telle que $n_1=1$ et $n_{k+1}=n_k+4n_k^2$.
$P\left(\left|\tilde{S}_{n_{k+1}}-\tilde{S}_{n_k}\right|\geq a\right) = P\left(\left|\tilde{S}_{n_{k+1}-n_k}\right|\geq a\right)$.
On considère les événements $B_k=\left\{\tilde{S}_{n_{k+1}}-\tilde{S}_{n_k}\geq A\sqrt{n_{k+1}-n_k}\right\}=\left\{\tilde{S}_{4n_k}\geq 2An_k\right\}$.
B_k dépend de $f_{n_k+1}\dots f_{n_{k+1}}$ donc les (B_k) sont indépendants.
Comme $P(B_k)\geq\delta$, d'après BC2, $P(\limsup B_k)=1$.
Pour un point $\omega\in B_k$ i.o., on a donc une infinité de k tels que

$$\begin{aligned}
\tilde{S}_{n_{k+1}} &\geq \tilde{S}_{n_k}+2An_k\\
&\geq n_k(2A+\min f)(\text{pour } A\geq|minf|) \quad \geq An_k
\end{aligned}$$

Or, $n_{k+1}\leq 16n_k^2$ donc $n_k\geq\frac{\sqrt{n_{k+1}}}{4}$. Donc $\tilde{S}_{n_{k+1}}\geq\frac{A}{4}\sqrt{n_{k+1}}$ pour une infinité de k. Donc $\limsup\frac{\tilde{S}_n}{\sqrt{n}}\geq\frac{A}{4}$.

Donc $P(\limsup\frac{\tilde{S}_n}{\sqrt{n}}=\infty)=1$.

□

2.7.4 Quelques extensions

On a montré avec le second lemme de Borel-Cantelli l'existence de séquences qui apparaissent presque partout. Il se peut malgré tout que le temps d'attente soit très long.

Lemme Estimation du temps d'attente : $a\in \mathrm{Im}f, P(a)>0$. D'après le second lemme de Borel-Cantelli, la valeur a est prise une infinité de fois.
Soit T le premier temps pour lequel $f_T=a$. Alors $E(T)=\frac{1}{P(a)}$

Démonstration : On montre facilement que T suit une loi géométrique, et en utilisant $E(T) = \sum_{k \geq 1} P(T \geq k)$, on conclut rapidement.

$\square$

Chapitre 3

Intégration

3.1 Quelques remarques sur les tribus

On a déjà défini la tribu la plus petite possible.
Définition : Si $\mathcal{B} \subset \mathcal{P}(X)$, on appelle tribu engendrée par $\mathcal{B}$ la plus petite tribu contenu $\mathcal{B}$.

Lemme : Elle existe car une intersection de tribus est une tribu.

Démonstration : Soit $\mathcal{A}_{i,i\in I}$ des tribus sur X. Alors $A = \bigcap_{i\in I} \mathcal{A}_i$ est une tribu.

Soit en effet $(A_n) \in \mathcal{A}^{\mathbb{N}}$, alors $\forall i, \forall n, \quad A_n \in \mathcal{A}_i$, donc pour tout i, l'union des A_n appartient à A_i. Donc l'union, des A_n appartient à l'intersection des $\mathcal{A}_i$.

□

Avec ce vocabulaire, on a le théorème :

Théorème 16
Soit $\mathcal{A}$ une algèbre des parties de X et $\underline{m}$ une mesure de probabilités additive sur $\mathcal{A}$, qui de plus est σ-additive. Alors, il existe une unique mesure de probabilité m sur τ, la tribu engendrée par $\mathcal{A}$.

On peut, de la même façon, définir l'algèbre engendrée par $\mathcal{B}$. L'algèbre engendré est inclus dans la tribu engendrée.

On peut définir l'algèbre engendrée de façon plus explicite : on note $\mathcal{B}' = \{B \subset X; B \in \mathcal{B} \text{ ou } B^{\complement} \in \mathcal{B}\}$. L'algèbre engendrée est l'ensemble des unions finies d'intersections finies d'éléments de $\mathcal{B}'$. Il suffit en effet de vérifier que cette famille d'ensemble est stable par union, intersection et complémentaire.

— La stabilité par union est tautologique.

— De même pour la stabilité par complémentaire une fois qu'on a mointré la stabilité par intersection.

Montrons la stabilité par intersection :
Démonstration : $A = \bigcup_i \bigcap_j B_{ij}$, $\tilde{A} = \bigcup_k \bigcap_l \tilde{B}_{kl}$ et

$$\begin{aligned} A \bigcap A &= (\bigcup_i *)(\cup_k *) \\ &= \bigcup_{i,k} \left(\left(\bigcap_j B_{ij} \right) \cap \left(\bigcap_l B_{kl} \right) \right) \\ &= \bigcup_{i,k} \left(\bigcap_{j,l} B_{ij} \cap B_{kl} \right) \end{aligned}$$

est une union finie d'intersections finies donc le complémentaire de l'union des intersections des B_{ij} qui est l'intersection de l'union des complémentaires des B_{ij} est une intersection finie d'éléments de notre ensemble et donc est dedans.

□

Il est plus difficile de décrire les éléments de la tribu engendrée, notamment car l'ensemble des unions dénombrables d'intersections dénombrables n'est pas stable par intersection dénombrable.

Propriété : Soit $f : (\Omega, \mathcal{T}) \to X$ une application et soit $\mathcal{B} \subset \mathcal{T}$ engendrant $\mathcal{T}$.
Si $f^{-1}(B) \in \mathcal{T}, \forall B \in \mathcal{B}$, alors f est mesurable.

Démonstration : On considère l'ensemble ε des partitions de X dont la pré-image par f est mesurable. $\varepsilon = \{Y \in \mathcal{P}(X),\ f^{-1}(Y) \in \mathcal{T}\}$. On a déjà démontré que c'est une tribu (notée $f_*\mathcal{T}$. Cette tribu contient $\mathcal{B}$, donc elle contient $\mathcal{T}$.

□

Propriété : Les parties de $\mathbb{R}$ suivantes engendrent la même tribu.

1. Les intervalle ;
2. Les intervalles fermées ;
3. Les intervalles ouverts ;
4. les ouverts ;
5. Les fermées ;
6. Les intervalles $]-\infty, a], a \in \mathbb{D}$ (ou n'importe quel ensemble dense dans $\mathbb{R}$).

On l'appelle la tribu Borélienne, et on la note Borel($\mathbb{R}$).

Démonstration :
On note $\mathcal{T}_i$ la tribu engendrée par (1).
La tribu engendrée par les ouverts contient les fermés, donc $\mathcal{T}_5 \subset \mathcal{T}_4$ et de même, nous avons l'autre inclusion, donc égalité.
L'inclusion de $\mathcal{T}_2$ dans $\mathcal{T}_1$ est claire. Montrons que tout intervalle est dans $\mathcal{T}_2$. Tout intervalle est une réunion dénombrable d'intervalles fermés, donc appartient à $\mathcal{T}_2$.
$\mathcal{T}_3 = \mathcal{T}_1$ car tout intervalle est une intersection dénombrable d'intervalles ouverts. Réciproquement, tout ouvert est une union dénombrable d'intervalles ouvrets.
$\forall a < b, [a, b] = \bigcap\limits_n]-\infty, b_n] \setminus \bigcup\limits_n]-\infty, a_n]$ où b_n tend en décroissant vers b et a_n tend en croissant vers a sans jamais l'atteindre.

□

Propriété : Sur $[0, 1[$, les trois tribus suivantes sont les mêmes :

1. La tribu engendrée par les intervalles ;
2. La tribu engendrée par les $[0, a], a \in \mathbb{D} \cap [0, 1[$;
3. La tribu $\{B \cap [0, 1[,\ B \in \text{Borel}(\mathbb{R})\} = \iota^*(\text{Borel}(\mathbb{R}))$ où $\iota : [0, 1[\to \mathbb{R}$ est l'inclusion ($\iota(A) = A \cap [0, 1[$).

Définition : Soit $f_i : \Omega \to (X_i, \mathcal{T}_i), i \in I$ une famille d'applications.
La tribu engendrée par les $(f_i)_{i \in I}$ est la tribu de Ω engendrée par l'union sur $i \in I$ des $f_i^*(\mathcal{T}_i)$. C'est la plus petite tribu pour laquelle toutes les applications f_i sont mesurables.
Définition Tribu produit :
Soit $(X_i, \mathcal{T}_i)$ des espaces mesurables, $i \in I$.
La tribu produit sur $\Omega := \prod\limits_{i \in I} X_i$ est la tribu engendrée par les projections de Ω sur les X_i.
Autrement dit, c'est par définition la tribu engendrée par les cylindres $Z_i \times \prod\limits_{j \neq i} X_j$, avec $Z_i \in \mathcal{T}_i$.
Remarque :
Si I est fini ou dénombrable, la tribu produit est la tribu engendrée par le produit des Z_i.
Remarque :
Si I n'est pas dénombrable, les éléments de la tribu produit dépendent d'une famille au plus dénombrable de coordonnées, c'est à dire que si $A \in \mathcal{T}$, alors il existe une partie $J \subset I$ telle que $I \setminus J$ soit au plus dénombrable et telle que $\pi_J : \prod\limits_I X_i \to \prod\limits_J X_j$ vérifie $\pi_J(A) = \prod\limits_J X_j$, autrement dit, A ne dépend pas des coordonnées de J.

Propriété :
Si $f : (X_i, \mathcal{T}_i) \to \Omega$ et $F : \Omega \to (\prod X_i, \prod \mathcal{T}_i)$ l'application dont les f_i sont les coordonnées, alors la tribu engendrée par les f_i est $F^*(\prod \mathcal{T}_i)$.

Propriété :
L'application $F : (\Omega, \mathcal{T}) \to (\prod X_i, \prod \mathcal{T}_i)$ est mesurable si et seulement si ses coordonnées le sont.

3.1.1 Tribus de $\mathbb{R}^d$

$d \geq 1$

Théorème 17
Les tribus suivantes sont identiques :

1. Les tribus engendrées par les fermées ;
2. Les tribus engendrées par les ouverts ;
3. Le produit des tribus boréliennes ;
4. La tribu engendrée par les produits d'intervalles ;
5. La tribu engendrée par les produits des $]-\infty, a_i], a_i \in \mathbb{D}$ (ou n'importe quel ensemble dense).
6. La tribu engendrée par les produits d'intervalles ouverts.

On l'appelle la tribu borélienne de $\mathbb{R}^d$.

Remarque :
C'est un cas particulier de la tribu produit et c'est un cas particulier de tribu borélienne. Dans le cas général, sur un espace métrique, on appelle tribu borélienne la tribu engendrée par les ouverts. $\mathcal{T}_1 = \mathcal{T}_2$ est toujours vrai.
Démonstration :
$\mathcal{T}_3$ est la tribu engendrée par les produits de boréliens, donc $\mathcal{T}_3 \subset T_4 \subset T_5$, mais $(\pi_i)^* \mathcal{T}_5$ contient les intervalles $]-\infty, a], a \in \mathbb{D}$, donc contient la tribu borélienne.
En conséquence, $\pi_i^{-1}(Z_i)$ est dans $\mathcal{T}_5$ pour tout Z_i mesurable, donc $\mathcal{T}_5$ est la tribu produit. Donc $\mathcal{T}_5 = \mathcal{T}_4 = \mathcal{T}_3$, et c'est aussi la tribu engendrée par les produits d'intervalles ouvertes ($\mathcal{T}_6$).
Donc $\mathcal{T}_6 \subset \mathcal{T}_1$, et réciproquement. Donc $\mathcal{T}_6 = \mathcal{T}$, car tout ouvert est une réunion dénombrable de pavés ouverts.

□

3.2 La mesure de Lebesgue sur $\mathbb{R}$

Soit $\mathcal{B} := \text{Borel}(\mathbb{R})$.

Théorème 18
Il existe une unique mesure $\lambda : \mathcal{B} \to [0, +\infty]$ telle que $\lambda([a, b]) = b - a \quad \forall a \leq b$.

C'est la mesure de Lebesgue.
Remarque : λ n'est pas une mesure finie. Mais elle est σ-finie : on peut recouvrir $\mathbb{R}$ par un nombre dénombrable d'ensemble de mesure finie, par exemple $[-n; n]$, ou $[n, n+1]$.
On va commencer par construire la mesure de Lebesgue sur $[0, 1]$.

Théorème 19
Il existe une unique mesure borélienne $\lambda : \mathcal{B} \to [0, 1]$ telle que $\lambda([a, b]) = b - a \quad \forall a \leq b$ dans $[0; 1]$..

Démonstration :
On considère l'algèbre $\mathcal{A}$ engendrée par les intervalles $[a, b[$, c'est-à-dire les réunions finies d'intervalles de ce type. On peut supposer cette réunion disjointe.
On a une mesure additive $\underline{\lambda} : \mathcal{A} \to [0, 1]$ donnée par la somme des longueurs. On peut essayer d'appliquer le théorème d'extension de Hahn - Kolmogorov pour étendre $\underline{\lambda}$ en une mesure borélienne λ. Ceci implique l'unicité de λ.
On peut montrer l'existence par deux méthodes : ou bien on montre que $\underline{\lambda}$ est σ-additive sur $\mathcal{A}$, *ce qui n'est pas vrai pour toute mesure additive sur* $\mathcal{A}$; ou bien, ce que nous allons faire : On pose

$$Y := \{0, 1, \ldots, 9\}$$

$$\Omega = Y^{\mathbb{N}}$$

$$\begin{aligned} f : \Omega &\to [0, 1[\\ (y_i) &\mapsto \sum_{i=1}^{\infty} y_i 10^{-i} \end{aligned}$$

On met sur Ω la probabilité $\mathbb{P}$ qui est le produit des lois uniformes sur Y. Alors $f_i\mathbb{P}$ est la mesure cherchée. Il suffit de vérifier que pour tout $a \in \mathbb{D}$, $f^{-1}([0, a])$ est mesurable et que $\mathbb{P}\left(f^{-1}([0, a])\right) = a$ avec $a = 0, a_1 a_2 \dots a_k$.

$\{f \leq a\} = \{y_1 < a_1\} \cup (\{y_1 = a_1\} \cap \{y_2 < a_2\}) \cup \dots \cup \{y = a\}$

Ceci implique que $\{f \leq a\}$ est dans $\mathcal{A}_\infty$ et $P(f \leq a) = a_1 \cdot 10^{-1} + \dots + a_k \cdot 10^{-K} + 0 = a$.

□

3.2.1 Retour à $\mathbb{R}$

Si λ est une mesure sur $\mathbb{R}$, alors pour tout $B \in \mathcal{B}$, $\lambda(B) = \sum_{n \in \mathbb{Z}} \lambda(B \cap [n, n+1[)$

Notons temporairement P la mesure de Lebesgue sur $[n, n+1[$, alors on dit avoir :

$$\lambda(B) = \sum_{n \in \mathbb{Z}} P(B \cap [n, n+1[)$$

Il reste à vérifier que la formule ci-dessus définit bien une mesure sur $\mathbb{R}$, l'autre propriété étant facile à vérifier.

Soit B_k une suite de boréliens disjoints de $\mathbb{R}$:

Propriété :
λ est invariante par translation : c'est-à-dire que pour toute translation τ, $\tau_* \lambda = \lambda$.

Démonstration :
$(\tau_* \lambda)([a, b]) = \lambda\left(\tau^{-1}([a, b])\right) = b - a$

D'où le résultat désiré par unicité.

□

Propriété :
Soit μ une mesure borélienne sur $\mathbb{R}$, invariante par translation telle que $\mu([0, 1[) < \infty$, alors $\mu = \mu([0, 1[) \cdot \lambda)$.

Démonstration : Tous les points ont la même mesure, et c'est 0. Ensuite, comme $[0, 1[$ est la réunion disjointe de k translatés de $[0, \frac{1}{k}[$, on conclut que $\mu\left(\left[0, \frac{1}{k}\right[\right) = \frac{\mu([0, 1[)}{k}$. On pose alors $\tilde{\mu} = \frac{\mu}{\mu([0, 1[)}$.
Si $q = \frac{l}{k}$ est un rationnel, alors $[0, q[$ est l'union disjointe de l translatés de $[0, \frac{1}{k}[$, donc $\tilde{\mu}([0, q[) = q$.
Par passage à la limite croissante, $\tilde{\mu}([0, t[) = t \quad \forall t \geq 0$.
Et donc, $\tilde{\mu}(Z) = \lambda(Z)$ pour tout intervalle, donc $\tilde{\mu} = \lambda$.

□

3.3 Mesures de probabilité boréliennes sur $\mathbb{R}$

Si P est une mesure de probabilité sur $\mathbb{R}$, on définit sa *fonction de répartition* : $f(t) := P(]-\infty, t])$.

Propriété :
La fonction de répartition est
- croissante ;
- continue à droite ;
- de limite nulle en moins l'infini et de limite unitaire en plus l'infini.

Théorème 20
Si f est une fonction vérifiant les trois propriétés ci-dessus, alors il existe une unique probabilité borélienne $\mathbb{P}$ dont f est la fonction de répartition.

Démonstration :
— Unicité : si $\mathbb{P}$ et $\tilde{\mathbb{P}}$ engendrent la même fonction de répartition, alors elles ont les mêmes valeurs sur les intervalles de la forme $]a,b]$ donc sur tous les boréliens.
— Existence : on considère une fonction g qui va jouer le rôle d'"inverse" de f :

$$g(\omega) := inf\{x \in \mathbb{R}; f(x) \geq \omega\}$$

La condition 3 implique que pour tout ω dans $]0,1[$, $g(\omega) \in \mathbb{R}$ et g est croissante, donc pour tout $t \in \mathbb{R}$, $g(f(t)) \leq t$.

□

Lemme : Toute fonction croissante de $]0,1[$ dans $\mathbb{R}$ est mesurable

Démonstration :
Si I est un intervalle de $\mathbb{R}$, alors $g^-1(I)$ est un intervalle car si $x, y \in g^-1(I)$ alors $[x,y] \in f^-1(I)$. Comme les intervalles engendrent la tribu, on conclut que g est mesurable.

□

Il reste à vérifier que $g_*\lambda$ a bien f comme fonction de répartition. Dans le cas où f est inversible (f est continue et strictement croissante), alors $g = f^-1$.
Dans le cas général, on a encore $g^{-1}(]-\infty, t]) =]0, f(t)]$, $\omega \leq f(t) \Rightarrow g(\omega) \leq g \circ f(t) \leq t$. Réciproquement,

3.4 Intégration

On cherche à définir $\int_{\mathbb{R}} f(t)\mathrm{d}t$, et plus généralement $\int_{\Omega} f(\omega)\mathrm{d}m(\omega)$ lorsque $(\Omega, \mathcal{T}, m)$ est un espace mesuré, et $f : (\Omega, \mathcal{T}) \to (\mathbb{R}, \text{Borel})$ est mesurable.

Propriété : Si f, g sont mesurables, $a \in \mathbb{R}$, alors, $f+g$, af, fg, $\min(f,g)$, $\max(f,g)$ sont mesurables.

Démonstration : $\mathbb{R}^2$ muni de sa tribu Borel. $(f,g) : \Omega \to \mathbb{R}^2$ est mesurable donc si $h : \mathbb{R}^2 \to R$ est mesurable alors $h(f,g)$ est mesurable.
Or une fonction continue est mesurable (caractérisation des boréliens par les ouverts).

□

Remarque : Moyennant que cela ait un sens, ce qui n'est pas toujours le cas, cette propriété s'étend au cas de fonctions à valeurs dans $[-\infty, +\infty]$.
Remarque : Une fonction $f : \Omega \to \bar{\mathbb{R}}$ est mesurable ssii les préimages $f^{-1}(I)$ sont mesurables ou si $f^{-1}(\mathbb{R})$, $f^{-1}(+\infty)$, $f^{-1}(-\infty)$ sont mesurables et $f|_{f^{-1}(\mathbb{R})} : f^{-1}(\mathbb{R}) \to \mathbb{R}$ est mesurable.

Propriété : Soit (f_n) une suite de fonctions mesurables, à valeurs dans $\bar{\mathbb{R}}$, alors $\sup f_n$, $\inf f_n$, $\liminf f_n$, $\limsup f_n$ sont mesurables (donc en particulier $\lim f_n$ quand elle existe).

Démonstration : $\{\sup f_n > a\} = \bigcap_n \{f_n > a\}$ est mesurable. On conclut par passage à l'opposé et par composition de sup et d'inf.

□

Remarque : Surtout en analyse, on se préoccupe rarement de vérifier si quelque chose est mesurable ou pas, parce que presque tout ce qu'on a envie de considérer est mesurable.
Remarque : Si (f_α) est une famille quelconque de fonctions mesurables, $\sup_\alpha f_\alpha$ n'est pas fonrcément une fonction mesurable.
Exemple : On considère $f : x \in \mathbb{R} \mapsto \delta_{0,x}$ (symbole delta de Kronecker). Toute fonction $g : \mathbb{R} \to]0,1[$ est le sup d'une famille de fonctions de la forme $(x \mapsto f(x-a) - b)_{a,b}$.

Propriété :

1. Si $f : \Omega \to [0, +\infty]$ est mesurable alors f est la limite croissante d'une suite (f_n) de fonctions qui prennent un nombre fini de valeurs réelles positives (a.k.a des fonctions simples).
2. Si $f : \Omega \to \mathbb{R}$ est mesurable, alors c'est la limite uniforme d'une suite (f_n) de fonctions prenant un nombre dénombrable de valeurs réelles.

Démonstration :

1. Pour chaque n, on pose $f_n = \dfrac{k}{2^n}$ si $k \leq 4^k$ et $f \in [\dfrac{k}{2^n}, \dfrac{k+1}{2^n}[$ et $f_n = 2^n$ si $f \geq 2^n$.

   On fait du coloriage pour enfants vraiment jeunes, on s'en fiche de dépasser, les parents seront quand même contents à la fin. - Thibaud Raymond

   Alors f_n est croissante, $f_n \leq f$, $f_n \geq f - \dfrac{1}{2^n}$ sauf si $f \geq 2^n$ donc dans tous les cas $\forall x, f_n(x) \to f(x)$
2. On ne s'en servira pas donc on ne le démontre pas.

$\square$

Définition Intégrale d'une fonction positive :

1. Si f prend un nombre fini de valeurs : $\displaystyle\int f \, \mathrm{d}m := \sum_{y \in \mathrm{Im}(f)} y \cdot m(f = y)$. C'est une somme finie, mais certains termes peuvent être infinis.
   L'intégrale ici est bien définie, d'après la propriété précédente, pour
   — Si f est à valeurs dans $[0, +\infty]$
   — Si f est à valeurs réelles si m est finie, même lorsque f n'est pas positive.

On va noter S l'ensemble des fonctions mesurables réelles prenant un nombre fini de valeurs (les fonctions simples), et S^+ celles qui sont positives.

Propriété :

— Si m est finie, alors $f \in S \mapsto \displaystyle\int f \, \mathrm{d}m$ est linéaire.
— En général, pour tout m, $\displaystyle\int af + bf = a \int f + b \int g, \quad \forall a, b \geq 0, \forall f, g \in S^+$.

Définition : Pour $f : \Omega \to [0, +\infty]$ mesurable, on définit $\displaystyle\int f \, \mathrm{d}m = \sup_{\substack{g \in S^+ \\ g \leq f}} \int g \, \mathrm{d}m$

Propriété :

1. $0 \leq f \leq \tilde{f} \Rightarrow \displaystyle\int f \leq \int \tilde{f}$
2. $0 \leq f$ et $\displaystyle\int f \, \mathrm{d}m = 0 \Rightarrow m(f > 0) = 0$ (on dit $f = 0$ p.p.)
3. $f = 0$ p.p., $f \geq 0 \Rightarrow \displaystyle\int f \, \mathrm{d}m = 0$
4. $\displaystyle\int_A f \, \mathrm{d}m_A = \int_\Omega f \mathbb{1}_A \, \mathrm{d}m$ si $m_A(B) := m(A \cap B)$.

Démonstration :

1. Vient trivialement avec 2.
2. $0 = \displaystyle\int f \, \mathrm{d}m \geq \int \varepsilon \mathbb{1}_{\{f \geq \varepsilon\}} \, \mathrm{d}m = \varepsilon m(f \geq \varepsilon)$

   $\forall \varepsilon > 0, m(f \geq \varepsilon) = 0$. On prend l'intersection sur $\varepsilon = \dfrac{1}{n}$. Donc $m(f > 0) = 0$.
3. Soit $g \in S^+$ telle que $0 \leq g \leq f$ et $f = 0$ p.p. $\Rightarrow g = 0$ p.p. $\Rightarrow \displaystyle\int g \, \mathrm{d}m = 0$
4. Si $g \in S$, $\displaystyle\int_A g \, \mathrm{d}m_A = \sum_{y \in \mathrm{Im} g} y \cdot m_A(g = y)$

   $$\int_\Omega g \mathbb{1}_A \, \mathrm{d}m = \sum_{y \in \mathrm{Im} g \mathbb{1}_A} y \cdot m(g \mathbb{1}_A = y) = \sum_{y \in \mathrm{Im} y \cup \{0\}} y \cdot m(A \cap \{g = y\})$$

□

Théorème 21 Convergence monotone
Si (f_n) est une suite de fonctions mesurables à valeurs dans $[0,+\infty]$ croissante, alors

$$\int f_n \text{ d}m \underset{n\to+\infty}{\longrightarrow} \int \lim_{n\to+\infty} f_n \text{ d}m$$

Démonstration : $f_n \leq f$ donc $\int f_n \leq \int f \forall n$ donc $\lim \int f_n \leq \int f$ Il faut démontrer que $\lim \int f_n \geq \int f$ ce qui découle de $\lim \int f_n \geq \int g \quad \forall g \in S^+, g \leq f$ On se fixe $\lambda \in]0,1[$, et on définit $g_n := \begin{cases} \lambda g & \text{si } f_n \leq \lambda g \\ 0 & \text{si } f_n < \lambda g \end{cases}$

$g_n \leq f_n$, g_n est croissante et $g_n \to \lambda g$. On avait démontré le théorème de convergence monotone pour les fonctions simples. $\int g_n \text{ d}m \to \int \lambda g \text{ d}m$. $f_n \geq g_n \Rightarrow \lim \int f_n \geq \lim \int g_n \geq \int \lambda g \text{ d}m$

On a montré : $\forall \lambda \in]0,1[, \quad \lim \int f_n \geq \int \lambda g = \lambda \int g$. Donc, par passage au sup sur λ, $\lim \int f_n \geq \int g \quad, \forall g \in S^+, g \leq f$, et donc $\lim \int f_n \geq \int f$.

□

Lemme Fatou : Soit $(f_n : \Omega \to [0,+\infty])_n$ une suite de fonctions mesurables.
Alors $\int \liminf f_n \leq \liminf \int f_n$.

Remarque :

1. cf lemme de Fatou ensembliste.
2. Penser aux fonctions triangles pour un exemple où l'inégalité n'est pas respectée (une bosse qui part à l'infini).

Démonstration : $g_n = \inf_{k\geq n} f_k$. g_n est croissante, $\liminf f_n := f$. On montre par le théorème de convergence monotone $\int g_n \to \int f$, et $g_n \leq f_n$ donc $\liminf \int f_n \leq \lim \int g_n = \int f$.

□

Propriété : Si $a, b \geq 0$, $f, g \geq 0$ mesurables, alors $\int (af+bg) = a\int f + b\int g$

Démonstration : Soient $f_n, g_n \in S^+$ qui tendent simplement vers f et g respectivement.
Alors $af_n + bg_n \in S^+$, est croissante et tend vers $af+bg$.
Par théorème de croissance monotone, $\int f_n \to \int f$, $\int g_n \to \int g$, $\int af_n + bg_n \to \int af + bg$.
De plus ${}_intaf_n + bg_n = a\int f_n + b\int g_n$, donc on conclut par passage à la limite.

□

Propriété : Si $f : \Omega \to [0,+\infty]$ mesurable, alors la fonction $A \mapsto \int_A f \text{ d}m, \quad A \in \mathcal{T}$ est une mesure sur $(\Omega, \mathcal{T})$. On la note parfois fm.

Démonstration : On note $\mu(A) = \int_A f \text{ d}m$.
Soient $A, B \in \mathcal{T}, A \cap B = \varnothing$. Alors $\mathbb{1}_{A\sqcup B} = \mathbb{1}_A + \mathbb{1}_B$
$\mu(A \sqcup B) = \int_\Omega f\mathbb{1}_{A\sqcup B} \text{ d}m = \int_\Omega f\mathbb{1}_A \text{ d}m + \int_\Omega f\mathbb{1}_B \text{ d}m = \mu(A) + \mu(B)$
σ-additivité : $(A_n) \in \mathcal{T}^{\mathbb{N}}$ disjoints. On pose $B_n = \bigsqcup_{i=1}^n A_i$
$\mu(B_n) = \sum_{i=1}^n \mu(A_i)$ or $\sum_{i=1}^n \mu(A_i) \to \sum_{i=1}^\infty \mu(A_i)$.

$\mu(B_n) = \int_\Omega f\mathbb{1}_{B_n}\ \mathrm{d}m \to \int_\Omega f\mathbb{1}_B\ \mathrm{d}m$ (convergence monotone).
$f\mathbb{1}_{B_n}$ tend en croissant vers $f\mathbb{1}_B$

$\square$

Définition Fonctions mesurables à valeurs réelles :
$f^+ := \max(f, 0), \quad f^- := max(-f, 0)$
$|f| = f^+ + f^-, \quad f = f^+ - f^-$
On a envide définir $\int f := \int f^+ - \int f^-$. Ceci a un sens si l'une des intégrales de f^+ ou f^- est finie.
On dit que f est *intégrable* si elle est mesurable et si $\int f^+$ et $\int f^-$ sont finies, c'est à dire $\int |f| < +\infty$.
.

Définition : On note $\mathcal{L}^1(\Omega, m)$ l'ensemble des fonctions intégrales à valeurs réelles.

Propriété : $\mathcal{L}^1$ est un espace vectoriel et $f \mapsto \int_\Omega f\ \mathrm{d}m$ est linéaire sur $\mathcal{L}^1$.

Démonstration : Si f et g sont intégrable, alors $\int |af + bg|\ \mathrm{d}m \leq |a| \int |f| + |b| \int |g| < +\infty$.
Donc $af + bg \in \mathcal{L}^1$.
Si $_f \in \mathcal{L}^1, \alpha \in \mathbb{R}$, alors pour $a \geq 0$, $(af)^\pm = a(f^\pm)$
Donc $\int (af)^+ - (af)^- = \int a(f^+) - a(f^-) = a \int f^+ - f^- = a \int f$.
Idem pour $a \leq 0$.
$f, g \in \mathcal{L}^1$. A priori, $(f+g)^+ \neq f^+ + g^+$, donc on doit prendre quelques précautions.
$f + g = (f+g)^+ - (f = g)^- = f^+ - g^- + g^+ - g^-$. Le reste suit trivialement.

$\square$

Théorème 22 Convergence dominée
Soit $(f_n : (\Omega, \mathcal{T}, m) \to \mathbb{R})_n$ une suite de fonctions mesurables telles que :
- $f_n \to f$ presque partout.
- $\exists h \geq$ intégrable telle que $|f_n| \leq h$ presque partout pour tout n.

Sous ces hypothèses, $\int f_n \to \int f$

Démonstration : On applique le lemme de Fatou à $f_n + h$ et à $h - f_n$.
- $f_n \to f$ presque partout donc $\int f \leq \liminf \int (f_n + h)\ \mathrm{d}m = (\liminf \int f_n) + \int h$

  Donc $\int f \leq \liminf \int f_n$
- $h - f_n \to h - f$ (et sont positives)

  $\int h - f \leq \liminf \int h - f_n\ \mathrm{d}m = \int h - \limsup \int f_n\ \mathrm{d}m$

  Donc $\int f \geq \limsup \int f_n\ \mathrm{d}m \geq \liminf \int f_n \geq \int f$.

  Donc $\int f = \lim \int f_n$.

$\square$

Définition L'espace $\mathcal{L}^1$: On définit $\|f\|_1 = \int |f|\ \mathrm{d}m$ pour tout f mesurable.
$f \in \mathcal{L}^1 \Leftrightarrow \|f\|_1 < +\infty$.
$\|\cdot\|_1$ est une semi-norme (ne vérifie pas l'hypothèse de séparation).
Pour y remédier, on "identifie" (on quotiente par) les fonctions égales presque partout.
On note L^1 ce quotient.
Remarque : Attention, pour $f \in L^1$, $f(x)$ n'est pas défini, sauf si $m(x) > 0$.

Propriété : L^1 est un espace vectoriel : $f = \tilde{f}$ p.p., $g = \tilde{g}$ p.p. alors $af + bg = a\tilde{f} + b\tilde{g}$ p.p.

Démonstration : Évident.

$\square$

Propriété : $f = \tilde{f}$ p.p. $\Rightarrow \|f\|_1 = \|\tilde{f}\|_1$, et $\|\cdot\|_1$ induit une norme sur L^1, qu'on notera pareil. Donc $(L^1(\Omega, \mathcal{T}, m), \|\cdot\|_1)$ est un espace vectoriel normé.

Remarque : C'est une propriété générale : quotienter un EV par le noyau d'une semi-norme donne un espace vectoriel normé (on note π la surjection canonique).

Propriété : L'ensemble S des fonctions simples est dense dans L^1, plus exactement $S \cap L^1$ est dense dans L^1

Remarque : Plus exactement $\pi(S \cap \mathcal{L}^1)$ est dense dans L^1.
Démonstration : Posons $f \in \mathcal{L}^1$, $f = f^+ - f^-$.
$\exists (f_n^+), (f_n^-)$ des suites croissantes de fonctions simples positives qui convergent vers f^+ et f^-.
Montrons que $f_n^+ - f_n^- \underset{\mathcal{L}^1}{\to} f$. Par convergence monotone : $\displaystyle\int f_n^+ \to f^+, \int f_n^- \to f^-$.
Donc $\displaystyle\|f - (f_n^+ - f_n^-)\|_1 \leq \|f^+ - f_n^+\|_1 + \|f^- - f_n^-\|_1 = \int f^+ - f_n^+ + \int f^- - f_n^- \to 0$

$\square$

Le théorème de convergence dominée n'est pas optimal mais :

Propriété : Si $f_n \to f$ dans L^1 alors ils existe une sous-suite f_{n_k} qui satisfait les hypothèses du TCD :
1. $f_{n_k} \to f$ p.p.
2. $\exists h \geq 0$ intégrable telle que $|f_{n_k}| \leq h$ presque partout pour tout k.

Démonstration : On choisit n_k telle que $\|f_{n_k} - f\| \leq k^{-3}$.
Alors $m(|f_{n_k} - f| \geq \dfrac{1}{k}) \leq k^{-2}$ par Markov.
$\sum k^{-2} < +\infty$ donc on peut appliquer le premier lemme de Borel-Cantelli.
Donc $m(|f_{n_k} - f| \geq \dfrac{1}{k}\text{i.o.}) = 0$, et en dehorsde cet ensemble, $f_{n_k} \to f$, donc on a 1.
Pour 2., on pose $g = \sum |f_{n_k+1} - f_{n_k} : \Omega \to [0, +\infty]$.
$\displaystyle g_l = \sum_{k=1}^{l} |f_{n_k+1} - f_{n_k}|...$
$\displaystyle\|g\|_1 = \int g \underset{\text{CV monotonne}}{=} \lim \int g_l \leq \lim \sum_{k=1}^{l} \|f_{n_k+1} - f_{n_k}\| \leq \sum_{k=1}^{\infty} k^{-3} < +\infty$
Or, $\displaystyle|f_{n_k}| \leq |f| + g$ car $|f_{n_k} - f| = \sum_{k}^{\infty} |f_{n_{k+1}} - f_{n_k}| \leq g$

$\square$

Remarque : La mesure de Lebesgue sur $\mathbb{R}$ n'est pas finie mais σ-finie.

Lemme : Sur $(\Omega, \mathcal{T})$, la mesure m est σ-finie ssi il existe une probabilité $\mathbb{P}$ et une fonction $f : \Omega \to [0, +\infty[$ mesurable telle que $m = f\mathbb{P}$, où $f\mathbb{P}$ est la mesure $A \mapsto \displaystyle\int_A f \, \mathrm{d}\mathbb{P}$

Démonstration : Si $m = f\mathbb{P}$, on définit $A_n = \{f \leq n\}$. $\Omega = \displaystyle\bigcup_n A_n$ et

$$m(A_n) = \int_{A_n} f \, \mathrm{d}\mathbb{P} \leq \int_{A_n} n \, \mathrm{d}\mathbb{P} \leq n\mathbb{P}(A_n) \leq n$$

Réciproquement, si m est σ-finie, il existe (A_n) telle que $\Omega = \displaystyle\bigcup_n A_n$, $m(A_n) < \infty$. On pose $B_n = A_n \setminus \displaystyle\bigcup_{k=1}^{n-1} A_n$.
Alors $m(B_n) \leq m(A_n) < \infty, \bigsqcup B_n = \bigcup A_n = \Omega$ et les B_n disjoints.

On va considérer une fonction $f = a_n$ sur B_n, $a_n \in \mathbb{R}_+$. On définit $\mathbb{P} := \dfrac{1}{f} m$.

$$P(\Omega) = \sum_{k=1}^{\infty} P(B_k) = \sum_{k=1}^{\infty} \frac{1}{a_k} m(B_k).$$

On peut prendre $a_k = m(B_k) \cdot 2^{-k-1}$. Alors, $P(\Omega) = \sum_{k \geq 1} 2^{-(k+1)} = 1$.

□

Remarque : La mesure de Lebesgue est "localement finie", i.e. tout point admet un voisinage de mesure finie. On dit aussi que λ est une "mesure de Radon".

Mais ce concept concerne la structure topologique et la structure mesurée de $\mathbb{R}$.

3.5 Intégration et mesure image

$(\Omega, \mathcal{T}, m)$ et $\phi : \Omega \to X$ alors on a défni une tribu image $\phi_* \mathcal{T} = \{Y \subset X,\ \phi^{-1}(Y) \in \mathcal{T}\}$ et la mesure image (loi) $\phi_* m(Y) = m(\phi^{-1}(Y))$.

Propriété : La mesure $\mu = \phi_* m$ vérifie $\int f \,\mathrm{d}\mu = \int f \circ \phi \,\mathrm{d}m$ pour toute fonction f positive telle que $f \circ \phi$ est mesurable sur $(\Omega, \mathcal{T})$.

Remarque :

1. Cette relation est vraie aussi pour f à valeurs réelles si $f \circ \phi$ est m-intégrable.
2. Cette propriété caractérise μ car $\mu(Y) = \int \mathbb{1}_Y \,\mathrm{d}\mu = \int \mathbb{1}_Y \circ \phi \,\mathrm{d}m = \int \mathbb{1}_{\phi^{-1}(Y)} \,\mathrm{d}m = m(\phi^{-1}(Y))$.

Démonstration : Pour quelles fonctions f a-t-on $\int f \,\mathrm{d}\mu = \int f \circ \phi \,\mathrm{d}m$?

1. Pour les indicatrices de Y avec $Y \subset \phi_* \mathcal{T}$.
2. Par linéarité, pour toute fonction simple positive $\phi_* \mathcal{T}$-mesurable.
3. Pour toute fonction $\phi_* \mathcal{T}$-mesurable positive. Si f est une telle fonction, on considère (f_n) croissante de fonctions simples, $f_n \to f$.
   On peut passer à la limite dans l'égalité $\int f_n \,\mathrm{d}\mu = \int f_n \circ \phi \,\mathrm{d}m \overset{\text{CV monotone}}{\longrightarrow} \int f \,\mathrm{d}\mu = \int f \circ \phi \,\mathrm{d}m$.

□

Propriété : Si $\mu = fm$ et si g est mesurable positive alors $\int g \,\mathrm{d}\mu = \int fg \,\mathrm{d}m$.

Remarque : $\mathbb{P} = \frac{1}{f} m \Rightarrow m = f\mathbb{P}$ (même preuve)

3.6 Intégrales dépendant d'un paramètre

On étudie la fonction $F : x \mapsto \int_\Omega f(x, \omega) \,\mathrm{d}m(\omega)$, où $f : X \times \Omega \to \mathbb{R}$, $f(x, \cdot)$ intégrable sur $(\Omega, \mathcal{T}, m)$ pour tout x.

3.6.1 Continuité

On suppose que X est un espace métrique.

Théorème 23
Si :

— $x \mapsto f(x, \omega)$ est continue en x_0 pour presque tout ω
— $|f(x, \omega)| \leq g(\omega) \quad \forall x$ avec g intégrable.

Alors F est continue en x_0.

Démonstration : Si $x_n \to x_0$, alors $h_n(\omega) := f(x_n, \omega)$ converge presque partout vers $h(\omega) = f(x_0, \omega)$, et $|h_n(\omega)| \leq g$.

Théorème de convergence dominée : $F(x_n)$......

□

Exemple :

1. Si m est une mesure finie sur $\mathbb{R}$, on défjnit sa transformée de Fourier (fonction caractéristique en proba) : $\hat{m}(y) := \displaystyle\int e^{iyx}\ \mathrm{d}m(x)$. $y \mapsto \hat{m}(y)$ est continue sur $\mathbb{R}$, par le théorème qui précède : $y \mapsto e^{iyx}$ est $\mathcal{C}^0 \quad \forall x$ et $|e^{iyx}| \leq 1$, 1 est intégrable.
2. Convolutions. $f \in \mathcal{L}^1(\mathbb{R}, \lambda)$ et h continue et bornée. On définit $h{*}f : x \mapsto \displaystyle\int h(x-t)f(t)\ \mathrm{d}\lambda(t)$, $x \in \mathbb{R}$. $h{*}f$ est continue.
   $x, t \mapsto |h(x-t)f(t)| \leq \ (\sup h)|f|$
   Remarque : Si h et $f \geq 0$, alors $\forall x \in \mathbb{R}, h{*}f(x) = f{*}h(x) = \displaystyle\int f(x-t)h(t)\ \mathrm{d}t$. $\displaystyle\int_{\mathbb{R}} h(x-t)f(t)\ \mathrm{d}\lambda(t) = \displaystyle\int_{\mathbb{R}} h(-s)f(s-x)\ \mathrm{d}\lambda(s)$ car λ est invariante par translation. $= \displaystyle\int_{\mathbb{R}} h(s)f(x-s)\ \mathrm{d}\lambda(s)$ car λ est invariante par la symétrie $s \mapsto -s$ (ne change pas la longueur des intervalles).
3. $f \in \mathcal{L}^1 \Rightarrow x \mapsto \displaystyle\int_{-\infty}^{x} f(t)\ \mathrm{d}t \in \mathcal{C}^0$.

   Démonstration : $F : x \mapsto \displaystyle\int_{\mathbb{R}} f(t)\mathbb{1}_{]-\infty,x]}(t)\ \mathrm{d}t$. La fonction intégrée est dominée par $|f(t)|$ et $\forall x_0, x \mapsto f(t)\mathbb{1}_{]-\infty,x]}(t)$ est continue en x_0, $\forall t \neq x_0$ or $\lambda(x_0) = 0$. Donc F est continue en x_0.

   □

Remarque : Notation : $\displaystyle\int_A f(t)\ \mathrm{d}\lambda(t)\ /\ \int_a^b f(t)\ \mathrm{d}t$.

Définition :
$$\int_a^b f(t)\ \mathrm{d}t = \begin{cases} \displaystyle\int_{[a,b]} f(t)\ \mathrm{d}\lambda(t) & \text{si } a \leq b \\ -\displaystyle\int_{[a,b]} f(t)\ \mathrm{d}\lambda(t) & \text{si } a > b \end{cases}$$

3.6.2 Dérivabilité

$X \subset \mathbb{R}$ est un intervalle ouvert.

Théorème 24
- $\forall x \in X, f(x, \cdot)$ est intégrable.
- $x \mapsto f(x, \omega)$ est dérivable en x_0, $\forall \omega \in \Omega$
- $\exists g \in \mathcal{L}^1, |f(x,\omega) - f(x_0,\omega)| \leq |x - x_0|g(\omega), \quad \forall x \in X, \omega \in \Omega$.

Alors F est dérivable en x_0 et $F'(x_0) = \displaystyle\int_{\Omega} \partial_x f(x_0, \omega)\ \mathrm{d}m(\omega)$.

Remarque : $|f(x,\omega) - f(x_0,\omega)| \leq |x - x_0|g(\omega)$ est satisfaite si $\partial_x f$ existe partout et $|\partial_x f(x,\omega)| \leq g(\omega), \quad \forall x \in X, \omega \in \Omega$.

Démonstration : $\dfrac{F(x_n) - F(x_0)}{x_n - x_0} = \displaystyle\int_{\Omega} \frac{f(x_n,\omega) - f(x,\omega)}{x_n - x_0}\ \mathrm{d}m(\omega)$

$x_n \to x_0$.

On peut appliquer le TCD : $\displaystyle\int_{\Omega} \partial_x f(x_0,\omega)\ \mathrm{d}m(\omega)$.

□

Exemple :

1. $f \in \mathcal{L}^1$, h est $\mathcal{C}^1$, bornée de dérivée bornée, alors $f{*}h$ est $\mathcal{C}^1$ et $(f{*}h)'f{*}(h')$.

   Démonstration : $\displaystyle\int h(x-t)f(t)\ \mathrm{d}t$, $g(x,t) = h(x-t)f(t)$, $\partial_x g = h'(x)$.....

   □

2. Si m est une mesure finie et $\displaystyle\int |x|\ \mathrm{d}m < +\infty$ alors $\hat{m}$ est $\mathcal{C}^1$ et $\hat{m}' = \displaystyle\int ixe^{iyx}\ \mathrm{d}x$

   Démonstration : $f(y,x) = e^{iyx}$, $\partial_y f = iye^{iyx}$, $|\partial_y f| \leq |x|$.

   □

Propriété Théorème fondamental de l'analyse : Si f est continue, $F : x \mapsto \int_0^x f(t)\ \mathrm{d}t$ est dérivable de dérivée $f(x)$.

Démonstration : $F(x_0+x) - F(x_0) = \int_{x_0}^{x_0+x} f(t)\ \mathrm{d}t \leq \int_{x_0}^{x_0+x} f(x_0) + \varepsilon(|x|)\ \mathrm{d}t \leq f(x_0)x + |x|\varepsilon(|x|)$. De même, $F(x_0+x) - F(x_0) \geq \int_{x_0}^{x_0+x} (f(x_0) - \varepsilon(|x|))\ \mathrm{d}t \leq f(x_0)x - |x|\varepsilon(|x|)$.

$\varepsilon(x) = \sup_{t\in[x_0-x,x_0+x]} |f(t) - f(x_0)|$

$f(t) \leq f(x_0) + \varepsilon(x) \forall t [x_0 - x, x_0 + x]$.

□

3.6.3 Inégalité de Jensen

$\phi : \mathbb{R} \to \mathbb{R}$ convexe ($\forall t \in [0,1], \forall x, y \in \mathbb{R}, \phi(tx + (1-t)y) \leq t\phi(x) + (1-t)\phi(y), \phi : \mathbb{R} \to \bar{\mathbb{R}}$) On appelle domaine de ϕ l'ensemble $\{\phi < +\infty\} \subset \mathbb{R}$

Remarque : Si $J \subset \mathbb{R}$ est un intervalle, et si $\phi : J \to \bar{\mathbb{R}}$ est convexe, alors $\tilde{\phi} = \phi$ sur $J, +\infty$ hors de J est convexe.

Théorème 25 Inégalité de Jensen
Si μ est une mesure de probabilité sur Ω, et $\phi : \mathbb{R} \to \bar{\mathbb{R}}$ est convexe, alors $\forall f$ mesurable, intégrable (ou semi-intégrable), $\phi\left(\int_\Omega f\ \mathrm{d}\mu\right) \leq \int_\Omega \phi \circ f\ \mathrm{d}\mu$.

(on dit que f est semi-intégrable si $\int f^+\ \mathrm{d}\mu < +\infty$ ou $\int f^-\ \mathrm{d}\mu < +\infty$).

Remarque : Si f est semi-intégrable mais pas intégrable, alors $\int f\ \mathrm{d}\mu \in \{-\infty, +\infty\}$. Alors $\phi\left(\int f\ \mathrm{d}\mu\right) := \lim_{t\to\pm\infty} \phi(t)$

Remarque : $\phi(txx + (1-t)y) \leq t\phi(x) + (1-t)\phi(y) \Leftrightarrow \phi\left(\int \omega\ \mathrm{d}\mu\right) \leq \int \phi(\omega)\ \mathrm{d}\mu$ où $\mu = t\delta_x + (1-t)\delta_y$.

Par récurrence sur le nombre de points, on montre que si ϕ est convexe, alors $\phi(\mathrm{Bar}(\mathrm{x_1}, \dots, \mathrm{x_n}, \mathrm{t_1}, \dots, \mathrm{t_n})) \leq \mathrm{Bar}(\phi(\mathrm{x_1}), \dots, \phi(\mathrm{x_n}), \mathrm{t_1}, \dots, \mathrm{t_n})$ où Bar désigne le barycentre.

On pourrait donner une preuve de Jensen sur cette base. On fait autrement.

Propriété : Si $\phi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ est convexe.

1. Le domaines de ϕ est un intervalle I.
2. ϕ est continue sur $\mathring{I}$ et admet des dérivées et à droite en chaque point de $\mathring{I}$ telle que $\phi'(x^-) \leq \phi'(x^+) \forall x \in \mathring{I}$
3. $\forall x_0 \in \mathring{I}$, il existe une fonction affine $l(x)$ telle que $l(x_0) = \phi(x_0)$ et $l(x) \leq \phi \forall x$.
4. $\lim_{\pm\infty} \phi$ existe.

Démonstration : Fixons $x_0 \in \mathring{I}$, alors $p : x \mapsto \dfrac{\phi(x) - \phi(x_0)}{x - x_0}$ est croissante (découle de la convexité, laissée en exercice au lecteur).

Alors $\lim_{x\to x_0^+}$ et $\lim_{x\to x_0^-}$ existent donc, et $p(x_0^-) \leq p(x_0^+)$. De plus, ces limites sont finies : si $x_1 \leq x_0 \leq x_2$ dans I, alors $-\infty \leq p(x_1) \leq p(x_0^-) \leq p(x_0^+) \leq p(x_2) < +\infty$.

Pour $x_0 \in I$, on considère $a \in [\phi'(x_0^-), \phi'(x_0)^+]$. $l(x) := a(x - x_0) + \phi(x_0)$. $l(x_0) = \phi(x_0)$.

$l(x) \leq \phi(x) \Leftrightarrow a(x - x_0) \leq \phi(x) - \phi(x_0)$, ce qui est vrai par croissance du taux de variation : si $x \geq x_0, a \leq \dfrac{\phi(x) - \phi(x_0)}{x - x_0}$.

Existence des limites. Si $\exists x \in \mathring{I}, \phi'(x_0^+) > 0$, alors $\phi(x) \geq \phi(x80) + \phi'(x_0^+)(x - x_0) \forall x \geq x_0$ donc $\phi \underset{+\infty}{\to} +\infty$

Si $\forall x_0 \in \mathring{I}, \phi'(x_0) \leq 0 \Rightarrow \phi$ est décroissante $\Rightarrow \phi$ admet une limite.

Idem en $-\infty$.

□

Démonstration Jensen : I le domaine de ϕ.

Cas principal : $y_0 = \int f \, \mathrm{d}\mu \in \mathring{I}$.

$\exists l : y \mapsto ay + b$ tel que $l(x_0) = \phi(y_0), l \leq \phi$.

$\phi(y_0) = l(y_0) = l\left(\int f \, \mathrm{d}\mu\right) = a \int f \, \mathrm{d}\mu + b = \int af + b \, \mathrm{d}\mu = \int l \circ f \, \mathrm{d}\mu \leq \int \phi \circ f \, \mathrm{d}\mu$.

Cas où $y_0 = \int f \, \mathrm{d}\mu = \sup I$. Si $\int \phi \circ f \, \mathrm{d}\mu = +\infty$, on n'a rien à démontrer.

Si $\int \phi \circ f \, \mathrm{d}\mu < +\infty$, alors $\phi \circ f < +\infty$ presque partout, donc $f \in I$ presque partout, donc $f \leq y_0$ presque partout. Alors, $y_0 - f \geq 0$ et $\int y_0 - f = 0$ donc $f = y_0$ presque partout. Donc $\phi \circ f = \phi(y_0)$ presque partout donc $\int \phi \circ f = y_0$.

Idem si $y_0 = \inf I$.

Dernier cas : $y_0 \neq \bar{I}$. Alors $\int \phi \circ f \, \mathrm{d}\mu = +\infty$ car $\mu(\phi \circ f = +\infty) > 0$. En effet, sinon $\phi \circ f < +\infty$ presque partout donc $f \in I$ presque partout, donc $\int f \, \mathrm{d}\mu \in \bar{I}$.

□

3.7 L'espace L^2

On se donne $(\Omega, \mathcal{T}, m)$ mesuré.

Définition : $\mathcal{L}^2(\Omega, \mathcal{T}, m)$ est l'ensemble des fonctions mesurables de carré intégrable ($\int f^2 \, \mathrm{d}m < +\infty$).

Remarque : Si m est finie, $\mathcal{L}^2 \subset \mathcal{L}^1$ car $|f| \leq 1 + f^2$.

On note $\|f\|_2 = \sqrt{\int f^2 \, \mathrm{d}m}$, $< f, g >= \int fg \, \mathrm{d}m$.

Lemme : $\mathcal{L}^2$ est un espace vectoriel et $f, g \mapsto < f, g >$ est une forme bilinéaire symétrique positive.

Démonstration : $(f+g)^2 \leq 2f^2 + 2g^2$, $()\lambda f)^2 \leq \lambda^2 f^2$, donc $\mathcal{L}^2$ est un sous-espace vectoriel des fonctions mesurables.

$< \cdot, \cdot >$ est clairement bilinéaire, symétrique et positive. Il faut juste montrer que c'est bien défini.

$|fg| \leq \frac{1}{2}(f^2 + g^2)$ par inégalité arithmético-géométrique, donc fg est intégrable si $f, g \in \mathcal{L}^2$.

□

Ainsi, $\mathcal{L}^2$ est presque un pré-hilbertien, ce est bien mais pas top. Donc on quotiente encore une fois par la semi-norme $\| \cdot \|_2$: $L^2 = \mathcal{L}^2 / \| \cdot \|_2$.

Ainsi, L^2 est un espace pré-hilbertien, muni du produit scalaire $< \cdot, \cdot >$. En particulier, on a donc Cauchy-Schwarz ($< f, g > \leq \sqrt{\|f\|_2 \|g\|_2}$).

On a aussi l'égalité du parallélogramme ($\|f+g\|^2 + \|f-g\|^2 = 2(\|f\|^2 + \|g\|^2)$)

Définition : Un espace de Hilbert est un espace pré-hilbertien complet (i.e. les suite de Cauchy convergent).

Propriété : Un espace vectoriel normé E est complet ssi les séries les séries normalement convergentes sont convergentes.

$\sum \|f_n\| < +\infty \Rightarrow \sum_{k=1}^{n} f_k$ converge dans E.

Démonstration : Soit $(E, \| \cdot \|)$ un evn complet, et soit f_n une suite telle que $\sum \|f_n\| < +\infty$ alors $S_n := \sum_{k=1}^{n} f_n$ vérifie $\|X_n - S_m\| = \| \sum_{k=n+1}^{m} f_k \| \leq \sum_{k=n+1}^{m} \|f_k\| \leq \sum_{k=n+1}^{\infty} \|f_k\| = \varepsilon(n)$ i.e. $\forall \varepsilon, \exists n_0, \forall m \geq n_0, n \geq n_0, \|S_n - S_m\| \leq \varepsilon$ donc S_n converge par complétude.

Réciproque : soit f_n une suite de Cauchy. Il suffit e montrer que f_n a une sous-suite convergente.

On choisit n_k tel que : $\|f_{n_{k+1}} - f_{k_k}\| \leq 2^{-k}$. Alors $\sum \|f_{n_{k+1}} - f_{n_k}\| < +\infty$ donc $\sum (f_{n_{k+1}} - f_{n_k})$ converge. Donc f_{n_k} converge.

□