

Intégration et probabilités

Introduction

Références :

- BILLINGSLEY, *Probability 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 $\int f dx$ pour $f : \mathbb{R}^d \rightarrow \mathbb{R}$
- Définir, préciser la notion mathématique décrivant une suite infinie de jets de dés

Par exemple :

- Si $f : \mathbb{R} \rightarrow \mathbb{R}$, on peut définir $\int f$ 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 $\lambda(A) := \inf_{(R_i)} \sum_{i=1}^{\infty} \lambda(R_i)$ où R_i est une suite de rectangles recouvrant A .

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

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 $x : \mathbb{N} \rightarrow X, Y \subset X$. 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 $x : \mathbb{N} \rightarrow X$ 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 $\mathbb{N}^2 \rightarrow \mathbb{N}$.

□

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 $b_i : \mathbb{N} \rightarrow A_i$ bijection.
 $(i, j) \mapsto b_i(j)$

(nb : ceci requiert en fait l'axiome du choix dénombrable) Alors $\mathbb{N}^2 \rightarrow \bigcup_i A_i$ est surjective.

Donc $\bigcup_i A_i$ est au plus dénombrable.

Or $\bigcup_i A_i \supset A_i$.

Donc $\bigcup_i A_i$ est dénombrable.

□

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 $x : X \rightarrow \mathcal{P}(X)$ une bijection.
 $x \mapsto A_x$

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.

□

Exercice :

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

— \mathbb{R} est non dénombrable.

lim sup et lim inf

Définition :

Soit $(x_n)_{n \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ (plus généralement $\in \bar{\mathbb{R}}^{\mathbb{N}}$). Alors $s_n := \sup_{k \geq n} x_k$.

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 $x_n \rightarrow l$ 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 $s_n \rightarrow l$ et $i_n \rightarrow l$.

□

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 $s_n - \frac{1}{n} \leq x_{k_n} \leq s_n$. $n \mapsto x_{k_n}$ converge vers $\limsup x_n$.

□

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

□

À partir de maintenant, on considérera I dénombrable.

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

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, $\sum_{k=1}^N a_{\sigma(k)} = \sum_{i \in \sigma(\llbracket 1, N \rrbracket)} a_i \leq \sum_{i \in I} a_i$. 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 $\sigma : \mathbb{N} \rightarrow \mathbb{N}$ 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 $F \subset \llbracket 1, N \rrbracket^2$. 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 $(M \rightarrow +\infty), \sum_{i=1}^N \sum_{j=1}^{+\infty} a_{i,j} \leq \sum_{(i,j) \in \mathbb{N}^2} a_{i,j}$.

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

□

Séries absolument convergentes

Soit $(a_i)_{i \in I}$ une famille de réels tels que $\sum_{i \in I} |a_i|$ 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 : $\sum_{i \in I} a_i^+ - \sum_{i \in I} a_i^- = \sum_{k=1}^{+\infty} a_{\sigma(k)}$ et ce quel que soit $\sigma : \mathbb{N} \rightarrow I$ bijection.

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

$$\sum_{k=1}^n a_{\sigma(k)} = \sum_{k=1}^n a_{\sigma(k)}^+ - \sum_{k=1}^n a_{\sigma(k)}^- \xrightarrow{n \rightarrow +\infty} \sum_{k=1}^{+\infty} a_{\sigma(k)}^+ - \sum_{k=1}^{+\infty} a_{\sigma(k)}^-$$

□

Corollaire : Sous réserve de convergence absolue, on a :

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

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

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 $x \in B^c$ et alors $A(x) \subset B^c$ i.e. $A(x) \cap B = \emptyset$

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

□

Définition : Si \mathcal{A} est une algèbre de X et $m : \mathcal{A} \rightarrow [0, +\infty]$ une fonction.

On dit que m est une *mesure additive* si :

- $m(\emptyset) = 0$
- $m(A \sqcup B) = m(A) + m(B)$ ($A \cap B = \emptyset$)

Définition : Si $\mathcal{T} \subset \mathcal{P}(X)$ est une tribu, $m : \mathcal{T} \rightarrow [0, +\infty]$ 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 $m : \mathcal{A} \rightarrow [0, +\infty]$ 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 : $\tilde{A}_1 = A_1, \dots, \tilde{A}_n = A_n \setminus \tilde{A}_{n-1} \quad \forall n \geq 1$

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 $m(A) = m(\bigsqcup (A \cap \tilde{A}_i)) \leq \sum m(A_i)$.

(2) \Rightarrow (1) :

Si $A = \bigsqcup A_i \stackrel{(2)}{\Rightarrow} m(A) \leq \sum m(A_i)$.

$A \supset \bigsqcup_{i=1}^n A_i$ quel que soit n .

Donc $m(A) \geq \sum_{i=1}^n m(A_i)$. Donc $(n \rightarrow +\infty)$, $m(A) \geq \sum_{i=1}^{+\infty} m(A_i)$.

□

Définition : Soit $f : \Omega \rightarrow X$ 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 $f : (\Omega, \mathcal{A}, m) \rightarrow X$ 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 : $f : (\Omega, \tau) \rightarrow (X, \mathcal{T})$ 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 $f : \Omega \rightarrow X$ une application, $g : \Omega \rightarrow Y$ une application à valeurs dans un ensemble fini Y . Alors g est $f^*\mathcal{T}$ mesurable ssi $\exists h : (X, \mathcal{T}) \rightarrow (Y, \mathcal{P}(Y))$ mesurable telle que $g = h \circ f$. i.e. " g est f -mesurable ssi g ne dépend que de f ".

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 $p : Y \rightarrow [0, 1]$ 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 $f : \Omega \rightarrow Y$ telle que $f_*\mathbb{P} = P$.

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

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 $Y = Y_1 \times Y_2 \times \dots \times Y_k$.

La description en termes de variables aléatoires consiste donc à se donner une application mesurable $f : (\Omega, \mathcal{T}, P) \rightarrow Y$, c'est à dire k applications mesurables $f_i : (\Omega, \mathcal{T}, P) \rightarrow Y_i$, 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 sont les mesures images $m_i = (\Pi_i)_* m_i$ où $\Pi_i : Y \rightarrow Y_i$ est la projection.

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

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

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

Définition :

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

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

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

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

□

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

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

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

Démonstration : Il faut montrer que

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

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

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

□

Exemple : Tirage non indépendant :

On tire — chiffres dans $\llbracket 1, 6 \rrbracket$, 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 : $P_1(y_1) := \sum_{y_2, \dots, y_6} P(y_1, \dots, y_6) = \frac{5!}{6!} = \frac{1}{6}$. 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.

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 $f_i : (\Omega_n, \mathcal{T}_n, P_n) \rightarrow Y_i$ 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 (Ω, \mathcal{T}, P) et une suite de variables aléatoires $f_i : \Omega \rightarrow Y_i$ 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 $(y_1, \dots, y_n) \in Y_1 \times \dots \times Y_n$, l'événement $\{f_1 = y_1, \dots, f_n = y_n\}$ est mesurable ($\in \mathcal{T}$) et de mesure $P(f_1 = y_1, \dots, f_n = y_n) = P_1(y_1) \cdot \dots \cdot P_n(y_n)$.

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

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

Sur le produit $X = \prod Y_i$, pour n fixé, les ensembles de la forme $\{y_1\} \times \dots \times \{y_n\} \times Y_{n+1} \times \dots$ forment une partition finie (ce sont les cylindres finis), qui engendrent 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 $\Pi : X \rightarrow Y_1 \times \dots \times Y_n$ est la projection, alors $\mathcal{A}_n = \Pi^*(\mathcal{P}(Y_1 \times \dots \times Y_n))$.

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}_∞ 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}_∞ 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}_∞ , et qu'elle y est déterminée par la relation

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

Théorème 2

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

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

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}_∞ . On suppose $Y_i \subset \mathbb{R}$

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

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 :

- $\liminf A_n \subset \limsup A_n$
- $\liminf A_n^c = (\limsup A_n)^c$

Démonstration : $\forall m, M, \bigcap_{n \geq m} A_n \subset \bigcap_{n \geq M_n} A_n$. Donc $\bigcap_{n \geq m} A_n \subset \limsup A_n$, donc $\liminf A_n \subset \limsup A_n$.

□

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}_{∞}

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}_{∞} est σ -additive.

Démonstration : Soient $A \in \mathcal{A}_{\infty}$ et $A_i \in \mathcal{A}_{\infty}$ 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 $B_n = A \setminus \bigcup_{i=1}^n A_i$. 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 $B_n(y_1) := \Pi_1^{-1}(y_1) \cap B_n$, ce sont les éléments de B_n qui 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, $m^* : \mathcal{P}(X) \rightarrow [0, \infty]$ 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'ensemble \mathcal{T} des parties mesurables est une algèbre.

Démonstration :

- si $m^*(Y \Delta A) \leq \varepsilon$, alors $m^*(Y^c \cap A^c) \leq \varepsilon$, 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.

$$(A \cap B) = (Y \cup (A \setminus Y)) \cap (Z \cup (B \setminus Z)) \subset Y \cap Z \cup (B \setminus Z) \cup (A \setminus Y)$$

$$\text{donc } \underline{m}(A \cap B) \leq 2\varepsilon$$

$$A \cup B = (Y \cup (A \setminus Y)) \cup (Z \cup (B \setminus Z)) \subset Y \cup Z \cup (A \setminus Y) \cup (B \setminus Z)$$

$$\underline{m}(A \cup B) \leq m^*(Y \cup Z) + 2\varepsilon$$

$$\text{et } \underline{m}(A \cup B) = \underline{m}(A) + \underline{m}(B) - \underline{m}(A \cap B) \geq \underline{m}(A) + \underline{m}(B) - 2\varepsilon \geq m^*(Y) - \varepsilon + m^*(Z) - \varepsilon - 2\varepsilon.$$

$$\text{Finalement, } m^*(Y) + m^*(Z) \leq m^*(Y \cup Z) + 6\varepsilon$$

□

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 $\forall n, m^*(\bigcup_{i=1}^n Y_i) = \sum_{i=1}^n m^*(Y_i) \leq m^*(X) = 1$. 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

$$\forall Z \subset X, m^*(Z) = m^*(Z \cap Y) + m^*(Z \cap Y^c).$$

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 \Rightarrow 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} : $\underline{m}(A) = \sum \underline{m}(A_i) \geq m^*(A)$.

- Soit $A \in \mathcal{A}$ et $Z \in \mathcal{P}(X)$. On considère un recouvrement A_i de Z .

$$\sum_i \underline{m}(A_i) = \sum_i \underline{m}(A_i \cap A) + \underline{m}(A_i \cap A^c) \geq m^*(Z \cap A) + m^*(Z \cap A^c).$$

On prend l'inf : $m^*(Z) \geq m^*(Z \cap A) + m^*(Z \cap A^c)$. 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.

$A, B \in \mathcal{T} \Rightarrow \forall Y, m^*(Y) = m^*(Y \cap A) + m^*(Y \cap A^c) = m^*(Y \cap A \cap B) + m^*(Y \cap A \cap B^c) + m^*(Y \cap A^c \cap B^c) + m^*(Y \cap A^c \cap B)$.

Remarque : $(A \cap B)^c = (B^c \cap A) \cup (B \cap A^c) \cup (A^c \cap B^c)$ Donc $m^*(Y) \geq m^*(Y \cap (B \cup A)) + m^*(Y \cap (B \cap A)^c)$

□

2. m^* est additive sur \mathcal{T}

Démonstration : $A, B \in \mathcal{T}, A \cap B = \emptyset$.

$m^*(A \cup B) = m^*((A \cup B) \cap A) + m^*((A \cup B) \cap A^c) = m^*(A) + m^*(B)$

□

3. \mathcal{T} est une tribu.

Démonstration : soit A_n une suite d'éléments deux à deux disjoints de \mathcal{T} . Posons $B_n = \bigcup_{k=1}^n A_k$ et $B_\infty = \bigcup_{k=1}^\infty A_k$.

$\forall Y \subset X, m^*(Y \cap B_n) = m^*(Y \cap B_n \cap A_n) + m^*(Y \cap B_n \cap A_n^c) = m^*(Y \cap A_n) + m^*(Y \cap B_{n-1})$.

Donc $m^*(Y \cap B_n) = \sum_{k=1}^n m^*(Y \cap A_k)$.

Alors $m^*(Y) = m^*(Y) = m^*(Y \cap B_n) + m^*(Y \cap B_n^c) \geq \sum_{k=1}^n m^*(Y \cap A_k) + m^*(Y \cap B_{n-1}^c)$.

À la limite : $m^*(Y) \geq \sum_{n=1}^\infty m^*(Y \cap A_n) + m^*(Y \cap B_\infty^c)$

□

□

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

Théorème 5

Si $\mu : B \rightarrow [0, 1]$ 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 $\mathcal{A} \left(\bigcap_{\mathcal{T} \subset \mathcal{A}, \mathcal{T} \text{ tribu}} \mathcal{T} \right)$. Sur cette tribu, il

existe une unique mesure prolongeant \underline{m} .

Démonstration :

1. Si $B \subset \bigcup_i A_i, A_i \in \mathcal{A}, B \in \mathcal{B}$, 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 $\mu(B) \leq 1 - \mu(B^c)$, si $B \in \mathcal{T}$, on a $\mu(B) = m^*(B)$ donc $\mu(B) \geq m^*(B)$ et donc $\mu(B) = m^*(B)$ si $B \in \mathcal{T}$

□

Loi des grands nombres

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

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

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

On définit $S_n = \frac{f_1 + \dots + f_n}{n} : \Omega \rightarrow \mathbb{R}$. 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

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

Théorème 7 Loi forte des grands nombres

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

Théorème 8

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

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

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

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

□

Propriété : $E(XY) := E(X)E(Y) + Cov(X, Y)$,
 $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

Démonstration : Il suffit de l'écrire.

□

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

Démonstration : Trivial.

□

Propriété Convergence monotone : Soit (Ω, \mathcal{T}, m) un espace mesuré.

Si (A_n) est une suite décroissante et que $m(A_1)$ est fini, alors $m\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim m(A_n)$.

Si (A_n) est une suite croissante, alors $m\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim m(A_n)$.

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 $m(A_n) \rightarrow m(A)$

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

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

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

□

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

Démonstration : $B_n := \bigcap_{m \geq n} A_m$.

$$m(B_n) \leq \sum_{m \geq n} m(A_m) \rightarrow 0 \text{ (reste de série convergente)}$$

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

□

Théorème 11 Inégalité de Kolmogorov

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

Démonstration : $T(\omega) :=$ le premier temps pour lequel $|\tilde{S}_n| \geq a$. $T(\omega) \in \mathbb{N} \cup \{+\infty\}$

$$(T = k) = \{|\tilde{S}_1| < a\} \cap \dots \cap \{|\tilde{S}_{n-1}| < a\} \cap \{|\tilde{S}_n| \geq a\} \in \mathcal{A}_k.$$

T est ainsi un *temps d'arrêt*.

$$\text{Var} \tilde{S}_n = E(\tilde{S}_n^2) \geq \sum_{k=1}^n E(S_n^2 \mathbb{1}_{\{T=k\}})$$

$$= \sum_{k=1}^n E((\tilde{S}_n + \tilde{S}_k - \tilde{S}_k) \mathbb{1}_{\{T=k\}})$$

$$= \sum_{k=1}^n E(\tilde{S}_k^2 \mathbb{1}_{\{T=k\}}) + \sum_{k=1}^n E((\tilde{S}_n - \tilde{S}_k) \tilde{S}_k \mathbb{1}_{\{T=k\}}) \text{ et } (*) : \tilde{S}_n - \tilde{S}_k = f_{k+1} + \dots + \tilde{f}_n. \text{ Or}$$

$$\geq \sum_{k=1}^n a^2 P(T = k) + 0 + 0(*)$$

$$\geq a^2 P(M_n \geq a)$$

$\tilde{S}_k \mathbb{1}_{\{T=k\}}$ 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.

□

Démonstrations des lois des grands nombres