Mesure de Lebesgue et représentation de Riesz

On donne dans ce texte une démonstration unifiée et auto-contenue des théorèmes d'extension de Caratheodory et de représentation de Riesz. Le premier permet en particulier d'étendre uniquement une probabilité définie sur une algèbre de parties d'un ensemble à la σ-algèbre qu'il engendre. Il mène à une construction immédiate de la mesure de Lebesgue sur $[0,1]$ et donc de la mesure de Lebesgue sur $\mathbb{R}^d$. Tout ceci est expliqué dans la section 2. L'unicité dans le théorème d'extension repose sur un énoncé qui fait l'objet de la section 1, qui permet de vérifier que tous les éléments d'une σ-algèbre ont une propriété donnée, caractérisée par leur appartenance à une classe de parties. L'énoncé de représentation de Riesz affirme que la formule $\mathbb{E}(f) = \int f(x)\,\mathbb{P}(dx)$, pour $f \in \mathcal{C}_c(\mathbb{R}^d)$, définit une bijection entre l'espace des probabilités sur $\mathbb{R}^d$ et l'espace des formes linéaires continues de norme 1 sur l'espace $\mathcal{C}_c(\mathbb{R}^d)$ des fonctions continues à support compact, muni de la norme du supremum. Ceci est exposé dans la section 3.

1 – Classe monotone

On fixe dans cette partie un ensemble Ω dont on considère des familles de sous-ensembles. Un λ-système (de Ω) est une famille de parties de Ω stable par complémentation et réunion finie ou dénombrable disjointes. Un π-système (de Ω) est une famille de parties de Ω stable par intersection finie. Une famille de parties de Ω qui est à la fois un λ-système et un π-système est donc une σ-algèbre (= une tribu). On note $\sigma(\mathcal{P})$ la plus petite σ-algèbre contenant une famille $\mathcal{P}$ de parties d'un ensemble donné ; on parle de σ-algèbre engendrée par $\mathcal{P}$. Pensez ci-dessous à $\mathcal{L}$ comme la famille des parties de Ω ayant une certaine propriété.

Lemme 1. *Soit $\mathcal{L}$ un λ-système et $\mathcal{P}$ un π-système. Si $\mathcal{P} \subset \mathcal{L}$ alors $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Tous les éléments de $\sigma(\mathcal{P})$ ont donc la propriété caractérisant $\mathcal{L}$ si les éléments de $\mathcal{P}$ ont cette propriété.

Démonstration – Notons $\mathcal{L}(\mathcal{P})$ le plus petit λ-système contenant $\mathcal{P}$. Il nous suffit de voir que c'est un π-système ; ce sera alors une σ-algèbre qui contiendra $\mathcal{P}$ donc $\sigma(\mathcal{P})$. Pour montrer que $A \cap B \in \mathcal{L}(\mathcal{P})$ pour tout $A, B \in \mathcal{L}(\mathcal{P})$, on montre d'abord que la famille $\big\{A; A \cap B \in \mathcal{L}(\mathcal{P}),\ \forall\, B \in \mathcal{P}\big\}$ est un λ-système qui contient $\mathcal{P}$ – elle contient donc $\mathcal{L}(\mathcal{P})$; c'est élémentaire. La famille $\big\{B'; A' \cap B' \in \mathcal{L}(\mathcal{P}),\ \forall\, A' \in \mathcal{L}(\mathcal{P})\big\}$ contient donc $\mathcal{P}$, et forme un λ-système. Elle contient donc $\mathcal{L}(\mathcal{P})$, ce qui est la conclusion recherchée. $\triangleright$

L'énoncé précédent est dû à Dynkin. On utilisera dans notre démonstration du théorème de représentation de Riesz la version 'fonctionnelle' suivante du lemme de Dynkin.

Théorème 2. *Soit H un espace vectoriel de fonctions bornées définies sur Ω, contenant la fonction constante $\mathbf{1}$, et stable par convergence croissante bornée. Si H contient une famille G stable par multiplication alors H contient l'ensemble des fonctions bornées mesurables par rapport à la tribu engendrée par G.*

Démonstration – On montre sans mal que la famille $\mathcal{L} := \{A\,;\, \mathbf{1}_A \in H\}$ est un λ-système en conséquences de hypothèses faites sur H. Notons $\mathcal{P}$ la famille des parties de Ω de la forme $\big\{f_1 \in I_1, \dots, f_n \in I_n\big\}$, pour $n \geqslant 1$, des fonctions réelles $f_i \in G$, et I_i des intervalles de $\mathbb{R}$. C'est un π-système. Étant donné un élément de $\mathcal{P}$, désignons par c un majorant commun à $f_1, \dots, f_n$, et notons avec Weierstrass $(P^i_k)_{k \geqslant 0}$ une suite de polynômes convergeant ponctuellement et de facon bornée sur l'intervalle $[-c, c]$ vers l'indicatrice $\mathbf{1}_{I_i}$, pour chaque $1 \leqslant i \leqslant n$. Les f_i étant dans la famille 'multiplicative' G, on définit un élément de G, donc de H, en posant $g_k(\omega) := P^1_k(f_1(\omega)) \cdots P^n_k(f_n(\omega))$. Ces dernières convergent ponctuellement de facon croissante et bornée vers l'indicatrice de l'ensemble $\big\{f_1 \in I_1, \dots, f_n \in I_n\big\}$, qui est donc un élément de $\mathcal{L}$ puisque la famille H est stable par convergence bornée. Le lemme de Dynkin nous assure alors que $\sigma(\mathcal{P}) \subset \mathcal{L}$. On a par définition $\sigma(G) := \sigma(\mathcal{P})$. On conclut en rappelant que toute fonction f bornée mesurable par rapport à $\sigma(\mathcal{P})$ est limite croissante bornée des fonctions $\sum_k k2^{-n}\,\mathbf{1}_{k2^{-n}<f\leqslant(k+1)2^{-n}}$, lorsque n croit indéfiniment, en utilisant à nouveau que la famille H est stable par convergence bornée. $\triangleright$

2 – Mesure de Lebesgue

Je rappelle qu'une fonction additive réelle μ sur une algèbre de parties de Ω est une fonction μ sur $\mathcal{A}$ telle que $\mu(A\cup B)=\mu(A)+\mu(B)$ lorsque A et B sont des éléments disjoints de l'algèbre.

Théorème 3. *Soit $(\Omega,\mathcal{F})$ un espace mesurable, $\mathcal{A}\subset\mathcal{F}$ une algèbre et $\mu:\mathcal{A}\to[0,1]$ une fonction additive telle que $\mu(\varnothing)=0$ et $\mu(\Omega)=1$, et pour toute suite $(A_n)_{n\geqslant 0}$ d'éléments de $\mathcal{A}$ décroissant vers l'ensemble vide on a $\mu(A_n)\to 0$. Alors μ s'étend de facon unique en une probabilité sur $\sigma(\mathcal{A})$.*

Cet énoncé est dû à Caratheodory. Notez que la propriété satisfaite par μ équivaut à avoir $\mu(\cup_{n\geqslant 0}A_n)=\sum_{n\geqslant 0}\mu(A_n)$, pour toute suite d'éléments A_n de $\mathcal{A}$ disjoints deux à deux, de réunion $\cup_{n\geqslant 0}A_n$ un élément de $\mathcal{A}$. On dit que μ est σ-additive sur $\mathcal{A}$. Je rappelle qu'une pseudo-métrique vérifie tous les axiomes d'une métrique hormis le fait que deux éléments à distance nulle coïncident. Je rappelle aussi la notation $B\Delta C:=(B\cup C)\backslash(B\cap C)$.

Démonstration – Unicité. La collection des éléments de $\sigma(\mathcal{A})$ sur lesquels deux extensions possibles coïncident étant un λ-système, deux mesures sont égales sur $\sigma(\mathcal{A})$ si elles coïncident sur $\mathcal{A}$ d'après le lemme de Dynkin, lemme 1 – une algèbre est en particulier un π-système.

Existence. Notons $\mathfrak{P}(\Omega)$ la famille des sous ensembles de Ω. On définit sur $\mathfrak{P}(\Omega)$ la mesure extérieure $\overline{\mu}$ associée à μ par la formule

$$\overline{\mu}(B):=\inf\left\{\sum_{n\geqslant 0}\mu(A_n)\,;\,B\subset\bigcup_{n\geqslant 0}A_n,\ \ A_n\in\mathcal{A}\right\},\quad \forall\,B\in\mathfrak{P}(\Omega).$$

Replacant les A_n par $A_n\backslash\bigcup_{k=0}^{n}A_k$ si nécessaire, on peut toujours supposer les A_n disjoints dans cette définition. On voit que la fonction $\overline{\mu}$ est croissante et telle que $\overline{\mu}\big(\bigcup_{n\geqslant 0}B_n\big)\leqslant\sum_{n\geqslant 0}\overline{\mu}(B_n)$, pour toute suite $(B_n)_{n\geqslant 0}$ d'ensembles de Ω – on dit que $\overline{\mu}$ est σ-sous additive. Pour $A\in\mathcal{A}$, l'infimum définissant $\overline{\mu}(A)$ coïncide avec l'infimum portant sur des partitions $(A_n)_{n\geqslant 0}$ de A. Il s'ensuit que $\mu(A)\leqslant\overline{\mu}(A)$, et donc $\mu(A)=\overline{\mu}(A)$, pour $A\in\mathcal{A}$, puisque μ est σ-additive sur $\mathcal{A}$. On vérifie qu'on définit une pseudo-métrique sur l'ensemble $\mathfrak{P}(\Omega)$ de toutes les parties de Ω en posant

$$d(B,C)=\overline{\mu}(B\Delta C).$$

Puisque $\overline{\mu}$ est sous additive et $B\subset\big(B\Delta C\big)\cup C$ pour tous sous ensembles B,C de Ω, on a

$$\big|\overline{\mu}(B)-\overline{\mu}(C)\big|\leqslant\overline{\mu}(B\Delta C)=d(B,C),$$

si bien que $\overline{\mu}(B)=\overline{\mu}(C)$ si $d(B,C)=0$. On définit $\mathcal{A}^{\overline{\mu}}$ comme la famille des sous ensembles de Ω qui peuvent être approché arbitrairement bien par des éléments of $\mathcal{A}$ pour la pseudo-distance d. En d'autres termes, $\mathcal{A}^{\overline{\mu}}$ est le complété de $\mathcal{A}$ pour la pseudo-distance d. On a bien sûr $\mathcal{A}\subset\mathcal{A}^{\overline{\mu}}$. L'application 1-Lipschitz $\overline{\mu}$ admet donc un prolongement unique à $\mathcal{A}^{\overline{\mu}}$.

Lemme 4. *La famille $\mathcal{A}^{\overline{\mu}}$ est une σ-algèbre sur laquelle $\overline{\mu}$ est additive.*

Si l'on admet un instant ce résultat, on a donc $\sigma(\mathcal{A})\subset\mathcal{A}^{\overline{\mu}}$, et on dispose sur $\mathcal{A}^{\overline{\mu}}$ d'une application $\overline{\mu}$ croissante, additive et σ-sous additive. Elle vérifie donc pour tout réunion disjointe $(A_n)_{n\geqslant 0}$ d'éléments de $\mathcal{A}^{\overline{\mu}}$ et tout $N\geqslant 0$ l'inégalité $\sum_{n=0}^{N}\overline{\mu}(A_n)\leqslant\overline{\mu}(\cup_{n\geqslant 0}A_n)\leqslant\sum_{n\geqslant 0}\overline{\mu}(A_n)$. On en déduit que $\overline{\mu}$ est σ-additive sur $\mathcal{A}^{\overline{\mu}}$. Sa restriction à $\sigma(\mathcal{A})$ est une probabilité étendant μ. On démontre maintenant le lemme.

• On commence par démontrer le caractère additif de $\overline{\mu}$ sur $\mathcal{A}^{\overline{\mu}}$. Prenons deux éléments disjoints B et C de $\mathcal{A}^{\overline{\mu}}$, une constante $\varepsilon>0$, et notons A_B et A_C des éléments de $\mathcal{A}$ tels que $d\big(B,A_B\big)$, $d\big(C,A_C\big)\leqslant\varepsilon$. Comme ils satisfont l'inégalité $\overline{\mu}(A_B\cap A_C)\leqslant 2\varepsilon$, la sous additivité de $\overline{\mu}$ nous donne

$$\max\Big(d\big(B,A_B\backslash(A_B\cap A_C)\big),\ d\big(C,A_C\backslash(A_B\cap A_C)\big)\Big)\leqslant 3\varepsilon.$$

Il s'ensuit que

$$\begin{aligned}\overline{\mu}(B)+\overline{\mu}(C) \geqslant \overline{\mu}(B\cup C) &\geqslant \overline{\mu}\big(A\backslash(A_B\cap A_C)\,\cup\,A\backslash(A_B\cap A_C)\big)-3\varepsilon\\ &\geqslant \overline{\mu}(B)+\overline{\mu}(C)-5\varepsilon,\end{aligned}$$

ce dont on tire la conclusion puisque $\varepsilon>0$ est arbitraire.

• Il est clair que $\mathcal{A}^{\overline{\mu}}$ est stable par complémentation. On vérifie que $\mathcal{A}^{\overline{\mu}}$ est stable par réunion dénombrable disjointe ; cela implique que $\mathcal{A}^{\overline{\mu}}$ est stable par réunion ou intersection dénombrable, et qu'il s'agit donc d'une σ-algèbre. Étant donné $\varepsilon>0$ et une suite $(B_n)_{n\geqslant 0}$ d'éléments de $\mathcal{A}^{\overline{\mu}}$ disjoints deux à deux, associons à chaque B_n un $A_n\in\mathcal{A}$ tel que $d(A_n,B_n)\leqslant 2^{-n-1}\varepsilon$. Comme $\overline{\mu}$ est additive sur $\mathcal{A}^{\overline{\mu}}$ on a $\sum_{n=0}^N\overline{\mu}(B_n)=\overline{\mu}\Big(\bigcup_{n=0}^N B_n\Big)\leqslant 1$, pour tout $N\geqslant 0$, si bien que $\sum_{n\geqslant N+1}\overline{\mu}(B_n)\leqslant\varepsilon$, pour N assez grand. Pour un tel choix de N on a

$$\begin{aligned}d\Big(\bigcup_{n\geqslant 0}B_n,\bigcup_{n=0..N}A_n\Big) &\leqslant d\Big(\bigcup_{n\geqslant 0}B_n,\bigcup_{n=0..N}B_n\Big)+d\Big(\bigcup_{n=0..N}B_n,\bigcup_{n=0..N}A_n\Big)\\ &\leqslant \overline{\mu}\Big(\bigcup_{n\geqslant N+1}B_n\Big)+\overline{\mu}\Big(\Big\{\bigcup_{n=0..N}B_n\Big\}\Delta\Big\{\bigcup_{n=0..N}A_n\Big\}\Big)\\ &\leqslant \varepsilon+\overline{\mu}\Big(\bigcup_{n=0..N}(B_n\Delta A_n)\Big)\leqslant\varepsilon+\sum_{n=0}^N 2^{-n-1}\varepsilon\leqslant 2\varepsilon.\end{aligned}$$

L'arbitraire sur $\varepsilon>0$ montre que $\bigcup_{n\geqslant 0}B_n$ est un élément de $\mathcal{A}^{\overline{\mu}}$. ▷

À titre d'exemple, regardons l'espace $\Omega=\{0,1\}^{\mathbb{N}}$ et son algèbre $\mathcal{A}$ des parties 'cylindriques' $\big\{\omega=(\omega_n)_{n\geqslant 0};\omega_i=a_i,\,\forall\,i\in I\big\}$, où I est un sous ensemble fini de $\mathbb{N}$ et les $a_i\in\{0,1\}$ fixés. Cette algèbre est dénombrable. Chaque partie cylindrique est à la fois ouverte et fermée, donc compacte, pour la topologie produit. Une intersection décroissante de parties cylindrique d'intersection vide est donc vide à partir d'un certain rang, si bien qu'une fonction additive définie sur $\mathcal{A}$ vérifie automatiquement la condition $\mu(A_n)\to 0$, du théorème d'extension. La σ-algèbre engendrée par les ouverts élémentaires donnés par les parties cylindriques s'appelle la tribu borelienne.

Théorème 5. *Toute fonction additive sur $\mathcal{A}$ définit une unique probabilité sur la tribu borelienne de $\{0,1\}^{\mathbb{N}}$.*

On définit une fonction additive sur $\mathcal{A}$ en posant

$$\mu(A):=2^{-|I|},\quad\text{pour}\quad A=\big\{\omega=(\omega_n)_{n\geqslant 0};\omega_i=a_i,\,\forall\,i\in I\big\}$$

On note $\mu^{\otimes\mathbb{N}}$ la probabilité associée définie sur la tribu borelienne de $\{0,1\}^{\mathbb{N}}$. La mesure image de $\mu^{\otimes\mathbb{N}}$ par l'application $(\omega_n)_{n\geqslant 0}\to\sum\omega_n 2^{-n-1}\in[0,1]$, est la mesure de Lebsegue sur l'intervalle $[0,1]$. (Elle attribue à chaque segment dyadique la mesure attendue, et ces derniers forment une algèbre engendrant la tribu borelienne de $[0,1]$, engendrée en permier lieu par les ouverts de $[0,1]$.) Il est élémentaire de construire à partir de là la mesure de Lebesgue sur $\mathbb{R}$ et $\mathbb{R}^d$.

3 – Représentation de Riesz

Notons $\mathcal{C}_c(\mathbb{R}^d)$ l'ensemble des fonctions réelles continues à support compact, définies sur $\mathbb{R}^d$, muni de la norme du supremum. Notez que la σ-algèbre sur $\mathbb{R}^d$ engendrée par $\mathcal{C}_c(\mathbb{R}^d)$ coïncide avec la tribu borelienne de $\mathbb{R}^d$.

Théorème 6. *Soit $\mathbb{E}:\mathcal{C}_c(\mathbb{R}^d)\to\mathbb{R}$ une forme linéaire positive de norme 1. Il existe une unique probabilité $\mathbb{P}$ sur la tribu borelienne de $\mathbb{R}^d$ telle que $\mathbb{E}(f)=\int f(x)\,\mathbb{P}(dx)$, pour toute $f\in\mathcal{C}_c(\mathbb{R}^d)$.*

Démonstration – On vérifie d'abord que des conditions analogues à celles du théorème de Caratheodory sont satisfaites ici. On a bien $\mathbb{E}(0)=0$, et on étend $\mathbb{E}$ à l'ensemble $\underline{\mathbb{R}}$

des fonctions constantes en posant $\mathbb{E}(\mathbf{1}) = 1$. Cette extension à $\mathcal{C}_c(\mathbb{R}^d) \oplus \underline{\mathbb{R}}$ est toujours linéaire positive et de norme 1. La σ-additivité de $\mathbb{E}$ sur $\mathcal{C}_c(\mathbb{R}^d)$ est automatique ! Toute suite décroissante de $\mathcal{C}_c(\mathbb{R}^d)$ convergeant vers 0 ponctuellement vers 0 converge en fait uniformément - c'est un énoncé élémentaire dû à Dini. Comme $\mathbb{E}$ est positive et bornée en tant qu'opérateur, il s'ensuit que les $\mathbb{E}(f_n)$ décroissent vers 0. La σ-additivité de $\mathbb{E}$ sur $\mathcal{C}_c(\mathbb{R}^d) \oplus \underline{\mathbb{R}}$ s'ensuit.
On peut alors copier mot à mot la démonstration du théorème d'extension de Caratheodory, avec $\mathcal{C}_c^+(\mathbb{R}^d) \cup \underline{\mathbb{R}}$ dans le rôle de l'algèbre $\mathcal{A}$, l'ensemble des fonctions réelles positives sur $\mathbb{R}^d$ dans le rôle de $\mathfrak{P}(\Omega)$, et l'opération $f\Delta g := f + g - 2f \wedge g$ dans le rôle de $A\Delta B$. On définit pour $f \geqslant 0$

$$\overline{\mathbb{E}}(f) := \inf \left\{ \sum_{n\geqslant 0} \mathbb{E}(f_n)\,;\; f \leqslant \sum_{n\geqslant 0} f_n,\; f_n \in \mathcal{C}_c(\mathbb{R}^d) \right\}.$$

On étend $\overline{\mathbb{E}}$ à l'ensemble des fonctions bornées en décomposant $f = f_+ - f_-$, avec $f_+ = f \vee 0$ et $f_- = f \wedge 0$, et en posant $\overline{\mathbb{E}}(f) := \overline{\mathbb{E}}(f_+) - \overline{\mathbb{E}}(f_-)$. On vérifie que $|\overline{\mathbb{E}}(f)| \leqslant \sup |f|$, et que $\overline{\mu} = \mu$ sur $\mathcal{C}_c(\mathbb{R}^d) \oplus \underline{\mathbb{R}}$. La pseudo-distance d est définie par $d(f,g) := \overline{\mathbb{E}}(f\Delta g)$, pour f, g réelles bornées. L'analogue $\mathcal{B}^{\overline{\mu}}$ de la classe $\mathcal{A}^{\overline{\mu}}$ est alors défini comme la famille des fonctions bornées de $\mathbb{R}^d$ qui peuvent être approché arbitrairement bien par des éléments de $\mathcal{C}_c(\mathbb{R}^d) \oplus \underline{\mathbb{R}}$ pour la pseudo-distance d ; c'est un espace vectoriel qui contient $\mathcal{C}_c(\mathbb{R}^d) \oplus \underline{\mathbb{R}}$. Le lemme 4 devient alors l'assertion selon laquelle $\mathcal{B}^{\overline{\mu}}$ est un espace vectoriel de fonctions bornées sur $\mathbb{R}^d$, contenant la constante $\mathbf{1}$, et stable par convergence croissante bornée, sur lequel on dispose d'une extension de $\mathbb{E}$ qui est additive et donnée par $\overline{\mathbb{E}}$. C'est donc une application linéaire sur $\mathcal{B}^{\overline{\mu}}$, de norme 1 pour la norme sup. La famille $G = \big(\mathcal{C}_c(\mathbb{R}^d) \oplus \underline{\mathbb{R}}\big) \subset \mathcal{B}^{\overline{\mu}}$ étant stable par multiplication, le théorème 2 nous assure que $\mathcal{B}^{\overline{\mu}}$ contient la classe des fonctions boreliennes bornées, engendrée par G. Si $\mathbb{E}'$ désigne une autre extension de $\mathbb{E}$ en une forme linéaire sur de norme 1 définie sur l'espace des applications boreliennes bornées, il est élémentaire de voir que la classe H des fonctions boreliennes bornées où $\overline{\mathbb{E}}$ et $\mathbb{E}'$ coïncident est un espace vectoriel contenant la constante $\mathbf{1}$, stable par convergence bornée. Le théorème 2 s'applique donc et montre que les deux extensions coïncident sur l'ensemble des fonctions boreliennes bornées. La restriction $\mathbb{P}$ de $\overline{\mathbb{E}}$ aux indicatrices d'ensembles boreliens est une probabilité. (Notez que la σ-additivité de $\mathbb{P}$ est conséquence du fait que $\overline{\mathbb{E}}$ est un opérateur linéaire borné.) On a par construction $\overline{\mathbb{E}}(f) = \int f(x)\, \mathbb{P}(dx)$ pour les fonctions élémentaires $f = \sum a_i \mathbf{1}_{A_i}$, et puisque toute fonction $f \in \mathcal{C}_c(\mathbb{R}^d)$ est limite uniforme de fonctions élémentaires et que $\overline{\mathbb{E}}$ est continue pour la norme du supremum, l'identité a lieu par continuité sur tout $\mathcal{C}_c(\mathbb{R}^d)$. $\triangleright$