

Borel versus Lebesgue: la complétion des mesures

Résumé

Entre intégrale de Lebesgue des fonctions boréliennes et mesure de Borel sur la tribu de Lebesgue, il est permis de se sentir un peu perdu dans le vocabulaire de la théorie de la mesure et de l'intégration sur \mathbb{R} . A quoi s'ajoute l'absence de consensus quand il s'agit de nommer les choses, ce qui fait que votre professeur du second semestre ne semble pas nécessairement parler la même langue que celui du premier ! Ce document vise à proposer une clarification.

Borel, boréliens, borélitude

Rappelons quelques définitions :

Définition. • La tribu de Borel, notée $\mathcal{B}(\mathbb{R})$ est la plus petite tribu sur \mathbb{R} qui contient les ouverts de \mathbb{R} . Ses éléments sont appelés (ensembles) boréliens.

- Plus généralement, si X est un ensemble muni d'une topologie (c'est-à-dire, si X a des ouverts et des fermés), on appelle tribu borélienne sur X la tribu engendrée par les ouverts de X , et on la note $\mathcal{B}(X)$. X est typiquement un espace vectoriel normé (généralement \mathbb{R}^n), ou une partie d'un espace vectoriel normé avec la topologie induite (par exemple un intervalle de \mathbb{R}).
- Si (X, \mathcal{T}_X) et (Y, \mathcal{T}_Y) sont deux espaces mesurables, une fonction $f : X \rightarrow Y$ est mesurable ssi l'image réciproque de tout ensemble mesurable de Y est mesurable dans X , autrement dit ssi

$$\forall B \in \mathcal{T}_Y, f^{-1}(B) \in \mathcal{T}_X$$

- Dans le cas particulier où X et Y ont une topologie, et où $\mathcal{T}_X = \mathcal{B}(X)$, $\mathcal{T}_Y = \mathcal{B}(Y)$, on dit plutôt que f est borélienne.

Remarque : Dans la plupart des cas qui nous intéressent, l'espace d'arrivée des fonctions mesurables est $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Puisque les fonctions boréliennes sont des cas particuliers de fonctions mesurables, on appelle parfois les fonctions boréliennes "mesurables", d'où un premier facteur de confusion.

Pourquoi introduire cette tribu ? Rappelons que, pour généraliser l'intégrale de Riemann afin d'avoir une théorie de l'intégration plus maniable, on commence par tenter de généraliser la notion de longueur d'intervalle. On souhaiterait pouvoir mesurer tout sous-ensemble $A \subset \mathbb{R}$. Une première tentative consiste à approcher A par une famille d'intervalles $(I_k)_k$ telle que $A \subset \bigcup_k I_k$, et à choisir des familles d'intervalles qui se "resserrent" autour de A , l'approchant de plus en plus finement. Cela donne la *mesure extérieure* de A :

$$|A| = \inf \left\{ \sum_k \ell(I_k), A \subset \bigcup_k I_k \right\}$$

où ℓ désigne la longueur des intervalles.

Mais là, on tombe sur un os : la mesure extérieure n'est pas additive. Autrement dit, il existe des sous-ensembles de \mathbb{R} A et B disjoints tels que $|A \cup B| \neq |A| + |B|$. Ce semble problématique, pour une mesure d'ensembles. Et il y a pire : on peut en fait montrer le résultat suivant :

Théorème. Il n'existe aucune application $\mu : \mathcal{P}(\mathbb{R}) \rightarrow [0, +\infty]$ telle que

1. $\mu(I) = \ell(I)$ pour tout intervalle ouvert I ;
2. Pour toute famille dénombrable d'ensembles disjoints $(A_k)_{k \in \mathbb{N}}$, $\mu(\bigcup_k A_k) = \sum_k \mu(A_k)$;
3. Pour tout $A \in \mathcal{P}(\mathbb{R})$, pour tout $t \in \mathbb{R}$, $\mu(A + t) = \mu(A)$.

Autrement dit, toute tentative de généraliser la longueur d'intervalles à tout sous-ensemble de \mathbb{R} est irrémédiablement vouée à l'échec. Le problème vient du fait que des parties de \mathbb{R} , il y en a vraiment beaucoup. Il y en a *monstrueusement* beaucoup, et certaines sont complètement biscornues. Il nous faut donc réduire nos prétentions à des ensembles pas trop sauvages, mais on ne veut pas non plus *trop* se restreindre. On veut au moins

- * savoir mesurer l'ensemble vide ;
- * savoir mesurer $\bigcup A_n$ si on sait mesurer chaque élément d'une suite $(A_n)_n$ (sinon, inutile d'essayer de récupérer le point 2. du théorème ci-dessus!)
- * savoir mesurer ${}^c A$ si on sait mesurer A (ici, on pense aux probabilités : il serait ridicule de savoir calculer la probabilité d'un événement, mais pas la probabilité qu'il ne se produise pas)

C'est ce qui nous donne la notion de tribu. $\mathcal{P}(\mathbb{R})$ est une tribu, dont on a dit qu'elle était trop grosse ; $\{\emptyset, \mathbb{R}\}$ est aussi une tribu... mais elle est ridiculement pauvre. Entre les deux, il y a $\mathcal{B}(\mathbb{R})$: on peut montrer que c'est la plus petite tribu qui contient tous les intervalles. Comme on cherche à étendre la longueur d'intervalles, ça semble prometteur...et ça marche :

Théorème. *Il existe une unique mesure μ sur la tribu $\mathcal{B}(\mathbb{R})$ telle que pour tout $a < b$, $\mu(]a, b[) = b - a$. On l'appelle mesure de Borel sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*

Quelques précisions :

Existence : Pour tout borélien $A \in \mathcal{B}(\mathbb{R})$, $\mu(A) = |A|$. Autrement dit, la mesure de Borel n'est autre que la mesure extérieure, restreinte aux boréliens. La preuve de la partie "existence" du théorème consiste donc à montrer que la mesure extérieure est bien une mesure sur les boréliens. Il s'agit de vérifier l'additivité, ce que l'on fait en 4 étapes :

1. On montre que si $A \in \mathcal{P}(\mathbb{R})$ et O ouvert de \mathbb{R} sont disjoints, alors $|A \cup O| = |A| + |O|$;
2. On en déduit que si $A \in \mathcal{P}(\mathbb{R})$ et F fermé de \mathbb{R} sont disjoints, alors $|A \cup F| = |A| + |F|$;
3. On montre qu'on peut "approximer les boréliens par des fermés" : si $B \in \mathcal{B}(\mathbb{R})$ alors pour tout $\varepsilon > 0$, il existe $F \subset B$ fermé tel que $|B \setminus F| < \varepsilon$.
4. Enfin, on en déduit que si $A \in \mathcal{P}(\mathbb{R})$ et B borélien de \mathbb{R} sont disjoints, alors $|A \cup B| = |A| + |B|$.

De là, on montre assez facilement qu'on a en fait l'additivité dénombrable : si $(B_n)_n$ est une suite de boréliens disjoints, alors $|\cup B_n| = \sum |B_n|$. Ce qui montre que la mesure extérieure est bien une mesure sur les boréliens.

Unicité : On montre que deux mesures sur \mathbb{R} qui coïncident sur les intervalles $]a, b[$ doivent en fait être égales. On utilise le résultat suivant :

Théorème. *Soit (X, \mathcal{T}) un espace mesurable et μ_1, μ_2 deux mesures. On suppose que $X = \bigcup_{n \geq 1} X_n$ avec $\mu_1(X_n) = \mu_2(X_n) < +\infty$ (autrement dit, les mesures sont σ -finies). Si μ_1 et μ_2 sont égales sur une famille \mathcal{F} de sous-ensembles mesurables stable par intersection et qui contient contenant les ensembles X_n . Alors $\mu_1 = \mu_2$ sur la tribu $\sigma(\mathcal{F})$ engendrée par \mathcal{F} .*

Ici, on prend donc $\mathcal{F} = \{]a, b[, a < b\}$ et $X_n =]-n, n[$. Alors \mathcal{F} vérifie les hypothèses du théorème et $\sigma(\mathcal{F}) = \mathcal{B}(\mathbb{R})$.

Complétion des tribus et des mesures

On a donc résolu notre problème initial de mesure des parties de \mathbb{R} , du moins tant qu'on se restreint aux boréliens. Mais peut-on faire encore mieux ?

Il se trouve que oui : pour une mesure donnée, on va pouvoir définir une tribu maximale sur laquelle on peut raisonnablement l'étendre.

L'idée est la suivante. Certains ensembles sont trop petits pour être "vus" par la mesure de Borel. Ainsi, pour tout $x \in \mathbb{R}$, $\mu(\{x\}) = \mu([x, x]) = x - x = 0$. Avec la propriété d'additivité des

mesures, on montre ainsi que tout ensemble dénombrable est borélien, de mesure de Borel nulle. Il y a aussi des ensembles boréliens non dénombrables mais néanmoins de mesure nulle (par exemple, l'ensemble de Cantor). Maintenant, étant donné que $A \subset B \Rightarrow \mu(A) \leq \mu(B)$, si B est de mesure nulle et $A \subset B$, il semble raisonnable de poser $\mu(A) = 0$, même si A n'est pas un borélien. Ceci nous conduit à la définition suivante :

Définition. Soit (X, \mathcal{T}, μ) un espace mesuré.

1. On dit que $A \subset X$ est négligeable s'il existe $\tilde{A} \in \mathcal{T}$ tel que $A \subset \tilde{A}$ et $\mu(\tilde{A}) = 0$.
2. On dit que la mesure μ est complète si \mathcal{T} contient tous les ensembles négligeables.

Or, la mesure de Borel μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ n'est pas complète : on peut montrer que l'ensemble de Cantor contient des sous-ensembles non boréliens. On peut donc étendre la mesure de Lebesgue aux ensembles négligeables en posant $\mu(N) = 0$ pour tout N négligeable. Mais on ne s'arrête pas là : on remarque que de même que μ ne voit pas les ensembles trop petits, elle ne fait pas la différence entre des ensembles qui diffèrent par un ensemble de mesure nulle. Par exemple $\mu(]a, b[) = \mu([a, b])$: ajouter $\{a\}$ ne change pas la mesure. On pourrait donc se dire que si μ ne "fait pas la différence" entre deux boréliens A et B , alors tout sous-ensemble "coincé entre A et B " devrait avoir la même mesure que A (et B). C'est l'idée de complétion des mesures :

Théorème. Soit (X, \mathcal{T}, μ) un espace mesuré. On définit

$$\begin{aligned} \mathcal{T}^* &= \{E \subset X \mid \exists A, B \in \mathcal{T}, A \subset E \subset B \text{ et } \mu(B \setminus A) = 0\} \\ &= \{E \subset X \mid \exists A \in \mathcal{T} \text{ et } N \text{ négligeable tq } E = A \cup N\} \end{aligned}$$

On étend μ à \mathcal{T}^* en posant $\mu^*(E) = \mu(A)$. Alors \mathcal{T}^* est une tribu sur X , appelée tribu complétée de \mathcal{T} , et μ^* est une mesure complète sur (X, \mathcal{T}^*) appelée mesure complétée.

Preuve. • μ^* est bien définie : soit $E \in \mathcal{T}^*$, et supposons que A_1, A_2, B_1, B_2 soient 4 ensembles mesurables tels que

$$\begin{cases} A_1 \subset E \subset B_1, \mu(B_1 \setminus A_1) = 0 \\ A_2 \subset E \subset B_2, \mu(B_2 \setminus A_2) = 0. \end{cases}$$

Alors $A_1 \subset B_2$ donc $\mu(A_1) \leq \mu(B_2) = \mu(A_2)$. Réciproquement $A_2 \subset B_1$ donc $\mu(A_2) \leq \mu(B_1) = \mu(A_1)$. Donc $\mu(A_1) = \mu(A_2)$, et la mesure de E ne dépend pas du choix de A et B .

• \mathcal{T}^* est une tribu sur X :

1. $X \in \mathcal{T}^*$: $X \subset X \subset X$, $X \in \mathcal{T}$ et $\mu(X \setminus X) = \mu(\emptyset) = 0$.
2. Soit $E \in \mathcal{T}^*$, alors il existe A et B tels que $A \subset E \subset B$ et $\mu(B \setminus A) = 0$. Mais alors $B^c \subset E^c \subset A^c$ et $\mu(A^c \setminus B^c) = \mu(B \setminus A) = 0^1$. Donc $E^c \in \mathcal{T}^*$.
3. Si (E_n) est une suite d'éléments de \mathcal{T}^* , et $A_n, B_n \in \mathcal{T}$ tels que $A_n \subset E_n \subset B_n$ pour tout n , notons $E = \bigcup_n E_n$, $A = \bigcup_n A_n$ et $B = \bigcup_n B_n$. Montrons que $E \in \mathcal{T}^*$. On a

$$A \subset E \subset B \text{ et } (B \setminus A) \subset \bigcup_n (B_n \setminus A_n).$$

Donc $\mu(B \setminus A) \leq \sum \mu(B_n \setminus A_n) = 0$. Donc $E \in \mathcal{T}^*$.

• μ^* est une mesure sur (X, \mathcal{T}^*) : $\mu^*(\emptyset) = \mu(\emptyset) = 0$ et si (E_n) est une suite disjointe de \mathcal{T}^* , alors les A_n correspondants sont aussi disjoints, donc

$$\mu^*(\bigcup E_n) = \mu(\bigcup A_n) = \sum \mu(A_n) = \sum \mu^*(E_n). \quad \square$$

En peut donc compléter $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$, où μ est la mesure de Borel :

Définition. On appelle tribu de Lebesgue la tribu qui complète $\mathcal{B}(\mathbb{R})$ pour μ , et on la note $\mathcal{L}(\mathbb{R})$. On note $\lambda = \mu^*$ la mesure complétée de la mesure de Borel, et on l'appelle mesure de Lebesgue.

La mesure de Lebesgue est donc égale à celle de Borel sur tous les boréliens, et peut mesurer des sous-ensembles de \mathbb{R} supplémentaires. Comme dans la pratique, on a généralement affaire à des boréliens, on utilise (relativement) indifféremment les noms "mesure de Borel" et "mesure de Lebesgue" pour désigner la mesure sur \mathbb{R} qui généralise la longueur d'intervalles.

1. car $A \setminus B = A \cap B^c = (A^c)^c \cap B^c = B^c \setminus A^c$

Intégrale de Lebesgue et Lebesgue-intégrabilité

Enfin, on a vu qu'à partir de toute mesure μ sur un ensemble mesurable (A, \mathcal{T}) , on peut définir une intégrale :

1. D'abord on définit l'intégrale supérieure d'une fonction mesurable *positive* f en posant

$$\int_X^* f d\mu = \sup \left\{ \sum_i \mu(A_i) \inf_{x \in A_i} f(x), (A_i)_i \text{ partition mesurable finie de } X \right\} \in [0, \infty]$$

2. Quand l'intégrale supérieure de f est finie, on dit que f est intégrable. Dans ce cas, son intégrale supérieure est appelée intégrale tout court, et notée $\int_X f d\mu$.
3. Pour f mesurable à valeurs dans \mathbb{R} (pas forcément positive), on dit que f est intégrable si $|f|$ l'est. Or $|f| = f_+ + f_-$, avec $f_+ = \max(f, 0)$ et $f_- = \max(-f, 0)$. Donc ceci implique que f_+ et f_- sont intégrables aussi (remarquons que ce sont des fonctions mesurables positives), et vérifient $f = f_+ - f_-$. On définit donc l'intégrale de f par

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

L'intégrale obtenue sur les fonctions $f : \mathbb{R} \rightarrow \mathbb{R}$ en appliquant ce procédé avec la mesure de Lebesgue est appelée *intégrale de Lebesgue*. Une fonction f intégrable par rapport à λ est dite Lebesgue-intégrable.