Pré-requis de statistique : Essentiels de théorie de la mesure

Notes de cours, Master Maths & IA

Guillermo Durand

15 septembre 2025

Tribu ou σ-algèbre sur Ω : $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ telle que $\mathcal{A} \neq \varnothing$, $\mathcal{A}$ stable par complémentaire et par union dénombrable.

$(\Omega, \mathcal{A})$ s'appelle un espace mesurable, et $A \in \mathcal{A}$ une partie mesurable ou un ensemble mesurable.

Propriétés : $\varnothing \in \mathcal{A}$, $\Omega \in \mathcal{A}$, $\mathcal{A}$ stable par intersection dénombrable (Exercice).

Tribu engendrée par $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ noté $\sigma(\mathcal{C})$: plus petite tribu qui contient $\mathcal{C}$, intersection de toutes les tribus qui contiennent $\mathcal{C}$.

Exemples : $\mathcal{P}(\Omega)$, $\{\varnothing, \Omega\}$, $\{\varnothing, A, A^c, \Omega\} = \sigma(\{A\})$, $\mathcal{B}(\mathbb{R}^d)$ la tribu borélienne de $\mathbb{R}^d$ engendrée par les ouverts (tout espace topologique a une tribu borélienne engendrée par les ouverts).

Tribu produit de $\mathcal{A}_1$ et $\mathcal{A}_2$ notée $\mathcal{A}_1 \otimes \mathcal{A}_2$ engendrée par les $A_1 \times A_2$, $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$. **Remarque :** $\mathcal{B}(\mathbb{R}^d) = \bigotimes_{i=1}^d \mathcal{B}(\mathbb{R})$ (ça se démontre).

Soit $f : \Omega_1 \to (\Omega_2, \mathcal{A}_2)$, l'ensemble $\left\{f^{-1}(B), B \in \mathcal{A}_2\right\}$, noté $\sigma(f)$ ou $f^{-1}(\mathcal{A}_2)$, est une tribu, appelée tribu image réciproque de $\mathcal{A}_2$ (Exercice).

Soit $f : (\Omega_1, \mathcal{A}_1) \to \Omega_2$, $\{B \in \mathcal{P}(\Omega_2) : f^{-1}(B) \in \mathcal{A}_1\}$ est une tribu appelée tribu image de $\mathcal{A}_1$ (Exercice). **ATTENTION** ce n'est pas $f(\mathcal{A}_1) = \{f(A) : A \in \mathcal{A}_1\}$! ! (Contre-exemple : $f : \mathbb{R} \to \mathbb{R}$ constante, $f(\mathcal{A}_1)$ n'est même pas une tribu.)

Application de la tribu image : soit $f : \Omega_1 \to (\Omega_2, \sigma(\mathcal{C}))$, alors $\sigma(f) = \sigma(\{f^{-1}(C), C \in \mathcal{C}\})$, autrement dit $f^{-1}(\sigma(\mathcal{C})) = \sigma(f^{-1}(\mathcal{C}))$. En effet, $\sigma(f)$ est une tribu qui contient les $f^{-1}(C), C \in \mathcal{C}$, donc $\sigma(\{f^{-1}(C), C \in \mathcal{C}\}) \subset \sigma(f)$. Réciproquement soit $\mathcal{D}$ la tribu image de $\sigma(\{f^{-1}(C), C \in \mathcal{C}\})$, $\mathcal{C} \subset \mathcal{D}$ donc $\sigma(\mathcal{C}) \subset \mathcal{D}$, donc tout $f^{-1}(B)$, $B \in \sigma(\mathcal{C})$ est dans $\sigma(\{f^{-1}(C), C \in \mathcal{C}\})$.

La tribu engendrée par une famille d'applications $f_i : \Omega \to (\Omega_i, \mathcal{A}_i)$, $i \in I$, est $\sigma\left(\bigcup_{i\in I} \sigma(f_i)\right) = \sigma\left(\bigcup_{i\in I} f_i^{-1}(\mathcal{A}_i)\right)$.

$f : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$ est mesurable si $f^{-1}(\mathcal{A}_2) \subseteq \mathcal{A}_1$. Trivialement, $f : (\Omega_1, \sigma(f)) \to (\Omega_2, \mathcal{A}_2)$ et $f : (\Omega_1, \mathcal{P}(\Omega_1)) \to (\Omega_2, \mathcal{A}_2)$ sont toujours mesurables.

Mesure : fonction $\mu : \mathcal{A} \to \mathbb{R}_+ \cup \{\infty\}$ telle que $\mu(\varnothing) = 0$ et $\mu\left(\bigcup_{n\in\mathbb{N}} E_n\right) = \sum_{n\in\mathbb{N}} \mu(E_n)$ si les E_n sont deux à deux disjoints.

$(\Omega, \mathcal{A}, \mu)$ s'appelle un espace mesuré.

Propriétés : $\mu\left(\bigcup_{1\le n\le N} E_n\right) = \sum_{1\le n\le N} \mu(E_n)$ si les E_n sont deux à deux disjoints. Si $A \subseteq B$, $\mu(A) \le \mu(B)$, et si de plus $\mu(A) < \infty$, $\mu(B \setminus A) = \mu(B) - \mu(A)$. $\mu\left(\bigcup_{n\in\mathbb{N}} E_n\right) \le \sum_{n\in\mathbb{N}} \mu(E_n)$ (borne d'union). $\mu\left(\bigcup_{n\in\mathbb{N}} E_n\right) = \lim_{n\to\infty} \mu(E_n)$ si les E_n sont croissants pour l'inclusion, $\mu\left(\bigcap_{n\in\mathbb{N}} E_n\right) = \lim_{n\to\infty} \mu(E_n)$ si les E_n sont décroissants pour l'inclusion et de mesure finie à partir d'un certain rang (Exercice).

μ est σ-finie si $\Omega = \bigcup_{n\in\mathbb{N}} E_n$ avec $E_n \in \mathcal{A}$, $\mu(E_n) < \infty$ pour tout n. μ est finie si $\mu(\Omega) < \infty$. μ est une mesure de probabilité si $\mu(\Omega) = 1$.

Si μ est de probabilité, on la note plutôt $\mathbb{P}$. On dit que $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace probabilisé, Ω est appelé un univers, et une partie mesurable $A \in \mathcal{A}$ est plutôt appelée un événement.

Exemples : mesure de comptage sur $(\Omega, \mathcal{P}(\Omega))$, σ-finie si et seulement si Ω est dénombrable (Exercice). Mesure de Dirac $\delta_y : A \mapsto \mathbb{1}_{\{y\in A\}}$, de proba. Combinaison dénombrable et positive de Dirac $\sum_{n\in\mathbb{N}} \alpha_n \delta_{y_n}$ toujours σ-finie (on suppose que la tribu contient les $\{y_n\}$), finie ou de proba selon la convergence de la série $\sum \alpha_n$. Mesure de Lebesgue sur les boréliens de $\mathbb{R}$ (resp. $\mathbb{R}^d$) : la seule mesure λ (resp. λ_d) invariante par translation et telle que $\lambda([0,1]) = 1$ (resp. $\lambda_d([0,1]^d) = 1$). Elle est σ-finie. Toutes les mesures de probabilité induites par les lois réelles usuelles, par exemple $\mathcal{N}(0,1) : A \mapsto \int \mathbb{1}_{\{x\in A\}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \mathrm{d}x$.

Exercice : pour une tribu qui contient tous les singletons, toute μ σ-finie ne peut charger au plus qu'un nombre dénombrable de points : $\{x : \mu(\{x\}) > 0\}$ est au plus dénombrable.

Mesure produit de μ_1 et μ_2 sur $\mathcal{A}_1 \otimes \mathcal{A}_2$: μ telle que $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$. Existe toujours, et est unique si μ_1 et μ_2 σ-finies (admis), dans ce dernier cas on la note $\mu_1 \otimes \mu_2$. Exemple : la mesure de Lebesgue sur $\mathbb{R}^d$: $\lambda_d = \lambda \otimes \cdots \otimes \lambda = \bigotimes_{i=1}^d \lambda = \lambda^{\otimes d}$.

Si $f : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$ est mesurable et μ est une mesure sur $(\Omega_1, \mathcal{A}_1)$, on note $f_{\#\mu}$ la mesure "push-forward", ou mesure induite, ou mesure image, définie sur $(\Omega_2, \mathcal{A}_2)$ par $f_{\#\mu}(A_2) = \mu\left(f^{-1}(A_2)\right)$. Dans le cas où $\mu = \mathbb{P}$ est une mesure de probabilité et $f = X$, on dit que X est une variable aléatoire, et sa mesure induite est souvent notée $\mathbb{P}_X$ ou $\mathcal{L}(X)$ plutôt que $X_{\#\mathbb{P}}$. La loi de X ou distribution de X sont en fait formellement définies comme étant exactement sa push-forward $\mathbb{P}_X$. Exercice : la loi de X est elle-même une mesure de probabilité.

Ensemble négligeable $N \subseteq \Omega$: il existe $Z \in \mathcal{A}$ de mesure nulle avec $N \subseteq Z$. On dit que la tribu $\mathcal{A}$ est complète pour la mesure μ ou que l'espace mesuré $(\Omega, \mathcal{A}, \mu)$ est complet si $\mathcal{A}$ contient tous les ensembles négligeables. Tribu complétée $\bar{\mathcal{A}} = \{A \cup N : A \in \mathcal{A}, N \text{ négligeable}\}$, c'est bien une tribu (Exercice), μ s'étend dessus en $\bar{\mu}$ (Exercice) et $\bar{\mathcal{A}}$ contient tous les négligeables de la mesure complétée donc $(\Omega, \bar{\mathcal{A}}, \bar{\mu})$ est complet (Exercice). Exemple : tribu de Lebesgue (parfois la mesure non-complétée est appelée mesure de Borel-Lebesgue pour la distinguer). Une propriété $P(\omega)$ dite vraie μ-presque partout si $\{\omega \in \Omega : \neg P(\omega)\}$ est négligeable (dépend de la mesure μ). On dit plutôt "presque sûrement" si de plus μ est une mesure de probabilité.

Intégration de fonctions mesurables à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$:

— si f étagée positive, nombre fini de valeurs $\alpha_1, \ldots, \alpha_N$:

$$\int f(x)\mathrm{d}\mu(x) = \int f\mathrm{d}\mu = \int f(x)\mu(\mathrm{d}x) = \sum_{i=1}^{N} \alpha_i f_{\#\mu}(\{\alpha_i\}) = \sum_{i=1}^{N} \alpha_i \mu\left(f^{-1}(\{\alpha_i\})\right) \in [0, \infty],$$

— si f positive,

$$\int f(x)\mathrm{d}\mu(x) = \sup_{\substack{h \text{ étagée positive} \\ h \leq f}} \int h(x)\mathrm{d}\mu(x) \in [0, \infty],$$

— si $\int |f|\mathrm{d}\mu < \infty$, on dit que f est μ-intégrable et

$$\int f\mathrm{d}\mu = \int f_+\mathrm{d}\mu - \int f_-\mathrm{d}\mu \in \mathbb{R}.$$

Théorème (admis) : chacune de ces définitions existe, est bien posée, et cohérente avec les précédentes.

FAIRE UN DESSIN

Pour $p > 0$, on note $\mathcal{L}^p(\Omega, \mu)$ l'ensemble des fonctions f mesurables telles que $|f|^p$ est intégrable. En particulier, $\mathcal{L}^1(\Omega, \mu)$ est l'ensemble des fonctions f mesurables intégrables.

Propriétés : linéarité, croissance. $\mu(A) = \int \mathbb{1}_A \mathrm{d}\mu$. Si $f \geq 0$, $\int f\mathrm{d}\mu = 0 \Leftrightarrow f = 0$ presque partout (admis). Conséquence : $\int |f - g|\mathrm{d}\mu = 0 \Leftrightarrow f = g$ presque partout.

Si $\mu =$ la mesure de Lebesgue λ, c'est l'intégrale de Lebesgue et on écrit en général juste $\mathrm{d}x$ au lieu de $\lambda(\mathrm{d}x)$. Si $\mu = \sum_{n\in\mathbb{N}} \alpha_n \delta_{y_n}$, $\int f \mathrm{d}\mu = \sum_{n\in\mathbb{N}} \alpha_n f(y_n)$ (si défini). Si $\mu = \mathbb{P}$, l'intégrale de la v.a. $X : \Omega \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ par rapport à $\mathbb{P}$ (si elle existe) est ce que l'on appelle aussi son espérance $\mathbb{E}[X] = \int X \mathrm{d}\mathbb{P} = \int_{\omega\in\Omega} X(\omega)\mathrm{d}\mathbb{P}(\omega)$ mais on ne calcule jamais une espérance comme cela, on utilise le théorème de transfert qui dit que $\mathbb{E}[X] = \int_{x\in F} x \mathrm{d}\mathbb{P}_X(x)$: on passe donc d'une intégrale sur Ω par rapport à la mesure $\mathbb{P}$ à une intégrale sur $\mathbb{R}$ par rapport à la mesure induite $\mathbb{P}_X$.

Soit μ et ν deux mesures. ν domine μ, ou μ est absolument continue par rapport à ν, et on note $\mu \ll \nu$, si $\forall A \in \mathcal{A}, \nu(A) = 0 \Rightarrow \mu(A) = 0$. **Remarque :** toute mesure est dominée par la mesure de comptage (mais ce n'est pas très intéressant car elle est rarement σ-finie).

Théorème de Radon-Nikodym-Lebesgue : Soit μ et ν deux mesures σ-finies, avec $\mu \ll \nu$. Il existe une fonction h réelle positive mesurable, et unique ν-presque partout, telle que $\forall A \in \mathcal{A}, \mu(A) = \int \mathbb{1}_A h \mathrm{d}\nu$. h s'appelle la dérivée de Radon-Nikodym de μ par rapport à ν ou la **densité** de μ par rapport à ν. On la note aussi $h = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$. La réciproque est vraie.

Exemple : la densité de $\mathcal{N}(0,1)$ par rapport à λ est $x \mapsto \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$. **Remarque :** on doit en principe dire "une" densité mais on peut se permettre de dire "la" grâce à l'unicité λ-p.p.

Remarque : la théorie de la mesure donne un cadre unifié pour traiter des distributions de probabilité discrètes et continues.

Lemme des classes monotones. π-système : une partie de $\mathcal{P}(\Omega)$ stable par intersection. Exemple : toute tribu. Classe monotone ou λ-système : une partie $\mathcal{M}$ de $\mathcal{P}(\Omega)$ telle que :

1. $\Omega \in \mathcal{M}$
2. $A, B \in \mathcal{M}$ et $A \subseteq B$ entraînent que $B \setminus A \in \mathcal{M}$
3. $A_i \in \mathcal{M}$, $i \in \mathbb{N}$ avec $A_i \subseteq A_{i+1}$ entraînent que $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{M}$

$\mathcal{M}$ est donc stable par différence ensembliste et par union dénombrable croissante. Exemple : toute tribu. Lemme des classes monotones (ou théorème π-λ de Sierpiński–Dynkin) : soit $\mathcal{C}$ un π-système, alors la classe monotone engendrée par $\mathcal{C}$ contient la tribu engendrée par $\mathcal{C}$ (elles sont même égales vu qu'une tribu est une classe monotone).

Application très utile : soit 2 mesures de probabilité $\mathbb{P}$ et $\mathbb{Q}$ qui coïncident sur un π-système $\mathcal{C} \subseteq \mathcal{A}$, alors elles coïncident sur $\sigma(\mathcal{C})$. En effet, $\{A \in \mathcal{A} : \mathbb{P}(A) = \mathbb{Q}(A)\}$ est une classe monotone (Exercice). Conséquence : la cdf d'une loi caractérise cette loi.

Tribu cylindrique. Soit $F \subset \mathbb{R}^{\mathbb{T}}$ un sous-ensemble de fonctions à valeurs réelles définies sur un ensemble $\mathbb{T}$ quelconque. Pour $n \geq 1$, $t_1, \ldots, t_n \in \mathbb{T}$, $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$, on définit le cylindre $C_{t_1,\ldots,t_n}(B_1, \ldots, B_n) = \{f \in F : \forall i \in [\![1, n]\!], f(t_i) \in B_i\}$. On pose $\Sigma_{t_1,\ldots,t_n} = \sigma\left(\{C_{t_1,\ldots,t_n}(B_1, \ldots, B_n), \forall i \in [\![1, n]\!], B_i \in \mathcal{B}(\mathbb{R})\}\right)$. L'ensemble $\mathcal{F}_F = \bigcup_{n\geq 1} \bigcup_{t_1,\ldots,t_n \in \mathbb{T}} \Sigma_{t_1,\ldots,t_n}$ est appelé l'algèbre cylindrique de F et $\sigma\left(\mathcal{F}_F\right)$ est la tribu cylindrique de F. Application : cette construction nous fournit une tribu sur tout sous-ensemble de suites réelles, où l'on prend $\mathbb{T} = \mathbb{N}$. Cette construction est d'ailleurs plus simple car on a alors $\mathcal{F}_F = \bigcup_{n\geq 1} \Sigma_{1,\ldots,n}$ (Exercice), en prenant de plus $F = \mathbb{R}^{\mathbb{N}}$ tout entier, on a donc une tribu $\sigma\left(\mathcal{F}_{\mathbb{R}^{\mathbb{N}}}\right)$ sur l'espace des suites réelles.