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

Notes de cours, Master Maths & IA

Guillermo Durand

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.

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 : $\mathcal{P}(\Omega)$, $\{\varnothing, \Omega\}$, $\{\varnothing, A, A^{\mathsf{c}}, \Omega\} = \sigma(\{A\})$, $\mathcal{B}(\mathbb{R}^d)$ la tribu borélienne de $\mathbb{R}^d$ 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).

Exercice : soit $f : \Omega_1 \to (\Omega_2, \mathcal{A}_2)$, $f^{-1}(\mathcal{A}_2)$ est une tribu appelée tribu image réciproque de $\mathcal{A}_2$. (Moins utile :) 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 $\mathcal{A}_1$. **ATTENTION** ce n'est pas $f(\mathcal{A}_1)$! ! (Contre-exemple : $f : \mathbb{R} \to \mathbb{R}$ constante, $f(\mathcal{A}_1)$ n'est même pas une tribu)

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

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

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.

Propriétés : $\mu\left(\bigcup_{1 \leq n \leq N} E_n\right) = \sum_{1 \leq n \leq N} \mu(E_n)$ si les E_n sont deux à deux disjoints. $\mu(B \setminus A) = \mu(B) - \mu(A)$ si $A \subseteq B$, en particulier $\mu(A) \leq \mu(B)$. $\mu\left(\bigcup_{n \in \mathbb{N}} E_n\right) \leq \sum_{n \in \mathbb{N}} \mu(E_n)$. $\mu\left(\bigcup_{n \in \mathbb{N}} E_n\right) = \lim_{n \to \infty} \mu(E_n)$ 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 $\mu(\Omega) = 1$.

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 $\mathbb{R}$: la seule mesure λ invariante par translation et telle que $\lambda([0,1]) = 1$. Elle est σ-finie. Toutes les mesures de probabilité induites par les lois réelles usuelles, par exemple $\mathcal{N}(0,1) : A \mapsto \int \mathbb{1}_{\{x\in A\}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \mathrm{d}x$.

Mesure produit de μ_1 et μ_2 sur $\mathcal{A}_1 \otimes \mathcal{A}_2$: μ telle que $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$. Unique si μ_1 et μ_2 σ-finies (admis). Exemple : la mesure de Lebesgue sur $\mathbb{R}^d$.

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 $\mathbb{P}_X$.

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), et μ s'étend dessus (Exercice). $\bar{\mathcal{A}}$ contient tous les négligeables de la mesure complétée et donc $(\Omega, \bar{\mathcal{A}}, \mu)$ est complet. Exemple : tribu de Lebesgue. Une propriété $P(\omega)$ est vraie presque partout si $\{\omega \in \Omega : \neg P(\omega)\}$ est négligeable. 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 $\alpha_1, \ldots, \alpha_N$:

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

— si f positive,

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

— si $\int |f| \mathrm{d}\mu < \infty$,

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

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

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).

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. $X : \Omega \to F$ par rapport à $\mathbb{P}$ (si elle existe) est ce que l'on appelle aussi son espérance $\mathbb{E}[X] = \int X \mathrm{d}\mathbb{P} = \int_{\omega\in\Omega} X(\omega)\mathbb{P}(\mathrm{d}\omega)$ mais on ne calcule jamais une espérance comme cela, on utilise le théorème de transfert qui dit que $\mathbb{E}[X] = \int_{x\in F} x\mathbb{P}_X(\mathrm{d}x)$: on passe donc d'une intégrale sur Ω par rapport à la mesure $\mathbb{P}$ à une intégrale sur F 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.

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.

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 $\mathcal{C}$. 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).