

Fondements de la Modélisation Statistique : Théorie de la Mesure et Applications

Statistique SD2 202526, Master DS, amU
Hadrien Lorenzo

Table des matières

1 Définitions essentielles en théorie de la mesure	2
1.1 Éléments de topologie	2
1.2 Tribus et espaces mesurables	3
1.3 Mesure	4
1.4 Tribu et mesure produit	6
2 Théorème de Radon-Nikodym	7
2.1 Convergence presque partout et domination	7
2.2 Deux théorèmes de Radon-Nikodym	7
3 Modélisation statistique	8
3.1 Modélisations statistique et théorie de la mesure	8
3.2 Application aux modèles statistiques dominés	9
3.3 Exemple fondamental 1 : Les ampoules qui grillent, ou pas	10
3.4 Exemple fondamental 2 : Capteur saturé	11
4 Exercices d'application	12
5 Conclusion	14

L'intuition derrière la « densité »

Imaginez deux matériaux de densités différentes. Pour trouver la masse d'une région, vous intégrez une fonction de densité sur cette région. De même, en théorie de la mesure, on souhaite parfois exprimer une mesure μ (la « masse ») à l'aide d'une autre mesure ν (le «

volume ») via une fonction f (la « densité ») :

$$\mu(A) = \int_{\mathbf{X}} \mathbf{1}_A(x)f(x)d\nu(x) \quad \text{pour tout ensemble mesurable } A.$$

Le théorème de Radon-Nikodym donne les conditions sous lesquelles une telle fonction f , appelée dérivée de Radon-Nikodym, existe.

Ce précis ne propose aucune démonstration, il s'attache surtout à proposer un grand nombre de définitions permettant de comprendre la formalisation du théorème de Radon-Nikodym d'une part. D'autre part, il particularise ces résultats à de nombreux exemples et notamment au cas de la statistique. Explicitant ainsi le calcul de vraisemblance dans des cas pathologiques comme celui d'un capteur de température qui sature.

1 Définitions essentielles en théorie de la mesure

La théorie des ensembles, nécessaire à la construction de la théorie de l'intégration et de la mesure, a nécessité l'axiomatisation de certaines notions. Sans axiomatisation solide, la théorie se retrouve avec des paradoxes, comme le paradoxe du barbier : un barbier doit raser tous les hommes qui ne se rasent pas, mais alors peut-il se raser lui-même ? Nous ne détaillerons pas ici ces axiomes mais renvoyons le lecteur à deux sources : une première source, bibMath, sans epsilon. Une seconde source, la page de Caroline Vernier, agrégée de mathématiques à l'Université de Strasbourg, avec des epsilon bien détaillés.

Ces axiomes sont

1.1 Éléments de topologie

Définition (Topologie et ouvert). *Soit \mathbf{X} un ensemble non vide et \mathcal{T} une famille de parties de \mathbf{X} si :*

- \mathbf{X} et \emptyset sont dans \mathcal{T} ,
- toute réunion d'éléments de \mathcal{T} est dans \mathcal{T} ,
- toute intersection finie d'éléments de \mathcal{T} est dans \mathcal{T} ,

alors on dit que \mathcal{T} est une topologie pour \mathbf{X} .

Le couple $(\mathbf{X}, \mathcal{T})$ est appelé espace topologique.

Un élément de \mathcal{T} est appelé ouvert de \mathbf{X} .

Définition (Fermé). *Soit \mathbf{X} un ensemble non vide et \mathbf{Y} une partie de \mathbf{X} . \mathbf{Y} est dit fermé si \mathbf{Y} est le complémentaire d'un ouvert dans \mathbf{X} .*

Définition (Ensemble des parties d'un ensemble : $\mathcal{P}(\cdot)$). *Soit \mathbf{X} un ensemble non vide. L'ensemble des parties de \mathbf{X} , notée $\mathcal{P}(\mathbf{X})$, est l'ensemble de tous les sous-ensembles de \mathbf{X} , y compris \emptyset et lui-même :*

$$\mathcal{P}(\mathbf{X}) = \{A \mid A \subset \mathbf{X}\}.$$

1.2 Tribus et espaces mesurables

Définition (Tribu ou σ -algèbre). Soit \mathbf{X} un ensemble non vide. Une tribu (ou σ -algèbre) sur \mathbf{X} est une collection \mathcal{A} de parties de \mathbf{X} vérifiant :

- \mathbf{X} est dans \mathcal{A} ,
- stabilité par complémentation : si $A \in \mathcal{A}$, alors $A^c \in \mathcal{A}$,
- stabilité par union dénombrable : si $(A_n)_{n \in \mathbb{N}}$ est une suite d'éléments de \mathcal{A} , alors $\bigcup_{n=0}^{\infty} A_n \in \mathcal{A}$.

Remarque. Les propriétés d'une tribu impliquent automatiquement que $\emptyset \in \mathcal{A}$ puisque $\emptyset = \mathbf{X}^c$.

Définition (Tribu engendrée). Soit une collection \mathcal{E} de parties de \mathbf{X} . La tribu engendrée par \mathcal{E} , notée $\sigma(\mathcal{E})$, est la plus petite tribu sur \mathbf{X} contenant \mathcal{E} . Elle est égale à l'intersection de toutes les tribus contenant \mathcal{E} .

Définition (Espace mesurable). Un espace mesurable est un couple $(\mathbf{X}, \mathcal{A})$ où \mathbf{X} est un ensemble non vide appelé univers et \mathcal{A} est une tribu sur \mathbf{X} .

Définition (Ensemble mesurable). Soit $(\mathbf{X}, \mathcal{A})$ un espace mesurable. Un sous-ensemble $E \subset \mathbf{X}$ est dit mesurable si $E \in \mathcal{A}$.

Exemple 1 (Un ensemble à trois éléments). Soit $\mathbf{X} = \{a, b, c\}$ un ensemble à trois éléments. Montrons que $\mathcal{A} = \{\emptyset, \{a\}, \{b, c\}, \mathbf{X}\}$ est une tribu.

On a bien $\mathbf{X} \in \mathcal{A}$ (1/3). Vérifions la stabilité par complémentation (2/3) :

$$\emptyset^c = \mathbf{X} \in \mathcal{A}, \quad \{a\}^c = \{b, c\} \in \mathcal{A}, \quad \{b, c\}^c = \{a\} \in \mathcal{A} \quad \text{et} \quad \mathbf{X}^c = \emptyset \in \mathcal{A}.$$

Vérifions la stabilité par union dénombrable (3/3), les unions possibles sont :

$$\begin{aligned} \emptyset \cup \{a\} &= \{a\} \in \mathcal{A}, & \emptyset \cup \{b, c\} &= \{b, c\} \in \mathcal{A}, \\ \{a\} \cup \{b, c\} &= \mathbf{X} \in \mathcal{A} & \text{et} \quad \emptyset \cup \{a\} \cup \{b, c\} \cup \mathbf{X} &= \mathbf{X} \in \mathcal{A}. \end{aligned}$$

Donc \mathcal{A} est bien une tribu.

Les ensembles mesurables sont exactement les éléments de \mathcal{A} , c'est-à-dire $\emptyset, \{a\}, \{b, c\}$ et \mathbf{X} . Les ensembles $\{b\}$ et $\{c\}$ ne sont pas mesurables car ils n'appartiennent pas à \mathcal{A} .

Définition (Tribu borélienne ou tribu de Borel). Soit $(\mathbf{X}, \mathcal{A})$ un espace mesurable. La tribu borélienne $\mathcal{B}(\mathbf{X})$ est la tribu engendrée par les ouverts de \mathbf{X} . Ses éléments sont appelés boréliens.

Exemple 2 (La tribu de Borel sur \mathbb{R}). Soit $\mathbf{X} = \mathbb{R}$. La tribu de Borel $\mathcal{B}(\mathbb{R})$ est la plus petite tribu contenant tous les intervalles ouverts de \mathbb{R} . En vertu de la Définition 1.2 :

- Les intervalles ouverts (a, b) sont mesurables.
- Les intervalles fermés $[a, b]$ sont mesurables.
- Les singletons $\{x\}$ sont mesurables.
- Les ensembles dénombrables (comme \mathbb{Q}) sont mesurables.

L'espace $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est l'espace mesurable le plus utilisé en théorie de la mesure.

1.3 Mesure

Après avoir introduit les tribus qui déterminent *ce qui* est mesurable, nous définissons maintenant les mesures qui spécifient *comment* mesurer.

Une mesure assigne systématiquement une “taille” (un nombre positif ou $+\infty$) à chaque ensemble mesurable, avec deux propriétés fondamentales :

- L’ensemble vide est de taille nulle.
- La taille d’une union disjointe est la somme des tailles.

Définition (Mesure et espace mesuré). *Soit $(\mathbf{X}, \mathcal{A})$ un espace mesurable. Une mesure sur $(\mathbf{X}, \mathcal{A})$ est une application :*

$$m : \mathcal{A} \rightarrow [0, +\infty]$$

vérifiant les deux propriétés suivantes :

1. L’ensemble vide est à mesure nulle : $m(\emptyset) = 0$.
2. σ -additivité de l’application m : pour toute famille dénombrable A_1, A_2, \dots de parties de \mathbf{X} appartenant à \mathcal{A} . Si ces parties sont deux à deux disjointes (i.e. $A_i \cap A_j = \emptyset$ pour $i \neq j$) alors on a :

$$m \left(\bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} m(A_n).$$

Le triplet $(\mathbf{X}, \mathcal{A}, m)$ est alors appelé espace mesuré

Maintenant que nous avons défini la notion générale de mesure, intéressons-nous à des exemples concrets.

Définition (Mesure de Dirac). *Soit $(\mathbf{X}, \mathcal{A})$ un espace mesurable et $a \in \mathbf{X}$. La mesure de Dirac au point a est définie par :*

$$\delta_a(A) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{si } a \notin A \end{cases} \quad \text{pour tout } A \in \mathcal{A}$$

À l’opposé de la mesure de Dirac qui concentre toute la masse en un point, la mesure de comptage considère chaque élément individuellement.

Définition (Mesure de comptage). *Soit $(\mathbf{X}, \mathcal{P}(\mathbf{X}))$ un espace mesurable. La mesure de comptage est définie par :*

$$\mu_c(A) = \begin{cases} \text{nombre d’éléments de } A & \text{si } A \text{ est fini} \\ +\infty & \text{si } A \text{ est infini.} \end{cases}$$

Ces mesures discrètes sont importantes, mais en analyse réelle, nous travaillons principalement avec des mesures continues. La plus fondamentale d’entre elles est la mesure de Lebesgue.

Définition (Mesure de Lebesgue). Soit l'espace mesurable $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, voir example 2, la mesure de Lebesgue λ est l'unique mesure telle que :

$$\lambda([a, b]) = b - a \quad \text{pour tout intervalle } [a, b].$$

Cette mesure généralise la notion de longueur aux ensembles boréliens.

Après ces exemples fondamentaux, caractérisons maintenant certaines propriétés importantes que peuvent avoir les mesures.

Définition (Mesure finie ou bornée). Soit $(\mathbf{X}, \mathcal{A}, m)$ un espace mesuré. La mesure m est dite finie si $m(\mathbf{X})$ est finie.

La finitude est une propriété forte. Une notion plus faible est la σ -finitude. Elle est la condition d'application du théorème de Radon-Nikodym.

Définition (Mesure σ -finie). Soit $(\mathbf{X}, \mathcal{A}, m)$ un espace mesuré. La mesure m sur \mathcal{A} est dite σ -finie s'il existe une suite $(A_k)_{k \in \mathbb{N}}$ d'ensembles mesurables telle que :

1. $\mathbf{X} = \bigcup_{k \in \mathbb{N}} A_k$
2. $m(A_k) < +\infty$ pour tout $k \in \mathbb{N}$

Le triplet $(\mathbf{X}, \mathcal{A}, m)$ est appelé espace mesuré σ -fini.

Remarque. En statistique, la σ -finitude assure que nos modèles sont "bien comportés" :

- La mesure de Lebesgue sur \mathbb{R}^n est σ -finie : $\mathbb{R}^n = \bigcup_{k \in \mathbb{N}} [-k, k]^n$.
- La mesure de comptage sur \mathbb{N} est σ -finie : $\mathbb{N} = \bigcup_{k \in \mathbb{N}} \{0, 1, \dots, k\}$.

Exemple 3 (Mesure non σ -finie pathologique). La mesure de comptage sur l'ensemble mesurable $(E, \mathcal{P}(E))$, où E est infini non dénombrable n'est pas σ -finie.

Un cas particulier très important en probabilités est celui des mesures de probabilité.

Définition (Mesure de probabilité). Soit $(\mathbf{X}, \mathcal{A}, m)$ un espace mesuré. La mesure m sur \mathcal{A} est une mesure de probabilité si $m(\mathbf{X}) = 1$.

Pour compléter notre panorama, il nous faut aborder la notion d'ensembles négligeables, fondamentale en théorie de l'intégration.

Définition (Ensembles négligeables). Soit $(\mathbf{X}, \mathcal{A}, m)$ un espace mesuré. Un ensemble $N \subset \mathbf{X}$ est dit négligeable (ou m -négligeable) s'il existe $A \in \mathcal{A}$ tel que $N \subset A$ et $m(A) = 0$.

Cette notion nous amène naturellement à la propriété de complétude d'un espace mesuré.

Définition (Espace mesuré complet). Un espace mesuré $(\mathbf{X}, \mathcal{A}, m)$ est dit complet si tout ensemble négligeable est mesurable, c'est à dire appartient à \mathcal{A} .

Exemple 4. L'espace mesuré $(\mathbf{X} = \{a, b, c\}, \mathcal{A} = \mathcal{P}(\mathbf{X}), \delta_a)$, δ_a voir Définition 1.3, est complet. En effet, les ensembles de mesure nulle sont exactement les ensembles qui ne contiennent pas a . Soit $N \subset \mathbf{X}$ avec $\delta_a(N) = 0$ (i.e., $a \notin N$), et soit $B \subset N$. Puisque $a \notin N$, on a aussi $a \notin B$, donc $\delta_a(B) = 0$. De plus, $B \in \mathcal{P}(\mathbf{X})$ car $\mathcal{P}(\mathbf{X})$ contient toutes les parties de \mathbf{X} .

Exemple 5. L'espace mesuré $(\mathbf{X} = \{a, b, c\}, \mathcal{A} = \mathcal{P}(\mathbf{X}), \mu_c)$, μ_c la mesure de comptage voir Définition 1.3, est complet. En effet, les seuls ensembles de mesure nulle sont \emptyset (qui a 0 élément) et c'est le seul. En effet $\mu_c(\emptyset) = 0$. Le seul sous-ensemble de \emptyset est \emptyset lui-même. $\emptyset \in \mathcal{P}(\mathbf{X})$ car $\mathcal{P}(\mathbf{X})$ contient toutes les parties.

Exemple 6 (Un exemple d'espace mesuré non complet). Soit $\mathbf{X} = \{1, 2, 3\}$ et considérons la tribu :

$$\mathcal{A} = \{\emptyset, \{1\}, \{2, 3\}, \mathbf{X}\}$$

Définissons la mesure :

$$m(A) = \begin{cases} 0 & \text{si } A = \emptyset \\ 1 & \text{si } A = \{1\} \\ 0 & \text{si } A = \{2, 3\} \\ 2 & \text{si } A = \mathbf{X} \end{cases}$$

Cet espace mesuré n'est pas complet. En effet, $\{2, 3\}$ est de mesure 0. Considérons $\{2\} \subset \{2, 3\}$ où $m(\{2, 3\}) = 0$ qui est mesurable de mesure nulle, donc $\{2\}$ est négligeable (voir Définition 1.3). Mais pourtant $\{2\}$ n'est pas mesurable puisque $\{2\} \notin \mathcal{A}$. L'espace mesuré n'est donc pas complet.

Exemple 7 (Mesure de Lebesgue complétée). Soit $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, λ voir Définition 1.3. Cet espace n'est pas complet. En effet, il existe des sous-ensembles d'ensembles de mesure nulle qui ne sont pas boreliens. La complétion de cet espace est $(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda)$, où :

$$\mathcal{L}(\mathbb{R}) = \{A \cup N : A \in \mathcal{B}(\mathbb{R}), N \subset B \in \mathcal{B}(\mathbb{R}) \text{ avec } \lambda(B) = 0\}$$

et on étend λ par $\lambda(A \cup N) = \lambda(A)$. Cet espace est complet : c'est la tribu de Lebesgue.

1.4 Tribu et mesure produit

Définition (Tribu produit). Soient $(\mathbf{X}_1, \mathcal{A}_1)$ et $(\mathbf{X}_2, \mathcal{A}_2)$ deux espaces mesurables. On pose $\mathbf{X} = \mathbf{X}_1 \times \mathbf{X}_2$ et on appelle tribu produit la tribu sur \mathbf{X} engendrée par $\mathcal{A}_1 \otimes \mathcal{A}_2 = \{A_1 \times A_2 \mid A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$.

Remarque. Il est possible de démontrer, mais c'est hors-programme, que :

$$\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^2)$$

et plus généralement que :

$$\bigotimes_{i=1}^n \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^n),$$

où la notation introduite est évidente.

Théorème 1 (Mesure produit). Soient $(\mathbf{X}_1, \mathcal{A}_1, \mu_1)$ et $(\mathbf{X}_2, \mathcal{A}_2, \mu_2)$ deux espaces mesurés σ -finis. Il existe une unique mesure μ sur l'espace mesurable produit $(\mathbf{X}_1 \times \mathbf{X}_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ telle que pour tous $A_1 \in \mathcal{A}_1$ et $A_2 \in \mathcal{A}_2$:

$$\mu(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2)$$

Cette mesure est de plus σ -finie et est appelée mesure produit.

Elle est notée $\mu = \mu_1 \otimes \mu_2$.

2 Théorème de Radon-Nikodym

2.1 Convergence presque partout et domination

Définition (Fonction mesurable, fonction borélienne). Soient $(\mathbf{X}, \mathcal{A})$ et $(\mathbf{Y}, \mathcal{B})$ deux espaces mesurables. Une fonction $f : \mathbf{X} \rightarrow \mathbf{Y}$ est dite mesurable (ou encore « $\mathcal{A}-\mathcal{B}$ mesurable si $f^{-1}(B) \in \mathcal{A}$ pour tout $B \in \mathcal{B}$. On dira que f est borélienne si \mathcal{B} est une tribu borélienne.

Remarque. Une fonction de \mathbb{R} dans \mathbb{R} est non mesurable si elle n'est pas l'image réciproque d'un borélien de \mathbb{R} par un borélien de \mathbb{R} . Ces fonctions sont pathologiques et difficiles à exhiber explicitement. La mesurabilité est une propriété fondamentale en théorie de la mesure, essentielle pour définir proprement l'intégrale de Lebesgue. Nous n'expliciterons pas plus. Cependant, il est nécessaire d'introduire la notion de fonction mesurable afin de définir un ingrédient essentiel au théorème de Radon-Nikodym, la convergence presque partout.

Définition (Convergence presque partout). Soit $(\mathbf{X}, \mathcal{A}, \mu)$ un espace mesuré et $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables de \mathbf{X} dans \mathbb{R} . On dit que (f_k) converge presque partout vers une fonction f s'il existe un ensemble $N \in \mathcal{A}$ de mesure nulle ($\mu(N) = 0$) tel que :

$$\forall X \in \mathbf{X} \setminus N, \quad \lim_{k \rightarrow \infty} f_k(X) = f(X)$$

On note alors $f_k \xrightarrow{p.p.} f$.

Définition (Absolue continuité/domination d'une mesure). Soient $(\mathbf{X}, \mathcal{A})$ un espace mesurable et μ, ν deux mesures positives sur cet espace. On dit que μ est absolument continue par rapport à ν , noté $\mu \ll \nu$, si :

$$\forall A \in \mathcal{A}, \quad \nu(A) = 0 \quad \Rightarrow \quad \mu(A) = 0.$$

On trouve aussi dans la littérature, surtout probabiliste et statistique, que la mesure ν domine la mesure μ ou encore que la mesure μ est dominée par la mesure ν .

2.2 Deux théorèmes de Radon-Nikodym

Théorème 2 (Théorème de Radon-Nikodym). Soient $(\mathbf{X}, \mathcal{A})$ un espace mesurable et μ et ν deux mesures σ -finies sur $(\mathbf{X}, \mathcal{A})$ et si $\mu \ll \nu$, alors il existe une fonction $f : (\mathbf{X}, \mathcal{A}) \rightarrow (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ mesurable telle que :

$$\forall A \in \mathcal{A}, \quad \mu(A) = \int_{\mathbf{X}} \mathbf{1}_A(x) f(x) d\nu(x) \tag{1}$$

La fonction f est unique à égalité ν -presque partout près et est notée $\frac{d\mu}{d\nu}$ que l'on appelle dérivée de Radon-Nikodym ou encore densité de μ par rapport à ν .

Théorème 3 (Théorème de Radon-Nikodym pour une mesure produit). Soient $(\mathbf{X}_i, \mathcal{A}_i, \mu_i)$, $i = 1 \dots n$, n espaces mesurés et μ_i , $i = 1 \dots n$, n mesures de référence σ -finies sur chacun

de ces espaces respectifs. Si $P = P_1 \otimes \cdots \otimes P_n$ et chaque $P_i \ll \mu_i$, alors $P \ll \mu_1 \otimes \cdots \otimes \mu_n$ et :

$$\frac{dP}{d(\mu_1 \otimes \cdots \otimes \mu_n)}(x_1, \dots, x_n) = \frac{dP_1}{d\mu_1}(x_1) \cdots \frac{dP_n}{d\mu_n}(x_n)$$

pour tout $(x_1, \dots, x_n) \in \mathbf{X}_1 \times \cdots \times \mathbf{X}_n$.

Corollaire (Dérivée de Radon-Nikodym pour les mesures discrètes). Soit $(\mathbf{X}, \mathcal{P}(\mathbf{X}))$ un espace mesurable discret, et soit μ_c la mesure de comptage sur \mathbf{X} . Soit P une mesure de probabilité sur \mathbf{X} telle que $P \ll \mu_c$.

Alors la dérivée de Radon-Nikodym de P par rapport à μ_c est donnée par la fonction de masse de probabilité :

$$\frac{dP}{d\mu_c}(x) = P(\{x\}) \quad \text{pour tout } x \in \mathbf{X}$$

Démonstration. Par définition de la dérivée de Radon-Nikodym, pour tout $A \in \mathcal{P}(\mathbf{X})$ on doit avoir :

$$P(A) = \int_{\mathbf{X}} \mathbf{1}_A(x) \frac{dP}{d\mu_c}(x) d\mu_c(x) = \int_{x \in A} \frac{dP}{d\mu_c}(x) d\mu_c(x) = \sum_{x \in A} \frac{dP}{d\mu_c}(x)$$

pour la mesure de comptage. D'un autre côté, par σ -additivité de P :

$$P(A) = \sum_{x \in A} P(\{x\})$$

L'égalité $P(A) = \int_A \frac{dP}{d\mu_c} d\mu_c$ étant vraie pour tout $A \subset \mathbf{X}$, on en déduit que nécessairement $\frac{dP}{d\mu_c}(x) = P(\{x\})$ pour tout $x \in \mathbf{X}$. \square

En statistique, on souhaite souvent exprimer comment une mesure de probabilité P_θ (dépendant d'un paramètre θ) se "comporte" par rapport à une mesure de référence ν . Le théorème de Radon-Nikodym fournit l'outil mathématique pour réaliser ceci. On remarquera que le théorème que l'on vient d'énoncer ne dépend pas de la nature discrète ou continue de l'ensemble \mathbf{X} , c'est ce qui en fait sa force. La suite de ce document transfère les notions vues jusqu'ici en terminologie statistique afin, principalement, de justifier/d'apprécier l'utilisation de la vraisemblance dans un grand nombre de problèmes statistiques.

3 Modélisation statistique

3.1 Modélisations statistique et théorie de la mesure

Définition (Modèle statistique). Un modèle statistique est un couple $(\mathbf{X}, (P_\theta)_{\theta \in \Theta})$ où \mathbf{X} est un ensemble appelé espace des observations, Θ est un ensemble appelé espace des paramètres et $(P_\theta)_{\theta \in \Theta}$ est une famille de probabilités définies sur une tribu fixée de parties de \mathbf{X} .

Dans la suite nous ne fixerons pas de tribu \mathcal{A} sur \mathbf{X} , sauf mention contraire. Ainsi si $\mathbf{X} = \mathbb{R}^n$ nous choisirons implicitement $\mathcal{A} = \mathbb{R}^n$ la tribu des boréliens sur \mathbb{R}^n et si \mathbf{X} est fini ou dénombrable, nous prendrons $\mathcal{A} = \mathcal{P}(\mathbf{X})$ l'ensemble des parties de \mathbf{X} .

Exemple 8 (Une expérience statistique, le lancé de 20 pièces). *Il s'agit ici de produire un modèle statistique qui représente l'expérience réalisée.*

On peut tout d'abord supposer que les pièces ont toutes les mêmes deux faces. Ainsi, en posant $n = 20$, il vient $\mathbf{X} : \{0, 1\}^n$ où 1 signifie "Pile" et 0 signifie "Face". On va munir $\{0, 1\}^n$ de la tribu discrète $\mathcal{P}(\{0, 1\}^n)$.

On peut alors supposer que tous les lancés suivent le même modèle de Bernoulli de paramètre $\theta \in [0, 1]$ et noté π_θ .

Le modèle statistique est ainsi $(\{0, 1\}^n, \pi_\theta^{\otimes 20})$.

Définition (Statistique sur un modèle). *Soit un modèle statistique $(\mathbf{X}, (P_\theta)_{\theta \in \Theta})$. On appelle statistique sur ce modèle toute application φ , indépendante de θ , mesurable relativement à la tribu \mathcal{A} considérée sur \mathbf{X} et à valeurs dans un certain espace \mathbf{Y} muni d'une tribu \mathcal{B} . Le modèle $(\mathbf{Y}, (Q_\theta)_{\theta \in \Theta}) = (\mathbf{Y}, (\varphi(P_\theta))_{\theta \in \Theta})$ où, pour tout $\theta \in \Theta$, la loi $Q_\theta = \varphi(P_\theta)$ est la loi image par φ de la loi P_θ , est dit modèle image par φ du modèle $(\mathbf{X}, (P_\theta)_{\theta \in \Theta})$.*

Exemple 9 (Suite du lancé de 20 pièces). *On reprend l'exemple 8. Dans le cas présent on va compter le nombre de "Pile". Il vient de considérer l'image par l'application φ : $(x_i)_{i=1..n} \mapsto x = \sum_{i=1}^n x_i$.*

On construit ainsi un nouveau modèle statistique qui prend ses valeurs dans $\mathbf{Y} = \llbracket 0, 20 \rrbracket$ et dont la loi est la loi binomiale de paramètre $(n, \theta) \in \mathbb{N}^ \times [0, 1]$.*

3.2 Application aux modèles statistiques dominés

Définition (Modèle statistique dominé). *Un modèle statistique $(\mathbf{X}, (P_\theta)_{\theta \in \Theta})$ est dit dominé par une mesure σ -finie ν sur $(\mathbf{X}, \mathcal{A})$ si pour tout $\theta \in \Theta$, on a $P_\theta \ll \nu$.*

Une conséquence directe du théorème de Radon-Nikodym, est l'existence d'une fonction de vraisemblance appliquée à chaque mesure P_θ par rapport à une mesure dominante ν .

Définition (Fonction de vraisemblance). *Soit $(\mathbf{X}, (P_\theta)_{\theta \in \Theta})$ un modèle statistique dominé par ν . On appelle fonction de vraisemblance la famille de fonctions $(L(\cdot, \theta))_{\theta \in \Theta}$ définie pour chaque $\theta \in \Theta$ par la dérivée de Radon-Nikodym :*

$$L(\cdot, \theta) = \frac{dP_\theta}{d\nu}$$

Ainsi, pour tout $A \in \mathcal{A}$:

$$P_\theta(A) = \int_{\mathbf{X}} \mathbf{1}_A(x) L(x, \theta) d\nu(x)$$

Remarque (Notation et terminologie). *Il vient l'ensemble des définitions supplémentaires suivant :*

- Pour une observation $x \in \mathbf{X}$ fixée, la fonction $\theta \mapsto L(x, \theta)$ est appelée vraisemblance du paramètre θ étant donné l'observation x .
- La notation $L(x, \theta)$ souligne que x est fixé (les données observées) tandis que θ varie (le paramètre inconnu). On est bien dans un contexte statistique où les données sont acquises et où il convient de savoir comment en tirer un maximum d'information sur Θ .

- Le théorème de Radon-Nikodym garantit l’existence et l’unicité (à un ensemble ν -négligeable près) de cette fonction.
- Aucune hypothèse n’est faite sur la nature de l’espace probabilisé. Ce dernier peut donc reposer sur un univers continu ou bien discret. C’est toute la force de résultat.

Corollaire (Vraisemblance dans le cas discret). *Pour un modèle statistique $(\mathbf{X}, (P_\theta)_{\theta \in \Theta})$ où \mathbf{X} est discret et dominé par la mesure de comptage μ_c , la fonction de vraisemblance s’identifie à la fonction de masse de probabilité :*

$$L(x, \theta) = \frac{dP_\theta}{d\mu_c}(x) = P_\theta(\{x\})$$

Exemple 10 (Retour aux pièces de monnaie). Reprenons l’exemple du lancé de $n = 20$ pièces 8. Le modèle statistique est $(\{0, 1\}^n, (\pi_\theta^{\otimes n})_{\theta \in [0, 1]})$. La loi du modèle s’écrit :

$$P_\theta(\{x\}) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} \quad (2)$$

Ce modèle est dominé par la mesure de comptage μ_c^n sur $\{0, 1\}^n$ qui est une mesure produit : $\mu_c^n = \bigotimes_{i=1}^n \mu_c$. Donc par application du théorème de Radon-Nikodym pour une mesure produit 3, la vraisemblance $L(x, \theta)$ du problème s’écrit :

$$L(x, \theta) = \frac{dP_\theta}{d\mu_c^n}(x) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}.$$

Exemple 11 (Échantillon gaussien indépendant). Supposons l’observation de n valeurs indépendantes et identiquement distribuées suivant chacune une loi normale de moyenne θ et de variance $\sigma^2 = 1$. On considère alors l’espace des observations \mathbb{R}^n et la tribu des boréliens sur \mathbb{R}^n comme tribu associée. Le modèle statistique ainsi construit est $(\mathbb{R}^n, (\mathcal{N}(\theta, 1)^{\otimes n})_{\theta \in \mathbb{R}})$ avec donc $P_\theta = \mathcal{N}(\theta, 1)^{\otimes n}$ pour tout $\theta \in \mathbb{R}$. Ce modèle est dominé par la mesure de Lebesgue $\lambda^n = \bigotimes_{i=1}^n \lambda$ sur \mathbb{R}^n . Pour chaque $\theta \in \mathbb{R}$, la mesure $(\mathcal{N}(\theta, 1))_{\theta \in \mathbb{R}}$ admet une densité par rapport à λ donnée par :

$$\frac{d\mathcal{N}(\theta, 1)}{d\lambda}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}},$$

il vient alors, par application du théorème de Radon-Nikodym pour une mesure produit 3, pour tout $x = (x_1, \dots, x_n)$:

$$L(x, \theta) = \frac{dP_\theta}{d\lambda^n}(x) = (2\pi)^{-n/2} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2},$$

3.3 Exemple fondamental 1 : Les ampoules qui grillent, ou pas

On allume n ampoules à $t = 0$ et on ferme la porte. On ré-ouvre la porte à $t = T > 0$ et on observe n_1 ampoules allumées et n_2 ampoules grillées. Quelle est la vraisemblance du problème ?

On construit l’espace des observations $\mathbf{X} = \{“éteinte à T”, “allumée à T”\}^n$ où $\forall x = (x_1, \dots, x_n) \in \mathbf{X}$, x_i décrit si l’ampoule i est allumée ou éteinte à l’instant T . On y associe la tribu $\mathcal{P}(\mathbf{X})$ et la mesure de comptage $\mu_c^n = \bigotimes_{i=1}^n \mu_c$ sur \mathbf{X} . Il nous reste à mesurer chacun des deux événements, dans la suite on note T_i la date d’extinction de l’ampoule i (attention cette date n’est pas connue, on dit qu’elle est latente). On suppose que les T_i sont toutes indépendantes et identiquement distribuées selon une loi exponentielle de paramètre $\theta > 0$, ainsi :

- $P_\theta(\{\text{“allumée après T”}\}) = \mathbb{P}(T_i > T) = \int_{t=T}^{\infty} \theta e^{-\theta t} dt = e^{-\theta T}.$
- $P_\theta(\{\text{“éteinte avant T”}\}) = \mathbb{P}(T_i \leq T) = \int_{t=0}^T \theta e^{-\theta t} dt = 1 - e^{-\theta T}.$

On voit que $P_\theta \ll \mu_c$, car $\{\text{“allumée après T”}\}$ et $\{\text{“éteinte après T”}\}$ sont des singletons et donc des ensembles de mesure de comptage unité. Aussi, $P_\theta^n \ll \mu_c^n$ et donc par application du théorème de Radon-Nikodym pour une mesure produit 3 :

$$L(x, \theta) = \frac{dP_\theta}{d\mu_c^n}(x) = P_\theta(\{\text{“allumée après T”}\})^{n_1} P_\theta(\{\text{“éteinte avant T”}\})^{n_2}.$$

Au total :

$$L(x, \theta) = e^{-n_1\theta T} (1 - e^{-\theta T})^{n_2}.$$

Et pour aller plus loin...

En dérivant par rapport à θ la log-vraisemblance que l'on cherche à annuler en $\hat{\theta}$, on obtient :

$$\frac{\partial \ell}{\partial \theta}(x, \hat{\theta}) = -Tn_1 + n_2 \frac{T e^{-\hat{\theta} T}}{1 - e^{-\hat{\theta} T}} = 0$$

ce qui équivaut à :

$$\hat{\theta} = \frac{1}{T} \log \left(\frac{n}{n_1} \right).$$

3.4 Exemple fondamental 2 : Capteur saturé

On accède aux données d'un capteur de température qui sature à partir d'une température a . On sait que la température suit une loi normale de moyenne θ et de variance 1. Parmi les $n = n_1 + n_2$ données, les n_1 premières ne sont pas saturées ($x_i < a$) alors que les n_2 suivantes sont saturées ($x_i \geq a$). Exprimer la vraisemblance de ce problème.

Nous avons deux types d'observations indépendantes :

- n_1 observations non saturées : $X_1, \dots, X_{n_1} \sim \mathcal{N}(\theta, 1)$ avec $X_i \in \mathbb{R}$.
- n_2 observations saturées : Y_1, \dots, Y_{n_2} avec $Y_j = a$ (valeur observée) mais la vraie valeur suit $\mathcal{N}(\theta, 1)$ tronquée pour des valeurs $\geq a$.

On construit ainsi deux ensembles d'observation

- $\mathbf{X} = \mathbb{R}^{n_1}$, muni de la tribu borélienne et de la mesure de Lebesgue associée. On utilise alors la mesure de Lebesgue λ^{n_1} comme mesure de référence, c'est une mesure produit. On cherche à construire la famille de probabilités $(P_{1,\theta})_{\theta \in \Theta}$ sur \mathbf{X} .
- $\mathbf{Y} = \{a\}^{n_2}$, muni de la mesure de comptage et de l'unique tribu accessible : $\{\emptyset, \mathbf{Y}\}$. On utilise alors la mesure de comptage $\mu_c^{n_2}$ comme mesure de référence, c'est une mesure produit. On cherche à construire la famille de probabilités $(P_{2,\theta})_{\theta \in \Theta}$ sur \mathbf{Y} .

On travaille alors sur l'espace des observations $\mathbf{Z} = \mathbf{X} \times \mathbf{Y}$ et on construit la mesure produit suivante :

$$\nu = \lambda^{n_1} \otimes \mu_c^{n_2}.$$

Pour $i = 1 \dots n_1$, nous sommes en présence d'une famille de mesures produit, $(P_{1,\theta})_\theta$, chacune dominée par la mesure de Lebesgue λ^{n_1} . Ainsi par application du théorème de Radon-Nikodym pour une mesure produit 3 pour les n_1 observations indépendantes :

$$\frac{dP_{1,\theta}}{d\lambda^{n_1}}(x_1, \dots, x_{n_1}) = \prod_{i=1}^{n_1} \varphi(x_i - \theta),$$

où φ est la densité de $\mathcal{N}(0, 1)$.

Pour $i = 1 \dots n_2$, la loi de Y_i sous $P_{2,\theta}$ est discrète : c'est une masse ponctuelle en a . On introduit une variable aléatoire $W \sim \mathcal{N}(\theta, 1)$ non observée, aussi qualifiée de latente. Il vient :

$$P_{2,\theta}(\{a\}) = \mathbb{P}(W \geq a) = 1 - \Phi(a - \theta),$$

où Φ est la fonction de répartition de $\mathcal{N}(0, 1)$. Cette loi discrète est dominée par la mesure de comptage μ_c sur l'ensemble $\{a\}$ et ainsi :

$$\frac{dP_{2,\theta}}{d\mu_c}(y_j) = 1 - \Phi(a - \theta).$$

Par application du théorème de Radon-Nikodym pour une mesure produit 3 pour les n_2 observations indépendantes :

$$\frac{dP_{2,\theta}}{d\mu_c^{n_2}}(y_1, \dots, y_{n_2}) = (1 - \Phi(a - \theta))^{n_2}.$$

Par le théorème de Radon-Nikodym pour une mesure produit 3, puisque $P_\theta = P_{1,\theta} \otimes P_{2,\theta}$, la fonction de vraisemblance du problème s'écrit :

$$L(z, \theta) = \frac{dP_{1,\theta}}{d\lambda^{n_1}}(x_1, \dots, x_{n_1}) \frac{dP_{2,\theta}}{d\mu_c^{n_2}}(y_1, \dots, y_{n_2}) = \left(\prod_{i=1}^{n_1} \varphi(x_i - \theta) \right) (1 - \Phi(a - \theta))^{n_2}.$$

Et pour aller plus loin...

Supposons que nous souhaitons maximiser cette log-vraisemblance en un point $\hat{\theta}$:

$$\frac{\partial \ell}{\partial \theta}(x, \hat{\theta}) = \sum_{i=1}^{n_1} (x_i - \hat{\theta}) + n_2 \frac{\varphi(a - \hat{\theta})}{1 - \Phi(a - \hat{\theta})} = 0.$$

Ce qui conduit à l'équation implicite

$$\hat{\theta} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i + \frac{n_2}{n_1} \frac{\varphi(a - \hat{\theta})}{1 - \Phi(a - \hat{\theta})}.$$

La résolution de cette équation nécessite l'utilisation de méthodes numériques.

4 Exercices d'application

On note EMV l'estimateur du maximum de vraisemblance dans la suite.

Exercice 1. Soit $\mathbf{X} = \{1, 2, 3, 4\}$ un ensemble à quatre éléments.

- (a) Vérifier que $\mathcal{A}_1 = \{\emptyset, \{1, 2\}, \{3, 4\}, \mathbf{X}\}$ est une tribu.
- (b) La collection $\mathcal{A}_2 = \{\emptyset, \{1\}, \{2, 3, 4\}, \mathbf{X}\}$ est-elle une tribu ?
- (c) Déterminer la tribu engendrée par $\{\{1, 4\}\}$.
- (d) Combien d'éléments contient la tribu $\mathcal{P}(\mathbf{X})$?

Exercice 2. Soit $\mathbf{X} = \{a, b, c\}$ et $\mathcal{A} = \mathcal{P}(\mathbf{X})$.

- (a) Définir la mesure de Dirac δ_a et calculer $\delta_a(\{a\})$, $\delta_a(b, c)$.
- (b) Définir la mesure de comptage μ_c et calculer $\mu_c(\{a, b\})$, $\mu_c(\mathbf{X})$.
- (c) Soit μ définie par $\mu(\{a\}) = 0.5$, $\mu(\{b\}) = 0.3$, $\mu(\{c\}) = 0.2$. Vérifier que μ est une mesure de probabilité.

Exercice 3. Soit $(\mathbf{X}, \mathcal{A}, \mu)$ un espace mesuré.

- (a) Montrer que si $A, B \in \mathcal{A}$ avec $A \subset B$, alors $\mu(A) \leq \mu(B)$ (monotonie).
- (b) Montrer la sous-additivité : pour toute suite $(A_n) \subset \mathcal{A}$, $\mu(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \mu(A_n)$.

Exercice 4. (a) Montrer que la mesure de Lebesgue sur \mathbb{R} est σ -finie.

- (b) Soit $\mathbf{X} = \{1, 2, 3\}$, $\mathcal{A} = \{\emptyset, \{1\}, \{2, 3\}, \mathbf{X}\}$ et μ définie par $\mu(\{1\}) = 1$, $\mu(\{2, 3\}) = 0$. Cet espace est-il complet ?

Exercice 5. Pour chaque paire de mesures suivante, déterminer si $\mu \ll \nu$:

- (a) $\mu = \delta_0$, ν = mesure de Lebesgue sur \mathbb{R} .
- (b) $\mu \sim \mathcal{N}(0, 1)$, ν = mesure de Lebesgue sur \mathbb{R} .
- (c) $\mu \sim \mathcal{N}(0, 1)$, $\nu \sim \mathcal{N}(1, 1)$.

Exercice 6. (a) Soit $\mu \sim \mathcal{N}(\theta, \sigma^2)$ et ν la mesure de Lebesgue sur \mathbb{R} . Calculer $\frac{d\mu}{d\nu}$.

- (b) Soit $\mathbf{X} = \{1, 2, 3\}$ et P définie par $P(\{1\}) = 0.2$, $P(\{2\}) = 0.5$, $P(\{3\}) = 0.3$. Calculer $\frac{dP}{d\mu_c}$.

- (c) Soit μ la mesure exponentielle de paramètre λ et ν la mesure de Lebesgue sur \mathbb{R} . Calculer $\frac{d\mu}{d\nu}$.

Exercice 7. Soient $(\mathbf{X}_1, \mathcal{A}_1, \mu_1)$ et $(\mathbf{X}_2, \mathcal{A}_2, \mu_2)$ deux espaces mesurés σ -finis.

- (a) Montrer que si $\mu_1 \ll \nu_1$ et $\mu_2 \ll \nu_2$, alors $\mu_1 \otimes \mu_2 \ll \nu_1 \otimes \nu_2$.
- (b) Exprimer $\frac{d(\mu_1 \otimes \mu_2)}{d(\nu_1 \otimes \nu_2)}$ en fonction des dérivées individuelles.
- (c) Application : soit $P = \mathcal{N}(0, 1) \otimes \mathcal{N}(1, 2)$ et $\nu = \lambda \otimes \lambda$. Calculer $\frac{dP}{d\nu}(x, y)$.

Exercice 8. On considère n lancés indépendants d'une pièce de paramètre θ .

- (a) Écrire le modèle statistique formel.
- (b) Identifier la mesure dominante.
- (c) Exprimer la fonction de vraisemblance $L(x, \theta)$.
- (d) Calculer l'EMV de θ , noté $\hat{\theta}$.

Exercice 9. On reprend l'exemple des ampoules 3.3. Que se passe-t-il si n_1 est proche de n ? de 0 ?

Exercice 10. On considère un capteur qui satire en dessous de b et au-dessus de a ($b < a$). Parmi n observations indépendantes, n_1 observations dans $[b, a]$ (non saturées), n_2 observations égales à b (saturation basse), n_3 observations égales à a (saturation haute). La vraie valeur suit $\mathcal{N}(\theta, 1)$.

- (a) Écrire la vraisemblance complète.
- (b) Écrire l'équation vérifiée par l'EMV de θ , noté $\hat{\theta}$.

Exercice 11. On observe des temps de survie T_1, \dots, T_n avec censure à droite :

- Si $T_i \leq C_i$: on observe T_i
- Si $T_i > C_i$: on observe C_i et on sait que $T_i > C_i$

Les T_i sont indépendants et suivent une loi exponentielle de paramètre θ .

- (a) Écrire le modèle statistique formel.
- (b) Écrire la vraisemblance.
- (c) Trouver l'EMV de θ , noté $\hat{\theta}$.

Exercice 12. On observe des données provenant d'un mélange de deux populations :

$$f(x, \theta) = p\varphi(x - \mu_1) + (1 - p)\varphi(x - \mu_2)$$

où $\theta = (p, \mu_1, \mu_2)$ avec $p \in [0, 1]$ et φ est la dérivée de Radon-Nikodym de $\mathcal{N}(0, 1)$ par rapport à la mesure de Lebesgue sur \mathbb{R} .

- (a) Écrire le modèle statistique formel.
- (b) Identifier la mesure dominante.
- (c) Écrire la vraisemblance pour un échantillon i.i.d. x_1, \dots, x_n .
- (d) Peut-on explicitement trouver l'EMV de ce modèle ?

5 Conclusion

Ce cours a permis d'établir les fondements théoriques de la modélisation statistique moderne en s'appuyant sur la théorie de la mesure. Le fil conducteur a été le théorème de Radon-Nikodym, qui fournit un cadre unificateur et rigoureux pour définir la vraisemblance, concept central en inférence statistique.

Les exemples des ampoules et du capteur saturé ont illustré la puissance de cette approche. Ils ont montré comment construire rigoureusement une vraisemblance dans des situations où les observations sont incomplètes ou tronquées. Dans les deux cas, une partie de l'information est manquante (le temps d'extinction exact des ampoules, la valeur réelle des températures saturées). Ces informations manquantes, que l'on qualifie de variables latentes, complexifient l'expression de la vraisemblance et rendent sa maximisation directe difficile, comme l'a montré l'équation implicite obtenue pour le capteur saturé.

C'est précisément pour résoudre ce type de problèmes que la suite du cours introduira l'algorithme EM (Expectation-Maximisation). Cet algorithme permet de trouver une estimation du maximum de vraisemblance de manière itérative en « complétant » cycliquement les données observées par les variables latentes. La compréhension des fondements mesurables que nous venons d'établir est donc une préparation indispensable pour appréhender la puissance et les limites de cet algorithme fondamental en statistique.