Problèmes inverses — partie 2
Estimation par maximum de vraisemblance
Chapitre 1: contexte et rappels en probabilités et statitstiques

Nicolas Champagnat

Ecole des Mines de Nancy, 14/04/2020

Problèmes inverses : point de vue EDP

$$\begin{cases} \partial_t u(t,x) = \mu \partial_x u + \frac{\sigma^2}{2} \partial^2_{xx} u, & (t,x) \in \mathbb{R}^*_+ \times \mathbb{R} \\ u(0,x) = f(x) & x \in \mathbb{R}. \end{cases}$$

- **Objectif :** trouver les paramètres $\mu \in \mathbb{R}$ et $\sigma > 0$ à partir de $u(t,x)$ (ou d'observation partielles de $u(t,x)$)
- **Solution de l'EDP :**

$$u(t,x) = \int_{-\infty}^{+\infty} f(y) \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(x-y-\mu t)^2}{2\sigma^2 t}}\, dy$$

- **Approche possible :** on pose $M_k(f) = \int x^k f(x)\, dx$ et on calcule

$$\int x^2 u(t,x)\, dx = (\sigma^2 t + \mu^2 t) M_0(f) - 2\mu t M_1(f) + M_2(f)$$

$$\int x u(t,x)\, dx = M_1(f) - \mu t M_0(f).$$

- **Conclusion :** la connaissance des deux premiers moments de la condtion initiale f et de $u(t,x)$ permet de retrouver μ et σ

Problème inverse en statistiques

- Interprétation probabiliste de l'EDP :

$$u(t,x) = \mathbb{E}f(x+G),$$

où G est une variable aléatoire gaussienne de moyenne μt et de variance $\sigma^2 t$.

- Problème inverse en EDP : retrouver μ et σ à partir de u (avec l'espérance)
- Problème inverse en statistiques : retrouver μ et σ sans l'espérance, c-à-d à partir de G $\rightsquigarrow$ inférence statistique (plus précisément estimation paramétrique)
- En pratique il faut plusieurs mesures, c-à-d plusieurs observations indépendantes de G
- Ce cours : étude de la méthode d'estimation par maximum de vraisemblance (EMV)

Sources bibliographiques

[1] B. Cadre et C. Vial. *Statistique Mathématique. Cours & Exercices Corrigés*. Ellipse, 2012.

[2] V. Rivoirard et G. Stoltz. *Statistique Mathématique en Action*. Vuibert, 2012.

Mesure et intégration (1)

- On munit $\mathbb{R}^k$ de sa tribu borélienne $\mathcal{B}(\mathbb{R}^k)$
- Pour μ mesure positive sur $\mathbb{R}^k$,
  $\mathbb{L}^p(\mu) = \left\{f : \mathbb{R}^k \to R \text{ tel que } \int_{\mathbb{R}^k} |f|^p \, d\mu < +\infty\right\}.$

Définition (mesure produit)

- *Soit μ_1 et μ_2 deux mesures de probabilité sur $\mathbb{R}^k$ et $\mathbb{R}^d$, respectivement. On définit et on note $\mu_1 \otimes \mu_2$ la* **mesure produit** *de μ_1 et μ_2 comme l'unique mesure de probabilité sur $\mathbb{R}^{k+d}$ telle que*

$$\mu_1 \otimes \mu_2(A \times B) = \mu_1(A)\mu_2(B), \quad \forall A \in \mathcal{B}(\mathbb{R}^k) \text{ et } B \in \mathcal{B}(\mathbb{R}^d).$$

- *Soit μ une mesure de probabilité sur $\mathbb{R}^k$ et $n \in \mathbb{N}$. On note*

$$\mu^{\otimes n} = \underbrace{\mu \otimes \ldots \otimes \mu}_{n \text{ fois}}.$$

Mesure et intégration (2)

Théorème (convergence dominée)

Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions boréliennes de $\mathbb{R}^k$ dans $\mathbb{R}$ qui converge μ-presque partout pour une certaine mesure μ sur $\mathbb{R}^k$. S'il existe $g \in \mathbb{L}^1(\mu)$ telle que $|f_n| \leq g$ μ-presque partout pour tout $n \in \mathbb{N}$, alors

$$\int_{\mathbb{R}^k} \lim_{n\to+\infty} f_n \, d\mu = \lim_{n\to+\infty} \int_{\mathbb{R}^k} f_n \, d\mu.$$

Mesure et intégration (3)

Théorème (continuité sous le signe somme)

Soit $f : \mathbb{R} \times \mathbb{R}^k \to \mathbb{R}$ *tel que* $f(x, \cdot) \in \mathbb{L}^1(\mu)$ *pour tout* $x \in \mathbb{R}$, $f(\cdot, y)$ *est continue pour* μ*-presque tout* $y \in \mathbb{R}^k$ *et, pour tout* $x \in \mathbb{R}$, *il existe* $g \in \mathbb{L}^1(\mu)$ *et* $V \subset \mathbb{R}$ *un voisinage de* x *tel que*

$$\sup_{z \in V} |f(z, \cdot)| \leq g.$$

Alors l'application

$$x \mapsto \int_{\mathbb{R}^k} f(x, y)\, \mu(dy)$$

est continue sur $\mathbb{R}$.

Mesure et intégration (3)

Théorème (dérivation sous le signe somme)

Soit $f : \mathbb{R} \times \mathbb{R}^k \to \mathbb{R}$ tel que $f(x, \cdot) \in \mathbb{L}^1(\mu)$ pour tout $x \in \mathbb{R}$, $f(\cdot, y) \in \mathcal{C}^1$ pour μ-presque tout $y \in \mathbb{R}^k$ et pour tout $x \in \mathbb{R}$, il existe $g \in \mathbb{L}^1(\mu)$ et $V \subset \mathbb{R}$ un voisinage de x tel que

$$\sup_{z \in V} \left| \frac{\partial}{\partial x} f(z, \cdot) \right| \leq g.$$

Alors

$$\frac{d}{dx} \int_{\mathbb{R}^k} f(x, \cdot)\, d\mu = \int_{\mathbb{R}^k} \frac{\partial}{\partial x} f(x, \cdot)\, d\mu.$$

Mesure et intégration (4)

Définition

On dit qu'une mesure ν sur $\mathbb{R}^k$ est absolument continue par rapport à la mesure μ sur $\mathbb{R}^k$, et on note $\nu \ll \mu$, ssi pour tout $A \in \mathcal{B}(\mathbb{R}^k)$ tel que $\mu(A) = 0$, on a $\nu(A) = 0$.

Théorème (Radon-Nikodym)

Soit ν mesure sur $\mathbb{R}^k$ telle que $\nu \ll \mu$.
Alors il existe $f : \mathbb{R}^k \to \mathbb{R}$ mesurable telle que $\nu(A) = \int_A f \, d\mu$ pour tout $A \in \mathcal{B}(\mathbb{R}^k)$. La fonction f est unique μ-presque partout et $f \in \mathbb{L}^1(\mu)$ si ν est finie (c'est-à-dire $\nu(\mathbb{R}^k) < +\infty$).
On appelle f la densité de ν par rapport à μ. Si μ est la mesure de Lebesgue sur $\mathbb{R}^k$, f est simplement appelée densité de ν.

Mesure et intégration (5)

Définition (convergence étroite)

Une suite $(\mu_n)_{n\in\mathbb{N}}$ de mesures de probabilité sur $\mathbb{R}^k$ converge étroitement vers la mesure de probabilité μ sur $\mathbb{R}^k$, ce qu'on note $\mu_n \Rightarrow \mu$, ssi pour tout $f : \mathbb{R}^k \to \mathbb{R}$ continue bornée,

$$\lim_{n\to+\infty} \int_{\mathbb{R}^k} f\, d\mu_n = \int_{\mathbb{R}^k} f\, d\mu.$$

Théorème (Portmanteau)

$\mu_n \Rightarrow \mu$ ssi $\liminf_{n\to+\infty} \mu_n(O) \geq \mu(O)$ pour tout ouvert $O \subset \mathbb{R}^k$.

Si la mesure limite μ est absolument continue par rapport à la mesure de Lebesgue, alors
$\mu_n \Rightarrow \mu$ ssi $\lim_{n\to+\infty} \mu_n(O) = \mu(O)$ pour tout ouvert $O \subset \mathbb{R}^k$,
ssi $\lim_{n\to+\infty} \mu_n(F) = \mu(F)$ pour tout fermé $F \subset \mathbb{R}^k$.

Variables aléatoires, loi, indépendance (1)

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité.

Définition (variable aléatoire, loi)

- *On dit que X est une variable aléatoire (abrégé en v.a.) sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $\mathbb{R}^k$ si $X : \Omega \to \mathbb{R}^k$ est mesurable.*
- *La loi de la v.a. X est la mesure de probabilité sur $\mathbb{R}^k$ notée $\mathbb{P}_X$ définie pour tout $A \in \mathcal{B}(\mathbb{R}^k)$ par*

$$\mathbb{P}_X(A) = \mathbb{P}\{\omega \in \Omega \text{ tels que } X(\omega) \in A\}.$$

Variables aléatoires, loi, indépendance (2)

Définition (espérance, variance)

Soit X une v.a. à valeurs dans $\mathbb{R}^k$.

- *Si $p \in [1, +I)$, on note $X \in \mathbb{L}^p = \mathbb{L}^p(\mathbb{P})$ si $\int_\Omega |X(\omega)|^p \mathbb{P}(d\omega) < +\infty$.*
- *Si $g : \mathbb{R}^k \to \mathbb{R}^d$ est telle que $g(X) \in \mathbb{L}^1$, on appelle espérance de $g(X)$, et on note $\mathbb{E}(g(X)) = \int_\Omega g(X)\, d\mathbb{P}$.*
- *Si $X \in \mathbb{L}^2$, on appelle matrice de variance-covariance de X (ou simplement variance en dimension $k = 1$) et note*

$$\mathbb{V}(X) = \mathbb{E}(X - \mathbb{E}(X))(X - \mathbb{E}(X))^T,$$

où M^T est la matrice transposée de M.

On appelle écart-type d'une v.a. réelle la racine carrée de sa variance.

Variables aléatoires, loi, indépendance (3)

Définition (indépendance)

- *Si X et Y sont des v.a. à valeurs dans $\mathbb{R}^k$ et $\mathbb{R}^d$ respectivement, on dit que X et Y sont indépendantes ssi $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$, c'est-à-dire si*

$$\mathbb{P}((X,Y) \in A \times B) = \mathbb{P}(X \in A)\mathbb{P}(X \in B), \quad \forall A \in \mathcal{B}(\mathbb{R}^k),\ B \in \mathcal{B}(\mathbb{R}^d),$$

ou, de façon équivalente, si pour toutes fonctions mesurables bornées $f : \mathbb{R}^k \to \mathbb{R}$ et $g : \mathbb{R}^k \to \mathbb{R}$,

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y)).$$

- *Les v.a. $X_1, \ldots, X_n$ à valeurs dans $\mathbb{R}^{k_1}, \ldots, \mathbb{R}^{k_n}$ respectivement sont dites indépendantes ssi $\mathbb{P}_{(X_1,\ldots,X_n)} = \mathbb{P}_{X_1} \otimes \ldots \otimes \mathbb{P}_{X_n}$.*
- *Les v.a. $X_1, \ldots, X_n$ à valeurs dans $\mathbb{R}^k$ sont dites indépendantes et identiquement distribuées (abrégé en i.i.d.) de loi μ ssi la loi du vecteur $(X_1, \ldots, X_n)$ est $\mu^{\otimes n}$. On a alors en particulier que $\mu = \mathbb{P}_{X_i}$ pour tout i.*

Variables aléatoires, loi, indépendance (4)

Proposition

Si les v.a. $X_1, \ldots, X_n$ à valeurs dans $\mathbb{R}^k$ sont indépendantes et $\mathbb{L}^2$, alors

$$\mathbb{V}(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \mathbb{V}(X_i).$$

Quelques inégalités (1)

Inégalité de Markov Si $X \in \mathbb{L}^1$, pour tout $t > 0$,

$$\mathbb{P}(|X| \geq t) \leq \frac{\mathbb{E}|X|}{t}.$$

Inégalité de Tchebytchev Si $X \in \mathbb{L}^2$, pour tout $t > 0$,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\mathbb{V}|X|}{t^2}.$$

Quelques inégalités (2)

Inégalité de Cauchy-Schwarz Si $X, Y \in \mathbb{L}^2$,

$$(\mathbb{E}XY)^2 \leq (\mathbb{E}X^2)(\mathbb{E}Y^2).$$

De plus, l'inégalité précédente est une égalité ssi $\exists C \in \mathbb{R}$ telle que $X = CY$ presque sûrement (en abrégé, *p.s.*).

Inégalité de Jensen Si $X \in \mathbb{L}^1$ est une v.a. à valeurs réelles et $\psi : \mathbb{R} \to \mathbb{R}$ est convexe, alors

$$\psi(\mathbb{E}(X)) \leq \mathbb{E}\psi(X).$$

Si de plus ψ est strictement convexe, l'inégalité précédente est une égalité ssi $\exists C \in \mathbb{R}$ telle que $X = C$ p.s.

Convergence stochastique (1)

Soit $(Z_n)_{n\in\mathbb{N}}$ une suite de v.a. à valeurs dans $\mathbb{R}^k$ sur une suite d'espaces de probabilités $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$.

Définition (convergence en probabilité, en loi)

- *On dit que Z_n converge en probabilité vers une constante $z \in \mathbb{R}^k$, et on note $Z_n \xrightarrow{\mathbb{P}_n} z$, si $\forall \varepsilon > 0$, $\lim \mathbb{P}_n(|Z_n - z| > \varepsilon) = 0$.*
- *On dit que Z_n converge en loi vers la probabilité μ (resp. la v.a. Z de loi μ), et on note $Z_n \xrightarrow{loi} \mu$ (resp. $Z_n \xrightarrow{loi} Z$) si la suite des lois de Z_n converge étroitement vers μ, c'est-à-dire si, pour toute fonction $f : \mathbb{R}^k \to \mathbb{R}$ continue bornée, $\mathbb{E}_n f(X_n) \to \int_{\mathbb{R}^k} f\, d\mu = EE(Z)$.*

Convergence stochastique (2)

Si de plus tous les espaces de probabilité $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ sont identiques, égaux à $(\Omega, \mathcal{F}, \mathbb{P})$, on a la définition suivantes.

Définition (convergence presque sûre)

On dit que Z_n converge presque sûrement vers la v.a. Z sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans RR^k, et on note $Z_n \to Z$ p.s., si $\mathbb{P}(\lim_{n\to+\infty} Z_n = Z) = 1$.

Convergence stochastique (3)

Proposition

- *Toutes ces notions de convergence sont préservées par la composition par une fonction continue.*
- *La convergence presque sûre de Z_n vers Z implique la convergence en probabilité de $Z_n - Z$ vers 0, qui implique elle-même la convergence en loi de Z_n vers Z.*
- *Si $Z = C$ p.s. pour une certaine constante C, alors la convergence en loi de Z_n vers Z est équivalente à la convergence en probabilité de Z_n vers Z.*

Convergence stochastique (4)

Théorème (Loi forte des grands nombres (LGN))

Soit $(X_i)_{\in \mathbb{N}}$ *une suite de v.a. i.i.d. sur* $(\Omega, \mathcal{F}, \mathbb{P})$ *à valeurs dans* $\mathbb{R}^k$ *et de même loi* μ *telle que* $\int_{\mathbb{R}^k} |x| \, \mu(dx) < +\infty$. *Alors*

$$\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mathbb{E}(X_1) = \int_{\mathbb{R}^k} x \, \mu(dx)$$

et la convergence a même lieu presque sûrement.

Convergence stochastique (5)

Lemme (Slutsky)

Soit $(X_n)_{n\in\mathbb{N}}$ *et* $(Y_n)_{n\in\mathbb{N}}$ *deux suites de v.a. à valeurs dans* $\mathbb{R}^k$ *et* $\mathbb{R}^d$*, respectivement, telles que* X_n *converge en loi vers une v.a.* $X \in \mathbb{R}^k$ *et* Y_n *converge en probabilité vers une constante* $y \in \mathbb{R}^d$*.*
Alors (X_n, Y_n) *converge en loi vers* (X, y)*.*

Convergence stochastique (6)

Théorème (théorème central limite (TCL))

Soit $(X_i)_{\in\mathbb{N}}$ *une suite de v.a. i.i.d. sur* $(\Omega, \mathcal{F}, \mathbb{P})$ *à valeurs dans* $\mathbb{R}^k$ *et de même loi* μ *telle que* $\int_{\mathbb{R}^k} |x|^2 \, \mu(dx) < +\infty$*. On note*

$$m = \int_{\mathbb{R}^k} x \, \mu(dx) \in \mathbb{R}^k \quad \textit{et} \quad \Sigma = \mathbb{V}(\mu) = \int_{\mathbb{R}^k} (x-m)(x-m)^T \, \mu(dx)$$

sa matrice de variance-covariance. Alors

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - m\right) \xrightarrow{loi} \mathcal{N}_k(0, \Sigma),$$

où $\mathcal{N}_k(0, \Sigma)$ *désigne la loi gaussienne de dimension k, de moyenne nulle et de matrice de variance-covariance* Σ*.*

Vecteurs gaussiens (1)

Définition (vecteurs gaussiens)

On dit qu'un v.a. X à valeurs dans $\mathbb{R}^d$ est un vecteur gaussien s'il exite un vecteur $m \in \mathbb{R}^d$ et un matrice $d \times d$ symétrique positive Σ tels que

$$\mathbb{E}e^{i\langle u,X\rangle} = \exp\left(i\langle u,m\rangle - \frac{1}{2}u^T\Sigma u\right), \quad \forall u \in \mathbb{R}^d,$$

où $\langle\cdot,\cdot\rangle$ désigne le produit scalaire usuel dans $\mathbb{R}^d$.
On note $\mathcal{N}_d(m,\Sigma)$ la loi de X.

Vecteurs gaussiens (2)

- Si $X \sim \mathcal{N}_d(m, \Sigma)$, alors, pour tout $u \in \mathbb{R}^d$, $\langle u, X \rangle$ suit la loi gaussienne en dimension 1 de moyenne $\langle u, m \rangle$ et de variance $u^T \Sigma u$.
- De plus
$$\mathbb{E}(X) = m \quad \text{et} \quad \mathbb{V}(X) = \Sigma.$$
- Si la matrice Σ est inversible, alors X a pour densité (par rapport à la mesure de Lebesgue)
$$\frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x-m)^T \Sigma^{-1}(x-m)\right), \quad \forall x \in \mathbb{R}^d.$$
- En dimension $d = 1$, $\mathcal{N}(m, \sigma^2)$ a pour densité
$$\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \quad \forall x \in \mathbb{R}.$$

Vecteurs gaussiens (3)

Proposition (caractérisation d'un vecteur gaussien)

Un vecteur aléatoire X est gaussien ssi toute combinaison linéaire de ses composantes (c'est-à-dire toute v.a. de la forme $\langle u, X \rangle$ pour un vecteur u) est une v.a. réelle gaussienne.

Proposition

Si A est une matrice réelle $k \times d$, $b \in \mathbb{R}^k$ et $X \sim \mathcal{N}_d(m, \Sigma)$,

$$AX + b \sim \mathcal{N}_k(Am + b, A\Sigma A^T).$$

Proposition (caractérisation de l'indépendance)

Soit X un vecteur gaussien. Les composantes de X sont des v.a. réelles indépendantes ssi la matrice de variance-convariance de X est diagonale.

Lois usuelles (1)

Lois discrètes :

- Bern(p) : loi de Bernoulli de paramètre $p \in [0,1]$.
  $X \sim \mathrm{Bern}(p)$ si $X \in \{0,1\}$, $\mathbb{P}(X=1) = p$, $\mathbb{P}(X=0) = 1-p$.
- Bin(n,p) : loi binomiale de paramètres $n \in \mathbb{N}^*$ et $p \in [0,1]$.
  C'est la loi de la somme de n Bern(p) indépendante.
  $X \sim \mathrm{Bin}(n,p)$ si $X \in \{0,1,\ldots,n\}$ et $\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$.
- Geom(p) : loi géométrique de paramètre $p \in]0,1]$.
  C'est la loi de l'instant de premier succès (la première valeur 1 obtenue) dans une suite i.i.d. Bern(p).
  $X \sim \mathrm{Geom}(p)$ si $X \in \mathbb{N}^*$ et $\mathbb{P}(X=k) = p(1-p)^{k-1}$.
- Poi(λ) : loi de Poisson de paramètre $\lambda > 0$.
  $X \sim \mathrm{Poi}(\lambda)$ si $X \in \mathbb{N}$ et $\mathbb{P}(X=k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$.

Lois usuelles (2)

Lois continues :

- $\mathcal{U}([a,b])$: loi uniforme sur $[a,b]$ pour $a<b$.
  $X \sim \mathcal{U}([a,b])$ si $X \in [a,b]$ de densité $\frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$.
- $\mathrm{Exp}(\lambda)$: loi exponentielle de paramètres $\lambda > 0$.
  $X \sim \mathrm{Exp}(\lambda)$ si $X \in \mathbb{R}_+$ de densité $\lambda e^{-\lambda x}\mathbb{1}_{\mathbb{R}_+}(x)$.
- $\mathcal{N}_d(m,\Sigma)$: loi gaussienne de moyenne $m \in \mathbb{R}^d$ et de matrice de variance-covariance Σ de taille $d \times d$ (cf. ci-dessus).
- $\chi^2(d)$: loi du khi-deux à $d \geq 1$ degrés de liberté.
  C'est la loi de $|G|^2$ où $G \sim \mathcal{N}_d(0,\mathrm{Id})$.
  $X \sim \chi^2(d)$ si $X \in \mathbb{R}_+$ de densité

$$\frac{(1/2)^{d/2}}{\Gamma(d/2)} x^{d/2-1} e^{-x/2}\mathbb{1}_{\mathbb{R}_+}(x).$$

Vocabulaire (1)

Pour une étude statistique donnée, on appelle

Population l'ensemble des objets sur lesquels l'étude porte

Individu chaque objet dont est constituée la population

Caractère ou variable toute propriété des individus sur laquelle porte l'étude ; un caractère peut être quantitatif s'il prend ses valeurs dans $\mathbb{R}^k$ pour un certain $k \geq 1$, ou qualitatif (ou catégoriel) s'il prend ses valeurs dans un ensemble discret

Échantillon un ensemble d'individus sélectionnés selon un certain processus parmi la population ; si l'échantillon est de taille n, on parle de n-échantillon ;

On verra qu'on appelle également échantillon une variable aléatoire dont les observations sont une réalisation. Ce sera le sens principal de ce mot dans ce cours.

Vocabulaire (2)

Processus de sélection l'échantillon est obtenu avec un certain processus de sélection des individus dans la population.

Par exemple

- la totalité des individus de la population
- tirage aléatoire sans remise dans la population
- tirage aléatoire avec remise dans la population

Une sélection par tirage aléatoire permet de supposer une propiété d'indépendance des observations.

D'autres processus de sélection sont possibles, par exemple des mesures temporelles rapprochées chez un patient, qui ne permettent pas de supposer les observations indépendantes.

Observations ou données le vecteur $(x_1, \ldots, x_n)$ des caractères observés dans le n-échantillon

Exemple 1

Essai clinique portant sur une liste identifiée de patients diabétiques, visant à collecter des données de concentration sanguine d'insuline.

Population l'ensemble des n patients de l'essai clinique

Individu chaque patient

Caractère quantitatif la concentration d'insuline dans le sang d'un patient

Échantillon tous les individus de la population

Processus de sélection tous les individus de la population

Observations $(x_1, \ldots, x_n) =$ concentrations d'insuline chez les patients de la population

Exemple 2

Sondage d'opinion en vue d'une pochaine élection en France.

Population l'ensemble des français

Individu chaque français

Caractère qualitatif son opinion, à valeurs dans l'ensemble $\{$vote A; vote B; ne vote pas; sans opinion$\}$

Échantillon un ensemble de n individus dans la population

Processus de sélection par tirage aléatoire sans remise (on ne sonde une personne qu'une fois)

Observations $(x_1, \ldots, x_n)$ = opinions des français de l'échantillon

Exemple 3

Mesure de longueurs des nageoires des poissons d'un lac.

Population l'ensemble des poissons du lac

Individu chaque poisson du lac

Caractère quantitatif la longueur de ses nageoires

Échantillon un ensemble de n poissons du lac

Processus de sélection les n premiers poissons pêchés dans le lac
$\leadsto$ tirage aléatoire avec remise (si les poissons pêchés sont relâchés après avoir mesuré leurs nageoires)

Observations $(x_1, \ldots, x_n) =$ longueurs des nageoires des poissons pêchés

Vocabulaire (3)

Estimation paramétrique : évaluer des paramètres inconnus liés aux caractères considérés (par exemple leur moyenne ou leur écart-type) à partir des observations $x_1, \ldots, x_n$.

On distingue :

- estimation ponctuelle : l'estimation directe de la valeur des paramètres inconnus
- estimation par intervalles de confiance : la définition de bornes entre lesquelles les paramètres inconnus se trouvent avec un certain niveau de confiance
- tests statistiques sur les paramètres : la validation ou le rejet d'hypothèses sur les paramètres inconnus avec un certain seuil de confiance

Modélisation : ex. du jeu de pile-ou-face (1)

- Jeu de pile-ou-face répété n fois
- But de l'étude statistique : évaluer les caractéristiques de la pièce
- Population $=$ n-échantillon $=$ l'ensemble des n lancers de la pièce
- Caractère $=$ résultat du lancer, noté 0 pour pile et 1 pour face
- Observations $= (x_1, \ldots, x_n) \in \{0, 1\}^n$

Modélisation : ex. du jeu de pile-ou-face (2)

- Pour pouvoir estimer les caractéristiques de la pièce, il faut introduire une modélisation du jeu de pile-ou-face, c-à-d de la loi des observations.
- On suppose que $(x_1, \dots, x_n)$ est la réalisation d'une v.a. $(X_1, \dots, X_n)$ qui décrit les résultats du jeu de pile-ou-face :

$$(x_1, \dots, x_n) = (X_1(\omega), \dots, X_n(\omega)) \text{ pour un certain } \omega \in \Omega.$$

- Ici, on suppose que les v.a. $X_1, \dots, X_n$ sont i.i.d. d'une certaine loi de Bernoulli dont le paramètre est l'objet de l'étude.
  Autrement dit, $(X_1, \dots, X_n) \sim \text{Bern}(\theta)^{\otimes n}$ pour un certain $\theta \in [0, 1]$ inconnu.

Modélisation : ex. du jeu de pile-ou-face (3)

Le modèle statistique associé à cette étude est donc donné par

$$\left(\{0,1\}^n, \{\mathrm{Bern}(\theta)^{\otimes n}\}_{\theta \in [0,1]}\right),$$

où $\{0,1\}^n$ est l'ensemble des valeurs possibles de la v.a. X, θ est le paramètre du modèle, $[0,1]$ l'ensemble des valeurs du paramètre, et $\mathrm{Bern}(\theta)^{\otimes n}$ la loi des observations lorsque le paramètre vaut θ.

Modélisation : ex. du jeu de pile-ou-face (4)

Estimation paramétrique : à partir des observations $(x_1, \dots, x_n) \in \{0, 1\}^n$, identifier la valeur réelle θ_0 du paramètre inconnu θ.

- Loi des grands nombres :

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to+\infty]{} \mathbb{E}(X_1) = \theta_0 \quad \text{p.s.} \tag{1}$$

- Attention ! ici, la taille de l'échantillon ne tend pas vers l'infini
- Cependant, on peut considérer

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \approx \theta_0. \tag{2}$$

- Attention ! $(x_1, \dots, x_n)$ est une réalisation de $(X_1, \dots, X_n)$ pour un certain $\omega \in \Omega$. Puisque (1) a seulement lieu *presque sûrement*, il se pourrait que cette convergence n'est pas lieu ou la limite soit différente de θ_0 $\rightsquigarrow$ il faut quantifier l'approximation et l'incertitude

Modélisation : ex. du jeu de pile-ou-face (5)

- On calcule le risque quadratique :

$$\begin{aligned}\mathbb{E}[(\bar{X}_n-\theta_0)^2] = \mathbb{V}(\bar{X}_n) &= \frac{1}{n^2}\mathbb{V}(X_1+\ldots+X_n)\\ &= \frac{1}{n}\mathbb{V}(X_1) = \frac{1}{n}\theta_0(1-\theta_0)\\ &\leq \frac{1}{4n}.\end{aligned}$$

- Inégalité de Tchebytchev : pour tout $\varepsilon > 0$,

$$\mathbb{P}(|\bar{X}_n-\theta_0| \geq \varepsilon) \leq \frac{\mathbb{V}(\bar{X}_n)}{\varepsilon^2} \leq \frac{1}{4\varepsilon^2 n}.$$

- Intervalle de confiance pour la valeur de θ_0 : supposons par ex. que $n = 1000$ et qu'on veuille un intervalle de confiance à 95%. On résout $1/(4\varepsilon^2 n) = 1 - 95\% = 0,05$, ce qui donne $\varepsilon = 0,08$. On a alors

$$\mathbb{P}\left(\theta_0 \in [\bar{X}_n - 0,08\ ,\ \bar{X}_n + 0,08]\right) \geq 0,95.$$

- Conclusion : l'intervalle $[\bar{x}_n - 0,08\ ,\ \bar{x}_n + 0,08]$ est un intervalle de confiance à 95% pour le paramètre inconnu θ.

Modèle statistique (1)

Définition

Un modèle statistique est un couple

$$(\mathcal{H}^n, \mathcal{P}),$$

où $\mathcal{H} \in \mathcal{B}(\mathbb{R}^k)$ *pour un certain* $k \geq 1$ *et* $\mathcal{P}$ *est une famille de mesures de probabilité sur* $(\mathcal{H}^n, \mathcal{B}(\mathcal{H}^n))$

Modèle statistique (2)

Remarque (observations i.i.d.)

Dans le cas d'un processus de sélection par tirage aléatoire avec remise, ou bien par tirage aléatoire sans remise mais avec une population beaucoup plus grande que n, on suppose généralement que

$$\mathcal{P} = \{\mathbb{Q}^{\otimes n}\}_{\mathbb{Q} \in \mathcal{Q}},$$

où $\mathcal{Q}$ est une famille de mesures de probabilités sur $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$. Ceci correspond à l'hypothèse que les caractéristiques des individus de l'échantillon sont i.i.d. Dans l'exemple du jeu de pile ou face de la section précédente, on avait

$$\mathcal{H} = \{0, 1\} \quad et \quad \mathcal{Q} = \{Bern(\theta)\}_{\theta \in [0,1]}.$$

Modèle statistique (3)

Exemple : premier pile au jeu de pile-ou-face
On considère un joueur de pile-ou-face qui répète n fois l'expérience suivante : il lance sa pièce autant de faois que nécessaire jusqu'à obtenir pile une fois, et il note le nombre de tirages effectués. Si on suppose (comme dans la section précédente) que les résultats des lancers de pile-ou-face sont i.i.d., on obtient le modèle statistique suivant :

$$\left((\mathbb{N}^*)^n, \left\{\mathrm{Geom}(\theta)^{\otimes n}\right\}_{\theta\in[0,1]}\right).$$

Modèle statistique (4)

Définition

Le modèle statistique $(\mathcal{H}^n, \mathcal{P})$ *est un modèle statistique paramétré par* Θ *si* $\mathcal{P} = \{\mathbb{P}_\theta\}_{\theta \in \Theta}$.
C'est un modèle statistique paramétrique si $\Theta \subset \mathbb{R}^d$ *pour un certain* $d \geq 1$.
Sinon, le modèle est non-paramétrique.

Modèle statistique (5)

Définition (identifiabilité)

Le modèle statistique paramétré $(\mathcal{H}^n, \{\mathbb{P}_\theta\}_{\theta\in\Theta})$ *est* *identifiable* *si* $\theta \mapsto \mathbb{P}_\theta$ *est injective sur* Θ.

Exemple : Le modèle stastistique gaussien

$$\left(\mathbb{R}^n, \{\mathcal{N}(m,\sigma^2)^{\otimes n}\}_{m\in\mathbb{R},\sigma\in\mathbb{R}}\right)$$

n'est pas identifiable. En revanche, le modèle

$$\left(\mathbb{R}^n, \{\mathcal{N}(m,\sigma^2)^{\otimes n}\}_{m\in\mathbb{R},\sigma\in[0,+\infty)}\right)$$

est identifiable.

Principe fondamental de la statistique (1)

- Supposons que l'on dispose d'une suite infinie d'observations $x_1, x_2, \ldots$ réalisation d'une suite de v.a. $(X_n)_{n\in\mathbb{N}}$ i.i.d. à valeurs dans $\mathbb{R}^k$.
- La reconstruction de la loi des observations à partir des x_i se formule comme la reconstruction d'une loi inconnue $\mathbb{Q}$ sur $\mathbb{R}^k$ à partir de la suite $(X_n)_{n\in\mathbb{N}}$ i.i.d. de loi $\mathbb{Q}$.
- On introduit la mesure empirique

$$Q_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i},$$

où δ_x est la mesure de Dirac en x.
- En vertu de la loi des grands nombres, pour tout $A \in \mathcal{B}(\mathbb{R}^k)$,

$$\lim_{n\to+\infty} Q_n(A) = \lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_A(X_i) = \mathbb{P}(X_1 \in A) = \mathbb{Q}(A) \quad \text{p.s.},$$

- On aimerait pouvoir intervertir la limite et le $\forall A \subset \mathcal{B}(\mathbb{R}^k)$.

Principe fondamental de la statistique (2)

Théorème (Varadarajan)

Avec probabilité 1, la suite de mesures Q_n converge étroitement vers $\mathbb{Q}$:

$$Q_n \Rightarrow \mathbb{Q} \quad \textit{quand } n \to +\infty \textit{ p.s.}$$

Ce résultat garantit que le problème de reconstruction de la loi $\mathbb{Q}$ est en théorie possible presque sûrement si on dispose d'une suite infinie d'observations i.i.d.