1 Modèle statistique

Définition

$(\mathbb{X},\mathfrak{F},\mathbb{P})$ espace probabilisé décrivant une expérience aléatoire.
$(\mathbb{X},\mathfrak{F},\mathcal{P})$ modèle statistique décrivant une famille d'expériences aléatoires.
- $\mathbb{X}$ espace des observations.
- $\mathfrak{F}$ tribu sur cet espace.
- $\mathbb{P}$ mesure de probabilité.
- $\mathcal{P}$ famille de mesures de probabilité.

On cherche $\mathbb{P} \in \mathcal{P}$ qui modélise au mieux les observations et prend en compte notre connaissance du phénomène aléatoire. Dépend souvent d'un paramètre inconnu θ qui caractérise la loi. Trouver $\mathbb{P}_\theta$ est alors équivalent à trouver la valeur de θ.

Définition

Variable aléatoire (VA) c'est une fonction mesurable $X : \mathbb{X} \longrightarrow \mathbb{R}$.
Mesurable signifie $\forall B \in \mathfrak{B}_{\mathbb{R}}$, $X^{-1}(B) \in \mathfrak{F}$.

Ex: tirage de Bernoulli sur $[0,1]$.
Ex: loi uniforme sur $[0,\theta]$.
Ex: loi de Poisson.
Ex: loi équirépartie.

Définition

n-échantillon: $(X_1,...,X_n)$ n v.a.i.i.d.
Modèle d'échantillonnage associé:

$$\left(\mathbb{X}^n,\mathfrak{F}^{\otimes n},\{\mathbb{P}^{\otimes n},\mathbb{P}\in\mathcal{P}\}\right)$$

$\mathbb{P}^{\otimes n}(dx_1,...,dx_n) = \prod_{i=1}^n \mathbb{P}(dx_i)$ loi jointe des X_i.

Ex: répétition de n tirages de Bernoulli.
Ex: répétition de n tirages d'une loi uniforme.

2 Statistique sur un modèle

Définition

C'est une application mesurable S sur $(\mathbb{X},\mathfrak{F},\mathcal{P})$ à valeurs dans $(\mathbb{Y},\mathcal{G})$ qui ne dépend que des observations et pas du paramètre inconnu θ.

$$S : \mathbb{X} \longrightarrow \mathbb{Y},\ x \mapsto y = S(x)$$

Ex: somme $S = X_1 + ... + X_n$ des éléments d'un échantillon de Bernoulli (loi binomiale).
Ex: $S = \sup(X_1,...,X_n)$ des éléments d'un échantillon de loi uniforme.

3 Modèle paramétrique et homogène

Définition

Chaque mesure de probabilité $\mathbb{P}_\theta \in \mathcal{P}$ est caractérisée par un paramètre $\theta \in \Theta \subset \mathbb{R}^d$. La fonction $\Lambda : \Theta \longrightarrow \mathcal{P}$ est alors bijective.

Ex: tous les exemples précédents.
Ex: échantillon gaussien $\mathcal{N}(m,\sigma^2)$. Ici, $\theta = (m,\sigma)$.
Cex: ensemble des lois à densité continue.

Définition

- Modèle homogène si le support des mesures ne dépend pas de θ.
- Modèle dominé si toutes les mesures de $\mathcal{P}$ possèdent une densité par rapport à une mesure commune.

On pourra toujours considérer qu'un modèle continu est dominé par la mesure de Lebesgue $\lambda(dx) = dx$ et qu'un modèle discret est dominé par la mesure de comptage $\delta_x(B) = 1 \iff x \in B$.

4 Vraisemblance

Pour la définir convenablement, il est nécessaire d'avoir une mesure dominée et de connaître le théorème de Radon-Nikodym.

Définition

$\mathbb{P}_\theta(dx) = f(x)\,dx = L(x,\theta)\,dx$
f densité (par rapport à la mesure dominante).
$L(x,\theta)$ vraisemblance: c'est la densité, mais vue comme fonction de θ:

$$L(x,\theta) = \frac{\mathbb{P}_\theta(dx)}{\lambda(dx)}$$

Ex: pour un modèle à densité continue par rapport à la mesure de Lebesgue, $L(x,\theta) = f(x)$.
Ex: pour une loi discrète, $L(x,\theta) = \mathbb{P}_\theta[X = x]$.
Ex: pour un échantillon discret: $L(x_1,...,x_n,\theta) = \mathbb{P}([X_1 = x_1] \cap ... \cap [X_n = x_n]) = \prod_{i=1}^n \mathbb{P}[X_i = x_i]$

5 Modèle exponentiel

Définition

C'est un modèle paramétrique, dominé, homogène dont la vraisemblance s'écrit

$$L(x,\theta) = h(x)\, e^{\langle \lambda(\theta), S(x)\rangle - \beta(\theta)}$$

- h fonction mesurable positive.
- $\lambda : \Theta \to \mathbb{R}^k$ paramètre canonique.
- $S : \mathbb{X} \to \mathbb{R}^k$ statistique naturelle.
- $\beta : \mathbb{R}^p \to \mathbb{R}$ fonction de log-partition.

$\lambda \bullet S = \langle \lambda(\theta), S(x)\rangle = \sum_{i=1}^{k} \lambda_i(\theta) \times S_i(x)$
Ex: échantillon ou variable de Bernoulli.
Ex: échantillon ou variable uniforme.
Ex: échantillon multinomial.
Ex: échantillon gaussien.
Ex: échantillon de Poisson, binomial, etc.

6 Modèle bayésien

Définition

C'est un modèle statistique $(\mathbb{X}, \mathfrak{F}, \mathcal{P})$ paramétrique muni d'une loi de probabilité π sur l'espace Θ des paramètres.
- π s'appelle la loi *a priori*.
- θ est la réalisation d'une variable aléatoire que l'on notera H et dont la loi est π.
- $L(x|\theta)$ loi conditionnelle de X sachant H.
- $L(\theta|x)$ loi *a posteriori*.

D'après la formule de Bayes,

Théorème

$L(\theta|x) \propto L(x|\theta) \times \pi(\theta)$

Ex: échantillon gaussien avec *a priori* gaussien.
Ex: échantillon de Bernoulli avec *a priori* béta.