Chapitre 1

Généralités-Maximum de Vraisemblance
Familles Exponentielles

I.1. Hypothèses sur le modèle

Soient $\mathcal{E}$ une expérience aléatoire,

$(\Omega, \mathcal{A}, P)$ l'espace de probabilité associé à $\mathcal{E}$,

X une variable aléatoire (v.a.) associée définie sur $(\Omega, \mathcal{A})$ à valeurs dans $\mathcal{X} = \mathbb{R}$ ou $\mathbb{R}^d$,

$(X, \mathfrak{B})$ l'espace des observations,

P_θ la loi de probabilité de X, caractérisée par un paramètre $\theta \in \Theta$,

Θ l'espace des paramètres (sous-ensemble mesurable de $\mathbb{R}^P$),

$(X, \mathfrak{B}, P^X)$ le modèle image par X du modèle $(\Omega, \mathcal{A}, P)$.

Définition 1 Modèle d'échantillonnage

On appelle modèle d'échantillonnage de taille n, le produit $(\mathcal{X}, P_\theta)^{\otimes n}$ associé à n expériences aléatoires indépendantes et de même loi P.

On suppose que P appartient à une famille de probabilités $\mathcal{P} = \{P_\theta$ (ou f_θ), $\theta \in \Theta\}$. (On dit que le modèle est paramétrique).

Notations : - X_i est la v.a. de même loi que X associée à l'expérience numéro « i ».

- x_i est la réalisation de X_i.

- Au modèle $(\mathcal{X}, P_\theta)^{\otimes n}$ correspondent n v.a.i.i.d. $X_1, \ldots, X_n$ de loi P_θ.

Définition 2 Echantillon (aléatoire)

- On appelle échantillon de taille n ou n-échantillon, le n-uple $(X_1, \ldots, X_n)$ où les X_i sont n v.a.i.i.d.
- Le n-uple $(x_1, \ldots, x_n)$ est appelé échantillon d'observations (ce sont n réalisations indépendantes d'une v.a. X).

Définition 3 Modèle paramétrique

a- Un modèle statistique $(\Omega, \mathcal{A}, P)_{P \in \mathcal{P}}$, est une famille de lois de probabilités sur un espace mesuré d'observations.

b- Un modèle (statistique) paramétrique $(\Omega, \mathcal{A}, P_\theta)_{\theta \in \Theta}$, consiste en l'observation d'une v.a. X de loi $P_\theta(x)$ (ou $f_\theta(x)$), où seul le paramètre θ est inconnu.

Exemples $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \{\mathcal{N}(0,\sigma^2)\}_{\sigma \in \mathbb{R}^{+*}})$, $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \{\mathcal{U}[a,b]\}_{(a,b) \in \mathbb{R}^2})$.

Le but de l'estimation paramétrique est d'identifier la probabilité inconnue P_θ (ou f_θ), ou bien le paramètre inconnu θ à partir de la réalisation d'un échantillon. On étudie essentiellement des modèles d'échantillonnage où la famille de probabilités $\mathcal{P}$ est constituée de lois possédant une densité ($P(x,\theta) = f_\theta(x)$), ou de lois discrètes ($P(x,\theta) = P_\theta(X=x)$).

Notation $E_\theta(f(X_1, \ldots, X_n))$ désigne l'espérance de $f(X_1, \ldots, X_n)$ où le vecteur $(X_1, \ldots, X_n)$ a pour loi $P_\theta^{\otimes n}$.

I.2. Statistique, Estimateur

Définition 4 Une **statistique** S est une v.a. fonction de $(X_1, \ldots, X_n)$, indépendante de la loi P de X. ($S : (\Omega, \mathcal{A}) \longrightarrow (\mathcal{X}, \mathfrak{B})$)

(*Si cette fonction est utilisée pour évaluer un ou plusieurs paramètres inconnus de la loi de X, elle est appelée* ***estimateur****, et ses réalisations sont des* ***estimations***.)

Exemples Les fonctions $S_1(X) = \sum_{i=1}^n X_i$, $S_2(X) = \max_{1\le i\le n} X_i$, $S_3(X) = X_1$ et

$S_4(X) = (h(X_3-X_1), X_2)$, sont des statistiques de l'échantillon $X = (X_1, \ldots, X_n)$.

Le premier objectif est d'estimer le paramètre θ. Plus généralement, on peut chercher à estimer une fonction $g(\theta) \in \mathbb{R}^P$ de ce paramètre.

Définition 5

a- Un **estimateur** δ de $g(\theta)$ est une fonction de l'échantillon X (indépendante de $P \in \mathcal{P}$) à valeurs dans $g(\Theta)$ (*c'est une fonction des v.a. observables ne dépendant pas des paramètres inconnus*).

b- Estimation : c'est la valeur prise par l'estimateur sur l'échantillon observé ($x_1, \ldots, x_n$).

Exemple La moyenne empirique $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ est un estimateur de E(X).

Définition 6 Convergences

a- Convergence forte : Une suite d'estimateurs $(\hat{g}_n)_{n\in N^*}$ de $g(\theta)$ (où $\hat{g}_n$ est une fonction de l'échantillon de taille n $(X_1, \ldots, X_n)$) est fortement convergente si

$$\forall \theta \in \Theta, \lim_n \hat{g}_n = g(\theta), \ P_\theta\text{-presque sûrement.}$$

Remarque Pour simplifier, on confondra estimateur et suite d'estimateurs.

Exemples

E₁. Si X est intégrable ($E(|X|) < +\infty$) alors $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{PS} E(X)$ c'est la loi forte des grands nombres

E₂. Dans le modèle d'échantillonnage exponentiel $\mathcal{P} = \{\mathcal{E}(\lambda), \lambda > 0\}$ les estimateurs

$\hat{\lambda}_1 = \dfrac{n}{\sum_{i=1}^{n} X_i}$ et $\hat{\lambda}_2 = \dfrac{1}{\sqrt{\dfrac{1}{2n}\sum_{i=1}^{n} X_i^2}}$ sont 2 estimateurs convergents (presque sûrement) de λ.

Ceci découle de la loi forte des grands nombres et le théorème de la continuité pour la convergence presque sûre suivant :

Lemme : Théorème de continuité pour la convergence presque sûre

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de v.a. qui converge presque sûrement (PS) vers X.

Soit g une fonction mesurable et $\mathcal{C}$ l'ensemble des points de continuité de g.

Si $P(X \in \mathcal{C}) = 1$, alors $g(X_n) \xrightarrow{PS} g(X)$.

I.3. Construction d'estimateurs convergents

I.3.3. Le maximum de vraisemblance

Soit $x = (x_1, x_2, ..., x_n)$ une réalisation de l'échantillon $(X_1, X_2, \ldots, X_n)$ d'une v.a. X. On suppose que X suit une loi de probabilité dépendant d'un paramètre θ. On note

$$L(\theta, x_1, ..., x_n) = \begin{cases} \prod_{i=1}^{n} f_\theta(x_i) & \text{si } X \text{ est absolument continue} \\ \prod_{i=1}^{n} P_\theta(X = x_i) & \text{si } X \text{ est discrète} \end{cases}$$

D**éfinition 8** **a-** L'application $x = (x_1, x_2, ..., x_n) \longmapsto L(\theta, x)$ est appelée **loi conjointe** de x.

b- L'application $\theta \longmapsto L(\theta, x)$ est appelée **vraisemblance** de θ.

Remarques

1. De la fonction $L(\theta, x_1, x_2, ..., x_n)$ on peut définir la variable aléatoire $L(\theta, X_1, X_2, ..., X_n)$.

2. Les observations x_i jouent le rôle de paramètre de la vraisemblance, c'est-à-dire que la vraisemblance n'est définie qu'après l'observation des réalisations de la variable X ; La vraisemblance est donc une notion statistique, alors que la loi conjointe est une notion probabiliste.

D**éfinition 9** **Estimateur du Maximum de Vraisemblance**

Si pour tout $x \in \mathcal{X}^n$, il existe une et une seule valeur de $\theta \in \Theta$ (notée $\hat{\theta}_{MV}$) telle que la vraisemblance soit maximale, alors la v.a. $\hat{\theta}_{MV} = \hat{\theta}_{MV}(X_1, ..., X_n)$ est appelée estimateur du maximum de vraisemblance (en abrégé EMV) de θ.

Principe Si L est deux fois dérivable par rapport à θ, on peut obtenir $\hat{\theta}_{MV}$ en résolvant le système :

$$\begin{cases} \dfrac{\partial L}{\partial \theta} = 0 & (\text{équation de vraisemblance}) \\ \dfrac{\partial^2 L}{\partial \theta^2} \prec 0 \end{cases}$$

Définition 10 La Log - Vraisemblance est le logarithme de la vraisemblance L(θ, *x*), c'est-à-dire la fonction :

$$\theta \to Log\big(L(\theta, x)\big) = \sum_{i=1}^{n} Log\left(f_\theta(x_i)\right) \left(= \sum_{i=1}^{n} Log\left(P_\theta(X = x_i)\right)\right).$$

Remarque Maximiser la Log-vraisemblance revient à maximiser la vraisemblance. Les calculs sont parfois plus simples pour la Log-vraisemblance.

Proposition (Propriété d'invariance de l'EMV)

Soit $\hat{\theta}_{MV}$ l'EMV de θ associé à la famille de lois $\mathcal{P} = \{P_\theta\ (f_\theta), \theta \in \Theta\}$.

Soit $g : \Theta \longrightarrow g(\Theta)$ mesurable bijective. Alors $g(\hat{\theta}_{MV})$ est l'EMV de g(θ) associé à la famille $\mathcal{P}_g = \{P_{g^{-1}(\beta)}, \beta \in g(\Theta)\}$

Remarque On dit que $g(\hat{\theta}_{MV})$ est l'EMV de g(θ) même si g n'est pas bijective.

Exemple Pour le modèle de lois exponentielles $\mathcal{P} = \{\mathcal{E}(\lambda), \lambda > 0\}$, la Log-vraisemblance est : $Log\, L(\lambda, x) = -\lambda \sum_{i=1}^{n} x_i + nLog\lambda \implies$

L'EMV de λ est $\hat{\lambda}_n = \dfrac{n}{\sum_{i=1}^{n} X_i}$, et l'EMV de $\theta = \dfrac{1}{\lambda}$ est $\hat{\theta}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.

I.4. Choix d'un estimateur

I.4.1. Comparaison d'estimateurs

Soit une famille paramétrique de lois $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, et $X_1, \ldots, X_n$ n v.a.i.i.d. de loi $P_\theta \in \mathcal{P}$. Soit δ un estimateur de g(θ). Il est naturel de minimiser le risque d'erreur, qui est une mesure de la qualité des estimations envisagées.

Définition 12

- L'estimateur δ (de g(θ)) est dit sans biais, s'il est intégrable et si $E_\theta(\delta) = g(\theta)$, $\forall\, \theta \in \Theta$.
- Le biais d'un estimateur intégrable δ de g(θ) est défini par : $b(\delta) = E_\theta(\delta) - g(\theta)$, $\forall\, \theta \in \Theta$.

Exemples La moyenne empirique $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ est un estimateur sans biais de E(X).

La variance empirique $S_n^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ n'est pas un estimateur sans biais de Var(X).

Définition 13 Pour un estimateur δ (de g(θ)), on définit le risque quadratique ou écart (erreur) quadratique moyen (Mean Square Error ou MSE, en anglais) par : R(δ,θ) ou

$$R_\theta(\delta) = E_\theta(\delta - g(\theta))^2.$$

Propriété L'erreur quadratique moyenne de δ se décompose en deux termes, le carré du biais et la variance de δ :

$$\mathbf{R_\theta(\delta)} = [\, E_\theta(\delta) - g(\theta)\,]^2 + E_\theta[\, \delta - E_\theta(\delta)\,]^2 = \mathbf{[\, b(\delta)\,]^2 + Var_\theta(\delta).}$$

I.5. Amélioration d'estimateurs

On considère un modèle régulier de taille n.

I.5.3. Familles exponentielles (ou modèles exponentiels)

I.5.3.1. Famille exponentielle d'ordre 1

Définition 26 Une famille de lois $\mathcal{P}=\{P_\theta, \theta\in\Theta\}$.est une famille exponentielle d'ordre 1 si :

a- Le support Δ de X ne dépend pas de θ,

b- La loi de X sur Δ est de la forme

$$f_\theta(x) = exp\{a(x)\alpha(\theta) + b(x) + \beta(\theta)\} = h(x).c(\theta)exp\{a(x)\alpha(\theta)\} \qquad \textbf{(I)}$$

où α(.), β(.) et c(.) ne dépendent que de *θ*,

a(.), b(.) et h(.) ne dépendent que de *x*.

Exemples

E$_3$. Les familles des lois de Bernoulli $\mathcal{B}$(p), gaussiennes $\mathcal{N}(0, \sigma^2)$, σ > 0, gamma (α, λ_0), λ_0 connu, correspondent à des familles exponentielles :

a- X↝ $\mathcal{B}$(p) $\Rightarrow$ $P(X = x) = p^x(1-p)^{1-x} = \exp\left\{x.Ln\left(\frac{p}{1-p}\right) + Ln(1-p)\right\}$, *x* = 0,1

ici *a(x) = x, α(p) = Ln(p)-Ln(1-p), b(x) = 0, β(p) = Ln(1-p).*

b- X↝ (0, σ²) $\Rightarrow$

c- X ↝ $\mathcal{G}$amma (α, λ_0) $\Rightarrow$

E$_4$. La famille de loi de Cauchy (θ, 1) n'est pas exponentielle, en effet

$$f_\theta(x) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2} = \frac{1}{\pi}exp\{-Ln[1+(\theta-x)^2]\}$$

avec $-Ln[1+(\theta-x)^2] \neq a(x).\alpha(\theta)$ car la fonction logarithme n'est pas linéaire.

I.5.3.1. Généralisation

Soient **i /** Y∈ $\mathbb{R}^p$ un vecteur aléatoire, X=(X_1,…, X_n) ∈ $\mathbb{R}^{np}$ un n-échantillon de Y, et θ∈ $\mathbb{R}^k$,

ii / le support Δ de Y ne dépend pas de θ.

Définition 27 (Définition exacte d'une famille exponentielle)

La famille des lois P_θ caractérisée par une densité de la forme (II),

$$f_\theta(y) = \exp\left\{\sum_{i=1}^{k} a_i(y)\alpha_i(\theta) + b(y) + \beta(\theta)\right\} = h(y).c(\theta)\exp\left\{\sum_{i=1}^{k} a_i(y)\alpha_i(\theta)\right\} \qquad \textbf{(II)}$$

sur un ensemble Δ indépendant de θ (et nulle en dehors de Δ), est appelée famille exponentielle.

Remarque On peut réécrire une famille exponentielle comme suit :

$$f_\theta(x) = h(x).c(\theta)exp\{T(x)Q(\theta)\}$$

T et Q sont des fonctions respectivement de $\mathcal{X}$ et Θ dans $\mathbb{R}^k$.

Définition 29 Dans le cas particulier où $\mathcal{X} \subset \mathbb{R}^k$ et $\Theta \subset \mathbb{R}^k$, et

$$f_\theta(x) = h(x).c(\theta)\exp\{\theta.x\} \qquad \textbf{(III)}$$

la famille est dite (exponentielle) naturelle.

Remarque

R$_{10}$. Un changement de variable de X en T(X) et une re-paramétrisation de θ en $\eta = Q(\theta)$ nous permettent de considérer principalement la forme naturelle **(III)**, bien que les espaces T($\mathcal{X}$) et Q(Θ) soient difficiles à décrire et à utiliser.

Définition 30 (Famille régulière)

Soit une famille exponentielle naturelle. Elle est dite régulière si l'espace naturel des paramètres $\mathcal{N}$ est un ensemble ouvert ($\mathcal{N}=\{\theta \,/\, \int_{\mathcal{X}} e^{\theta x} h(x) dP_X(x) \prec \infty\}$).

Propriété Les familles exponentielles naturelles peuvent être réécrites

$$f_\theta(x) = h(x).c(\theta)e^{\theta x} = h(x)e^{\theta x - \varphi(\theta)}$$

où $\varphi(\theta) = -Ln\, c(\theta)$ est dite fonction cumulante des moments.

Lemme Si θ appartient à l'intérieur de $\mathcal{N}$, alors la fonction φ est $\mathcal{C}^\infty$ et :

$$E_\theta(X) = \nabla\varphi(\theta) \left(= \frac{\partial}{\partial\theta_i}\varphi(\theta) \right)$$ (∇ est l'opérateur gradient),

et $$Cov(X_i, X_j) = \frac{\partial^2\varphi(\theta)}{\partial\theta_i\partial\theta_j}.$$

Exemples

E$_5$. X $\rightsquigarrow \mathcal{P}(\lambda) \Longrightarrow P_\lambda(X = x) = e^{-\lambda}\dfrac{\lambda^x}{x!} =$

E$_6$. X $\rightsquigarrow \mathcal{B}(n, \theta) \Longrightarrow P_\theta(X = x) = C_n^x \theta^x (1-\theta)^{n-x} =$

Proposition On considère une famille exponentielle définie par :

$$f_\theta(x) = h(x).c(\theta)e^{\theta T(x)} = h(x)e^{\theta T(x) - \varphi(\theta)}$$

Si θ appartient à l'intérieur de $\mathcal{N}$, alors **$E_\theta(T(X)) = \varphi'(\theta)$**

et **$Var_\theta(T(X)) = \varphi''(\theta)$**.

De plus T(X) est un estimateur efficace de $E_\theta(T(X)) = g(\theta)$, c'est-à-dire que $Var_\theta(T(X)) = \varphi''(\theta)$ est la borne de Cramer-Rao de $g(\theta)$ (égale à $\dfrac{(g'(\theta))^2}{I(\theta)}$).