Chapitre 2

Exhaustivité et Information

2.1 Introduction

Nous abordons avec ce chapitre deux outils de base de la Statistique Mathématique. Après avoir défini précisément ce que nous entendons par le mot "statistique", nous introduisons le concept d'information. En effet, puisque le statisticien est souvent placé dans une situation de décision, il va avoir besoin de l'information apportée par un échantillon. S'il peut avoir la même information avec un deuxième échantillon de taille inférieure, il optera pour cette solution, simple question de coût.

Il est donc naturel de tenter de réduire les données, c'est-à-dire d'essayer de représenter l'information contenue dans un échantillon par une fonction $T(x_1, ..., x_n)$ que l'on appellera une *statistique*. Cela nous conduira à donner un sens au mot "information".

Comment savoir si la réduction des données opérée par la statistique T n'a pas fait perdre l'information ?. C'est ce type de problème que cherche à résoudre la notion d'exhaustivité.

2.2 Exhaustivité

2.2.1 Modèles statistiques

Définition 1 : On appelle *structure statistique* ou *modèle statistique* le triplet $(\mathcal{X}, \mathcal{B}, \mathcal{P})$ où :

$\cdot\mathcal{X}$: ensemble des observations ($\equiv$ensemble des valeurs de la v.a X)

$\cdot\mathcal{B}$: tribu des parties de $\mathcal{X}$.

· $\mathcal{P}$: famille de lois de proba de X.

Définition 2 : Un modèle statistique est paramétrique $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\})$.

Définition 3 : Lorsque la famille des lois $\{P_\theta, \theta \in \Theta\}$ admettent des densités $f_\theta(x)$, leurs modèles correspondant sont dit *dominés :* $(\mathcal{X}, \mathcal{B}, \{f_\theta, \theta \in \Theta\})$

Exemples : Donner le modèle statistique associé à :

1°. une observation d'une v.a $X \rightsquigarrow \mathcal{P}(\lambda)$.

2°. Un échantillon, $(X_1, ..., X_n)$, généré à partir d'une v.a entière X.

2.2.2 Statistique

Soit X une variable aléatoire dans $(\mathcal{X}, a)$ et $(\mathcal{Y}, B)$ un espace mesurable auxiliaire quelconque.

Définition 4 : On appelle statistique toute application T mesurable de $\mathcal{X}^n$ dans $\mathcal{Y}$:

$\forall n, \quad T : \mathcal{X}^n \to \mathcal{Y} \subset \mathcal{R}^p.$

$(x_1, ..., x_n) \to T(x_1, ..., x_n)$

Remarques : 1°.Une statistique n'est rien d'autre qu'une fonction mesurable des observations, autrement dit, est un résumé d'un échantillon.

2°.Si $p = 1$, T est dite *statistique réelle.*

Si $p > 1$, T est dite *statistique mesurable.*

3°. T ne dépend jamais du paramètre.

Exemples : 1). $\mathcal{X} = \mathbb{R}, \mathcal{Y} = \mathbb{R}$:

$$(x_1, ..., x_n) \to T(x_1, ..., x_n) = \sum_{i=1}^{n} x_i.$$

$$(x_1, ..., x_n) \to T(x_1, ..., x_n) = \sum_{i=1}^{n} x_i^2.$$

2) $\mathcal{X} = \mathbb{R}, \mathcal{Y} = \mathbb{R}^n$: $(x_1, ..., x_n) \to T(x_1, ..., x_n) = (x_{(1)}, x_{(2)}, ..., x_{(n)})$

où $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$.

Cette statistique porte le nom de *statistique d'ordre (échantillon ordonné).*

2.2.3 Statistique exhaustive

Définition 5 : Soit le modèle $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$. La statistique T sera dite *exhaustive* si $P(X = x/T = t)$ est indépendante du paramètre θ.

Remarque : x désigne soit l'observation x soit l'échantillon $(x_1, ..., x_n)$.

Exemple : X suite une loi de Bernoulli $\mathcal{B}(p)$. Montrer que $T = \sum_{i=1}^{n} X_i$ est une statistique exhaustive.

2.2.4 Critère de factorisation

Notation : On notera $\mathcal{L}(x_1, ..., x_n, \theta) = \mathcal{L}(\underline{x}, \theta)$ la densité de $(X_1, ..., X_n)$ si X est absolument continue ou bien la probabilité conjointe de cet échantillon si X est une v.a.d. Elle est appelée *vraisemblance* de θ :

· vraisemblance=Likelihood=$\mathcal{L}$

$$\cdot\mathcal{L}(x_1, ..., x_n, \theta) = \prod_{i=1}^{n} f(\theta, x_i) \quad \text{si } X_i \rightsquigarrow f(\theta, x_i)$$

$$= \prod_{i=1}^{n} P(X = x_i) \quad \text{si } X_i \rightsquigarrow P(\theta, x_i).$$

Théorème (Fisher-Neyman) Soit $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$ un modèle statistique paramétrique et T une statistique sur $\mathcal{X}$:

T exhaustive$\Leftrightarrow \mathcal{L}(\underline{x}, \theta) = g(T(x), \theta) \times h(x)$.

Exemples : Trouver une statistique exhaustive pour :

1°. n échantillon de $X \rightsquigarrow \mathcal{B}(p)$.

2°. n échantillon de $X \rightsquigarrow \mathcal{P}(\lambda)$.

3°. n échantillon de $X \rightsquigarrow \mathcal{N}(m, \sigma^2)$.

4°. n échantillon de $X \rightsquigarrow \mathcal{U}_{[0,\theta]}$.

2.2.5 Cas de la famille exponentielle

Théorème Soit X une v.a de densité $f(x,\theta)$ t.q le support est indépendant de θ. Pour qu'il existe une statistique exhaustive pour θ il suffit que $f(x,\theta)$ s'écrit sous la forme suivante :

$$f(x,\theta) = c(\theta)\exp\left\{\sum_{j=1}^{k}\alpha_j(\theta)\,a_j(x)\right\}h(x)$$

La statistique exhaustive est donnée par : $T(x) = \sum_{i=1}^{n} a(x_i)$.

Dans le cas d'un échantillon indépendant de taille n :

$$f(\underline{x},\theta) = c^n(\theta)\exp\left\{\sum_{j=1}^{k}\alpha_j(\theta)\sum_{i=1}^{n}a_j(x_i)\right\}h(x)$$

La statistique exhaustive est alors $T(\underline{x}) = \left(\sum_{i=1}^{n} a_1(x_i), \sum_{i=1}^{n} a_2(x_i), ..., \sum_{i=1}^{n} a_k(x_i)\right)$.

Exemples : Trouver une statistique exhaustive pour :

1°. n échantillon de $X \rightsquigarrow \mathcal{B}(p)$.

2°. n échantillon de $X \rightsquigarrow \mathcal{N}(m,\sigma^2)$.

2.3 Elements de théorie de l'information

2.3.1 Information au sens de Fisher (Cas réel)

Soit le modèle statistique $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta \subset \mathbb{R}\})$. On fait les hypothèses suivantes (dites de régularité) :

$\cdot H_1$: le support de $f(x,\theta)$ est indépendant de θ.

$\cdot H_2$: $\forall x, \forall \theta$, $f(x,\theta)$ est dérivable au moins deux fois par rapport à θ.

$\cdot H_3$: On peut dériver au moins deux fois $\int_A f(x,\theta)\,dx$ par rapport à θ sous le signe d'intégration, pour tout $A \in \mathcal{A}$.

Définition 6 : On appelle quantité d'information de Fisher $I_n(\theta)$ apporter par un n-éch sur le paramètre θ, la quantité suivante positive ou nulle (si elle $\exists$) :

$$I_n(\theta) = E\left(\frac{\partial}{\partial\theta}\log\mathcal{L}(x,\theta)\right)^2$$

$I(\theta) = E\left(\frac{\partial}{\partial\theta}\log f(x,\theta)\right)^2$ pour une réalisation

Définition 7 : On appelle score la quantité $S(x,\theta) = \frac{\partial}{\partial\theta}\log f(x,\theta)$ donc $I(\theta) = E(S^2)$

Théorème 1 : Si le support de x ne dépend pas de θ, alors : pour n-éch: $S_n(x,\theta) = \frac{\partial}{\partial\theta}\log L(x,\theta)$

$$I_n(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\log\mathcal{L}(x,\theta)\right) = -E\left(\frac{\partial}{\partial\theta}S(x,\theta)\right)$$

Propriétés de $I(\theta)$:

a/. **positivité :** $I_n(\theta) = V(S) \geq 0$.

b/. **additivité :** Soient les modèles statistiques $(\mathcal{X}, P_\theta)$ et $(\mathcal{Y}, Q_\theta)$, et X et Y deux variables aléatoires indépendantes à valeurs dans les espaces $\mathcal{X}$ et $\mathcal{Y}$. On note :

$\cdot f(x,\theta)$ la densité de X. $\cdot f(y,\theta)$ la densité de Y.

$\cdot I_X(\theta), I_Y(\theta)$ et $I_{(X,Y)}(\theta)$ les informations au point θ fournis respect par X, Y et (X,Y).

Théorème 2 : sous les hypothèses H_1, H_2, H_3 : $I_{(X,Y)}(\theta) = I_X(\theta) + I_Y(\theta)$.

Corollaire : Soit un n-éch issu du modèle $(\mathcal{X}, P_\theta)$. Alors, sous les hypothèses H_1, H_2, H_3 :

$$I_n(\theta) = I_{X_1,\dots,X_n} = I_{X_1}(\theta) + I_{X_2}(\theta) + \dots + I_{X_n}(\theta) = nI(\theta).$$

où $I(\theta)$ est l'information apportée par une réalisation x de X.

Propriétés de la fonction score:

Soit $S_n(\theta)$ le score sur θ apporté par un échantillon iid de X, $(X_1, \dots, X_n)$, les hypothèses H_1, H_2 et H_3 vérifiées : $\frac{1}{n} S_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} S(x_i, \theta)$

Théorème 3 :

$$\frac{1}{n} S_n(\theta) \overset{PS}{\to} 0.$$

Démonstration : en cours.

Théorème 4 :

$$\frac{1}{\sqrt{n}} S_n \to \cancel{\mathcal{N}(0,\theta)}. \quad \mathcal{N}(0, I(\theta))$$

Exemples : Donner l'information au sens de Fisher apportée par :

1°. n échantillon de $X \rightsquigarrow \mathcal{P}(\lambda)$.

2°. une réalisattion de $X \rightsquigarrow \mathcal{U}_{[0,\theta]}$.

3°. n échantillon de $X \rightsquigarrow \mathcal{U}_{[0,\theta]}$.

2.3.2 Information au sens de Fisher (Cas vectoriel)

Soit X une v.a de loi P_θ et de densité f_θ dont $\theta = (\theta_1, \dots, \theta_k)$. On fait les hypothèses suivantes :

$\cdot H'_1$: $f(x, \theta) > 0, \forall x, \forall \theta$.

$\cdot H'_2$: grad de f existe, $\forall x, \forall \theta$, c'est-à-dire que l'on peut dériver f par rapport à $\theta_i, \forall i = 1, \dots, k$.

$\cdot H'_3$: on peut dériver au moins deux fois $\int_A f(x, \theta)\, dx$ par rapport à $\theta_i, \forall i = 1, \dots, k, \forall A \in \mathcal{A}$.

Remarques : 1°. Le support de X est indépendant de θ.

2°. grad $f = \left(\frac{\partial f}{\partial \theta_1}, \dots, \frac{\partial f}{\partial \theta_q} \right) = S(X, \theta)$.

Définition 8 : On appelle *matrice d'information de Fisher* la matrice :

$\cdot I(\theta) = var\ Grad_\theta \left[\log f(X,\theta)\right]$ si on a une réalisation de la v.a X.

$\cdot\ I_n(\theta) = var\ Grad_\theta \left[\log \mathcal{L}(X,\theta)\right]$ si elle existe.

Remarques : 1°. Le vecteur $Grad_\theta(\log \mathcal{L}(X,\theta))$ est P_θ centré :

$$E\left(Grad_\theta\left(\log \mathcal{L}(X,\theta)\right)\right) = 0 \Leftrightarrow \begin{cases} E\left(\dfrac{\partial}{\partial\theta_1}\log \mathcal{L}(X,\theta)\right) = 0 \\ \vdots \\ E\left(\dfrac{\partial}{\partial\theta_k}\log \mathcal{L}(X,\theta)\right) = 0 \end{cases}$$

2°. La matrice $I(\theta)$ a pour terme général :

$$I_{i,j}(\theta) = E\left(\frac{\partial}{\partial\theta_i}\log f(x,\theta) \times \frac{\partial}{\partial\theta_j}\log f(x,\theta)\right), 1 \le i \le k, 1 \le j \le k.$$

Autrement dit : $I_{i,j}(\theta) = cov\left(\dfrac{\partial}{\partial\theta_i}\log f(x,\theta), \dfrac{\partial}{\partial\theta_j}\log f(x,\theta)\right)$ qui est une matrice symétrique définie positive.

Théorème 5 : Sous les hypothèses de régularité (dérivation deux fois sous le signe somme) :

$$I(\theta) = -E\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f(x,\theta)\right), 1 \le i \le k, 1 \le j \le k.$$

Exemple : Donner l'information au sens de Fisher apportée par une réalisation de $X \rightsquigarrow \mathcal{N}(m,\sigma^2)$.

2.4 Compléments sur la notion d'exhaustivité

2.4.1 Statistique complète

Définition 9 : Soit $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$ un modèle statistique. On dit que ce modèle est complet ou la famille de loi P_θ est complète si :

$\forall h : \mathcal{X} \rightarrow \mathbb{R}$ vérifiant $E_\theta(h(X)) = 0, \forall\theta \in \Theta \implies P_\theta(h(X) = 0) = 1 \iff h \equiv 0\ P_\theta$ *PS.*

Une statistique T est dite complète si la famille de loi P_θ^T est complète, où :

P_θ^T : la famille de loi de T.

Techniques de calcul :

▷ *première technique* : si $\sum_n a_n x^n = 0, \forall x \in \mathbb{R} \Rightarrow a_n = 0$.

▷ *deuxième technique* : $F(x) = \int_a^x f(t)\,dt = 0, \forall f \in I \Rightarrow f(x) = 0, \forall x \in I$.

Exemples : 1°. Montrer que la famille de loi $B(n,p)$ est complète.

2°. Montrer que $\sum_i X_i$ est complète.où $X \rightsquigarrow B(p)$.

3°. Montrer que $T = \sum_i x_i$ est une stat exh complète tq $X \rightsquigarrow \mathcal{E}(\lambda)$.

Cas de la famille exponentielle :

▷$\theta \in \mathbb{R}$: $f(x,\theta) = c(\theta)\exp\left\{\sum_{j=1}^{k} \alpha_j(\theta)\,a_j(x)\right\} h(x)$: $T(\underline{x}) = \sum_{i=1}^{n} a(x_i)$ est une statistique exhaustive et complète.

▷$\theta \in \mathbb{R}^k$: $T(\underline{x}) = \left(\sum_{i=1}^{n} a_1(x_i), \sum_{i=1}^{n} a_2(x_i), \ldots, \sum_{i=1}^{n} a_k(x_i)\right)$ est une statistique exhaustive et complète si dimension de la statistique=.dimension du paramètre.

2.4.2 Statistique exhaustive minimale

Soit $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$ un modèle statistique.et S une statistique exhaustive. On dira que S est exh minimale s'il existe une fonction mesurable g tel que $S = g(T)$ pour une stat exh q.q T. La statistique obtenue dans la famille des lois exp est minimale.

▷ *troisième technique* : transformé de laplace : soit $f : \mathbb{R} \to \mathbb{R}$ une fonction mesurable et $L_p(t) = \int_{\mathbb{R}} e^{-tx} f(x)\,dx$ la transformé de laplace de la fonction f.

Théorème : Si $L_n(t) = 0, \forall t \in I$ (I : ouvert de $\mathbb{R}$) $\Rightarrow f(x) = 0$ (sauf éventuellement pour un ensemble fini dénombrable).