N° d'ordre : 124

THÈSE

présentée à la

FACULTÉ DES SCIENCES DE L'UNIVERSITÉ DE LILLE

pour obtenir le

Titre de Docteur de Spécialité

(Mathématiques Appliquées)

par

Joseph LOSFELD

*

INFORMATION ET STATISTIQUE

Généralisation de l'Inégalité de Cramer-Rao

Thèse soutenue le 17 Décembre 1968, devant la Commission d'Examen

Monsieur P. POUZET, Président
Monsieur J. VAILLANT, Examinateur
Monsieur BUI-TRONG-LIEU, Examinateur
Mademoiselle S. MARQUET, Rapporteur

FACULTE DES SCIENCES

UNIVERSITE DE LILLE

LISTE DES PROFESSEURS

-oOo-

DOYENS HONORAIRES

MM. H. LEFEBVRE, M. PARREAU.

PROFESSEURS HONORAIRES

MM. ARNOULT, BEGHIN, BROCHARD, CAU, CHAPPELON, CHAUDRON, CORDONNIER, DEHEUVELS, DEHORNE, DOLLE, FLEURY, P. GERMAIN, KAMPE DE FERIET, KOURGANOFF, LAMOTTE, LELONG, Mme LELONG, MM. MAZET, MICHEL, NORMANT, PARISELLE, PASCAL, PAUTHENIER, ROIG, ROSEAU, ROUBINE, ROUELLE, WIEMAN, ZAMANSKY.

PROFESSEURS

BACCHUS P.	Mathématiques appliquées
BEAUFILS J.P.	Chimie
BECART M.	Physique
BONNEMAN P.	Chimie
BLOCH V.	Biologie et Physiologie Animales
BONTE A.	Sciences de la Terre
BOUGHON P.	Mathématiques pures
BOUISSET S.	Biologie et Physiologie Animales
BOURIQUET R.	Biologie végétale
CELET P.	Sciences de la Terre
CONSTANT E.	Electronique, Electrotechnique et Automatique
CORSIN P.	Sciences de la Terre
DECUYPER M.	Mathématiques pures
DEDECKER P.	Mathématiques pures

DEFRETIN R.	Biologie et Physiologie Animales
DEHORS R.	Electronique, Electrotechnique et Automatique
DELATTRE C.	Sciences de la Terre
DELEAU P.	Sciences de la Terre
DELHAYE M.	Chimie
DERCOURT J.M.	Sciences de la Terre
DESCOMBES R.	Mathématiques pures
DURCHON M.	Biologie et Physiologie Animales
FOURET R.	Physique
GABILLARD R.	Electronique, Electrotechnique et Automatique
GLACET C.	Chimie
GONTIER G.	Mathématiques appliquées
HEIM DE BALSAC H.	Biologie et Physiologie Animales
HEUBEL J.	Chimie
HOCQUETTE M.	Biologie végétale
LEBRUN A.	Electronique, Electrotechnique et Automatique
Mlle LENOBLE	Physique
LIEBAERT R.	Electronique, Electrotechnique et Automatique
LINDER R.	Biologie végétale
LUCQUIN M.	Chimie
MARTINOT-LAGARDE A.	Mathématiques appliquées
Mlle MARQUET S.	Mathématiques pures
MONTARIOL F.	Chimie
MONTREUIL J.	Chimie
MORIAMEZ M.	Physique
PARREAU M.	Mathématiques pures
PEREZ J.P.	Physique
PHAM MAU QUAN	Mathématiques pures
POUZET P.	Mathématiques appliquées
PROUVOST J.	Sciences de la Terre
SAVARD J.	Chimie
SCHILTZ R.	Physique
SCHALLER F.	Biologie et Physiologie Animales
Mme SCHWARTZ M.H.	Mathématiques pures
TILLIEU J.	Physique
TRIDOT G.	Chimie

VAILLANT J.	Mathématiques pures
VIVIER E.	Biologie et Physiologie Animales
WATERLOT G.	Sciences de la Terre
WERTHEIMER R.	Physique

MAITRES DE CONFERENCES

BELLET J.	Physique
BENABOU J.	Mathématiques pures
BILLARD J.	Physique
BOILLET P.	Physique
BUI TRONG LIEU	Mathématiques pures
CHERRUAULT Y.	Mathématiques pures
DEVRAINNE P.	Chimie
Mme DRAN R.	Chimie
GOUDMAND P.	Chimie
GUILBAULT P.	Biologie et Physiologie Animales
GUILLAUME J.	Biologie végétale
HUARD DE LA MARRE P.	Mathématiques appliquées
JOLY R.	Biologie et Physiologie Animales
LABLACHE-COMBIER A.	Chimie
LACOSTE L.	Biologie végétale
LANDAIS J.	Chimie
LEHMANN D.	Mathématiques pures
Mme LEHMANN J.	Mathématiques pures
LOUCHEUX C.	Chimie
MAES S.	Physique
MONTEL M.	Physique
PANET M.	Electronique, Electrotechnique et Automatique
PARSY F.	Mathématiques appliquées
RACZY L.	Physique
SAADA G.	Physique
SEGARD E.	Chimie
VIDAL P.	Electronique, Electrotechnique et Automatique
Mme ZINN-JUSTIN N.	Mathématiques pures

INTRODUCTION

L'un des buts de la théorie de l'estimation statistique est la détermination d'un critère de choix entre plusieurs estimateurs d'une même quantité. Une méthode employée pour résoudre cette question utilise la notion d'efficacité introduite à partir de l'inégalité de Cramer - Rao et de la quantité d'information de Fisher. Parmi les conditions de régularité exigées pour obtenir ces résultats, il est nécessaire que la famille de probabilités considérée soit homogène. Dans cet exposé, nous généralisons les techniques employées par S. Kullback dans "Information theory and statistics" ; cela nous permet de définir une quantité d'information de Fisher généralisée et de démontrer l'inégalité de Cramer - Rao sous des conditions moins restrictives que dans le cas classique ; en particulier la famille de probabilités peut être non-homogène.

Dans le premier chapitre nous définissons les notations, nous démontrons des résultats utilisés par la suite et nous énonçons les hypothèses employées.

Dans le chapitre deux nous généralisons au cas non-homogène la notion d'information statistique introduite par S. Kullback dans le cas homogène.

Nous en déduisons, dans le troisième chapitre, la définition d'une quantité d'information de Fisher généralisée. Cette quantité possède les propriétés de la quantité d'information de Fisher classique et ces deux notions se confondent quand les conditions de régularité sont vérifiées. Nous nous inspirons de la démonstration de Chapman et Robbins [1] pour montrer que l'on obtient l'inégalité de Cramer - Rao même dans le cas non-homogène. Nous traitons enfin un exemple où l'on ne peut appliquer les méthodes classiques, mais où cette technique donne des résultats intéressants.

Je tiens à exprimer ma reconnaissance à Monsieur le Professeur Pouzet qui m'a fait l'honneur d'accepter la présidence du jury.

Je remercie vivement Monsieur le Professeur Vaillant qui a bien voulu faire partie de la commission d'examen.

Je souhaite que Monsieur le Professeur Bui-Trong-Lieu trouve ici l'expression de ma gratitude pour les encouragements qu'il m'a prodigués.

Je voudrais exprimer ma profonde reconnaissance à Mademoiselle le Professeur Marquet qui m'a fourni les moyens d'effectuer ce travail dans les meilleures conditions possibles, et dont les conseils m'ont été précieux pour la préparation de cette thèse.

Je remercie également Mademoiselle Louis et tous ceux qui ont participé à la réalisation matérielle de cette thèse.

INFORMATION ET STATISTIQUE

GENERALISATION DE L'INEGALITE DE CRAMER - RAO

Ch I INTRODUCTION

§ 1 RAPPELS ET NOTATIONS

Dans ce paragraphe nous rappelons certaines propriétés classiques du calcul des probabilités [7] et nous définissons précisément les notations employées dans la suite de l'exposé.

1.1 RAPPELS.

Un espace de probabilité $(\mathcal{X}, \mathcal{A}, P)$ est constitué par la donnée d'un ensemble non vide $\mathcal{X}$, d'une tribu $\mathcal{A}$ de parties de $\mathcal{X}$ et d'une mesure de probabilité P sur $(\mathcal{X}, \mathcal{A})$. Soit ν une mesure signée sur $(\mathcal{X}, \mathcal{A})$ et λ une mesure positive sur le même espace mesurable, λ *domine* ν *en mesure ou* ν *est absolument continue par rapport à* λ, si :

$$N \in \mathcal{A} \text{ et } \lambda(N) = 0 \Rightarrow \nu(N) = 0$$

on note $\lambda >> \nu$.

Soient $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ une famille de probabilités sur $(\mathcal{X}, \mathcal{A})$ et $(\theta_1, \theta_2) \in \Theta^2$:

- si $P_{\theta_1} >> P_{\theta_2}$ et $P_{\theta_1} << P_{\theta_2}$ alors P_{θ_1} et P_{θ_2} sont *des probabilités équivalentes en mesure*, on note $P_{\theta_1} \equiv P_{\theta_2}$
- si $\forall (\theta_1, \theta_2) \in \Theta^2$ on a $P_{\theta_1} \equiv P_{\theta_2}$ alors $\mathcal{P}$ forme *une famille homogène de probabilités*.
- si $\forall \theta \in \Theta$ on a $P_\theta << \lambda$, on dit que $\mathcal{P}$ est *une famille dominée* par la mesure λ.

Soient ν une mesure signée, σ- finie et λ une mesure positive, σ - finie, sur $(\mathcal{X},\mathcal{A})$ avec $\nu \ll \lambda$, *le théorème de Radon-Nikodym* nous montre qu'il existe une application mesurable $f = \frac{d\nu}{d\lambda}$, de $(\mathcal{X},\mathcal{A})$ dans $(\mathbb{R},\mathcal{R})$, λ - presque surement (λ -p.s.) définie, telle que :

$$\forall A \in \mathcal{A} \qquad \nu(A) = \int_A f \, d\lambda$$

Si de plus ν est une mesure positive, σ- finie, alors

$$0 \leq f < +\infty \qquad \lambda \text{ - p.s.}$$

L'espace mesurable $(\mathbb{R},\mathcal{R})$ est formé par l'ensemble $\mathbb{R}$ des nombres réels et la tribu $\mathcal{R}$ engendrée par les boreliens de $\mathbb{R}$.

Etant donné un espace probabilisé $(\mathcal{X},\mathcal{A},P)$, une mesure λ positive et σ- finie sur $(\mathcal{X},\mathcal{A})$ telle que $\lambda \gg P$, si $f = \frac{dP}{d\lambda}$ et si g est une application mesurable de $(\mathcal{X},\mathcal{A})$ dans $(\mathbb{R},\mathcal{R})$ dire que

$g \in \mathcal{L}^p(\mathcal{X},\mathcal{A},P)$ signifie que les intégrales suivantes ont un sens

$$\int_{\mathcal{X}} |g|^p dP = \int_{\mathcal{X}} |g|^p f \, d\lambda$$

1.2 NOTATIONS.

Dans la suite de l'exposé, sauf indications contraires, nous employons les notations et les conventions suivantes.

L'espace mesuré $(\mathcal{X},\mathcal{A},\lambda)$ est complet pour la mesure positive et σ - finie λ. Soit $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}\}$ une famille de probabilités sur $(\mathcal{X},\mathcal{A})$, dominée par λ, notons $f(.,\theta) = f_\theta$ la dérivée de Radon-Nikodym de P_θ par rapport à λ

$$0 \leq f_\theta = \frac{dP_\theta}{d\lambda} < +\infty \qquad \lambda \text{ - p.s.}$$

Nous supposons $\phi \neq \Theta \subset \mathbb{R}$ et que le paramètre sépare les probabilités, c'est à dire que pour $(\theta_1, \theta_2) \in \Theta^2$ on a :

$$\theta_1 \neq \theta_2 \iff P_{\theta_1} \neq P_{\theta_2} \iff f_{\theta_1} \neq f_{\theta_2} \quad \lambda - \text{p.s.}$$

Quand il n'y a pas d'ambiguïté possible, nous posons pour $\theta_i \in \Theta$ et $i \in \mathbb{N}$ un indice quelconque :

- $P_i = P_{\theta_i}$, $f_i = f(.,\theta_i), \ldots$
- $\mathcal{X}_i = \mathcal{X}_{\theta_i} = \{x \in \mathcal{X} : f_i(x) > 0\}$

$\mathcal{X}_{12} = \mathcal{X}_1 . \mathcal{X}_2 = \{x \in \mathcal{X} : f_1(x) > 0 \text{ et } f_2(x) > 0\}$

Comme $(\mathcal{X}, \mathcal{A}, \lambda)$ est complet : $\mathcal{X}_i \in \mathcal{A}$ et $\mathcal{X}_{12} \in \mathcal{A}$.
Avec les notations ainsi définies, il vient :

1.2.1 Proposition.

a) $P_1 >> P_2 \iff \mathcal{X}_1 \supset \mathcal{X}_2 \quad \lambda - \text{p.s.}$

b) $P_1 \equiv P_2 \iff \mathcal{X}_1 = \mathcal{X}_2 \quad \lambda - \text{p.s.}$

c) La famille $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ est homogène si et seulement si il existe $\mathcal{X}^o \in \mathcal{A}$, indépendant de $\theta \in \Theta$, tel que $\theta \in \Theta \Rightarrow \mathcal{X}_\theta = \mathcal{X}^o \quad \lambda - \text{p.s.}$

Montrons rapidement 1.2.1. a)
Soit $\mathcal{X}_1 \supset \mathcal{X}_2$ λ- p.s. et $A \in \mathcal{A}$ tel que $P_1(A) = 0$. De proche en proche nous avons :

$$0 = P_1(A) = \int_A f_1 \, d\lambda = \int_{A.\mathcal{X}_1} f_1 \, d\lambda \Rightarrow \lambda(A.\mathcal{X}_1) = 0 \Rightarrow \lambda(A.\mathcal{X}_2) = 0$$

$$\Rightarrow P_2(A) = \int_A f_2 \, d\lambda = \int_{A.\mathcal{X}_2} f_2 \, d\lambda = 0 \text{ donc } P_1 >> P_2.$$

Inversement supposons $P_1 \gg P_2$:

$P_1(\mathcal{X}_2 - \mathcal{X}_1) = 0 \Rightarrow P_2(\mathcal{X}_2 - \mathcal{X}_1) = 0 \Rightarrow \lambda(\mathcal{X}_2 - \mathcal{X}_1) = 0$

$\Rightarrow \mathcal{X}_1 \supset \mathcal{X}_2 \quad \lambda$- p.s.

1.3. VARIABLE ALEATOIRE.

Soit X *une variable aléatoire* définie sur un espace probabilisé abstrait quelconque, à valeurs dans $(\mathcal{X}, \mathcal{A})$ et de loi P_θ. Une valeur particulière x de X est appelée *réalisation* de X. Si φ est une application mesurable et intégrable de $(\mathcal{X}, \mathcal{A}, P_\theta)$ dans $(\mathbb{R}, \mathcal{R})$ nous noterons selon les cas pour la variable aléatoire $Z = \varphi(X) = \varphi o X$

$$E_\theta(Z) = E_\theta(\varphi o X) = \int \varphi \, dP_\theta = \int_{\mathcal{X}} \varphi(x) \; f_\theta(x) \; d\lambda(x)$$

Soit T *une variable aléatoire* définie sur $(\mathcal{X}, \mathcal{A})$ à valeurs dans l'espace mesurable $(\mathcal{Y}, \mathcal{B})$. Par définition

$$\mathcal{A}_T = T^{-1}(\mathcal{B}) \subset \mathcal{A}$$

et suivant les conventions habituelles nous pouvons définir sur $(\mathcal{Y}, \mathcal{B})$ la mesure γ et la probabilité Q_θ (loi de T) de la manière suivante :

- $(\mathcal{X}, \mathcal{A}, \lambda) \xrightarrow{T} (\mathcal{Y}, \mathcal{B}, \gamma)$ avec
  $\forall B \in \mathcal{B} \quad \gamma(B) = \lambda(T^{-1}(B))$
- $(\mathcal{X}, \mathcal{A}, P_\theta) \xrightarrow{T} (\mathcal{Y}, \mathcal{B}, Q_\theta)$ avec $\forall B \in \mathcal{B} \quad Q_\theta(B) = P_\theta(T^{-1}(B))$

Comme $P_\theta \ll \lambda$ il est facile de voir que $Q_\theta \ll \gamma$ et l'on peut ainsi définir les dérivées de Radon - Nikodym

$$f_\theta = \frac{dP_\theta}{d\lambda} \quad \text{et} \quad g_\theta = \frac{dQ_\theta}{d\gamma}$$

Suivant la terminologie habituelle en statistique, nous dirons que f_θ et g_θ sont les densités des lois de X et de T.
Nous noterons $\mathcal{Q} = \{Q_\theta : \theta \in \Theta\}$.

Si ψ est une application mesurable et intégrable de $(\mathcal{Y}, \mathcal{B}, Q_\theta)$ dans $(\mathbb{R}, \mathcal{R})$ alors $\varphi = \psi \circ T$ est une application mesurable et intégrable de $(\mathcal{X}, \mathcal{A}, P_\theta)$ dans $(\mathbb{R}, \mathcal{R})$. Le théorème de transfert nous donne en particulier :

$$B \in \mathcal{B} \int_{T^{-1}(B)} \psi \circ T . f_\theta . d\lambda = \int_B \psi . g_\theta . d\lambda$$

Enfin pour $\theta_i \in \Theta$, $i = 1,2$, définissons

$$\mathcal{Y}_i = \{y \in \mathcal{Y} : g_i(y) > 0\} \qquad \mathcal{Y}_{12} = \mathcal{Y}_1 . \mathcal{Y}_2$$

$$\mathcal{X}_i = \{x \in \mathcal{X} : f_i(x) > 0\} \qquad \mathcal{X}_{12} = \mathcal{X}_1 . \mathcal{X}_2$$

il vient - $\mathcal{X}_i = T^{-1}(\mathcal{Y}_i) = \{x \in \mathcal{X} : g_i \circ T(x) > 0\}$ λ - p.s.

- $\mathcal{X}_{12} = T^{-1}(\mathcal{Y}_{12}) = \{x \in \mathcal{X} : g_1 \circ T(x) > 0$ et $g_2 \circ T(x) > 0\}$ λ - p.s.

- $Q_i(\mathcal{Y}_{12}) = P_i(\mathcal{X}_{12})$
- $\mathcal{P}$ homogène <=> $\mathcal{Q}$ homogène

1.3.1 Définition.

La variable aléatoire T est *un résumé exhaustif de* X *concernant le paramètre* si : $\forall \theta \in \Theta$ $f_\theta(x) = g_\theta \circ T(x).h(x)$ λ- p.s.
où g_θ est la densité de T et h ne dépend pas de θ

Pour la démonstration de la proposition 1.3.2 rappelons le lemme suivant.

Lemme. Une condition nécessaire et suffisante pour qu'une famille de probabilités $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ soit dominée par une mesure positive et σ- finie, est qu'il existe une partie finie ou dénombrable $\{P_{\theta_i}\}_{i \in I}$ de $\mathcal{P}$ telle que

$A \in \mathcal{A}$ et $P_{\theta_i}(A) = 0 \;\; \forall\, i \in I \Rightarrow P_\theta(A) = 0 \;\forall\, \theta \in \Theta$

La famille $\mathcal{P}$ que nous étudions, étant dominée par λ, nous pouvons appliquer le lemme. Supposons $I = \{1,\ldots,n\}$ ou $I = \{1,\ldots,n,\ldots\}$ et notons

$$\mathcal{X}_\theta = \{x : f_\theta(x) > 0\}$$
$$\mathcal{X}^1 = \mathcal{X}_1,\ \mathcal{X}^2 = \mathcal{X}_2 - \mathcal{X}_1, \ldots, \mathcal{X}^n = \mathcal{X}_n - \bigcup_{i=1}^{n-1} \mathcal{X}_i \qquad n \in I.$$

La famille des ensembles mesurables $\{\mathcal{X}^i\}_{i \in I}$ forme une partition finie ou dénombrable de $\mathcal{X}^o = \bigcup_{i \in I} \mathcal{X}_i = \bigcup_{i \in I} \mathcal{X}^i$.

Extrayons de cette famille une sous-famille d'indices $J \subset I$ telle que :

$$\lambda(\mathcal{X}^i) > 0 \qquad i \in J$$
$$\lambda(\mathcal{X}^i) = 0 \qquad i \in I-J$$

Il est clair que $\mathcal{X}^o = \bigcup_{i \in J} \mathcal{X}^i \quad \lambda$ -p.s.

et par une démonstration analogue à celle de 1.2.1, il vient

$$\theta \in \Theta \Rightarrow \mathcal{X}_\theta \subset \mathcal{X}^o \qquad \lambda \text{ - p.s.}$$

La famille $\{\mathcal{X}_\theta^j = \mathcal{X}^j . \mathcal{X}_\theta\}_{j \in J}$ forme λ- p.s. une partition mesurable finie ou dénombrable de $\mathcal{X}_\theta$.

1.3.2 **Proposition.** Une condition nécessaire et suffisante pour que T soit un résumé exhaustif de X concernant le paramètre $\theta \in \Theta$ est que :
$\forall\, (\theta_1, \theta_2) \in \Theta^2$

$$\frac{f_{\theta_1}}{g_{\theta_1} \circ T} = \frac{f_{\theta_2}}{g_{\theta_2} \circ T} \qquad \lambda\text{- p.s. sur } \mathcal{X}_{12} = \mathcal{X}_{\theta_1} . \mathcal{X}_{\theta_2}$$

Les rapports, ainsi définis, ont toujours un sens car $g_{\theta_1} oT$ et $g_{\theta_2} oT$ sont λ- p.s. strictement positives sur $\mathcal{X}_{12}$.

La condition nécessaire est évidente à partir de la définition. Pour montrer la condition suffisante nous allons considérer les deux cas suivants :

<u>1er cas</u>. La famille $\mathcal{P}$ est homogène, fixons-nous $\theta_o \in \Theta$. Nous savons que $\forall\ \theta \in \Theta \quad P_\theta \equiv P_{\theta_o}$ et $\mathcal{X}_\theta = \mathcal{X}_{\theta_o}$ $\quad \lambda$ - p.s. donc T est exhaustif par définition en prenant

$$h = 1_{\mathcal{X}_{\theta_o}} \frac{f_{\theta_o}}{g_{\theta_o} oT}$$

<u>2ème cas</u>. La famille $\mathcal{P}$ n'est pas homogène, mais $\forall\ \theta \in \Theta$ $\mathcal{X}_\theta \subset \mathcal{X}^o = \bigcup_{j \in J} \mathcal{X}^j$ $\quad \lambda$ - p.s. Il est facile de vérifier que T est exhaustif en prenant

$$h = \sum_{j \in J} 1_{\mathcal{X}^j} \frac{f_{\theta_o}}{g_{\theta_j} oT}$$

1.4 ECHANTILLON, STATISTIQUE.

Notons $(\mathcal{X}^n, \overset{n}{\otimes}\mathcal{A}, P_\theta^n)$ le produit de n espaces probabilisés identiques à $(\mathcal{X}, \mathcal{A}, P_\theta)$, $(\mathcal{X}^n, \overset{n}{\otimes}\mathcal{A}, \lambda)$ le produit de n espaces mesurés identiques à $(\mathcal{X}, \mathcal{A}, \lambda)$ avec la convention classique qui consiste à noter λ la mesure sur $(\mathcal{X}, \mathcal{A})$ et par la même lettre la mesure sur $(\mathcal{X}^n, \overset{n}{\otimes}\mathcal{A})$ construite à partir de λ.

La variable aléatoire $(X_1, X_2, \ldots, X_n)$ définie sur un espace probabilisé abstrait quelconque à valeurs dans $(\mathcal{X}^n, \overset{n}{\otimes}\mathcal{A})$ forme *un n - échantillon de* X si $X_1, X_2, \ldots, X_n$ sont des variables aléatoires indépendantes, à valeurs dans $(\mathcal{X}, \mathcal{A})$, de même loi que X.

Une valeur $x = (x_1, x_2, \ldots, x_n)$ de $(X_1, X_2, \ldots, X_n)$ est *une réalisation du n - échantillon*. La loi du n - échantillon est P^n_θ et sa densité par rapport à λ est :

$$\frac{dP^n_\theta}{d\lambda} = L(x,\theta) = L(x_1,x_2,\ldots,x_n;\theta) = \prod_{i=1}^{n} f_\theta(x_i) \qquad \theta \in \Theta$$

et $\quad 0 \leq L(x,\theta) < +\infty \qquad \lambda-$ p.s.

Si $x = (x_1,x_2,\ldots,x_n)$ est une réalisation du n - échantillon, $L(x,.)$ est une fonction de θ appelée *la vraisemblance* associée à cette réalisation.

Soit φ une application mesurable et intégrable de $(\mathcal{X}^n, \overset{n}{\otimes}\mathcal{A}, P^n_\theta)$ dans $(\mathbb{R}, \mathcal{R})$. Nous noterons selon les cas pour la variable aléatoire Z :

$$Z = \varphi(X_1,\ldots,X_n) = \varphi \circ (X_1,\ldots,X_n) \text{ et}$$

$$E_\theta(Z) = E_\theta(\varphi \circ (X_1,\ldots,X_n)) = \int_{\mathcal{X}^n} \varphi \, dP^n_\theta = \int_{\mathcal{X}^n} \varphi(x).L(x,\theta)\, d\lambda(x)$$

$$= \int_{\mathcal{X}^n} \varphi(x_1,\ldots,x_n)\, L(x_1,\ldots,x_n\,;\theta)\, d\lambda(x_1)\ldots d\lambda(x_n)$$

Soit S une variable aléatoire définie sur $(\mathcal{X}^n, \overset{n}{\otimes}\mathcal{A})$ à valeurs dans l'espace mesurable $(\mathcal{Y}, \mathcal{B})$. Nous appellerons *statistique* de $(X_1,\ldots,X_n)$ la variable aléatoire $S \circ (X_1,\ldots,X_n)$ définie sur le même espace probabilisé abstrait que le n - échantillon et à valeurs dans $(\mathcal{Y}, \mathcal{B})$. Par extension nous appellerons statistique la variable aléatoire S elle-même.

Il est facile de généraliser les notations, et les propositions du paragraphe 1.3 en remplaçant respectivement $(\mathcal{X}, \mathcal{A})$ par $(\mathcal{X}^n, \overset{n}{\otimes}\mathcal{A})$; X, P_θ, f_θ par $(X_1,\ldots,X_n)$, P^n_θ, $L(.,\theta)$; et T par S.

§ 2 ETUDES PRELIMINAIRES

Dans ce paragraphe nous rappelons et démontrons les inégalités employées par la suite. Ces inégalités sont toutes déduites de certaines propriétés de la fonction logarithme et leurs démonstrations au fur et à mesure des besoins alourdiraient inutilement le texte tout en masquant le lien qui les unit. Dans la dernière partie du paragraphe nous énonçons les hypothèses utilisées dans les chapitres ultérieurs.

2.1 **RAPPEL.** Pour tout $u \in]0,+\infty[$ on a

$$1 - \frac{1}{u} \leq \operatorname{Log} u \leq u-1$$

avec égalité, dans une des deux inégalités, si et seulement si $u = 1$.

Il est clair que $\operatorname{Log} u \leq u-1$ et que $\operatorname{Log} u = u-1 \iff u = 1$.
On en déduit la première relation par $-\operatorname{Log} u = \operatorname{Log} \frac{1}{u} \leq \frac{1}{u} - 1$.

Dans les propositions et démonstrations qui suivent, supposons que les intégrales qui interviennent, ont un sens. Cette hypothèse sera étudiée dans la remarque 2.6.

2.2 **Lemme.** Soient f_1 et f_2 des applications mesurables et intégrables définies sur $(\mathcal{X}, \mathcal{A}, \lambda)$, strictement positives λ- p.s. sur $A \in \mathcal{A}$. Si l'on note :

$$\mu_i(A) = \int_A f_i \, d\lambda \qquad i = 1,2$$

on a $$\mu_1(A) - \mu_2(A) \leq \int_A f_1 \operatorname{Log} \frac{f_1}{f_2} \, d\lambda \leq \mu_1(A) - \mu_2(A) + \int_A \frac{(f_1-f_2)^2}{f_2} \, d\lambda$$

avec égalité, dans une des deux inégalités, si et seulement si $f_1=f_2$ λ - p.s. sur A.

Si $\lambda(A) = 0$ la proposition est vraie, supposons $\lambda(A)>0$.
Les applications f_1 et f_2 étant strictement positives λ- p.s. sur A, la relation 2.1 avec $u = \frac{f_1}{f_2}$ devient :

$$1 - \frac{f_2}{f_1} \leq \text{Log}\,\frac{f_1}{f_2} \leq \frac{f_1}{f_2} - 1 \qquad \lambda\text{- p.s. sur } A$$

Multiplions par f_1 et intégrons sur A

$$\int_A (f_1-f_2)d\lambda \leq \int_A f_1\text{Log}\,\frac{f_1}{f_2}\,d\lambda \leq \int_A \frac{f_1}{f_2}\,(f_1-f_2)d\lambda$$

$$\mu_1(A) - \mu_2(A) \leq \int_A f_1\text{Log}\,\frac{f_1}{f_2}\,d\lambda \leq \mu_1(A) - \mu_2(A) + \int_A \frac{(f_1-f_2)^2}{f_2}\,d\lambda$$

De plus f_1 étant strictement positive λ- p.s. sur A, il y a égalité si et seulement si $f_1=f_2$ λ- p.s. sur A.

La démonstration montre de plus que si les applications f_1 et f_2 sont λ - intégrables sur A et si l'application $f_1\text{Log}\,\frac{f_1}{f_2}$ n'est pas λ- intégrable, alors $\int_A f_1\text{Log}\,\frac{f_1}{f_2}\,d\lambda = +\infty$.

Le lemme 2.2 est une généralisation du lemme de l'information :

2.3 Lemme de l'information. Si, avec les notations précédentes, f_1 et f_2 sont les densités des probabilités P_1 et P_2 par rapport à λ et si $P_1 \equiv P_2$ (ou simplement $P_1 << P_2$) on obtient l'inégalité de l'information en prenant

$$A = \mathcal{X}_1 = \{x \in \mathcal{X} : f_1(x)>0\}$$

cela donne $E_1\,\{\text{Log}\,f_1(X)\} \geq E_1\,\{\text{Log}\,f_2(X)\}$

ou $\displaystyle\int_{\mathcal{X}} \text{Log}\,\frac{f_1}{f_2}\,dP_1 = \int_{\mathcal{X}_1} f_1\text{Log}\,\frac{f_1}{f_2}\,d\lambda \geq 0$

avec égalité si et seulement si $P_1 = P_2$ (c'est à dire $f_1=f_2$ λ- p.s.)

Nous savons (1.2.1) que $P_1 \ll P_2 \iff \mathcal{X}_1 \subset \mathcal{X}_2$ λ -p.s.

le lemme 2.2 $\Rightarrow \int_{\mathcal{X}_1} f_1 \mathrm{Log}\frac{f_1}{f_2}\,d\lambda \geq 1 - P_2(\mathcal{X}_1) \geq 0$

Si $f_1 = f_2$ λ- p.s. il est clair que $\int_{\mathcal{X}_1} f_1 \mathrm{Log}\frac{f_1}{f_2}\,d\lambda = 0$ inversement

$$\int_{\mathcal{X}_1} f_1 \mathrm{Log}\frac{f_1}{f_2}\,d\lambda = 0 \Rightarrow P_2(\mathcal{X}_1) = 1 \Rightarrow \int_{\mathcal{X}_1} f_1 \mathrm{Log}\frac{f_1}{f_2}\,d\lambda = P_1(\mathcal{X}_1) - P_2(\mathcal{X}_1) = 1 - 1 = 0$$

le lemme 2.2 $\Rightarrow$ $f_1 = f_2$ λ- p.s.

Avec les notations de 2.2 on a symétriquement

$$\mu_2(A) - \mu_1(A) \leq \int_A f_2 \mathrm{Log}\frac{f_2}{f_1}\,d\lambda \leq \mu_2(A) - \mu_1(A) + \int_A \frac{(f_1-f_2)^2}{f_1}\,d\lambda$$

en sommant cette inégalité et celle du lemme 2.2 on obtient

$$0 \leq \int_A (f_1-f_2)\,\mathrm{Log}\frac{f_1}{f_2}\,d\lambda \leq \int_A (f_1-f_2)^2 \left(\frac{1}{f_1} + \frac{1}{f_2}\right) d\lambda$$

Appelons $\inf f_i$ et $\sup f_i$ les applications mesurables définies par

$\inf f_i(x) = \inf (f_1(x), f_2(x))$

$\sup f_i(x) = \sup (f_1(x), f_2(x))$ on a alors

$$0 \leq \int_A (f_1-f_2)\,\mathrm{Log}\frac{f_1}{f_2}\,d\lambda \leq 2\int_A \frac{(f_1-f_2)^2}{\inf f_i}\,d\lambda$$

La proposition suivante va nous donner un encadrement plus fin du terme central.

2.4 Proposition. Avec les notations et les hypothèses du lemme 2.2, on a :

2.4.1

2.4.2

$$0 \leq \int_A \frac{(f_1-f_2)^2}{\sup f_i}\,d\lambda \leq \left\{ \begin{array}{l} 2\int_A f_1 \mathrm{Log}\frac{f_1}{f_2}\,d\lambda + 2\mu_2(A) - 2\mu_1(A) \\ \int_A (f_1-f_2)\,\mathrm{Log}\frac{f_1}{f_2}\,d\lambda \end{array} \right\} \leq \int_A \frac{(f_1-f_2)^2}{\inf f_i}\,d\lambda$$

avec égalité dans une quelconque des inégalités, si et seulement si $f_1=f_2$ λ - p.s. sur A. En particulier si f_1 et f_2 sont les densités des probabilités P_1 et P_2 par rapport à λ et si $P_1 \equiv P_2$ alors en prenant $A = \mathcal{X}^0 = \{x : f_1(x)>0\} = \{x : f_2(x)>0\}$ λ - p.s., on a :

2.4.3

2.4.4

$$0 \leq \int_{\mathcal{X}^0} \frac{(f_1-f_2)^2}{\sup f_i}\, d\lambda \leq \left\{ \begin{array}{l} 2\int_{\mathcal{X}^0} f_1 \mathrm{Log}\, \frac{f_1}{f_2}\, d\lambda \\ \\ \int_{\mathcal{X}^0} (f_1-f_2)\, \mathrm{Log}\, \frac{f_1}{f_2}\, d\lambda \end{array} \right\} \leq \int_{\mathcal{X}^0} \frac{(f_1-f_2)^2}{\inf f_i}\, d\lambda$$

avec égalité, dans une quelconque des inégalités, si et seulement si $f_1=f_2$ λ - p.s. c'est à dire $P_1 = P_2$.

Toutes les fonctions à intégrer sont λ - p.s. finies sur A car par hypothèse f_1 et f_2 sont λ- p.s. strictement positives sur A. Cette hypothèse est essentielle pour la démonstration . Si $\lambda(A) = 0$ les relations sont vraies, supposons $\lambda(A) > 0$.

Il est clair que 2.4.1 $\Rightarrow$ 2.4.2 , mais nous allons donner la démonstration directe de 2.4.2 qui est très simple. Avec $u = \frac{f_1}{f_2}$ l'inégalité 2.1 devient

$$\frac{f_1-f_2}{f_1} \leq \mathrm{Log}\, \frac{f_1}{f_2} \leq \frac{f_1-f_2}{f_2} \qquad \lambda \text{ - p.s. sur A.}$$

En multipliant par (f_1-f_2) les trois membres de la relation, en examinant les deux cas possibles, on voit que :

2.4.5 $$0 \leq \frac{(f_1-f_2)^2}{\sup f_i} \leq (f_1-f_2)\, \mathrm{Log}\, \frac{f_1}{f_2} \leq \frac{(f_1-f_2)^2}{\inf f_i} \qquad \lambda \text{ - p.s. sur A}$$

L'intégration donne la relation 2.4.2. Il est facile de montrer que $f_1=f_2$ λ - p.s. sur A, entraîne toutes les égalités dans 2.4.2 et 2.4.5 et qu'une égalité quelconque entraîne $f_1=f_2$ λ - p.s. sur A.

Pour la démonstration de 2.4.1 nous avons besoin des résultats suivants :

1) $h(u) = 2u \operatorname{Log} u - u^2 + 1 \qquad u \in]0,+\infty[$

$h(u) > 0 \iff u < 1$

$h(u) = 0 \iff u = 1$

$h(u) < 0 \iff u > 1$

en effet $h'(u) = 2(1+\operatorname{Log} u - u) < 0$ pour $u \neq 1$ par (2.2), $h'(1) = h(1) = 0$, cela nous donne le résultat désiré.

2) $g(u) = 2 \operatorname{Log} u + (1-u)(3-u)$

$g(u) < 0 \iff u < 1$

$g(u) = 0 \iff u = 1$

$g(u) > 0 \iff u > 1$

en effet $g'(u) = 2\left(\frac{1}{\sqrt{u}} - \sqrt{u}\right)^2 > 0$ pour $u \neq 1$,

$g'(1) = g(1) = 0$ et nous avons le résultat cherché.

Montrons que sur A nous avons

2.4.6 $$0 \leq \frac{(f_1 - f_2)^2}{\sup f_i} \leq 2 f_1 \operatorname{Log} \frac{f_1}{f_2} + 2 f_2 - 2 f_1 \leq \frac{(f_1-f_2)^2}{\inf f_i} \qquad \lambda - \text{p.s.}$$

avec une des trois inégalités en égalité si et seulement si $f_1 = f_2$ λ - p.s. sur A.

<u>1er cas.</u> $f_1 > f_2 > 0 \qquad u = \frac{f_1}{f_2}$ et $v = \frac{f_2}{f_1}$.

L'inégalité de droite (en divisant par $f_2 > 0$) devient

$2 u \operatorname{Log} u + 2 - 2u < (u-1)^2$ ou

$h(u) = 2 u \operatorname{Log} u - u^2 + 1 < 0$ ce qui est vrai car $u = \frac{f_1}{f_2} > 1$.

L'inégalité de gauche (en divisant par $f_1 > 0$) devient

$$(1-v)^2 < -2 \operatorname{Log} v + 2v - 2 \quad \text{ou}$$

$$g(v) = 2 \operatorname{Log} v + (1-v)(3-v) < 0 \quad \text{ce qui est vrai car } v = \frac{f_2}{f_1} < 1.$$

<u>2ème cas.</u> $f_2 > f_1 > 0$ d'une manière identique.

L'inégalité de droite est $-2 \operatorname{Log} v + 2v - 2 < (1-v)^2$ ou

$$g(v) > 0 \text{ ce qui est vrai car } v = \frac{f_2}{f_1} > 1.$$

L'inégalité de gauche est $(u-1)^2 < 2u \operatorname{Log} u - 2(u-1)$ ou

$$h(u) > 0 \text{ ce qui est vrai car } u = \frac{f_1}{f_2} < 1.$$

<u>3ème cas.</u> $f_1 = f_2 > 0$ les trois membres sont égaux à zéro.
Inversement si une des inégalités est une égalité, par des raisonnements analogues à ceux du 1er cas et du 2ème cas on montre que $f_1 = f_2$ λ - p.s. sur A.

Comme par hypothèse f_1 et f_2 sont λ - p.s. strictement positives sur A, la relation 2.4.6 est vraie λ - p.s. et par intégration sur A nous trouvons les inégalités de 2.4.1 désirées.

Les inégalités de 2.4.3 et 2.4.4 sont alors évidentes ; de plus remarquons :

<u>2.4.7</u> $$0 \leq \frac{(f_1-f_2)^2}{\sup f_i} \leq \frac{(f_1-f_2)^2}{f_i} \leq \frac{(f_1-f_2)^2}{\inf f_i} \qquad \lambda \text{ - p.s. sur A} \qquad i = 1,2$$

<u>2.4.8</u> $$0 \leq \int_A \frac{(f_1-f_2)^2}{\sup f_i}\, d\lambda \leq \int_A \frac{(f_1-f_2)^2}{f_i}\, d\lambda \leq \int_A \frac{(f_1-f_2)^2}{\inf f_i}\, d\lambda \qquad i = 1,2$$

avec une quelconque des inégalités en égalité si et seulement si $f_1 = f_2$ λ - p.s. sur A.

<u>2.5</u> <u>Proposition.</u> Avec les notations et les hypothèses du lemme 2.2 et de plus si : $\lambda(A) > 0$

: $f_1 \neq f_2$ λ - p.s. sur A

: $\exists\ \varepsilon \in\ ]0,1[$ tel que $1-\varepsilon \leq \frac{f_i}{f_j} \leq 1+\varepsilon \qquad i \neq j, \quad i = 1,2,\ j = 1,2$

λ - p.s. sur A

$$\text{alors } 1-\varepsilon \leq \frac{\int_A \frac{(f_1-f_2)^2}{\sup f_i}\, d\lambda}{\int_A \frac{(f_1-f_2)^2}{\inf f_i}\, d\lambda} < 1 < \frac{\int_A \frac{(f_1-f_2)^2}{\inf f_i}\, d\lambda}{\int_A \frac{(f_1-f_2)^2}{\sup f_i}\, d\lambda} \leq 1+\varepsilon$$

et cela entraîne que le rapport de deux quelconques des quantités suivantes, est compris entre $1-\varepsilon$ et $1+\varepsilon$

$$\alpha = \int_A \frac{(f_1-f_2)^2}{\inf f_i}\, d\lambda, \qquad \beta = \int_A \frac{(f_1-f_2)^2}{\sup f_i}\, d\lambda$$

$$a_i = \int_A \frac{(f_1-f_2)^2}{f_i}\, d\lambda \qquad i = 1,2$$

$$b_i = 2\int_A f_i \operatorname{Log} \frac{f_i}{f_j} + 2\ \mu_j(A) - 2\ \mu_i(A) \qquad i \neq j\ ;\ i = 1,2,\quad j = 1,2$$

$$c = \int_A (f_1-f_2) \operatorname{Log} \frac{f_1}{f_2}\, d\lambda$$

Les conditions $f_1 \neq f_2$ λ - p.s. sur A et $\lambda(A) > 0$ entraînent que toutes les quantités α, β, a_i, b_i, c sont strictement positives, 2.4.1, 2.4.2, 2.4.8. Les rapports ont donc toujours un sens. D'autre part λ - p.s. sur A :

$$\left.\begin{array}{l} 1-\varepsilon \leq \frac{f_1}{f_2} \leq 1+\varepsilon \Rightarrow 0 \leq \frac{|f_1-f_2|}{f_2} \leq \varepsilon \\ 1-\varepsilon \leq \frac{f_2}{f_1} \leq 1+\varepsilon \Rightarrow 0 \leq \frac{|f_1-f_2|}{f_1} \leq \varepsilon \end{array}\right\} \Rightarrow 0 \leq \frac{|f_1-f_2|}{\sup f_i} \leq \frac{|f_1-f_2|}{\inf f_i} \leq \varepsilon$$

(rappelons que f_1 et f_2 donc $\inf f_i$ et $\sup f_i$ sont λ - p.s. strictement positives sur A)

Nous avons $0 < \alpha-\beta = \int_A (f_1-f_2)^2 \left(\frac{1}{\inf f_i} - \frac{1}{\sup f_i} \right)$ soit

$$0 < \alpha-\beta = \int_A (f_1-f_2)^2 \frac{|f_1-f_2|}{f_1 f_2} \leq \varepsilon \beta \leq \varepsilon \alpha$$

et nous obtenons la relation désirée car

$$0 < \alpha-\beta \leq \varepsilon \beta \Rightarrow 1 < \frac{\alpha}{\beta} \leq 1+\varepsilon \quad \text{et}$$

$$0 < \alpha-\beta \leq \varepsilon \alpha \Rightarrow 1-\varepsilon \leq \frac{\beta}{\alpha} < 1$$

Comme le rapport de deux quelconques des sept quantités ci-dessus précisées est toujours compris entre $\frac{\beta}{\alpha}$ et $\frac{\alpha}{\beta}$ par 2.4.1, 2.4.2, 2.4.8, ce rapport est donc compris entre $1-\varepsilon$ et $1+\varepsilon$.

2.6 <u>Remarque</u>. Soient f_1, f_2 les densités de probabilité de P_1, P_2 par rapport à la mesure λ sur $(\mathcal{X}, \mathcal{A})$. Dans les hypothèses des propositions 2.2, 2.4, 2.5 nous avons supposé que les applications f_1 et f_2 sont λ - p.s. strictement positives sur A c'est à dire que :

$$A \subset \mathcal{X}_{12} = \mathcal{X}_1 . \mathcal{X}_2 \qquad \lambda \text{ - p.s. avec}$$
$$\mathcal{X}_i = \{x : f_i(x) > 0\} \qquad i = 1,2$$

Il est facile d'établir les deux résultats suivants par des démonstrations et sous des hypothèses identiques à celles du lemme 2.2.

2.6.1 Si $\frac{f_1}{f_2} \in \mathcal{L}^1(\mathcal{X}, \mathcal{A}, P_1)$ alors : $\text{Log} \frac{f_1}{f_2} \in \mathcal{L}^1(\mathcal{X}, \mathcal{A}, P_1)$, $P_1 << P_2$,

et les intégrales du lemme 2.2 ont un sens $\forall A \in \mathcal{A}$ et $A \subset \mathcal{X}_1 \subset \mathcal{X}_2$ λ - p.s.

Si de plus $\frac{f_2}{f_1} \in \mathcal{L}^1(\mathcal{X}, \mathcal{A}, P_2)$ alors :

Log $\frac{f_2}{f_1} \in \mathcal{L}^1(\mathcal{X}, \mathcal{A}, P_2)$, $P_2 << P_1$ (et donc $P_1 \equiv P_2$) et toutes les intégrales α, β, a_i, b_i, c de la proposition 2.5 ont un sens $\forall A \in \mathcal{A}$ et $A \subset \mathcal{X}_1 = \mathcal{X}_2$ λ - p.s.

Les hypothèses de 2.6.1 sont très restrictives. Elles nous amènent à supposer $P_1 \equiv P_2$. Quand $P_1 \not\equiv P_2$ on a le résultat suivant :

2.6.2 Pour que toutes les intégrales α, β, a_i, b_i, c de la proposition 2.5 aient un sens $\forall A \in \mathcal{A}$ et $A \subset \mathcal{X}_{12}$ λ - p.s, il faut et il suffit que

$$\frac{f_1}{f_2} \in \mathcal{L}^1(\mathcal{X}_{12}, \mathcal{A}.\mathcal{X}_{12}, P_1) \quad \text{et}$$

$$\frac{f_2}{f_1} \in \mathcal{L}^1(\mathcal{X}_{12}, \mathcal{A}.\mathcal{X}_{12}, P_2)$$

$(\mathcal{X}_{12}, \mathcal{A}.\mathcal{X}_{12}, P_i)$ est l'espace mesuré formé par l'ensemble $\mathcal{X}_{12}$, par la tribu $\mathcal{A}.\mathcal{X}_{12}$ trace de $\mathcal{A}$ sur $\mathcal{X}_{12}$, par la mesure restriction à $(\mathcal{X}_{12}, \mathcal{A}.\mathcal{X}_{12})$ de la probabilité P_i sur $(\mathcal{X}, \mathcal{A})$; par abus d'écriture nous notons par la même lettre P_i cette mesure.

2.7 Hypothèses.

Cette étude nous conduit, pour la suite de l'exposé, à définir les conditions ci-dessous :

Soit $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}\}$ une famille de probabilités sur $(\mathcal{X}, \mathcal{A})$, dominée par la mesure λ positive et σ - finie :

$$0 \leq f_\theta = \frac{dP_\theta}{d\lambda} < +\infty \qquad \lambda \text{ - p.s.}$$

Notons $\mathcal{X}_\theta = \{x \in \mathcal{X} : f_\theta(x) > 0\}$ et

$$\mathcal{X}_{12} = \mathcal{X}_{\theta_1 \theta_2} = \mathcal{X}_{\theta_1} . \mathcal{X}_{\theta_2}$$

Définissons les hypothèses suivantes :

A.0 $\forall\ (\theta_1,\theta_2) \in \Theta^2 \qquad \text{Log}\ \dfrac{f_1}{f_2} \in \mathcal{L}^1(\mathcal{X}_{12}, \mathcal{A}.\mathcal{X}_{12}, P_1)$

A.1 $\forall\ (\theta_1,\theta_2) \in \Theta^2 \qquad \dfrac{f_1}{f_2} \in \mathcal{L}^1(\mathcal{X}_{12}, \mathcal{A}.\mathcal{X}_{12}, P_1)$

A.2 1) Θ est un ouvert de $\mathbb{R}$

2) $\forall\ \theta_1 \in \Theta, \quad \forall\ \varepsilon > 0, \quad \exists\ \nu(\theta_1,\varepsilon) > 0,$

tels que : $\qquad \theta_2 \in \Theta, \quad \theta_2 \neq \theta_1, \quad |\theta_1-\theta_2| < \nu \quad \Rightarrow$

$$\begin{cases} - P_1(\mathcal{X}_{12}) \geq 1-\varepsilon \\ - \dfrac{f_1}{f_2} \neq \text{Constante} & \lambda - \text{p.s. sur } \mathcal{X}_{12} \\ - 1-\varepsilon \leq \dfrac{f_1}{f_2} \leq 1+\varepsilon & \lambda - \text{p.s. sur } \mathcal{X}_{12} \end{cases}$$

Nous avons : A-1 => A-0. Ces deux hypothèses ont pour but de donner un sens aux diverses intégrales précédemment étudiées. Avec les notations de 2.5, $\forall\ A \in \mathcal{A}$ et $A \subset \mathcal{X}_{12}$ λ - p.s., il vient :

A-0 => existence des intégrales : β, b_1, b_2 et c,

A-1 => existence des intégrales : α, β, a_1, a_2, b_1, b_2 et c.

Pour $(\theta_1,\theta_2) \in \Theta^2$, si $\lambda(\mathcal{X}_{12}) > 0$, définissons

$$f_i^* = \frac{f_i}{P_i(\mathcal{X}_{12})} \qquad i = 1,2$$

On montre facilement que :

Si l'hypothèse A.0 est vérifiée, les intégrales suivantes ont un sens :

$$b_i^* = 2\int_{\mathcal{X}_{12}} f_i^* \,\text{Log}\, \frac{f_i^*}{f_j^*}\, d\lambda \qquad i \neq j, \quad i = 1,2, \quad j = 1,2$$

$$\overset{*}{c} = \int_{\mathcal{X}_{12}} (\overset{*}{f_1} - \overset{*}{f_2}) \quad \mathrm{Log} \frac{\overset{*}{f_1}}{\overset{*}{f_2}} \, d\lambda$$

et $$\overset{*}{\beta} = \int_{\mathcal{X}_{12}} \frac{(\overset{*}{f_1} - \overset{*}{f_2})^2}{\sup \overset{*}{f_i}} \, d\lambda$$

Si l'hypothèse A.1 est vérifiée, les intégrales suivantes ont un sens :

$$\overset{*}{\alpha} = \int_{\mathcal{X}_{12}} \frac{(\overset{*}{f_1} - \overset{*}{f_2})^2}{\inf \overset{*}{f_i}} \, d\lambda, \qquad \overset{*}{\beta}, \quad \overset{*}{c}$$

$$\overset{*}{a_i} = \int_{\mathcal{X}_{12}} \frac{(\overset{*}{f_1} - \overset{*}{f_2})^2}{\overset{*}{f_i}} \, d\lambda, \qquad \overset{*}{b_i} \qquad i = 1,2$$

Définissons l'hypothèse B

B

1) $\mathcal{P}$ vérifie l'hypothèse A-1
2) Θ est un ouvert de $\mathbb{R}$.
3) $\forall \, \theta_1 \in \Theta, \quad \forall \, \varepsilon > 0, \quad \exists \, \nu \, (\theta_1, \varepsilon) > 0$

tel que $\theta_2 \in \Theta, \quad \theta_2 \neq \theta_1, \quad |\theta_1 - \theta_2| < \nu \quad \Rightarrow$

- $P_1(\mathcal{X}_{12}) \geq 1-\varepsilon$
- les quantités $\overset{*}{a_1}, \overset{*}{a_2}, \overset{*}{b_1}, \overset{*}{b_2}, \overset{*}{c}$, sont strictement positives et le rapport de deux quelconques d'entre elles est compris entre $1-\varepsilon$ et $1+\varepsilon$.

Différentes hypothèses peuvent être posées sur les densités f_i de manière à obtenir la condition B, en particulier la proposition 2.5 appliquée aux $\overset{*}{f_i}$ avec $A = \mathcal{X}_{12}$, nous donne :

2.7.1 Proposition. Si $\mathcal{P}$ vérifie les hypothèses A-1 et A-2 alors $\mathcal{P}$ vérifie l'hypothèse B.

Remarquons : exiger a_i^*, b_i^*, c^* strictement positives, c'est exiger $\lambda(\mathcal{X}_{12}) > 0$ et $f_1^* \neq f_2^*$ λ - p.s. sur $\mathcal{X}_{12}$.

Comme: $f_1^* \neq f_2^*$ λ - p.s. sur $\mathcal{X}_{12}$ $\Leftrightarrow$ $\frac{f_1}{f_2} \neq$ constante λ - p.s. sur $\mathcal{X}_{12}$), la proposition est alors évidente.

Ch II GENERALISATION DE LA NOTION D'INFORMATION

Dans ce chapitre nous rappelons et généralisons certains résultats sur l'information telle qu'elle a été définie par Kullback [4],[5],[6]. La théorie traitée est une application à la statistique de la notion d'information introduite par Shannon et Wiener.

§ 3 DEFINITIONS

3.1 NOTATIONS.

Soient $(\mathcal{X},\mathcal{A},P_i)$, $i = 1,2$, deux espaces probabilisés ; λ une mesure positive, σ- finie, sur $(\mathcal{X},\mathcal{A})$ telle que $\lambda >> P_i$, $i = 1,2$;
f_1, f_2 les dérivées de Radon-Nikodym de P_1 et P_2 par rapport à λ :

$$0 \leq f_i = \frac{dP_i}{d\lambda} < +\infty \qquad \lambda - \text{p.s.}$$

Notons $\mathcal{X}_i = \{x \in \mathcal{X} : f_i(x) > 0\}$, $i = 1,2$,
et $\mathcal{X}_{12} = \mathcal{X}_1 . \mathcal{X}_2$. Soit X une variable aléatoire définie sur un espace probabilisé quelconque, à valeurs dans $(\mathcal{X},\mathcal{A})$.

3.2 NOTIONS D'INFORMATION.

Considérons les hypothèses suivantes $(i = 1,2)$:
H_i : la loi de probabilité de X est P_i.

Soit $A \in \mathcal{A}$ tel que $P_1(A).P_2(A) > 0$, notons $(i = 1,2)$:

$P(H_i)$: probabilité a priori de H_i
$P(H_i/A)$: probabilité a posteriori de H_i, c'est à dire la probabilité de H_i étant donné A
$P_i(A) = P(A/H_i)$: probabilité de A dans l'hypothèse H_i.

Par le *théorème de Bayes* nous obtenons :

$$P(H_i/A) = \frac{P(H_i)\ P_i(A)}{P(H_1)\ P_1(A) + P(H_2)\ P_2(A)} \qquad i = 1,2$$

ce qui nous donne :

$$\text{Log}\ \frac{P_1(A)}{P_2(A)} = \text{Log}\ \frac{P(H_1/A)}{P(H_2/A)} - \text{Log}\ \frac{P(H_1)}{P(H_2)}$$

Le terme Log $\frac{P(H_1/A)}{P(H_2/A)}$ traduit ce qui est en faveur de H_1 contre H_2 après la donnée de A.

Le terme Log $\frac{P(H_1)}{P(H_2)}$ traduit ce qui est en faveur de H_1 contre H_2 avant la donnée de A.

Le terme Log $\frac{P_1(A)}{P_2(A)}$ représente donc *l'information globale apportée par* $A \in \mathcal{A}$ *, en faveur de* H_1 *contre* H_2.

De la même manière, en écrivant la formule de Bayes avec les densités, c'est à dire en ne considérant que la réalisation x de X, on obtient :

$$\text{Log}\ \frac{f_1(x)}{f_2(x)} = \text{Log}\ \frac{P(H_1/x)}{P(H_2/x)} - \text{Log}\ \frac{P(H_1)}{P(H_2)} \qquad \lambda - \text{p.s.}$$

Le terme Log $\frac{f_1(x)}{f_2(x)}$ représente *l'information apportée par l'observation de la réalisation* x *de* X, *en faveur de* H_1 *contre* H_2. Ces notions intuitives donnent une première justification de la définition suivante (selon Kullback) :

3.2.0 *L'information moyenne apportée par* $A \in \mathcal{A}$ *en faveur de* H_1 *contre* H_2 est la quantité :

$$I(1,2\ ;\ A) = \frac{1}{P_1(A)} \int_A f_1 \text{Log}\ \frac{f_1}{f_2}\ d\lambda \qquad \text{si } P_1(A) > 0$$

$$I(1,2\ ;\ A) = 0 \qquad \text{si } P_1(A) = 0$$

avec la convention :

$$\int_A f_1 \mathrm{Log}\frac{f_1}{f_2}\, d\lambda = \int_A \mathrm{Log}\frac{f_1}{f_2}\, dP_1 = \int_{A.\mathcal{X}_1} f_1 \mathrm{Log}\frac{f_1}{f_2}\, d\lambda, \quad \mathcal{X}_1 = \{x : f_1(x) > 0\}$$

Quand la condition $P_1 << P_2$, posée par Kullback, est vérifiée, l'intégrale ainsi définie a un sens $\forall A \in \mathcal{A}$ (en supposant : $\mathrm{Log}\frac{f_1}{f_2} \in \mathcal{L}^1(\mathcal{X}, \mathcal{A}, P_1)$) mais dans le cas général où P_1 n'est pas dominée par P_2 une étude plus précise s'impose.

Considérons la partition mesurable suivante de $A \in \mathcal{A}$:

$$A = (A.\mathcal{X}_{12}) \cup (A.(\mathcal{X}_1 - \mathcal{X}_{12})) \cup (A - \mathcal{X}_1) = A_{12} \cup A'_{12} \cup A''_{12}$$

3.2.1 $A_{12} = A.\mathcal{X}_{12} = \{x \in A : f_1(x) > 0 \quad \text{et} \quad f_2(x) > 0\}$

A partir du lemme (2.2) on voit que :

$$I(1,2\ ;\ A_{12}) = 0 \qquad \text{si } P_1(A_{12}) = 0$$

$$I(1,2\ ;\ A_{12}) = \frac{1}{P_1(A_{12})}\int_{A_{12}} f_1 \mathrm{Log}\frac{f_1}{f_2}\, d\lambda \geq \frac{P_1(A_{12}) - P_2(A_{12})}{P_1(A_{12})} > -\infty$$

$$\text{si } P_1(A_{12}) > 0$$

Notons : $P_1(A_{12}) = 0 \iff \lambda(A_{12}) = 0 \iff P_2(A_{12}) = 0$ et

$$P_1(A_{12}) > 0 \iff \lambda(A_{12}) > 0 \iff P_2(A_{12}) > 0$$

La signification de ce cas est claire, A_{12} est un ensemble de probabilité strictement positive dans les hypothèses H_1 et H_2, la donnée de A_{12} ne permet pas de rejeter une des hypothèses mais $I(1,2\ ;\ A_{12}) \in]-\infty,+\infty]$ est un coefficient qui "exprime" l'information apportée en faveur de H_1 contre H_2 par A_{12}. C'est ce terme qui sera étudié ultérieurement.

3.2.2 $A'_{12} = A.(\mathcal{X}_1 - \mathcal{X}_{12}) = \{x \in A : f_1(x) > 0 \quad \text{et} \quad f_2(x) = 0\}$

Nous avons $P_2(A'_{12}) = 0$

$$I(1,2\ ;\ A'_{12}) = 0 \qquad\qquad \text{si } P_1(A'_{12}) = 0$$

$$I(1,2\ ;\ A'_{12}) = \frac{1}{P_1(A'_{12})} \int_{A'_{12}} f_1 \text{Log} \frac{f_1}{0}\, d\lambda = +\infty \qquad \text{si } P_1(A'_{12}) > 0$$

A'_{12} est un ensemble de probabilité nulle dans l'hypothèse H_2, il est de probabilité strictement positive dans l'hypothèse H_1, la donnée de A'_{12} entraîne le rejet de l'hypothèse H_2 et l'information apportée par A'_{12} en faveur de H_1 contre H_2 est totale, infinie.

3.2.3 $A''_{12} = A - \mathcal{X}_1 = \{x \in A : f_1(x) = 0\}$

Nous avons $P_1(A''_{12}) = 0$ et $I(1,2\ ;\ A''_{12}) = 0$

On ne sait rien sur $P_2(A''_{12})$.

A''_{12} est un ensemble de probabilité nulle dans l'hypothèse H_1 et l'information en faveur de H_1 contre H_2 apportée par la donnée de A''_{12} est nulle, quelle que soit l'hypothèse H_2.

3.2.4 Proposition. Avec les notations ainsi définies, il vient :

$$P_1(A)\ I(1,2\ ;\ A) = P_1(A_{12})\ I(1,2\ ;\ A_{12}) + P_1(A'_{12})\ I(1,2\ ;\ A'_{12})$$

3.2.5 Remarque. Dans la pratique on ne connait pas $A \in \mathcal{A}$ mais la réalisation x de la variable aléatoire X et $x \in A$; il faut donc distinguer les trois cas possibles $(P_1(A_{12}).P_1(A'_{12}) > 0)$

$x \in A_{12}$: On ne peut rejeter ni H_1 ni H_2, mais l'étude de $I(1,2\ ;\ A_{12})$ donne des indications sur la manière de choisir entre H_1 et H_2.

$x \in A'_{12}$: On rejette H_2.

$x \in A''_{12}$: On rejette H_1.

Cette étude préliminaire nous conduit à poser les définitions du paragraphe suivant.

3.3 INFORMATION APPORTEE PAR $A \in \mathcal{A}$.

3.3.1 **Définition** : On appelle *partie finie de l'information apportée par* $A \in \mathcal{A}$ *en faveur de* H_1 *contre* H_2, la quantité finie, si elle existe, définie par :

$$\psi(1,2\ ;\ A) = \frac{1}{P_1(A_{12})} \int_{A_{12}} f_1 \mathrm{Log}\, \frac{f_1}{f_2}\, d\lambda \qquad \text{si } P_1(A_{12}) > 0$$

$$\psi(1,2\ ;\ A) = 0 \qquad \text{si } P_1(A_{12}) = 0$$

avec $A_{12} = A.\mathcal{X}_{12} = \{x \in A : f_1(x) > 0 \text{ et } f_2(x) > 0\}$

L'information moyenne apportée par A *en faveur de* H_1 *contre* H_2 est :

$$I(1,2\ ;\ A) = \psi(1,2\ ;\ A) \qquad \text{si } P_1(A) = P_1(A_{12})$$

$$I(1,2\ ;\ A) = +\infty \qquad \text{si } P_1(A) > P_1(A_{12})$$

Cette définition est conforme à la définition 3.2.0 par suite de 3.2.4.

L'hypothèse A-0 est une condition nécessaire et suffisante pour que l'intégrale définissant $\psi(1,2\ ;\ A)$ ait un sens $\forall A \in \mathcal{A}$, c'est à dire qu'elle soit finie $\forall A \in \mathcal{A}$. Le lemme 2.2 montre que si $\psi(1,2\ ;\ A)$ n'est pas finie alors $\psi(1,2\ ;\ A) = +\infty$ (un exemple de ce type est donné dans [4]), cela n'est pas génant mais dans ce cas toutes les propriétés que nous montrons dans ce chapitre, sont triviales ; d'autre part il est nécessaire dans les chapitres ultérieurs que $\psi(1,2\ ;\ A)$ soit finie. C'est pourquoi nous limitons la définition 3.3.1 au cas où $\psi(1,2\ ;\ A)$ est effectivement finie.

3.3.2 **Définition.** Soit $A \in \mathcal{A}$, $A \subset \mathcal{X}_{12}$ λ - p.s. et $\lambda(A) > 0$. La partie finie de l'information apportée par A en faveur de H_1 contre H_2 est la somme de deux termes

$$\psi(1,2\ ;\ A) = \mathrm{Log}\,\frac{P_1(A)}{P_2(A)} + \int_A f_1^* \,\mathrm{Log}\,\frac{f_1^*}{f_2^*}\, d\lambda \quad \text{avec } f_i^* = \frac{f_i}{P_i(A)} \qquad i = 1,2$$

Le terme $\mathrm{Log}\,\frac{P_1(A)}{P_2(A)}$ représente *l'information globale apportée par* A *en faveur de* H_1 *contre* H_2.

Le terme $\psi^*(1,2\ ;\ A) = \int_A f_1^* \,\mathrm{Log}\,\frac{f_1^*}{f_2^*}\, d\lambda$ représente *l'information fine apportée par* A *en faveur de* H_1 *contre* H_2.

Comme $A \subset \mathcal{X}_{12}$ λ - p.s. $\Rightarrow$ $A = A_{12}$ λ - p.s. et $\lambda(A) > 0 \Rightarrow P_1(A) > 0$ et $P_2(A) > 0$ donc

$$\psi(1,2\ ;\ A) = \frac{1}{P_1(A)} \int_A f_1 \mathrm{Log}\,\frac{f_1}{f_2}\, d\lambda = \int_A f_1^* \,\mathrm{Log}\,\frac{f_1^*\, P_1(A)}{f_2^*\, P_2(A)}\, d\lambda$$

$$= \int_A f_1^* \,\mathrm{Log}\,\frac{P_1(A)}{P_2(A)}\, d\lambda + \int_A f_1^* \,\mathrm{Log}\,\frac{f_1^*}{f_2^*}\, d\lambda$$

ce qui nous donne la relation désirée.

Le terme $\mathrm{Log}\,\frac{P_1(A)}{P_2(A)}$, (3.2), représente l'information apportée par le rapport des masses globales de probabilité, portées par A, dans les hypothèses H_1 et H_2.

Le terme $\psi^*(1,2\ ;\ A)$ représente l'information apportée par la manière différente dont ces masses de probabilités sont distribuées sur A, dans les hypothèses H_1 et H_2.

3.3.3 <u>Proposition.</u> Si $\psi(1,2\ ;\ A)$ est définie $\forall\ A \in \mathcal{A}$, une condition nécessaire et suffisante pour que $P_1 << P_2$ est que $I(1,2\ ;\ A) = \psi(1,2\ ;\ A)$ $\forall\ A \in \mathcal{A}$, ou seulement que $I(1,2\ ;\mathcal{X}) = \psi(1,2\ ;\ \mathcal{X})$

En effet $\mathcal{X} \in \mathcal{A}$ et $I(1,2\ ;\mathcal{X}) = \psi(1,2\ ;\ \mathcal{X}) \iff P_1(\mathcal{X}_{12}) = P_1(\mathcal{X}) = 1$ $\iff \mathcal{X}_1 = \mathcal{X}_{12} \subset \mathcal{X}_2$ λ - p.s. $\iff P_1 << P_2$.

3.3.4 <u>Proposition.</u> Soit $\{A_n : A_n \in \mathcal{A},\ n \in I\}$ une famille finie ou dénombrable d'ensembles mesurables, disjoints deux à deux, alors :

$$P_1(\bigcup_{n\in I} A_n).I(1,2\ ;\ \bigcup_{n\in I} A_n) = \sum_{n\in I} P_1(A_n).\ I(1,2\ ;\ A_n)$$

$$P_1(\bigcup_{n\in I} A_n^{12}).\psi\ (1,2\ ;\ \bigcup_{n\in I} A_n) = \sum_{n\in I} P_1(A_n^{12}).\psi\ (1,2\ ;\ A_n)$$

avec $A_n^{12} = \mathcal{X}_{12}.A_n$

La démonstration de cette proposition est évidente à partir des définitions de 3.3.1.

3.4 <u>INFORMATION.</u>

Conformément aux définitions et notations précédentes, en prenant $A = \mathcal{X}$, on peut définir les diverses informations apportées par la donnée de l'espace $\mathcal{X}$ tout entier.

Notons $\mathcal{X}_i = \{x \in \mathcal{X} : f_i(x) > 0\}$ $\quad i = 1,2$ et

$\mathcal{X}_{12} = \mathcal{X}_1.\mathcal{X}_2$, $\quad P_1(\mathcal{X}_{12}) = P_{12}$, $\ P_2(\mathcal{X}_{12}) = P_{21}$

3.4.1 **Définition.** On appelle *Partie finie de l'information en faveur de* H_1 *contre* H_2, la quantité finie, si elle existe, définie par :

$$\psi(1,2) = \frac{1}{P_{12}} \int_{\mathcal{X}_{12}} f_1 \mathrm{Log} \frac{f_1}{f_2}\, d\lambda \qquad \text{si } P_{12} > 0$$

$$\psi(1,2) = 0 \qquad \text{si } P_{12} = 0$$

L'information en faveur de H_1 *contre* H_2 est :

$$I(1,2) = \psi(1,2) \qquad \text{si } P_{12} = 1 \text{ c'est à dire si } P_1 << P_2$$

$$I(1,2) = +\infty \qquad \text{sinon}$$

L'information globale en faveur de H_1 *contre* H_2 est :

$$\mathrm{Log} \frac{P_{12}}{P_{21}}$$

L'information fine en faveur de H_1 *contre* H_2 est :

$$\psi^*(1,2) = \int_{\mathcal{X}_{12}} f_1^* \mathrm{Log} \frac{f_1^*}{f_2^*}\, d\lambda \qquad \text{avec } f_i^* = \frac{f_i}{P_i(\mathcal{X}_{12})} \qquad i = 1,2$$

et nous avons :

$$\psi(1,2) = \mathrm{Log} \frac{P_{12}}{P_{21}} + \psi^*(1,2)$$

3.4.2 **Rappel (2.7).** Le couple (P_1,P_2) vérifie l'hypothèse A-0 si

$$\mathrm{Log} \frac{f_1}{f_2} \in \mathcal{L}^1(\mathcal{X}_{12}, \mathcal{A}\cdot\mathcal{X}_{12}, P_i) \qquad i = 1,2$$

3.4.3 **Proposition.** Une condition nécessaire et suffisante pour que $\psi(1,2)$ existe est que $\mathrm{Log} \frac{f_1}{f_2} \in \mathcal{L}^1(\mathcal{X}_{12}, \mathcal{A}\cdot\mathcal{X}_{12}, P_1)$. ; de plus l'existence de $\psi(1,2)$ entraîne l'existence de $\psi^*(1,2)$ de $\psi(1,2 \,;\, A)$ et de $\psi^*(1,2 \,;\, A)$ $\forall\, A \in \mathcal{A}$.

Une condition nécessaire et suffisante pour que $\psi(1,2)$ et $\psi(2,1)$ existent, est que (P_1,P_2) vérifie l'hypothèse A-0.

La proposition 3.4.3 est évidente à partir de la définition 3.4.1 et de la proposition 3.3.3, il en est de même de la remarque suivante.

3.4.4 Remarque.

- $P_1 \equiv P_2 \iff \mathcal{X}_1 = \mathcal{X}_2 \quad \lambda\text{-p.s.} \iff P_{12} = P_{21} = 1$

$$\iff \psi(1,2) = I(1,2) = \psi^*(1,2)$$

de plus $\psi(1,2) = \int_{\mathcal{X}} f_1 \mathrm{Log}\, \frac{f_1}{f_2}\, d\lambda$

- $P_1 \underset{st.}{<<} P_2 \iff \mathcal{X}_1 \underset{st.}{\subset} \mathcal{X}_2 \quad \lambda\text{-p.s.} \iff P_{12} = 1$ et $P_{21} < 1$

$$\iff \psi(1,2) = I(1,2) \neq \psi^*(1,2)$$

de plus $\psi(1,2) = \int_{\mathcal{X}} f_1 \mathrm{Log}\, \frac{f_1}{f_2}\, d\lambda = \psi^*(1,2) - \mathrm{Log}\, P_{21}$

- si P_1 n'est pas dominée par P_2 alors $I(1,2) = +\infty$ et les trois autres quantités sont en général distinctes.

Kullback suppose que $P_1 << P_2$ et il prend la valeur commune de ψ et de I comme définition de l'information en faveur de H_1 contre H_2. Il n'introduit rien de comparable à ψ^*. Notre généralisation consiste à définir ψ, ψ^* et I même quand $P_1 << P_2$ n'est pas vérifiée et à montrer que les propriétés trouvées par Kullback se conservent dans la généralisation pour l'une ou plusieurs des informations I, ψ, ψ^*.

3.5 DIVERGENCE.

Supposons que (P_1,P_2) vérifie l'hypothèse A-0

3.5.1 <u>Définition</u>. On appelle *divergence* entre H_1 et H_2 :

$$J(1,2) = I(1,2) + I(2,1)$$

On appelle *partie finie de la divergence* entre H_1 et H_2 :

$$\varphi(1,2) = \psi(1,2) + \psi(2,1)$$

Notre définition est cohérente car il vient immédiatement

3.5.2 <u>Proposition.</u>

$$J(1,2) = \varphi(1,2) \qquad \text{si } P_1 \equiv P_2$$

$$J(1,2) = +\infty \qquad \text{sinon}$$

De plus nous avons la symétrie : $J(1,2) = J(2,1)$ et $\varphi(1,2) = \varphi(2,1)$

3.5.3 <u>Proposition</u>. Avec les notations du paragraphe 3.4, nous avons :

$$\varphi(1,2) = \psi^*(1,2) + \psi^*(2,1) \text{ c'est à dire}$$

$$\varphi(1,2) = \int_{\mathcal{X}_{12}} (f_1^* - f_2^*) \operatorname{Log} \frac{f_1}{f_2}\, d\lambda = \int_{\mathcal{X}_{12}} (f_1^* - f_2^*) \operatorname{Log} \frac{f_1^*}{f_2^*}\, d\lambda$$

$$f_i^* = \frac{f_i}{P_i(\mathcal{X}_{12})} \qquad i = 1,2.$$

En effet
$$\varphi(1,2) = \psi(1,2) + \psi(2,1)$$
$$= \psi^*(1,2) + \operatorname{Log}\frac{p_{12}}{p_{21}} + \psi^*(2,1) + \operatorname{Log}\frac{p_{21}}{p_{12}}$$
$$= \psi^*(1,2) + \psi^*(2,1)$$

3.6 <u>REMARQUE</u>.

Les différentes informations, ainsi définies, sont relatives aux densités f_1 et f_2 des lois de probabilité P_1 et P_2 de la variable aléatoire X, dans les hypothèses H_1 et H_2. Dans la suite si nous nous référons à d'autres variables aléatoires Y,T...

si les lois de ces variables aléatoires peuvent être choisies dans un ensemble indicé par $\theta \in \Theta$, le cas échéant, pour lever toutes ambiguïtés, nous noterons $\psi_X(\theta_1,\theta_2)$, $I_X(\theta_1,\theta_2)$, $\overset{*}{\psi}_X(\theta_1,\theta_2)$, $\psi_X(\theta_1,\theta_2;\, A)$ etc....
pour indiquer précisément qu'il s'agit des quantités ψ, I, $\overset{*}{\psi}$, $\psi(\ \ ,\ \ ;\, A)$, ... calculées à partir des densités f_{θ_1} et f_{θ_2} de la variable aléatoire X.

§ 4 PROPRIETES

Dans ce paragraphe et le suivant, nous démontrons les principales propriétés des différentes notions d'information précédemment introduites. Plus exactement nous généralisons les propriétés trouvées par Kullback en montrant que telle propriété correspond à telle notion d'information et que telle autre propriété correspond à une autre notion d'information.

Reprenons les notations de 3.1 et 3.4, supposons que les quantités employées existent (hypothèse A-0), et pour éviter un cas trivial supposons $P_i(\mathcal{X}_{12}) > 0, \quad i = 1,2.$

La remarque suivante va alors nous permettre de déduire directement les propriétés de $\psi^*(1,2)$ de celles de $I(1,2)$ telles qu'on les trouve dans [4] sous l'hypothèse $P_1 \equiv P_2$.

4.1 REMARQUE FONDAMENTALE.

Soient les espaces probabilisés $(\mathcal{X}, \mathcal{A}, P_i)$ $i = 1,2$, et la mesure λ positive et σ - finie sur $(\mathcal{X}, \mathcal{A})$, $\lambda >> P_i$, $i = 1,2$.

Nous avons déjà défini :

$$0 \leq f_i = \frac{dP_i}{d\lambda} < + \infty \qquad \lambda - \text{p.s.}$$

$$\mathcal{X}_i = \{x \in \mathcal{X} : f_i(x) > 0\} \qquad \text{et} \quad \mathcal{X}_{12} = \mathcal{X}_1 . \mathcal{X}_2$$

$$P_{12} = P_1(\mathcal{X}_{12}) \text{ et } P_{21} = P_2(\mathcal{X}_{12})$$

Définissons de plus les deux espaces probabilisés

$$(\mathcal{X}_{12}, \mathcal{A} . \mathcal{X}_{12}, P_i^*) \qquad i = 1,2$$

formés par l'ensemble $\mathcal{X}_{12}$, la tribu $\mathcal{A} . \mathcal{X}_{12}$ trace de $\mathcal{A}$ sur $\mathcal{X}_{12}$ et la probabilité conditionnelle de P_i par rapport à $\mathcal{X}_{12}$.

Nous avons $\forall A \in \mathcal{Q}$ $\quad P_i(A/\mathcal{X}_{12}) = \dfrac{P_i(A.\mathcal{X}_{12})}{P_i(\mathcal{X}_{12})}$ car $P_i(\mathcal{X}_{12}) > 0$

et l'on définit $\overset{*}{P}_i$ par

$$\forall A \in \mathcal{Q}.\mathcal{X}_{12} \qquad \overset{*}{P}_i(A) = P_i(A/\mathcal{X}_{12}) = \frac{P_i(A)}{P_i(\mathcal{X}_{12})}$$

Il est clair que la restriction $\overset{*}{\lambda}$ à $(\mathcal{X}_{12}, \mathcal{Q}.\mathcal{X}_{12})$ de λ est une mesure positive, σ - finie, telle que $\overset{*}{\lambda} >> \overset{*}{P}_i$, $\quad i = 1,2$.
On peut donc définir la densité $\overset{*}{f}_i$ de $\overset{*}{P}_i$ par :

$$0 \leq \overset{*}{f}_i = \frac{d\overset{*}{P}_i}{d\overset{*}{\lambda}} < +\infty$$

On montre facilement que $\overset{*}{f}_i = \dfrac{f_i}{P_i(\mathcal{X}_{12})}$ $\quad \lambda$ - p.s. sur $\mathcal{X}_{12}$. L'application mesurable $\overset{*}{f}_i$ est définie sur $(\mathcal{X}_{12}, \mathcal{Q}.\mathcal{X}_{12})$, mais nous utiliserons la même notation pour l'application mesurable définie sur $(\mathcal{X}, \mathcal{Q})$ par

$$\overset{*}{f}_i = \frac{f_i}{P_i(\mathcal{X}_{12})} \qquad \lambda \text{ - p.s. sur } \mathcal{X}.$$

4.1.1 Remarque. Nous avons $\overset{*}{P}_1 \equiv \overset{*}{P}_2$.
La quantité $\quad \overset{*}{\psi}(1,2) = \displaystyle\int_{\mathcal{X}_{12}} \overset{*}{f}_1 \operatorname{Log} \frac{\overset{*}{f}_1}{\overset{*}{f}_2} d\lambda$ est l'information I(1,2) en faveur de H_1 contre H_2 telle qu'elle est définie par 3.2.0 dans [4], en se plaçant dans les espaces $(\mathcal{X}_{12}, \mathcal{Q}.\mathcal{X}_{12}, \overset{*}{P}_i)$ $\quad i = 1,2$.

Toutes les propriétés de $\overset{*}{\psi}(1,2)$ (et par suite celles de $\psi(1,2)$ et I(1,2)) se démontrent comme dans le cas connu I(1,2) avec $P_1 \equiv P_2$.
Nous nous contenterons d'énoncer sans démonstration les propriétés dont nous aurons besoin ultérieurement, sauf quand nous possédons une démonstration originale ou quand nous énonçons une propriété que Kullback a laissée dans l'ombre.

4.2 PROPRIETES DE CONVEXITE.

4.2.1 Propriété. Avec les notations de 3.4, nous avons :

- $\overset{*}{\psi}(1,2) \geq 0$
- $\psi\,(1,2) \geq \text{Log}\,\dfrac{P_{12}}{P_{21}}$

avec égalité si et seulement si

$$\overset{*}{f_1} = \overset{*}{f_2} \qquad \lambda - \text{p.s. sur } \mathcal{X}_{12}.$$

Notons que : $\overset{*}{f_1} = \overset{*}{f_2}$ λ - p.s. sur $\mathcal{X}_{12}$ $\iff$

$$\frac{f_1}{f_2} = \frac{P_{12}}{P_{21}} = \text{constante} \qquad \lambda - \text{p.s. sur } \mathcal{X}_{12}$$

Le lemme 2.2 et la relation $\psi(1,2) = \overset{*}{\psi}(1,2) + \text{Log}\,\dfrac{P_{12}}{P_{21}}$ nous procurent les résultats désirés.

Le lemme 2.2 appliqué à ψ, et le rappel 2.1, nous donnent les propriétés suivantes.

4.2.2 Propriété. $\psi(1,2) \geq \text{Log}\,\dfrac{P_{12}}{P_{21}} \geq \dfrac{P_{12} - P_{21}}{P_{12}}$ avec en particulier

- $\psi(1,2) = \text{Log}\,\dfrac{P_{12}}{P_{21}} \geq \dfrac{P_{12} - P_{21}}{P_{12}} \iff \overset{*}{f_1} = \overset{*}{f_2} \qquad \lambda - \text{p.s. sur } \mathcal{X}_{12}$

  $\iff \overset{*}{\psi}(1,2) = 0$

- $\psi(1,2) \geq \text{Log}\,\dfrac{P_{12}}{P_{21}} = \dfrac{P_{12} - P_{21}}{P_{12}} \iff P_{12} = P_{21}$

  $\iff \psi(1,2) = \overset{*}{\psi}(1,2)$

- $\psi(1,2) = \text{Log}\,\dfrac{P_{12}}{P_{21}} = \dfrac{P_{12} - P_{21}}{P_{12}} \iff f_1 = f_2 \qquad \lambda - \text{p.s. sur } \mathcal{X}_{12}$

  $\iff \psi(1,2) = \overset{*}{\psi}(1,2) = \text{Log}\,\dfrac{P_{12}}{P_{21}} = 0$

De la même manière $\forall A \in \mathcal{A}$, $A \subset \mathcal{X}_{12}$ et $\lambda(A) > 0$ nous avons $P_1(A) > 0$, $P_2(A) > 0$ et le lemme (2.2) nous donne

$$\psi^*(1,2 ; A) = \int_A \frac{f_1}{P_1(A)} \operatorname{Log} \frac{f_1 \, P_2(A)}{P_1(A) \, f_2} \, d\lambda \geq 0$$

c'est à dire $\psi(1,2 ; A) = \displaystyle\int_A \frac{f_1}{P_1(A)} \operatorname{Log} \frac{f_1}{f_2} \, d\lambda \geq \operatorname{Log} \frac{P_1(A)}{P_2(A)}$

avec égalité si et seulement si $\dfrac{f_1}{f_2} = \dfrac{P_1(A)}{P_2(A)}$ λ- p.s. sur A.

De plus par 2.1 et 2.2 : $\psi(1,2 ; A) \geq \operatorname{Log} \dfrac{P_1(A)}{P_2(A)} \geq \dfrac{P_1(A) - P_2(A)}{P_1(A)}$

et une étude analogue à celle de 4.2.2 est possible. Nous déduisons de cela et de 3.3.4 la proposition suivante :

4.2.3 Proposition. Soit $\{A_i : A_i \in \mathcal{A}, A_i \subset \mathcal{X}_{12}$ λ-p.s., $\lambda(A_i) > 0$, $i \in I\}$ une famille finie ou dénombrable d'ensembles mesurables deux à deux disjoints, tels que $A = \bigcup_{i \in I} A_i$ nous avons :

$$P_1(A) \quad \psi(1,2 ; A) \geq \sum_{i \in I} P_1(A_i) \operatorname{Log} \frac{P_1(A_i)}{P_2(A_i)} \geq P_1(A) - P_2(A)$$

La première inégalité est une égalité si et seulement si

$$\frac{f_1}{f_2} = \frac{P_1(A_i)}{P_2(A_i)} \qquad \lambda \text{ - p.s. sur } A_i \qquad \forall i \in I$$

La deuxième inégalité est une égalité si et seulement si

$$P_1(A_i) = P_2(A_i) \qquad\qquad \forall i \in I$$

Les deux inégalités sont des égalités si et seulement si

$$f_1 = f_2 \qquad \lambda \text{ - p.s. sur A.}$$

D'autres propriétés analogues, concernant ψ^* et I, peuvent être énoncées.

4.3 PROPRIETE D'ADDITIVITE.

Soient X et Y deux variables aléatoires indépendantes, à valeurs respectivement dans $(\mathcal{X},\mathcal{A})$ et $(\mathcal{Y},\mathcal{B})$, dont les lois sont P_i et Q_i dans l'hypothèse H_i $(i = 1,2)$.

Supposons $P_i << \lambda$ et $Q_i << \gamma$, λ et γ mesures positives et σ - finies sur $(\mathcal{X},\mathcal{A})$ et $(\mathcal{Y},\mathcal{B})$. Nous avons donc :

$$f_i = \frac{dP_i}{d\lambda} \quad \lambda - \text{p.s.} \qquad \text{et } g_i = \frac{dQ_i}{d\gamma} \quad \gamma - \text{p.s.}$$

La variable aléatoire (X,Y) à valeurs dans $(\mathcal{X}\times\mathcal{Y},\mathcal{A}\otimes\mathcal{B})$ a pour loi R_i, dans l'hypothèse H_i, R_i est telle que :

- $\forall A \in \mathcal{A}$, $\forall B \in \mathcal{B}$: $\quad R_i(A\times B) = P_i(A).Q_i(B)$
- $R_i << \lambda.\gamma$ mesure positive et σ - finie sur $(\mathcal{X}\times\mathcal{Y},\mathcal{A}\otimes\mathcal{B})$
- $\dfrac{dR_i}{d(\lambda.\gamma)} = h_i \quad \lambda.\gamma - \text{p.s.}$ de plus

$$\forall (x,y) \in \mathcal{X}\times\mathcal{Y} : h_i(x,y) = f_i(x).g_i(y) \qquad \lambda.\gamma - \text{p.s.}$$

- $\{(x,y) \in \mathcal{X}\times\mathcal{Y} \; : h_i(x,y) > 0\} = \mathcal{X}_i\times\mathcal{Y}_i \qquad \lambda.\gamma - \text{p.s.}$

  et $R_i(\mathcal{X}_{12}\times\mathcal{Y}_{12}) = P_i(\mathcal{X}_{12}).Q_i(\mathcal{Y}_{12})$

On a donc : $\psi_{X,Y}(1,2) = \dfrac{1}{R_{12}}\displaystyle\int_{\mathcal{X}_{12}\times\mathcal{Y}_{12}} h_1 \text{Log}\,\frac{h_1}{h_2}\, d(\lambda.\gamma)$

$$\overset{*}{\psi}_{X,Y}(1,2) = \int_{\mathcal{X}_{12}\times\mathcal{Y}_{12}} \overset{*}{h}_1 \text{ Log } \frac{\overset{*}{h}_1}{\overset{*}{h}_2}\, d(\lambda.\gamma) \qquad \text{avec}$$

$$\overset{*}{h}_i = \frac{h_i}{R_i(\mathcal{X}_{12}\times\mathcal{Y}_{12})} = \overset{*}{f}_i.\overset{*}{g}_i \qquad \lambda.\gamma - \text{p.s.} \qquad i = 1,2$$

4.3.1 **Proposition.** Avec les notations ainsi définies, X et Y étant des variables aléatoires indépendantes :

$$\psi^*_{X,Y}(1,2) = \psi^*_X(1,2) + \psi^*_Y(1,2)$$

$$\psi_{X,Y}(1,2) = \psi_X(1,2) + \psi_Y(1,2)$$

$$\mathrm{Log}\,\frac{R_{12}}{R_{21}} = \mathrm{Log}\,\frac{P_{12}}{P_{21}} + \mathrm{Log}\,\frac{Q_{12}}{Q_{21}}$$

La démonstration est immédiate et se conduit comme dans le cas classique (4.1.1).

4.3.2 **Proposition.** Avec les notations de 1.4, soit $(X_1,X_2,\ldots,X_n)$ un n-échantillon de X :

- $\psi^*_n(1,2) = \psi^*_{X_1,\ldots,X_n}(1,2) = n \cdot \psi^*_X(1,2)$

- $\psi_n(1,2) = \psi_{X_1,\ldots,X_n}(1,2) = n \cdot \psi_X(1,2)$

Avec $\psi_n(1,2) = \dfrac{1}{P^n_{12}} \displaystyle\int_{\mathcal{X}^n_{12}} L(x,1)\ \mathrm{Log}\,\frac{L(x,1)}{L(x,2)}\ d\lambda(x)$

$$\psi^*_n(1,2) = \int_{\mathcal{X}^n_{12}} L^*(x,1)\ \mathrm{Log}\,\frac{L^*(x,1)}{L^*(x,2)}\ d\lambda(x)$$

$x = (x_1,\ldots,x_n)$, $L(x,i) = \prod_{k=1}^{n} f_i(x_k)$ et $L^*(x,i) = \dfrac{L(x,i)}{[P_i(\mathcal{X}_{12})]^n}$

car $\mathcal{X}^n_i = \{x \in \mathcal{X}^n : L(x,i) > 0\}$ $i = 1,2,$

la proposition se déduit immédiatement de 4.3.1.

4.4 INFORMATION ET EXHAUSTIVITE.

4.4.1 Proposition. Avec les notations de 1.3, soit T une application mesurable de $(\mathcal{X},\mathcal{A})$ dans $(\mathcal{Y},\mathcal{B})$, nous avons :

- $\overset{*}{\psi}_X(1,2) \geq \overset{*}{\psi}_T(1,2)$
- $\psi_X(1,2) \geq \psi_T(1,2)$
- L'information globale reste identique à elle-même :

$$\mathrm{Log}\,\frac{P_{12}}{P_{21}} = \mathrm{Log}\,\frac{Q_{12}}{Q_{21}}$$

Dans les deux premières relations on a égalité si et seulement si T est un résumé exhaustif de X.

La démonstration est identique au cas classique (4.1.1).
On montre, avec les notations de 1.3 que : $P_{12} = Q_{12}$, $P_{21} = Q_{21}$ et que

$$\overset{*}{\psi}(1,2) = \int_{\mathcal{X}_{12}} \overset{*}{f_1}\,\mathrm{Log}\,\frac{\overset{*}{g_1}\circ T}{\overset{*}{g_2}\circ T}\,d\lambda + \int_{\mathcal{X}_{12}} \overset{*}{f_1}\,\mathrm{Log}\,\frac{\overset{*}{f_1}\cdot\overset{*}{g_2}\circ T}{\overset{*}{f_2}\cdot\overset{*}{g_1}\circ T}\,d\lambda = \alpha + \beta$$

Le théorème de transfert donne $\alpha = \overset{*}{\psi}_T(1,2)$, le lemme 2.2 appliqué à β avec $\overset{*}{f_1}$ et $\dfrac{\overset{*}{f_2}\cdot\overset{*}{g_1}\circ T}{\overset{*}{g_2}\circ T}$ montre que $\beta \geq 0$ avec égalité si et seulement si

$$\frac{f_1}{f_2} = \frac{g_1\circ T}{g_2\circ T} \text{ sur } \mathcal{X}_{12}$$

c'est à dire par 1.3.2 si T est exhaustif.

On en déduit :

4.4.2 Proposition. Avec les notations de 1.4, soient $(X_1,\ldots,X_n)$ un n-échantillon de X et S une statistique du n-échantillon, on a :

$$\overset{*}{\psi}_n(1,2) = n\cdot\overset{*}{\psi}_X(1,2) \geq \overset{*}{\psi}_S(1,2)$$

$$\psi_n(1,2) = n\cdot\psi_X(1,2) \geq \psi_S(1,2)$$

avec égalité si et seulement si S est un résumé exhaustif du n-échantillon.

4.5 PROPRIETES DE LA DIVERGENCE.

4.5.1 **Proposition.** Avec les notations de 3.5, nous avons :

$$\varphi(1,2) \geq 0$$

avec égalité si et seulement si $f_1^* = f_2^*$ λ - p.s. sur $\mathcal{X}_{12}$

c'est à dire si $\dfrac{f_1}{f_2} = \dfrac{p_{12}}{p_{21}}$ = constante sur $\mathcal{X}_{12}$ λ - p.s..

4.5.2 **Proposition.** Avec les notations de 4.3, soient X et Y deux variables aléatoires indépendantes, nous avons :

$$\varphi_{X,Y}(1,2) = \varphi_X(1,2) + \varphi_Y(1,2)$$

4.5.3 **Proposition.** Avec les notations de 1.3, soit T une application mesurable de $(\mathcal{X},\mathcal{A})$ dans $(\mathcal{Y},\mathcal{B})$, nous avons :

$$\varphi_X(1,2) \geq \varphi_T(1,2)$$

avec égalité si et seulement si T est exhaustif.

4.5.4 **Proposition.** Avec les notations de 1.4, soient $(X_1,\ldots,X_n)$ un n-échantillon de X et S une statistique du n-échantillon on a :

$$\varphi_n(1,2) = n.\varphi_X(1,2) \geq \varphi_S(1,2)$$

avec égalité si et seulement si S est un résumé exhaustif du n-échantillon.

Les démonstrations de ces quatre propositions sont évidentes à partir de la définition 3.5.1 de φ et de 4.2.1, 4.3.1, 4.4.1 et 4.4.2.

§ 5 INEGALITE DE KULLBACK

Dans ce paragraphe, nous donnons la généralisation de la notion de "minimum discrimination information" dans le cas où $\mathcal{P}$ n'est pas homogène. A partir de cette inégalité Kullback démontre un certain nombre de relations et en particulier, sous certaines conditions, l'inégalité de Cramer-Rao ; mais il semble plus simple et plus général de démontrer directement cette inégalité (7.2) que de la déduire de la proposition 5.1.

Avec les notations de 3.1, soient f_1, f_2 les densités de P_1, P_2, et T une variable aléatoire réelle définie sur $(\mathcal{X}, \mathcal{A})$, P_1 - intégrable et vérifiant :

$$- \quad 0 < M_2 = \int_{\mathcal{X}_{12}} e^T . f_2^* \; d\lambda < + \infty$$

$$- \quad \theta = \int_{\mathcal{X}_{12}} T \cdot f_1^* \; d\lambda$$

5.1 **Proposition.** L'information en faveur de H_1 contre H_2 vérifie :

$$- \quad \psi^*(1,2) \geq \theta - \text{Log } M_2$$

$$- \quad \psi(1,2) \geq \theta - \text{Log } M_2 + \text{Log } \frac{p_{12}}{p_{21}}$$

avec égalité si et seulement si

$$f_1 = \frac{p_{12}}{p_{21}} \cdot \frac{e^T}{M_2} \cdot f_2 \qquad \lambda - \text{p.s. sur } \mathcal{X}_{12}.$$

Il suffit d'appliquer la remarque 4.1.1 au résultat de [5] si l'on note $\tilde{P}$ la mesure de probabilité sur $(\mathcal{X}, \mathcal{A})$ dont la densité est :

$$\begin{cases} \tilde{f} = \dfrac{P_{12}}{P_{21}} \dfrac{e^T}{M_2} f_2 & \text{sur } \mathcal{X}_{12}. \\ \tilde{f} = f_1 & \text{sur } \mathcal{X} - \mathcal{X}_{12}. \end{cases}$$

la proposition 5.1 peut s'exprimer sous la forme

5.2 <u>Proposition.</u> Si $\tilde{P}$ vérifie $\theta = \int_{\mathcal{X}_{12}} T \, \tilde{f}^* \, d\lambda$ alors

$$\psi^*(1,2) \geq \psi^*(\sim,2) = \theta - \text{Log } M_2.$$

$$\psi(1,2) \geq \psi(\sim,2) = \theta - \text{Log } M_2 + \text{Log } \frac{P_{12}}{P_{21}}$$

avec égalité si et seulement si $f_1 = \tilde{f}$ λ - p.s.

$$\tilde{f}^* = \frac{\tilde{f}}{\tilde{P}(\mathcal{X}_{12})} = \frac{\tilde{f}}{P_{12}} \qquad \text{car } \mathcal{X}_{12} = \mathcal{X}_{\sim}.\mathcal{X}_2$$

5.3 <u>Proposition.</u> Si $\tilde{P}$ vérifie

$$\theta_1 = \int_{\mathcal{X}_{12}} T \, f_1^* \, d\lambda = \int_{\mathcal{X}_{12}} T \, \tilde{f}^* \, d\lambda$$

$$\theta_2 = \int_{\mathcal{X}_{12}} T \, f_2^* \, d\lambda$$

On a : $\varphi(\sim,2) = \theta_1 - \theta_2$

car $\varphi(\sim,2) = \theta_1 - \text{Log } M_2 + \int_{\mathcal{X}_{12}} f_2^* \text{ Log } \dfrac{f_2^*}{\tilde{f}^*} = \theta_1 - \int_{\mathcal{X}_{12}} f_2^* . T \, d\lambda$

Ch III QUANTITE D'INFORMATION DE FISHER GENERALISEE

Dans ce chapitre nous proposons une définition originale de la quantité d'information de Fisher (Q.I.F.) directement dérivée de la divergence et de l'information selon Kullback, étudiées au chapitre II.

Cette notion se confond avec la notion classique quand les conditions de régularité, habituellement demandées pour obtenir l'existence et les principales propriétés de la Q.I.F., sont vérifiées ; mais nous montrerons que dans le cas général, nous conservons la presque totalité des "bonnes" propriétés de la Q.I.F. et en particulier l'inégalité de Cramer - Rao.

§ 6 DEFINITIONS ET PROPRIETES

6.1 NOTATIONS.

Soit $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}\}$ une famille de probabilités sur $(\mathcal{X}, \mathcal{A})$, dominée par la mesure positive σ - finie λ :

$$0 \leq f_\theta = \frac{dP_\theta}{d\lambda} < + \infty \qquad \lambda - \text{p.s.}$$

Nous supposons dans ce chapitre que Θ *est un ouvert de* $\mathbb{R}$, *que* $\mathcal{P}$ *vérifie l'hypothèse* A-0 et comme précédemment nous notons :

- $\mathcal{X}_\theta = \{x \in \mathcal{X} : f_\theta(x) > 0\}$ et $\mathcal{X}_{01} = \mathcal{X}_{\theta_0} . \mathcal{X}_{\theta_1}$

- $f_i^* = \dfrac{f_i}{P_i(\mathcal{X}_{01})}$ $i = 0,1,$ $P_{01} = P_0(\mathcal{X}_{01})$ et $P_{10} = P_1(\mathcal{X}_{01})$

- $\varphi(\theta_0,\theta_1)$ la partie finie de la divergence entre θ_1 et θ_0 c'est à dire

$$\varphi(\theta_0,\theta_1) = \int_{\mathcal{X}_{01}} (f_0^* - f_1^*) \operatorname{Log} \frac{f_0^*}{f_1^*} \, d\lambda$$

6.2 DEFINITION.

6.2.1 Définition. On appelle *quantité d'information de Fisher généralisée* (Q.I.F.G.) *sur le paramètre* $\theta_0 \in \Theta$, la quantité finie, si elle existe, définie par :

$$F(\theta_0) = \frac{1}{2} \frac{d^2}{d\theta^2} \varphi(\theta_0,\theta) \Big|_{\theta=\theta_0} = \frac{1}{2} \frac{d^2}{d\theta^2} \varphi(\theta,\theta_0) \Big|_{\theta=\theta_0}$$

C'est à dire, pour $\theta_0 \in \Theta$ fixé, nous supposons que $\varphi(\theta_0,\theta)$ possède une dérivée seconde par rapport à θ en $\theta=\theta_0$. La symétrie de φ nous montre que $\varphi''(\theta_0,\theta)\Big|_{\theta=\theta_0} = \varphi''(\theta,\theta_0)\Big|_{\theta=\theta_0}$

6.2.2 Proposition. Si la Q.I.F.G. existe pour $\theta_0 \in \Theta$ alors

$$F(\theta_0) = \lim_{\theta \to \theta_0} \frac{\varphi(\theta_0,\theta)}{(\theta-\theta_0)^2} = \lim_{\theta \to \theta_0} \frac{\varphi(\theta,\theta_0)}{(\theta-\theta_0)^2}$$

Si $F(\theta_0)$ existe, nous pouvons écrire la 2ème formule de Taylor pour tout θ assez voisin de θ_0 :

$$\varphi(\theta_0,\theta) = \varphi(\theta_0,\theta_0) + (\theta-\theta_0)\,\frac{d}{d\theta}\varphi(\theta_0,\theta)\Big|_{\theta=\theta_0} + \frac{(\theta-\theta_0)^2}{2}\,\frac{d^2}{d\theta^2}\varphi(\theta_0,\theta)\Big|_{\theta=\theta_0}$$

$$+ \frac{(\theta-\theta_0)^2}{2}\,f(\theta)$$

avec $\lim\limits_{\theta\to\theta_0} f(\theta) = 0$

La proposition 4.5.1 nous donne : $\varphi(\theta_0,\theta) \geq 0$ et $\varphi(\theta_0,\theta_0) = 0$. $\varphi(\theta_0,\theta)$ passe donc par un minimum en θ_0, comme la dérivée première $\varphi'(\theta_0,\theta)$ existe dans un voisinage de θ_0, il vient : $\varphi'(\theta_0,\theta)\Big|_{\theta=\theta_0} = 0$

et par suite :

$$F(\theta_0) = \frac{\varphi(\theta_0,\theta)}{(\theta-\theta_0)^2} - \frac{1}{2}\,f(\theta)$$

en passant à la limite on obtient le résultat désiré.

<u>Remarque</u> : Si l'on considère $\varphi(\theta_0,\theta)$ comme une fonction positive, nulle pour $\theta=\theta_0$, alors la réciproque de cette proposition n'est pas vraie car la limite peut exister sans que φ soit deux fois dérivable /θ en θ_0, par exemple :

pour $\varphi(\theta_0,\theta) = (\theta-\theta_0)^4\,(\cos\frac{1}{\theta-\theta_0} + 2)$ nous avons

$$\lim \frac{\varphi(\theta_0,\theta)}{(\theta-\theta_0)^2} = 0 \text{ mais } \frac{d^2}{d\theta^2}\varphi(\theta_0,\theta)\Big|_{\theta=\theta_0} \text{ n'est pas définie.}$$

6.3 PROPRIETE.

Comme dans les chapitres précédents (3.6), nous noterons le cas échéant : $\varphi_X(\theta_0,\theta)$ et $F_X(\theta_0)$ pour indiquer précisément que φ et F sont calculées à partir des densités f_{θ_0}, f_θ correspondant à la variable aléatoire X.

6.3.1 Proposition. $F(\theta_0) \geq 0$

Si $\dfrac{f_\theta}{f_{\theta_0}} = C(\theta_0,\theta)$ λ - p.s. sur $\mathcal{X}_{\theta\theta_0}$, pour tous les θ appartenant à un voisinage de θ_0, alors $F(\theta_0) = 0$. $C(\theta_0,\theta)$ étant une constante, indépendante λ - p.s. de $x \in \mathcal{X}_{\theta\theta_0}$, qui peut dépendre de (θ_0,θ).

De la proposition 4.5.1 : $\varphi(\theta_0,\theta) \geq 0$ on déduit

$$F(\theta_0) = \lim_{\theta\to\theta_0} \frac{\varphi(\theta_0,\theta)}{(\theta-\theta_0)^2} \geq 0$$

La condition $\dfrac{f_\theta}{f_{\theta_0}}$ = constante sur $\mathcal{X}_{\theta\theta_0}$ => $\varphi(\theta_0,\theta) = 0$ et $\varphi(\theta_0,\theta) = 0$ pour tous les θ dans un voisinage de θ_0 => $F(\theta_0) = 0$.

6.3.2 Proposition. Avec les notations du paragraphe 4.3, si X et Y sont des variables aléatoires indépendantes, on a :

$$F_{X,Y}(\theta_0) = F_X(\theta_0) + F_Y(\theta_0)$$

La relation est obtenue immédiatement à partir de la définition et de la proposition 4.5.2 :

$$\varphi_{X,Y}(\theta_0,\theta) = \varphi_X(\theta_0,\theta) + \varphi_Y(\theta_0,\theta)$$

6.3.3 Proposition. Avec les notations du paragraphe 1.3, si T est une application mesurable de $(\mathcal{X}, \mathcal{A})$ dans $(\mathcal{Y}, \mathcal{B})$ nous avons :

$$F_X(\theta_0) \geq F_T(\theta_0)$$

De plus, si T est un résumé exhaustif de X pour $\theta_0 \in \Theta$, alors $F_X(\theta_0) = F_T(\theta_0)$.

De la proposition 4.5.3 : $\varphi_X(\theta_0,\theta) \geq \varphi_T(\theta_0,\theta)$ on déduit

$$F_X(\theta_0) = \lim_{\theta\to\theta_0} \frac{\varphi_X(\theta_0,\theta)}{(\theta-\theta_0)^2} \geq \lim_{\theta\to\theta_0} \frac{\varphi_T(\theta_0,\theta)}{(\theta-\theta_0)^2} = F_T(\theta_0)$$

Si T est exhaustif $\varphi_X = \varphi_T \Rightarrow F_X = F_T$.

De 6.3.1, 6.3.2 et 6.3.3 avec les notations de 1.4, on déduit

6.3.4 Proposition. Si $(X_1,X_2,\ldots,X_n)$ est un n-échantillon de X et S une statistique de ce n-échantillon, on a :

$$F_{X_1,\ldots,X_n}(\theta_0) = F_n(\theta_0) = n.F_X(\theta_0) \geq F_S(\theta_0) \geq 0$$

De plus, si S est un résumé exhaustif du n-échantillon pour $\theta_0 \in \Theta$, alors : $F_n(\theta_0) = F_S(\theta_0)$.

6.3.5 Remarque : Notre Q.I.F.G. possède donc les propriétés de la Q.I.F. classique sauf le fait que $F_X = F_T$, dans 6.3.3, n'entraîne pas nécessairement l'exhaustivité. Il ne semble pas exister de conditions simples pour conserver cette propriété.

6.4 HYPOTHESE B.

Si $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ vérifie les hypothèses A.1 et A.2 de 2.7, $\mathcal{P}$ vérifie l'hypothèse B :

B

1) $\mathcal{P}$ vérifie A.1
2) Θ est un ouvert de $\mathbb{R}$
3) $\forall\, \theta_0 \in \Theta$, $\forall\, \varepsilon > 0$, $\exists\, \nu(\theta_0,\varepsilon) > 0$ tel que

$\theta_1 \in \Theta$, $\theta_1 \neq \theta_0$, $|\theta_1-\theta_0| < \nu \Rightarrow$

- $P_0(\mathcal{X}_{01}) > 1-\varepsilon$
- le rapport de deux quelconques des cinq quantités suivantes, supposées strictement positives, est compris entre $1-\varepsilon$ et $1+\varepsilon$.

$$A(\theta_0,\theta_1) = \int_{\mathcal{X}_{01}} \frac{(f_0^* - f_1^*)^2}{f_0^*}\, d\lambda \quad \text{et} \quad A(\theta_1,\theta_0)$$

$$2\psi^*(\theta_0,\theta_1) = 2\int_{\mathcal{X}_{01}} f_0^* \operatorname{Log} \frac{f_0^*}{f_1^*}\, d\lambda \quad \text{et } 2\psi^*(\theta_1,\theta_0)$$

$$\varphi(\theta_0,\theta_1) = \int_{\mathcal{X}_{01}} (f_0^* - f_1^*) \operatorname{Log} \frac{f_0^*}{f_1^*}\, d\lambda$$

6.4.1 Proposition. Notons $B(\theta_0,\theta)$ une quelconque des quatre quantités $A(\theta_0,\theta)$, $A(\theta,\theta_0)$, $2\psi^*(\theta_0,\theta)$, $2\psi^*(\theta,\theta_0)$, si $\mathcal{P}$ vérifie l'hypothèse B et si la Q.I.F.G. $F(\theta_0)$ existe alors :

$$F(\theta_0) = \lim_{\theta\to\theta_0} \frac{B(\theta_0,\theta)}{(\theta-\theta_0)^2}$$

De plus, si $B(\theta_0,\theta)$ est deux fois dérivable par rapport à θ en θ_0, nous avons :

$$F(\theta_0) = \frac{1}{2} \frac{d^2}{d\theta^2} B(\theta_0,\theta)\Big|_{\theta=\theta_0}$$

Si $F(\theta_0)$ existe, 6.2.2 => $\quad F(\theta_0) = \lim\limits_{\theta\to\theta_0} \dfrac{\varphi(\theta_0,\theta)}{(\theta-\theta_0)^2}$

mais l'hypothèse B entraîne :

$$1 = \lim_{\theta\to\theta_0} \frac{\varphi(\theta_0,\theta)}{B(\theta_0,\theta)} = \lim_{\theta\to\theta_0} \frac{\varphi(\theta_0,\theta)}{(\theta-\theta_0)2} \cdot \lim_{\theta\to\theta_0} \frac{(\theta-\theta_0)^2}{B(\theta_0,\theta)}$$

on en déduit le premier résultat demandé.

De plus, si $B(\theta_0,\theta)$ est deux fois différentiable /θ enθ_0, la formule de Taylor et une démonstration analogue à celle de la proposition 6.2.2 nous donnent $\quad F(\theta_0) = \dfrac{1}{2} \dfrac{d^2}{d\theta^2} B(\theta_0,\theta) \Big|_{\theta=\theta_0}$

On peut faire une remarque similaire à celle de la proposition 6.2.2 indiquant qu'il est indispensable de supposer $B(\theta_0,\theta)$ deux fois différentiable /θ en θ_0 pour que la limite soit égale à la moitié de la dérivée seconde.

6.5 DEFINITION CLASSIQUE DE LA QUANTITE D'INFORMATION DE FISHER (Q.I.F.).

Pour définir la Q.I.F., au sens classique, et obtenir les propriétés équivalentes à celles des paragraphes 6 et 7, on suppose habituellement un certain nombre de conditions de régularité, en particulier $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ doit être homogène, [2],[8].

Cette dernière condition, nous restreint au cas de l'information et de la divergence étudiées par Kullback où l'on a :

$$I(\theta_0,\theta) = \psi^*(\theta_0,\theta) \quad \text{et} \quad J(\theta_0,\theta) = \varphi(\theta_0,\theta).$$

On trouvera dans [4] les hypothèses et la démonstration qui conduisent aux résultats suivants (voir aussi [3]) :

6.5.1 La Q.I.F. au sens classique est définie par

$$\mathcal{F}(\theta) = \int \left(\frac{\partial}{\partial\theta} \text{ Log } f(.,\theta)\right)^2 f(.,\theta)\, d\lambda$$

$$= \text{Var}_\theta \left\{ \frac{\partial}{\partial\theta} \text{Log } f(X,\theta)\right\}$$

car la variable aléatoire $\frac{\partial}{\partial\theta}$ Log $f(X,\theta)$ est centrée.

6.5.2 Sous les conditions de régularité de [4] :

$$\mathcal{F}(\theta_0) = \lim_{\theta\to\theta_0} \frac{J(\theta_0,\theta)}{(\theta-\theta_0)^2} = 2 \lim_{\theta\to\theta_0} \frac{I(\theta_0,\theta)}{(\theta-\theta_0)^2}$$

La démonstration se conduit en montrant d'abord que :

$$\lim_{\theta\to\theta_0} \frac{A(\theta_0,\theta)}{(\theta-\theta_0)^2} = \lim_{\theta\to\theta_0} \frac{J(\theta_0,\theta)}{(\theta-\theta_0)^2} = 2 \lim_{\theta\to\theta_0} \frac{I(\theta_0,\theta)}{(\theta-\theta_0)^2}$$

et ensuite que :

$$\mathcal{F}(\theta_0) = \lim_{\theta\to\theta_0} \frac{A(\theta_0,\theta)}{(\theta-\theta_0)^2}$$

Nous pouvons donc en conclure

6.5.3 Si $\mathcal{P}$ vérifie les conditions de régularité de [4] :

- $F(\theta_0) = \mathcal{F}(\theta_0)$
- $\mathcal{P}$ vérifie l'hypothèse B.

§ 7 INEGALITE DE CRAMER - RAO

Dans ce paragraphe nous présentons des inégalités qui généralisent l'inégalité dite de "Cramer - Darmois - Frechet - Rao". Nous nous inspirons de la méthode de Chapman et Robbins dans [1], notre contribution porte d'une part sur le fait que nous ne supposons pas $\mathcal{P}$ homogène et d'autre part nous montrons qu'une quantité introduite dans [1],[3] et [8], n'est autre que la Q.I.F.G. du paragraphe 6 et cela dans le cas général et non pas uniquement sous les conditions de régularité de 6.5.

Reprenons les notations et les hypothèses de 6.1 et supposons de plus que $\mathcal{P}$ vérifie l'hypothèse A-1 de 2.7, c'est à dire en particulier que l'intégrale suivante existe :

$$A_X(\theta_0,\theta_1) = \int_{\mathcal{X}_{01}} \frac{(f_0^* - f_1^*)^2}{f_0^*} \, d\lambda$$

avec $f_i^* = \dfrac{f_i}{P_i(\mathcal{X}_{01})}$ $\quad i = 0,1$ et $P_{01} = P_0(\mathcal{X}_{01})$, $P_{10} = P_1(\mathcal{X}_{01})$

Nous supposerons que T est une variable aléatoire réelle définie sur $(\mathcal{X}, \mathcal{A}, P_\theta)$, du second ordre $\forall\ \theta \in \Theta$ et nous noterons :

$$E_\theta(T) = \int_{\mathcal{X}_\theta} T\, f_\theta \, d\lambda, \qquad \mathrm{Var}_\theta\, T = \int_{\mathcal{X}_\theta} (T - E_\theta(T))^2 \, f_\theta \, d\lambda$$

$$E_{01} = E_{01}(T) = \int_{\mathcal{X}_{01}} T\, f_0^* \, d\lambda \quad \text{et} \quad E_{10} = E_{10}(T) = \int_{\mathcal{X}_{01}} T\, f_1^* \, d\lambda$$

7.1 INEGALITES FONDAMENTALES.

7.1.1 Proposition. Avec les notations précisées ci-dessus, si $\mathcal{P}$ vérifie l'hypothèse A-1 et si T est du second ordre $\forall\ \theta_0 \in \Theta$ alors $\forall\ (\theta_0,\theta_1) \in \Theta^2$:

$$\mathrm{Var}_{\theta_0}\ T.A_X(\theta_0,\theta_1) \geq \left(E_{01}(T) - E_{10}(T)\right)^2 .P_{01}$$

Dans le cas particulier où $\mathcal{P}$ est homogène l'inégalité devient :

$$\mathrm{Var}_{\theta_0}\ T.A_X(\theta_0,\theta_1) \geq \left(E_0\ (T) - E_1(T)\right)^2$$

L'inégalité de Schwartz nous donne :

$$\int_{\mathcal{X}_{01}} (T-E_{01})^2\ f_0^*\ d\lambda\ .\int_{\mathcal{X}_{01}} \frac{(f_0^* - f_1^*)^2}{f_0^{*2}}\ f_0^*\ d\lambda \geq \left[\int_{\mathcal{X}_{01}} (T-E_{01})\ \frac{f_0^* - f_1^*}{f_0^*}\ f_0^*\ d\lambda\right]^2$$

1) Nous avons $\dfrac{\mathrm{Var}_{\theta_0}\ T}{P_{01}} \geq \displaystyle\int_{\mathcal{X}_{01}} (T-E_{01})^2\ f_0^*\ d\lambda$ en effet

$$\frac{\mathrm{Var}_{\theta_0}\ T}{P_{01}} = \int_{\mathcal{X}_0 - \mathcal{X}_{01}} (T-E_0(T))^2\ f_0^*\ d\lambda + \int_{\mathcal{X}_{01}} (T-E_0(T))^2\ f_0^* d\lambda = a + b$$

$$b = \int_{\mathcal{X}_{01}} (T-E_{01})^2\ f_0^*\ d\lambda + [E_{01} - E_0(T)]^2 = b_1 + b_2$$

$$\frac{\mathrm{Var}_{\theta_0}\ T}{P_{01}} = a + b_1 + b_2 \qquad \text{avec } a \geq 0,\ b_2 \geq 0 \qquad \text{donc}$$

$$\frac{\mathrm{Var}_{\theta_0}\ T}{P_{01}} \geq b_1 \qquad \text{ce que nous cherchions}$$

$$2) \int_{\mathcal{X}_{01}} (T-E_{01}) \frac{f_0^* - f_1^*}{f_0^*} f_0^* \, d\lambda = E_{01} - E_{10}$$ il vient

$$\mathrm{Var}_{\theta_0} T . A_X(\theta_0,\theta_1) \geq (E_{01}(T) - E_{10}(T))^2 . P_{01}$$

Si $\mathcal{P}$ est homogène nous avons $P_{\theta_1} \equiv P_{\theta_0} \Rightarrow f_0^* = f_0$, $f_1^* = f_1$,

$E_{01}(T) = E_0(T)$, $E_{10}(T) = E_1(T)$ et $P_{01} = 1$.

Remarquons que : $A_X(\theta_0,\theta_1) = 0 \iff f_0^* = f_1^*$ λ - p.s.

$\iff \frac{f_0}{f_1} =$ constante λ - p.s. $\Rightarrow E_{10} = E_{01}$ et l'inégalité dans ce cas est sans intérêt $0 \geq 0$.

7.1.2 **Proposition.** On déduit immédiatement de 7.1.1 :

$$\mathrm{Var}_{\theta_0} T \geq \sup_{\theta_1 \in \Theta_0} \left\{ \frac{(E_{01}(T) - E_{10}(T))^2}{A_X(\theta_0,\theta_1)} P_{01} \right\}$$

avec $\Theta_0 = \{\theta_1 \in \Theta : A(\theta_0,\theta_1) > 0\}$.

Dans le cas particulier où $\mathcal{P}$ est homogène, cette inégalité devient :

$$\mathrm{Var}_{\theta_0} T \geq \sup_{\theta_1 \in \Theta} \frac{(E_0(T) - E_1(T))^2}{A_X(\theta_0,\theta_1)}$$

Dans le cas $\mathcal{P}$ homogène $\Theta_0 = \Theta$ car $\forall(\theta_0,\theta_1) \in \Theta^2$ nous avons $P_{01} = P_0(\mathcal{X}_{01}) = 1$ et de plus nous avons supposé $f_{\theta_0} \neq f_\theta$ λ -p.s. si $\theta \neq \theta_0$ (1.2)

7.2 INEGALITE DE CRAMER - RAO GENERALISEE.

Nous étudions maintenant ce que deviennent les inégalités de 7.1.1 quand $\theta_1 \to \theta_0$, et quand l'hypothèse B de 2.7 est vérifiée.

7.2.1 **Proposition.** Si T est du second ordre pour $\theta_0 \in \Theta$, si $\mathcal{P}$ vérifie l'hypothèse B et si la Q.I.F.G. $F_X(\theta_0)$ existe alors :

$$\mathrm{Var}_{\theta_0} T . F_X(\theta_0) \geq \limsup_{\theta_1 \to \theta_0} \left(\frac{E_{01}(T) - E_{10}(T)}{\theta_0 - \theta_1} \right)^2$$

Dans le cas particulier où $\mathcal{P}$ est homogène, l'inégalité s'écrit :

$$\mathrm{Var}_{\theta_0} T . F_X(\theta_0) \geq \limsup_{\theta_1 \to \theta_0} \left(\frac{E_0(T) - E_1(T)}{\theta_0 - \theta_1} \right)^2$$

si de plus dans ce cas, $E_\theta(T)$ est dérivable par rapport à θ en θ_0 alors :

$$\mathrm{Var}_{\theta_0} T . F_X(\theta_0) \geq \left(\frac{d}{d\theta} E_\theta(T) \Big|_{\theta = \theta_0} \right)^2$$

Cette dernière inégalité est très exactement *l'inégalité de Cramer - Rao* mais établie par une démonstration originale sous des hypothèses plus faibles que dans la démonstration classique.

En divisant les deux membres des inégalités de 7.1.1 par $(\theta_0 - \theta_1)^2$ et en remarquant que sous l'hypothèse B

$$- \lim_{\theta_1 \to \theta_0} P_{01} = P_0(\mathcal{X}_{01}) = 1$$

$$- \lim_{\theta_1 \to \theta_0} \frac{A_X(\theta_0, \theta_1)}{(\theta_0 - \theta_1)^2} = F_X(\theta_0) \qquad \text{par 6.4.1}$$

le passage à la limite nous procure les deux premières inégalités.

De plus supposer $E_\theta(T)$ dérivable $/\theta$ en θ_0 c'est dire

$$\lim_{\theta_1 \to \theta_0} \frac{E_{\theta_1}(T) - E_{\theta_0}(T)}{\theta_1 - \theta_0} = \frac{d}{d\theta} E_\theta(T) \Big|_{\theta = \theta_0}$$

cela nous donne l'inégalité de Cramer - Rao.

Dans ce paragraphe T est une variable aléatoire réelle du second ordre telle que : $(\mathcal{X}, \mathcal{A}, P_\theta) \xrightarrow{T} (\mathbb{R}, \mathcal{R}, Q_\theta)$, avec les notations des paragraphes 1.3 et 6.3, en identifiant $(\mathcal{Y}, \mathcal{B})$ à $(\mathbb{R}, \mathcal{R})$, il vient :

$$E_\theta(T) = \int_{\mathcal{X}_\theta} T\, f_\theta\, d\lambda = \int_{\mathcal{Y}_\theta} y\, g_\theta(y)\, d\gamma(y)$$

$$\mathrm{Var}_\theta T = \int_{\mathcal{Y}_\theta} (y - E_\theta(T))^2\, g_\theta(y)\, d\gamma(y)$$

$$E_{01}(T) = \int_{\mathcal{Y}_{01}} y\, g_1^*(y)\, d\gamma(y) \qquad \text{etc}$$

Nous pouvons alors énoncer sans difficulté les propositions 7.1.1, 7.1.2, 7.2.1, dans l'espace $(\mathbb{R}, \mathcal{R}, Q_\theta)$ et non plus $(\mathcal{X}, \mathcal{A}, P_\theta)$ en substituant $\mathcal{Q} = \{Q_\theta : \theta \in \Theta\}$ à $\mathcal{P}$ et en remplaçant $A_X(\theta_0,\theta_1)$, $F_X(\theta_0)$ par les quantités $A_T(\theta_0,\theta_1)$, $F_T(\theta_0)$ calculées à partir des g_θ. (Rappelons que : $\mathcal{P}$ homogène $\Leftrightarrow$ $\mathcal{Q}$ homogène). Contentons-nous d'énoncer les deux propositions suivantes déduites de 7.2.1.

7.2.2 **Proposition.** Si T est du second ordre, si $\mathcal{P}$ et $\mathcal{Q}$ vérifient l'hypothèse B et si les Q.I.F.G. $F_X(\theta_0)$ et $F_T(\theta_0)$ existent alors :

$$\mathrm{Var}_{\theta_0} T . F_X(\theta_0) \geq \mathrm{Var}_{\theta_0} T . F_T(\theta_0) \geq \limsup_{\theta_1 \to \theta_0} \left(\frac{E_{01}(T) - E_{10}(T)}{\theta_0 - \theta_1} \right)^2$$

Dans le cas particulier où $\mathcal{P}$ est homogène et où $E_\theta(T)$ est dérivable /θ en θ_0, les inégalités s'écrivent :

$$\mathrm{Var}_{\theta_0} T . F_X(\theta_0) \geq \mathrm{Var}_{\theta_0} T . F_T(\theta_0) \geq \left(\frac{d}{d\theta} E_\theta(T) \Big|_{\theta=\theta_0} \right)^2$$

7.2.3 | <u>Proposition</u>. Avec les notations de 1.4 et 6.3, soient $(X_1,\ldots,X_n)$ un n-échantillon de X et S une statistique réelle de ce n-échantillon. Si S est du second ordre, si $\mathcal{P}$ et $\mathcal{Q}$ vérifient l'hypothèse B et si les Q.I.F.G. $F_X(\theta_0)$ et $F_S(\theta_0)$ existent, alors :

$$\mathrm{Var}_{\theta_0} S.F_n(\theta_0) = n.\mathrm{Var}_{\theta_0} S.F_X(\theta_0) \geq \mathrm{Var}_{\theta_0} S.F_S(\theta_0)$$
$$\geq \limsup_{\theta_1 \to \theta_0} \left(\frac{E_{01}(S) - E_{10}(S)}{\theta_0 - \theta_1}\right)^2$$

Dans le cas particulier où $\mathcal{P}$ est homogène et où $E_\theta(S)$ est dérivable /θ en θ_0, ces inégalités s'écrivent :

$$\mathrm{Var}_{\theta_0} S.F_n(\theta_0) = n.\mathrm{Var}_{\theta_0} S.F_X(\theta_0) \geq \mathrm{Var}_{\theta_0} S.F_S(\theta_0)$$
$$\geq \left(\frac{d}{d\theta} E_\theta(S)\bigg|_{\theta=\theta_0}\right)^2$$

La première inégalité de 7.2.2 est évidente car

$$F_X(\theta_0) \geq F_T(\theta_0) \qquad \text{par 6.3.3}$$

et la seconde inégalité est la transposition dans $(\mathbb{R}, \mathcal{R}, Q_\theta)$ de 7.2.1.
La proposition 7.2.3 se déduit directement de 7.2.2.

7.3 EFFICACITE.

Plaçons-nous dans le cas où $\mathcal{P}$ est homogène, sous les hypothèses définies dans la proposition 7.2.2, nous avons :

$$\mathrm{Var}_\theta T \geq \frac{\left(\frac{d}{d\theta} E_\theta(T)\right)^2}{F_T(\theta)} \geq \frac{\left(\frac{d}{d\theta} E_\theta(T)\right)^2}{F_X(\theta)}$$

Le terme central donne une minoration plus fine de $Var_\theta T$ que le terme de droite, mais celui-ci sous certaines conditions peut donner une minoration de $Var_\theta T$ indépendante de T. Supposons que T est un estimateur sans biais de $h(\theta)$ de variance finie, $h(\theta)$ étant une fonction réelle dérivable de θ.

Nous aurons $\quad E_\theta(T) = h(\theta) \qquad \forall\ \theta \in \Theta \qquad$ et

$$Var_\theta T \geq \frac{\left(\frac{d}{d\theta} h(\theta)\right)^2}{F_T(\theta)} \geq \frac{\left(\frac{d}{d\theta} h(\theta)\right)^2}{F_X(\theta)}$$

Définissons

$$M_X(\theta,h) = \sup_{\theta_1 \in \Theta_0} \frac{\left(h(\theta) - h(\theta_1)\right)^2}{A_X(\theta,\theta_1)} = \sup_{\theta_1 \in \Theta_0} \frac{\left(E_\theta(T) - E_{\theta_1}(T)\right)^2}{A_X(\theta,\theta_1)}$$

pour $\theta \in \Theta$ fixé : $\quad \Theta_0 = \{\theta_1 \in \Theta \ : A_X(\theta,\theta_1) > 0\}$

A partir de 7.1.2 et 7.2.1 il vient :

$$Var_\theta T \geq M_X(\theta,h) \geq \frac{\left(\frac{d}{d\theta} h(\theta)\right)^2}{F_X(\theta)}$$

Les deux termes de droite sont indépendants de T, nous pouvons alors donner une définition de l'efficacité plus précise que dans le cas classique.

7.3.1 <u>Définition</u> : Si $\mathscr{P}$ est homogène et vérifie l'hypothèse B, si la Q.I.F.G. $F_X(\theta)$ existe, si T est un estimateur du second ordre sans biais de $h(\theta)$, $h(\theta)$ étant une fonction réelle dérivable de $\theta \in \Theta$ alors :

a) $\mathscr{P}$ *permet un estimateur efficace de* $h(\theta)$, si :

$$M_X(\theta,h) = \frac{\left(\frac{d}{d\theta} h(\theta)\right)^2}{F_X(\theta)} \qquad \forall\ \theta \in \Theta$$

et T est un *estimateur efficace de* h(θ), si :

$$\mathrm{Var}_\theta T = M_X(\theta,h) = \frac{\left(\frac{d}{d\theta}\ h(\theta)\right)^2}{F_X(\theta)} \qquad \forall\ \theta \in \Theta$$

b) $\mathcal{P}$ *ne permet pas d'estimateur efficace de* h(θ), si :

$$\exists \theta \in \Theta : M_X(\theta,h) > \frac{\left(\frac{d}{d\theta}\ h(\theta)\right)^2}{F_X(\theta)}$$

et T est un *estimateur de variance minimum* si :

$$\mathrm{Var}_\theta T = M_X(\theta,h) \qquad \forall\ \theta \in \Theta$$

Une étude analogue à partir de la proposition 7.2.3 nous conduit à énoncer d'une autre manière les définitions précédentes.

7.3.2 Définition. Avec les notations de 1.4, soient $(X_1,\ldots,X_n)$ un n-échantillon de X et S une statistique réelle de ce n-échantillon. Si S est un estimateur du second ordre sans biais de h(θ), h(θ) étant une fonction réelle dérivable, si $\mathcal{P}$ est homogène et vérifie l'hypothèse B et si la Q.I.F.G. $F_X(\theta_0)$ existe, nous pouvons définir :

$$A_n(\theta,\theta_1) = \int_{\mathcal{X}^n_{\theta\theta_1}} \frac{[L(.,\theta) - L(.,\theta_1)]^2}{L(.,\theta)}\ d\lambda$$

$$M_n(\theta,h) = \sup_{\theta_1 \in \Theta_0} \frac{\left(h(\theta) - h(\theta_1)\right)^2}{A_n(\theta,\theta_1)}$$

pour $\theta \in \Theta$ fixé : $\Theta_0 = \{\theta_1 \in \Theta : A_n(\theta,\theta_1) > 0\}$

Nous savons que $E_\theta(S) = h(\theta)$, alors il vient :

a) $\mathcal{P}$ *permet un estimateur efficace de* h(θ) si :

$$M_n(\theta,h) = \frac{\left(\frac{d}{d\theta}\,h(\theta)\right)^2}{n.F_X(\theta)} \qquad \forall\ \theta \in \Theta$$

et S est *un estimateur efficace de* h(θ) si :

$$\mathrm{Var}_\theta S = M_n(\theta,h) = \frac{\left(\frac{d}{d\theta}\,h(\theta)\right)^2}{n.F_X(\theta)} \qquad \forall\ \theta \in \Theta$$

b) $\mathcal{P}$ *ne permet pas d'estimateur efficace de* h(θ) si :

$$\exists\theta \in \Theta : \ M_n(\theta,h) > \frac{\left(\frac{d}{d\theta}\,h(\theta)\right)^2}{n.F_X(\theta)}$$

et S est un *estimateur de variance minimum de* h(θ) si :

$$\mathrm{Var}_\theta S = M_n(\theta,h) \qquad \forall\ \theta \in \Theta$$

Dans un problème concernant l'estimation de h(θ)à partir d'un échantillon de taille n variable, il est normal d'ajouter à la définition 7.3.2 le fait que les relations doivent être vraies quelque soit n.
D'autre part le cas échéant, on peut généraliser ces définitions en ajoutant le qualificatif asymptotique quand une relation de 7.3.2 est vraie seulement pour $n \to \infty$.

Notons que sous les hypothèses de 7.3.1 et 7.3.2, s'il existe un estimateur efficace alors $F_X(\theta) = F_T(\theta)$ ou

$$F_n(\theta) = n.F_X(\theta) = F_S(\theta)$$

et s'il existe un estimateur de variance minimum (éventuellement efficace) cet estimateur est unique [2].

7.3.3 Remarque.

Dans le cas où $\mathcal{P}$ n'est pas homogène, il ne semble pas exister de conditions simples et générales pour que $\mathrm{Var}_\theta T$ soit minorée par une quantité indépendante de T. Il n'est donc pas possible, par cette méthode, de définir une notion suffisamment générale et satisfaisante d'efficacité ou d'estimateur de variance minimum.

Mais il est possible, dans certains cas où $\mathcal{P}$ n'est pas homogène, de montrer que les quantités de la forme (7.2.2) :

$$\limsup_{\theta_1 \to \theta_0} \left(\frac{E_{01}(T) - E_{10}(T)}{\theta_0 - \theta_1} \right)^2$$

sont indépendantes de T appartenant à une certaine classe d'estimateurs, les inégalités de 7.2.2 et 7.2.3 permettent alors de définir un critère d'efficacité pour les estimateurs de cette classe.

§ 8 EXEMPLES

Dans ce paragraphe nous traitons succintement un exemple dans lequel nous mettons en oeuvre les techniques définies dans les chapitres précédents. La loi choisie ne vérifie pas les conditions de régularité de 6.5 et en particulier l'application de la formule classique 6.5.1 conduit à des absurdités. Par contre cette loi vérifie les conditions A et B de 2.7 et il est possible d'obtenir la quantité d'information de Fisher généralisée et l'inégalité de Cramer - Rao.

8.1 Etude préliminaire.

Soit X une variable aléatoire réelle dont la densité de la loi par rapport à la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{R})$ est :

$$f(x,\theta) = \frac{\theta}{e^{\theta^2}-1} \quad e^{\theta x}\, 1_{[0,\theta]}(x) \qquad \text{avec} \quad \theta \in \,]0,+\infty[$$

Nous avons :

$$F_\theta(x) = P_\theta(X<x) = \frac{e^{\theta x}-1}{e^{\theta^2}-1}$$

$$E_\theta(X) = \frac{\theta\, e^{\theta^2}}{e^{\theta^2}-1} - \frac{1}{\theta} \qquad\qquad \mathrm{Var}_\theta(X) = \frac{1}{\theta^2} - \frac{\theta^2\, e^{\theta^2}}{\left(e^{\theta^2}-1\right)^2}$$

Si $x = (x_1,\ldots,x_n)$ est une réalisation du n-échantillon $(X_1,\ldots,X_n)$ de X, la vraisemblance de θ est :

$$L(x,\theta) = \prod_{i=1}^{n} f(x_i,\theta) = \left(\frac{\theta}{e^{\theta^2}-1}\right)^n e^{\theta(x_1+\dots+x_n)} \qquad \text{si } x \in [0,\theta]^n$$

$$L(x,\theta) = 0 \qquad \text{ailleurs.}$$

Enfin on montre facilement que la famille de probabilités considérée, vérifie les hypothèses A-0, A-1, A-2 donc l'hypothèse B de 2.7.

8.2 Quantité d'information de Fisher au sens classique.

Pour $x_i \in]0,\theta[$, $i = 1,\dots,n$, nous avons :

$$\frac{\partial}{\partial\theta} \operatorname{Log} f(x_i,\theta) = x_i - E_\theta(X) - \frac{\theta\, e^{\theta^2}}{e^{\theta^2}-1}$$

$$\frac{\partial}{\partial\theta} \operatorname{Log} L(x,\theta) = \sum_{i=1}^{n} \left(x_i - E_\theta(X)\right) - \frac{n\,\theta\, e^{\theta^2}}{e^{\theta^2}-1}$$

Calculons la quantité d'information au sens classique (6.5.1)

$$\mathcal{F}_X(\theta) = \operatorname{Var}_\theta X + \frac{\theta^2 e^{2\theta^2}}{\left(e^{\theta^2}-1\right)^2}$$

$$\mathcal{F}_{X_1,\dots,X_n}(\theta) = \operatorname{Var}_\theta\left(\sum_{i=1}^{n} X_i\right) + n^2 \frac{\theta^2 e^{2\theta^2}}{\left(e^{\theta^2}-1\right)^2} \qquad \text{donc}$$

$$\mathcal{F}_{X_1,\dots,X_n}(\theta) = n\,\mathcal{F}_X(\theta) + n(n-1) \frac{\theta^2 e^{2\theta^2}}{\left(e^{\theta^2}-1\right)^2}$$

La propriété d'additivité est en défaut car $\frac{\partial}{\partial\theta}$ Log $f(X,\theta)$ n'est pas une variable aléatoire centrée. Cela provient du fait que l'on ne peut intervertir, dans ce cas, le signe de dérivation par rapport à θ avec le signe d'intégration. Pour la même raison on ne peut obtenir l'inégalité de Cramer - Rao par le théorème classique.

8.3 Partie finie de la divergence.

Pour $\theta_1, \theta_2 \in]0,+\infty[$ et $\theta_1 \leq \theta_2$ nous avons

$$f^*_{\theta_1} = f_{\theta_1} \qquad \text{et} \qquad f^*_{\theta_2} = \frac{\theta_2 \, e^{\theta_2 x}}{e^{\theta_1\theta_2} - 1} \qquad \text{car} \quad \chi_{12} = [0,\theta_1]$$

Définissons

$$\alpha(\theta_1,\theta_2) = \int_0^{\theta_1} f_{\theta_1}(x) \,\text{Log}\, \frac{f_{\theta_1}(x)}{f^*_{\theta_2}(x)} \, dx$$

$$= \text{Log}\left(\frac{\theta_1}{\theta_2} \cdot \frac{e^{\theta_1\theta_2} - 1}{e^{\theta_1^2} - 1}\right) + (\theta_1 - \theta_2)\left(\frac{\theta_1 \, e^{\theta_1^2}}{e^{\theta_1^2} - 1} - \frac{1}{\theta_1}\right)$$

$$\beta(\theta_1,\theta_2) = \int_0^{\theta_1} f^*_{\theta_2}(x) \,\text{Log}\, \frac{f^*_{\theta_2}(x)}{f_{\theta_1}(x)} \, dx$$

$$= \text{Log}\left(\frac{\theta_2}{\theta_1} \cdot \frac{e^{\theta_1^2} - 1}{e^{\theta_1\theta_2} - 1}\right) + (\theta_2 - \theta_1)\left[\frac{\theta_1 e^{\theta_1\theta_2}}{e^{\theta_1\theta_2} - 1} - \frac{1}{\theta_2}\right]$$

Il vient $\quad \forall\ (\theta,\theta_0) \in\ ]0,+\infty[^2$

$$\overset{*}{\psi}_X(\theta_0,\theta) = 1_{]0,\theta_0]}(\theta)\ \beta(\theta,\theta_0) + 1_{]\theta_0,+\infty[}(\theta)\ \alpha(\theta_0,\theta)$$

$$\overset{*}{\psi}_X(\theta,\theta_0) = 1_{]0,\theta_0]}(\theta)\ \alpha(\theta,\theta_0) + 1_{]\theta_0,+\infty[}(\theta)\ \beta(\theta_0,\theta)$$

$$\varphi_X(\theta,\theta_0) = 1_{]0,\theta_0]}(\theta)\ (\alpha(\theta,\theta_0) + \beta(\theta,\theta_0))$$

$$+ 1_{]\theta_0,+\infty[}(\theta)\ (\alpha(\theta_0,\theta) + \beta(\theta_0,\theta))$$

Une autre écriture est possible en prenant les indicatrices : $1_{]0,\theta[}(\theta_0)$ et $1_{]\theta,+\infty[}(\theta_0)$. Dans le paragraphe suivant nous allons supposer θ_0 fixé et faire tendre θ vers θ_0, l'écriture choisie est donc plus explicite. A partir de $\overset{*}{\psi}$ et de φ_X, le calcul de ψ_X, I_X et J_X est immédiat mais ne présente pas d'intérêt dans cet exemple.

8.4 Quantité d'information de Fisher généralisée.

Un calcul direct montre que pour $\theta_0 \in\ ]0,+\infty[$ fixé :

$$\frac{d^2}{d\theta^2}\ \alpha(\theta,\theta_0)\Big|_{\theta=\theta_0} = \frac{d^2}{d\theta^2}\ \alpha(\theta_0,\theta)\Big|_{\theta=\theta_0} = \frac{d^2}{d\theta^2}\ \beta(\theta_0,\theta)\Big|_{\theta=\theta_0}$$

$$= \frac{d^2}{d\theta^2}\ \beta(\theta,\theta_0)\Big|_{\theta=\theta_0} = \frac{1}{\theta_0^2} - \frac{\theta_0^2\ e^{\theta_0^2}}{\left(e^{\theta_0^2}-1\right)^2}$$

et donc que :

$$F_X(\theta_0) = \frac{1}{2}\frac{d^2}{d\theta^2}\varphi(\theta,\theta_0)\Big|_{\theta=\theta_0} = \frac{d^2}{d\theta^2}\psi^*(\theta,\theta_0)\Big|_{\theta=\theta_0} = \frac{d^2}{d\theta^2}\psi^*(\theta_0,\theta)\Big|_{\theta=\theta_0}$$

$$F_X(\theta_0) = \frac{1}{\theta_0^2} - \frac{\theta_0^2\, e^{\theta_0^2}}{\left(e^{\theta_0^2}-1\right)^2}$$

Ce calcul direct est compliqué, mais en remarquant que :

$$\alpha(\theta_1,\theta_2) + \beta(\theta_1,\theta_2) = \frac{(\theta_1-\theta_2)^2}{\theta_1\theta_2} + \frac{\theta_1(\theta_1-\theta_2)\, e^{\theta_1^2}\left(e^{\theta_1(\theta_2-\theta_1)}-1\right)}{(e^{\theta_1^2}-1)\ (e^{\theta_1\theta_2}-1)}$$

et que

$$\lim_{\theta_1\to\theta_2=\theta_0}\frac{\alpha(\theta_1,\theta_2)+\beta(\theta_1,\theta_2)}{(\theta_1-\theta_2)^2} = \lim_{\theta_2\to\theta_1=\theta_0}\frac{\alpha(\theta_1,\theta_2)+\beta(\theta_1,\theta_2)}{(\theta_1-\theta_2)^2}$$

$$= \frac{1}{\theta_0^2} - \frac{\theta_0^2\, e^{\theta_0^2}}{\left(e^{\theta_0^2}-1\right)^2}$$

La proposition 6.4.1 et la formule de Taylor par une technique analogue à celle de 6.2.2 nous prouvent directement les résultats désirés.

8.5 Inégalité de Cramer - Rao généralisée.

Soit la statistique $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ du n-échantillon $(X_1,\dots,X_n)$ de X,

$\bar{X}$ est un estimateur sans biais de $h(\theta)$:

$$E_\theta(\bar{X}) = E_\theta(X) = h(\theta) = \frac{\theta\, e^{\theta^2}}{e^{\theta^2}-1} - \frac{1}{\theta}$$

Nous avons l'inégalité de Cramer - Rao :

$$\mathrm{Var}_{\theta_0}\bar{X}\,.\;F_{X_1,\ldots,X_n}(\theta_0) \geq \limsup_{\theta_1\to\theta_0}\left(\frac{E_{01}(\bar{X}) - E_{10}(\bar{X})}{(\theta_0-\theta_1)}\right)^2$$

Un calcul simple donne

$$\mathrm{Var}_{\theta_0}\bar{X}\,.\;F_{X_1,\ldots,X_n}(\theta_0) \geq \left(\frac{d}{d\theta}h(\theta)\Big|_{\theta=\theta_0} + \frac{\theta_0^2\, e^{\theta_0^2}}{\left(e^{\theta_0^2}-1\right)^2} - \frac{e^{\theta_0^2}}{e^{\theta_0^2}-1}\right)^2$$

et il vient :

$$\mathrm{Var}_{\theta_0}\bar{X}\,.\;F_{X_1,\ldots,X_n}(\theta_0) \geq \left(F_X(\theta_0)\right)^2$$

comme $\quad F_{X_1,\ldots,X_n}(\theta_0) = n.F_X(\theta_0) \qquad (6.3.4)$

et $\quad \mathrm{Var}_{\theta_0}\bar{X} = \frac{1}{n}\,\mathrm{Var}_{\theta_0}X = \frac{1}{n}\,F_X(\theta_0) \qquad (8.1)$

Nous avons donc :

$$\mathrm{Var}_{\theta_0}\bar{X}\,.\;F_{X_1,\ldots,X_n}(\theta_0) = \left(F_X(\theta_0)\right)^2$$

c'est à dire égalité dans l'inégalité de Cramer - Rao.

On en déduit de plus : $F_{X_1,\ldots,X_n}(\theta_0) = F_{\bar{X}}(\theta_0)$.

BIBLIOGRAPHIE

[1] D.G. CHAPMAN et H. ROBBINS : *"Minimum variance estimation without regularity assumptions"*. Ann. Math. Statist., 1951, 22, pp 281-286.

[2] C. FOURGEAUD et A. FUCHS : *"Statistique"*. Dunod - 1967.

[3] D.A.S. FRASER : *"On information in statistics"*. Ann. Math. Statist., 1965, 36, pp 890-896.

[4] S. KULLBACK : *"Information theory and statistics"*. J. WILEY, 1959.

[5] S. KULLBACK : *"A lower bound for discrimination information in terms of variation"*. Information theory, 1967, 13, pp 126-127.

[6] S. KULLBACK et M.A. KHAIRAT : *"A note on minimum discrimination information"*. Ann. Math. Statist., 1966, 37, pp 279-280.

[7] J. NEVEU : *"Bases mathematiques du calcul des probabilités"*. MASSON, 1964.

[8] C.R. RAO : *"Linear statistical inference and its applications"*. J. WILEY, 1965.