N° d'ordre : 1356

THÈSE

présentée à

L'UNIVERSITÉ DES SCIENCES ET TECHNIQUES DE LILLE FLANDRES ARTOIS

pour obtenir

LE GRADE DE DOCTEUR DE 3e CYCLE
SPÉCIALITÉ : MATHÉMATIQUES APPLIQUÉES

par

DE ALMEIDA Rui Manuel

DECANTATION DANS LES CHAINES DE MARKOV

Membres du Jury : J. GEFFROY, Président (Université de Paris VI)
A. HILLION, Rapporteur (E.N.S.T. de Bretagne)
R. MOCHÉ, Rapporteur et Directeur de Recherche
D. BOSQ
C. LANGRAND } Examinateurs

Soutenue le 15 octobre 1986

À minha Mãe.

Monsieur Jean GEFFROY a accepté de présider le jury de cette thèse, me faisant ainsi un grand honneur. Par ailleurs, c'est lui qui a introduit les méthodes de décantation en statistique, qui sont naturellement fondamentales pour ce travail.

Monsieur Raymond MOCHÉ m'a initié à cette technique et m'a proposé de l'appliquer aux chaînes de Markov. Il n'est pas exagéré de dire que sans son aide et sa compétence cette thèse n'aurait pas été accomplie.

Monsieur Alain HILLION, lui aussi spécialiste de la décantation, a accepté d'être un des rapporteurs de cette thèse, lui donnant ainsi un aval précieux.

Monsieur Denis BOSQ apporte sa caution de spécialiste renommé d'estimation fonctionnelle ; je n'oublie pas non plus qu'il m'a permis d'entreprendre des recherches à Lille, en tant que directeur du Laboratoire de Statistique et Probabilités.

Monsieur Claude LANGRAND cautionne lui aussi mon travail : avant de se consacrer à l'analyse des données, il a lui-même travaillé sur les processus de Markov. Son exemple me permet de penser que je pourrai peut-être m'orienter vers un travail plus appliqué à la Faculdade de Economia da Universidade de Coimbra.

Madame Arlette LENGAIGNE a soigneusement dactylographié mon manuscrit, en faisant preuve à mon égard de beaucoup de patience et de gentillesse.

Les membres du Conseil Scientifique et du Conseil Directif de la F.E.U.C. m'ont toujours soutenu et encouragé à poursuivre ce travail.

La Fundação Calouste Gulbenkian (Lisbonne), en m'accordant une bourse pendant trois ans, m'a donné les moyens matériels nécessaires à mon séjour en France.

De nombreux amis et parents m'ont soutenu moralement pendant ces recherches, ce qui s'est révélé essentiel.

A tous, je tiens à exprimer ici ma très vive reconnaissance.

TABLE DES MATIERES

CHAPITRE I

INTRODUCTION.

I.1. - DESCRIPTION DU MODELE.-

On va considérer une suite $X = (X_n \; ; \; n \geq 1)$ d'observations, X_n à valeurs dans un espace probabilisable $(\mathcal{X}_n, \mathcal{B}_n)$ quelconque $(n \geq 1)$.

Les notations suivantes seront utilisées :

- $\mathcal{X}^{(n)} = \prod_{k=1}^{n} \mathcal{X}_k \quad , \quad \mathcal{X}^{(\infty)} = \prod_{k \geq 1} \mathcal{X}_k \; ;$
- $\mathcal{B}^{(n)} = \bigotimes_{k=1}^{n} \mathcal{B}_k \quad , \quad \mathcal{B}^{(\infty)} = \bigotimes_{k \geq 1} \mathcal{B}_k \; ;$
- $\bar{\mathcal{B}}^{(n)} = \mathcal{B}^{(n)} \times \prod_{j>n} \mathcal{X}_j$.

$(\bar{\mathcal{B}}^{(n)} \; ; \; n \geq 1)$ est une suite croissante de sous-tribus de $\mathcal{B}^{(\infty)} = \sigma(\bigcup_{n \geq 1} \bar{\mathcal{B}}^{(n)})$.

On admettra que les X_n sont des variables aléatoires définies sur un espace probabilisable $(\Omega, \mathcal{A})$ et l'on choisira pour X le modèle statistique $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, \mathcal{H})$, où $\mathcal{H}$ est un ensemble de lois de probabilité sur $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)})$.

Un sous-ensemble de $\mathcal{H}$ est appelé une hypothèse ; $\mathcal{H}$ est dite hypothèse générale.

Si (E,d) est un espace métrique et $f : \mathcal{H} \to E$ est un paramètre fonctionnel, on essayera d'obtenir des conditions qui assurent l'estimabilité de f, c'est-à-dire l'existence d'une suite d'estimateurs de f qui converge vers f dans un sens à préciser.

On se reportera quelquefois à un modèle $(X^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta, \theta \in \Theta))$ et à l'estimabilité du paramètre θ. Le sens d'une telle expression doit être éclairci.

$(P_\theta, \theta \in \Theta)$ est un ensemble de lois sur $(X^{(\infty)}, \mathcal{B}^{(\infty)})$. Comme il est d'usage, chaque élément d'un tel ensemble n'est repéré qu'une fois : $\theta_1, \theta_2 \in \Theta, \theta_1 \neq \theta_2 \Longrightarrow P_{\theta_1} \neq P_{\theta_2}$.

Si un paramètre $f : (P_\theta, \theta \in \Theta) \to (E,d)$ est injectif, on peut considérer sur Θ la métrique $d_1 : (\theta_1,\theta_2) \mapsto d(f(P_{\theta_1}), f(P_{\theta_2}))$. L'estimabilité de θ se traduit ainsi par l'estimabilité de tout paramètre injectivement relié à la loi. Par exemple, si $\Theta = \mathbb{R}$, $P_\theta = N(\theta,1)$, dire que θ est estimable est une façon abrégée de désigner l'estimabilité de $f : N(\theta,1) \mapsto \theta$.

La notation $(X^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta, \theta \in \Theta))$ s'avère commode lorsque l'on veut désigner la loi de X_n associée à θ, qui sera notée $P_{\theta,n}$, ou celle du vecteur $X^{(n)} = (X_1,\dots,X_n)$, qui sera notée $P_\theta^{(n)}$.

On se réfèrera souvent à la loi sur $\mathcal{B}_n$ associée à θ, conditionnée par $X^{(n-1)} = x^{(n-1)} = (x_1,\dots,x_{n-1}) \in X^{(n-1)}$. Une telle loi, en admettant qu'elle existe, sera notée $P_{\theta,X_n}^{X^{(n-1)}=x^{(n-1)}}$.

I.2. - <u>DECANTATION ET SEPARATION ASYMPTOTIQUE UNIFORME DE FAMILLES DE LOIS</u>.-

Soit $(X,\mathcal{B})$ un espace probabilisable. Une application mesurable $\phi : (X,\mathcal{B}) \longrightarrow ([0,1], \mathcal{B}_{[0,1]})$ est appelée un test.

<u>Définition I.1</u>.- Soit $(X,\mathcal{B})$ un espace probabilisable, H_0 et H_1 deux familles de probabilités sur $(X,\mathcal{B})$. On dit qu'elles sont décantées s'il

existe un test φ défini sur X tel que

$$\omega_o = \sup(\int \varphi \, dP \ ; \ P \in H_o) < \inf(\int \varphi \, dP \ ; \ P \in H_1) = \omega_1 \ . \tag{1}$$

$\omega_1 - \omega_o$ sera quelquefois appelé seuil de décantation associé au test φ.

Remarque I.1.- Puisque $1-\varphi$ est aussi un test, la notion de décantation est symétrique par rapport aux familles H_o et H_1.

Remarque I.2.- Si φ est l'indicatrice d'un événement dans B, on parle de décantation par des événements. Mais cette notion est strictement plus forte.

Exemple I.1.- Soit $X = \{1,2,3\}$, $B = P(X)$, H_o = {loi uniforme P sur X}, H_1 = {lois Q sur B tels que $2Q(\ 1\ ) + Q(\ 2\ ) \leq 0{,}8$}. On voit facilement que ces deux familles ne peuvent pas être décantées par aucun événement. Pourtant, si $\varphi(1) = 1$, $\varphi(2) = \frac{2}{3}$, $\varphi(3) = \frac{1}{3}$

$$\sup(\int \varphi \, dQ \ ; \ Q \in H_1) = \frac{3}{5} < \frac{2}{3} = \int \varphi \, dP \ .$$

Etant donné le modèle $(X^{(\infty)}, B^{(\infty)}, (P_\theta \ , \ \theta \in \Theta))$ une suite $(C_n \ ; \ n \geq 1)$ est dite $B^{(n)}$-adaptée si $C_n \in B^{(n)} (n \geq 1)$.

Définition I.2. - Soient $(P_\theta, \ \theta \in \Theta_i)$ $(i = 0,1)$ deux familles de lois sur $(X^{(\infty)}, B^{(\infty)})$. On dit qu'elles se séparent asymptotiquement (resp. asymptotiquement uniformément) s'il existe une suite $(C_n \ ; \ n \geq 1)$ $B^{(n)}$-adaptée telle que :

$$\forall \theta \in \Theta_o \quad \lim_{n\to\infty} P_\theta^{(n)}(C_n) = 0 \quad , \quad \forall \theta \in \Theta_1 \ \lim_{n\to\infty} P_\theta^{(n)}(C_n) = 1 \tag{2}$$

(resp. telle que

$$\lim_{n\to\infty} \sup_{\theta \in \Theta_o} P_\theta^{(n)}(C_n) = 0 \quad , \quad \lim_{n\to\infty} \inf_{\theta \in \Theta_1} P_\theta^{(n)}(C_n) = 1). \tag{3}$$

Remarque I.3.- Pour vérifier que ces notions sont aussi symétriques par rapport aux familles de lois, on n'a qu'à remplacer C_n par $X^{(n)}-C_n$.

Remarque I.4.- Compte tenu de la remarque I.2, on pourrait envisager une définition de séparation asymptotique faisant intervenir une suite $(\varphi_n \; ; \; n \geq 1)$ de tests définis sur $X^{(n)}$ au lieu des événements C_n , pour laquelle on ait :

$$\forall\theta \in \Theta_0 \quad \lim_{n\to\infty} \int \varphi_n \, dP_\theta^{(n)} = 0 \; , \quad \forall\theta \in \Theta_1 \quad \lim_{n\to\infty} \int \varphi_n \, dP_\theta^{(n)} = 1 \; . \tag{4}$$

Pourtant, cette condition équivaut à (2) : si $(\varphi_n \; ; \; n \geq 1)$ vérifie (4), et si l'on pose $C_n = \mathbb{1}_{\{\phi_n > \frac{1}{3}\}}$, alors $(C_n \; ; \; n \geq 1)$ vérifie (2).

Les mêmes considérations sont extensibles à la séparation asymptotique uniforme.

Si H_0 et H_1 se séparent asymptotiquement, $\forall(P,Q) \in H_0 \times H_1$ P est orthogonale à Q , ce qui se notera $P \perp Q$.

Les notions de décantation et de séparation asymptotique uniforme sont reliées par le fait suivant. Lorsqu'il existe des versions régulières de $P_{\theta,X_n}^{X^{(n-1)}=x^{(n-1)}}$ $(\theta \in \Theta_0 \cup \Theta_1 \; , \; n \geq 2 \; , \; x^{(n-1)} \in X^{(n-1)})$, si $\forall n \geq 2$, $\forall x^{(n-1)} \in X^{(n-1)}$ $(P_{\theta,X_n}^{X^{(n-1)}=x^{(n-1)}} \; ; \; \theta \in \Theta_i)$ $(i = 0,1)$ sont décantées et si les seuils de décantation n-dimensionnels satisfont certaines contraintes, on en déduira la séparation asymptotique uniforme de $(P_\theta \; ; \; \theta \in \Theta_i)$ $(i = 0,1)$.

I. 3. - SEPARATION ASYMPTOTIQUE UNIFORME ET ESTIMATION CONVERGENTE.-

Considérons un modèle statistique $(X^{(\infty)}, \mathcal{B}^{(\infty)}, H)$, un espace métrique (E,d) et un paramètre $f : H \to E$. Soit E muni de sa tribu borélienne.

Pour $n \geq 1$ soit $\hat{f}_n : X^{(n)} \to E$ mesurable. $(\hat{f}_n\ ;\ n \geq 1)$ est appelée une suite d'estimateurs de f. Une telle suite est dite :

- convergente si

$$\forall \varepsilon > 0 \quad \forall P \in H \quad \lim_{n\to\infty} P^{(n)} \{d\ (\hat{f}_n, f(P)) > \varepsilon\} = 0 \quad ; \tag{5}$$

- presque-sûrement convergente si

$$\forall \varepsilon > 0 \quad \forall P \in H \quad \lim_{n\to\infty} P\{\bigcup_{k\geq n} [\{d(\hat{f}_k\ , f(P)) > \varepsilon\} \times \prod_{j>k} X_j]\} = 0 \ ; \tag{6}$$

- uniformément convergente si

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} \sup_{P\in H} \quad P^{(n)} \{d(\hat{f}_n, f(P)) > \varepsilon\} = 0 \quad ; \tag{7}$$

- uniformément presque-sûrement convergente si ,

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} \sup_{P\in H} P[\bigcup_{k\geq n} (\{d(\hat{f}_k, f(P)) > \varepsilon\} \times \prod_{j>k} X_j)] = 0 \quad . \tag{8}$$

Définition I.3.- Soit $(X^{(\infty)}, \mathcal{B}^{(\infty)}, H)$ un modèle statistique, (E,d) un espace métrique, $f : H \to E$ un paramètre. f est dit estimable (resp. p.-s. estimable, resp. uniformément estimable, resp. uniformément p.-s. estimable) s'il existe une suite $(\hat{f}_n\ ;\ n \geq 1)$ d'estimateurs de f vérifiant (5) (resp. (6), resp. (7), resp. (8)).

Remarque I.5.- On verra dans le lemme II.1 que s'il existe une suite d'estimateurs de f vérifiant (7), il en existe une autre vérifiant (8) et réciproquement, c'est-à-dire f est uniformément estimable $\Longleftrightarrow$ f est uniformément p.-s. estimable.

Sous des hypothèses topologiques sur (E,d) et des hypothèses de séparation asymptotique uniforme de certaines familles de lois, on peut démontrer que certains paramètres sont estimables. La démonstration de l'estimabilité est constructive : on parvient à exhiber une suite convergente d'estimateurs du paramètre en question. Les résultats les plus importants à ce propos feront l'objet du chapitre II.

I.4. - ORTHOGONALITE DE DEUX HYPOTHESES.-

Soit $(\mathcal{X},\mathcal{B})$ un espace probabilisable, P et Q deux probabilités sur cet espace. L'expression

$$d_V(P,Q) = \sup\,(P(A) - Q(A)\ ;\ A \in \mathcal{B}) \tag{9}$$

définit la distance en variation d_V entre P et Q. On peut toujours trouver un événement B tel que $P(B) - Q(B) = d_V(P,Q)$. En particulier, si $d_V(P,Q) = 1$, P et Q sont orthogonales : $\exists A \in \mathcal{B}$ tel que $P(A^c) = Q(A) = 0$.

Si P et Q sont des probabilités sur un espace produit infini $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)})$, $d_V(P,Q)$ est la limite lorsque $n \to \infty$ de la distance en variation des projections de P et Q sur $\mathcal{B}^{(n)}$, que l'on notera $P^{(n)}$ et $Q^{(n)}$ respectivement :

$$d_V(P,Q) = \lim_{n\to\infty} d_V(P^{(n)}, Q^{(n)})\ . \tag{10}$$

Si $P^{(n)} = \underbrace{P \otimes \ldots \otimes P}_{n \text{ facteurs}}$ et $Q^{(n)} = \underbrace{Q \otimes \ldots \otimes Q}_{n \text{ facteurs}}$ (cas d'observations indépendantes identiquement distribuées), et si $P \neq Q$, on vérifie que $d_V(P^{(n)}, Q^{(n)}) \geq 1 - e^{-n/2\, d_V^2(P,Q)} \xrightarrow[n\to\infty]{} 1$. D'après (9), $P \perp Q$ et il est évident que $\{P\}$ et $\{Q\}$ se séparent asymptotiquement.

Naturellement, la vérification de l'orthogonalité est plus difficile si l'on a affaire à des observations qui, bien qu'indépendantes, ne sont pas identiquement distribuées, et pire encore si les observations ne sont même pas indépendantes.

Pour ce qui est de la vérification de l'orthogonalité de deux hypothèses simples, deux résultats doivent être soulignés :

a) Pour le cas d'observations indépendantes, avec $H_o = \{P\}$, $H_1 = \{Q\}$, $P = \bigotimes_{i\geq 1} P_i$, $Q = \bigotimes_{i\geq 1} Q_i$, S. Kakutani ([34]) a établi que, si $P_i \sim Q_i$ $(i \geq 1)$, P et Q sont ou bien orthogonales ou bien équivalentes.

b) Si les observations ne sont pas indépendantes mais les probabilités P et Q sont déterminées par les lois initiales P_1 , Q_1 et les lois conditionnelles $P_{X_n}^{X^{(n-1)}=x^{(n-1)}}$, $Q_{X_n}^{X^{(n-1)}=x^{(n-1)}}$ $(n > 1,\ x^{(n-1)} \in \mathcal{X}^{(n-1)})$, l'orthogonalité de P et Q peut se déduire d'un critère énoncé par J. Kabanov, R. Lipcer et A. Širjaev ([33]), applicable si $P_{X^{(n)}}$ et $Q_{X^{(n)}}$ sont équivalentes $(n \geq 1)$.

Si l'on passe de la considération d'hypothèses simples à celle d'hypothèses multiples, notamment lorsqu'elles sont infinies, le problème de la vérification de leur séparation asymptotique uniforme se complique beaucoup plus. En effet, le fait que $\forall (P,Q) \in \mathcal{H}_o \times \mathcal{H}_1$ $P \perp Q$ n'implique pas la séparation asymptotique uniforme de $\mathcal{H}_o$ et $\mathcal{H}_1$, si l'une de telles hypothèses est infinie. Sauf si $\mathcal{H}_o$ et $\mathcal{H}_1$ sont au plus infinies dénombrables, l'orthogonalité de tout couple de $\mathcal{H}_o \times \mathcal{H}_1$ n'induit même pas la simple séparation asymptotique de $\mathcal{H}_o$ et $\mathcal{H}_1$.

Or le plus souvent en statistique on est confronté avec le problème de vérifier la séparation d'hypothèses infinies non dénombrables. Dans le cas où les éléments P de ces hypothèses sont définis par les lois initiales P_1 et les lois conditionnelles $P_{X_n}^{X^{(n-1)}=x^{(n-1)}}$ $(n > 1,\ x^{(n-1)} \in \mathcal{X}^{(n-1)})$, le problème peut se résoudre en faisant appel à un outil plus fort que les antérieurs, à savoir le théorème fondamental de la décantation (formulé initialement par J. Geffroy ([23]) ; amélioré plus tard par R. Moché ([43])).

Des résultats concernant la séparation de lois seront analysés dans le chapitre III.

I.5. - RAPPORTS ENTRE CERTAINES DISTANCES SUR DES ENSEMBLES DE PROBABILITES.-

On s'est sans doute aperçu qu'il existe un lien étroit entre la séparation asymptotique des lois et la distance en variation. Mais il se peut que, dans certains cas, on ne sache pas évaluer celle-ci, tandis que l'on réussit à calculer d'autres distances entre probabilités.

Le fait qu'une fonction de ces dernières minore la distance en variation s'avère ainsi important pour l'étude de la séparation des lois.

Dans le chapitre IV on exhibera des fonctions qui relient de cette façon la distance en variation à d'autres distances usuellement employées, telles que la distance de Hellinger ou la distance de Prokhorov.

I.6. - SEPARATION ASYMPTOTIQUE DANS LES CHAINES DE MARKOV.-

Dans le chapitre V on considérera le cas où $X = (X_n \ ; n \geq 1)$ est une chaîne de Markov à valeurs dans un espace $\mathcal{X}$ fini ou dénombrable, muni de la tribu de ses parties.

Les résultats concernant le rapport entre la décantation des lois conditionnelles et la séparation asymptotique uniforme de familles de lois y trouvent un domaine d'application privilégié, vu que :

- les problèmes de mesurabilité d'applications définies sur $\mathcal{X}^n$ ne se posent pas ;

- les lois conditionnelles de X_n sachant $X^{(n-1)}$ existent, sont assujetties à moins de contraintes que dans le cas général, et une des versions de ces lois est supposée fixée dès le début de l'analyse.

Les résultats obtenus sont, dans le cas particulier des chaînes homogènes, des conséquences de conditions relativement souples.

On pourra ainsi énoncer des conditions suffisantes pour que des paramètres tels que la matrice de transition et la distribution stationnaire d'une chaîne homogène irréductible récurrente positive soient uniformément p.-s. estimables.

Les estimateurs convergents dont la construction est déduite de la connaissance de suites $(C_n ; n \geq 1)$ qui séparent asymptotiquement uniformément certains couples d'hypothèses ne sont pas mis en oeuvre facilement. Pour éviter ce problème, on énoncera une condition suffisante de convergence de l'estimateur du maximum de vraisemblance.

I.7. - APPORTS DE CE TRAVAIL.-

Pendant l'élaboration de cette thèse l'application des méthodes de décantation à l'étude de la séparabilité de familles de lois et de l'estimabilité convergente de certains paramètres a connu des développements intéressants. Notamment, et sans prétention d'exhaustivité :

a) R. Moché ([41], [43]) a affaibli les conditions du théorème fondamental de la décantation énoncé originalement par J. Geffroy ([23]).

b) W. Pieczynski ([47]) a énoncé une condition suffisante d'estimabilité d'un paramètre lorsque l'espace paramétrique est compact. Cette condition bien qu'équivalente dans ce cas à une autre introduite antérieurement par R. Moché ([39]), est d'application plus facile. Par ailleurs, W. Pieczynski a remarqué que, sous les hypothèses en question, la condition est aussi nécessaire.

Dans le même travail, W. Pieczynski a aussi exhibé une condition suffisante de convergence de l'estimateur du maximum de vraisemblance pour des observations quelconques.

c) R. Moché ([42]) a précisé le rapport entre la condition nécessaire et suffisante d'estimabilité d'un paramètre de W. Pieczynski et celle qu'il avait présentée antérieurement et a par ailleurs étendu quelques résultats de la convergence en probabilité à la convergence presque-sûre.

Sauf pour ce qui concerne la condition suffisante de convergence de l'estimateur du maximum de vraisemblance, que l'on examinera sous les conditions particulières qui nous intéressent dans le chapitre V, on essaiera d'exposer de façon organisée ces résultats dans les chapitres II et III. Naturellement, d'autres résultats seront analysés, de manière à bien encadrer les premiers et permettre de mieux comprendre leur importance.

L'apport de ce travail doit donc être recherché dans les deux derniers chapitres :

i) dans le chapitre IV, suivant une suggestion de R. Moché, on analyse le rapport entre la distance en variation et d'autres distances sur des familles de probabilités sur un même espace.

En particulier, les distances de Hellinger, de Prokhorov, de Kolmogorov, de Lévy, de Forter-Mourier et celle associée à la norme ∞ entre les fonctions caractéristiques y sont contemplées.

Naturellement, l'analyse ne pourrait jamais être exhaustive, mais des majorations de la distance en variation par des fonctions d'autres distances sont exhibées, et ces fonctions se font par ailleurs remarquer par leur simplicité. Quelquefois elles permettent d'obtenir quelques résultats connus de façon presque immédiate. C'est le cas du calcul de la distance de Fortet-Mourier tel qu'il est présenté dans le corollaire 1 de la proposition IV.7, à comparer avec celui de R. Dobrushin ([17]).

ii) Dans le chapitre V , le théorème fondamental de la décantation et les théorèmes d'existence d'estimateurs convergents décrits dans le chapitre II sont appliquées aux chaînes de Markov à espace des états dénombrable . Le travail de J. Geffroy ([24]) a constitué le point de départ de ce qui y est présenté. La notion de décantation de familles de matrices de transition a été reformulée en tenant compte de la version améliorée (R. Moché, [43]) du théorème fondamental de la décantation. La notion de temps de décantation est utilisée, comme dans ([24]), pour conclure que certaines hypothèses se séparent asymptotiquement uniformément, mais on a aussi envisagé les cas où l'espace des états est infini dénombrable, la chaîne est d'ordre $k > 1$ ou irréductible récurrente positive périodique. Par ailleurs, l'application des résultats de R. Moché ([42]) mentionnés à la fin de c) a permis l'obtention de résultats plus forts que ceux de ([24]), notamment parce que l'existence d'estimateurs convergents est remplacée par celle d'estimateurs p.-s. convergents, pour ce qui est de l'estimation de la distribution stationnaire d'une chaîne de Markov homogène irréductible récurrente positive apériodique. De plus, on a aussi établi des conditions pour que la matrice de transition d'une chaîne de Markov homogène soit p.-s. estimable. Des extensions au cas où l'espace des états est infini dénombrable ont aussi été envisagées.

Finalement, en suivant des démarches semblables à celles de W. Pieczynsky ([47]), une condition suffisante de convergence de l'estimateur du maximum de vraisemblance est énoncée. En effet, elle s'avère dans ce cadre plus facile à vérifier que celle de ([47]). Par ailleurs, le fait de travailler avec des chaînes de Markov n'étant pas essentiel à notre résultat (proposition V.13), on peut même dire que, moyennant le renforcement d'une hypothèse, on affaiblit la condition de W. Pieczynski dans un cadre aussi général que le sien.

I.8. - COMPARAISON DES RESULTATS OBTENUS AVEC CEUX QUI SONT USUELLEMENT PRESENTES.-

Dans le domaine où se place cette thèse, on essaie essentiellement d'obtenir, à partir de la décantation de lois conditionnelles et de la séparation asymptotique uniforme qu'elle peut induire, des conditions qui assurent l'existence d'estimateurs convergents de certains paramètres. Donc, le caractère de ce travail est naturellement plus théorique que celui de la plupart des travaux concernant l'inférence statistique sur les chaînes de Markov.

Pour exemplifier, mentionnons la question suivante. Il est connu que l'on peut passer d'une chaîne de Markov d'ordre $k > 1$ à une autre d'ordre 1, en modifiant convenablement l'espace des états. Cela est décrit en détail dans le chapitre V, surtout pour souligner que la décantation de familles de matrices de transition n'est pas affectée par le passage en question. Les résultats statistiques concernant des observations indépendantes sont naturellement plus nombreux que ceux qui s'appliquent aux chaînes de Markov. Une des questions que les praticiens se sont posés d'abord a été celle de tester l'hypothèse d'indépendance des observations dans ("dans" au lieu de "contre" pour suivre la terminologie usuelle, motivée par le fait qu'une suite d'observations indépendantes est aussi une chaîne de Markov) l'hypothèse générale de dépendance markovienne d'ordre 1, puis celle de tester la dépendance markovienne d'ordre $k \in \mathbb{N}^*$ dans celle d'ordre $r > k$ (voir [3], [6], [7], [26], [27], [28], [31]). Cela parce qu'il est clair que l'analyse d'une chaîne de Markov d'ordre $k > 1$ peut se ramener à celle plus simple d'une chaîne d'ordre 1 , mais seulement si k est connu au préalable.

Remarquons que le comportement asymptotique des estimateurs proposés pour certains paramètres est analysé selon l'une de deux voies :

- en considérant T échantillons de taille fixe N et étudiant leur comportement lorsque $T \to \infty$ (par exemple, [1], sauf dans la brève section 5) ;

- en considérant un seul échantillon de taille N et étudiant leur comportement lorsque $N \to \infty$.

Naturellement, seule la seconde de ces voies est comparable au cadre où nous nous sommes placés.

On essaiera de décrire le procédé d'analyse habituel dans un cas simple, pour le comparer avec les résultats obtenus dans ce travail, plus particulièrement avec les propositions V.5 et V.6. On considère un échantillon $(x_1,\dots,x_{n+1})$ d'une chaîne de Markov homogène d'ordre 1, à valeurs dans X fini. Pour alléger l'écriture posons $X = \{1,\dots,\nu\}$. Le problème est celui d'estimer les probabilités de transition P_{xy} $(x, y = 1,\dots, \nu)$.

C. Derman ([16]) et L. Goodman ([27]) considèrent aussi le cas où X est infini dénombrable, mais pour des problèmes de test. Naturellement, il n'est pas question d'estimer les P_{xy} $(x, y \geq 1)$ à partir de l'échantillon. Ils proposent de choisir au hasard b valeurs parmi les P_{xy} et, à partir de leurs valeurs estimées, de tester si elles coïncident avec les valeurs correspondantes de l'hypothèse nulle $H_o : P_{xy} = P^o_{xy}$ $(x, y \geq 1)$. On voit que, si la concordance des b valeurs sélectionnées P^o_{xy} avec leurs estimatives n'est pas très conclusive, par contre leur discordance implique le rejet de H_o : ici, comme d'habitude dans les problèmes de test, le rejet de l'hypothèse nulle est plus important que son acceptation. Remarquons que le passage d'un espace des états X infini à un autre X fini, par groupement en classes, peut détruire le caractère markovien du processus (voir [7] ; pourtant, dans [27], un tel procédé est suggéré).

Reprenons le problème d'estimation des P_{xy} $(x, y \in \{1,\dots,\nu\})$ à partir d'un échantillon $(x_1,\dots,x_{n+1})$ d'une chaîne de Markov d'ordre 1. Dans la plupart des travaux sur ce sujet, la chaîne est supposée irréductible récurrente positive (voir [3], [7], [8], [16], [26] , [27], [31]).

Soit n_{xy} $(x, y = 1,\dots,\nu)$ le nombre (non excédant n) de fois auxquelles à l'état x s'est succédé l'état y. La matrice $F = (n_{xy})_{x,y=1,\dots,\nu}$ est appelée matrice de comptage de transitions.

Soit $n_{x\cdot} = \sum_{y=1}^{\nu} n_{xy}$, $n_{\cdot y} = \sum_{x=1}^{\nu} n_{xy}$ $(x, y = 1,\dots,\nu)$ et admettons pour simplifier que $n_{x\cdot} > 0$, $n_{\cdot x} > 0$ $(1 \leq x \leq \nu)$, en rappelant que les méthodes employées supposent n large. Les P_{xy} $(x,y = 1,\dots,\nu)$ sont estimées par le maximum de vraisemblance sous les contraintes $\sum_{y=1}^{\nu} P_{xy} = 1$ $(1 \leq x \leq \nu)$, soit par $\hat{P}_{xy} = \dfrac{n_{xy}}{n_{x\cdot}}$ $(x, y = 1,\dots,\nu)$.

En utilisant une formule de P. Whittle ([55]) et des résultats connus sur les lois multinomiales et l'analyse de tables de contingence avec des fréquences marginales données, on obtient des tests de l'hypothèse d'indépendance dans celle de dépendance markovienne d'ordre 1 (voir [3], [7]).

Ou alors on suppose que les P_{xy} sont des fonctions d'un paramètre $\theta \in \Theta \subset \mathbb{R}^k$ et on estime aussi θ par le maximum de vraisemblance, en admettant des conditions de régularité (voir [7], [8]). Par exemple, dans ([8]) on admet que :

- Θ est un ouvert de $\mathbb{R}^k$;

- l'ensemble $D = \{(x,y) \in \{1,\ldots,\nu\}^2 : P_{xy}(\theta) > 0\}$ est indépendant de θ ;

- chaque $P_{xy}(\theta)$ admet des dérivées partielles continues par rapport à θ jusqu'à l'ordre 3 ($\theta \in \Theta$) ;

- la matrice d'entrées $\dfrac{\partial P_{xy}(\theta)}{\partial \theta_u}$ $((x,y) \in D\ ;\ 1 \leq u \leq k)$ est de rang $k (\theta \in \Theta)$;

- la chaîne est irréductible récurrente positive.

On établit le système des équations de vraisemblance

$$\frac{\partial}{\partial \theta_u} \sum_{(x,y)\in D} n_{xy} \log P_{xy}(\theta) = 0 \quad (1 \leq u \leq k) \tag{10}$$

et on conclut que, $\theta^o \in \Theta$ étant la vraie valeur du paramètre, il existe une solution $\hat{\theta}_n$ de (10) qui converge vers θ^o lorsque $n \to \infty$.

Par ailleurs, $\left(\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\partial}{\partial \theta_u} \log P_{x_j\, x_{j+1}}(\theta)\right),\ 1 \leq u \leq k\right)$ est asymptotiquement gaussien non dégénéré $N(0, \Sigma(\theta))$, où $\Sigma(\theta) = (\sigma_{uv}(\theta))_{u,v=1,\ldots,k}$ est définie par

$$\sigma_{uv}(\theta) = E_\theta \left\{\frac{\partial}{\partial \theta_u} \log P_{x_1\, x_2}(\theta) \frac{\partial}{\partial \theta_v} \log P_{x_1\, x_2}(\theta)\right\} .$$

On remarque tout de suite la ressemblance avec les propriétés de l'estimateur du maximum de vraisemblance pour des observations indépendantes équidistribuées, sous les hypothèses classiques, par exemple celles énoncées par H. Cramér ([13] ; la méthode du χ^2 minimum pourrait aussi être utilisée, mais ne présenterait pas de grands avantages : voir la référence à ce propos dans [8], p. 27).

En effet, le fait que $P_{xy}(\theta)$ soit de classe C^3 par rapport à θ permet (localement) de dériver sous le signe de l'intégrale et on voit que $\left(\sum_{j=1}^{n} \frac{\partial}{\partial\theta_u} \log P_{x_j\, x_{j+1}}(\theta),\ n \geq 1\right)$ est une martingale dont le terme général possède des moments d'ordre 2.

Les conclusions sont obtenues en tenant compte de l'ergodicité de la chaîne et en appliquant un théorème central limite pour martingales ([9]).

On voit immédiatement que dans les propositions V.5 et V.6 l'irréductibilité de la chaîne n'est pas invoquée, et qu'il n'est pas question non plus d'exiger que les $P_{xy}(\theta)$ soient différentiables, donc que les équations de vraisemblance ne sont pas utilisées.

Comparer les résultats usuels sur l'analyse statistique de chaînes de Markov à ceux auxquels nous arrivons par la théorie de la décantation nous amène à une discussion, qui forcément ne peut être que concise, de l'utilisation des estimateurs du maximum de vraisemblance. Les propriétés de ces estimateurs sont en général obtenues en supposant des conditions de différentiabilité (ici, des $P_{xy}(\theta)$) et considérant les équations de vraisemblance (10).

Pour le cas d'observations indépendantes équidistribuées, les hypothèses classiques telles que celles de H. Cramér ([13]) permettent de conclure qu'une solution $\hat{\theta}_n$ de (10) converge vers la vraie valeur θ^o du paramètre et que $\sqrt{n}(\hat{\theta}_n-\theta^o)$ est asymptotiquement normal centré. $\hat{\theta}_n$ est par ailleurs asymptotiquement efficace.

L'approche de J. Wolfowitz ([57]) concerne aussi des observations indépendantes et impose des conditions de régularité semblables à celles de ([13]). Aussi dans le cas indépendant, L. Le Cam ([36]) envisage surtout la normalité asymptotique de l'estimateur du maximum de vraisemblance en affaiblissant les hypothèses de différentiabilité.

Des extensions des résultats classiques aux cas d'observations dépendantes ont aussi été énoncées.

i) Dans ([52]), A. Wald considère un processus discret où les observations sont éventuellement dépendantes. En imposant des conditions semblables à celles de ([13]), il conclut par la convergence et l'efficacité asymptotique de l'estimateur du maximum de vraisemblance (efficacité au sens large, c'est-à-dire n'exigeant pas que $\sqrt{n}\ (\hat{\theta}_n - \theta^o)$ soit asymptotiquement gaussien).

ii) S. Silvey ([50]) fait essentiellement une discussion des restrictions qui assurent la convergence et la normalité asymptotique de l'estimateur du maximum de vraisemblance, et de la façon selon laquelle elles pourraient être affaiblies. Pour ce qui est de la convergence, qui est plutôt en rapport avec ce que nous avons fait, ses conditions supposent l'existence des moments d'ordre 2 de $\log \frac{L_n(\theta^o)}{L_n(\theta)}$, où $L_n(\theta)$ est la vraisemblance d'un échantillon $(x_1,\ldots,x_n)$ sous la loi associée à θ, et que le quotient entre l'écart-type de $\log \frac{L_n(\theta^o)}{L_n(\theta)}$ et son espérance sous θ^o converge vers 0 uniformément sur un ensemble dépendant de θ^o, bien que d'autres conditions de régularité. Ces mentions n'ont pour but qu'illustrer la complexité du problème.

iii) Y. Bar-Shalom ([4]) analyse le problème de la convergence et de l'efficacité asymptotique de l'estimateur du maximum de vraisemblance pour des observations dépendantes suivant les démarches de H. Cramér ([13]), mais en utilisant une loi des grands nombres pour observations dépendantes de E. Parzen (mentionnée dans [46], p. 419) plus restrictive que celle de Khinchine (voir [25], p. 203) utilisée par H. Cramér. D'autre part, il utilise le théorème central limite pour martingales de ([9]), plus facile à mettre en oeuvre que l'énoncé de Lévy sur lequel ([9]) est basé, tandis que S. Silvey s'est reporté à l'énoncé original de Lévy.

Le travail de Y. Bar-Shalom est aussi à mettre en rapport avec celui de A. Wald ([52]).

iv) B. Bhat ([5]) a étendu les résultats de Y. Bar-Shalom pour obtenir la normalité asymptotique, en utilisant un théorème central limite pour martingales de M. Loève ([37], p. 387). Pour qu'une racine de (10) soit convergente, les logarithmes des densités conditionnelles $p_k(\theta) = P_{\theta,X^{(k)}} / P_{\theta,X^{(k-1)}}$ $(k > 1)$ doivent être dérivables sur un intervalle contenant la vraie valeur du paramètre. Ainsi, ses conditions sont encore lointaines de celles induites par la théorie de la décantation.

v) M. Bad ([2]) souligne une erreur de B. Bhat et utilise un théorème central limite pour martingales de D. Scott ([49]) pour obtenir des résultats pareils à ceux mentionnés dans iv) ; pourtant, son travail a aussi pour but de démontrer l'équivalence de l'estimateur du maximum de vraisemblance et de l'estimateur de Bayes, sous des conditions supplémentaires.

vi) M. Crowder ([14]) présente des résultats semblables aux autres antérieurement cités, utilisant les équations de vraisemblance. Pourtant, il n'exige que l'existence des dérivées de 2ème ordre des logarithmes des densités, moyennant l'utilisation d'une forme du théorème du point fixe de Brower. Pour la normalité asymptotique, il fait appel à un théorème central limite de B. Brown ([11]).

On doit aussi faire mention d'autres travaux qui, n'ayant a priori pour but d'étendre la connaissance des conditions de convergence de l'estimateur du maximum de vraisemblance, ont quand même abouti à des résultats sur cette question. Ayant en vue des problèmes de robustesse et des propriétés des estimateurs dans des modèles mal spécifiés (c'est-à-dire on suppose que les observations suivent une loi P appartenant à un ensemble donné H, tandis que la vraie loi n'est pas un élément de H), on a remarqué qu'une solution du système (10) serait tout de même raisonnable comme estimateur, vu qu'elle correspondrait à l'estimateur intuitivement proposé par le critère du minimum de l'information de Kullback (en notation abrégée KLIC ; voir H. White ([54]).

Or il se peut que parmi les erreurs de spécification du modèle soit incluse l'hypothèse d'indépendance d'observations qui sont en effet dépendantes. Sous certaines conditions de régularité, on peut conclure qu'une des solutions de (10) est convergente. Cela conduit aux estimateurs dits du pseudo-maximum de vraisemblance (voir C. Gouriéroux, A. Monfort, A. Trognon ([29])).

Des hypothèses de différentiabilité sont aussi invoquées, naturellement, d'où qu'elles ne soient pas comparables à celles que nous émettrons. C'est vrai pourtant que ces travaux ont pour objet des recherches plus appliquées et donc que la connaissance des lois limites des estimateurs y occupe une place importante.

Ceci dit, on peut se demander si, par exemple, les hypothèses qui assurent la convergence de l'estimateur du maximum de vraisemblance mentionnées dans la proposition V.13 ne sont pas trop lourdes. On doit plutôt les comparer à celles de A. Wald dans ([53] ; voir aussi [56]), où il n'est question que d'assurer la convergence de l'estimateur du maximum de vraisemblance basé sur des observations indépendantes (A. Wald souligne que des extensions au cas dépendant doivent être possibles ; cette approche, que je sache, n'a pas été développée ultérieurement). On remarque que 8 hypothèses y sont présentées (7, pour des lois discrètes), et qu'elles peuvent aussi être considérées trop restrictives.

Eventuellement, on parviendra en utilisant des conditions de décantation et en admettant des hypothèses convenables, à énoncer des lois des grands nombres, des théorèmes central limite ... Pour le moment, l'auteur avoue ne pas être en position de le faire.

Le bilan de toutes ces considérations est que, si par les méthodes employées ici on peut aboutir à des théorèmes d'existence d'estimateurs convergents là où des exemples d'estimateurs convergents ont déjà été proposés, les hypothèses qui assurent la convergence des estimateurs sont, dans les deux cas, de nature bien diverse, ce qui rend la comparaison des deux approches, la classique et celle de la décantation, bien difficile à accomplir.

CHAPITRE II

CONSTRUCTION D'ESTIMATEURS CONVERGENTS.

II.1. - GENERALITES.-

Dans ce chapitre, on se reportera à un modèle statistique $(X^{(\infty)}, B^{(\infty)}, H)$, à un espace métrique (E,d) et à un paramètre fonctionnel $f : H \to E$.

B(e,r) désignera la boule fermée de E de centre $e \in E$ et rayon $r \geq 0$; la boule ouverte de même centre et rayon sera notée b(e,r).

Sous des conditions portant sur les caractéristiques topologiques de E et la séparation asymptotique uniforme de certaines familles de lois, on exhibera une suite convergente d'estimateurs de f.

On ne présentera que l'énoncé de la plupart des résultats. Les démonstrations seront fournies dans certains cas, pour illustrer le type de construction des estimateurs ou éclaircir le rapport entre certaines conditions.

Commençons par établir un résultat déja annoncé.

Lemme II.1.- Soit $(X^{(\infty)}, B^{(\infty)}, H)$ un modèle statistique, (E,d) un espace métrique, $f : H \to E$ un paramètre. f est uniformément estimable si et seulement s'il est uniformément p.-s. estimable.

Preuve : Il est évident que si f est uniformément p.-s. estimable, il est uniformément estimable.

Réciproquement soit $(\hat{f}_n \; ; \; n \geq 1)$ une suite d'estimateurs de f vérifiant I.(7).

$$\forall r \in \mathbb{N}^* \quad \exists n(r) \in \mathbb{N}^* : n \geq n(r) \Longrightarrow \sup_{P \in \mathcal{H}} P^{(n)} \{d(\hat{f}_n, f(P)) > \frac{1}{r}\} \leq \frac{1}{r^2} \; .$$

La suite $(n(r) \; ; \; r \geq 1)$ peut être choisie strictement croissante. $\forall n \geq n(1)$ notons r_n l'unique entier tel que $n(r_n) \leq n < n(r_n + 1)$.

Soit $(\hat{g}_n \; ; \; n \geq 1)$ définie comme suit :

- si $n < n(1)$ $\hat{g}_n$ est une application mesurable de $\mathcal{X}^{(n)}$ dans E quelconque ;

- si $n \geq n(1)$ $\hat{g}_n(x^{(n)}) = \hat{f}_{n(r_n)}(x^{(n(r_n))})$.

Soit $\varepsilon > 0$. $\exists n(\varepsilon) \in \mathbb{N}^* : n \geq n(\varepsilon) \Longrightarrow \frac{1}{r_n} < \varepsilon$. On a

$$\bigcup_{k \geq n} (\{d(\hat{g}_k, f(P)) > \varepsilon\} \times \prod_{j>k} \mathcal{X}_j) = \bigcup_{p \geq r_n} (\{d(\hat{f}_{n(p)}, f(P)) > \varepsilon\} \times \prod_{j>n(p)} \mathcal{X}_j) \; .$$

Si $n \geq n(\varepsilon)$, cet ensemble est contenu dans

$$\bigcup_{p \geq r_n} (\{d(\hat{f}_{n(p)}, f(P) > \frac{1}{p}\} \times \prod_{j>n(p)} \mathcal{X}_j) \quad .$$

Si $n \geq n(\varepsilon)$, $\forall P \in \mathcal{H}$

$$P(\bigcup_{p \geq r_n} [\{d(\hat{f}_{n(p)}, f(P)) > \frac{1}{p}\} \times \prod_{j>n(p)} \mathcal{X}_j]) \leq \sum_{p=r_n}^{\infty} P[\{d(\hat{f}_{n(p)}, f(P)) > \frac{1}{p}\} \times \prod_{j>n(p)} \mathcal{X}_j] .$$

Vu que $n \to \infty \Longrightarrow r_n \to \infty$, on en déduit

$$\sup_{P \in \mathcal{H}} P[\bigcup_{p \geq r_n} (\{d(\hat{f}_{n(p)}, f(P)) > \varepsilon\} \times \prod_{j>n(p)} \mathcal{X}_j] \leq \sum_{p=r_n}^{\infty} \frac{1}{p^2} \xrightarrow[n\to\infty]{} 0 \; . \blacksquare$$

II.2. - CAS OU (E,d) EST PRECOMPACT OU σ-PRECOMPACT (R. Moché [39], [42] ; voir aussi J. Geffroy et R. Moché [21] et [22]).

Considérons la condition

(S1) $\forall E^* \subset E$, E^* précompact $\forall P_o, P_1 \in f^{-1}(E^*)$ $\forall \alpha_o, \alpha_1 > 0$ tels que $d(f(P_o),f(P_1)) > \alpha_o + \alpha_1$ alors $(f^{-1}(E^* \cap B(f(P_i), \alpha_i)))$ $(i = 0,1)$ se séparent asymptotiquement uniformément. (1)

Proposition II.1.- Soit $(X^{(\infty)}, \mathcal{B}^{(\infty)}, \mathcal{H})$ un modèle statistique, (E,d) un espace métrique, $f : \mathcal{H} \to d$ un paramètre.

Supposons (S1) vérifiée. Alors :

a) Si (E,d) est précompact, f est uniformément p.-s. estimable ;

b) Si (E,d) est σ-précompact, f est p.-s. estimable.

Preuve de (a) : $f(\mathcal{H})$ est précompact. Si f est constant, il est trivialement uniformément estimable. Sinon, soit $\varepsilon > 0$, $\varepsilon <$ diam $f(\mathcal{H})$.

$$\exists P_1 ,\ldots, P_r \in \mathcal{H} : f(\mathcal{H}) \subset \bigcup_{i=1}^{r} B(f(P_i), \varepsilon/5) \quad .$$

Pour $i = 1,\ldots,r$ posons :

$$\mathcal{H}_i = f^{-1}(B(f(P_i), \varepsilon/5)) \ ;$$

$$\mathcal{H}'_i = \bigcup_{j:d(f(P_i),f(P_j)) \geq \frac{3\varepsilon}{5}} \mathcal{H}_j \quad .$$

Par hypothèse, ces ensembles ne sont pas vides et, si $i, j \in \{1,\dots,r\}$, $d(f(P_i), f(P_j)) \geq 3\,\varepsilon/5$, H_i et H_j se séparent asymptotiquement uniformément. On peut donc trouver une suite $(C^n_{ij}, n \geq 1)$ $B^{(n)}$-adaptée telle que

$$\inf(P^{(n)}(C^n_{ij})\ ;\ P \in H_i) \xrightarrow[n\to\infty]{} 1\ ,\quad \sup(P^{(n)}(C^n_{ij})\ ;\ P \in H_j) \xrightarrow[n\to\infty]{} 0\ .$$

$\forall i \in \{1,\dots,r\}$ $\forall n \geq 1$ posons $C^n_i = \bigcap_{j:d(f(P_i),f(P_j))\geq 3\varepsilon/5} C^n_{ij}$.

On a :

$$\inf_{H_i} P^{(n)}(C^n_i) \xrightarrow[n\to\infty]{} 1\ ,\quad \sup_{H'_i} P^{(n)}(C^n_i) \xrightarrow[n\to\infty]{} 0\ . \tag{2}$$

Pour $n \geq 1$ définissons un estimateur $\hat{f}_{n,\varepsilon}$ comme suit :

- si $x^{(n)} \notin \bigcup_{i=1}^{r} C^n_i$ ou si $x^{(n)} \in C^n_1$, $\hat{f}_{n,\varepsilon}(x^{(n)}) = f(P_1)$;

- si $x^{(n)} \in \bigcup_{i=1}^{r} C^n_i$ et $i_o = \inf(i\in\{1,\dots,r\} : x^{(n)} \in C^n)$, $\hat{f}_{n,\varepsilon}(x^{(n)})=f(P_{i_o})$.

Soit $P \in H$. $\exists i \in \{1,\dots,r\}$: $P \in H_i$. Cela implique $d(f(P), f(P_i)) \leq \varepsilon/5$.

Vu que

$$\{d(\hat{f}_{n,\varepsilon},f(P)) \geq \varepsilon\} \subset (X^{(n)} - \bigcup_{i=1}^{r} C^n_i) \cup (\bigcup_{j:d(f(P_i),f(P_j))\geq 3\varepsilon/5} C^n_j)\ ,$$

$$\sup_{H} P^{(n)}\{d(\hat{f}_{n,\varepsilon},f(P)) \geq \varepsilon\} \leq \sup_{H} P^{(n)}(X^n - \bigcup_{i=1}^{r} C^n_i) + \sum_{j=1}^{r} \sup_{H'_j} P^{(n)}(C^n_j) \xrightarrow[n\to\infty]{} 0\ ,$$

d'après (2).

Considérons la famille $(\hat{f}_{n,\varepsilon}\ ;\ n \geq 1,\ 0 < \varepsilon < \mathrm{diam}\, f(H))$. Soit $p_o \geq 1$ tel que $\frac{1}{p_o} < \mathrm{diam}\, f(H)$. Donnons à ε successivement les valeurs $\frac{1}{p_o}, \frac{1}{p_o+1}, \dots$ et construisons une suite $(n_i\ ;\ i \geq p_o)$ strictement croissante vérifiant

$$\sup_{H} P^{(n_p)}\{d(\hat{f}_{n_p,1/p},f(P)) \geq \frac{1}{p}\} \leq \frac{1}{p} \quad .$$

Si $n \geq n(p_o)$, désignons par $r(n)$ l'unique entier $r \geq p_o$ tel que

$n_r \leq n < n_{r+1}$. $\forall x^{(n)} \in X^{(n)}$ posons $\hat{f}_n(x^{(n)}) = \hat{f}_{n_{r(n)},\frac{1}{r(n)}}(x_1,\ldots,x_{n_{r(n)}})$.

Soit $\varepsilon > 0$ et p_ε un entier vérifiant $p_\varepsilon \geq p_o$, $\frac{1}{p_\varepsilon} \leq \varepsilon$.

Si $n \geq n_{p_\varepsilon}$, $r(n) \geq p_\varepsilon$ et $\frac{1}{r(n)} \leq \varepsilon$. Alors $\forall \varepsilon > 0$ $\forall n \geq n_{p_\varepsilon}$

$$P^{(n)}\{d(\hat{f}_n,f(P)) \geq \varepsilon\} = P^{(n_{r(n)})}\{d(\hat{f}_{n_{r(n)},\frac{1}{r(n)}},f(P)) \geq \varepsilon\} \leq$$

$$\leq P^{(n_{r(n)})}\{d(\hat{f}_{n_{r(n)},\frac{1}{r(n)}},f(P)) \geq \frac{1}{r(n)}\} \quad .$$

D'où $\sup_{H} P^{(n)}\{d(\hat{f}_n,f(P)) \geq \varepsilon\} \leq \frac{1}{r(n)} \xrightarrow[n\to\infty]{} 0$ et f est donc uniformément p.-s. estimable. ∎

La partie (b) de la proposition II.1. se déduit du lemme suivant dont la démonstration qui ne rapporterai rien est omise (cf. [42]).

Lemme II.2.- Soit $(X^{(\infty)}, \mathcal{B}^{(\infty)}, H)$ un modèle statistique, (E,d) un espace métrique, $f : H \to E$ un paramètre, $(H_n\ ;\ n \geq 1)$ une suite croissante de parties non vides de H dont la réunion est H.

On suppose que $\forall n \geq 1$ f restreint à H_n , noté $f \upharpoonright H_n$, est uniformément estimable pour le modèle statistique $(X^{(\infty)}, \mathcal{B}^{(\infty)}, H_n)$. Alors f est p.-s. estimable.

Considérons la condition

(S2) $\forall E^* \subset E$, E^* précompact $\forall e_o$, $e_1 \in E^*$ $\forall \alpha_o, \alpha_1$ tels que $d(e_o,e_1) > \alpha_o+\alpha_1$ $(f^{-1}(E^* \cap B(e_i,\alpha_i)))$ $(i = 0,1)$ se séparent asymptotiquement uniformément . (3)

Il est évident que (S2) $\Longrightarrow$ (S1), d'où que les conclusions de la proposition II.1. soient valables si (S2) est vérifiée.

II.3. - CAS OU (E,d) *EST COMPLET.-*

Considérons la condition

(S3) $\forall E^* \subset E$, E^* précompact $\forall e_o$, $e_1 \in E^*$, $e_o \neq e_1$ $\exists \alpha > 0$ tel que $(f^{-1}(E^* \cap B(e_i,\alpha)))$ $(i = 0,1)$ se séparent asymptotiquement uniformément. (4)

Lemme II.3. ([42]).- Si (E,d) est complet, (S2) $\Longleftrightarrow$ (S3).

Preuve : Evidemment (S2) $\Longrightarrow$ (S3), même si (E,d) n'est pas complet. Réciproquement, soit (E,d) complet, supposons (S3) vérifiée.

i) Nous allons d'abord montrer que la propriété suivante est satisfaite :

$\forall E^* \subset E$, E^* précompact $\forall \varepsilon > 0$ $\exists \alpha \in]0,\varepsilon[$ $\exists N \in \mathbb{N}^*$ $\exists e_1,\dots,e_N$ points distincts de E^* tels que $E^* \subset \bigcup_{i=1} B(e_i,\alpha)$ et $d(e_i,e_j) \geq \varepsilon \Longrightarrow f^{-1}(E^* \cap B(e_i,\alpha))$ et $f^{-1}(E^* \cap B(e_j,\alpha))$ se séparent asymptotiquement uniformément. (5)

Soit E^* une partie précompacte de E. Si ε n'est pas inférieur au diamètre de E^* , (5) est trivialement vérifiée. Cette possibilité ne sera pas considérée dans la suite.

Si (5) n'est pas vérifiée, $\exists E^* \subset E$, E^* précompact $\exists \varepsilon \in \,]0, \operatorname{diam} E^*[$ tels que $\forall \alpha \in \,]0,\varepsilon[$ $\forall e_1,\ldots,e_{N_\alpha}$ points distincts de E^* tels que $E^* \subset \bigcup_{i=1}^{N_\alpha} B(e_i,\alpha)$ on a

$$\left.\begin{array}{l} d(e_i,e_j) < \varepsilon (i,\ j \in \{1,\ldots,N_\alpha\}) \text{ ou } \exists i,\ j \in \{1,\ldots,N_\alpha\} \\ \text{tels que } d(e_i,e_j) \geq \varepsilon\ ,\ f^{-1}(E^* \cap B(e_i,\alpha)) \text{ et} \\ f^{-1}(E^* \cap B(e_j,\alpha)) \text{ ne se séparent pas asymptotiquement} \\ \text{uniformément} \end{array}\right| \quad (6)$$

Soit $\alpha \in \,]0,\varepsilon[$, $\alpha < \frac{1}{2}\operatorname{diam}(E^*-\varepsilon)$, $E^* \subset \bigcup_{i=1}^{N} B(e_i,\alpha)$, $d(e_i,e_j) < \varepsilon$ $(i,\ j \in \{1,\ldots,N\})$. Alors, $\sup(d(e_i,e_j)\ ;\ i,\ j \in \{1,\ldots,N\}) \leq \varepsilon$ et $\operatorname{diam} \bigcup_{i=1}^{N} B(e_i,\alpha) \leq \varepsilon + 2\alpha < \operatorname{diam} E^*$, ce qui est absurde. Donc la négation de la propriété (5) se traduit par :

$$\left.\begin{array}{l} \exists E^* \subset E,\ E^* \text{ précompact } \exists \varepsilon > 0 \text{ tels que } \forall \alpha \in \,]0,\varepsilon[\\ \exists e'_\alpha\ ,\ e''_\alpha \in E^* \text{ vérifiant simultanément} \\ -\ d(e'_\alpha,e''_\alpha) \geq \varepsilon\ ; \\ -\ f^{-1}(E^* \cap B(e'_\alpha,\alpha)) \text{ et } f^{-1}(E^* \cap B(e''_\alpha,\alpha)) \text{ ne se séparent} \\ \text{pas asymptotiquement uniformément.} \end{array}\right| \quad (7)$$

Donnons à α successivement les valeurs $\frac{\varepsilon}{2}$, $\frac{\varepsilon}{3}$, ..., $\frac{\varepsilon}{n+1}$, ... On en déduit l'existence de deux suites $(e'_n\ ;\ n \geq 1)$ et $(e''_n\ ;\ n \geq 1)$ de E^* telles que :

- $d(e'_n,e''_n) \geq \varepsilon$;

- $f^{-1}(E^* \cap B(e'_n, \frac{\varepsilon}{n+1}))$ et $f^{-1}(E^* \cap B(e''_n, \frac{\varepsilon}{n+1}))$ ne se séparent

pas asymptotiquement uniformément.

E^* étant précompact, de $(e'_n \,;\, n \geq 1)$ et $(e''_n \,;\, n \geq 1)$ respectivement, on peut extraire deux sous-suites de Cauchy $(e'_{n_p} \,;\, p \geq 1)$, $(e''_{n_p} \,;\, p \geq 1)$ convergeant dans l'espace complet (E,d) vers e', e''. Puisque $\forall p \geq 1$ $d(e'_{n_p}, e''_{n_p}) \geq \varepsilon$, on a $d(e',e'') \geq \varepsilon$, $e' \neq e''$.

$E^{**} = E^* \cup \{e',e''\}$ est précompact ; d'après (S3), $\exists \alpha > 0$ tel que $f^{-1}(E^{**} \cap B(e',\alpha))$ et $f^{-1}(E^{**} \cap B(e'',\alpha))$ se séparent asymptotiquement uniformément. Or à partir d'un certain rang $B(e'_{n_p}, \frac{\varepsilon}{n_p+1}) \subset B(e',\alpha)$ et $B(e''_{n_p}, \frac{\varepsilon}{n_p+1}) \subset B(e'',\alpha)$. Donc $f^{-1}(E^* \cap B(e'_{n_p}, \frac{\varepsilon}{n_p+1}))$ et $f^{-1}(E^* \cap B(e''_{n_p}, \frac{\varepsilon}{n_p+1}))$ se sépareraient asymptotiquement uniformément, ce qui contredit (7) et permet de conclure (5).

ii) Montrons maintenant que si (E,d) est complet, (S3) $\Longrightarrow$ (S2).

Soit $E^* \subset E$, E^* précompact, e_o, $e_1 \in E^*$, α_o , $\alpha_1 > 0$ tels que $\varepsilon = \frac{1}{3}\left[d(e_o,e_1) - \alpha_o - \alpha_1\right] > 0$.

Plaçons-nous dans le cas non trivial où $f(H) \cap E^* \cap B(e_i,\alpha_i) \neq \emptyset$ $(i = 0,1)$.

D'après (i), on sait que $\exists \beta \in \left]0,\varepsilon\right[$ $\exists e'_1, \ldots, e'_N$ points de E^* tels que $E^* \subset \bigcup_{i=1}^{N} B(e'_i, \ )$ et $(i, j \in \{1,\ldots,N\}$, $d(e'_i,e'_j) \geq \varepsilon) \Longrightarrow f^{-1}(E^* \cap B(e'_i,\beta))$ et $f^{-1}(E^* \cap B(e'_j,\beta))$ se séparent asymptotiquement uniformément.

Pour $i = 0,1$ faisons

$$I_i = \{j \in \{1,\ldots,N\} : f(H) \cap E^* \cap B(e_i,\alpha_i) \cap B(e'_j,\beta) \neq \emptyset\} .$$

Alors $f^{-1}(E^* \cap B(e_i,\alpha_i)) \subset \bigcup_{j \in I_i} (f^{-1}(E^* \cap B(e'_j,\beta)))$ pour $i = 0,1$.

Comme $j_o \in I_o$, $j_1 \in I_1 \Longrightarrow d(e'_{j_o},e'_{j_1}) \geq \varepsilon$, on en déduit que

$(f^{-1}(E^* \cap B(e_i,\alpha_i)))$ $(i = 0,1)$ se séparent asymptotiquement uniformément. ■

Une suite $(\hat{f}_n \,;\, n \geq 1)$ d'estimateurs d'un paramètre f à valeurs dans (E,d) est dite uniformément convergente sur tout compact de E si

$$\forall K \subset E \;,\; K \text{ compact } \forall \varepsilon > 0 \quad \sup_{P \in f^{-1}(K)} P^{(n)}\{d(\hat{f}_n, f(P)) > \varepsilon\} \xrightarrow[n\to\infty]{} 0 \;. \qquad (8)$$

Proposition II.2. ([47], [42]).- Soit $(X^{(\infty)}, \mathcal{B}^{(\infty)}, \mathcal{H})$ un modèle statistique, (E,d) un espace métrique, $f : \mathcal{H} \to E$ un paramètre.

a) Si (E,d) est complet, (S3) est une condition nécessaire pour que f soit uniformément estimable sur tout compact de E.

b) Si (E,d) est complet et s'il existe une suite croissante $(E_p \,;\, p \geq 1)$ de parties compactes de E telles que $E = \bigcup_{p\geq 1} E_p$ et $\forall K \subset E$, K compact $\exists n \in \mathbb{N}_* : K \subset E_n$, alors (S3) est une condition nécessaire et suffisante pour que f soit uniformément p.-s. estimable sur tout compact de E.

Preuve :

a) f étant uniformément estimable sur tout compact de E , soit $(\hat{f}_n \,;\, n \geq 1)$ une suite d'estimateurs de f telle que $\forall K \subset E$, K compact $\forall \varepsilon > 0$

$$\lim_{n\to\infty} \sup_{P \in f^{-1}(K)} P^{(n)}\{d(\hat{f}_n, f(P)) > \varepsilon\} = 0 \;.$$

Soit $E^* \subset E$, E^* précompact, $e_o, e_1 \in E^*$, $e_o \neq e_1$, $\lambda = d(e_o,e_1)$, $\delta = \frac{1}{4}\lambda$, $H_i = f^{-1}(\bar{E}^* \cap B(e_i,\delta))$ $(i = 0,1)$, où $\bar{E}^*$ désigne l'adhérence de E^* . $\bar{E}^* \cap B(e_i,\delta)$ $(i = 0,1)$ sont des compacts de (E,d).

Plaçons-nous dans le cas non trivial où $H_i \neq \emptyset$ $(i = 0,1)$.

Posons $C_n = \{d(\hat{f}_n, e_o) < \frac{\lambda}{2}\}$ $(n \geq 1)$.

Si $P \in H_o$ et $x \in \{d(\hat{f}_n, f(P)) < \delta\}$, alors
$d(\hat{f}_n(x), e_o) \leq d(\hat{f}_n(x), f(P)) + d(f(P), e_o) < \frac{\lambda}{2}$. Donc $C_n \supset \{d(\hat{f}_n, f(P)) < \delta\}$
et $\inf\limits_{P\in H_o} P^{(n)}(C_n) \xrightarrow[n\to\infty]{} 1$.

Si $P \in H_1$ et $x \in C_n$, alors $d(\hat{f}_n(x), f(P)) \geq d(e_o, f(P)) - d(\hat{f}_n(x), e_o) \geq \delta$.
Ainsi $C_n \subset \{d(\hat{f}_n, f(P)) \geq \delta\}$ et $\sup\limits_{P\in H_1} P^{(n)}(C_n) \xrightarrow[n\to\infty]{} 0$.

b) <u>*Condition nécessaire.*</u> Soit $(\hat{f}_n \; ; \; n \geq 1)$ une suite d'estimateurs de f convergeant uniformément p.-s. sur tout compact de E. Evidemment, cette même suite converge uniformément sur tout compact de E. Vu que (E,d) est supposé complet, on n'a qu'à appliquer (a) pour conclure que (S3) est nécessaire.

<u>*Condition suffisante.*</u> (E,d) étant complet, (S3) $\Longleftrightarrow$ (S2) $\Longrightarrow$ (S1). On peut ainsi appliquer la proposition II.1. : $\forall p \geq 1$ E_p est un compact, d'où il existe une suite uniformément p.-s. convergente d'estimateurs de $f \restriction f^{-1}(E_p)$.

Par hypothèse, si K est un compact de E , $\exists n(K) \in \mathbb{N}^*$: $K \subset E_{n(K)}$ et une suite d'estimateurs de f convergeant uniformément p.-s. sur $E_{n(K)}$ converge a fortiori uniformément p.-s. sur K. f est donc uniformément p.-s. estimable sur tout compact K de E. ■

II.4. - CAS OU (E,d) EST SEPARABLE.-

II.4.1. - L'algèbre des hypothèses testables.-

Définition II.1.- Soit $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, \mathcal{H})$ un modèle statistique, H une hypothèse. H est dite testable (resp. p.-s. testable, resp. uniformément testable) si le paramètre $f : P \longmapsto \mathbb{1}_H(P)$, à valeurs dans $(\mathbb{R}, |.|)$, est estimable (resp. p.-s. estimable, resp. uniformément estimable).

Lemme II.4.- Une hypothèse H du'n modèle statistique $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, \mathcal{H})$ est testable (resp. uniformément testable) si et seulement si H et $\mathcal{H}$ - H se séparent asymptotiquement (resp. asymptotiquement uniformément).

Preuve : Si $(\hat{f}_n ; n \geq 1)$ est une suite convergente (resp. uniformément convergente) d'estimateurs de $f = \mathbb{1}_H$, il suffit de poser $C_n = \{|\hat{f}_n - 1| \leq \frac{1}{4}\}$ $(n \geq 1)$ pour obtenir une suite $\mathcal{B}^{(n)}$-adaptée qui sépare asymptotiquement (resp. asymptotiquement uniformément) H et $\mathcal{H}$-H .

Réciproquement, si $(C_n ; n \geq 1)$ est une suite $\mathcal{B}^{(n)}$-adaptée telle que

$$\forall P \in H \quad P^{(n)}(C_n) \xrightarrow[n\to\infty]{} 1 \quad , \quad \forall P \in \mathcal{H}\text{-H} \quad P^{(n)}(C_n) \xrightarrow[n\to\infty]{} 0 \tag{9}$$

(resp.

$$\liminf_{n\to\infty} \inf_{P\in H} P^{(n)}(C_n) = 1 \quad , \quad \lim_{n\to\infty} \sup_{P\in \mathcal{H}-H} P^{(n)}(C_n) = 0) \quad , \tag{10}$$

faisons $\hat{f}_n = \mathbb{1}_{C_n}$ $(n \geq 1)$ pour conclure . ■

Remarque II.1.-

a) Si H est uniformément testable, elle est a fortiori testable.

b) Les C_n qui figurent dans (9) ou (10) peuvent s'interpréter comme régions d'acceptation de tests de H contre $\mathcal{H}$-H basés sur des échantillons de taille n , ce qui justifie le qualificatif testable de l'hypothèse H.

Lemme II.5.- La classe $\mathcal{H}_t$ des hypothèses testables et la classe $\mathcal{H}_u$ des hypothèses uniformément testables sont des algèbres de parties de $\mathcal{H}$.

Le résultat dérive du fait que ϕ et $\mathcal{H}$ sont uniformément testables et que :

i) $H_o \in \mathcal{H}_t$ (resp. $H_o \in \mathcal{H}_u$) $\Longrightarrow \mathcal{H}-H_o \in \mathcal{H}_t$ (resp. $\mathcal{H}-H_o \in \mathcal{H}_u$) ;

ii) $H_o, H_1 \in \mathcal{H}_t$ (resp. $H_o, H_1 \in \mathcal{H}_u$) $\Longrightarrow H_o \cap H_1$, $H_o \cup H_1 \in \mathcal{H}_t$ (resp. $H_o \cap H_1$, $H_o \cup H_1 \in \mathcal{H}_u$) .

En effet, si $(C_o^n\ ;\ n \geq 1)$, $(C_1^n\ ;\ n \geq 1)$ sont des suites vérifiant les conditions (9) (ou (10)) par rapport à H_o et H_1 , respectivement, on prend $(X^{(n)} - C_o^n\ ;\ n \geq 1)$ pour vérifier (i) et $C_o^n \cap C_1^n$ (resp. $C_o^n \cup C_1^n$) $(n \geq 1)$ pour vérifier (ii) en ce qui concernce $H_o \cap H_1$ (resp. $H_o \cup H_1$).

Remarque II.2.- En général, $\mathcal{H}_t$ n'est pas une tribu, comme le montre l'exemple suivant (dû à J. Hodges, Jr., cité dans [35]). Soit $(X_n\ ;\ n \geq 1)$ une suite d'observations indépendantes équidistribuées, $\mathcal{H} = \{N(\theta,1)^{\otimes(\infty)}\ ;\ \theta \in [0,1]\}$. Alors ; $\forall\theta \in \mathbb{Q} \cap [0,1]$ $\{N(\theta,1)^{\otimes(\infty)}\} \in \mathcal{H}_t$; pourtant, $H = \{N(\theta,1)^{\otimes(\infty)} : \theta \in \mathbb{Q} \cap [0,1]\} \notin \mathcal{H}_t$.

Lemme II.6.- Soit $H \in \mathcal{H}_t$ (resp. $\mathcal{H}_u$) pour le modèle statistique $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, \mathcal{H})$ et $\{H_1,\dots,H_k\}$ une partition de H en éléments de $\mathcal{H}_t$ (resp. de $\mathcal{H}_u$). Soit $(\mathbb{1}_{C_n}\ ;\ n \geq 1)$ une suite convergente (resp. uniformément convergente) de tests pour H. A chaque H_j $(1 \leq j \leq k)$ on peut associer une suite convergente (resp. uniformément convergente) $(\mathbb{1}_{C_{nj}}\ ;\ n \geq 1)$ de tests pour H_j de manière que $\{C_{n1},\dots, C_{nk}\}$ est une partition de C_n $(n \geq 1)$.

On le vérifie en considérant, pour $j \in \{1,\dots,k\}$, une suite $(\mathbb{1}_{C'_{nj}}\ ;\ n \geq 1)$ de tests convergente pour H_j, en posant pour $1 \leq j \leq k$ $C''_{nj} = C'_{nj} \cup (C_n - \bigcup_{j=1}^{k} C'_{nj})$, puis $C_{n1} = C''_{n1} \cap C_n$, $C_{nj} = (C''_{nj} - \bigcup_{\ell=1}^{j-1} C_n) \cap C_n$ $(1 < j \leq k)$. Il ne reste qu'à tenir compte du lemme II.4. et des considérations qui le suivent.

II.4.2. - Condition suffisante d'estimabilité d'un paramètre à valeurs dans un espace métrique séparable.

Définition II.2.- Soit E un ensemble, $\mathcal{B}$ une classe de sous-ensembles de E. Si $\mathcal{Z}_j = \{Z_{j,1},\dots, Z_{j,k(j)}\}$, avec $k(j) \in \mathbb{N}^*$, est une partition de E en éléments de $\mathcal{B}$ et, pour $j \geq 2$, $\forall B \in \mathcal{Z}_{j-1}\ \exists p \in \mathbb{N}_*\ \exists B_1,\dots,B_p \in \mathcal{Z}_j : \bigcup_{i=1}^{p} B_i = B$, alors $(\mathcal{Z}_j\ ;\ j \geq 1)$ est appelée une suite de partitions emboîtées $\mathcal{B}$-mesurables de E.

Remarque II.3.- $(\mathcal{Z}_j\ ;\ j \geq 1)$ étant une suite de partitions emboîtées de E, $\forall j \geq 1\ \forall e \in E\ \exists h(j,e) : e \in Z_{j,h(j,e)}$.

Soit (E,d) un espace métrique, notons $\mathcal{A}(E)$ l'algèbre engendrée par les boules fermées de E.

Lemme II.7. ([48]).- Si (E,d) est séparable, il existe une suite $(\mathcal{Z}_j\ ;\ j \geq 1)$ de partitions emboîtées finies $\mathcal{A}(E)$-mesurables de E telle que, avec les notations précédentes, on ait : $\forall j \geq 1\quad Z_{j,i} \neq 0\quad (1 \leq i \leq k < j)$ et

$$\forall e \in E \qquad \lim_{j\to\infty} \operatorname{diam} Z_{j,h(j,e)} = 0 \ . \tag{11}$$

En se basant sur cette propriété, on peut démontrer la

Proposition II.3. ([48]).- Soit $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, \mathcal{H})$ un modèle statistique, (E,d) un espace métrique séparable, $f : \mathcal{H} \to E$ un paramètre. Si f est $(\mathcal{H}_t - \mathcal{A}(E))$ - mesurable, f est estimable.

Preuve : (E,d) étant séparable, d'après le lemme précédent il existe une suite $(\mathcal{Z}_j\ ,\ j \geq 1)$ de partitions emboîtées finies $\mathcal{A}(E)$-mesurables de E , notées $\mathcal{Z}_j = (Z_{j,1}\ ,\dots,\ Z_{j,k(j)})$, telles que :

- $\forall j \geq 1 \qquad Z_{j,1}\ ,\dots,\ Z_{j,k(j)}$ sont non vides ;

- $\forall e \in E$, si $\forall j \geq 1$ $h(j,e)$ désigne l'unique élément h de $\{1,\dots,k(j)\}$ tel que $e \in Z_{j,h}$,

$$\operatorname{diam}\,(Z_{j,h(j,e)}) \xrightarrow[j\to\infty]{} 0 \ . \tag{12}$$

$\forall j \geq 1$ notons $\mathcal{Z}'_j$ l'image réciproque par f de la partition $\mathcal{Z}_j$:

$$\mathcal{Z}'_j = \{f^{-1}(Z_{1,j})\ ,\dots,\ f^{-1}(Z_{j,k(j)})\} = \{Z'_{j,1}\ ,\dots,\ Z'_{j,k(j)}\} \ .$$

$\mathcal{Z}'_j$ est une partition finie $\mathcal{H}_t$-mesurable de $\mathcal{H}$ dont certains éléments peuvent être vides.

$\mathcal{Z}'_j$ étant une partition finie $\mathcal{H}_t$-mesurable de $\mathcal{H}$, $\forall n \geq 1$ on peut, en utilisant le lemme II.6., trouver une partition $\mathcal{B}^{(n)}$-mesurable de $\mathcal{X}^{(n)}$, $(C_n^{1,1}, \dots, C_n^{1,k(1)})$, telle que :

- $\forall i \in \{1,\dots,k(1)\}$ tel que $Z'_{1,i} = \emptyset$ on a $\forall n \geq 1$ $C_n^{1,i} = \emptyset$;
- $\forall i \in \{1,\dots,k(1)\}$ $\forall P \in Z'_{1,i}$ $P^{(n)}(C_n^{1,i}) \xrightarrow[n\to\infty]{} 1$. (13)

a) Montrons que $\forall n \geq 1$ $\forall j \geq 1$ $\exists (C_n^{j,1}, \dots, C_n^{j,k(j)})$ partition $\mathcal{B}^{(n)}$-mesurable de $\mathcal{X}^{(n)}$ dont certains éléments peuvent être vides telle que :

- $\forall n$, $m \geq 1$ $(C_n^{1,1}, \dots, C_n^{1,k(1)})$, ..., $(C_n^{m,1}, \dots, C_n^{m,k(m)})$ sont des partitions emboîtées $\mathcal{B}^{(n)}$-mesurables de $\mathcal{X}^{(n)}$. (14)

- $\forall m \geq 1$ $\forall j \in \{1,\dots,m\}$ $\forall i \in \{1,\dots,k(j)\}$ $\forall P \in Z'_{j,i}$ $P^{(n)}(C_n^{j,i}) \xrightarrow[n\to\infty]{} 1$. (15)

i) Pour $m = 1$, les partitions $(C_n^{1,1}, \dots, C_n^{1,k(1)})_{n\geq 1}$ introduites ci-dessus vérifient (14) et (15), d'après (13).

ii) Supposons que pour $j = 1,\dots,m$ $(m \geq 1)$ on ait construit des suites $(C_n^{j,1}, \dots, C_n^{j,k(j)})_{n\geq 1}$ de partitions $\mathcal{B}^{(n)}$-mesurables de $\mathcal{X}^{(n)}$ de façon que (14) et (15) soient satisfaites.

Considérons $\mathcal{Z}'_m$ et $\mathcal{Z}'_{m+1}$.

Pour $i \in \{1,\dots,k(m)\}$, $Z'_{m,i}$ est une hypothèse testable qui admet une partition finie $\mathcal{H}_t$-mesurable composée d'éléments de $\mathcal{Z}'_{m+1}$, notés $Z'_{m+1,\alpha(i,1)}, \dots, Z'_{m+1,\alpha(i,\beta(i))}$ (images réciproques par f de la trace sur $Z_{m,i}$ de la partition $\mathcal{Z}_{m+1}$).

Si $Z'_{m,i} \neq \emptyset$, en appliquant le lemme II.6 au modèle statistique $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, Z'_{m,i})$ on peut trouver une suite de partitions finies $\mathcal{B}^{(n)}$-mesurables de $\mathcal{X}^{(n)}$, $(D^1_{n,i}, \ldots, D^{\beta(i)}_{n,i})_{n \geq 1}$ (en prenant $\emptyset$ si l'hypothèse correspondante est vide), telles que :

$$\forall r \in \{1,\ldots,\beta(i)\} \quad \forall P \in Z'_{m+1,\alpha(i,r)} \quad P^{(n)}(D^r_{n,i}) \xrightarrow[n\to\infty]{} 1 \quad . \tag{16}$$

$\forall n \geq 1 \quad \forall r \in \{1,\ldots,\beta(i)\}$ on pose

$$C_n^{m+1,\alpha(i,r)} = C_n^{m,i} \cap D^r_{n,i} \; . \tag{17}$$

Si $Z'_{m,i} = \emptyset$, on sait que $\forall n \geq 1 \quad C_n^{m,i} = \emptyset$ et on pose $C_n^{m+1,\alpha(i,r)} = \emptyset \quad (1 \leq r \leq \beta(i))$.

Dans tous les cas $(C_n^{m+1,\alpha(i,r)} \; ; \; 1 \leq r \leq \beta(i))$ est une partition $\mathcal{B}^{(n)}$-mesurable de $C_n^{m,i}$, et il en résulte que $\{C_n^{m+1,\alpha(i,r)} \; ; \; 1 \leq i \leq k(m), \; 1 \leq r \leq \beta(i)\}$ est une partition $\mathcal{B}^{(n)}$-mesurable de $\mathcal{X}^{(n)}$, que l'on peut noter $(C_n^{m+1,i} \; ; \; 1 \leq i \leq k(m+1))$ car $\{\alpha(i,r) \; ; \; 1 \leq i \leq k(m) \; , \; 1 \leq r \leq \beta(i)\} = \{1,\ldots,k(m+1)\}$.

Par construction, les partitions $(C_n^{j,i} \; ; \; 1 \leq i \leq k(j)) \; (n \geq 1 \; ; \; 1 \leq j \leq m+1)$ vérifient la propriété (14) à l'ordre m+1 .

Par ailleurs $\forall i \in \{1,\ldots,k(m)\} \; \forall r \in \{1,\ldots,\beta(i)\} \; \forall P \in Z'_{m+1,\alpha(i,r)}$ nous avons :

- $P^{(n)}(C_n^{m,i}) \xrightarrow[n\to\infty]{} 1$, d'après (15), car $P \in Z'_{m+1,i}$, vu que $Z'_{m+1,\alpha(i,r)} \subset Z'_{m,i}$;

- $P^{(n)}(D^r_{n,i}) \xrightarrow[n\to\infty]{} 1$, d'après (16), car $P \in Z'_{m+1,\alpha(i,r)}$.

D'après (17), on en déduit $P^{(n)}(C_n^{m+1,\alpha(i,r)}) \xrightarrow[n\to\infty]{} 1$ et la propriété (15) est aussi vérifiée à l'ordre m+1 .

Par récurrence sur m, on peut donc construire une famille de cylindres $(C_n^{j,i}\ ;\ j \geq 1\ ,\ 1 \leq i \leq k(j),\ n \geq 1)$ tels que :

- $\forall n \geq 1\quad \forall j \geq 1\quad (C_n^{j,i}\ ;\ 1 \leq i \leq k(j))$ est une partition finie $\mathcal{B}^{(n)}$-mesurable de $\mathcal{X}^{(n)}$ dont certains éléments peuvent être vides ;

- les propriétés (14) et (15) sont vérifiées.

b) *Construction d'une suite convergente d'estimateurs de* f.

Soit $n \geq 1$. Pour définir $\hat{f}_n : (\mathcal{X}^{(n)}, \mathcal{B}^{(n)}) \longrightarrow (E,d)$ on considère Z'_n et $(C_n^{n,1}, \ldots, C_n^{n,k(n)})$. Pour $i = 1,\ldots,k(n)$ on choisit arbitrairement un point $P_{n,i} \in Z'_{n,i}$ et $\forall x^{(n)} \in \mathcal{X}^{(n)}$ on pose $\hat{f}_n(x^{(n)}) = f(P_{n,i})$ si $x^{(n)} \in C_n^{n,i}$ (en remarquant que $C_n^{n,i} \neq \emptyset \Longrightarrow Z'_{n,i} \neq \emptyset$, par construction).

Montrons que la suite d'estimateurs $(\hat{f}_n\ ;\ n \geq 1)$ ainsi définie est convergente.

Soient $P \in \mathcal{H}$, $\varepsilon > 0$. D'après (12), $\operatorname{diam} Z_{m,h(m,f(P))} \xrightarrow[m\to\infty]{} 0$.

Donc :

$$\exists M_\varepsilon \in \mathbb{N}^* : m \geq M_\varepsilon \Longrightarrow Z_{m,h(m,f(P))} \subset B(f(P),\varepsilon)\ . \tag{18}$$

$\forall m \geq M_\varepsilon\quad \forall i \in \{1,\ldots,k(m)\}$ si $C_m^{m,i} \cap C_m^{M_\varepsilon,h(M_\varepsilon,f(P))} \neq \emptyset$ alors $C_m^{m,i} \subset C_m^{M_\varepsilon,h(M_\varepsilon,f(P))}$, $Z'_{m,i} \subset Z'_{M_\varepsilon,h(M_\varepsilon,f(P))}$ et $\forall x^{(m)} \in C_m^{m,i}\quad \hat{f}_m(x^{(m)}) \in B(f(P),\varepsilon)$. On en déduit que $\forall m \geq M_\varepsilon\quad \{d(\hat{f}_m,f(P)) \leq \varepsilon\} \supset C_m^{M_\varepsilon,h(M_\varepsilon,f(P))}$, puis que $P^{(m)}\{d(\hat{f}_m,f(P)) \leq \varepsilon\} \geq P^{(m)}(C_m^{M_\varepsilon,h(M_\varepsilon,f(P))})$, ce qui permet de conclure puisque $P \in Z'_{M_\varepsilon,h(M_\varepsilon,f(P))}$ et que, d'après (15), il en résulte que $P^{(m)}(C_m^{M_\varepsilon,h(M_\varepsilon,f(P))}) \xrightarrow[m\to\infty]{} 1$. ∎

On voit que la $(\mathcal{H}_t - \mathcal{A}(E))$-mesurabilité de f nous permet de construire une suite convergente d'estimateurs de f. On voudrait traduire cette condition en termes de séparation asymptotique des images réciproques par f de certaines boules de (E,d). Pour cela, on tient compte des résultats suivants.

Lemme II.8. ([39]).- Si (E,d) est un espace métrique séparable, $\forall B = B(e,\alpha)$ $(e \in E,\ \alpha \geq 0)$ boule fermée de E telle que $E - B \neq \emptyset$ on peut trouver une suite $(B(e_n,\alpha_n)\ ;\ n \geq 1)$ de boules fermées de E vérifiant $\forall n \geq 1 \quad d(e,e_n) > \alpha + \alpha_n$ et $E - B = \bigcup_{n\geq 1} B(e_n,\alpha_n)$.

D'autre part, par des méthodes standard déjà utilisées il est facile de montrer que :

Lemme II.9. ([39]).- Soit $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, \mathcal{H})$ un modèle statistique. H_o et H_1 deux hypothèses. S'il existe deux suites d'hypothèses $(H_{or}\ ;\ r \geq 1)$ et $(H_{1s}\ ;\ s \geq 1)$ vérifiant :

$$H_o = \bigcup_{r\geq 1} H_{or} \quad , \quad H_1 = \bigcup_{s\geq 1} H_{1s} \quad , \quad \forall r\ ,\ s \in \mathbb{N}^* \quad H_{or} \text{ et } H_{1s}$$

se séparent asymptotiquement uniformément, alors H_o et H_1 se séparent asymptotiquement.

En utilisant ces résultats, on démontre la

Proposition II.4. ([42]).- Soit $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, \mathcal{H})$ un modèle statistique, (E,d) un espace métrique séparable, $f : \mathcal{H} \to E$ un paramètre. Si la condition (S4) ci-après est satisfaite, f est $(\mathcal{H}_t - \mathcal{A}(E))$-mesurable :

(S4) $\forall e_o , e_1 \in E \quad \forall \alpha_o , \alpha_1 > 0 : d(e_o,e_1) > \alpha_o + \alpha_1$
$(f^{-1}(B(e_i,\alpha_i)))$ $(i=0,1)$ se séparent asymptotiquement uniformément. (19)

Compte tenu de la proposition II.3., lorsque (E,d) est séparable (S4) est une condition suffisante pour l'estimabilité de f. Elle peut pourtant être affaiblie, vu que si (E,d) est séparable il en est de même pour $(f(H),d)$ et qu'en effet cet espace est celui qui nous intéresse.

Proposition II.5. ([42]).- Soit $(X^{(\infty)}, B^{(\infty)}, H)$ un modèle statistique, (E,d) un espace métrique séparable, $f : H \to E$ un paramètre. Pour qu'il existe une suite convergente d'estimateurs de f il suffit que la condition (S5) ci-dessous soit vérifiée :

(S5) $\forall P_o , P_1 \in H \quad \forall \alpha_o , \alpha_1 > 0 : d(f(P_o),f(P_1)) > \alpha_o + \alpha_1$
$(f^{-1}(B(f(P_i),\alpha_i)))$ $(i=0,1)$ se séparent asymptotiquement uniformément. (20)

II.4.3. - Le point de vue bayesien.

Proposition II.6. ([48]).- Soit $(X^{(\infty)}, B^{(\infty)}, H)$ un modèle statistique. On suppose que H est muni de la tribu $\mathcal{P}$ engendrée par les applications $P \longmapsto P^{(n)}(A)$ $(P \in H , n \geq 1 , A \in B^{(n)})$ et que $(H,\mathcal{P})$ est muni d'une probabilité μ . Soit (E,d) un espace métrique, $f : H \to E$ un paramètre. Si f est estimable, il existe une suite $(\hat{f}_n ; n \geq 1)$ d'estimateurs de f convergeant P-p.-s. vers f pour μ-presque toute probabilité $P \in H$.

Corollaire 1. ([42]).- Dans les conditions de la proposition II.6., si (E,d) est séparable et (S5) est satisfaite, f est P-p.-s. estimable pour μ-presque toute probabilité $P \in \mathcal{H}$.

CHAPITRE III

SEPARATION ASYMPTOTIQUE DE LOIS DE PROBABILITE .

III. 1. - NOTIONS PRELIMINAIRES : LA DISTANCE EN VARIATION ET LA DISTANCE DE HELLINGER.-

Définition III.1.- Soit $(X,\mathcal{B})$ un espace probabilisable, $\mathcal{H}$ un ensemble de probabilités sur $\mathcal{B}$, $P, Q \in \mathcal{H}$, μ une mesure σ-finie sur $\mathcal{B}$ dominant à la fois P et Q , $p \in \frac{dP}{d\mu}$, $q \in \frac{dQ}{d\mu}$. La distance en variation d_V entre P et Q est définie par

$$d_V(P,Q) = \sup_{B\in\mathcal{B}} (P(B) - Q(B)) = \frac{1}{2} \int |p-q| \, d\mu \, . \tag{1}$$

Remarque III.1.-

a) La définition ne dépend pas de μ vu que, si μ domine P et Q , μ domine $\nu = P + Q$ et $\frac{dP}{d\mu} = \frac{dP}{d\nu} \cdot \frac{d\nu}{d\mu}$.

b) Le supremum qui intervient dans (1) est atteint pour $B = \{p > q\}$.

Si P et Q sont des lois sur un espace produit infini $(X^{(\infty)}, \mathcal{B}^{(\infty)})$ on a évidemment, pour $n \geq 1$, $d_V(P^{(n)}, Q^{(n)}) = \sup_{B\in\bar{\mathcal{B}}^{(n)}} (P(B) - Q(B)) \leq$

$\leq \sup_{B\in\bar{\mathcal{B}}^{(n+1)}} (P(B) - Q(B)) = d_V(P^{(n+1)}, Q^{(n+1)})$.

Lemme III.1.- Si P et Q sont des lois sur $(X^{(\infty)}, \mathcal{B}^{(\infty)})$

$$d_V(P^{(n)}, Q^{(n)}) \underset{n\to\infty}{\uparrow} d_V(P,Q) \quad .$$

Preuve

a) Pour $n \geq 1$ soit $\mu = P + Q$, $\mu^{(n)} = P^{(n)} + Q^{(n)}$, $p \in \frac{dP}{d\mu}$, $q = 1 - p$, $p^{(n)} = \frac{dP^{(n)}}{d\mu^{(n)}}$, $q^{(n)} = 1 - p^{(n)}$, $0 \leq p \leq 1$, $0 \leq p^{(n)} \leq 1$ $(n \geq 1)$.

Si $f : (X^{(\infty)}, \mathcal{B}^{(\infty)}) \longrightarrow \mathbb{R}$ est $\bar{\mathcal{B}}^{(n)}$-mesurable, on désignera aussi par f la fonction $(x_1,\ldots,x_n) \longmapsto f(x_1,\ldots,x_n,y_{n+1},\ldots)$, où $y_{n+i} \in X_{n+i}$ $(i \geq 1)$, suivant un abus de notation commode.

Commençons par vérifier que $p^{(n)} = E_\mu^{\bar{\mathcal{B}}^{(n)}}(p)$ μ-p.p. :

$\forall A \in \bar{\mathcal{B}}^{(n)}$ $\int_A E_\mu^{\bar{\mathcal{B}}^{(n)}}(p)\, d\mu = \int_A p\, d\mu = P(A)$; si $A' \in \mathcal{B}^{(n)}$ est la base du cylindre A, on a aussi $P(A) = P^{(n)}(A') = \int_{A'} p^{(n)}\, d\mu^{(n)} = \int_A p^{(n)}\, d\mu$.

Donc, puisque $p^{(n)} = E_\mu^{\bar{\mathcal{B}}^{(n)}}(p)$ μ-p.p., $(p^{(n)}\ ;\ n \geq 1)$ est une martingale positive intégrable, qui converge μ-p.-s. et dans $L^1(\mu)$ vers p (cf. [45]).

b) D'après a), $d_V(P^{(n)}, Q^{(n)}) = \frac{1}{2}\int |p^{(n)}-q^{(n)}|\, d\mu^{(n)} =$

$$= \frac{1}{2}\int |2p^{(n)}-1|\, d\mu^{(n)} \xrightarrow[n\to\infty]{} \frac{1}{2}\int |2p-1|\, d\mu = \frac{1}{2}\int |p-q|\, d\mu = d_V(P,Q) \quad . \blacksquare$$

Passons maintenant à la distance de Hellinger.

Avec les mêmes notations, l'affinité de Hellinger entre P et Q, notée $\rho(P,Q)$, est donnée par $\rho(P,Q) = \int \sqrt{pq}\, d\mu$.

Définition III.2.- Soit $(X,\mathcal{B})$ un espace probabilisable, $\mathcal{H}$ un ensemble de lois sur $(X,\mathcal{B})$, P , $Q \in \mathcal{H}$, μ une mesure σ-finie sur $\mathcal{B}$ dominant à la fois P et Q , $p \in \frac{dP}{d\mu}$, $q \in \frac{dQ}{d\mu}$. La distance de Hellinger d_H entre P et Q est donnée par :

$$d_H(P,Q) = \left[\frac{1}{2}\int (\sqrt{p}-\sqrt{q})^2\, d\mu\right]^{1/2} = \left[1-\rho(P,Q)\right]^{1/2} \quad . \tag{2}$$

Remarque III.2.- Pour des raisons similaires à celles indiquées dans la remarque III.1.a) , d_H ne dépend pas de μ .

Lemme III.2.- Soient P et Q deux lois sur un espace produit $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)})$. Notons $\rho(P^{(n)}, Q^{(n)})$ par $\rho_n(P,Q)$. Alors, si $P^{(n)}$ et $Q^{(n)}$ sont des lois-produit, $P^{(n)} = \bigotimes_{i=1}^{n} P_i$ et $Q^{(n)} = \bigotimes_{i=1}^{n} Q_i$, on a :

$$\rho_n(P,Q) = \prod_{i=1}^{n} \rho(P_i,Q_i) \ .$$

Preuve : Soit $\mu_i = P_i + Q_i$, $p_i \in \frac{dP_i}{d\mu_i}$, $q_i = 1 - p_i$ $(i \geq 1)$. Alors,

$$\rho_n(P,Q) = \int \sqrt{p_1(x_1)\ q_1(x_1)} \ldots \sqrt{p_n(x_n)\ q_n(x_n)}\ (\bigotimes_{i=1}^{n} \mu_i)(dx_1,\ldots,dx_n) = \prod_{i=1}^{n} \rho(P_i,Q_i). \blacksquare$$

Lemme III.3.- Soient P et Q deux lois sur un espace produit $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)})$. On a $\forall n \geq 1$ $\rho_n(P,Q) \geq \rho_{n+1}(P,Q)$.

Preuve : Rappelons les notations et résultats de la partie (a) de la preuve du lemme III.1.

$$\rho_n(P,Q) = \int \sqrt{p^{(n)} q^{(n)}\ d\mu^{(n)}} = \int \sqrt{E_\mu^{\bar{\mathcal{B}}^{(n)}}(p^{(n+1)})\ E_\mu^{\bar{\mathcal{B}}^{(n)}}(q^{(n+1)})}\ d\mu \geq$$

$$\geq \int E_\mu^{\bar{\mathcal{B}}^{(n)}} \sqrt{p^{(n+1)} q^{(n+1)}}\, d\mu = \int \sqrt{p^{(n+1)} q^{(n+1)}}\ d\mu^{(n+1)} = \rho_{n+1}(P,Q)$$

L'inégalité ci-dessus résulte d'une double application de l'inégalité de Jensen pour les espérances conditionnelles :

i) $x \longmapsto x^2$ est convexe $\Longrightarrow E^{\mathcal{B}}(p^2) \geq [E^{\mathcal{B}}(p)]^2 \Longrightarrow$

$\Longrightarrow E^{\mathcal{B}}(p) - [E^{\mathcal{B}}(p)]^2 \geq E^{\mathcal{B}}(p-p^2)$, soit

$E^{\mathcal{B}}(p)\ [1 - E^{\mathcal{B}}(p)] = E^{\mathcal{B}}(p)\ E^{\mathcal{B}}(q) \geq E^{\mathcal{B}}(pq) = E^{\mathcal{B}}(p-p^2)$;

ii) $x \longmapsto \sqrt{x}$ est concave $\Longrightarrow \sqrt{E^{\mathcal{B}}(p,q)} \geq E^{\mathcal{B}}(\sqrt{pq})$. ■

De façon similaire à celle du lemme III.1, on en déduit le

Lemme III.4.- Soient P et Q deux lois sur un espace produit infini $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)})$. Alors,

$$\rho_n(P,Q) \underset{n\to\infty}{\downarrow} \rho(P,Q) \quad , \quad d_H(P^{(n)},Q^{(n)}) \underset{n\to\infty}{\uparrow} d_H(P,Q) \quad .$$

Lemme III.5. ([35]).- Soit $(\mathcal{X},\mathcal{B})$ un espace probabilisable, P et Q deux probabilités sur $\mathcal{B}$. On a

$$d_H^2(P,Q) \leq d_V(P,Q) \leq [1-\rho^2(P,Q)]^{1/2} \leq \sqrt{2}\ d_H(P,Q) \quad . \tag{3}$$

Preuve : $|p-q| = |\sqrt{p}-\sqrt{q}|\ |\sqrt{p}+\sqrt{q}| \geq |\sqrt{p}-\sqrt{q}|^2$ permet de vérifier la première inégalité.

La deuxième résulte de

$$\int |p-q|\ d\mu \leq \sqrt{\int (\sqrt{p}-\sqrt{q})^2\ d\mu}\ \sqrt{\int (\sqrt{p}+\sqrt{q})^2\ d\mu} = 2\sqrt{[1-\rho(P,Q)]\ [1+\rho(P,Q)]} \ .$$

L'inégalité de Schwarz entraînant $\rho(P,Q) \leq 1$, la troisième inégalité dans (3) est une conséquence immédiate de la dernière inégalité ci-dessus. ■

Remarque III.3.- d_V et d_H prennent leurs valeurs dans $[0,1]$. Par ailleurs, $d_V(P,Q) = 1 \Longleftrightarrow d_H(P,Q) = 1 \Longleftrightarrow P \perp Q$. En effet, $P \perp Q \Longleftrightarrow \exists A \in \mathcal{B} : Q(A) = P(A^c) = 0 \Longleftrightarrow \exists A \in \mathcal{B} : P(A) - Q(A) = 1 \Longleftrightarrow$ $\Longleftrightarrow d_V(P,Q) = 1$ et $d_V(P,Q) = 1 \Longrightarrow 1 \leq \sqrt{1-\rho^2(P,Q)} \Longrightarrow d_H(P,Q) = 1 \Longrightarrow$ $\Longrightarrow d_V(P,Q) = 1$, d'après (3).

III.2. - LE THEOREME DE KAKUTANI ([34]).-

Restreignons notre attention au cas d'observations indépendantes, c'est-à-dire à des lois-produit $P^{(n)} = \bigotimes_{i=1}^{n} P_i$, $P = \bigotimes_{i=1}^{\infty} P_i$ définies su un espace probabilisable $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)})$.

Remarque III.4.- Pour ne pas alourdir les notations, on désignera par le même symbole les fonctions $\varphi_n : \mathcal{X}^{(n)} \to [0,1] : (x_1,\dots,x_n) \longmapsto \varphi_n(x_1,\dots,x_n)$ et $\varphi_n^* : \mathcal{X}^{(\infty)} \to [0,1] : (x_1,\dots,x_n,x_{n+1},\dots) \longmapsto \varphi_n(x_1,\dots,x_n)$.

Théorème III.1. (de Kakutani).- Soient $P = \bigotimes_{i\geq 1} P_i$ et $Q = \bigotimes_{i\geq 1} Q_i$ deux lois-produit sur $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)})$.

i) $P \perp Q \Longleftrightarrow \prod_{n=1}^{\infty} \rho(P_n,Q_n) = 0$.

ii) Si $\forall n \geq 1$ $P_n \sim Q_n$, c'est-à-dire P_n est dominée par Q_n et réciproquement, alors $P \not\perp Q \Longleftrightarrow P \sim Q$.

Preuve :

i) $\prod_{n=1}^{\infty} \rho(P_n,Q_n) = \lim_{k\to\infty} \rho_k(P,Q) = \rho(P,Q) = 0 \Longleftrightarrow d_H(P,Q) = 1 \Longleftrightarrow P \perp Q$.

ii) Il est évident que $P \sim Q \Longrightarrow P \not\perp Q$.

Réciproquement, si $P \not\perp Q$, $\prod_{n=1}^{\infty} \rho(P_n,Q_n) = \alpha > 0$.

$\forall n \geq 1$ P_n est dominée par Q_n , ce que l'on notera $P_n \prec Q_n$.

Soit $q_n = 1$, $p_n \in \frac{dP_n}{dQ_n}$. Soient m , $n \in \mathbb{N}^*$, $m > n$. On a

$$\prod_{j=1}^{m} \rho(P_j,Q_j) = \int \sqrt{p_1 \dots p_n}\ \sqrt{p_{n+1} \dots p_m}\ dQ_1 \dots dQ_m = \rho_n(P,Q) \int \sqrt{p_{n+1} \dots p_m}\ dQ_{n+1} \dots dQ_m .$$

Puisque $\rho_n(P,Q) \xrightarrow[n\to\infty]{} \alpha$, $\int \sqrt{p_{n+1} \dots p_m}\ dQ_{n+1} \dots dQ_m \xrightarrow[m>n,n\to\infty]{} 1$

Soit $\psi_n = \sqrt{p_1 \dots p_n}$, considérée à la fois comme fonction sur $X^{(n)}$ ou comme fonction sur $X^{(\infty)}$ ne dépendant que des n premières cordonnées, selon la remarque III.4. On a

$$\int (\psi_m - \psi_n)^2\ dQ = \int \psi_n^2\ (\sqrt{p_{n+1} \dots p_m} - 1)^2\ dQ =$$

$$= \int \psi_n^2\ dQ^{(n)} \int (\sqrt{p_{n+1} \dots p_m} - 1)^2\ dQ_{n+1} \dots dQ_m = 2 - 2 \int \sqrt{p_{n+1} \dots p_m}\ dQ_{n+1} \dots dQ_m \xrightarrow[m>n,n\to\infty]{} 0.$$

Donc $(\psi_n ; n\ 1)$ est une suite de Cauchy dans $L^2(Q)$. Puisque $\psi_n \geq 0 (n \geq 1)$, $\exists \psi \in L^2(Q)$, $\psi \geq 0$: $\int (\psi_n - \psi)^2\ dQ \xrightarrow[n\to\infty]{} 0$. Par ailleurs, $1 = ||\psi_n||_2^2 \xrightarrow[n\to\infty]{} ||\psi||_2^2$ et par conséquent ψ^2 est une densité de probabilité par rapport à Q.

Remarquons que $\forall m \in \mathbb{N}^*$

$$\int (\psi_m + \psi)^2\ dQ = 2 + 2 \int \psi\ \psi_m\ dQ \leq 2 + 2 \left(\int \psi_m^2\ dQ \int \psi^2\ dQ \right)^{1/2} = 4 .$$

Soient $m, n \in \mathbb{N}^*$, $m > n$, $A \in \mathcal{B}^{(n)}$.

$$|P(A \times \prod_{j>n} X_j) - \int_{A \times \prod_{j>n} X_j} \psi^2\ dQ| = |P^{(n)}(A) - \int_{A \times \prod_{j>n} X_j} \psi^2\ dQ| =$$

$$= |\int_{A \times \prod_{j>n} X_j} \psi_n^2\ dQ - \int_{A \times \prod_{j>n} X_j} \psi^2\ dQ| \leq \int_{A \times \prod_{j>n} X_j} |\psi_n^2 - \psi^2|\ dQ \leq$$

$$\leq \int |\psi_n - \psi|\ |\psi_n + \psi|\ dQ \leq \left[\int (\psi_n - \psi)^2\ dQ \quad (\psi_n + \psi)^2\ dQ \right]^{1/2} \leq 2\ ||\psi_n - \psi||_2 \xrightarrow[n\to\infty]{} 0$$

Ceci nous permet d'affirmer que $P \prec Q$, $\rho^2 \leq \frac{dP}{dQ}$.

Compte tenu de ce que $\forall n \geq 1$, $Q_n \prec P_n$ et de ce que dans la démonstration ci-dessus les rôles de P et Q peuvent être intervertis, on conclut que $P \sim Q$. ■

III.3. - CRITERE D'ORTHOGONALITE DE DEUX HYPOTHESES SIMPLES A PARTIR DES LOIS CONDITIONNELLES.-

III.3.1. - Introduction - Notations.

Soit (Ω,A,μ) un espace probabilisé, P et Q deux probabilités sur A dominées par μ , $\mathcal{B}_o = \{\emptyset,\Omega\}$, $(\mathcal{B}_n \; ; \; n \geq 1)$ une suite croissante de sous-tribus de A, $\mathcal{B}_\infty = \sigma(\bigcup\limits_{n\geq 1} \mathcal{B}_n)$.

Choisissons des versions non négatives f et g de $\frac{dP}{d\mu}$ et $\frac{dQ}{d\mu}$ respectivement. Puis, pour $n \geq 1$, des versions non négatives de $E_\mu^{\mathcal{B}_n}(f)$ et de $E_\mu^{\mathcal{B}_n}(g)$, notées respectivement f_n et g_n , et des versions non négatives f_∞ de $E_\mu^{\mathcal{B}_\infty}(f)$ et g_∞ de $E_\mu^{\mathcal{B}_\infty}(g)$. Finalement, faisons $f_o = g_o = 1$. On a $f_o \in E_\mu^{\mathcal{B}_o}(f)$, $g_o \in E_\mu^{\mathcal{B}_o}(g)$.

Si $n \in \mathbb{N} \cup \{\infty\}$ et ν est une probabilité sur A, la restriction de ν à $\mathcal{B}_n$ sera notée $\nu_{\mathcal{B}_n}$. Il est clair que $\forall n \in \mathbb{N} \cup \{\infty\}$ on a

$P_{\mathcal{B}_n} < \mu_{\mathcal{B}_n}$, $Q_{\mathcal{B}_n} < \mu_{\mathcal{B}_n}$, $f_n \in \frac{dP_{\mathcal{B}_n}}{d\mu_{\mathcal{B}_n}}$, $g_n \in \frac{dQ_{\mathcal{B}_n}}{d\mu_{\mathcal{B}_n}}$. En effet,

$$\forall B \in \mathcal{B}_n \quad P(B) = \int_B f \, d\mu = \int_B E_\mu^{\mathcal{B}_n}(f) \, d\mu_{\mathcal{B}_n} = \int_B f_n \, d\mu_{\mathcal{B}_n} .$$

III.3.2. - Calcul de $\rho(P_{\mathcal{B}_\infty}, Q_{\mathcal{B}_\infty})$ *([40]).*

Remarquons que $\rho(P_{\mathcal{B}_o}, Q_{\mathcal{B}_o}) = 1$. D'autre part, $\forall n \geq 1$

$$\rho(P_{\mathcal{B}_n}, Q_{\mathcal{B}_n}) = \rho(P_{\mathcal{B}_o}, Q_{\mathcal{B}_o}) + \sum_{k=1}^{n} [\rho(P_{\mathcal{B}_k}, Q_{\mathcal{B}_k}) - \rho(P_{\mathcal{B}_{k-1}}, Q_{\mathcal{B}_{k-1}})] =$$

$$= 1 + \sum_{k=1}^{n} \left[\int \sqrt{f_k g_k}\, d\mu_{\mathcal{B}_k} - \int \sqrt{f_{k-1} g_{k-1}}\, d\mu_{\mathcal{B}_{k-1}}\right] =$$

$$= 1 - \sum_{k=1}^{n} \int (\sqrt{f_{k-1} g_{k-1}} - \sqrt{f_k g_k})\, d\mu \quad . \tag{4}$$

On voudrait mettre en facteur dans l'expression entre parenthèses $\sqrt{f_{k-1} g_{k-1}}$, ce qui pose problème vu que $f_{k-1} g_{k-1}$ peut s'annuler. Mais on peut remarquer que $P_{\mathcal{B}_k}(f_{k-1}=0) = \int_{\{f_{k-1}=0\}} f_k \, d\mu_{\mathcal{B}_k} = \int_{\{f_{k-1}=0\}} E_\mu^{\mathcal{B}_{k-1}}(f_k)\, d\mu_{\mathcal{B}_{k-1}} =$

$= \int_{\{f_{k-1}=0\}} E_\mu^{\mathcal{B}_{k-1}}(f)\, d\mu_{\mathcal{B}_{k-1}} = \int_{\{f_{k-1}=0\}} f_{k-1}\, d\mu_{\mathcal{B}_{k-1}} = 0$. Donc

$\int_{\{f_{k-1}=0\}} f_k \, d\mu_{\mathcal{B}_k} = 0$ et $\mu_{\mathcal{B}_k}(f_k > 0 \, , \, f_{k-1} = 0) = 0$.

Par conséquent, en convenant que $\frac{o}{o} = 1$, $\frac{f_k}{f_{k-1}}$ est $\mu_{\mathcal{B}_k}$-p.-s. définie. Il en est de même pour $\frac{g_k}{g_{k-1}}$. Il faut pourtant retenir que ces fonctions sont $\mu_{\mathcal{B}_k}$-p.-s. définies au moyen d'une convention.

Sur $\{f_{k-1} > 0 \, , \, g_{k-1} > 0\}$ on a évidemment

$$\sqrt{f_{k-1} g_{k-1}} - \sqrt{f_k g_k} = \sqrt{f_{k-1} g_{k-1}} \left(1 - \sqrt{\frac{f_k}{f_{k-1}} \frac{g_k}{g_{k-1}}}\right) \quad .$$

Sur $\{f_{k-1}=0\} \cup \{g_{k-1}=0\}$ on a encore cette égalité $\mu_{\mathcal{B}_k}$-p.-s., puisque $\sqrt{f_{k-1}g_{k-1}} - \sqrt{f_k g_k} \underset{\mu_{\mathcal{B}_k}\text{-p.-s.}}{=} 0 \underset{\mu_{\mathcal{B}_k}\text{-p.-s.}}{=} \sqrt{f_{k-1}g_{k-1}}\ (1 - \sqrt{\frac{f_k}{f_{k-1}}\frac{g_k}{g_{k-1}}})$.

En reportant dans (4), on obtient

$$\forall n \geq 1 \quad \rho(P_{\mathcal{B}_n}, Q_{\mathcal{B}_n}) = 1 - \sum_{k=1}^{n} \int \sqrt{f_{k-1}g_{k-1}}\ (1 - \sqrt{\frac{f_k}{f_{k-1}}\frac{g_k}{g_{k-1}}})\ d\mu_{\mathcal{B}_k} \ . \tag{5}$$

Montrons maintenant que $\sqrt{\frac{f_k}{f_{k-1}}\frac{g_k}{g_{k-1}}}$ est $\mu_{\mathcal{B}_k}$-intégrable.

$\forall k \geq 0$ soient $\hat{f}_k = f_k + \mathbb{1}_{\{f_k=0\}}$, $\hat{g}_k = g_k + \mathbb{1}_{\{g_k=0\}}$.

$\forall k \geq 1$ soient $\bar{f}_k = f_k + \mathbb{1}_{\{f_{k-1}=0\}}$, $\bar{g}_k = g_k + \mathbb{1}_{\{g_{k-1}=0\}}$.

On a $\frac{f_k}{f_{k-1}} = \frac{f_k + \mathbb{1}_{\{f_{k-1}=0\}}}{f_{k+1} + \mathbb{1}_{\{f_{k-1}=0\}}} = \frac{\bar{f}_k}{\hat{f}_{k-1}}$, cette dernière expression étant un quotient de fonctions défini sans aucune convention sous-jacente. Or

$$\forall k \geq 1 \quad \int \frac{f_k}{f_{k-1}}\ d\mu_{\mathcal{B}_k} = \int \frac{\bar{f}_k}{\hat{f}_{k-1}}\ d\mu_{\mathcal{B}_k} = \int_{\{f_{k-1}=0\}} E_\mu^{\mathcal{B}_{k-1}}(\frac{\bar{f}_k}{\hat{f}_{k-1}})\ d\mu_{\mathcal{B}_{k-1}} +$$

$$+ \int_{\{f_{k-1}>0\}} E_\mu^{\mathcal{B}_{k-1}}(\frac{\bar{f}_k}{\hat{f}_{k-1}})\ d\mu_{\mathcal{B}_{k-1}} = \int_{\{f_{k-1}=0\}} E_\mu^{\mathcal{B}_{k-1}}(1)\ d\mu_{\mathcal{B}_{k-1}} +$$

$$+ \int_{\{f_{k-1}>0\}} \frac{1}{\hat{f}_{k-1}} E_\mu^{\mathcal{B}_{k-1}}(\bar{f}_k)\ d\mu_{\mathcal{B}_{k-1}} = \mu_{\mathcal{B}_{k-1}}\{f_{k-1}=0\} +$$

$$+ \int_{\{f_{k-1}>0\}} \frac{1}{f_{k-1}} E_\mu^{\mathcal{B}_{k-1}}(f_k)\ d\mu_{\mathcal{B}_{k-1}} = \mu_{\mathcal{B}_{k-1}}\{f_{k-1}=0\} +$$

$$+ \int_{\{f_{k-1}>0\}} \frac{f_{k-1}}{f_{k-1}}\ d\mu_{\mathcal{B}_{k-1}} = \mu_{\mathcal{B}_{k-1}}\{f_{k-1}=0\} + \mu_{\mathcal{B}_{k-1}}\{f_{k-1}>0\} = 1 \ .$$

$\frac{f_k}{f_{k-1}}$ est donc la densité, par rapport à $\mu_{\mathcal{B}_k}$, d'une probabilité que l'on identifiera ultérieurement (voir remarque III.6 ci-dessous). Il en est de même de $\dfrac{g_k}{g_{k-1}}$. Ainsi, $\sqrt{\dfrac{f_k}{f_{k-1}}}$ et $\sqrt{\dfrac{g_k}{g_{k-1}}}$ sont de carré intégrable par rapport à $\mu_{\mathcal{B}_k}$ et leur produit est donc $\mu_{\mathcal{B}_k}$-intégrable. Donc, pour $k \geq 1$:

- $\sqrt{f_{k-1}g_{k-1}}\,(1 - \sqrt{\dfrac{f_k}{f_{k-1}}\dfrac{g_k}{g_{k-1}}})$ est $\mu_{\mathcal{B}_k}$-intégrable, car égal $\mu_{\mathcal{B}_k}$-p.-s. à $\sqrt{f_{k-1}g_{k-1}} - \sqrt{f_k g_k}$;

- $\sqrt{f_{k-1}g_{k-1}}\,(1 - \sqrt{\dfrac{f_k}{f_{k-1}}\dfrac{g_k}{g_{k-1}}})$ est le produit de deux fonctions $\mu_{\mathcal{B}_k}$-intégrables.

En conditionnant dans (5), on obtient :

$$\forall n \geq 1 \quad \rho(P_{\mathcal{B}_n}, Q_{\mathcal{B}_n}) = 1 - \sum_{k=1}^{n} \int \sqrt{f_{k-1}g_{k-1}}\; E_\mu^{\mathcal{B}_{k-1}}(1 - \sqrt{\frac{f_k}{f_{k-1}}\frac{g_k}{g_{k-1}}})\, d\mu_{\mathcal{B}_{k-1}} . \quad (6)$$

Puisque $\rho(P_{\mathcal{B}_\infty}, Q_{\mathcal{B}_\infty}) = \lim_{n\to\infty} \downarrow \rho(P_{\mathcal{B}_n}, Q_{\mathcal{B}_n})$,

$$\rho(P_{\mathcal{B}_\infty}, Q_{\mathcal{B}_\infty}) = 1 - \sum_{k=1}^{\infty} \int \sqrt{f_{j-1}g_{k-1}}\; E_\mu^{\mathcal{B}_{k-1}}\,(1 - \sqrt{\frac{f_k}{f_{k-1}}\frac{g_k}{g_{k-1}}})\, d\mu_{\mathcal{B}_{k-1}} . \quad (7)$$

Comme $d_H^2(P_{\mathcal{B}_\infty}, Q_{\mathcal{B}_\infty}) = 1 - \rho(P_{\mathcal{B}_\infty}, Q_{\mathcal{B}_\infty})$, on a finalement

$$d_H^2(P_{\mathcal{B}_\infty}, Q_{\mathcal{B}_\infty}) = \sum_{n=1}^{\infty} \int \sqrt{f_{n-1}g_{n-1}}\; E_\mu^{\mathcal{B}_{n-1}}\,(1 - \sqrt{\frac{f_n}{f_{n-1}}\frac{g_n}{g_{n-1}}})\, d\mu_{\mathcal{B}_{n-1}} . \quad (8)$$

III.3.3. - Rappels sur les lois conditionnelles.

On veut exprimer $E_\mu^{\mathcal{B}_{n-1}}\,(1 - \sqrt{\dfrac{f_n}{f_{n-1}}\dfrac{g_n}{g_{n-1}}})$ $(n \geq 1)$ en fonction de la distance de Hellinger entre certaines probabilités conditionnelles.

Définition III.3.- Soit $(\Omega, \mathcal{A}, \mu)$ un espace probabilisé, $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. On appelle version régulière de la probabilité μ conditionnée par $\mathcal{B}$ toute fonction $\mu^{\mathcal{B}} : \Omega \times \mathcal{A} \longrightarrow [0,1]$ telle que :

- $\forall A \in \mathcal{A} \quad \mu^{\mathcal{B}}(.,A) \in E_{\mu}^{\mathcal{B}}(\mathbb{1}_A)$;
- $\forall \omega \in \Omega \quad \mu^{\mathcal{B}}(\omega,.)$ est une probabilité sur $\mathcal{A}$.

Lemme III.6.- Soit $(\Omega, \mathcal{A}, \mu)$ un espace probabilisé, $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. Si la tribu $\mathcal{A}$ est de type dénombrable et s'il existe au moins une version régulière $\mu^{\mathcal{B}}$ de μ conditionnée par $\mathcal{B}$, alors pour toute autre version régulière μ_1 de μ conditionnée par $\mathcal{B}$ on peut trouver $\Omega_1 \in \mathcal{B}$ tel que $\mu(\Omega_1) = 1$ et $\forall \omega \in \Omega_1 \quad \mu^{\mathcal{B}}(\omega,.) = \mu_1^{\mathcal{B}}(\omega,.)$.

Preuve : Soit $(A_n \; ; \; n \geq 1)$ une sous-algèbre de $\mathcal{A}$ engendrant $\mathcal{A}$. $\forall n \geq 1 \quad \mu^{\mathcal{B}}(.,A_n)$ et $\mu_1^{\mathcal{B}}(.,A_n)$ sont des versions de $E_{\mu}^{\mathcal{B}}(\mathbb{1}_{A_n})$. Donc, $C_n = \{\omega : \mu^{\mathcal{B}}(\omega,A_n) = \mu_1^{\mathcal{B}}(\omega,A_n)\}$ est un élément de $\mathcal{B}$ $\mu_{\mathcal{B}}$-certain et il en est de même de $\Omega_1 = \bigcap_{n \geq 1} C_n$. De plus, $\forall \omega \in \Omega_1 \quad \mu^{\mathcal{B}}(\omega,.) = \mu_1^{\mathcal{B}}(\omega,.)$ sur $(A_n \; ; \; n \geq 1)$. Le fait que $\mathcal{A} = \sigma(A_n \; ; \; n \geq 1)$ permet de conclure. ∎

On rappelle le résultat suivant, dont la démonstration, qui se fait par des méthodes standard, est omise.

Lemme III.7.- Soit $(\Omega, \mathcal{A}, \mu)$ un espace probabilisé, $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. Supposons qu'il existe une version régulière $\mu^{\mathcal{B}}$ de μ conditionnée par $\mathcal{B}$. Alors, pour toute fonction $f : (\Omega, \mathcal{A}) \longrightarrow [0,\infty[$ mesurable, $\omega \longmapsto \int f(\omega') \, \mu^{\mathcal{B}}(\omega, d\omega') \in E_{\mu}^{\mathcal{B}}(f)$.

Remarque III.5.- Dans les conditions du lemme précédent, $\omega \longmapsto \int f(\omega') \, \mu^{\mathcal{B}}(\omega, d\omega')$ est une fonction à valeurs dans $[0,\infty]$.

Lemme III.8.- Soient P et Q deux probabilités sur un espace probabilisable $(\Omega,\mathcal{A})$, $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. On suppose qu'il existe une version régulière $P^{\mathcal{B}}$ (resp. $Q^{\mathcal{B}}$) de P (resp. de Q) conditionnée par $\mathcal{B}$. Alors, il existe une version régulière $\mu^{\mathcal{B}}$ de $\mu = \frac{1}{2}(P+Q)$ conditionnée par $\mathcal{B}$.

Preuve : $\forall B \in \mathcal{B} \quad \forall A \in \mathcal{A} \int_B 1_A \, d\mu = \frac{1}{2}\left(\int_B 1_A \, dP + \int_B 1_A \, dQ\right) =$

$= \frac{1}{2}\left[\int_B P^{\mathcal{B}}(\omega,A)\, P_{\mathcal{B}}(d\omega) + \int_B Q^{\mathcal{B}}(\omega,A)\, Q_{\mathcal{B}}(d\omega)\right]$.

Choisissons $f_{\mathcal{B}} \in \frac{dP_{\mathcal{B}}}{d\mu_{\mathcal{B}}}$ et $g_{\mathcal{B}} \in \frac{dQ_{\mathcal{B}}}{d\mu_{\mathcal{B}}}$ telles que $f_{\mathcal{B}} + g_{\mathcal{B}} = 2$, $f_{\mathcal{B}} \geq 0$, $g_{\mathcal{B}} \geq 0$.

Alors on a :

$$\int_B 1_A \, d\mu = \int_B \frac{1}{2}\left[P^{\mathcal{B}}(\omega,A)\, f_{\mathcal{B}}(\omega) + Q^{\mathcal{B}}(\omega,A)\, g_{\mathcal{B}}(\omega)\right] \mu_{\mathcal{B}}(d\omega) \quad .$$

Posons $\Psi(\omega,A) = \frac{1}{2}\left[P^{\mathcal{B}}(\omega,A)\, f_{\mathcal{B}}(\omega) + Q^{\mathcal{B}}(\omega,A)\, g_{\mathcal{B}}(\omega)\right]$.

$\forall \omega \in \Omega$ $\Psi(\omega,.)$ est une probabilité sur $(\Omega,\mathcal{A})$ puisque

$$\Psi(\omega,\Omega) = \frac{1}{2}\left[P^{\mathcal{B}}(\omega,\Omega)\, f_{\mathcal{B}}(\omega) + Q^{\mathcal{B}}(\omega,\Omega)\, g_{\mathcal{B}}(\omega)\right] = \frac{1}{2}\left[f_{\mathcal{B}}(\omega) + g_{\mathcal{B}}(\omega)\right] = 1 \quad .$$

De plus, $\forall A \in \mathcal{A}$ $\Psi(.,A)$ est $\mathcal{B}$-mesurable et le calcul précédent permet de conclure que $\Psi(.,.)$ est une version régulière de μ conditionnée par $\mathcal{B}$. ■

Lemme III.9.- Soit $(\Omega,\mathcal{A})$ un espace probabilisable, $\mathcal{A}$ de type dénombrable, $\mathcal{A}$ une sous-tribu de $\mathcal{A}$, P et Q deux probabilités sur $\mathcal{A}$ telles que $P_{\mathcal{B}} \sim Q_{\mathcal{B}}$.

On suppose qu'il existe des versions régulières $P^{\mathcal{B}}$ et $Q^{\mathcal{B}}$ de P et Q conditionnées par $\mathcal{B}$, respectivement.

Soit $\mu = \frac{1}{2}(P+Q)$. Alors, $\exists \Omega_* \in \mathcal{B} : \mu(\Omega_*) = 1$ et l'application $\omega \longmapsto d_H^2(P^{\mathcal{B}}(\omega,.), Q^{\mathcal{B}}(\omega,.))\, 1_{\Omega_*}(\omega)$ est $\mathcal{B}$-mesurable.

<u>*Preuve*</u> : Soit $f \in \frac{dP_{\mathcal{B}}}{d\mu_{\mathcal{B}}}$. Si $P^{\mathcal{B}}$ est une version régulière de P conditionnée par $\mathcal{B}$, il en est de même de $P_1^{\mathcal{B}}$ définie sur $\Omega \times \mathcal{A}$ par

$$P_1^{\mathcal{B}}(\omega,A) = P^{\mathcal{B}}(\omega,1)\, \mathbb{1}_{\{f_{\mathcal{B}}>0\}}(\omega) + P(A)\, \mathbb{1}_{\{f_{\mathcal{B}}=0\}}(\omega) \quad .$$

Définissons $Q_1^{\mathcal{B}}$ de façon analogue.

Si Ψ est la version régulière de μ conditionnée par $\mathcal{B}$ construite comme dans le lemme précédant à partir de $P_1^{\mathcal{B}}$ et $Q_1^{\mathcal{B}}$, $\forall \omega \in \Omega$ $\Psi(\omega,.)$ domine à la fois $P_1^{\mathcal{B}}(\omega,.)$ et $Q_1^{\mathcal{B}}(\omega,.)$.

Soit $f_\omega \in \frac{dP_1^{\mathcal{B}}(\omega,.)}{d\Psi(\omega,.)}$, $g_\omega \in \frac{dQ_1^{\mathcal{B}}(\omega,.)}{d\Psi(\omega,.)}$. D'après le lemme III.7., $\vartheta : \omega \longmapsto d_H^2(P_1^{\mathcal{B}}(\omega,.), Q_1^{\mathcal{B}}(\omega,.))$ est une version de $E_\mu^{\mathcal{B}}[\frac{1}{2}(\sqrt{f_\omega}-\sqrt{g_\omega})^2]$, donc $\mathcal{B}$-mesurable.

Or, d'après le lemme III.6., $\exists \Omega_1 \in \mathcal{B} : P(\Omega_1) = 1$, $P_1^{\mathcal{B}}(\omega,.) = P^{\mathcal{B}}(\omega,.)$ sur Ω_1 et $\exists \Omega_2 \in \mathcal{B} : Q(\Omega_2) = 1$, $Q_1^{\mathcal{B}}(\omega,.) = Q^{\mathcal{B}}(\omega,.)$ sur Ω_2.

Vu que $P_{\mathcal{B}} \sim Q_{\mathcal{B}}$, $\Omega_* = \Omega_1 \cap \Omega_2$ est simultanément P-certain et Q-certain et, sur Ω_* , ϑ coïncide avec $\omega \mapsto d_H^2(P^{\mathcal{B}}(\omega,.), Q^{\mathcal{B}}(\omega,.))$. Donc, $\exists \Omega_* \in \mathcal{B} : \mu(\Omega_*) = 1$ et $\omega \mapsto d_H^2(P^{\mathcal{B}}(\omega,.), Q^{\mathcal{B}}(\omega,.))\, \mathbb{1}_{\Omega_*}(\omega)$ est $\mathcal{B}$-mesurable, comme nous voulions démontrer. ■

<u>*Lemme III.10*</u> ([40]).- Soit $(\Omega,\mathcal{A},\mu)$ un espace probabilisé, $\mathcal{B}$ une sous-tribu de $\mathcal{A}$, P une probabilité sur $\mathcal{A}$ dominé par μ. S'il existe une version régulière $\mu^{\mathcal{B}}$ de μ conditionnée par $\mathcal{B}$, il existe une version régulière de P conditionnée par $\mathcal{B}$.

Remarque III.6.- Une telle version est construite ci-dessous.

On choisit d'abord une version f non négative de $\frac{dP}{d\mu}$, puis une version $f_{\mathcal{B}}$ non négative de $E_{\mu}^{\mathcal{B}}(f)$. On a $f_{\mathcal{B}} \in \frac{dP_{\mathcal{B}}}{d\mu_{\mathcal{B}}}$. La version régulière $P_{*}^{\mathcal{B}}$ de P conditionnée par $\mathcal{B}$ que nous allons construire vérifie : pour $\mu_{\mathcal{B}}$-presque tout , $P_{*}^{\mathcal{B}}(\omega,.)$ est la probabilité sur A de densité $\frac{f}{f_{\mathcal{B}}}$ par rapport à $\mu^{\mathcal{B}}(\omega,.)$, $\frac{f}{f_{\mathcal{B}}}$ étant $\mu^{\mathcal{B}}(\omega,.)$ - p.-s. définie grâce à la convention $\frac{0}{0} = 1$.

Preuve du lemme III.10.- Sous la convention $\frac{0}{0} = 1$, $\frac{f}{f_{\mathcal{B}}}$ est μ-p.-s. définie comme l'on a déjà souligné. En effet, cette fonction est définie sur $\Omega_1 = \Omega - \{f > 0 , f_{\mathcal{B}} = 0\}$ et

$$P\{f_{\mathcal{B}}=0\} = \int_{\{f_{\mathcal{B}}=0\}} f\, d\mu = \int_{\{f_{\mathcal{B}}=0\}} E_{\mu}^{\mathcal{B}}(f)\, d\mu_{\mathcal{B}} = \int_{\{f_{\mathcal{B}}=0\}} f_{\mathcal{B}}\, d\mu_{\mathcal{B}} = 0 \Longrightarrow \mu(f>0, f_{\mathcal{B}}=0) = 0 .$$

Introduisons, comme ci-dessus, des fonctions auxiliaires qui permettent de substituer à $\frac{f}{f_{\mathcal{B}}}$ un vrai quotient :

$$\bar{f} = f + \mathbb{1}_{\{f_{\mathcal{B}}=0\}} , \qquad \hat{f}_{\mathcal{B}} = f_{\mathcal{B}} + \mathbb{1}_{\{f_{\mathcal{B}}=0\}} .$$

Alors $\frac{f}{f_{\mathcal{B}}} \underset{\mu\text{-p.-s.}}{=} \frac{\bar{f}}{\hat{f}_{\mathcal{B}}}$, donc $E_{\mu}^{\mathcal{B}}(\frac{f}{f_{\mathcal{B}}}) = E_{\mu}(\frac{\bar{f}}{\hat{f}_{\mathcal{B}}})$. De plus,

sur $\{f_{\mathcal{B}} > 0\}$ on a $E_{\mu}^{\mathcal{B}}(\frac{\bar{f}}{\hat{f}_{\mathcal{B}}}) \underset{\mu_{\mathcal{B}}\text{-p.-s.}}{=} \frac{1}{\hat{f}_{\mathcal{B}}} E_{\mu}^{\mathcal{B}}(\bar{f}) = \frac{1}{f_{\mathcal{B}}} E_{\mu}^{\mathcal{B}}(f) \underset{\mu_{\mathcal{B}}\text{-p.s.}}{=} 1$, et

sur $\{f_{\mathcal{B}}=0\}$, on a $E_{\mu}^{\mathcal{B}}(\frac{\bar{f}}{\hat{f}_{\mathcal{B}}}) = E_{\mu}^{\mathcal{B}}(1) \underset{\mu_{\mathcal{B}}\text{-p.-s.}}{=} 1$. On en déduit que $1 \in E_{\mu}^{\mathcal{B}}(\frac{f}{f_{\mathcal{B}}})$

D'autre part, on peut calculer $E_{\mu}^{\mathcal{B}}(\frac{\bar{f}}{\hat{f}_{\mathcal{B}}})$ à partir de $\mu^{\mathcal{B}}$.

Plus précisément, $\omega \longmapsto \int \frac{\bar{f}(\omega')}{\hat{f}_{\mathcal{B}}(\omega')}\, \mu^{\mathcal{B}}(\omega, d\omega') \in E_{\mu}^{\mathcal{B}}(\frac{\bar{f}}{\hat{f}})$. On a donc,

pour $\mu_{\mathcal{B}}$-presque tout ω, $1 = \int \frac{\bar{f}(\omega')}{\hat{f}_{\mathcal{B}}(\omega')}\, \mu^{\mathcal{B}}(\omega, d\omega')$.

$\Omega_1 = \{\frac{f}{f_{\mathcal{B}}} = \frac{\bar{f}}{\hat{f}_{\mathcal{B}}}\}$ étant μ-certain, nous avons $1 = \mu(\Omega_1) = \int \mathbb{1}_{\Omega_1}\, d\mu =$ $= \int E_\mu^{\mathcal{B}}(\mathbb{1}_{\Omega_1})\, d\mu_{\mathcal{B}} = \int \mu^{\mathcal{B}}(\omega,\Omega_1)\, \mu_{\mathcal{B}}(d\omega)$. Donc, pour $\mu_{\mathcal{B}}$-presque tout ω, $\mu^{\mathcal{B}}(\omega,\Omega_1) = 1$; autrement dit, pour $\mu_{\mathcal{B}}$-presque tout ω $\frac{f}{f_{\mathcal{B}}} = \frac{\bar{f}}{\hat{f}_{\mathcal{B}}}$ $\mu^{\mathcal{B}}(\omega,.)$-p.-s.

En résumé : on peut trouver $\Omega_* \in \mathcal{B}$ tel que $\mu(\Omega_*) = 1$, $\forall\omega \in \Omega_*$ $\frac{\bar{f}}{\hat{f}_{\mathcal{B}}} = \frac{f}{f_{\mathcal{B}}}$ $\mu^{\mathcal{B}}(\omega,.)$-p.-s. et $\forall\omega \in \Omega_*$

$$1 = \int \frac{\bar{f}(\omega')}{\hat{f}_{\mathcal{B}}(\omega')}\, \mu^{\mathcal{B}}(\omega, d\omega') = \int \frac{f}{f_{\mathcal{B}}}(\omega')\, \mu^{\mathcal{B}}(\omega, d\omega') \quad .$$

Considérons $P_*^{\mathcal{B}} : \Omega \times \mathcal{A} \longrightarrow [0,1]$ définie par :

$- \forall A \in \mathcal{A}\ \forall\omega \in \Omega_*\quad P_*^{\mathcal{B}}(\omega,A) = \int_A \frac{f}{f_{\mathcal{B}}}(\omega')\, \mu^{\mathcal{B}}(\omega, d\omega')$;

$- \forall A \in \mathcal{A}\ \forall\omega \notin \Omega_*\quad P_*^{\mathcal{B}}(\omega,A) = P(A)$.

Il est clair que par construction $\forall\omega \in \Omega$ $P_*^{\mathcal{B}}(\omega,.)$ est une probabilité sur $\mathcal{A}$. De plus, $\forall A \in \mathcal{A}$ $P_*^{\mathcal{B}}(.,A)$ est constante sur $\Omega - \Omega_* \in \mathcal{B}$; sur Ω_*, $P_*(.,A)$ coïncide avec $\omega \longmapsto \frac{\bar{f}(\omega')}{\hat{f}_{\mathcal{B}}(\omega')} \mathbb{1}_A(\omega')\, \mu^{\mathcal{B}}(\omega,d\omega') \in E_\mu^{\mathcal{B}}(\frac{\bar{f}}{\hat{f}_{\mathcal{B}}} \mathbb{1}_A)$, laquelle est $\mathcal{B}$-mesurable. C'est-à-dire : $\forall A \in \mathcal{A}$ $P_*^{\mathcal{B}}(.,A)$ est $\mathcal{B}$-mesurable. Vérifions qu'il s'agit d'une version de $E_P^{\mathcal{B}}(\mathbb{1}_A)$.

$$\forall B \in \mathcal{B} \quad \int_B P_*^{\mathcal{B}}(\omega,A)\, P_{\mathcal{B}}(d\omega) =$$

$$= \int_{B \cap \Omega_*} \left(\int \frac{\bar{f}(\omega')}{\hat{f}_{\mathcal{B}}(\omega')} \mathbb{1}_A(\omega')\, \mu^{\mathcal{B}}(\omega,d\omega')\right) P_{\mathcal{B}}(d\omega) =$$

$$= \int_B E_\mu^{\mathcal{B}}\left(\frac{\bar{f}}{\hat{f}_{\mathcal{B}}} \mathbb{1}_A\right) dP_{\mathcal{B}} = \int_B E_\mu^{\mathcal{B}}\left(\frac{\bar{f}}{\hat{f}_{\mathcal{B}}} \mathbb{1}_A\right) f_{\mathcal{B}}\, d\mu_{\mathcal{B}} =$$

$$\left(\text{car } f_{\mathcal{B}} \in \frac{dP_{\mathcal{B}}}{d\mu_{\mathcal{B}}}\right)$$

$$= \int_{B \cap \{f_{\mathcal{B}}>0\}} E_\mu^{\mathcal{B}}\left(\frac{f}{\hat{f}_{\mathcal{B}}} \mathbb{1}_A f_{\mathcal{B}}\right) d\mu_{\mathcal{B}} = \int_{B \cap \{f_{\mathcal{B}}>0\}} E_\mu^{\mathcal{B}}(f\, \mathbb{1}_A)\, d\mu =$$

$$= \int_{B \cap \{f_{\mathcal{B}}>0\}} \mathbb{1}_A\, dP = P(A \cap B \cap \{f_{\mathcal{B}}>0\}) =$$

$$= P(A \cap B) - P(A \cap B \cap \{f_{\mathcal{B}} = 0\} \cap \{f>0\}) = P(A \cap B) = \int_B \mathbb{1}_A\, dP\ .$$

On peut donc conclure que $P_*^{\mathcal{B}}$ est une version régulière de P conditionnée par $\mathcal{B}$ et que $\forall \omega \in \Omega_*$ $P_*^{\mathcal{B}}(\omega,.)$ est la probabilité sur $\mathcal{A}$ de densité $\frac{f}{f_{\mathcal{B}}}$ par rapport à $\mu^{\mathcal{B}}(\omega,.)$, compte tenu de la convention $\frac{0}{0} = 1$. ■

III.3.4. - La condition suffisante d'orthogonalité de deux lois de J. Kabanov, R. Lipcer et A. Širjaev ([33]).

Reprenons le calcul de $d_H(P_{\mathcal{B}_\infty}, Q_{\mathcal{B}_\infty})$, en utilisant les notations des sous-sections antérieures et en introduisant une hypothèse supplémentaire :

h_1') $\forall n \geq 1$ il existe une version régulière $\mu_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}$ de la probabilité $\mu_{\mathcal{B}_n}$ sur $(\Omega, \mathcal{B}_n)$ conditionnée par $\mathcal{B}_{n-1}$. (9)

Soit $n \geq 1$. Considérons l'espace probabilisé $(\Omega, \mathcal{B}_n, \mu_{\mathcal{B}_n})$ et la probabilité $P_{\mathcal{B}_n}$ sur $(\Omega, \mathcal{B}_n)$, absolument continue par rapport à $\mu_{\mathcal{B}_n}$ puisque $P \prec \mu$.

On a $f_n \in \frac{dP_{\mathcal{B}_n}}{d\mu_{\mathcal{B}_n}}$ et $f_{n-1} = E_\mu^{\mathcal{B}_{n-1}}(f) = E_\mu^{\mathcal{B}_{n-1}}(E_\mu^{\mathcal{B}_n}(f)) = E_\mu^{\mathcal{B}_{n-1}}(f_n)$.

D'après le lemme III.10, on peut construire une version régulière $P_{*_{\mathcal{B}_n}}^{\mathcal{B}_{n-1}}$ de $P_{\mathcal{B}_n}$ conditionnée par $\mathcal{B}_{n-1}$ qui vérifie : pour $\mu_{\mathcal{B}_{n-1}}$ - presque tout ω , $P_{*_{\mathcal{B}_n}}^{\mathcal{B}_{n-1}}(\omega,.)$ est la probabilité sur $(\Omega,\mathcal{B}_n)$ de densité $\frac{f_n}{f_{n-1}}$ par rapport à $\mu_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}(\omega,.)$, la définition $\mu_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}$ -p.-s. de $\frac{f_n}{f_{n-1}}$ faisant intervenir la convention $\frac{o}{o} = 1$.

On introduit de même $Q_{*_{\mathcal{B}_n}}^{\mathcal{B}_{n-1}}$.

Pour $\mu_{\mathcal{B}_{n-1}}$ - presque tout ω on a donc

$$\rho(P_{*_{\mathcal{B}_n}}^{\mathcal{B}_{n-1}}(\omega,.) , Q_{*_{\mathcal{B}_n}}^{\mathcal{B}_{n-1}}(\omega,.)) = \int \sqrt{\frac{f_n}{f_{n-1}}(\omega')\ \frac{g_n}{g_{n-1}}(\omega')}\ \mu_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}(\omega,d\omega') \quad \text{et}$$

$$\omega \longmapsto \int \sqrt{\frac{f_n}{f_{n-1}}(\omega')\ \frac{g_n}{g_{n-1}}(\omega')}\ \mu_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}(\omega,d\omega') \in E_{\mu_{\mathcal{B}_n}}^{\mathcal{B}_{n-1}}\left(\sqrt{\frac{f_n}{f_{n-1}}\frac{g_n}{g_{n-1}}}\right) = E_\mu^{\mathcal{B}_{n-1}}\left(\sqrt{\frac{f_n}{f_{n-1}}\frac{g_n}{g_{n-1}}}\right).$$

En reportant dans (8) on obtient :

$$d_H^2(P_{\mathcal{B}_\infty}, Q_{\mathcal{B}_\infty}) = \sum_{n=1}^{\infty} \int \sqrt{f_{n-1}(\omega)g_{n-1}(\omega)}\ d_H^2(P_{*_{\mathcal{B}_n}}^{\mathcal{B}_{n-1}}(\omega,.),Q_{*_{\mathcal{B}_n}}^{\mathcal{B}_{n-1}}(\omega,.))\ \mu_{\mathcal{B}_{n-1}}(d\omega) \ . \qquad (10)$$

Faisons deux nouvelles hypothèses supplémentaires :

h_2') $\forall n \geq 1$ $\mathcal{B}_n$ est une tribu de type dénombrable . (11)

h_3') $\forall n \geq 1$ $P_{\mathcal{B}_n} \sim Q_{\mathcal{B}_n}$. (12)

Alors, si $\forall n \geq 1$ $P_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}$ (resp. $Q_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}$) désigne une version régulière quelconque de la probabilité $P_{\mathcal{B}_n}$ (resp. $Q_{\mathcal{B}_n}$) sur $(\Omega,\mathcal{B}_n)$ conditionnée par $\mathcal{B}_{n-1}$, compte tenu des lemmes III.6. et III.9. on a :

$$d_H^2(P_{\mathcal{B}_\infty},Q_{\mathcal{B}_\infty}) = \sum_{n\geq 1} \int \sqrt{f_{n-1}(\omega)g_{n-1}(\omega)}\; d_H^2(P_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}(\omega,.),\, Q_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}(\omega,.))\; \mu_{\mathcal{B}_{n-1}}(d\omega) \quad . (13)$$

Remarquons que, dans cette formule, on peut aussi remplacer f_{n-1} (resp. g_{n-1}) par une version quelconque de $\frac{dP_{\mathcal{B}_{n-1}}}{d\mu_{\mathcal{B}_{n-1}}}$ (resp. de $\frac{dQ_{\mathcal{B}_{n-1}}}{d\mu_{\mathcal{B}_{n-1}}}$) et que, grâce au théorème de la convergence monotone, $\sum$ et $\int$ peuvent être intervertis.

Avant de faire cette intervention, calculons le premier terme de la série (13). Nous avons :

- $f_o = g_o = 1$;
- $\forall A \in \mathcal{B}_1$ $P_{\mathcal{B}_1}^{\mathcal{B}_o}(.,A) \in E_{P_{\mathcal{B}_1}}^{\mathcal{B}_o}(1_A) = P_{\mathcal{B}_1}(A)$, car $\mathcal{B}_o = \{\emptyset,\Omega\}$;
- de même, $\forall A \in \mathcal{B}_1$ $Q_{\mathcal{B}_1}^{\mathcal{B}_o}(.,A) = Q_{\mathcal{B}_1}(A)$.

On en déduit que

$$\int \sqrt{f_o(\omega)g_o(\omega)}\; d_H^2(P_{\mathcal{B}_1}^{\mathcal{B}_o}(\omega,.),\, Q_{\mathcal{B}_1}^{\mathcal{B}_o}(\omega,.))\; \mu_{\mathcal{B}_o}(d\omega) = \int d_H^2(P_{\mathcal{B}_1},Q_{\mathcal{B}_1})\; \mu_{\mathcal{B}_o}(d\omega) = d_H^2(P_{\mathcal{B}_1},Q_{\mathcal{B}_1}).$$

En résumant l'étude précédente, qui évidemment permet aussi le calcul de $d_H^2(P_{\mathcal{B}_n},Q_{\mathcal{B}_n})$ $(n\geq 2)$, on peut écrire :

<u>*Théorème III.2.*</u>- Soit $(\Omega,\mathcal{A},\mu)$ un espace probabilisé, $(\mathcal{B}_n\ ;\ n \geq 1)$ une suite croissante de sous-tribus de $\mathcal{A}$, $\mathcal{B}_\infty = \sigma(\bigcup_{n\geq 1} \mathcal{B}_n)$, P et Q deux probabilités sur $\mathcal{A}$ dominées par μ .

On suppose que :

- $\forall n \geq 2$ $\mathcal{B}_n$ est une tribu de type dénombrable ;

- $\forall n \geq 2$ il existe une version régulière de la probabilité $\mu_{\mathcal{B}_n}$ sur $(\Omega,\mathcal{B}_n)$ conditionnée par $\mathcal{B}_{n-1}$.

- $\forall n \geq 1$ $P_{\mathcal{B}_n} \sim Q_{\mathcal{B}_n}$

Alors :

- $\forall n \geq 1$ $P_{\mathcal{B}_n}$ (resp. $Q_{\mathcal{B}_n}$) est absolument continue par rapport à $\mu_{\mathcal{B}_n}$; on note f_n (resp. g_n) une version non négative de $\dfrac{dP_{\mathcal{B}_n}}{d\mu_{\mathcal{B}_n}}$ (resp. de $\dfrac{dQ_{\mathcal{B}_n}}{d\mu_{\mathcal{B}_n}}$) .

- $\forall n \geq 2$ il existe au moins une version régulière de la probabilité $P_{\mathcal{B}_n}$ (resp. $Q_{\mathcal{B}_n}$) sur $(\Omega,\mathcal{B}_n)$ conditionnée par $\mathcal{B}_{n-1}$; on note $P_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}$ (resp. $Q_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}$) l'une quelconque de ces versions.

- $\forall n \geq 2$

$$d_H^2(P_{\mathcal{B}_n},Q_{\mathcal{B}_n}) = d_H^2(P_{\mathcal{B}_1},Q_{\mathcal{B}_1}) + \int \Big[\sum_{k=2}^{n} \sqrt{f_{k-1}(\omega)g_{k-1}(\omega)}\; d_H^2(P_{\mathcal{B}_k}^{\mathcal{B}_{k-1}}(\omega,.),Q_{\mathcal{B}_k}^{\mathcal{B}_{k-1}}(\omega,.))\Big]\mu(d\omega), \tag{14}$$

$$d_H^2(P_{\mathcal{B}_\infty},Q_{\mathcal{B}_\infty}) = d_H^2(P_{\mathcal{B}_1},Q_{\mathcal{B}_1}) + \int \Big[\sum_{k\geq 2} \sqrt{f_{k-1}(\omega)g_{k-1}(\omega)}\; d_H^2(P_{\mathcal{B}_k}^{\mathcal{B}_{k-1}}(\omega,.),Q_{\mathcal{B}_k}^{\mathcal{B}_{k-1}}(\omega,.))\Big]\mu(d\omega). \tag{15}$$

<u>*Corollaire 1.*</u> (*J. Kabanov, R. Lipcer et A. Širjaev* [33], *J. Mémin* [38]).-

Pour que P et Q soient orthogonales il suffit que

$$\mu\,\{\omega : \sum_{n\geq 2} d_H^2(P_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}(\omega,.),\; Q_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}(\omega,.)) = \infty\} = 1 \quad .$$

<u>*Preuve*</u> : D'après (15) et puisqu'une fonction non négative $\mu_{\mathcal{B}_\infty}$-intégrable est $\mu_{\mathcal{B}_\infty}$-p.p. finie, nous avons :

- pour $\mu_{\mathcal{B}_\infty}$-presque tout ω

$$\sum_{n\geq 2} \sqrt{f_{n-1}(\omega) g_{n-1}(\omega)}\; d_H^2(P_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}(\omega,.),\; Q_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}(\omega,.)) < \infty \quad . \tag{16}$$

Soit f_∞ (resp. g_∞) une version non négative de $\dfrac{dP_{\mathcal{B}_\infty}}{d\mu_{\mathcal{B}_\infty}} = E_\mu^{\mathcal{B}_\infty}(f)$ (resp. de $\dfrac{dQ_{\mathcal{B}_\infty}}{d\mu_{\mathcal{B}_\infty}} = E_\mu^{\mathcal{B}_\infty}(g)$) .

$(f_n \; ; \; n \geq 1)$ (resp. $(g_n \; ; \; n \geq 1)$) converge $\mu_{\mathcal{B}_\infty}$-p.-s. vers f_∞ (resp. vers g_∞) (voir [45]).

Compte-tenu du lemme III.9, pour $\mu_{\mathcal{B}_\infty}$-presque tout ω on a donc, en plus de (16) :

- $\displaystyle\sum_{n\geq 2} d_H^2(P_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}(\omega,.),\; Q_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}(\omega,.)) = \infty$;
- $f_{n-1}(\omega) \xrightarrow[n\to\infty]{} f_\infty(\omega)$;
- $g_{n-1}(\omega) \xrightarrow[n\to\infty]{} g_\infty(\omega)$.

On en déduit que pour $\mu_{\mathcal{B}_\infty}$-presque tout ω $(f_\infty\, g_\infty)(\omega) = 0$.

Donc $\mu_{\mathcal{B}_\infty}(f_\infty > 0,\ g_\infty > 0) = 0$ et par conséquent $P_{\mathcal{B}_\infty}(f_\infty > 0,\ g_\infty = 0) = 0$. D'où $P_{\mathcal{B}_\infty}(f_\infty > 0,\ g_\infty = 0) = P_{\mathcal{B}_\infty}(f_\infty > 0) = 1$, tandis qu'il est évident que $Q_{\mathcal{B}_\infty}(f_\infty > 0 \ ,\ g_\infty = 0) = 0$: $P_{\mathcal{B}_\infty}$ et $Q_{\mathcal{B}_\infty}$, et a fortiori P et Q , sont orthogonales. ■

Corollaire 2. ([40]).-

$$P_{\mathcal{B}_\infty} < Q_{\mathcal{B}_\infty} \Longrightarrow P_{\mathcal{B}_\infty}\{\omega : \sum_{n\geq 2} d_H^2(P_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}(\omega,.), Q_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}(\omega,.)) < \infty\} = 1 .$$

Preuve : En reprenant les arguments de la démonstration précédente, et vu que $P_{\mathcal{B}_\infty} \prec \mu_{\mathcal{B}_\infty}$, on a, pour $P_{\mathcal{B}_\infty}$ - presque tout ω :

- $\sum_{n\geq 2} \sqrt{f_{n-1}(\omega)g_{n-1}(\omega)}\, d_H^2(P_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}(\omega,.), Q_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}(\omega,.)) < \infty$;
- $f_{n-1}(\omega) \xrightarrow[n\to\infty]{} f_\infty(\omega)$;
- $g_{n-1}(\omega) \xrightarrow[n\to\infty]{} g_\infty(\omega)$;
- $(f_\infty\, g_\infty)(\omega) > 0$, vu que $P_{\mathcal{B}_\infty} \prec Q_{\mathcal{B}_\infty} \Longrightarrow P_{\mathcal{B}_\infty}(f_\infty > 0, g_\infty = 0) = 0$.

On en déduit que, pour $P_{\mathcal{B}_\infty}$ - presque tout ω , $\sum_{n\geq 2} d_H^2(P_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}(\omega,.), Q_{\mathcal{B}_n}^{\mathcal{B}_{n-1}}(\omega,.)) < \infty$. ■

III.3.5. - Application à l'orthogonalité de deux hypothèses simples en statistique.

Soient P et Q deux probabilités sur un espace produit infini $(X^{(\infty)}, \mathcal{B}^{(\infty)}) = \bigotimes_{n\geq 1} (X_n,\mathcal{B}_n)$. Pour $n \geq 1$ posons $\mathcal{B}^{(n)} = \bigotimes_{k=1}^{n} \mathcal{B}_k$, $\bar{\mathcal{B}}^{(n)} = \mathcal{B}^{(n)} \times \prod_{j>n} X_j$. $\mathcal{B}^{(n)}$ est une tribu sur $X^{(n)} = \prod_{k=1}^{n} X_k$, $\bar{\mathcal{B}}^{(n)}$ une tribu sur $X^{(\infty)}$. $(\bar{\mathcal{B}}^{(n)} ; n \geq 1)$ est une suite croissante de sous-tribus de $\mathcal{B}^{(\infty)}$ qui vérifie $\mathcal{B}^{(\infty)} = \sigma(\bigcup_{n\geq 1} \bar{\mathcal{B}}^{(n)})$.

Pour $n \geq 1$ on pose aussi $\bar{P}^{(n)} = P \restriction \bar{\mathcal{B}}^{(n)}$, $\bar{Q}^{(n)} = Q \restriction \bar{\mathcal{B}}^{(n)}$.

On peut interpréter P et Q comme deux hypothèses simples pour une suite d'observations $X = (X_n \; ; \; n \geq 1) : (\Omega, A) \longrightarrow (\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)})$. On notera $P_{X^{(n)}}$ (resp. $Q_{X^{(n)}}$) la loi de $X^{(n)} = (X_1, \dots, X_n)$ sous l'hypothèse P (resp. Q). On a :

$$\forall B \in \mathcal{B}^{(n)} \quad P_{X^{(n)}}(B) = \bar{P}^{(n)}(B \times \prod_{j>n} \mathcal{X}_j) = P(B \times \prod_{j>n} \mathcal{X}_j) \; .$$

Admettons que les hypothèses (h_1), (h_2) et (h_3) suivantes sont vérifiées :

h_1) $\forall n \geq 2$ on peut choisir une version $P^{(n)}(.,.)$ (resp. $Q^{(n)}(.,.)$) de la loi conditionnelle de $X^{(n)}$ sachant $X^{(n-1)}$ sous l'hypothèse P (resp. Q). (17)

h_2) $\forall n \geq 1$ $\mathcal{B}_n$ est de type dénombrable . (18)

h_3) $\forall n \geq 1$ $P_{X^{(n)}} \sim Q_{X^{(n)}}$. (19)

Rappelons que $\forall n \geq 2$ $P^{(n)}(.,.) : \mathcal{X}^{(n-1)} \times \mathcal{B}^{(n)} \longrightarrow [0,1]$ (et de même $Q^{(n)}(.,.)$) a par définition les propriétés suivantes :

a) $\forall B \in \mathcal{B}^{(n)}$ $P^{(n)}(.,B)$ est $\mathcal{B}^{(n-1)}$-mesurable et vérifie

$$\forall A \in \mathcal{B}^{(n-1)} \int_A P^{(n)}(x^{(n-1)},B)\, P_{X^{(n-1)}}(dx^{(n-1)}) =$$

$$= \Pr(X^{(n-1)} \in A, \, X^{(n)} \in B) = P_{X^{(n)}} \left[(A \times \mathcal{X}_n) \cap B\right] \, ,$$

où $\Pr$ désigne la probabilité sur (Ω, A) dont P est l'image par X.

b) $\forall x^{(n-1)} \in \mathcal{X}^{(n-1)}$ $P^{(n)}(x^{(n-1)},.)$ est une probabilité sur $(\mathcal{X}^{(n)}, \mathcal{B}^{(n)})$ (dite loi de $X^{(n)}$ sachant que $X^{(n-1)} = x^{(n-1)}$) .

Le produit de deux tribus de type dénombrable étant de type dénombrable, il résulte de (18) que $\forall n \geq 1$ $\mathcal{B}^{(n)}$ est de type dénombrable.

Donné $x^{(\infty)} \in \mathcal{X}^{(\infty)}$, soit $x^{(n-1)}$ le (n-1)-uple des n-1 premières coordonnées de $x^{(\infty)}$. $\forall n \geq 2$ $\forall x^{(\infty)} \in \mathcal{X}^{(\infty)}$ $\forall B \in \mathcal{B}^{(n)}$ on pose

$$P_*^{(n)}(x^{(\infty)}, B \times \prod_{j>n} \mathcal{X}_j) = P^{(n)}(x^{(n-1)}, B) .$$

On voit aisément que $P_*^{(n)}(.,.)$ est une version régulière de la probabilité $\bar{P}^{(n)}$ sur $(\mathcal{X}^{(\infty)}, \bar{\mathcal{B}}^{(n)})$ sachant $\bar{\mathcal{B}}^{(n-1)}$. Il en est de même de $Q_*^{(n)}(.,.)$, définie de manière analogue à partir de $Q^{(n)}(.,.)$.

De plus, si $\mu = \frac{1}{2}(P+Q)$ et si $\forall n \geq 1$ on pose $\bar{\mu}^{(n)} = \mu \upharpoonright \bar{\mathcal{B}}^{(n)}$, on a $\bar{\mu}^{(n)} = \frac{1}{2}(\bar{P}^{(n)} + \bar{Q}^{(n)})$, probabilité sur $(\mathcal{X}^{(\infty)}, \bar{\mathcal{B}}^{(n)})$ qui admet une version régulière sachant $\bar{\mathcal{B}}^{(n-1)}$, d'après le lemme III.8.

On peut donc appliquer le corollaire 1 du théorème III.2. : pour que $P \perp Q$ il suffit que $\mu\{x^{(\infty)} : \sum_{n\geq 2} d_H^2(P_*^{(n)}(x^{(\infty)},.),Q_*^{(n)}(x^{(\infty)},.)) = \infty\} = 1$ ou, ce qui est équivalent, que

$$\left.\begin{array}{l} P\{x^{(\infty)} : \sum_{n\geq 2} d_H^2(P_*^{(n)}(x^{(\infty)},.),\ Q_*^{(n)}(x^{(\infty)},.)) = \infty\} = 1 \\ \text{et} \\ Q\{x^{(\infty)} : \sum_{n\geq 2} d_H^2(P_*^{(n)}(x^{(\infty)},.),\ Q_*^{(n)}(x^{(\infty)},.)) = \infty\} = 1 . \end{array}\right| \qquad (20)$$

Nous voulons revenir aux données : $P^{(n)}(.,.)$ et $Q^{(n)}(.,.)$ ($n \geq 2$). Leur rapport avec les $P_*^{(n)}(.,.)$ et $Q_*^{(n)}(.,.)$, respectivement, conduit à formuler la condition suffisante d'orthogonalité présentée ci-dessus de la façon suivante :

pour que $P \perp Q$, il suffit que l'on ait, pour tout $x^{(\infty)} \in \mathcal{X}^{(\infty)}$ d'un ensemble simultanément P-certain et Q-certain,

$$\sum_{n\geq 2} d_H^2(P^{(n)}(x^{(n-1)},.),\ Q^{(n)}(x^{(n-1)},.)) = \infty$$

ou, avec une notation plus suggestive,

$$\sum_{n\geq 2} d_H^2(P_{X^{(n)}}^{X^{(n-1)}=x^{(n-1)}},\ Q_{X^{(n)}}^{X^{(n-1)}=x^{(n-1)}}) = \infty\ . \tag{21}$$

<u>*Lemme III.11.-*</u> Soit $(\Omega,\mathcal{A})$ un espace probabilisable, $\mathcal{B}$ une sous-tribu de $\mathcal{A}$, P et Q deux probabilités sur $\mathcal{A}$. On a :

$$d_H(P_{\mathcal{B}},\ Q_{\mathcal{B}}) \leq d_H(P,Q)\ . \tag{22}$$

Preuve : Suivant des démarches analogues à celles du lemme III.3., soit $\mu = P+Q$, $f \in \frac{dP}{d\mu}$, $g \in \frac{dQ}{d\mu}$. On a $E_\mu^{\mathcal{B}}(f) \in \frac{dP_{\mathcal{B}}}{d\mu_{\mathcal{B}}}$, $E_\mu^{\mathcal{B}}(g) \in \frac{dQ_{\mathcal{B}}}{d\mu_{\mathcal{B}}}$ et

$$\rho(P_{\mathcal{B}},Q_{\mathcal{B}}) = \int \sqrt{E_\mu^{\mathcal{B}}(f)\ E_\mu^{\mathcal{B}}(g)}\ d\mu \geq \int E_\mu^{\mathcal{B}}(\sqrt{fg})\ d\mu = \int \sqrt{fg}\ d\mu = \rho(P,Q)\ ,$$

ce qui permet de conclure. ■

Ce lemme nous amène à une autre condition suffisante d'orthogonalité de deux hypothèses, à savoir :

<u>*Proposition III.1.-*</u> Soient $(\mathcal{X}_n,\mathcal{B}_n)$ $(n\geq 1)$ des espaces probabilisables, $X = (X_n\ ;\ n \geq 1)$ une suite d'observations, X_n à valeurs dans $\mathcal{X}_n$ $(n \geq 1)$. On considère deux hypothèses pour X , P et Q , probabilités sur $(\mathcal{X}^{(\infty)},\ \mathcal{B}^{(\infty)})$, et l'on admet que :

- $\forall n \geq 1$ $\mathcal{B}_n$ est de type dénombrable ;
- $\forall n \geq 1$ $P_{X^{(n)}} \sim Q_{X^{(n)}}$;
- $\forall n \geq 2$ il existe une version régulière $P_{X^{(n)}}^{X^{(n-1)}}$ (resp. $Q_{X^{(n)}}^{X^{(n-1)}}$)

de la loi de $X^{(n)}$ conditionnée par $X^{(n-1)}$, sous l'hypothèse P (resp. Q).

Une condition suffisante pour que $P \perp Q$ est que, pour tout $x^{(\infty)} \in \mathcal{X}^{(\infty)}$ d'un événement simultanément P-certain et Q-certain, on ait :

$$\sum_{n\geq 2} d_H^2(P_{X_n}^{X^{(n-1)}=x^{(n-1)}} , Q_{X_n}^{X^{(n-1)}=x^{(n-1)}}) = \infty , \tag{23}$$

où $x^{(n-1)}$ est le (n-1)-uple des n-1 premières coordonnées de $x^{(\infty)}$.

Preuve : Commençons par remarquer que, $\forall n \geq 2$, l'existence des versions régulières $P_{X_n}^{X^{(n-1)}}$ (resp. $Q_{X_n}^{X^{(n-1)}}$) de la loi de X_n sachant $X^{(n-1)}$ sous l'hypothèse P (resp. Q) est assurée : on n'a qu'à les définir comme la restriction à la sous-tribu de $\mathcal{B}^{(n)}$ des événements de la forme $\mathcal{X}^{(n-1)} \times A$, avec $A \in \mathcal{B}_n$, de $P_{X^{(n)}}^{X^{(n-1)}}$ et $Q_{X^{(n)}}^{X^{(n-1)}}$, respectivement.

D'après (22),

$$d_H(P_{X_n}^{X^{(n-1)}=x^{(n-1)}} , Q_{X_n}^{X^{(n-1)}=x^{(n-1)}}) \leq d_H(P_{X^{(n)}}^{X^{(n-1)}=x^{(n-1)}} , Q_{X^{(n)}}^{X^{(n-1)}=x^{(n-1)}})$$

$(n \geq 2 \; ; \; x^{(n-1)} \in \mathcal{X}^{(n-1)})$.

Donc, (23) $\Longrightarrow$ (21), ce qui permet d'établir le résultat prétendu. ■

Si on se limite à des hypothèses sur $(\mathbb{R}^s , \mathcal{B}_{\mathbb{R}^s})^{(\infty)}$ $(s \in \mathbb{N}^*)$, le type dénombrable des tribus $\mathcal{B}^{(n)}$ est assuré et l'existence d'une version régulière de la loi de X_n conditionnée par $X^{(n-1)}$ $(n \geq 2)$ résulte d'un théorème de M. Jiřina ([32]). On peut ainsi énoncer le théorème suivant, qui est une conséquence immédiate de la proposition III.1. et de ces considérations.

Théorème III.3.- Soit $s \in \mathbb{N}^*$, $X = (X_n \,;\, n \geq 1)$ une suite d'observations de $\mathbb{R}^s$ dont la loi inconnue sur $\mathcal{B}^{(\infty)}_{\mathbb{R}^s}$ est P ou Q .

Supposons que $\forall n \geq 1 \quad P_{X^{(n)}} \sim Q_{X^{(n)}}$.

Alors, pour que $P \perp Q$ il suffit que, pour tout $x \in (\mathbb{R}^s)^{(\infty)}$ d'un événement simultanément P-certain et Q-certain on ait :

$$\sum_{n\geq 2} d_H^2 \left(P_{X_n}^{X^{(n-1)}=x^{(n-1)}} , Q_{X_n}^{X^{(n-1)}=x^{(n-1)}}\right) = \infty \, , \tag{24}$$

où $x^{(n-1)}$ désigne le (n-1)-uple des n-1 premières coordonnées de x.

III.4. - SEPARABILITE DE DEUX HYPOTHESES SIMPLES.-

Théorème III.4.- Soient $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, \mathcal{H})$ un modèle statistique, $H_o = \{P\}$ et $H_1 = \{Q\}$ deux hypothèses. H_o et H_1 se séparent asymptotiquement uniformément si et seulement si $P \perp Q$.

Preuve : Si $\exists (C_n \,;\, n \geq 1)$ $\mathcal{B}^{(n)}$-adaptée telle que $P^{(n)}(C_n) \xrightarrow[n\to\infty]{} 1$, $Q^{(n)}(C_n) \xrightarrow[n\to\infty]{} 0$ soit $(n_k \,;\, k \geq 1)$ une suite de naturels strictement croissante telle que $Q^{n_k}(C_{n_k}) \leq \frac{1}{k^2}$. Posons $C = \limsup\limits_{k\to\infty} C_{n_k}$. Alors

$$P(C) = \lim_{j\to\infty} P\Big[\bigcup_{k\geq j} (C_{n_k} \times \prod_{i>n_k} \mathcal{X}_i)\Big] \geq \lim_{j\to\infty} P^{(n_j)}(C_{n_j}) = 1 \quad \text{et}$$

$$Q(C) = \lim_{j\to\infty} Q\Big[\bigcup_{k\geq j} (C_{n_k} \times \prod_{i>n_k} \mathcal{X}_i)\Big] \leq \lim_{j\to\infty} \sum_{k\geq j} Q^{(n_k)}(C_{n_k}) \leq \lim_{j\to\infty} \sum_{k\geq j} \frac{1}{k^2} = 0 \, .$$

Donc $P \perp Q$.

Réciproquement, si $P \perp Q$, $d_V(P^{(n)}, Q^{(n)}) \xrightarrow[n\to\infty]{} 1$. Soient $\mu^{(n)} = P^{(n)} + Q^{(n)}$, $p^{(n)} \in \dfrac{dP^{(n)}}{d\mu^{(n)}}$, $C_n = \{p^{(n)} \geq \frac{1}{2}\}$. On a

$$P^{(n)}(C_n) = P^{(n)}(p^{(n)} \geq 1-p^{(n)}) = d_V(P^{(n)},Q^{(n)}) + Q^{(n)}(p^{(n)} \geq 1 - p^{(n)}) \geq$$

$$\geq d_V(P^{(n)},Q^{(n)}) \xrightarrow[n\to\infty]{} 1$$

et

$$Q^{(n)}(C_n) = P^{(n)}(C_n) - d_V(P^{(n)},Q^{(n)}) \leq 1 - d_V(P^{(n)},Q^{(n)}) \xrightarrow[n\to\infty]{} 0 \quad . \blacksquare$$

Corollaire 1.- Dans un modèle concernant une suite d'observations indépendantes identiquement distribuées, deux hypothèses simples distinctes sont toujours asymptotiquement séparées.

Preuve : $\rho_n(P,Q) = [\rho(P^{(1)},Q^{(1)})]^n \xrightarrow[n\to\infty]{} 0$. $\blacksquare$

III.5. - SEPARABILITE DE DEUX HYPOTHESES MULTIPLES.-

Théorème III.5.- Soit $(X^{(\infty)}, \mathcal{B}^{(\infty)}, \mathcal{H})$ un modèle statistique, $H_o = (P_i \ ; \ i \geq 1)$ et $H_1 = (Q_i \ ; \ i \geq 1)$ deux hypothèses dénombrables. Alors, H_o et H_1 se séparent asymptotiquement si et seulement si $\forall i, j \in \mathbb{N}^* \ P_i \perp Q_j$.

Preuve : La condition étant évidemment nécessaire, passons tout de suite à la démonstration de sa suffisance.

D'après le théorème III.2., $\forall i, j \in \mathbb{N}^* \ \exists (A^n_{ij} \ ; \ n \geq 1)$ $\mathcal{B}^{(n)}$-adaptée telle que $\forall i \geq 1 \quad P^{(n)}_i(A^n_{ij}) \xrightarrow[n\to\infty]{} 1$, $\forall j \geq 1 \quad Q^{(n)}_j(A^n_{ij}) \xrightarrow[n\to\infty]{} 0$.

Si $C^n_{rs} = \bigcup_{i=1}^{r} (\bigcap_{j=1}^{s} A^n_{ij})$, $(C^n_{rs} ; n \geq 1)$ sépare asymptotiquement (uniformément, vu qu'il s'agit d'ensembles finis) $H^r_o = \bigcup_{i=1}^{r} \{P_i\}$ et $H^s_1 = \bigcup_{j=1}^{s} \{Q_j\}$.

En effet, $\forall i \in \{1,\ldots,r\}$

$$P^{(n)}_i(C^n_{rs}) \geq P^{(n)}_i(\bigcap_{j=1}^{s} A^n_{ij}) = 1-P^{(n)}_i \left[\bigcup_{j=1}^{s} (X^{(n)}-A^n_{ij})\right] \geq 1 - \sum_{j=1}^{s} P^{(n)}(X^{(n)}-A^n_{ij}) \xrightarrow[n\to\infty]{} 1$$

et $\forall j \in \{1,\ldots,s\}$ $\quad Q^{(n)}_j(C^n_{rs}) \leq Q^{(n)}_j(\bigcup_{i=1}^{r} A^n_{ij}) \xrightarrow[n\to\infty]{} 0$.

$$\forall p \in \mathbb{N}^* \ \exists m_p \in \mathbb{N}^* : n \geq m_p \Longrightarrow \inf_{P\in H^p_o} P^{(n)}(C^n_{pp}) - \sup_{Q\in H^p_1} Q^{(n)}(C^n_{pp}) > 1 - \frac{1}{p} .$$

$(m_p ; p \geq 1)$ peut être choisie strictement croissante.

Pour $n \geq m_1$, soit $p(n)$ tel que $m_{p(n)} \leq n < m_{p(n)+1}$.

Posons
$$C_n = C^{m_{p(n)}}_{p(n),p(n)} \times \prod_{j=m_{p(n)}+1}^{n} X_j .$$

$$\forall P_i \in H_o \quad n \geq m_{p(i)} \Longrightarrow P^{(n)}_i(C_n) = P_i^{(m_{p(n)})}(C^{m_{p(n)}}_{p(n),p(n)}) \geq 1 - \frac{1}{p(n)} \xrightarrow[n\to\infty]{} 1$$

$$\text{et } \forall Q_j \in H_1 \quad n \geq m_{p(j)} \Longrightarrow Q^{(n)}_j(C_n) = Q_j^{(m_{p(n)})}(C^{m_{p(n)}}_{p(n),p(n)}) \leq \frac{1}{p(n)} \xrightarrow[n\to\infty]{} 0 . \blacksquare$$

Corollaire 1. ([35]).- Dans un modèle concernant une suite de variables aléatoires indépendantes identiquement distribuées, deux hypothèses dénombrables H_o et H_1 sont asymptotiquement séparées dès que disjointes.

Si l'on veut énoncer une condition suffisante pour que deux hypothèses dénombrables H_o et H_1 se séparent asymptotiquement uniformément on doit imposer des contraintes difficiles à vérifier. Dans les cas les plus courants, on est amené à considérer les enveloppes convexes $C(H_o)$ et $C(H_1)$ de ces hypothèses, comme le montre le théorème III.6. ci-dessous (dû à L. Le Cam, cité dans [35]). On abandonne alors l'hypothèse que H_o et H_1 sont dénombrables.

Remarque III.7.- Il est facile de vérifier que, pour tout espace probabilisable $(X,\mathcal{B})$ tout test φ défini sur X et toute paire H_o, H_1 de familles de probabilités sur $(X,\mathcal{B})$

$$\inf_{P\in H_1} \int \varphi \, dP - \sup_{P\in H_o} \int \varphi \, dP = \inf_{P\in C(H_1)} \int \varphi \, dP - \sup_{P\in C(H_o)} \int \varphi \, dP =$$

$$= \inf_{P\in \overline{C(H_1)}} \int \varphi \, dP - \sup_{P\in \overline{C(H_o)}} \int \varphi \, dP ,$$

où $\bar{A}$ désigne l'adhérence de A, par rapport à d_V.

Théorème III.6.- Soit $(X,\mathcal{B})$ un espace probabilisable, H_o et H_1 deux ensembles de probabilités sur cet espace dominées par une mesure σ-finie μ sur $(X,\mathcal{B})$, $C(H_i)$ $(i=0,1)$ leurs enveloppes convexes. Soit $\varepsilon > 0$.

Une condition nécessaire et suffisante pour qu'il existe un test φ défini sur $(X,\mathcal{B})$ tel que

$$\sup_{P\in H_1} \int \varphi \, dP + \varepsilon \leq \inf_{P\in H_o} \int \varphi \, dP \tag{25}$$

est que $d_V(C(H_o), C(H_1)) \geq \varepsilon$.

Preuve : Supposons (25) vérifiée. Notons p une version de $\frac{dP}{d\mu}$, q une version de $\frac{dQ}{d\mu}$, où $P \in C(H_o)$, $Q \in C(H_1)$.

$$\varepsilon \leq \int \varphi \, dP - \int \varphi \, dQ = \int \varphi \, p \, d\mu - \int \varphi \, q \, d\mu .$$

Soit $A = \{p \geq q\}$. On a

$$\varepsilon \leq \int_A \phi(p-q)\, d\mu \leq \int_A (p-q)\, d\mu = P(A) - Q(A) = d_V(P,Q) \quad .$$

La condition est donc nécessaire.

Pour voir qu'elle est suffisante, supposons $d_V(\mathcal{C}(H_o), \mathcal{C}(H_1)) \geq \varepsilon$. Soit $b(0,2\varepsilon)$ la boule ouverte de centre 0 et rayon 2ε dans l'espace de Banach $L^1_{\mathbb{R}}(\mathcal{X},\mathcal{B},\mu)$. Soit $C = \{\frac{dP}{d\mu} - \frac{dQ}{d\mu} + g : P \in \mathcal{C}(H_o),\ Q \in \mathcal{C}(H_1),\ g \in b(0,2\varepsilon)\}$.

C est évidemment convexe, est ouvert comme réunion d'ouverts :

$$C = \bigcup_{(P,Q)\in\mathcal{C}(H_o)\times\mathcal{C}(H_1)} \left(\frac{dP}{d\mu} - \frac{dQ}{d\mu} + b(0,2\varepsilon)\right).$$ De plus, C ne contient pas 0.

Sinon on pourrait trouver $P \in \mathcal{C}(H_o)$, $Q \in \mathcal{C}(H_1)$, $g \in b(0,2\varepsilon)$ tels que, avec $p \in \frac{dP}{d\mu}$ et $q \in \frac{dQ}{d\mu}$, $0 = p-q+g$. On aurait donc $||p-q||_1 = 2\, d_V(P,Q) < 2\varepsilon$. Or $d_V(P,Q) \geq \epsilon$.

Le dual de $L^1_{\mathbb{R}}(\mathcal{X},\mathcal{B},\mu)$ est l'espace M des classes d'équivalence de fonctions mesurables bornées, avec la norme

$$||f||_\infty = \mu - \text{ess. sup} . |f| = \inf \{\lambda > 0 : \mu(|f|>\lambda) = 0\} \quad (f \in M) \quad .$$

D'après le théorème de Hahn-Banach, il existe un hyperplan fermé qui ne rencontre pas C. Il existe donc $\Psi \in M$, Ψ de norme 1 vérifiant $\forall c \in C \int c\, \Psi\, d\mu > 0$. C'est-à-dire que :

$$\forall P \in \mathcal{C}(H_o)\ \forall Q \in \mathcal{C}(H_1)\ \forall g \in b(0,2\varepsilon) \quad \int \frac{dP}{d\mu}\Psi\, d\mu - \int \frac{dQ}{d\mu}\Psi\, d\mu > \int g\, \Psi\, d\mu \ . \tag{26}$$

Si l'on pose $\phi = \frac{1}{2}(1+\Psi)$, ϕ est un test défini sur $(\mathcal{X},\mathcal{B})$ et, d'après (26), $\int (\frac{dP}{d\mu} - \frac{dQ}{d\mu})\, \phi\, d\mu > \frac{1}{2}\int g\, \Psi\, d\mu$. On en déduit que

$$\inf_{(P,Q)\in C(H_o)\times C(H_1)} \int \left(\frac{dP}{d\mu} - \frac{dQ}{d\mu}\right) \varphi \, d\mu \geq \frac{1}{2} \sup_{g\in b(0,2\varepsilon)} \int g \Psi \, d\mu = \frac{1}{2}.2\varepsilon.||\Psi||_\infty = \varepsilon ,$$

comme nous voulions démontrer. ■

Remarque III.8.- Dans le théorème précédent, l'hypothèse de domination des lois est indispensable. Par exemple, si $(X,B) = (\mathbb{R}, B_{\mathbb{R}})$, $H_o = \{U_{[0,1]}\}$ et $H_1 = \{\delta_x : x \in [0,1]\}$, alors $d_V(C(H_o), C(H_1)) = 1$, mais il est évident qu'il n'existe aucun test φ vérifiant $\sup_{P\in H_1} \int \varphi \, dP + 1 \leq \int_{[0,1]} \varphi(x) \, dx$.

A partir du théorème III.6., on peut énoncer une condition nécessaire et suffisante de séparation asymptotique uniforme de deux hypothèses multiples.

Dans la proposition suivante et sa démonstration on utilisera librement les notations introduites dans la sous-section III.3.5.

Proposition III.2.- Soient (X_n,B_n) $(n\geq 1)$ des espaces probabilisables, H_o et H_1 deux hypothèses multiples sur $(X^{(\infty)}, B^{(\infty)})$, dominées par une mesure σ-finie μ sur $(X^{(\infty)}, B^{(\infty)})$.

$\forall n \geq 1$ notons $C_o^{(n)}$ (resp. $C_1^{(n)}$) l'enveloppe convexe de la famille de lois $H_o^{(n)} = (\bar{P}^{(n)} ; P \in H_o)$ (resp. de $H_1^{(n)}$: $(\bar{Q}^{(n)} ; Q \in H_1)$).

Une condition nécessaire et suffisante pour que H_o et H_1 se séparent asymptotiquement uniformément est que $d_V(C_o^{(n)}, C_1^{(n)}) \xrightarrow[n\to\infty]{} 1$.

Preuve : Vu que $H_o^{(n)}$ et $H_1^{(n)}$ sont dominées par $\bar{\mu}^{(n)} (n\geq 1)$, il résulte du théorème III.6. et de la remarque I.4. que la condition est suffisante.

D'autre part, si H_o et H_1 se séparent asymptotiquement uniformément $\exists (C_n \; ; \; n \geq 1)$ $\bar{\mathcal{B}}^{(n)}$-adaptée telle que

$$\inf_{(P,Q)\in H_o \times H_1} (\bar{P}^{(n)}(C_n) - \bar{Q}^{(n)}(C_n)) \xrightarrow[n\to\infty]{} 1 \quad . \tag{27}$$

Désignons par $\mathcal{C}(H_o)$ et $\mathcal{C}(H_1)$, respectivement, les enveloppes convexes de H_o et H_1 . Il résulte de (27) que

$$\inf_{(P,Q)\in \mathcal{C}(H_o)\times \mathcal{C}(H_1)} (\bar{P}^{(n)}(C_n) - \bar{Q}^{(n)}(C_n)) \xrightarrow[n\to\infty]{} 1 \quad ,$$

d'où $d_V(C_o^{(n)}, C_1^{(n)}) \xrightarrow[n\to\infty]{} 1$. Donc, la condition est aussi nécessaire . ∎

Soulignons que la condition de la proposition précédente n'est pas facile à mettre en oeuvre, vu qu'elle implique l'évaluation de $d_V(C_o^{(n)}, C_1^{(n)})$.

On peut présenter d'autres conditions qui assurent la séparation asymptotique uniforme de deux hypothèses infinies, mais ces conditions sont très contraignantes. Par exemple :

<u>*Proposition III.3.*</u> ([35]).- Soient H_o et H_1 deux hypothèses d'un modèle statistique $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, \mathcal{H})$. On suppose que

$\forall n \geq 1 \; \exists k_n \in \mathbb{N}^* \; \exists H'_{n,1}, \ldots, H'_{n,k_n} \subset H_o \; \exists H''_{n,1}, \ldots, H''_{n,k_n} \subset H_1$

$\exists \varepsilon_n > 0$ et $\forall i, j \in \{1,\ldots,k_n\} \; \exists \; \phi_{i,j}^{(n)}$ test défini sur $(\mathcal{X}^{(n)}, \mathcal{B}^{(n)})$

tels que :

- $H_o = \bigcup_{i=1}^{k_n} H'_{n,i}$.

- $H_1 = \bigcup_{i=1}^{k_n} H''_{n,i}$.

- $\forall P \in H'_{n,i}$ $\quad \int \psi_{i,j}^{(n)} \, dP^{(n)} > 1 - \varepsilon_n$.

- $\forall Q \in H''_{n,j}$ $\quad \int \psi_{i,j}^{(n)} \, dQ^{(n)} < \varepsilon_n$.

Alors, si $\varepsilon_n k_n \xrightarrow[n\to\infty]{} 0$, H_o et H_1 se séparent asymptotiquement uniformément.

<u>*Preuve*</u> : Posons $\psi_n = \sup_i (\inf_j \psi_{i,j}^{(n)})$. Si $P \in H'_{n,i_o}$, alors

$$\int (1-\psi_n) \, dP^{(n)} \leq \int \sup_j (1-\psi_{i_o,j}^{(n)}) \, dP^{(n)} \leq \sum_j \int (1-\psi_{i_o,j}^{(n)}) \, dP^{(n)} \leq k_n \, \varepsilon_n \xrightarrow[n\to\infty]{} 0 \, .$$

D'autre part, si $Q \in H''_{n,j_o}$, alors

$$\int \psi_n \, dQ^{(n)} \leq \sum_i \int \inf_j \psi_{i,j}^{(n)} \, dQ^{(n)} \leq \sum_i \int \psi_{i,j_o}^{(n)} \, dQ^{(n)} \leq k_n \, \varepsilon_n \xrightarrow[n\to\infty]{} 0 \, .$$

La remarque I.4. nous permet de conclure. ■

III.6. - <u>LE THEOREME FONDAMENTAL DE LA DECANTATION</u>.-

On présentera ici une version améliorée d'un résultat de J. Geffroy ([23]).

Commençons par rappeler une inégalité de W. Hoeffding.

<u>*Lemme III.12.*</u> ([30]).- Soit X une variable aléatoire réelle presque-sûrement à valeurs dans un intervalle $[a,b]$, $a \leq b$. Alors :

$$E(e^{X-E(X)}) \leq e^{(b-a)^2/8} \, . \tag{28}$$

On peut énoncer un résultat similaire en termes d'espérances conditionnelles.

<u>*Lemme III.13.*</u> ([15]).- Soit $(\Omega,\mathcal{A},P)$ un espace probabilisé, $\mathcal{B}$ une sous-tribu de $\mathcal{A}$ pour laquelle il existe une version régulière de P sachant $\mathcal{B}$. Soit Y une variable aléatoire réelle intégrable définie sur $(\Omega,\mathcal{A},P)$. Soit $\gamma \geq 0$.

On suppose que pour P-presque tout ω il existe un intervalle $I(\omega)$ de longueur γ tel que $P_Y^{\mathcal{B}}(\omega,I(\omega)) = 1$. Alors

$$E^{\mathcal{B}}(e^{Y-E^{\mathcal{B}}(Y)}) \leq e^{\gamma^2/8} \quad P\text{-p.-s.} \tag{29}$$

et par conséquent

$$E_P(e^{Y-E^{\mathcal{B}}(Y)}) \leq e^{\gamma^2/8} . \tag{30}$$

<u>*Preuve*</u> : Pour P -presque tout ω, d'après l'inégalité (28), $\int e^{y-\int y P_Y^{\mathcal{B}}(\omega,dy)} P_Y^{\mathcal{B}}(\omega,dy) \leq \gamma^2/8$. Or cela n'est autre que $\int e^{y-E^{\mathcal{B}}(Y)(\omega)} P_Y^{\mathcal{B}}(\omega,dy) = e^{-E^{\mathcal{B}}(Y)(\omega)} \int e^y P_Y^{\mathcal{B}}(\omega,dy) = e^{-E^{\mathcal{B}}(Y)(\omega)} [E^{\mathcal{B}}(e^Y)](\omega) =$ $= [E^{\mathcal{B}}(e^{Y-E^{\mathcal{B}}(Y)})](\omega)$. ■

On peut aussi obtenir une inégalité de Markov pour les espérances conditionnelles, à savoir :

<u>*Lemme III.14.*</u> ([43]).- Soit $(\Omega,\mathcal{A},P)$ un espace probabilisé, $\mathcal{B}$ une sous-tribu de $\mathcal{A}$ pour laquelle $P^{\mathcal{B}}$ existe. Si X est une variable aléatoire réelle $\mathcal{B}$-mesurable et si Y est une variable aléatoire réelle, X et Y définies sur $(\Omega,\mathcal{A},P)$, pour P-presque tout ω on a

$$P^{\mathcal{B}}(\omega, \{Y \geq X\}) \leq e^{-X(\omega)} [E^{\mathcal{B}}(e^Y)](\omega) . \tag{31}$$

Preuve :

$$P^{\mathcal{B}}(\omega, \{Y \geq X\}) = P^{\mathcal{B}}(\omega, \{e^Y \geq e^X\}) = P^{\mathcal{B}}(\omega, \{e^{Y-X} \geq 1\}) . \tag{32}$$

D'après l'inégalité de Markov, pour P-presque tout ω (32) est majorée par $\int e^{(Y-X)(\omega')} P^{\mathcal{B}}(\omega, d\omega') = [E^{\mathcal{B}}(e^{Y-X})](\omega) = e^{-X(\omega)} [E^{\mathcal{B}}(e^Y)](\omega)$. ∎

Proposition III.4. ([43]).- Soit (Ω, A) un espace probabilisable, P et Q deux probabilités sur cet espace. Soit $(\mathcal{B}_n ; n \geq 0)$, avec $\mathcal{B}_o = \{\emptyset, \Omega\}$, une suite croissante de sous-tribus de A, $(a_n ; n \geq 1)$ une suite réelle, $(\varphi_n ; n \geq 1)$ et $(p_n ; n \geq 1)$ deux suites de variables aléatoires réelles, telles que $\forall n \geq 1$:

- il existe une version régulière $P_{\mathcal{B}_{n+1}}^{\mathcal{B}_n}$ (resp. $Q_{\mathcal{B}_{n+1}}^{\mathcal{B}_n}$) de $P_{\mathcal{B}_{n+1}}$ (resp. de $Q_{\mathcal{B}_{n+1}}$) conditionnée par $\mathcal{B}_n$;
- $a_n \in [0,1]$, φ_n et p_n sont à valeurs dans $[0,1]$, $a_n + p_n \leq 1$;
- φ_n est $\mathcal{B}_n$-mesurable ;
- p_n est $\mathcal{B}_{n-1}$-mesurable ;
- $E_P^{\mathcal{B}_{n-1}}(\varphi_n) \leq p_n$ P-p.-s. ;
- $E_Q^{\mathcal{B}_{n-1}}(\varphi_n) \geq a_n + p_n$ Q-p.-s. .

Soit $Y_n = \sum_{j=1}^{n} a_j \varphi_j$ $(n \geq 1)$. Alors, $\forall n \geq 1$

$$P_{\mathcal{B}_n}\{Y_n \geq \sum_{j=1}^{n} a_j(p_j + \frac{1}{2} a_j)\} \leq \exp\{-\frac{3}{8} \sum_{j=1}^{n} a_j^2\} , \tag{33}$$

$$Q_{\mathcal{B}_n}\{Y_n \geq \sum_{j=1}^{n} a_j(p_j + \frac{1}{2} a_j)\} \geq 1 - \exp\{-\frac{3}{8} \sum_{j=1}^{n} a_j^2\} . \tag{34}$$

Preuve : Remarquons tout d'abord que (34) est une conséquence de (33), que l'on établira de suite.

En effet, si $\psi'_n = 1 - \psi_n$, $p'_n = 1 - p_n - a_n$, $Y'_n = \sum_{j=1}^{n} a_j \psi'_j$,

pour $n \geq 1$ $E_Q^{\mathcal{B}_{n-1}}(\psi'_n) = 1 - E_Q^{\mathcal{B}_{n-1}}(\psi_n) \leq 1 - p_n - a_n = p'_n$. D'après (33),

$$Q_{\mathcal{B}_n}\{Y'_n \geq \sum_{j=1}^{n} a_j\ (p'_j + \frac{1}{2}\, a_j)\} \leq \exp - \frac{3}{8} \sum_{j=1}^{n} a_j^2\} .$$

Or $Y'_n \geq \sum_{j=1}^{n} a_j\ (p'_j + \frac{1}{2}\, a_j) \Longleftrightarrow Y_n \leq \sum_{j=1}^{n} a_j (p_j + \frac{1}{2}\, a_j)$. Par passage au complémentaire,

$$Q_{\mathcal{B}_n}\{Y_n \geq \sum_{j=1}^{n} a_j (p_j + \frac{1}{2}\, a_j)\} \geq 1 - \exp\{- \frac{3}{8} \sum_{j=1}^{n} a_j^2\} .$$

Voyons maintenant comment établir l'inégalité (33).

D'après l'inégalité de Markov conditionnelle (31),

$$\forall n \geq 1 \quad P^{\mathcal{B}_{n-1}}(.,\{Y_n \geq \sum_{j=1}^{n} a_j (p_j + \frac{1}{2}\, a_j)\}) \leq \exp\{- \sum_{j=1}^{n} (a_j\, p_j + \frac{1}{2}\, a_j^2)\}\ E_P^{\mathcal{B}_{n-1}}(e^{Y_n})$$

P-p.-s.

En intégrant, on obtient :

$$P\{Y_n \geq \sum_{j=1}^{n} a_j (p_j + \frac{1}{2}\, a_j)\} \leq E_P(e^{- \sum_{j=1}^{n} (a_j p_j + \frac{1}{2} a_j^2)}\ E_P^{\mathcal{B}_{n-1}}(e^{Y_n})) . \tag{35}$$

Or $$E_P^{\mathcal{B}_{n-1}}(e^{Y_n}) = e^{Y_{n-1}}\ E_P^{\mathcal{B}_{n-1}}(e^{a_n \psi_n}) \quad (n>1).$$

D'après l'inégalité de Hoeffding conditionnelle (29),

$$E_P^{\mathcal{B}_{n-1}}(e^{a_n \psi_n - E_P^{\mathcal{B}_{n-1}}(a_n \psi_n)}) \leq e^{a_n^2/8} \quad \text{P-p.-s.} \quad , \quad \text{et}$$

ceci implique $$E_P^{\mathcal{B}_{n-1}}(e^{a_n\psi_n}) \underset{P\text{-p.-s.}}{\leqslant} e^{a_n^2/8+a_n E_P^{\mathcal{B}_{n-1}}(\psi_n)} \underset{P\text{-p.-s.}}{\leqslant} e^{a_n^2/8+a_np_n} .$$

En reportant dans (35), pour $n > 1$

$$P\{Y_n \geqslant \sum_{j=1}^{n} a_j(p_j+\frac{1}{2}a_j)\} \leqslant E_P(\exp\{-\sum_{j=1}^{n}(a_jp_j+\frac{1}{2}a_j^2) + a_np_n+\frac{a_n^2}{8}+Y_{n-1}\}) =$$

$$= e^{-3/8a_n^2} E_P(\exp\{-\sum_{j=1}^{n-1} a_j(p_j+\frac{1}{2}a_j)\} E_P^{\mathcal{B}_{n-2}}(e^{Y_{n-1}})) .$$

En particulier pour le dernier terme on a :

$$E_P^{\mathcal{B}_o}(e^{a_1\psi_1-E^{\mathcal{B}_o}(a_1\psi_1)}) = E_P(e^{a_1\psi_1-E(a_1\psi_1)}) \leqslant e^{a_1^2/8} ,$$

d'où $$E_P[e^{-a_1p_1-\frac{a_1^2}{2}}e^{Y_1}] \leqslant e^{-\frac{3}{8}a_1^2} . \blacksquare$$

Corollaire 1.- Dans les conditions de la proposition précédente, $\sum_{j\geqslant 1} a_j^2 = \infty$ est une condition suffisante pour que $P \perp Q$.

Preuve : La conclusion résulte du fait que

$$d_V(P^{(n)},Q^{(n)}) \geqslant 1-2\exp\{-\frac{3}{8}\sum_{j=1}^{n} a_j^2\} . \blacksquare$$

Théorème III.7. (Théorème fondamental de la décantation) ([43]).-

Soit $X = (X_n ; n \geqslant 1)$ une suite d'observations à valeurs dans $\mathbb{R}^s$, de loi inconnue.

Pour $n \geqslant 2$, soit $a_n \in [0,1]$, $p_n : \mathbb{R}^{s(n-1)} \longrightarrow [0,1]$ mesurable vérifiant $\forall x^{(n-1)} \in \mathbb{R}^{s(n-1)}$ $p_n(x^{(n-1)}) + a_n \leqslant 1$, $\psi_n : \mathbb{R}^{sn} \longrightarrow [0,1]$ mesurable.

$\forall n \geq 2$ posons $C_n = \{\sum_{j=2}^{n} a_j \varphi_j \geq \sum_{j=2}^{n} a_j (p_j + \frac{1}{2} a_j)\}$.

Pour tout $n \geq 2$ et toute hypothèse simple P sur $(\mathbb{R}^s, \mathcal{B}_{\mathbb{R}^s})^{(\infty)}$ on suppose choisie une version régulière $P_{X_n}^{X^{(n-1)}}$ de la loi de X_n sachant $X^{(n-1)}$ sous l'hypothèse P.

Alors, pour toute hypothèse P vérifiant

$$\forall n \geq 2 \quad \forall x^{(n-1)} \in \mathbb{R}^{s(n-1)} \quad \int \varphi_n(x^{(n-1)},x)\, P_{X_n}^{X^{(n-1)}=x^{(n-1)}}(dx) \leq p_n(x^{(n-1)}) \qquad (36)$$

on a
$$P_{X^{(n)}}(C_n) \leq \exp\{-\,3/8 \sum_{j=2}^{n} a_j^2\} \qquad (n\geq 2)\ ,$$

et pour toute hypothèse Q vérifiant

$$\forall n \geq 2 \quad \forall x^{(n-1)} \in \mathbb{R}^{s(n-1)} \quad \int \varphi(x^{(n-1)},x)\, Q^{X^{(n-1)}=x^{(n-1)}}(dx) \geq p_n(x^{(n-1)})+a_n \qquad (37)$$

on a

$$Q_{X^{(n)}}(C_n) \geq 1 - \exp\{-\,3/8 \sum_{j=2}^{n} a_j^2\} \qquad (n\geq 2)\ .$$

Si P vérifie (36) et Q vérifie (37), $\sum_{j\geq 2} a_j^2 = \infty$ est une condition suffisante pour que $P \perp Q$.

Preuve : Le résultat se déduit de la proposition III.4., en considérant les X_n des variables aléatoires définies sur un espace probabilisable $(\Omega,\mathcal{A})$ et en prenant $\mathcal{B}_o = \{\emptyset,\Omega\}$, $\mathcal{B}_n = \sigma(X^{(n)})$ $(n\geq 1)$, $\xi_1 = \Psi_1 = a_1 = 0$, $\xi_n = \varphi_n \circ X^{(n)}$, $\Psi_n = p_n \circ X^{(n-1)}$ $(n\geq 2)$.

Soit Pr la loi sur $\sigma(X)$ dont P est la loi image. D'après le théorème de Jiřina, il existe une version régulière de $Pr_{\mathcal{B}_{n+1}}$ conditionnée par $\mathcal{B}_n$ $(n\geq 1)$. Puisque $P_{X^{(n)}}(C_n) = Pr_{\mathcal{B}_n}\{\sum_{j=1}^{n} a_j \xi_j \geq \sum_{j=1}^{n} a_j(\Psi_j + \frac{1}{2} a_j)\}$

il nous suffit de vérifier que $\forall n \geq 2 \quad E_{Pr}^{\mathcal{B}_{n-1}}(\xi_n) \leq \Psi_n$ Pr-p.-s.

Or il est connu que $f \in E_{Pr}^{\mathcal{B}_{n-1}}(\xi_n) \Longleftrightarrow \exists g \in E[\xi_n \mid X^{(n-1)}]$ tel que $f = g \circ X^{(n-1)}$, d'où que l'on puisse se ramener à vérifier que $E[\xi_n \mid X^{(n-1)}] \leq p_n \quad P_{X^{(n-1)}}$-p.-s. .

ψ_n étant intégrable par rapport à $P_{(X^{(n-1)},X_n)}$, d'après le théorème de Fubini généralisé $x \longmapsto \psi_n(x^{(n-1)},x)$ est $P_{X_n}^{X^{(n-1)}=x^{(n-1)}}$-intégrable et $\forall A \in \mathcal{B}_{\mathbb{R}^{s(n-1)}}$.

$$\int_A \left(\int \psi(x^{(n-1)},x)\, P_{X_n}^{X^{(n-1)}=x^{(n-1)}}(dx)\right) P_{X^{(n-1)}}(dx^{(n-1)}) =$$

$$= \int_{A\times\mathbb{R}^s} \psi_n \, dP_{X^{(n)}} = \int_{[X^{(n)}]^{-1}(A\times\mathbb{R}^s)} \xi_n \, dPr \quad .$$

Donc $x^{(n-1)} \longmapsto \int \psi(x^{(n-1)},x)\, P_{X_n}^{X^{(n-1)}=x^{(n-1)}}(dx) \in E[\xi_n \mid X^{(n-1)}]$,

ce qui nous permet de conclure pour ce qui concerne P vérifiant (36).

La deuxième partie résulte de la première, en faisant des modifications pareilles à celles introduites dans la preuve de la proposition III.4., et la partie finale se déduit du corollaire 1 de cette proposition.

Remarque III.9.- On aurait pu, naturellement, introduire des conditions portant sur les lois initiales. Par exemple (cf. [41])

$\exists a_1, p_1 \in [0,1]$ tels que $a_1 + p_1 \leq 1 \;\exists\, \psi_1 : \mathbb{R}^s \longrightarrow [0,1]$ mesurable pour lesquels

$$\int \psi_1 \, dP_{X^{(1)}} \leq p_1 \qquad \text{(à ajouter à (36))} ,$$

$$\int \psi_1 \, dQ_{X^{(1)}} \geq p_1 + a_1 \qquad \text{(à ajouter à (37))} .$$

On remplacerait alors, dans l'énoncé du théorème III.7., $\sum_{j=2}^{n}$ par $\sum_{j=1}^{n}$.

On peut se demander pourquoi le théorème III.7. est qualifié de fondamental. Il y a des raisons qui justifient ce qualificatif.

Un des avantages qu'il présente est que la vitesse de séparation des lois P et Q est explicitement minorée par $2 \exp \{- 3/8 \sum_{j=2}^{n} a_j^2\}$.

Par ailleurs, les bases de cylindre C_n qui séparent P et Q sont aussi exhibées.

Mais, ce qui est essentiel en statistique, le théorème III.7 ne s'applique pas seulement à des hypothèses simples, mais aussi à des hypothèses multiples. En effet, si H_o et H_1 sont deux hypothèses quelconques, si on se donne des versions fixes $P_{X_n}^{X^{(n-1)}}$ et $Q_{X_n}^{X^{(n-1)}}$ des lois P dans H_o et Q dans H_1 qui vérifient, respectivement, (36) et (37), alors

$$\sup (P^{(n)}(C_n) \; ; \; P \in H_o) \leq \exp \{- 3/8 \sum_{j=2}^{n} a_j^2\} \ , \tag{38}$$

$$\inf (Q^{(n)}(C_n) \; ; \; Q \in H_1) \geq 1 - \exp \{- 3/8 \sum_{j=2}^{n} a_j^2\} \ , \tag{39}$$

et $\sum_{j \geq 2} a_j^2 = \infty \Longrightarrow H_o$ et H_1 se séparent asymptotiquement uniformément ([43]).

CHAPITRE IV

PROPRIETES DE DECANTATION DE DISTANCES SUR DES ENSEMBLES DE PROBABILITES.

IV. 1. - NOTION DE DISTANCE DECANTANTE SUR UN ENSEMBLE DE PROBABILITES.-

On a vu dans le chapitre précédent qu'il y a un rapport très étroit entre la séparation des lois et la distance en variation entre ces lois. Pourtant, la distance en variation n'est pas toujours facile à évaluer. Par exemple, on ne sait pas le faire pour deux lois gaussiennes non dégénérées quelconques sur $\mathbb{R}^s$. Par contre, on peut déterminer la distance de Hellinger entre ces lois.

Définition IV.1.- Soit $(\mathcal{X},\mathcal{B})$ un espace probabilisable, $\mathcal{P}$ un ensemble de probabilités sur cet espace, d une distance définie sur $\mathcal{P}$. d est dite décantante si

$$\left.\begin{array}{l} \exists f :]0,\infty[\longrightarrow]0,1] \quad \text{telle que } \forall\alpha > 0 \quad \forall P, Q \in \mathcal{P} \\ \qquad d(P,Q) \geq \alpha \implies d_V(P,Q) \geq f(\alpha) \quad . \end{array}\right| \qquad (1)$$

Une conséquence immédiate de la définition est le

Lemme IV.1.- Si d est décantante sur $\mathcal{P}$ et f vérifie (1), alors $\forall P, Q \in \mathcal{P}$, $P \neq Q$ $\quad d_V(P,Q) \geq f(d(P,Q))$.

Puisque suivant les cas il se peut qu'une distance décantante d soit plus facile à évaluer que d_V, on s'aperçoit que la connaissance d'une fonction f vérifiant (1) est importante en statistique asymptotique.

Naturellement, on a intérêt à choisir f de façon optimale, c'est-à-dire à déterminer f_d vérifiant (1) par rapport à d (supposée décantante) telle que, si f vérifie (1) aussi, $f \leq f_d$, en supposant que ce soit possible.

Proposition IV.1.-

a) Pour que la distance d définie sur un ensemble $\mathcal{P}$ de probabilités définies sur un même espace probabilisable soit décantante, il faut et il suffit que

$$\left.\begin{array}{l} \forall\alpha > 0 : \quad \{(P,Q) \in \mathcal{P}^2 : d(P,Q) \geq \alpha\} \neq \emptyset \\ \inf(d_V(P,Q) \; ; P,Q \in \mathcal{P} , \; d(P,Q) \geq \alpha) > 0 \; . \end{array}\right| \quad (2)$$

b) Si (2) est vérifiée, posons

$$f_d :]0,\infty[\longrightarrow]0,1] : x \longmapsto \begin{cases} 1, \text{ si } \{(P,Q) \in \mathcal{P}^2 : d(P,Q) \geq x\} = \emptyset \; . \\ \\ \inf(d_V(P,Q) \; ; P,Q \in \mathcal{P}, \; d(P,Q) \geq x), \\ \qquad \text{sinon} \; . \end{cases} \quad (3)$$

Alors f_d vérifie (1) et $\forall f$ vérifiant (1) $f_d \geq f$.

Preuve :

a) Il est évident que (2) est une condition nécessaire pour que d soit décantante. Qu'elle est aussi suffisante résulte de b).

b) Si (2) est vérifiée, f_d est en effet à valeurs dans $]0,1]$.

Sur $\{\alpha > 0 : \{(P,Q) \in \mathcal{P}^2 : d(P,Q) \geq \alpha\} = \emptyset\}$, (1) est trivialement vérifiée pour toute fonction f à valeurs dans $]0,1]$ et $f_d \geq f$ sur cet ensemble. Sur $\{\alpha > 0 : \{(P,Q) \in \mathcal{P}^2 : d(P,Q) \geq \alpha\} \neq \emptyset\}$, $d_V(P,Q) \geq f_d(\alpha)$. Donc f_d vérifie (1). D'autre part, $\forall f$ vérifiant (1) sur cet ensemble

$\forall P, Q \in \mathcal{P} : d(P,Q) \geq \alpha \quad d_V(P,Q) \geq f(\alpha) \Longrightarrow \inf\limits_{P,Q\in\mathcal{P}, d(P,Q)\geq\alpha} d_V(P,Q) = f_d(\alpha) \geq f(\alpha)$. ∎

Corollaire 1.- Soit $\mathcal{P}$ un ensemble de probabilités sur un espace probabilisable $(\mathcal{X},\mathcal{B})$, d une distance sur $\mathcal{P}$. Pour que d ne soit pas décantante il faut et il suffit que

$$\left.\begin{array}{l} \exists \alpha > 0 \quad \exists (P_n\ ;\ n \geq 1),\ (Q_n\ ;\ n \geq 1) \text{ suites d'éléments de } \mathcal{P} \\ \text{pour lesquelles} \\ \forall n \geq 1 \quad d(P_n,Q_n) \geq \alpha\ ,\quad \lim\limits_{n\to\infty} d_V(P_n,Q_n) = 0\ . \end{array}\right| \qquad (4)$$

Corollaire 2.- Si, dans les conditions du corollaire 1, $\mathcal{P}$ est fini, toutes les distances définies sur $\mathcal{P}$ sont décantantes.

Nous allons maintenant essayer de déterminer f_d , qui est la meilleure fonction de décantation possible, pour un certain nombre de distances usuelles d sur un ensemble donné $\mathcal{P}$ de lois sur $(\mathcal{X},\mathcal{B})$. Sauf mention expresse du contraire, $\mathcal{P}$ désignera toujours l'ensemble de toutes les lois de probabilité sur l'espace probabilisable considéré $(\mathcal{X},\mathcal{B})$.

Remarque IV.1.- Soit $(\mathcal{X},\mathcal{B})$ un espace probabilisable, $\mathcal{P}$ un ensemble de probabilités sur $\mathcal{B}$ contenant deux suites de lois $(P_n\ ;\ n \geq 1)$ et $(Q_n\ ;\ n \geq 1)$ telles que $\forall n \in \mathbb{N}^* \quad P_n \neq Q_n$, $d_V(P_n,Q_n) \xrightarrow[n\to\infty]{} 0$. Alors, la distance discrète $d : (P,Q) \longmapsto \mathbb{1}_{\{P\}}(Q)$ n'est pas décantante.

On verra dans la suite que la plupart des distances couramment employées sont décantantes.

IV.2. - LA DISTANCE DE HELLINGER.-

Elle a déjà été introduite et on sait que si $(X,\mathcal{B})$ est un espace probabilisable, $\mathcal{P}$ l'ensemble des probabilités sur $\mathcal{B}$ $\forall P, Q \in \mathcal{P}$ $d_V(P,Q) \geq d_H^2(P,Q)$. Cette inégalité n'est pas améliorable en toute généralité, comme il s'ensuit de l'exemple ci-dessous.

Exemple IV.1.- Soit (X,d) un espace métrique, $\mathcal{B}$ la tribu borélienne de X. Désignons par $b(x,r)$ la boule ouverte de centre $x \in X$ et rayon $r \in \mathbb{R}_+$.

Soit $(X,\mathcal{B})$ muni d'une mesure σ-finie μ.

Supposons que X possède deux points x et y tels que $d(x,y) > 1$ et que, avec $\nu \in \mathbb{N}^*$, $\forall \alpha \in [0,1]$ $\mu(b(x,\frac{\alpha}{2})) = \alpha^\nu$, $\mu(b(y,\frac{\alpha}{2})) = \alpha^\nu$.

Soit P la probabilité sur $\mathcal{B}$ de densité $1\!\!1_{b(x,\frac{1}{2})}$ par rapport à μ, Q celle de densité $1\!\!1_{b(x,\frac{1-\alpha}{2})} + 1\!\!1_{b(y,\frac{\sqrt[\nu]{1-(1-\alpha)^\nu}}{2})}$ par rapport à μ, où $\alpha \in]0,1[$.

On a $d_V(P,Q) = d_H^2(P,Q) = \frac{1}{2}\left[\mu(b(x,\frac{1}{2}) - b(x,\frac{1-\alpha}{2})) + \right.$

$$\left. + \mu(b(y, \frac{\sqrt[\nu]{1-(1-\alpha)^\nu}}{2}))\right] = 1-(1-\alpha)^\nu.$$

En particulier, si l'on munit $\mathbb{R}^\nu$ de $d : ((x_1,\dots,x_\nu), (y_1,\dots,y_\nu)) \longmapsto \max\limits_{1\leq i\leq \nu} |x_i-y_i|$ et λ^ν est la mesure de

Lebesgue sur $\mathcal{B}_{\mathbb{R}^\nu}$, on voit que les conditions de l'exemple IV.1. sont satisfaites pour $(X,\mathcal{B}) = (\mathbb{R}^\nu, \mathcal{B}_{\mathbb{R}^\nu})$, $\mu = \lambda^\nu$. Donc :

Proposition IV.2.-

i) d_H est décantante .

ii) Sur l'ensemble $\mathcal{P}$ de toutes les lois boréliennes sur $(\mathbb{R}^\nu, \mathcal{B}_{\mathbb{R}^\nu})$ on a :

$$f_{d_H} :]0,\infty[\longrightarrow]0,1] : x \longmapsto \min(x^2,1) \ . \tag{5}$$

IV.3. - LA DISTANCE DE PROKHOROV.-

Définition IV.2.- Soit (X,d) un espace métrique, $\mathcal{B}$ la tribu borélienne de X, $\mathcal{P}$ l'ensemble des probabilités sur $\mathcal{B}$. La distance de Prokhorov, d_P , est définie par

$$\forall P\ ,\ Q \in \mathcal{P} \qquad d_P(P,Q) = \inf\{\varepsilon > 0 : \forall A \in \mathcal{B} \ \ P(A) \leq Q(A^\varepsilon_\bullet) + \varepsilon\}\ , \tag{6}$$

où $A^\varepsilon_\bullet$ est le ε-dilaté ouvert de A, c'est-à-dire $A^\varepsilon_\bullet = \{x \in X : d(x,A) < \varepsilon\}$, si $A \neq \emptyset$, $\emptyset^\varepsilon_\bullet = \emptyset$.

Remarque IV.2.- On aurait pu considérer des ε-dilatés fermés, et se restreindre à des événements ouverts ou à des événements fermés. Dans tous les cas, $(P(A) \leq Q(A^\varepsilon_\bullet) + \varepsilon)$ pourrait être remplacée par $(P(A) \leq Q(A^\varepsilon_\bullet) + \varepsilon$ et $Q(A) \leq P(A^\varepsilon_\bullet) + \varepsilon)$ (cf. [19]).

Si (X,d) est séparable, d_P définit la topologie de la convergence faible sur $\mathcal{P}$.

Il est connu que $\{\varepsilon > 0 : \forall A \in \mathcal{B} \quad P(A) \leqslant Q(A^{\varepsilon}_{\bullet}) + \varepsilon\}$ est un intervalle qui contient $[1,\infty[$ et que d_P prends ses valeurs dans $[0,1]$, 1 étant atteinte si, par exemple, avec $\nu \in \mathbb{N}^*$, $\mathbb{R}^{\nu}$ est muni de $d : (x,y) \longmapsto \max\limits_{1 \leqslant i \leqslant \nu} |x_i - y_i|$ $(x, y \in \mathbb{R}^{\nu})$, $\mathcal{X} = \mathbb{R}^{\nu}$, $P = \delta_{\{(0,\ldots,0)\}}$ et $Q = \delta_{\{(2,0,\ldots,0)\}}$, où $(0,\ldots,0)$, $(2,0,\ldots,0) \in \mathbb{R}^{\nu}$.

Contrairement aux distances d_V et d_H, pour lesquelles $d_V(P,Q) = 1 \Longleftrightarrow d_H(P,Q) = 1 \Longleftrightarrow P \perp Q$, d_P peut prendre la valeur 1 maix $P \perp Q \not\Longrightarrow d_P(P,Q) = 1$. En effet, dans les conditions mentionnées ci-dessus, soit $(\mathcal{X},d) = (\mathbb{R}^{\nu},d)$, $P = \delta_{\{(0,\ldots,0)\}}$ (où $(0,\ldots,0) \in \mathbb{R}^{\nu}$), $Q = \mathcal{U}_{[0,1]^{\nu}}$, $A \in \mathcal{B}_{\mathbb{R}^{\nu}}$. On a :

- $(0,\ldots,0) \notin A \Longrightarrow P(A) = 0 < Q(A^{\varepsilon}_{\bullet}) + \varepsilon \quad \forall \varepsilon > 0$;
- $(0,\ldots,0) \in A \Longrightarrow P(A) = 1 \leqslant Q(A^{\varepsilon}_{\bullet}) + \varepsilon \quad$ si $\varepsilon \geqslant 1 - \dfrac{1}{2^{\nu}}$.

Donc, $d_P(P,Q) \leqslant 1 - \dfrac{1}{2^{\nu}}$ et $P \perp Q \not\Longrightarrow d_P(P,Q) = 1$.

Soient $P, Q \in \mathcal{P}$, $\beta = d_V(P,Q) = \sup\limits_{B \in \mathcal{B}} (P(B) - Q(B))$. Alors $\forall B \in \mathcal{B} \quad P(B) \leqslant Q(B) + \beta \leqslant Q(B^{\beta}_{\bullet}) + \beta$. Donc

$$d_V(P,Q) \geqslant d_P(P,Q). \tag{7}$$

Remarquons que (7) permet d'affirmer :

$$d_P(P,Q) = 1 \Longrightarrow P \perp Q. \tag{8}$$

Revenons à l'espace $(\mathbb{R}^{\nu},d)$. Soit $P = \mathcal{U}_{[0,1]^{\nu}}$, $Q = \dfrac{1}{2^{\nu}} \; \mathcal{U}_{[0,\frac{1}{2}]^{\nu}} + (1 - \dfrac{1}{2^{\nu}}) \; \delta_{\{(2,0,\ldots,0)\}}$, où $(2,0,\ldots,0) \in \mathbb{R}^{\nu}$.

On a $d_V(P,Q) = \frac{1}{2} \, (1 - \dfrac{1}{2^{\nu}} + \displaystyle\int_{[\frac{1}{2},1]^{\nu}} dx_1 \ldots dx_{\nu}) = \frac{1}{2}$ et, d'après (7),

$d_P(P,Q) \leq \frac{1}{2}$.

Soit $A = [0,1]^\nu$, $\gamma \in]0,\frac{1}{2}]$. Alors,

$P(A) = 1 \leq Q(A^\gamma_\bullet) + \gamma = \frac{1}{2^\nu} + \gamma \Longrightarrow \gamma \geq 1 - \frac{1}{2^\nu} \geq \frac{1}{2}$. Donc, $d_P(P,Q) = \frac{1}{2}$.

Compte tenu de (7) il s'ensuit que

Proposition IV.3.-

i) d_P est décantante .

ii) Sur l'ensemble $\mathcal{P}$ de toutes les lois boréliennes sur $(\mathbb{R}^\nu, \mathcal{B}_{\mathbb{R}^\nu})$ on a :

$$f_{d_P} :]0,\infty[\longrightarrow]0,1] : x \longmapsto \min(x,1) . \tag{9}$$

IV.4. - LA DISTANCE DE KOLMOGOROV.-

Définition IV.3.- Soit $\nu \in \mathbb{N}^*$, $(X,\mathcal{B}) = (\mathbb{R}^\nu , \mathcal{B}_{\mathbb{R}^\nu})$, $\mathcal{P}$ l'ensemble des lois sur $(X,\mathcal{B})$, $P, Q \in \mathcal{P}$, F et G les fonctions de répartition de P et Q , respectivement. La distance de Kolmogorov entre P et Q , notée $d_K(P,Q)$, est définie par

$$d_K(P,Q) = \sup_{x\in\mathbb{R}^\nu} |F(x) - G(x)| \quad . \tag{10}$$

d_K est la distance qui intervient dans le théorème de Glivenko-Cantelli. Elle prend ses valeurs dans $[0,1]$, et il est aussi évident que $d_K(P,Q) = 1 \Longrightarrow P \perp Q$. Pourtant, il suffit de considérer, sur $\mathcal{B}_{\mathbb{R}^\nu}$, $P = N(0,1)^{\otimes\nu}$ et $Q = \delta_{\{(0,\ldots,0)\}}$ (où $(0,\ldots,0) \in \mathbb{R}^\nu$) pour voir que $P \perp Q \not\Longrightarrow d_K(P,Q) = 1$, vu que dans ce cas $d_K(P,Q) = 1/2$.

Puisqu'il ne s'agit ici que de la valeur absolue des différences des probabilités d'événements de type particulier, à savoir $]-\infty,x_1] \times \ldots \times]-\infty, x_\nu]$ $(x=(x_1,\ldots,x_\nu) \in \mathbb{R}^\nu)$, et que $d_V(P,Q) = \sup (|P(A) - Q(A)| \; ; A \in \mathcal{B})$, on a

$$d_K(P,Q) \leq d_V(P,Q) \; . \tag{11}$$

Calcul de f_{d_K}. Considérons l'exemple suivant :

Exemple IV.2.- Soit $\nu \in \mathbb{N}^*$, $\alpha \in]0,1]$, $P = \delta_{\{(0,\ldots,0)\}}$, $Q_\alpha = (1-\alpha)\,\delta_{\{(0,\ldots,0)\}} + \alpha\,\delta_{\{(1,0,\ldots,0)\}}$, où $(0,\ldots,0)$, $(1,0,\ldots,0) \in \mathbb{R}^\nu$.

On a $d_V(P,Q_\alpha) = \frac{1}{2}(\alpha+\alpha) = \alpha = d_K(P,Q_\alpha)$. Compte tenu de (11),

Proposition IV.4.- d_K est décantante sur l'ensemble des lois boréliennes sur $(\mathbb{R}^\nu, \mathcal{B}_{\mathbb{R}^\nu})$ et

$$f_{d_K} :]0,\infty[\longrightarrow]0,1] \; : \; x \longmapsto \min(x,1). \tag{12}$$

IV.5. - LA DISTANCE DE LEVY.-

Définition IV.4.- Soit $\nu \in \mathbb{N}^*$, $(X,\mathcal{B}) = (\mathbb{R}^\nu, \mathcal{B}_{\mathbb{R}^\nu})$, $\mathcal{P}$ l'ensemble des probabilités sur $\mathcal{B}$. La distance de Lévy, d_L, est définie par

$$\forall P, Q \in \mathcal{P} \quad d_L(P,Q) = \inf\{h > 0 : \forall x \in \mathbb{R}^\nu \;\; G(x-\dot{h}) - h \leq F(x) \leq G(x+\dot{h}) + h\}, \tag{13}$$

où $\dot{h} = (h,\ldots,h) \in \mathbb{R}^\nu$ et F et G désignent respectivement les fonctions de répartition de P et Q.

d_L définit, comme d_P, la topologie de la convergence faible sur $\mathcal{P}$.

On voit qu'il s'agit d'un infimum pris sur un ensemble qui est un intervalle contenant $[1,\infty[$ et que d_L prend ses valeurs dans $[0,1]$. La valeur 1 est atteinte si, par exemple, $P = \delta_{\{(0,\dots,0)\}}$ et $Q = \delta_{\{(2,0,\dots,0)\}}$, où $(0,\dots,0), (2,0,\dots,0) \in \mathbb{R}^\nu$.

$P \perp Q \not\Longrightarrow d_L(P,Q) = 1$. Pour le vérifier on n'a qu'à considérer, sur $\mathcal{B}_{\mathbb{R}^\nu}$, $P = \delta_{\{(0,\dots,0)\}}$ et $Q = N(0,1)^{\otimes\nu}$. On a $d_L(P,Q) \leq \frac{1}{2}$, vu que $\forall x = (x_1,\dots,x_\nu) \in \mathbb{R}^\nu$, si $\exists i \in \{1,\dots,\nu\}$ tel que $x_i < 0$,

$$G(x - (\tfrac{1}{2},\dots,\tfrac{1}{2})) - \tfrac{1}{2} < 0 = F(x) < G(x + (\tfrac{1}{2},\dots,\tfrac{1}{2})) + \tfrac{1}{2}$$

et que, si $x_i \geq 0$ $(1 \leq i \leq \nu)$ $G(x - (\frac{1}{2},\dots,\frac{1}{2})) - \frac{1}{2} < 1 = F(x) <$ $< G(x + (\frac{1}{2},\dots,\frac{1}{2})) + \frac{1}{2}$.

On sait que $\forall x \in \mathbb{R}^\nu$, si $\beta = d_V(P,Q)$ $F(x) \leq G(x) + \beta \leq G(x+\dot{\beta}) + \beta$. De même, $G(x-\dot{\beta}) \leq F(x-\dot{\beta}) + \beta \leq F(x) + \beta$, d'où

$$d_V(P,Q) \geq d_L(P,Q) . \tag{14}$$

En particulier,

$$d_L(P,Q) = 1 \Longrightarrow P \perp Q . \tag{15}$$

Calcul de f_{d_L}. Reprenons l'exemple IV.2. On sait que $d_V(P,Q_\alpha) = \alpha$; donc, $d_L(P,Q) \leq \alpha$.

D'autre part, soit $0 < h < \alpha$. Désignons par G_α la fonction de répartition de Q_α. On a $-h = G_\alpha(-h,\dots,-h) - h < 1 = F(0,\dots,0)$, mais $1 > G_\alpha(h,\dots,h) + h = 1 - \alpha + h$. Donc $d_L(P,Q_\alpha) \geq \alpha$.

Finalement, $d_L(P,Q_\alpha) = \alpha$.

Compte tenu de (14), on peut affirmer :

Proposition IV.5.- d_L est décantante sur l'ensemble des lois boréliennes sur $(\mathbb{R}^\nu, \mathcal{B}_{\mathbb{R}^\nu})$ et

$$f_{d_L} :]0,\infty[\longrightarrow]0,1] : x \longmapsto \min(x,1) . \tag{16}$$

IV.6. - LA DISTANCE ASSOCIEE A LA $||\cdot||_\infty$ ENTRE LES FONCTIONS CARACTERISTIQUES.-

Définition IV.5.- Soit $\nu \in \mathbb{N}^*$, $(X,\mathcal{B}) = (\mathbb{R}^\nu , \mathcal{B}_{\mathbb{R}^\nu})$, $\mathcal{P}$ l'ensemble des probabilités sur $\mathcal{B}$. Considérons la distance d_φ définie par

$$\forall P , Q \in \mathcal{P} \qquad d_\varphi(P,Q) = \frac{1}{2} \sup_{t\in\mathbb{R}^\nu} |\varphi_P(t) - \varphi_Q(t)| , \tag{17}$$

où φ_P et φ_Q désignent respectivement les fonctions caractéristiques de P et Q.

On voit que d_φ prend ses valeurs dans $[0,1]$: $0 \leq d_\varphi \leq 1$ et la valeur 1 est atteinte si, par exemple , $P = \delta_{\{(0,\ldots,0)\}}$ et $Q = \delta_{\{(1,0,\ldots,0)\}}$, où $(0,\ldots,0)$, $(1,0,\ldots,0) \in \mathbb{R}^\nu$. En effet,
$d_\varphi(P,Q) = \frac{1}{2} \sup_{t\in\mathbb{R}} |1-e^{it}| = \frac{1}{2} \sup_{t\in\mathbb{R}} [2(1-\cos t)]^{1/2} = 1$.

Si $P = \delta_{\{(0,\ldots,0)\}}$, où $(0,\ldots,0) \in \mathbb{R}^\nu$ et $Q = N(0,1) \otimes \delta_{\{0\}}^{\otimes\nu-1}$, alors $d_\varphi(P,Q) = \frac{1}{2} \sup_{t\in\mathbb{R}} |1-e^{-t^2/2}| = \frac{1}{2}$. Donc $P \perp Q \not\Longrightarrow d_\varphi(P,Q) = 1$.

Avec $P, Q \in \mathcal{P}$, $\mu = \frac{P+Q}{2}$, $f \in \frac{dP}{d\mu}$, $g \in \frac{dQ}{d\mu}$, on a :

$$d_\psi(P,Q) = \frac{1}{2} \sup_{t\in\mathbb{R}^\nu} \left|\int e^{i<t,x>} (f-g)(x)\ \mu(dx)\right| \leq \frac{1}{2} \int |f-g|(x)\ \mu(dx) = d_V(P,Q) . \quad (18)$$

En particulier,

$$d_\psi(P,Q) = 1 \Longrightarrow P \perp Q . \quad (19)$$

Calcul de f_{d_ψ} . Reprenons l'exemple IV.2. :

$$d_\psi(P,Q_\alpha) = \frac{1}{2} \sup_{t\in\mathbb{R}} |1-(1-\alpha)-\alpha e^{it}| = \frac{1}{2} \sup_{t\in\mathbb{R}} \{\alpha|2(1-\cos t)|^{1/2}\} = \alpha = d_V(P,Q_\alpha).$$

Rappelant (18), on peut affirmer :

Proposition IV.6.- d_ψ est décantante sur l'ensemble des lois boréliennes sur $(\mathbb{R}^\nu, \mathcal{B}_{\mathbb{R}^\nu})$ et

$$f_{d_\psi} :]0,\infty[\longrightarrow]0,1] : x \longmapsto \min(x,1) . \quad (20)$$

IV.7. - LA DISTANCE DE FORTET-MOURIER ([20]).-

On examinera dans cette section une distance utilisée en statistique asymptotique ([20] généralise le théorème de Glivenko-Cantelli à des espaces autres que la droite réelle) qui n'est décantante que dans des cas particuliers.

Soit (X,d) un espace métrique, $\mathcal{B}$ la tribu borélienne de X, $\mathcal{P}$ l'ensemble des probabilités P sur $\mathcal{B}$ vérifiant :

$$\exists y \in X : \int d(y,x)\ P(dx) < \infty . \quad (21)$$

Si $z \in X$ et (21) est vérifiée pour $y \in X$, alors
$\int d(z,x)\ P(dx) \leqslant d(z,y) + \int d(y,x)\ P(dx) < \infty$. On peut donc fixer $a \in X$
et dire que $\mathcal{P}$ est l'ensemble des lois P sur $\mathcal{B}$ telles que

$$\int d(x,a)\ P(dx) < \infty\ . \tag{22}$$

Remarque IV.3.- L'ensemble $(\delta_x\ ;\ x \in X)$ est évidemment contenu dans $\mathcal{P}$.

Soit $\mathcal{M}_a$ l'espace vectoriel défini par

$$\mathcal{M}_a = \{f : X \longrightarrow \mathbb{R} \text{ telles que } f(a) = 0\ ,\ \sup_{x,y \in X, x \neq y} \frac{|f(x) - f(y)|}{d(x,y)} < \infty\} \tag{23}$$

Sur $\mathcal{M}_a$ définissons

$$||\cdot|| \ : f \longmapsto \sup(\frac{|f(x) - f(y)|}{d(x,y)}\ ;\ x,\ y \in X,\ x \neq y). \tag{24}$$

Vu que $f \in \mathcal{M}_a \Longrightarrow f(a) = 0$, $||\cdot||$ est une norme sur $\mathcal{M}_a$.

Soit $(f_n\ ;\ n \geqslant 1)$ une suite de Cauchy dans $(\mathcal{M}_a\ ,\ ||\cdot||)$.

$$\forall x \in X \quad \forall m\ ,\ n \in \mathbb{N}^* \quad |f_n(x)-f_m(x)| = |(f_n-f_m)(x) - (f_n-f_m)(a)| \leqslant$$
$$||f_n-f_m||\ d(x,a) \xrightarrow[n,m\to\infty]{} 0\ .$$

Donc, $(f_n(x)\ ;\ n \geqslant 1)$ est une suite de Cauchy réelle. Notons $f(x)$ sa limite. On définit ainsi une fonction réelle f sur X, qui appartient à $\mathcal{M}_a$ car :

i) $f(a) = 0$;

ii) $\forall \varepsilon > 0 \quad \exists N_\varepsilon \in \mathbb{N} : n,\ m \geqslant N_\varepsilon \Longrightarrow ||f_n-f_m|| \leqslant \varepsilon$.

$\forall x,\ y \in X \quad \forall n,\ m \geqslant N_\varepsilon \quad |(f_n-f_m)(x) - (f_n-f_m)(y)| \leqslant \varepsilon d(x,y)$.

Faisant $n \to \infty$, $\forall x, y \in X$ $\forall m \geq N_\varepsilon$ $|(f-f_m)(x) - (f-f_m)(y)| \leq \varepsilon d(x,y)$.
Donc $f-f_m \in M_a$, $f \in M_a$.

D'où

<u>*Lemme IV.2.*</u>- $(M_a , ||\cdot||)$ défini comme ci-dessus est un espace de Banach.

Soit $(M_a^* , ||\cdot||^*)$ le dual fort de $(M_a , ||\cdot||)$. On peut plonger P dans l'espace de Banach M_a^* en posant

$$\forall P \in P \qquad \psi(P) : f \in M_a \longmapsto [\psi(P)](f) = \int f \, dP . \tag{25}$$

(25) est justifiée par :

a) $f \in M_a$, $x \in X$ $\Longrightarrow |f(x)| = |f(x)-f(a)| \leq ||f|| \; d(x,a) \Longrightarrow$

$$\Longrightarrow \int |f(x)| \; P(dx) \leq ||f|| \int d(x,a) \; P(dx) < \infty . \tag{26}$$

f est donc P-intégrable.

b) (26) montre aussi que la forme linéaire $\psi(P)$ est continue et que $||\psi(P)||^* \leq \int d(x,a) \; P(dx)$. En effet, si $\exists x \in X - \{a\}$, posons $g : x \longmapsto d(x,a)$. On voit que $g \in M_a$, vu que $|d(x,a)-d(y,a)| \leq d(x,y)$ $(x,y \in X)$, et aussi que $||g|| = 1$, vu que $|g(x)-g(a)| = d(x,a)$.
Donc, $||\psi(P)||^* = \int d(x,a) \; P(dx)$.

<u>*Lemme IV.3.*</u> ([40]).- $\psi : P \longrightarrow M_a^*$ définie par (25) est injective.

<u>*Preuve*</u> : On sait que toute probabilité P sur (X,B) est régulière, c'est-à-dire

$$\forall A \in B \qquad P(A) = \sup(P(F) \; ; \; F \subset A \; , \; F \in F) \quad , \tag{27}$$

où F est la classe des fermés de (X,d) .

On peut donc se limiter à montrer que si $P, Q \in \mathcal{P}$, et si $\psi(P) = \psi(Q)$, alors P et Q coïncident sur $\mathcal{F}$.

Soit $F \in \mathcal{F} - \{\emptyset\}$. Pour $n \geq 1$ posons

$$f_n : X \longrightarrow \mathbb{R} \ : \ x \longmapsto \frac{1}{1+nd(x,F)} - \frac{1}{1+nd(a,F)} \ .$$

On a $f_n(a) = 0$ et $\forall x, y \in X$

$$|f_n(x)-f_n(y)| = \left|\frac{1}{1+nd(x,F)} - \frac{1}{1+nd(y,F)}\right| \leq n\ |d(x,F)-d(y,F)| \leq nd(x,y) \ .$$

Donc, $f_n \in M_a$ $(n \geq 1)$.

Si $\psi(P) = \psi(Q)$, $\forall n \geq 1$ $\int f_n\, dP = \int f_n\, dQ$. Or $\forall n \geq 1$ $\forall x \in X$ $|f_n(x)| \leq 2$ et $\forall x \in X$ $f_n(x) \underset{n\to\infty}{\longrightarrow} \mathbb{1}_F(x) - \mathbb{1}_F(a)$.

D'après le théorème de la convergence dominée, $P(F) - \mathbb{1}_F(a) = Q(F) - \mathbb{1}_F(a)$, $P(F) = Q(F)$. Donc $P = Q$ sur $\mathcal{F}$. ■

Définition IV.6.- Soit (X,d) un espace métrique, $\mathcal{B}$ la tribu borélienne de X, $\mathcal{P}$ l'ensemble des probabilités sur $\mathcal{B}$ vérifiant (22), M_a , $||\cdot||$, ψ et $||\cdot||^*$ définis comme ci-dessus. La distance de Fortet - Mourier, d_{FM} , est donnée par

$$\forall P, Q \in \mathcal{P} \qquad d_{FM}(P,Q) = ||\psi(P)-\psi(Q)||^* = \sup\left(\left|\int f dP - \int f dQ\right| \ ; \ f \in M_a \ , \ ||f||=1\right). \qquad (28)$$

Remarque IV.4.- d_{FM} ne dépend pas du point $a \in X$ fixé dans (22) : si $f : X \longrightarrow \mathbb{R}$ est telle que $f(a) \neq 0$, $\sup\limits_{x,y \in X, x\neq y} \dfrac{|f(x)-f(y)|}{d(x,y)} = 1$, alors $f' = f - f(a) \in M_a$, $||f'|| = 1$ et $\left|\int f' dP - \int f' dQ\right| = \left|\int f dP - \int f dQ\right|$. Par contre, d_{FM} dépend évidemment de la distance d dont X est muni.

R. Moché ([40]) a démontré que, si (X,d) est un espace métrique, $\mathcal{B}$ sa tribu borélienne, $\mathcal{P}$ l'ensemble des lois sur $(X,\mathcal{B})$ vérifiant (22), une condition nécessaire et suffisante pour que d_{FM} soit décantante sur $\mathcal{P}$ est que (X,d) soit borné.

La proposition suivante étend ce résultat.

Proposition IV.7.- Soit (X,d) un espace métrique, $\mathcal{B}$ la tribu borélienne de X, $\mathcal{P}$ l'ensemble des probabilités sur $\mathcal{B}$ vérifiant (22). Une condition nécessaire et suffisante pour que d_{FM} soit décantante sur $\mathcal{P}$ est que (X,d) soit borné. Dans ce cas $\mathcal{P}$ est l'ensemble de toutes les lois boréliennes sur $(X,\mathcal{B})$.

Si (X,d) est borné et $\operatorname{diam} X = D > 0$, on a :

$$f_{d_{FM}} :]0,\infty[\longrightarrow]0,1] \quad : \quad x \longmapsto \min(\tfrac{x}{D}\ ,\ 1) \ . \tag{29}$$

Preuve :

a) Supposons que (X,d) n'est pas borné. $\forall n \geq 1$, $\exists x_n \in X$: $d(a,x_n) \geq n$ $\exists y_n \in X : d(a,y_n) \geq d(a,x_n) + n$.

Pour $n \geq 1$ posons $P_n = (1 - \frac{1}{\sqrt{n}})\ \delta_{\{a\}} + \frac{1}{\sqrt{n}}\ \delta_{\{x_n\}}$; $Q_n = (1 - \frac{1}{\sqrt{n}})\ \delta_{\{a\}} + \frac{1}{\sqrt{n}}\ \delta_{\{y_n\}}$. On a :

$$d_V(P_n,Q_n) = \frac{1}{\sqrt{n}} \xrightarrow[n\to\infty]{} 0 \ . \tag{30}$$

On a vu que $f = d(a,.) \in \mathcal{M}_a$, $||f|| = 1$. Donc $d_{FM}(P_n,Q_n) \geq |\int f\ dP_n - \int f\ dQ_n| = \frac{1}{\sqrt{n}}\ |d(a,x_n) - d(a,y_n)| \geq \sqrt{n} \geq 1$.

On conclut que d_{FM} n'est pas décantante en utilisant le corollaire 1 de la proposition IV.1.

b) Soit (X,d) borné. Plaçons-nous dans le cas non trivial où $\operatorname{diam} X = D > 0$. On voit aisément que $\mathcal{P}$ est l'ensemble des lois boréliennes sur $(X,\mathcal{B})$.

Soient $\alpha > 0$, $P, Q \in \mathcal{P}$ telles que $d_{FM}(P,Q) \geq \alpha$.

$$\forall\beta \in]0,\alpha[\quad \exists f_\beta \in \mathcal{M}_a : ||f_\beta|| = 1 \ , \ \left|\int f_\beta dP - \int f_\beta dQ\right| \geq \beta \ . \tag{31}$$

$$\forall x \in X \qquad |f_\beta(x)| = |f_\beta(x) - f_\beta(a)| \leq d(x,a) \leq D \ .$$

Puisque f_β est bornée, $\forall\varepsilon > 0 \ \exists x_\varepsilon \in X : f_\beta(x_\varepsilon) \leq \inf_{x\in X} f_\beta(x) + \varepsilon$. Fixons un tel x_ε. On voit aisément que $f_{\beta,\varepsilon} = f_\beta - f_\beta(x_\varepsilon)$ est bornée et prend ses valeurs dans $[-\varepsilon,D]$. Donc $\frac{\varepsilon+f_{\beta,\varepsilon}}{D+\varepsilon}$ prend ses valeurs dans $[0,1]$. Etant donné que $d_V(P,Q) = \sup(\left|\int \psi dP - \int \psi dQ\right| \ ; \ \psi$ mesurable à valeurs dans $[0,1])$ et vu que $f_\beta \in \mathcal{M}_a \Longrightarrow f_\beta$ continue $\Longrightarrow f_\beta$ mesurable $\Longrightarrow \frac{\varepsilon+f_{\beta,\varepsilon}}{D+\varepsilon}$ mesurable,

$$d_V(P,Q) \geq \left|\int \frac{\varepsilon+f_{\beta,\varepsilon}}{D+\varepsilon} dP - \int \frac{\varepsilon+f_{\beta,\varepsilon}}{D+\varepsilon} dQ\right| = \left|\int \frac{f_\beta}{D+\varepsilon} dP - \int \frac{f_\beta}{D+\varepsilon} dQ\right| \geq \frac{\beta}{D+\varepsilon} \ .$$

ε étant un réel positif arbitraire ,

$$d_V(P,Q) \geq \frac{\beta}{D} \xrightarrow[\beta\to\alpha]{} \frac{\alpha}{D} > 0 \ . \tag{32}$$

D'après la proposition IV.1. a) d_{FM} est donc décantante sur $\mathcal{P}$.

c) *Calcul de* $f_{d_{FM}}$, *lorsque* (X,d) *est borné*, $\operatorname{diam} X = D > 0$.

Pour $\varepsilon \in]0,D[$ donné, fixons x'_ε , $x''_\varepsilon \in X$ tels que $d(x'_\varepsilon,x''_\varepsilon) \geq D-\varepsilon$.

Soient $P_\varepsilon = \delta_{\{x'_\varepsilon\}}$, $Q_\varepsilon = \delta_{\{x''_\varepsilon\}}$, $f_\varepsilon(x) = d(x,x'_\varepsilon)$. On a $f_\varepsilon \in \mathcal{M}_{x'_\varepsilon}$ et, vu que $f_\varepsilon \neq 0$, on voit comme dans b), p. 93, que

$$||f_\varepsilon|| = 1 \; . \tag{33}$$

D'où

$$d_{FM}(P_\varepsilon,Q_\varepsilon) \geqslant \left|\int f_\varepsilon \, dP_\varepsilon - \int f_\varepsilon \, dQ_\varepsilon\right| = d(x'_\varepsilon,x''_\varepsilon) \quad .$$

D'après (33), $d_{FM}(P_\varepsilon,Q_\varepsilon) \geqslant d(x'_\varepsilon,x''_\varepsilon) \geqslant D-\varepsilon$.

Compte tenu de (32), $1 = d_V(P_\varepsilon,Q_\varepsilon) \geqslant \dfrac{d_{FM}(P_\varepsilon,Q_\varepsilon)}{D} \geqslant \dfrac{D-\varepsilon}{D} \xrightarrow[\varepsilon\to 0]{} 1$.

Donc ,

$$f_{d_{FM}} :]0,\infty[\longrightarrow]0,1] : x \longmapsto (\tfrac{x}{D} , 1) \; . \; \blacksquare \tag{34}$$

Le résultat suivant a été démontré par R. Dobrushin ([17]) pour la distance qu'il a appelée distance de Wasserstein. Il en résulte que celle-ci n'est autre que d_{FM} , dans le cas considéré ci-dessous.

<u>*Corollaire 1.*</u>- Si X est un ensemble non vide muni de la distance discrète d , d_{FM} coïncide avec d_V sur l'ensemble $\mathcal{P}$ de toutes les lois sur $(X, \mathcal{P}(X))$.

<u>*Preuve*</u> : Il est clair que, si X est muni de la distance discrète d, sa tribu borélienne est $\mathcal{P}(X)$. D'autre part, vu que (X,d) est borné, l'ensemble des lois sur $\mathcal{P}(X)$ vérifiant (22) est l'ensemble $\mathcal{P}$ de toutes les lois sur $(X, \mathcal{P}(X))$.

Soit $a \notin X$, $X' = X \cup \{a\}$. Considérons X' muni aussi de la distance discrète d. Si P est une probabilité sur $\mathcal{P}(X)$, elle en est une sur $\mathcal{P}(\mathcal{X}')$ vérifiant $P(\{a\}) = 0$. De même, toute application $f : \mathcal{X} \longrightarrow \mathbb{R}$ peut s'identifier à une application de $\mathcal{X}'$ dans $\mathbb{R}$ telle que $f(a) = 0$.

Soient $P, Q \in \mathcal{P}$, $P \neq Q$, $\mu = P + Q$, $p \in \frac{dP}{d\mu}$, $0 \leq p \leq 1$ sur $\mathcal{X}$, $q = 1-p$ sur $\mathcal{X}$. Faisons $p(a) = q(a) = 0$. Si $f : \mathcal{X}' \longrightarrow \mathbb{R} : x \longmapsto \mathbb{1}_{\{p>q\}}(x)$, $f \in \mathcal{M}_a$ et $||f|| = 1$. Donc $d_{FM}(P,Q) \geq \left|\int f dP - \int f dQ\right| = d_V(P,Q)$.

Or, d'après (32), $d_V(P,Q) \geq d_{FM}(P,Q)$ ce qui nous permet de conclure. ■

<u>*Corollaire 2.*</u>- Sur $(\mathbb{R}, |\cdot|)$, $\mathcal{P}$ est l'ensemble des lois sur $\mathcal{B}_{\mathbb{R}}$ possédant une espérance, et

$$\forall P, Q \in \mathcal{P} \qquad d_{FM}(P,Q) = \int |F(x)-G(x)| \, dx , \tag{35}$$

où F et G désignent respectivement les fonctions de répartition de P et Q.

Ce résultat est démontré dans ([20], p. 275).

<u>*Remarque III.5.*</u>- d_{FM} telle que définie par (35) est aussi appelée distance de Wasserstein sur $\mathcal{P}$ (cf. [51]).

CHAPITRE V

SEPARATION ASYMPTOTIQUE DANS LES CHAINES DE MARKOV.

V.1. - PRELIMINAIRES.-

Soit (Ω, A, Pr) un espace probabilisé, X un ensemble non vide fini ou dénombrable, $\mathcal{B} = \mathcal{P}(X)$, $X = (X_n \; ; \; n \geq 1)$ une suite de variables aléatoires définies sur (Ω, A) et à valeurs dans $(X, \mathcal{B})$.

Soit $k \in \mathbb{N}^*$. Notons le k-uple $(x_1, \ldots, x_k) \in X^k$ par $x^{(k)}$.

$f = \{f(x^{(k)}) \; ; \; x^{(k)} \in X^k\}$ vérifiant $f(x^{(k)}) \geq 0$ $(x^{(k)} \in X^k)$ et $\sum_{x^{(k)} \in X^k} f(x^{(k)}) = 1$, caractérise une loi de probabilité sur X^k .

Pour $n \geq k+1$, soit $P_{n-1,n} : X^k \times X \longrightarrow [0,1]$ vérifiant $\forall x^{(k)} \in X^k \quad \sum_{x \in X} P_{n-1,n}(x^{(k)}, x) = 1$. $P_{n-1,n}$ est appelée fonction (stochastique) de transition de l'instant $n-1$ à l'instant n. Si $k=1$ on appellera aussi $P_{n-1,n}$ matrice de transition.

Supposons fixées une probabilité f sur X^k et une famille $(P_{n-1,n} \; ; \; n \geq k+1)$ de fonctions de transition.

X est dite une chaîne de Markov d'ordre k , définie sur (Ω, A, Pr) et à valeurs dans $(X, \mathcal{B})$, ayant f par loi de $(X_1, \ldots, X_k)$ et $(P_{n-1,n} \; ; \; n \geq k+1)$ par famille de fonctions de transition si :

i) $\forall x^{(k)} \in X^k \quad Pr[(X_1, \ldots, X_k) = x^{(k)}] = f(x^{(k)})$;

ii) $Pr\,[X_n = x_n \mid X_1 = x_1 \,, \ldots, \, X_{n-1} = x_{n-1}] = P_{n-1,n}((x_{n-k}, \ldots, x_{n-1}), x_n)$ $\forall n \geq k+1 \quad \forall x_n \in X \quad \forall x_1, \ldots, x_{n-1} \in X$ tels que le premier membre soit défini, c'est-à-dire tels que $Pr[X_1 = x_1 \,, \ldots, \, X_{n-1} = x_{n-1}] > 0$.

On vérifie aisément que :

$$\forall m \geq k \quad \forall x^{(k)} = (x_1,\ldots,x_k) \in X^k \quad \forall x_{k+1}, \ldots, x_m \in X$$

$$Pr[X_i = x_i (1 \leq i \leq m)] = f(x^{(k)}) \prod_{n=k+1}^{m} P_{n-1,n}((x_{n-k},\ldots,x_{n-1}), x_n) .$$

Si $k = 1$, on parlera simplement d'une chaîne de Markov.

S'il existe une fonction de transition P telle que $P_{n-1,n} = P (n \geq 1)$, on dit que la chaîne est homogène.

V.2. - DECANTATION DE FAMILLES DE FONCTIONS DE TRANSITION.-

Soit $X = (X_n \; ; \; n \geq 1)$ une suite d'observations à valeurs dans $\mathcal{X}$, ensemble fini ou dénombrable, $\mathcal{B} = \mathcal{P}(\mathcal{X})$. On admet que les X_n sont des variables aléatoires définies sur un certain espace probabilisable (Ω, A) et on choisit pour X le modèle statistique $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta \; ; \; \theta \in \Theta))$, où $\mathcal{X}^{(\infty)} = \prod_{i \geq 1} \mathcal{X}_i$, $\mathcal{B}^{(\infty)} = \bigotimes_{i \geq 1} \mathcal{B}_i$, avec $\mathcal{X}_i = \mathcal{X}$, $\mathcal{B}_i = \mathcal{B}$ $(i \geq 1)$, et où $\forall \theta \in \Theta$ X est chaîne de Markov d'ordre k (dans la pratique, le choix de k peut dériver des tests auxquels on a fait mention en I.8.) sous l'hypothèse P_θ.

f_θ , $(P_{n-1,n}(\theta) \; ; \; n \geq k+1)$, $P_{\theta,n}$, $P_\theta^{(n)}$ désigneront respectivement la loi de $(X_1,\ldots,X_k)$, la famille de fonctions de transition, la loi de X_n , la loi de $X^{(n)} = (X_1,\ldots,X_n)$ associées à $\theta \in \Theta$.

On s'abstiendra d'ajouter chaque fois que les modèles considérés dans la suite sont de ce genre.

Définition V.1.- Soient Θ_o et Θ_1 deux parties de Θ . On dit que les deux familles $(P_{n-1,n}(\theta) \; ; \; n \geq k+1 \; , \; \theta \in \Theta_i)$ $(i=0,1)$ sont décantées si à chaque $n \geq k+1$ on peut associer :

- $a_n \in [0,1]$, la suite $(a_n \; ; \; n \geq k+1)$ vérifiant $\sum_{n \geq k+1} a_n^2 = \infty$;
- une fonction $p_n : X^k \longrightarrow [0,1]$ telle que $p_n + a_n \leq 1$ $(n \geq k+1)$;
- à chaque $x^{(k)} \in X^k$ une fonction $\varphi_{x^{(k)},n} : X \longrightarrow [0,1]$, pour

lesquelles $\forall n \geq k+1$ $\forall x^{(k)} \in X^k$

$$\theta \in \Theta_o \Longrightarrow \sum_{y \in X} \varphi_{x^{(k)},n}(y) \cdot [P_{n-1,n}(\theta)]_{x^{(k)},y} \leq p_n(x^{(k)}) \; ,$$

$$\theta \in \Theta_1 \Longrightarrow \sum_{y \in X} \varphi_{x^{(k)},n}(y) \cdot [P_{n-1,n}(\theta)]_{x^{(k)},y} \geq p_n(x^{(k)}) + a_n \; ,$$

où $[P_{n-1,n}(\theta)]_{x^{(k)},y}$ est la valeur assignée par la fonction $P_{n-1,n}(\theta)$ à l'argument $(x^{(k)},y)$.

Cette définition est une généralisation et un raffinement de celle introduite par J. Geffroy dans ([24]). Dans cet article, on ne considère que le cas où la chaîne est homogène et k=1. Par ailleurs, au lieu de la décantation par des tests on y fait appel à la décantation par des événements, et les bornes $p_n(x^{(k)})$ et $p_n(x^{(k)}) + a_n$ sont supposées constantes.

Remarque V.1.- En choisissant le modèle statistique, on a fait un choix des fonctions de transition, autrement dit un choix des lois conditionnelles de X_n sachant $X^{(n-1)}$ sous l'hypothèse P_θ, $\forall n \geq k+1$ $\forall \theta \in \Theta$, puisque évidemment

$$\forall x^{(n-1)} \in X^{(n-1)} \qquad [P_{n-1,n}(\theta)]_{(x_{n-k},\ldots,x_{n-1})} \; ,$$

est un choix de la loi conditionnelle de X_n sachant que $X^{(n-1)}=x^{(n-1)}$.

Proposition V.1.- Si dans le modèle $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta\ ;\ \theta \in \Theta\ )$ les familles de fonctions de transition $(P_{n-1,n}(\theta)\ ;\ n \geq k+1\ ,\ \theta \in \Theta_i)$ $(i=0,1)$ sont décantées, alors les hypothèses $(P_\theta\ ;\ \theta \in \Theta_i)$ $(i=0,1)$ se séparent asymptotiquement uniformément.

Preuve : On n'a qu'à appliquer le théorème III.7., en faisant a'_n , p'_n , ψ'_n tous nuls, pour $1 \leq n \leq k$ et en posant, pour $n \geq k+1$, $x^{(n-1)} = (x_1,\ldots,x_{n-1}) \in \mathcal{X}^{(n-1)}$, $x \in \mathcal{X}$:

- $a'_n = a_n$;
- $p'_n(x^{(n-1)}) = p_n(x_{n-k}\ ,\ldots,\ x_{n-1})$;
- $\psi'_n(x^{(n-1)},x) = \psi_{(x_{n-k},\ldots,x_{n-1}),n}(x)$,

où a_n , p_n et $\psi_{x^{(k)},n}$ vérifient les conditions de la définition V.1.

Puisque $\mathcal{B} = \mathcal{P}(\mathcal{X})$, des questions de mesurabilité ne se présentent pas.

Pour tout $\theta \in \Theta$ et $n \geq k+1$ on aura

$$\left.\begin{aligned}\int_{\mathcal{X}} \psi_n(x^{(n-1)},x)\ P_{\theta,X_n}^{X^{(n-1)}=x^{(n-1)}}(dx) = \sum_{x_n \in \mathcal{X}} \psi_{(x_{n-k},\ldots,x_{n-1})}(x_n)\ \cdot \\ \cdot\ [P_{n-1,n}(\theta)]_{(x_{n-k},\ldots,x_{n-1}),x_n}\end{aligned}\right| \quad (1)$$

Les hypothèses du théorème III.7. sont donc satisfaites. ■

V.3. - PASSAGE D'UNE CHAINE DE MARKOV D'ORDRE $k > 1$ A UNE CHAINE DE MARKOV D'ORDRE 1.-

Soit X une chaîne de Markov d'ordre $k > 1$, définie sur (Ω,A,Pr) et à valeurs dans $(\mathcal{X},\mathcal{B})$, $\mathcal{X}$ fini ou dénombrable, $\mathcal{B} = \mathcal{P}(\mathcal{X})$. Soit f la loi de $(X_1,\ldots,X_k)$ et $(P_{n-1,n}\ ;\ n \geq k+1)$ la famille de fonctions de transition de X.

Soit $Y_n = (X_n,\ldots,X_{n+k-1})$ $(n\geq 1)$. Il s'agit d'une variable aléatoire définie sur (Ω,A) et à valeurs dans $(\mathcal{X}^k,\ \mathcal{P}(\mathcal{X}^k))$. Montrons que $Y = (Y_n\ ;\ n\geq 1)$ est une chaîne de Markov d'ordre 1, ayant f pour loi initiale et $(P^*_{n-1,n}\ ;\ n\geq 2)$ pour famille de fonctions de transition, où
$\forall n \geq 2$ $\quad \forall x^{(k)} = (x_1,\ldots,x_k)$, $y^{(k)} = (y_1,\ldots,y_k) \in \mathcal{X}^k$

$$P^*_{n-1,n}(x^{(k)},y^{(k)}) = \begin{cases} P_{n+k-2,n+k-1}(x^{(k)},y_k), & \text{si } y_i = x_{i+1}\ (1\leq i\leq k-1); \\ 0\ , & \text{sinon}. \end{cases} \tag{2}$$

En effet, il n'y a que deux propriétés à vérifier :

i) que $\forall n \geq 2$ $\quad \forall x^{(k)} \in \mathcal{X}^k$ $\quad \sum\limits_{y^{(k)}\in\mathcal{X}^k} P^*_{n-1,n}(x^{(k)},y^{(k)}) = 1$;

ii) que $\forall n \geq 2$

$$Pr[Y_n=y^{(k)} \mid Y_1=y_1^{(k)},\ldots,Y_{n-1}=y_{n-1}^{(k)}] = P^*_{n-1,n}(y_{n-1}^{(k)},y^{(k)}) \ , \tag{3}$$

$\forall y^{(k)} \in \mathcal{X}^k$ $\quad \forall y_1^{(k)},\ldots,y_{n-1}^{(k)} \in \mathcal{X}^k$ tels que le premier membre de (3) soit défini.

Pour i), il suffit de remarquer que $\forall n \geq 2$ $\quad \forall x^{(k)} \in \mathcal{X}^k$

$$\sum_{y^{(k)}\epsilon X^k} P^*_{n-1,n}(x^{(k)},y^{(k)}) = \sum_{y^{(k)}\epsilon X^k, y_i=x_{i+1}(1\leq i\leq k-1)} P^*_{n-1,n}(x^{(k)},y^{(k)}) =$$

$$= \sum_{y_k\epsilon X} P_{n+k-2,n+k-1}(x^{(k)},y_k) = 1 .$$

Quand à ii), signalons que pour que le premier membre de (3) soit défini, il est nécessaire et suffisant que $Pr\{Y_1=y_1^{(k)},\ldots,Y_{n-1}=y_{n-1}^{(k)}\} > 0$. D'où il résulte que, si pour j donné X_j apparaît plusieurs fois dans l'événement en question, à chaque fois la valeur qui lui est assignée est identique. L'événement est donc la forme $\{X_1=x_1,\ldots,X_{n+k-2}=x_{n+k-2}\}$ et le premier membre de (3) peut s'écrire :

$$Pr[X_n=y_1,\ldots,X_{n+k-1}=y_k|X_1=x_1,\ldots,X_{n+k-2}=x_{n+k-2}], \tag{4}$$

avec $Pr\{X_1=x_1,\ldots,X_{n+k-2}=x_{n+k-2}\} > 0$.

Si $\exists j \in \{1,\ldots,k-1\} : y_j \neq x_{n+j-1}$, la probabilité (4) est nulle, comme la valeur de $P^*_{n-1,n}((x_{n-1},\ldots,x_{n+k-2}),(y_1,\ldots,y_k))$, d'après (12). Sinon (4) devient

$$Pr\{X_n=x_n,\ldots,X_{n+k-2}=x_{n+k-2},X_{n+k-1}=y_k|X_1=x_1,\ldots,X_{n+k-2}=x_{n+k-2}\} =$$

$$= \frac{Pr\{X_1=x_1,\ldots,X_{n+k-2}=x_{n+k-2},X_{n+k-1}=y_k\}}{Pr\{X_1=x_1,\ldots,X_{n+k-2}=x_{n+k-2}\}} = Pr\{X_{n+k-1}=y_k|X_1=x_1,\ldots,X_{n+k-2}=x_{n+k-2}\} =$$

$$= P_{n+k-2,n+k-1}((x_{n-1},\ldots,x_{n+k-2}),y_k) = P^*_{n-1,n}((x_{n-1},\ldots,x_{n+k-2}),(x_n,\ldots,x_{n+k-2},y_k)),$$

d'après (2).

Ceci dit, pour ce qui est de la décantation de familles de fonction de transition, on pourrait se restreindre à considérer des chaînes de Markov d'ordre 1, pourvu que la décantation des familles $(P_{n-1,n}(\theta);\ n\geq k+1,\ \theta \in \Theta_i)(i=0,1)$

associées à une chaîne X d'ordre $k>1$ soit assurée dès que les familles $(P^*_{n-1,n}(\theta)\ ;\ n\geq 2,\ \theta\in\Theta_i)$ $(i=0,1)$ leur correspondant selon le procédé ci-dessus sont décantées, et réciproquement. C'est bien le cas.

En effet, supposons que X est une chaîne de Markov d'ordre $k>1$ à valeurs dans $\mathcal{X}$ et que a_n, p_n, $\psi_{x^{(k)},n}$ $(n\geq k+1,\ x^{(k)}\in\mathcal{X}^k)$ vérifient les conditions de décantation des familles de fonctions de transition de la définition V.1.

$\forall n\geq 2$ $\forall x^{(k)}\in\mathcal{X}^k$ $\forall y^{(k)}=(y_1,\ldots,y_k)\in\mathcal{X}^k$ posons $a^*_n=a_{n+k-1}$, $p^*_n(x^{(k)})=p_{n+k-1}(x^{(k)})$, $\psi^*_{x^{(k)},n}(y^{(k)})=\psi_{x^{(k)},n+k-1}(y_k)$. Alors

$\forall n\geq 2$ $\forall x^{(k)}\in\mathcal{X}^k$ $\forall\theta\in\Theta_o$

$$\sum_{y^{(k)}\in\mathcal{X}^k}\psi^*_{x^{(k)},n}(y^{(k)})\,[P^*_{n-1,n}(\theta)]_{x^{(k)},y^{(k)}}=\sum_{y_k\in\mathcal{X}}\psi_{x^{(k)},n+k-1}(y_k)\ .$$

$$.\ [P_{n+k-2,n+k-1}(\theta)]_{x^{(k)},y_k}\leq p_{n+k-1}(x^{(k)})=p^*_n(x^{(k)}).$$

De même, $\forall\theta\in\Theta_1$ $\displaystyle\sum_{y^{(k)}\in\mathcal{X}^k}\psi^*_{x^{(k)},n}(y^{(k)})[P^*_{n-1,n}(\theta)]_{x^{(k)},y^{(k)}}\geq p^*_n(x^{(k)})+a^*_n$.

Inversement, si Y est une chaîne de Markov d'ordre 1 sur $\mathcal{X}^{(k)}$ associée à une chaîne X d'ordre $k>1$ sur $\mathcal{X}$ selon le procédé décrit ci-dessus et si $\forall n\geq 2$ $\forall x^{(k)}\in\mathcal{X}^k$ a^*_n, p^*_n, $\psi^*_{x^{(k)},n}:\mathcal{X}^k\longrightarrow[0,1]$ décantent les familles $(P^*_{n-1,n}(\theta)\ ;\ n\geq 2\ ,\ \theta\in\Theta_i)$ $(i=0,1)$ dans le sens de la définition V.1., on voit que $(P_{n-1,n}(\theta)\ ;\ n\geq k+1\ ,\ \theta\in\Theta_i)$ $(i=0,1)$ sont décantées en posant $\forall n\geq k+1$ $\forall x^{(k)}=(x_1,\ldots,x_k)\in\mathcal{X}^k$ $\forall y\in\mathcal{X}$

$a_n=a^*_{n-k-1}$, $p_n(x^{(k)})=p^*_{n-k+1}(x^{(k)})$, $\psi_{x^{(k)},n}(y)=\psi^*_{x^{(k)},n-k+1}(x_2,\ldots,x_k,y)$.

Ceci justifie que dans la suite nous ne nous occuperons que des chaînes de Markov d'ordre 1.

V.4. - CAS DES CHAINES DE MARKOV HOMOGENES.-

On examinera dans cette section et les trois suivantes des particularités des chaînes homogènes.

Dans la première sous-section, on rappellera quelques notions et résultats qui seront utilisés plus tard et on fera quelques considérations d'ordre pratique. Dans la deuxième, on verra comment se traduit dans le cas homogène la notion de décantation de familles de fonctions de transition et on introduira la notion de temps de décantation.

V.4.1. - Rappels et considérations générales.

Soit $\mathcal{X}$ un ensemble fini ou dénombrable, $\mathcal{B} = \mathcal{P}(\mathcal{X})$. Soit X une chaîne de Markov homogène définie sur (Ω,A,Pr) et à valeurs dans $(\mathcal{X},\mathcal{B})$, de loi initiale f et de matrice de transition $P = (P_{x,y})_{x,y\in\mathcal{X}}$.

Rappel V.1.- Un état $x \in \mathcal{X}$ est dit récurrent si $\sum_{n\geq 1} P^{(n)}_{x,x} = \infty$, où $P^{(n)}_{x,y}$ désigne l'élément de la ligne x et colonne y de la matrice P^n.

Un étant $x \in \mathcal{X}$ est dit transitoire s'il n'est pas récurrent.

Rappel V.2.- Si $\forall x, y \in \mathcal{X}$ $\exists n \in \mathbb{N}^* : P^{(n)}_{x,y} > 0$, X est dite irréductible. Alors, tous les états sont transitoires ou tous les états sont récurrents, et X est dite, respectivement, irréductible transitoire ou irréductible récurrente. Si X est irréductible récurrente,

$\forall x, y \in X \quad \sum_{n\geq 1} P_{x,y}^{(n)} = \infty$ (cf. [10], chapitre 1, section 8).

Rappel V.3.- Si X est irréductible récurrente, il existe une mesure $\mu = (\mu_x \; ; \; x \in X)$ strictement positive sur X telle que $\forall y \in X \quad \sum_{x\in X} \mu_x P_{x,y} = \mu_y$ (cf. [44], proposition III.36). Naturellement, μ n'est définie qu'à une constante multiplicative près. Si $\mu(X) = \sum_{x\in X} \mu_x = \infty$, X est dite irréductible récurrente nulle. Sinon, $\pi_x = \frac{\mu_x}{\mu(X)}$ $(x \in X)$, $\pi(A) = \sum_{x\in A} \pi_x$ $(A \subset X)$ définit une probabilité π sur $P(X)$ que l'on appellera distribution stationnaire de X.

Rappel V.4.- Soit X irréductible récurrente. $\forall x \in X$ $\exists s \in \mathbb{N}^* : \Pr(X_s = x) > 0$. Soit m le plus petit entier vérifiant cette condition pour un état $x \in X$, $n \in \mathbb{N}^*$. Posons $f_x^{(n)} = \Pr[X_{m+k} \neq x (1 \leq k < n), X_{m+n} = x | X_m = x]$. On voit aisément que $f_x^{(n)}$ ne dépend pas de m. On appelle temps moyen de retour à x la quantité $\lambda_x = \sum_{n\geq 1} n \, f_x^{(n)}$. Alors (cf. [44], définition et corollaire III.37) si X est nulle $\lambda_x = \infty$ $(x \in X)$, tandis que si X est posivite $\lambda_x = \frac{1}{\pi_x}$ $(x \in X)$.

Rappel V.5.- Soit $x \in X$. Si $\exists n \in \mathbb{N}^*$ tel que $P_{x,x}^{(n)} > 0$ et $d = \text{P.G.C.D.} \; \{n \in \mathbb{N}^* : P_{x,x}^{(n)} > 0\}$, on appelle d la période de l'état x.

Si X est irréductible, tous les états ont même période d. Si $d > 1$, X est dite irréductible périodique de période d. Si $d = 1$, X est dite irréductible apériodique. Si X est irréductible de période d, $\forall x, y \in X \quad \exists ! \; r \in \{0,\dots,d-1\}$ tel que $P_{x,y}^{(n)} > 0 \Longrightarrow n = r (\text{mod } d)$. Posons alors $y \in C_r(x)$. On a $\forall s \in \{0,\dots,d-1\} \quad C_s(y) = C_{r+s(\text{mod } d)}(x)$. Par ailleurs, si $y \in C_r(x)$ $\exists N_y \in \mathbb{N}^*$ tel que $n \geq N_y \Longrightarrow P_{x,y}^{(nd+r)} > 0$. D'autre part, $\forall x \in X \quad X = \sum_{r=0}^{d-1} C_r(x)$ (cf. [12], I.3).

Rappel V.6.- Si X est irréductible récurrente positive de période d et si $r \in \{0,\dots,d-1\}$, $x, y \in \mathcal{X}$, $y \in C_r(x)$, alors $\lim_{n\to\infty} P_{x,y}^{(nd+r)} = d\,\pi_y$, où π est la distribution stationnaire de X (cf. [12], I.6.). Il s'ensuit : $\forall x, y \in \mathcal{X}$ $\frac{1}{n} \sum_{m=1}^{n} P_{x,y}^{(m)} \xrightarrow[n\to\infty]{} \pi_y$.

Rappel V.7.- Si $\mathcal{X}$ est fini et si X est irréductible, X est irréductible récurrente positive. Si, de plus, X est apériodique, $\exists c > 0$ $\exists r \in [0,1[$ tels que $\forall x , y \in \mathcal{X}$ $|P_{x,y}^{(n)} - \pi_y| \leq c\, r^n$ (cf. [10], section citée).

Compte tenu de ces rappels, considérons maintenant une chaîne de Markov homogène X dont l'espace des états $\mathcal{X}$ est fini ou dénombrable et dont la loi est inconnue. La connaissance a priori du phénomène qui a engendré la chaîne de Markov, l'expérience du praticien qui enregistre les observations permettent souvent de restreindre considérablement l'hypothèse générale, parce que les différents types de chaînes de Markov ont des comportements très différenciés.

Ainsi on sait que les chaînes de Markov rencontrées dans les applications sont souvent irréductibles et l'on peut raisonnablement penser que l'on est capable de préciser si l'hypothèse générale $(P_\theta \; ; \; \theta \in \Theta)$ vérifie :

- soit $\forall \theta \in \Theta$ X est irréductible récurrente sous l'hypothèse P_θ ;

- soit $\forall \theta \in \Theta$ X est irréductible transitoire sous l'hypothèse P_θ ,

puisqu'un cas où X pourrait être à la fois irréductible récurrente et irréductible transitoire pour deux valeurs différentes de θ est pratiquement à exclure.

Si l'on a été amené à supposer que X est irréductible récurrente pour toute hypothèse P_θ , on sera aussi capable normalement de préciser si l'on a :

- soit $\forall\theta \in \Theta$ X est positive sous l'hypothèse P_θ ;
- soit $\forall\theta \in \Theta$ X est nulle sous l'hypothèse P_θ ,

un cas mélangé étant encore à exclure pratiquement vu la différence significative de comportement des temps moyens de retour à chaque état $x \in \mathcal{X}$ (voir rappel V.4.).

Par ailleurs, dans certains cas, on peut avoir des raisons pour admettre que $\forall\theta \in \Theta$ X est irréductible périodique de période d sous l'hypothèse P_θ (cf. [12], I.5.).

V.4.2. - Décantation de familles de matrices de transition, dans le cas homogène.

$P(\theta)$ étant une matrice stochastique, on désignera par $P_{x,y}(\theta)$ (resp. $P^{(t)}_{x,y}(\theta)$) l'élément de la ligne x et colonne y de $P(\theta)$ (resp. $[P(\theta)]^t$).

Supposons que X est une chaîne de Markov homogène à valeurs dans $(\mathcal{X}, \mathcal{P}(\mathcal{X}))$, $\mathcal{X}$ étant au plus dénombrable, et que $(\mathcal{X}^{(\infty)}, (\mathcal{P}(\mathcal{X}))^{(\infty)}, (P_\theta ; \theta \in \Theta))$ est un modèle statistique pour X , les familles $(P(\theta) ; \theta \in \Theta_i \subset \Theta)$ (i=0,1) étant décantées. Ceci veut dire qu'à tout $n \geq 2$ on peut associer :

- $a_n \in [0,1]$, $(a_n ; n \geq 2)$ vérifiant $\sum_{n\geq 2} a_n^2 = \infty$;
- $p_n : \mathcal{X} \longrightarrow [0,1]$ telle que $p_n + a_n \leq 1$;

- à tout $x \in X$ une application $\varphi_{x,n} : X \longrightarrow [0,1]$ pour lesquels $\forall n \geq 2 \quad \forall x \in X$

$$\theta \in \Theta_o \Longrightarrow \sum_{y \in X} \varphi_{x,n}(y)\, P_{x,y}(\theta) \leq p_n(x) \quad , \tag{5}$$

$$\theta \in \Theta_1 \Longrightarrow \sum_{y \in X} \varphi_{x,n}(y)\, P_{x,y}(\theta) \geq p_n(x) + a_n \ . \tag{6}$$

Vu qu'il existe $N \geq 2$ tel que $a_N = a > 0$, on peut substituer aux données précédentes les nouvelles données $a_n = a$, $p_n = p_N$, $\varphi_{x,n} = \varphi_{x,N}$ $(n \geq 2 \ , \ x \in X)$ pour conclure que $(P_\theta \ ; \ \theta \in \Theta_i)$ $(i=0,1)$ se séparent asymptotiquement uniformément : en effet, on a bien $\sum_{n=2}^{k} a_n^2 = (k-1)\, a^2 \xrightarrow[k\to\infty]{} \infty$ et $\forall n \geq 2 \quad \exists C_n \subset X^n$ pour lesquels
$\sup_{\theta \in \Theta_o} P_\theta^{(n)}(C_n) \leq \exp\{- 3/8(n-1)a^2\}$, $\inf_{\theta \in \Theta_1} P_\theta^{(n)}(C_n) \geq 1 - \exp\{- 3/8(n-1)a^2\}$,
d'après le théorème III.7. (voir aussi III. (38) et (39)).

Donc, dans le cas homogène, la notion de décantation de deux familles $(P(\theta) \ ; \ \theta \in \Theta_i \subset \Theta)$ $(i=0,1)$ de matrices de transition se traduit par :

$$\left.\begin{array}{l} \exists a > 0 \quad \exists p : X \longrightarrow [0,1-a] \quad \text{et} \quad \forall x \in X \ \exists \ \varphi_x : X \longrightarrow [0,1] \\ \text{tels que} \quad \forall x \in X \quad : \\ \qquad \theta \in \Theta_o \Longrightarrow \sum_{y \in X} \varphi_x(y)\, P_{x,y}(\theta) \leq p(x) \ , \\ \qquad \theta \in \Theta_1 \Longrightarrow \sum_{y \in X} \varphi_x(y)\, P_{x,y}(\theta) \geq p(x) + a \ . \end{array}\right| \tag{7}$$

<u>*Définition V.2.*</u> ([24]).- $t \in \mathbb{N}^*$ est appelé un temps de décantation pour $(P(\theta) \ ; \ \theta \in \Theta_i \subset \Theta)$ $(i=0,1)$ si $([P(\theta)]^t \ ; \ \theta \in \Theta_i)$ $(i=0,1)$ sont décantées.

Proposition V.2. ([24]).- S'il existe un temps de décantation $t \in \mathbb{N}^*$ pour $(P(\theta)\ ;\ \theta \in \Theta_i \subset \Theta)$ $(i=0,1)$, alors les hypothèses $(P_\theta\ ;\ \theta \in \Theta_i)$ $(i=0,1)$ se séparent asymptotiquement uniformément.

Preuve : Soit $X^*_m = X_{(m-1)t+1}$ $(m \geq 1)$. $X^* = (X^*_m\ ;\ m \geq 1)$ est une chaîne de Markov homogène dont la matrice de transition $P^*(\theta)$ associée à $\theta \in \Theta$ est donnée par $P^*(\theta) = [P(\theta)]^t$.

Puisque $(P^*(\theta)\ ;\theta \in \Theta_i)$ $(i=0,1)$ sont décantées, $\exists a > 0$ $\exists (C^*_m\ ;\ m \geq 2)$ suite de cylindres $(\mathcal{P}(X))^{(m)}$-adaptée pour lesquels

$$\sup_{\theta\in\Theta_0} P_\theta^{*(m)}(C^*_m) \leq \exp\{-3/8(m-1)a^2\}, \quad \inf_{\theta\in\Theta_1} P_\theta^{*(m)}(C^*_m) \geq 1-\exp\{-3/8(m-1)a^2\}\ , \tag{8}$$

où P^*_θ désigne la loi de X^* sous l'hypothèse P_θ.

$\forall n \geq t+1 \quad \exists!\ m_n \in \mathbb{N}^*$: $m_n\, t+1 \leq n < (m_n+1)\, t+1$. On a $m_n > \dfrac{n-1}{t} -1$.

Donc

$$n \longrightarrow \infty \implies m_n \longrightarrow \infty\ . \tag{9}$$

$\forall n \geq t+1$ posons

$C_n = \{x^{(n)}=(x_1,x_2,\ldots,x_n) \in X^n\ :\ (x_1,x_{t+1},x_{2t+1},\ldots,x_{m_n t+1}) \in C^*_{m_n+1}\}$. Alors $\forall\theta \in \Theta \quad P_\theta^{(n)}(C_n) = P_\theta^{*(m_n+1)}(C^*_{m_n+1})$ et (8) et (9) nous permettent de conclure. ■

V.5. - CAS DES CHAINES IRREDUCTIBLES RECURRENTES POSITIVES APERIODIQUES.-

L'idée de décanter les distributions stationnaires est dûe à J. Geffroy ([24]).

Proposition V.3.- Si $\Theta_o , \Theta_1 \subset \Theta$ et $\forall \theta \in \Theta' = \Theta_o \cup \Theta_1$ X est irréductible récurrente positive sous l'hypothèse P_θ , notons $\pi(\theta)$ sa distribution stationnaire. Supposons que :

a) $\forall t \in \mathbb{N}^* \quad \exists (\eta_t(x) \ ; \ x \in X)$ mesure positive finie sur X telle que $\eta_t(X) = \sum_{x \in X} \eta_t(x) \xrightarrow[t\to\infty]{} 0$ et $\forall \theta \in \Theta' \quad \forall x,y \in X \quad |P^{(t)}_{x,y}(\theta) - \pi_y(\theta)| \leq \eta_t(y)$;

b) $(\pi(\theta) \ ; \ \theta \in \Theta_i)$ $(i=0,1)$ sont décantées par un test ψ,

$$\omega_o = \sup_{\theta \in \Theta_o} \sum_{x \in X} \psi(x)\ \pi_x(\theta) < \inf_{\theta \in \Theta_1} \sum_{x \in X} \psi(x)\ \pi_x(\theta) = \omega_1 .$$

Alors les hypothèses $(P_\theta \ ; \ \theta \in \Theta_i)$ $(i=0,1)$ se séparent asymptotiquement uniformément.

Remarque V.2.- Dans l'énoncé ci-dessus, l'apériodicité de la chaîne X n'est pas posée en hypothèse. Mais on suppose que $P^{(t)}_{x,y}(\theta) \xrightarrow[t\to\infty]{} \pi_y(\theta)$, ce qui implique que X est apériodique (voir rappel V.5.).

Preuve de la proposition V.3.- $\forall t \geq 1 \quad \forall \theta \in \Theta_o \quad \forall x, y \in X$

$P^{(t)}_{x,y}(\theta) \leq \pi_y(\theta) + \eta_t(y)$, d'où $\sum_{y \in X} \psi(y)\ P^{(t)}_{x,y}(\theta) \leq \omega_o + \eta_t(X)$.

De même, pour $\theta \in \Theta_1$ on a $\sum_{y \in X} \psi(y)\ P^{(t)}_{x,y}(\theta) \geq \omega_1 - \eta_t(X)$.

Si $\eta_t(X) < \frac{\omega_1 - \omega_o}{2}$, alors t est un temps de décantation pour $(P(\theta) \ ; \ \theta \in \Theta_i)$ $(i=0,1)$, ce qui permet de conclure. ∎

Remarque V.3.- Si X est fini et si X est irréductible apériodique sous P_θ ($\theta \in \Theta'$), d'après le rappel V.7. $\exists c : \Theta' \longrightarrow \mathbb{R}_+^*$ $\exists r : \Theta' \longrightarrow [0,1[$ vérifiant :

$$\forall \theta \in \Theta' \quad \forall x, y \in \mathcal{X} \qquad |P_{x,y}^{(t)}(\theta) - \pi_y(\theta)| \leq c(\theta)\, r^t(\theta) \quad .$$

S'il existe une fonction $\eta(t)$, convergeant vers 0 lorsque $t \to \infty$, qui majore $c(\theta)\, r^t(\theta)$ uniformément sur Θ', on obtient a) dans la proposition V.3. en posant $\eta_t(x) = \eta(t)$ ($x \in \mathcal{X}$).

Sous l'hypothèse b) seulement, on pourra démontrer par les procédés standard que (P_θ ; $\theta \in \Theta_i$) (i=0,1) se séparent asymptotiquement.

Corollaire 1. ([24]).- Si $\mathcal{X}$ est fini, $\Theta = \{\theta_o, \theta_1\}$, X est irréductible apériodique sous l'hypothèse P_{θ_i} (i=0,1) et si $\pi(\theta_o) \neq \pi(\theta_1)$, alors $P_{\theta_o} \perp P_{\theta_1}$.

Pour le cas où $\mathcal{X}$ est infini dénombrable, on peut faire appel au théorème suivant, qui étend un résultat de J. Neveu ([44], proposition III.38) en explicitant une minoration de la vitesse de convergence, obtenue selon un procédé classique (cf. [18], p. 173).

Théorème V.1.- Soit X une chaîne de Markov dont l'espace des états $\mathcal{X}$ est dénombrable et dont la matrice de transition P vérifie la propriété suivante :

. $\exists\, k \in \mathbb{N}^*$ $\exists \rho = (\rho_x \; ; \; x \in \mathcal{X})$ mesure strictement positive sur $\mathcal{X}$ tels que $\forall x, y \in \mathcal{X}$ $P_{x,y}^{(k)} \geq \rho_y$.

Alors X est irréductible récurrente positive apériodique. π désignant sa distribution stationnaire, on a :

$$\forall t > k \quad \forall x\,,\, y \in \mathcal{X} \qquad |P_{x,y}^{(t)} - \pi_y| \leq [1-\rho(\mathcal{X})]^{\frac{t}{k}-1} \xrightarrow[t\to\infty]{} 0 \quad .$$

Preuve : Puisque $\forall x, y \in X \quad P^{(k)}_{x,y} > 0$, X est irréductible .

Remarquons que $\forall x, y \in X \quad P^{(k+1)}_{x,y} = \sum_{z\in X} P_{x,z}\, P^{(k)}_{z,y} \geq \rho_y \sum_{z\in X} P_{x,z} = \rho_y$.

Par récurrence, on conclut que $\forall n \geq k \quad \forall x, y \in X \quad P^{(n)}_{x,y} \geq \rho_y$. Donc :

i) Vu que P.G.C.D. $\{n \in \mathbb{N}^* : n \geq k\} = 1$, X est apériodique.

ii) $\forall x \in X \quad \sum_{n\geq 1} P^{(n)}_{x,x} \geq \sum_{n\geq k} P^{(n)}_{x,x} \geq \sum_{n\geq k} \rho_x = \infty$, vu que $\rho_x > 0$.

Ainsi, X est irréductible récurrente.

Montrons que X est irréductible récurrente positive. Si μ est une mesure strictement positive sur $(X, \mathcal{P}(X))$ vérifiant $\forall y \in X$ $\mu_y = \sum_{x\in X} \mu_x\, P_{x,y}$, alors $\forall y \in X \quad \mu_y = \sum_{x\in X} \mu_x\, P^{(k)}_{x,y} \geq \rho_y\, \mu(X)$, et $\mu_y , \rho_y \in \mathbb{R}^*_+ \Longrightarrow \mu(X) < \infty$.

Le fait que $1 = \sum_{y\in X} P^{(k)}_{x,y} \geq \sum_{y\in X} \rho_y = \rho(X)$ permet d'affirmer que $[1 - \rho(X)]^{\frac{t}{k}-1} \xrightarrow[t\to\infty]{} 0$.

Montrons maintenant que :

$$\forall t > k \quad \forall x, y \in X \qquad |P^{(t)}_{x,y} - \pi_y| \leq [1 - \rho(X)]^{\frac{t}{k}-1} .$$

Soient $m^r_y = \inf_{x\in X} P^{(r)}_{x,y}$, $M^r_y = \sup_{x\in X} P^{(r)}_{x,y}$ $(y \in X , r \geq 1)$.

Puisque $m^{r+1}_y = \inf_{x\in X} \sum_{z\in X} P_{x,z}\, P^{(r)}_{z,y} \geq m^r_y$ et de même $M^{r+1}_y \leq M^r_y$, on a $m^1_y \leq m^2_y \leq \ldots \leq M^2_y \leq M^1_y$.

Soient x,y fixés dans X. Définissons sur $\mathcal{P}(X)$ une mesure à signe $\emptyset$ par $\forall z \in X \quad \emptyset(z) = P^{(k)}_{x,z} - P^{(k)}_{y,z}$, $\forall B \subset X \quad \emptyset(B) = \sum_{b\in B} \emptyset(b)$.

Soit $B_{x,y} = \{b \in X : \emptyset(b) \geq 0\}$. On a $\emptyset(B_{x,y}) + \emptyset(B^c_{x,y}) = \emptyset(X) = 0$.

D'autre part,

$$\psi(B_{x,y}) = \sum_{b\in B_{x,y}} P^{(k)}_{x,b} - \sum_{b\in B_{x,y}} P^{(k)}_{y,b} = 1 - \sum_{d\in B^c_{x,y}} P^{(k)}_{x,d} - \sum_{b\in B_{x,y}} P^{(k)}_{y,b} \leq 1 - \rho(X) < 1 .$$

On a

$$M^k_z - m^k_z = \sup_{x,y\in X} (P^{(k)}_{x,z} - P^{(k)}_{y,z}) \leq \sup_{x,y\in X} \sum_{b\in B_{x,y}} (P^{(k)}_{x,b} - P^{(k)}_{y,b}) = \sup_{x,y\in X} \psi(B_{x,y}) \leq 1-\rho(X) .$$

Par ailleurs, $\forall t \geq 1$

$$M^{t+k}_z - m^{t+k}_z = \sup_{x,y\in X} (P^{(t+k)}_{x,z} - P^{(t+k)}_{y,z}) = \sup_{x,y\in X} \Big(\sum_{\omega\in X} (P^{(k)}_{x,\omega} - P^{(k)}_{y,\omega})\, P^{(t)}_{\omega,z}\Big) \leq$$

$$\leq \sup_{x,y\in X} \Big(\sum_{\omega\in B_{x,y}} (P^{(k)}_{x,\omega} - P^{(k)}_{y,\omega})\, M^t_z + \sum_{\omega\in B^c_{x,y}} (P^{(k)}_{x,\omega} - P^{(k)}_{y,\omega})\, m^t_z\Big) =$$

$$= \sup_{x,y\in X} (\psi(B_{x,y})\, M^t_z - \psi(B_{x,y})\, m^t_z) = (\sup_{x,y\in X} \psi(B_{x,y}))\,(M^t_z - m^t_z) \leq (1-\rho(X))(M^t_z - m^t_z).$$

D'où $M^{rk}_z - m^{rk}_z \leq (1-\rho(X))^r$ $(r \geq 1,\ z \in X)$. Donc $\forall z \in X$ m^t_z et M^t_z ont une limite commune lorsque $t \to \infty$.

Puisque $\forall x, y \in X$ $\forall n \in \mathbb{N}^*$ $M^n_y \geq \lim_{t\to\infty} M^t_y \geq \lim_{t\to\infty} P^{(t)}_{x,y} =$

$= \pi_y \geq \lim_{t\to\infty} m^t_y \geq m^n_y$, $\forall t > k$ $|P^{(t)}_{x,y} - \pi_y| \leq M^t_y - m^t_y \leq M^{[\frac{t}{k}].k}_y - m^{[\frac{t}{k}].k}_y \leq$

$\leq (1-\rho(X))^{[\frac{t}{k}]} \leq (1-\rho(X))^{\frac{t}{k}-1}$, ce qu'il fallait démontrer . ■

Revenons au modèle statistique introduit précédemment. Le théorème V.1. nous permet d'obtenir un résultat dont la preuve est analogue à celle de la proposition V.3.

Corollaire 1.- Si $\forall\theta \in \Theta' = \Theta_o \cup \Theta_1$ $\exists k_\theta \in \mathbb{N}^*$ $\exists \rho(\theta)$ mesure strictement positive sur X telle que :

- $P_{x,y}^{(k_\theta)} \geq \rho_y(\theta)$ $(x, y \in X)$;

- $k = \sup\limits_{\theta\in\Theta'} k_\theta < \infty$;

- $R = \inf\limits_{\theta\in\Theta'} \sum\limits_{x\in X} \rho_x(\theta) > 0$,

et si $(\pi(\theta)\ ;\ \theta \in \Theta_i)$ $(i=0,1)$ sont décantées par un événement fini A , $\omega_o = \sup\limits_{\theta\in\Theta_o} \sum\limits_{x\in A} \pi_x(\theta) < \inf\limits_{\theta\in\Theta_1} \sum\limits_{x\in A} \pi_x(\theta) = \omega_1$, alors les hypothèses

$(P_\theta\ ;\ \theta \in \Theta_i)$ $(i=0,1)$ se séparent asymptotiquement uniformément.

Preuve : Remarquons qu'en vertu des hypothèses, le théorème V.1. nous permet d'assurer l'existence des distributions stationnaires $\pi(\theta)$ $(\theta \in \Theta')$, la chaîne étant irréductible récurrente positive apériodique $\forall\theta \in \Theta'$. De plus, $\forall\theta \in \Theta'$ $\forall t > k$ $\forall x, y \in X$ $|P_{x,y}^{(t)}(\theta) - \pi_y(\theta)| \leq (1-R)^{\frac{t}{k}-1}$. D'où :

- $\forall t > k$ $\forall\theta \in \Theta_o$ $P_{x,y}^{(t)}(\theta) \leq \pi_y(\theta) + (1-R)^{\frac{t}{k}-1}$ et

$$\sup_{\theta\in\Theta_o} \sum_{y\in A} P_{x,y}^{(t)}(\theta) \leq \omega_o + (1-R)^{\frac{t}{k}-1} .\ \mathrm{card}(A) ;$$

- de même, $\forall t > k$ $\inf\limits_{\theta\in\Theta_1} \sum\limits_{y\in A} P_{x,y}^{(t)}(\theta) \geq \omega_1 - (1-R)^{\frac{t}{k}-1} .\ \mathrm{card}(A)$.

Si $(1-R)^{\frac{t}{k}-1} < \dfrac{\omega_1 - \omega_o}{2\ \mathrm{card}(A)}$, t est un temps de décantation pour

$(P(\theta)\ ;\ \theta \in \Theta_i)$ $(i=0,1)$, ce qui permet de conclure. ■

V.6. - CAS DES CHAINES IRREDUCTIBLES RECURRENTES POSITIVES PERIODIQUES.-

Soit $X = (X_n\ ;\ n \geq 1)$ une chaîne de Markov homogène définie sur (Ω, A, Pr) de loi initiale f et de matrice de transition P, à valeurs dans $(\mathcal{X}, \mathcal{B})$, où $\mathcal{X}$ est un ensemble dénombrable, $\mathcal{B} = \mathcal{P}(\mathcal{X})$.

Supposons X irréductible récurrente positive de période $d > 1$. Sous certaines conditions, on peut associer à X une chaîne de Markov Y irréductible récurrente positive apériodique. Ce fait nous permettra d'obtenir un résultat analogue à la proposition V.3., applicable au cas où X est irréductible périodique de période $d > 1$.

Supposons que $\exists x \in \mathcal{X}$ tel que f est concentrée sur $C_o(x)$.

Pour $m \geq 1$, soit $Y_m = X_{(m-1)d+1}$. On voit aisément que $\forall m \geq 1$ Y_m est p.-s. à valeurs dans $C_o(x)$. Soit $\Omega^* = \bigcap_{m \geq 1} (Y_m \in C_o(x))$, A^* la tribu trace de A sur Ω^*, $Pr^* = Pr \upharpoonright A^*$. Soit :

$$\left.\begin{array}{l} Y : (\Omega^*, A^*, Pr^*) \longrightarrow ((C_o(x))^{(\infty)}, (\mathcal{P}(C_o(x)))^{(\infty)}) \\ \qquad\qquad \omega \longrightarrow (Y_m(\omega)\ ;\ m \geq 1) \end{array}\right| \quad (10)$$

On va voir que Y est une chaîne de Markov irréductible récurrente positive apériodique, de loi initiale f et de matrice de transition $P^* = P^d \upharpoonright (C_o(x) \times C_o(x))$.

Il est évident que la loi initiale de Y est f. D'autre part, si $n \in \mathbb{N}^*$, $y_i \in C_o(x)$ $(1 \leq i \leq n)$ et $Pr^*[Y_i = y_i\ (1 \leq i \leq n)] > 0$, compte tenu de la propriété de Markov de X on a :

$$\forall y \in C_o(x) \qquad Pr^*[Y_{n+1} = y \mid Y_i = y_i (1 \leq i \leq n)] = Pr[X_{nd+1} = y \mid X_{(n-1)d+1} = y_n] = P^{(d)}_{y_n, y}\ .$$

Ceci montre que P^* est la matrice de transition de Y.

<u>*Remarque V.4.*</u>- On devrait dire plutôt que P^* est un des choix possibles pour la matrice de transition de Y. Pourtant, si $\forall y \in C_o(x)$ $\exists n \in \mathbb{N}^*$ tel que $Pr^*(Y_n=y) > 0$, on voit aisément que la matrice de transition est univoquement définie. Or on montrera ci-dessous que Y est irréductible, ce qui permet d'affirmer que P^* est la matrice de transition de Y.

Soit $y, z \in C_o(x)$. D'après le rappel V.5., $\exists n , m \in \mathbb{N}^*$ tels que $P^{(nd)}_{y,x} > 0$, $P^{(md)}_{x,z} > 0$. D'où : $P^{*(n+m)}_{y,z} = P^{[(n+m)d]}_{y,z} \geq P^{(nd)}_{y,x} P^{(md)}_{x,z} > 0$. Donc Y est irréductible.

Les résultats présentés dans le rappel V.5., impliquent, de plus, que $\forall y \in C_o(x)$ $\exists N \in \mathbb{N}^* : n \geq N \Longrightarrow P^{(nd)}_{x,y} > 0$, $P^{(nd)}_{y,x} > 0$. Il s'ensuit que, si $k \geq N$, $P^{*(k+N)}_{y,y} \geq P^{(kd)}_{y,x} \cdot P^{(Nd)}_{x,y} > 0$. Vu que P.G.C.D. $\{m \in \mathbb{N}^* : m \geq 2N\} = 1$, Y est irréductible apériodique.

Si $y \in C_o(x)$, $P^{(m)}_{y,y} > 0 \Longrightarrow m = 0 \pmod d$. Donc, $\forall y \in C_o(x)$ $\sum_{n\geq 1} P^{*(n)}_{y,y} = \sum_{n\geq 1} P^{(nd)}_{y,y} = \sum_{n\geq 1} P^{(n)}_{y,y}$. X étant par hypothèse irréductible récurrente, $\sum_{n\geq 1} P^{(n)}_{y,y} = \infty$. Il en résulte que Y est aussi irréductible récurrente.

Si π désigne la distribution stationnaire de X, $\forall y \in C_o(x)$ $\sum_{z\in X} \pi_z P^{(d)}_{z,y} = \pi_y$. Mais, vu que $P^{(d)}_{z,y} = 0$ si $z \notin C_o(x)$, on a aussi $\sum_{z\in C_o(x)} \pi_z P^{(d)}_{z,y} = \sum_{z\in C_o(x)} \pi_z P^*_{z,y} = \pi_y$. Puisque $\pi(C_o(x)) < \infty$, Y est irréductible récurrente positive.

Enfin, le lemme suivant permet de vérifier que $\pi(C_o(x)) = \frac{1}{d}$, ce qui implique que la distribution stationnaire π^* de Y est donnée par $\pi^*_y = d\ \pi_y$ $(y \in C_o(x))$.

Lemme V.1.- Soit X une chaîne de Markov irréductible récurrente positive de période d, dont l'espace des états $\mathcal{X}$ est dénombrable. Désignons par π la distribution stationnaire de X. Alors $\forall x \in \mathcal{X}$ $\forall r \in \{0,\dots,d-1\}$ $\pi(C_r(x)) = \frac{1}{d}$.

Preuve : D'après le rappel V.5., $\forall r \in \{0,\dots,d-1\}$ $y \in C_r(x) \Longrightarrow C_k(y) = C_{k+r(\text{mod } d)}(x)$. On en déduit que :

$$y \in C_r(x)\ ,\ y \in C_1(z) \Longrightarrow C_r(x) = C_1(z) \Longrightarrow z \in C_{d+r-1(\text{mod } d)}(x) \quad . \tag{11}$$

Ainsi, $\pi(C_r(x)) = \sum_{y\in C_r(x)} \pi_y = \sum_{y\in C_r(x)} \sum_{z\in\mathcal{X}} \pi_z P_{z,y} =$

$$= \sum_{y\in C_r(x)} \sum_{z\in\mathcal{X}, y\in C_1(z)} \pi_z P_{z,y} =$$

(puisque $y \notin C_1(z) \Longrightarrow P_{z,y} = 0$)

$$= \sum_{y\in C_r(x)} \sum_{z\in C_{d+r-1(\text{mod } d)}(x)} \pi_z P_{z,y} =$$

(d'après (11))

$$= \sum_{z\in C_{d+r-1(\text{mod } d)}(x)} \pi_z \Big(\sum_{y\in C_r(x)} P_{z,y}\Big) = \sum_{z\in C_{d+r-1(\text{mod } d)}(x)} \pi_z \Big(\sum_{y\in C_1(z)} P_{z,y}\Big) =$$

(car $C_r(x) = C_1(z)$ si $z \in C_{d+r-1(\text{mod } d)}(x)$)

$$= \sum_{z\in C_{d+r-1(\text{mod } d)}(x)} \pi_z = \pi(C_{d+r-1(\text{mod } d)}(x))$$

(vu que $\sum_{y\in C_1(z)} P_{z,y} = 1$).

Donnant successivement à r les valeurs $1,2,\dots,d-1,0$ on voit que $\pi(C_r(x))$ est indépendant de r, d'où $1 = \pi(\mathcal{X}) = \sum_{r=0}^{d-1} \pi(C_r(x)) = d\pi(C_i(x))$, où i est un élément arbitraire de $\{0,\dots,d-1\}$. Donc $\forall x \in \mathcal{X}$ $\forall r \in \{0,\dots,d-1\}$ $\pi(C_r(x)) = \frac{1}{d}$. ■

Corollaire 1.- $\forall x \in \mathcal{X} \quad \forall r \in \{0,\ldots,d-1\} \quad C_r(x) \neq \emptyset$.

Revenons au problème d'une chaîne de Markov X de loi inconnue.

Proposition V.4.- Soit X une chaîne de Markov à valeurs dans un ensemble fini ou dénombrable $\mathcal{X}$ muni de la tribu $\mathcal{B} = \mathcal{P}(\mathcal{X})$. Soit $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta \ ; \ \theta \in \Theta))$ un modèle statistique pour X.

Si $\Theta_i \subset \Theta$ (i=0,1) et si $\forall \theta \in \Theta' = \Theta_0 \cup \Theta_1$ X est une chaîne de Markov irréductible récurrente positive, notons $\pi(\theta)$ la distribution stationnaire associée à $\theta \in \Theta'$.

Supposons que :

- $\exists d > 1$ tel que $\forall \theta \in \Theta'$ X est périodique de période d.
- $\exists x \in \mathcal{X}$: $\forall \theta \in \Theta'$ $C_0(x)$ est indépendante de θ et $P_\theta^{(1)}$ est concentrée sur $C_0(x)$. (12)

Admettons de plus que :

a) $\forall t \in \mathbb{N}^* \quad \exists (\eta_t(y) \ ; \ y \in C_0(x))$ mesure positive sur $C_0(x)$ telle que $\eta_t(C_0(x)) = \sum_{y \in C_0(x)} \eta_t(y) \xrightarrow[t\to\infty]{} 0$ et $\forall \theta \in \Theta'$ $\forall z, y \in C_0(x)$ $|P_{z,y}^{(td)}(\theta) - d\ \pi_y(\theta)| \leq \eta_t(y)$;

b) les familles de probabilités $[(d\ \pi_y(\theta) \ ; \ y \in C_0(x)) \ ; \ \theta \in \Theta_i]$ (i=0,1) sont décantées par un test φ ,

$$\omega_0 = \sup_{\theta\in\Theta_0} \sum_{y\in C_0(x)} d\varphi(y)\ \pi_y(\theta) < \inf_{\theta\in\Theta_1} \sum_{y\in C_0(x)} d\varphi(y)\ \pi_y(\theta) = \omega_1 \ .$$

Alors les hypothèses $(P_\theta \ ; \ \theta \in \Theta_i)$ (i=0,1) se séparent asymptotiquement uniformément.

Preuve : Passons de X à Y de la façon décrite ci-dessus. On voit comme dans la démonstration de la proposition V.3. que, si $\eta_t(C_o(x)) < \frac{\omega_1 - \omega_o}{2}$, t est un temps de décantation pour $(P^*(\theta)\ ;\ \theta \in \Theta_i)$ (i=0,1), ce qui permet de conclure, compte tenu de la proposition V.2. ■

V.7. - ESTIMATION DE CERTAINS PARAMETRES.-

V.7.1. - Introduction.

Soit X une chaîne de Markov homogène à valeurs dans un ensemble $\mathcal{X}$ dénombrable.

On appliquera ici les résultats présentés dans le chapitre II pour trouver des conditions suffisantes pour l'estimabilité de la matrice de transition de X et de sa distribution stationnaire, en supposant dans ce cas que X est irréductible récurrente positive apériodique.

A notre connaissance, le seul résultat de ce genre publié antérieurement est le suivant, dû à J. Geffroy ([24], p. 13). Il s'applique au cas de l'estimation de la distribution stationnaire, lorsque $\mathcal{X}$ est fini.

Théorème V.2. ([24]).- Soit $\{P_\theta\}_{\theta\in\Theta}$ un modèle statistique pour une chaîne de Markov homogène finie $\{X_n\}_{n\in\mathbb{N}}$. On suppose :

a) que Θ est un espace métrique σ-compact ;

b) que la matrice de transition $P(\theta)$ est une fonction continue de θ ;

c) que pour tout θ, le processus $\{X_n\}$ est irréductible apériodique ;

d) que sur toute partie compacte Q de Θ, la distribution stationnaire $\pi(\theta)$ du processus - qui existe en vertu de (c) - satisfait une relation de la forme

$$\forall(\theta,\theta') \in Q^2 \quad , \quad h[d(\theta,\theta')] \leqslant d_V[\pi(\theta),\pi(\theta')] \leqslant h'[d(\theta,\theta')]$$

où h et h' sont deux fonctions continues monotones non décroissantes telles que : $\forall d \in]0, \mathrm{diam}(Q)[$, $0 < h(d) < h'(d)$ et $\lim\limits_{d\to 0} h'(d) = \lim\limits_{d\to 0} h(d) = 0$.

Dans ces conditions, il existe un estimateur convergent de θ.

Pour le cas où X est infini dénombrable, on tiendra compte des considérations suivantes.

Il est connu que $\mathbb{R}^\infty$ est en général muni de la métrique

$$d_1 : (x,y) \longmapsto \sum_{i\geqslant 1} 2^{-i} \frac{|x_i-y_i|}{1+|x_i-y_i|} \quad (x = (x_i \ ; \ i \geqslant 1), \ y = (y_i \ ; \ i \geqslant 1) \in \mathbb{R}^\infty).$$

Considérons $[0,1]^\infty$ muni de

$$d^* : (x,y) \longmapsto \sum_{i\geqslant 1} 2^{-i} |x_i-y_i| \quad . \tag{13}$$

On voit aisément que d_1 et d^* sont équivalentes sur $[0,1]^\infty$, vu que l'on a alors : $\frac{1}{2} d^* \leqslant d_1 \leqslant d^*$.

Il est aussi connu que $[0,1]^\infty$ est un compact pour l'une ou l'autre de ces distances.

D'autre part, on sait que $\mathbb{N}^* \times \mathbb{N}^*$ est infini dénombrable. On voit par des méthodes élémentaires que si :

$$g : \mathbb{N}^* \times \mathbb{N}^* \longrightarrow \mathbb{N}^* : (m,n) \longrightarrow 2^{n-1}(2m-1) \ , \tag{14}$$

g est bijective. Si g_1 est une autre bijection de $\mathbb{N}^* \times \mathbb{N}^*$ dans $\mathbb{N}^*$, il existe une permutation p de $\mathbb{N}^* \times \mathbb{N}^*$ telle que $g_1 = g \circ p$. Si $\sum_{i,j\geq 1} x_{i,j}$ est une série réelle absolument convergente, on peut intervertir l'ordre de ses termes. Donc $\sum_{i,j\geq 1} x_{i,j} = \sum_{k\geq 1} x_{g^{-1}(k)} = \sum_{k\geq 1} x_{g_1^{-1}(k)}$. Dans la suite, on utilisera l'application g définie ci-dessus.

Enfin, si l'on considère $\mathbb{R}^s$ muni de

$$d_{2,s} : (x,y) \longmapsto \sum_{i=1}^{s} |x_i - y_i| , \tag{15}$$

on sait que $d_{2,s}$ est équivalente aux distances déduites des normes sur $\mathbb{R}^s$ (en particulier à la distance euclidienne) et induit donc la topologie usuelle de $\mathbb{R}^s$.

V.7.2. - Estimation de la matrice de transition d'une chaîne de Markov homogène.-

Proposition V.5.- Soit $X = (X_n \; ; \; n \geq 1)$ une suite d'observations à valeurs dans un ensemble fini $\mathcal{X} = \{x_1,\ldots,x_\nu\}$, de loi inconnue, Soit $\mathcal{B} = \mathcal{P}(\mathcal{X})$. On suppose que le modèle statistique $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta \; ; \; \theta \in \Theta))$ associée à X vérifie les propriétés suivantes :

a) $\forall \theta \in \Theta$ X est, sous l'hypothèse P_θ , une chaîne de Markov homogène à laquelle on peut associer une matrice de transition, notée $P_{\cdot,\cdot}(\theta)$, identifiée à l'élément suivant de $[0,1]^{\nu^2}$, métrisé par $d = d_{2,\nu^2}$ (voir (15)) et muni donc de la topologie usuelle :

$$(P_{x_1,x_1}(\theta), P_{x_1,x_2}(\theta) ,\ldots, P_{x_1,x_\nu}(\theta), P_{x_2,x_1}(\theta) ,\ldots, P_{x_\nu,x_\nu}(\theta)) .$$

b) $\theta, \theta' \in \Theta$, $\theta \neq \theta' \Longrightarrow \forall x \in \mathcal{X}$ $P_{x,\cdot}(\theta) \neq P_{x,\cdot}(\theta')$.

Alors la matrice de transition de X est p.-s. estimable.

Preuve : i) Posons, pour $T \in \mathbb{N}^*$,

$$\Theta_T^* = \{(\theta,\theta') \in \Theta^2 : \forall x \in X \quad d_V(P_{x,.}(\theta) \ , \ P_{x,.}(\theta')) \geq \frac{1}{T}\}, \tag{16}$$

$$\Theta_T = \{\theta \in \Theta : \exists \theta' \in \Theta \quad \text{vérifiant} \quad (\theta,\theta') \in \Theta_T^*\}.$$

Soit $T_o = \inf\{T \in \mathbb{N}^* : \Theta_T = \emptyset\}$. On a $\Theta = \bigcup_{T \geq T_o} \Theta_T = \lim_{T\to\infty} \uparrow \Theta_T$.

Pour $T \geq T_o$, notons E_T l'adhérence de $P_{.,.}(\Theta_T)$. Vu que la convergence selon d implique la convergence coordonnée à coordonnée, notons $(e^1_{1,1}, e^1_{1,2}, \dots, e^1_{1,\nu}, e^1_{2,1}, \dots, e^1_{\nu,\nu})$ un élément e^1 de E_T .
On prolonge d à E_T en posant $\forall e^o, e^1 \in E_T \quad d(e^o,e^1) = \sum_{i,j=1}^{\nu} |e^o_{i,j} - e^1_{i,j}|$.
D'après (16), $\forall e^o, e^1 \in E_T$, $e^o \neq e^1$

$$d(e^o,e^1) \geq \frac{1}{T} \quad (T \geq T_o) \ . \tag{17}$$

D'après la proposition II.1.b) et le lemme II.3., une condition suffisante pour que la matrice de transition de X soit p.-s. estimable est que, $\forall T \geq T_o \quad \forall e^o, e^1 \in E_T$, $e^o \neq e^1 \quad \exists \alpha > 0$ tel que, dans le modèle statistique réduit $(X^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta \ ; \ \theta \in \Theta_T))$ les deux hypothèses $(P_\theta \ ; \ \theta \in \Theta_T \ , \ d(P_{.,.}(\theta), e^i) \leq \alpha) \ (i=0,1)$ se séparent asymptotiquement uniformément. D'après la proposition V.1., il suffit pour cela que les familles de matrices de transition associées aux hypothèses mentionnées soient décantées.

ii) Soient $\theta_o, \theta_1 \in \Theta_T$, $\theta_o \neq \theta_1$ tels que $d(P_{.,.}(\theta_i) \ , \ e^i) \leq \frac{1}{16\nu T}$ (i=0,1).

$\forall \theta \in \Theta \quad \forall x \in X \quad \forall A \subset X$ notons $P_{x,A}(\theta) = \sum_{y \in A} P_{x,y}(\theta)$.

D'après (16), $\forall x \in X \quad \exists A_x \in \mathcal{B} : P_{x,A_x}(\theta_o) + \frac{1}{T} \leq P_{x,A_x}(\theta_1)$.

Montrons que $H_i = (P_\theta \ ; \ \theta \in \Theta_T \ , \ d(P_{.,.}(\theta), e^i) \leq \frac{1}{16\nu T})$ $(i=0,1)$ se séparent asymptotiquement uniformément.

Si $P_\theta \in H_o$, $d(P_{.,.}(\theta), P_{.,.}(\theta_o)) \leq \frac{1}{8\nu T}$. Donc $\forall x \in X$ $d_V(P_{x,.}(\theta), P_{x,.}(\theta_o)) \leq \frac{1}{4\nu T}$ et ceci implique

$$\forall x \in X \qquad P_{x,A_x}(\theta) \leq P_{x,A_x}(\theta_o) + \frac{1}{4\nu T} \ .$$

On voit de même que si $P_{\theta'} \in H_1$ alors

$$\forall x \in X \qquad P_{x,A_x}(\theta') \geq P_{x,A_x}(\theta_1) - \frac{1}{4\nu T} \ .$$

Donc $\forall x \in X$ $P_{x,A_x}(\theta') - P_{x,A_x}(\theta) \geq \frac{1}{T} - \frac{1}{2\nu T} = \frac{2\nu-1}{2\nu T} > 0$ ce qui, compte tenu de (7), permet de conclure. ■

En passant, et vu la proposition II.1. a), on a démontré que :

Corollaire 1.- Si $\exists \ k \in \mathbb{N}^*$ tel que $\Theta = \Theta_k$, alors la matrice de transition de X est uniformément p.-s. estimable.

La condition de ce corollaire est vérifiée, notamment, d'après la propriété archimédienne de $\mathbb{R}$, si $\inf\limits_{\theta,\theta' \in \Theta, \theta \neq \theta'} \ \min\limits_{x \in X} d_V(P_{x,.}(\theta), P_{x,.}(\theta')) > 0.$

Proposition V.6.- Soit $X = (X_n \ ; \ n \geq 1)$ une suite d'observations dans les conditions générales de la proposition V.5.

On suppose que le modèle $(X^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta \ ; \ \theta \in \Theta))$ associé à X vérifie, en plus de la propriété a) de la proposition V.5., la propriété suivante :

b) $\exists f : \mathbb{R}_+^* \longrightarrow \mathbb{R}_+^*$ telle que $\forall \theta, \theta' \in \Theta \quad \forall \varepsilon > 0$

$$d(P_{.,.}(\theta), P_{.,.}(\theta')) \geq \varepsilon \implies \min_{x\in} \sum_{y\in} \frac{1}{2} |P_{x,y}(\theta) - P_{x,y}(\theta')| \geq f(\varepsilon) .$$

Alors la matrice de transition de X est uniformément p.-s. estimable.

Preuve : D'après la proposition II.1. a) et le lemme II.3., et puisque $([0,1]^{\nu^2}, d)$ est compact, il suffit de vérifier que $\forall e_o, e_1 \in [0,1]^{\nu^2}, e_o \neq e_1 \quad \exists \alpha > 0$ tel que les deux hypothèses $H_i = (P_\theta \; ; \theta \in \Theta \quad , \quad d\ (P_{.,.}(\theta), e_i) \leq \alpha)$ (i=0,1) se séparent asymptotiquement uniformément.

Si e_o (ou e_1) n'est pas adhérent à $P_{.,.}(\Theta)$, il suffit de prendre $\alpha \in \mathbb{R}_+^*$ assez petit pour que H_o (ou H_1) soit vide, auquel cas la séparation asymptotique uniforme de H_o et H_1 est trivialement vérifiée.

Si $e_i \in \overline{P_{.,.}(\Theta)}$ (i=0,1), alors $\forall \varepsilon \in]0, d\ (e_o, e_1)[$ on peut trouver $\theta_i \in \Theta$ (i=0,1) de manière à avoir simultanément

i) $d(P_{.,.}(\theta_i), e_i) < \frac{1}{6} f(\varepsilon)$ (i=0,1)

ii) $d(P_{.,.}(\theta_o), P_{.,.}(\theta_1)) > \varepsilon$.

D'après b) et ii), $\min_{x\in X} \sum_{y\in X} \frac{1}{2} |P_{x,y}(\theta_o) - P_{x,y}(\theta_1)| \geq f(\varepsilon)$,

soit $\min_{x\in X} d_V(P_{x,.}(\theta_o), P_{x,.}(\theta_1)) \geq f(\varepsilon)$. D'où $\forall x \in X \quad \exists A_x \in \mathcal{B}$ tel que $P_{x,A_x}(\theta_o) + f(\varepsilon) \leq P_{x,A_x}(\theta_1)$.

Posons $H_i' = (P_\theta \; ; \theta \in \Theta \; , d(P_{.,.}(\theta), e_i) \leq \frac{1}{6} f(\varepsilon))$ (i=0,1) .

Pour voir que H'_o et H'_1 se séparent asymptotiquement uniformément, il suffit de vérifier, d'après la proposition V.1., que les familles de matrices de transition correspondantes sont décantées. Or, si θ est tel que

$d(P_{\cdot,\cdot}(\theta), e_o) \leq \frac{1}{6} f(\varepsilon)$, on a

$d(P_{\cdot,\cdot}(\theta), P_{\cdot,\cdot}(\theta_o)) \leq \frac{1}{3} f(\varepsilon) \Longrightarrow \forall x \in X \;\; d_V(P_{x,\cdot}(\theta), P_{x,\cdot}(\theta_o)) \leq \frac{1}{6} f(\varepsilon) \Longrightarrow$

$\Longrightarrow \forall x \in X \qquad P_{x,A_x}(\theta) \leq P_{x,A_x}(\theta_o) + \frac{1}{6} f(\varepsilon)$.

On voit de même que si θ' est tel que $d\ (P_{\cdot,\cdot}(\theta'), e_1) \leq \frac{1}{6} f(\varepsilon)$ alors $\forall x \in X \quad P_{x,A_x}(\theta') \geq P_{x,A_x}(\theta_1) - \frac{1}{6} f(\varepsilon)$.

Il en résulte, pour θ et θ' dans ces conditions, que $\forall x \in X$ $P_{x,A_x}(\theta) \leq P_{x,A_x}(\theta_o) + \frac{1}{6} f(\varepsilon) \leq P_{x,A_x}(\theta_1) - \frac{5}{6} f(\varepsilon) < P_{x,A_x}(\theta_1) - \frac{1}{6} f(\varepsilon) \leq P_{x,A_x}(\theta')$, ce qui, compte tenu de (7), permet de conclure. ■

Corollaire 1.- La proposition V.6. reste exacte si b) est remplacée par les conditions suivantes :

$b_1)$ $\quad \theta, \theta' \in \Theta$, $\theta \neq \theta' \Longrightarrow \forall x \in X \;\; P_{x,\cdot}(\theta) \neq P_{x,\cdot}(\theta')$.

$c_1)$ Θ est muni d'une topologie qui en fait un espace compact et pour laquelle $\forall x$, $y \in X$ la fonction $\Theta \longrightarrow (\mathbb{R}\ , |.|) : \theta \longmapsto P_{x,y}(\theta)$ est continue.

Preuve : La fonction

$h : \Theta^2 \longrightarrow (\mathbb{R}, |.|) : (\theta,\theta') \longmapsto \min\limits_{x \in X} \sum\limits_{y \in X} \frac{1}{2} (P_{x,y}(\theta) - P_{x,y}(\theta'))$ est continue sur Θ^2 pour la topologie produit.

De plus, $\forall \varepsilon > 0 \quad \{(\theta,\theta') \in \Theta^2 : d(P_{\cdot,\cdot}(\theta), P_{\cdot,\cdot}(\theta')) \geq \varepsilon\}$ est vide ou est, sinon, une partie compacte de Θ^2 sur laquelle h ne prend que des valeurs strictement positives. On pose $f(\varepsilon) = 1$ dans le premier cas et $f(\varepsilon) = \inf \{h(\theta,\theta') \; ; \; \theta, \theta' \in \Theta^2 \, , \, d(P_{.,.}(\theta), P_{.,.}(\theta')) \geq \varepsilon\}$, dans le second. ∎

Pour le cas où X est infini dénombrable, on peut présenter les résultats suivants, où les hypothèses sont naturellement plus contraignantes.

Proposition V.7.- Soit $X = (X_n \; ; \; n \geq 1)$ une suite d'observations à valuers dans un ensemble infini dénombrable $\mathcal{X} = \{x_i \; ; \; i \geq 1\}$, de loi inconnue. Soit $\mathcal{B} = \mathcal{P}(\mathcal{X})$. On suppose que le modèle statistique $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta \; ; \; \theta \in \Theta))$ associé à X vérifie les propriétés suivantes :

a) $\forall \theta \in \Theta$ X est, sous l'hypothèse P_θ , une chaîne de Markov homogène à laquelle on peut associer une matrice de transition, notée $P_{.,.}(\theta)$, identifiée à l'élément suivant de $[0,1]^\infty$:

$$\begin{matrix} (P_{x_1,x_1}(\theta), & P_{x_1,x_2}(\theta) & ,\dots \\ P_{x_2,x_1}(\theta), & P_{x_2,x_2}(\theta) & ,\dots \\ \dots\dots & &) \; . \end{matrix}$$

b) $\forall \theta \, , \, \theta' \in \Theta \, , \, \theta \neq \theta' \quad \inf\limits_{x \in \mathcal{X}} d_V(P_{x,.}(\theta), P_{x,.}(\theta')) > 0$

On munit $P_{.,.}(\Theta)$ de la distance

$$d : (P_{.,.}(\theta), P_{.,.}(\theta') \longmapsto \sup_{x \in \mathcal{X}} d_V(P_{x,.}(\theta), P_{x,.}(\theta')) \; .$$

On admet de plus :

c) $\forall \theta \in \Theta \;\; \exists M(\theta) \in \mathbb{N}^* \;$ tel que $\forall i \geq 1 \;\; \forall n > M(\theta) \;\; P_{x_i,x_n}(\theta) = 0$.

Alors la matrice de transition de X est p.-s. estimable.

Preuve : i) $\forall \theta \, , \, \theta' \in \Theta \, , \, \theta \neq \theta'$ posons $M(\theta,\theta') = \max\{M(\theta), M(\theta')\}$. Soit $T \in \mathbb{N}^*$. Soit

$$\left.\begin{aligned} &\Theta^*_T = \{(\theta,\theta') \in \Theta^2 : M(\theta,\theta') \leq T \text{ et } \forall i \geq 1 \quad d_V(P_{x_{i,.}}(\theta), P_{x_{i,.}}(\theta')) \geq \tfrac{1}{T}\}, \\ &\Theta_T = \{\theta \in \Theta : \exists \theta' \in \Theta \text{ vérifiant } (\theta,\theta') \in \Theta^*_T\} \quad . \end{aligned}\right| \quad (18)$$

Soit $T_o = \inf\{T \in \mathbb{N}^* : \Theta_T \neq \emptyset\}$. En vertu de b) et c), $T_o \in \mathbb{N}^*$ et on a évidemment $\Theta = \bigcup\limits_{T \geq T_o} \Theta_T = \lim\limits_{T\to\infty} \uparrow \Theta_T$.

ii) $\forall T \geq T_o$ considérons le modèle statistique réduit $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta \; ; \; \theta \in \Theta_T))$. On va voir que, sous les hypothèses de l'énoncé, d est équivalente à d^* (voir (13)) sur $P_{.,.}(\Theta_T)$. Pour cela, on rappelle la fonction g définie par (14), et on associera aux éléments des matrices dans $P_{.,.}(\Theta_T)$ des indices dans $\mathbb{N}^*$, de la façon suivante :

$P_{x_i,x_j}(\theta) = y_{g(i,j)}(\theta)$. On a alors

$$d^*(P_{.,.}(\theta), P_{.,.}(\theta')) = \sum_{i,j\geq 1} 2^{-g(i,j)} |y_{g(i,j)}(\theta) - y_{g(i,j)}(\theta')| \quad \text{et}$$

$$\forall \theta , \theta' \in \Theta_T , \theta \neq \theta' \quad 2d(P_{.,.}(\theta), P_{.,.}(\theta')) = 2d(P_{.,.}(\theta), P_{.,.}(\theta')) \sum_{n\geq 1} 2^{-n} =$$

$$= \sum_{i,j\geq 1} 2^{-2^{j-1}(2i-1)} . 2d(P_{.,.}(\theta), P_{.,.}(\theta')) \geq$$

$$\geq \sum_{i,j\geq 1} 2^{-2^{j-1}(2i-1)} |P_{x_i,x_j}(\theta) - P_{x_i,x_j}(\theta')| =$$

$$= \sum_{i,j\geq 1} 2^{-g(i,j)} |y_{g(i,j)}(\theta) - y_{g(i,j)}(\theta')| = d^*(P_{.,.}(\theta), P_{.,.}(\theta')) .$$

D'autre part, d'après c) et vu (18), $\forall \theta , \theta' \in \Theta_T , \theta \neq \theta'$

$$d^*(P_{.,.}(\theta), P_{.,.}(\theta')) = \sum_{i\geq 1} \sum_{j=1}^{M(\theta,\theta')} 2^{-2^{j-1}(2i-1)} |P_{x_i,x_j}(\theta)-P_{x_i,x_j}(\theta')| \geq$$

$$\geq \sum_{i\geq 1} \sum_{j=1}^{M(\theta,\theta')} 2^{-2^{T-1}(2i-1)} |P_{x_i,x_j}(\theta)-P_{x_i,x_j}(\theta')| =$$

$$= \sum_{i\geq 1} 2^{-2^{T-1}(2i-1)} \cdot 2\, d_V(P_{x_i,.}(\theta), P_{x_i,.}(\theta')) \geq \frac{2}{T} \sum_{i\geq 1} 2^{-2^{T-1}(2i-1)} =$$

$$= \frac{2^{(2^T)}}{T} \sum_{i\geq 1} 2^{-i2^T} = \frac{1}{T(1-2^{-2^T})} \geq \frac{1}{T(1-2^{-2^T})}\, d(P_{.,.}(\theta), P_{.,.}(\theta'))\ , \quad \text{puisque}\ \ d \leq 1.$$

Donc, l'adhérence par rapport à d^* (ou d, vu que la convergence selon d est celle de la convergence coordonnée à coordonnée) de $P_{.,.}(\Theta_T)$ est un compact de $[0,1]^\infty$, noté E_T.

iii) $\forall T \geq T_o$ notons $\begin{pmatrix} e^1_{1,1}, & e^1_{1,2}, & \dots \\ e^1_{2,1}, & e^1_{2,2}, & \dots \\ \dots\dots\dots & & \end{pmatrix}$ un élément

e^1 de E_T. On peut prolonger d à E_T en posant
$d(e^o,e^1) = \sup\limits_{i\geq 1} \sum\limits_{j\geq 1} \frac{1}{2} |e^o_{i,j} - e^1_{i,j}|$.

Pour que la matrice de transition de X soit p.-s. estimable il suffit de vérifier, comme dans la preuve de la proposition V.5., que $\forall e^o, e^1 \in E_T$, $e^o \neq e^1$ $\exists \alpha > 0$ tel que dans le modèle statistique réduit $(X^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta\ ;\ \theta \in \Theta_T))$ les familles de matrices de transition associées aux hypothèses $H_i = (P_\theta\ ;\ \theta \in \Theta_T\ ,\ d(P_{.,.}(\theta), e^i) \leq \alpha)$ $(i=0,1)$ sont décantées. Posons $\alpha = \frac{1}{6T}$.

Soient θ_o, $\theta_1 \in \Theta_T$, $\theta_o \neq \theta_1$ tels que $d(P_{.,.}(\theta_i),e^i) \leq \frac{1}{6T}$.

Compte tenu de (18), $\forall x \in X$ $d_V(P_{x,.}(\theta_o), P_{x,.}(\theta_1)) \geq \frac{1}{T}$.

Donc $\forall x \in X$ $\exists A_x \in \mathcal{B}$ tel que $P_{x,A_x}(\theta_o) + \frac{1}{T} \leq P_{x,A_x}(\theta_1))$.

Si $P_\theta \in H_o$, $d(P_{.,.}(\theta), P_{.,.}(\theta_o)) \leq \frac{1}{3T}$. D'où $\forall x \in X$
$d_V(P_{x,.}(\theta), P_{x,.}(\theta_o)) \leq \frac{1}{3T}$ et $\forall x \in X$ $P_{x,A_x}(\theta) \leq P_{x,A_x}(\theta_o) + \frac{1}{3T}$.
On voit de même que si $P_{\theta'} \in H_1$, alors $\forall x \in X$ $P_{x,A_x}(\theta') \geq P_{x,A_x}(\theta_1) - \frac{1}{3T}$.
Donc $\forall x \in X$ $P_{x,A_x}(\theta') - P_{x,A_x}(\theta) \geq \frac{1}{T} - \frac{2}{3T} = \frac{1}{3T} > 0$, ce qu'il fallait démontrer.■

Corollaire 1.- Si $\exists k \in \mathbb{N}^*$ tel que $\Theta = \Theta_k$, alors la matrice de transition de X est uniformément p.-s. estimable.

Proposition V.8.- Soit $X = (X_n \; ; \; n \geq 1)$ une suite d'observations dans les conditions générales de la proposition V.7. On suppose que le modèle statistique $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta \; ; \; \theta \in \Theta))$ associé à X vérifie, en plus des conditions a) et b) de la proposition V.7., la propriété suivante :

c_1) $P_{.,.}(\Theta)$ est σ-précompact et fermé dans (B,d) , où

$$B = \left\{ b^o = \begin{pmatrix} b^o_{1,1} , b^o_{1,2} ,\dots \\ b^o_{2,1} , b^o_{2,2} ,\dots \\ \dots\dots\dots\dots \end{pmatrix} \in [0,1]^\infty : \forall i \geq 1 \sum_{j\geq 1} b^o_{i,j} = 1 \right\} \quad \text{et}$$

$$d(b^o,b^1) = \sup_{i\geq 1} \frac{1}{2} \sum_{j\geq 1} |b^o_{i,j} - b^1_{i,j}| \quad (b^o, b^1 \in B) .$$

Alors la matrice de transition de X est p.-s. estimable.

Preuve : On voit aisément que (B,d) est complet.

$F = (P_{.,.}(\Theta), d)$ étant σ-précompact et complet (puisque d'après c_1) il s'agit d'un fermé de B), d'après la proposition II.1.b) et le lemme II.3. il suffit de vérifier que pour toute partie précompacte E^* de F on ait : $\forall e_o, e_1 \in E^*$, $e_o \neq e_1$ $\exists \alpha > 0$ tel que les deux hypothèses $(P_\theta \; ; \; \theta \in \Theta$ et $P_{.,.}(\theta) \in E^* \cap B(e_i,\alpha))$ $(i=0,1)$ se séparent asymptotiquement uniformément.

On démontre la propriété plus forte suivante :

$\forall\theta_o$, $\theta_1 \in \Theta$, $\theta_o \neq \theta_1$ $\exists\alpha > 0$ tel que les deux hypothèses

$$H_i = (P_\theta \; ; \; \theta \in \Theta \; , \; d(P_{.,.}(\theta), P_{.,.}(\theta_i)) \leq \alpha) \; (i=0,1) \tag{19}$$

se séparent asymptotiquement uniformément.

D'après b), $\theta_o \neq \theta_1 \Longrightarrow \frac{1}{2} \inf_{x \in X} \sum_{y \in X} |P_{x,y}(\theta_o) - P_{x,y}(\theta_1)| = a > 0$.

D'où $\forall x \in X$ $\exists A_x \in \mathcal{B}$ tel que $P_{x,A_x}(\theta_o) + a \leq P_{x,A_x}(\theta_1)$.

Posons, dans (19), $\alpha = \frac{a}{3}$. On voit selon le procédé habituel que, si $P_\theta \in H_o$, alors $\forall x \in X$ $P_{x,A_x}(\theta) \leq P_{x,A_x}(\theta_o) + \frac{a}{3}$ et que, si $P_{\theta'} \in H_1$, alors $\forall x \in X$ $P_{x,A_x}(\theta') \geq P_{x,A_x}(\theta_1) - \frac{a}{3}$.

D'où $\forall x \in X$ $P_{x,A_x}(\theta) \leq P_{x,A_x}(\theta_1) - \frac{2a}{3} < P_{x,A_x}(\theta_1) - \frac{a}{3} \leq P_{x,A_x}(\theta')$,

ce qui permet de conclure. ■

Remarque V.5.- On a déjà dit que la matrice de transition de X est unique si, X étant définie sur un espace probabilisé (Ω, A, Pr), $\forall x \in X$ $\exists n \in \mathbb{N}^*$ tel que $Pr(X_n=x) > 0$. En particulier, si X est irréductible, cette condition est vérifiée, mais en toute généralité la réciproque n'a pas lieu.

V.7.3. - Estimation de la distribution stationnaire d'une chaîne de Markov homogène irréductible récurrente positive apériodique.

Dans les deux propositions suivantes, on tiendra compte du rappel V.7. (section V.4.).

D'autre part, signalons que si $X = \{x_1,\dots,x_\nu\}$ et si π^o (resp. π^1) est une probabilité sur $\mathcal{B} = \mathcal{P}(X)$ définie par les masses ponctuelles π^o_x $(x \in X)$ (resp. π^1_x $(x \in X)$) et identifiée à l'élément $(\pi^o_{x_1}, \pi^o_{x_2}, \dots, \pi^o_{x_\nu})$ (resp. $(\pi^1_{x_1}, \pi^1_{x_2}, \dots, \pi^1_{x_\nu})$) de $[0,1]^\nu$, on a

$d_V(\pi^o,\pi^1) = \frac{1}{2} \sum_{i=1}^{\nu} |\pi^o_{x_i} - \pi^1_{x_i}| = \frac{1}{2} d_{2,\nu}(\pi^o,\pi^1)$, où $d_{2,\nu}$ est définie par (15).

Il est donc naturel de métriser l'ensemble des lois sur $\mathcal{B}$ par d_V , vu que cette distance induit la topologie usuelle de $[0,1]^\nu$.

Proposition V.9.- Soit $X = (X_n \ ; n \geq 1)$ une suite d'observations à valeurs dans $X = \{x_1,\dots,x_\nu\}$, de loi inconnue. Soit $\mathcal{B} = \mathcal{P}(X)$. On suppose que le modèle statistique $(X^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta \ ; \theta \in \Theta))$ associée à X vérifie :

a) $\forall \theta \in \Theta$ X est, sous l'hypothèse P_θ , une chaîne de Markov homogène irréductible apériodique. On note $\pi(\theta)$ (resp. $P_{.,.}(\theta)$) la distribution stationnaire (resp. la matrice de transition) correspondante.

b) $\forall \theta_o , \theta_1 \in \Theta$, $\theta_o \neq \theta_1$ $\pi(\theta_o) \neq \pi(\theta_1)$.

On munit $\pi(\Theta)$ de la distance $d = d_V$.

Alors la distribution stationnaire de X est p.-s. estimable.

Preuve : i) $\forall \theta \in \Theta$ $\forall x, y \in X$ $\exists T(x,y,\theta) \in \mathbb{N}^*$ tel que $t \geq T(x,y,\theta) \Longrightarrow P^{(t)}_{x,y}(\theta) \geq \frac{1}{2} \pi_y(\theta) > 0$, vu que $P^{(t)}_{x,y}(\theta) \xrightarrow[t\to\infty]{} \pi_y(\theta)$.

Comme X est fini, $\forall \theta \in \Theta$ $\exists T(\theta) \in \mathbb{N}^*$ tel que $t \geq T(\theta) \Longrightarrow \forall x, y \in X$ $P^{(t)}_{x,y}(\theta) \geq \frac{1}{2} \pi_y(\theta)$.

D'après le théorème V.1.,

$$\forall\theta \in \Theta \quad \forall t > T(\theta) \quad \forall x, y \in X \quad |P_{x,y}^{(t)}(\theta) - \pi_y(\theta)| \leqslant 2\left(\frac{1}{2}\right)^{\frac{t}{T(\theta)}} .$$

$\forall T \in \mathbb{N}^*$ posons $\Theta_T = \{\theta \in \Theta : T(\theta) \leqslant T\}$. Soit $T_o = \inf(T \in \mathbb{N}^* : \Theta_T \neq \emptyset)$. Il est clair que $\Theta = \bigcup_{T \geq T_o} \Theta_T = \lim_{T \to \infty} \uparrow \Theta_T$.

ii) $\forall T \geqslant T_o$ nous considérons le modèle statistique réduit $(X^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta ; \theta \in \Theta_T))$ et nous allons montrer que dans ce modèle $\pi(\theta)$ est uniformément p.-s. estimable. Compte tenu du fait que $\pi(\Theta)$ (et donc $\pi(\Theta_T)$) est précompact et de la proposition II.1.a), il suffit de vérifier que :

$\forall\theta_o, \theta_1 \in \Theta_T \quad \forall\alpha_o, \alpha_1 > 0$ tels que $d(\pi(\theta_o), \pi(\theta_1)) > \alpha_o + \alpha_1$ les hypothèses $H_i = (P_\theta ; \theta \in \Theta_T, d(\pi(\theta), \pi(\theta_i)) \leqslant \alpha_i)$ $(i=0,1)$ se séparent asymptotiquement uniformément.

iii) La séparation asymptotique uniforme de H_o et H_1 sera démontrée à l'aide de la proposition V.3.

Notons d'abord que l'hypothèse a) de cette proposition est satisfaite, parce que X est fini (voir remarque V.3., section V.5.) et que nous avons

$$\forall\theta \in \Theta_T \quad \forall t > T \quad \forall x, y \in X \quad |P_{x,y}^{(t)}(\theta) - \pi_y(\theta)| \leqslant 2\left(\frac{1}{2}\right)^{\frac{t}{T}} ,$$

puisque $t > T \Longrightarrow t > T(\theta) \Longrightarrow |P_{x,y}^{(t)}(\theta) - \pi_y(\theta)| \leqslant 2\left(\frac{1}{2}\right)^{\frac{t}{T(\theta)}} \leqslant 2\left(\frac{1}{2}\right)^{\frac{t}{T}}$.

iv) Il reste à vérifier la seconde hypothèse de la proposition V.3., c'est-à-dire démontrer que $(\pi(\theta) ; \theta \in \Theta_T, d(\pi(\theta), \pi(\theta_i)) \leqslant \alpha_i)$ $(i=0,1)$ sont décantées. Ceci est une conséquence immédiate du fait que ces deux familles de lois sur $\mathcal{B}$ sont respectivement contenues dans les boules fermées $B_i = B(\pi(\theta_i), \alpha_i)$ $(i=0,1)$ de l'ensemble de toutes les lois sur $\mathcal{B}$

muni de la distance d_V . Or on sait (cf. [39], p. 67) que la condition $d_V(\pi(\theta_o), \pi(\theta_1)) > \alpha_o + \alpha_1$ assure la décantation de ces boules par un événement, ce qui veut dire que $\exists A \in \mathcal{B}$ tel que $\inf([P(A)-Q(A)] \; ; (P,Q) \in B_o \times B_1) > 0$.

Plus précisément, si $\forall A \in \mathcal{B}$ on pose $\pi_A(\theta) = \sum_{y \in A} \pi_y(\theta)$, on a (cf. [39], p. 68) :

$$\left.\begin{aligned} & d_V(\pi(\theta_o), \pi(\theta_1)) = \pi_A(\theta_o) - \pi_A(\theta_1) \Longrightarrow \\ & \inf_{(\pi(\theta),\pi(\theta'))\in B_o \times B_1} [\pi_A(\theta)-\pi_A(\theta')] \geq d_V(\pi(\theta_o),\pi(\theta_1))-(\alpha_o+\alpha_1) > 0 \; . \end{aligned}\right| \quad (20)$$

D'après les propositions V.3. et II.1. a), la distribution stationnaire de X est donc uniformément p.-s. estimable pour le modèle statistique réduit $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta \; ; \theta \in \Theta_T))$. Comme $\Theta = \lim_{T\to\infty} \uparrow \Theta_T$, on en déduit, d'après la proposition II.1. b), que la distribution stationnaire de X est p.-s. estimable pour le modèle initial. ■

Corollaire 1.- Si $\exists k \in \mathbb{N}^*$ tel que $\Theta = \Theta_k$, la distribution stationnaire de X est uniformément p.-s. estimable.

Proposition V.10.- Dans les conditions de la proposition V.9., si $\pi(\Theta)$ est fermé et si $\exists f : \mathbb{N}^* \to \mathbb{R}_+^*$ telle que $f(t) \xrightarrow[t\to\infty]{} 0$ et $\forall \theta \in \Theta \quad \forall x, y \in \mathcal{X} \quad |P_{x,y}^{(t)}(\theta) - \pi_y(\theta)| \leq f(t)$, alors la distribution stationnaire de X est uniformément p.-s. estimable.

Preuve : $\pi(\Theta)$ étant un sous-ensemble précompact de $([0,1]^\nu, d)$, il est compact dès que fermé.

Soient θ_o , $\theta_1 \in \Theta$, $\theta_o \neq \theta_1$. D'après b), $\exists x_o \in \mathcal{X}$ tel que $\pi_{x_o}(\theta_1) - \pi_{x_o}(\theta_o) > 0$.

$(\pi(\Theta),d)$ étant complet, pour que la distribution stationnaire de X soit uniformément p.-s. estimable il suffit que $\forall e_o = \pi(\theta_o)$, $e_1 = \pi(\theta_1) \in \pi(\Theta)$, $e_o \neq e_1$ $\exists \alpha > 0$ tel que les hypothèses $H_i = (P_\theta \; ; \; \theta \in \Theta \, , \, d(\pi(\theta), e_i) \leq \alpha)$ $(i=0,1)$ se séparent asymptotiquement uniformément. Si $\pi_{x_o}(\theta_1) - \pi_{x_o}(\theta_o) = a > 0$, posons $\alpha = \frac{a}{4}$.

Si $P_\theta \in H_o$, on a $d_V(\pi(\theta), \pi(\theta_o)) \leq \frac{a}{4}$ et $\forall y \in \mathcal{X}$ $\pi_y(\theta) \leq \pi_y(\theta_o) + \frac{a}{4}$.

Donc $\forall x, y \in \mathcal{X}$ $P^{(t)}_{x,y}(\theta) \leq \pi_y(\theta) + f(t) \leq \pi_y(\theta_o) + \frac{a}{4} + f(t)$.

De même, si $P_{\theta'} \in H_1$ $\forall x, y \in \mathcal{X}$ $P^{(t)}_{x,y}(\theta) \geq \pi_y(\theta_1) - \frac{a}{4} - f(t)$.

Donc $\forall x \in \mathcal{X}$ $P^{(t)}_{x,x_o}(\theta') - P^{(t)}_{x,x_o}(\theta) \geq \frac{a}{2} - 2\, f(t) > 0$ dès que $f(t) < \frac{a}{4}$. Si t_o vérifie $f(t_o) < \frac{a}{4}$, alors t_o est un temps de décantation pour $(P_{.,.}(\theta); P_\theta \in H_i)$ $(i=0,1)$, et la proposition V.2. nous permet de conclure. ■

Les résultats suivants concernent le cas où $\mathcal{X} = \{x_i \; ; \; i \geq 1\}$ est infini dénombrable. On identifiera une probabilité π sur $\mathcal{P}(\mathcal{X})$ à l'élément $(\pi_{x_1}, \pi_{x_2}, \ldots)$ de $[0,1]^\infty$, où $\pi_x = \pi(\{x\})$ $(x \in \mathcal{X})$.

Proposition V.11.- Soit $X = (X_n \; ; \; n \geq 1)$ une suite d'observations à valeurs dans un ensemble infini dénombrable $\mathcal{X} = (x_i \; ; \; i \geq 1)$, de loi inconnue. Soit $\mathcal{B} = \mathcal{P}(\mathcal{X})$. Supposons que le modèle $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta \; ; \; \theta \in \Theta))$ associé à X vérifie :

a) $\forall \theta \in \Theta$ X est, sous l'hypothèse P_θ, une chaîne de Markov homogène irréductible récurrente positive. On notera $\pi(\theta)$ (resp, $P_{.,.}(\theta)$) la distribution stationnaire (resp. la matrice de transition) correspondante.

b) θ , $\theta' \in \Theta$, $\theta \neq \theta' \Longrightarrow \pi(\theta) \neq \pi(\theta')$.

c) $\exists f : \mathbb{N}^* \longrightarrow \mathbb{R}_+^*$ telle que $f(t) \xrightarrow[t\to\infty]{} 0$ et $\forall \theta \in \Theta$ $\forall x, y \in X$

$$|P_{x,y}^{(t)}(\theta) - \pi_y(\theta)| \leqslant f(t) .$$

d) $\forall \theta, \theta' \in \Theta$, $\theta \neq \theta'$ $A_{(\theta,\theta')} = \{x \in X : \pi_x(\theta) > \pi_x(\theta')\}$

ou $A_{(\theta',\theta)} = \{x \in X : \pi_x(\theta') > \pi_x(\theta)\}$ est fini . Si $A_{(\theta,\theta')}$ (resp. $A_{(\theta',\theta)}$) est fini, notons $M_{(\theta,\theta')}$ (resp. $M_{(\theta',\theta)}$) l'entier $\max\{i \in \mathbb{N}^* : x_i \in A_{(\theta,\theta')}\}$ (resp. $\max\{i \in \mathbb{N}^* : x_i \in A_{(\theta',\theta)}\}$) . Posons, alors ,

$$M(\theta,\theta') = \begin{cases} M_{(\theta,\theta')} \text{ , si } M_{(\theta',\theta)} & \text{n'est pas défini .} \\ M_{(\theta',\theta)} \text{ , si } M_{(\theta,\theta')} & \text{n'est pas défini .} \\ \inf\{M_{(\theta,\theta')}, M_{(\theta',\theta)}\} \text{ ,} & \text{dans les autres cas .} \end{cases}$$

e) $\sup\limits_{\theta,\theta' \in \Theta, \theta\neq\theta'} M(\theta,\theta') = N < \infty$.

On munit $\pi(\Theta)$ de $d = d_V$.

Alors la distribution stationnaire de X est p.-s. estimable.

<u>*Preuve*</u> : i) Considérons $[0,1]^\infty$ muni de d^* (cf. (13)). On va voir que, sous les hypothèses ci-dessus, d^* est équivalente à d sur $\pi(\Theta)$.

Il est évident que $d^* \leqslant d$. D'autre part, si θ , $\theta' \in \Theta$, $\theta \neq \theta'$, supposons $M(\theta,\theta') = M_{(\theta,\theta')}$, le cas où $M(\theta,\theta') = M_{(\theta',\theta)}$ pouvant se traiter de façon tout à fait analogue. Alors, $d(\pi(\theta), \pi(\theta')) =$

$$= \sum_{x\in A_{(\theta,\theta')}} (\pi_x(\theta)-\pi_x(\theta')) = \sum_{x_i\in A_{(\theta,\theta')}} 2^{M(\theta,\theta')}\, 2^{-M(\theta,\theta')} |\pi_{x_i}(\theta)-\pi_{x_i}(\theta')| \leqslant$$

$$\leqslant \sum_{x_i\in A_{(\theta,\theta')}} 2^{M(\theta,\theta')}\, 2^{-i} |\pi_{x_i}(\theta)-\pi_{x_i}(\theta')| \leqslant 2^N d^*(\pi(\theta), \pi(\theta')) .$$

ii) Soit $T \in \mathbb{N}^*$, $\Theta_T^* = \{(\theta,\theta') \in \Theta^2 : d(\pi(\theta), \pi(\theta')) \geq \frac{1}{T}\}$,
$\Theta_T = \{\theta \in \Theta : \exists \theta' \in \Theta$ vérifiant $(\theta,\theta') \in \Theta_T^*\}$,
$T_o = \inf\{T \in \mathbb{N}^* : \Theta_T \neq \emptyset\}$. D'après b), $T_o \in \mathbb{N}^*$ et il est clair que $\Theta = \bigcup_{T \geq T_o} \Theta_T = \lim_{T\to\infty} \uparrow \Theta_T$.

Considérons le modèle statistique réduit $(X^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta ; \theta \in \Theta_T))$.

$\forall T \geq T_o$ notons E_T l'adhérence de $\pi(\Theta_T)$ par rapport à d^* . Il s'agit d'un compact de $([0,1]^\infty, d^*)$. Vu que la convergence selon d est celle de la convergence coordonnée à coordonnée, on peut prolonger d à E_T en posant $\forall e^o = (e_1^o, e_2^o ,\ldots)$, $e^1 = (e_1^1, e_2^1 ,\ldots) \in E_T$
$d(e^o,e^1) = \frac{1}{2} \sum_{i\geq 1} |e_i^o - e_i^1|$, et (E_T,d) est complet.

Ceci dit, pour que la distribution stationnaire de X soit p.-s. estimable il suffit, vu la proposition II.1. et le lemme II.3. que $\forall e^o, e^1 \in E_T$, $e^o \neq e^1$ $\exists \alpha > 0$ tel que les hypothèses $(P_\theta ; \theta \in \Theta_T , d(\pi(\theta), e^i) \leq \alpha)$ $(i=0,1)$ se séparent asymptotiquement uniformément, et d'après la proposition V.1. une condition suffisante pour cela est que les familles de matrices de transition associées à ces hypothèses soient décantées.

iii) Soient $\theta_o , \theta_1 \in \Theta_T$, $\theta_o \neq \theta_1$ tels que $d(\pi(\theta_i), e^i) \leq \frac{1}{6T}$ $(i=0,1)$. La définition de Θ_T entraîne $d(\pi(\theta_o), \pi(\theta_1)) \geq \frac{1}{T}$.

Si $M(\theta_o,\theta_1) = M_{(\theta_o,\theta_1)}$, posons $A = A_{(\theta_o,\theta_1)}$, et si $M(\theta_o,\theta_1) = M_{(\theta_1,\theta_o)}$, posons $A = A_{(\theta_1,\theta_o)}$.

Alors card $A = M(\theta_o,\theta_1) \leq N$. Admettons que $A = A_{(\theta_1,\theta_o)}$, l'autre cas pouvant se traiter de façon analogue. On a $\pi_A(\theta_o) + \frac{1}{T} \leq \pi_A(\theta_1)$.

Soit $H_i = (P_\theta\ ;\ \theta \in \Theta_T\ ,\ d(\pi(\theta), e^i) \leq \frac{1}{6T})$ $(i=0,1)$.

Si $P_\theta \in H_o$, alors $d(\pi(\theta), \pi(\theta_o)) \leq \frac{1}{3T}$ et $\pi_A(\theta) \leq \pi_A(\theta_o) + \frac{1}{3T}$.
D'où : $\forall x \in X \quad P^{(t)}_{x,A}(\theta) \leq \pi_A(\theta) + Nf(t) \leq \pi_A(\theta_o) + \frac{1}{3T} + Nf(t)$.

De même, si $P_{\theta'} \in H_1$, alors $\forall x \in X \quad P^{(t)}_{x,A}(\theta') \geq \pi_A(\theta_1) - \frac{1}{3T} - Nf(t)$.

Donc $\forall x \in X \quad P^{(t)}_{x,A}(\theta') - P^{(t)}_{x,A}(\theta) \geq \frac{1}{T} - \frac{2}{3T} - 2N\,f(t)$.

Si t_o vérifie $f(t_o) < \frac{1}{6NT}$, t_o est un temps de décantation pour $(P_{\cdot,\cdot}(\theta)\ ;\ P_\theta \in H_i)$ $(i=0,1)$, et la proposition V.2. permet d'affirmer que H_o et H_1 se séparent asymptotiquement uniformément. ∎

Corollaire 1.- Si $\exists k \in \mathbb{N}^*$ tel que $\Theta = \Theta_k$, la distribution stationnaire de X est uniformément p.-s. estimable.

Proposition V.12.- Soit X une suite d'observations dans les conditions générales de la proposition V.11. et dont le modèle statistique associé $(\mathcal{X}^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta\ ;\ \theta \in \Theta))$ vérifie les conditions a), b), c) et d) de la proposition V.11.

On munit $\pi(\Theta)$ de $d = d_V$.

Alors la distribution stationnaire de X est estimable.

Preuve : i) Soit ℓ^1 l'espace de Banach des suites réelles $y = (y_i\ ;\ i \geq 1)$ absolument sommables muni de la norme $||y|| = \sum_{i\geq 1} |y_i|$. Soit d_o la distance associée à cette norme. Il est évident que $\pi(\Theta)$ est un sous-ensemble de (ℓ^1, d_o). Puisque ℓ^1 est séparable, et que $d = \frac{1}{2} d_o$, il s'ensuit que $(\pi(\Theta), d)$ est séparable. Il est complet dès que fermé. Pourtant, il n'est pas précompact, en général.

ii) D'après la proposition II.5., une condition suffisante pour que la distribution stationnaire de X soit estimable est que

$\forall\theta_o$, $\theta_1 \in \Theta$ $\forall\alpha_o$, $\alpha_1 > 0$ tels que $\beta = d(\pi(\theta_o), \pi(\theta_1)) - \alpha_o - \alpha_1 > 0$ les hypothèses $H_i = (P_\theta \; ; \; \theta \in \Theta \; , \; d(\pi(\theta), \pi(\theta_i)) \leq \alpha_i)$ $(i=0,1)$ se séparent asymptotiquement uniformément.

Comme l'on a vu dans la preuve de la proposition V.9. (partie iv) il existe un événement A qui décante les familles $(\pi(\theta) \; ; \; \theta \in \Theta \; , \; d(\pi(\theta), \pi(\theta_i)) \leq \alpha_i)$ $(i=0,1)$. Par ailleurs, d'après (20) et d), cet événement peut être choisi fini, de cardinal $M(\theta_o,\theta_1)$, comme décrit dans la preuve de la proposition V.11, partie iii). On supposera A déterminé de cette façon.

Supposons $A = A_{(\theta_1,\theta_o)}$, l'autre cas pouvant se traiter de façon semblable.

Si $P_\theta \in H_o$, alors $d(\pi(\theta), \pi(\theta_o)) \leq \alpha_o$ et $\pi_A(\theta) \leq \pi_A(\theta_o) + \alpha_o$. D'où $\forall x \in X$ $P^{(t)}_{x,A}(\theta) \leq \pi_A(\theta) + M(\theta_o,\theta_1)\, f(t) \leq \pi_A(\theta_o) + \alpha_o + M(\theta_o,\theta_1)\, f(t)$.

De même, si $P_{\theta'} \in H_1$, alors $\forall x \in X$ $P^{(t)}_{x,A}(\theta') \geq \pi_A(\theta_1) - \alpha_1 - M(\theta_o,\theta_1)\, f(t)$.

Donc $\forall x \in X$ $P^{(t)}_{x,A}(\theta') - P^{(t)}_{x,A}(\theta) \geq \beta - 2M(\theta_o,\theta_1)\, f(t)$.

Si t_o vérifie $f(t_o) < \dfrac{\beta}{2M(\theta_o,\theta_1)}$, t_o est un temps de décantation pour $(P_{\theta,.}(\theta) \; ; \; P_\theta \in H_i)$ $(i=0,1)$, et la proposition V.2. nous permet d'établir le résultat annoncé. ■

Remarque V.6.- Si l'on admet de plus que X est un processus strictement stationnaire, sa loi initiale coïncide avec la distribution stationnaire. Donc , la loi initiale de X est estimable si la distribution stationnaire de X est estimable.

En nous reportant à la remarque III.9. et au théorème III.7., on voit que l'on peut toujours poser a_1 , p_1 , ϑ_1 tous nuls, $C_1 = \mathcal{X}$. L'exemple où le modèle statistique $(X^{(\infty)}, \mathcal{B}^{(\infty)}, (P_\theta \; ; \theta \in \Theta))$ associée à X vérifie, pour une matrice stochastique P donnée, $\forall\theta \in \Theta \quad P_{.,.}(\theta) = P$ et $\forall\theta, \theta' \in \Theta$, $\theta \neq \theta'$ $P_\theta^{(1)} \neq P_{\theta'}^{(1)}$ permet de vérifier qu'en général les méthodes employées dans ce travail ne peuvent aboutir à des conditions pour que la loi initiale soit estimable.

V.8. - UNE CONDITION SUFFISANTE DE CONVERGENCE DE L'ESTIMATEUR DU MAXIMUM DE VRAISEMBLANCE.-

Les suites d'estimateurs convergents d'un certain paramètre dont l'existence est assurée d'après les méthodes constructives présentées dans le chapitre II ne sont pas utilisables dans la pratique. Des conditions de convergence des estimateurs usuellement employés sont donc utiles. Ceci justifie le contenu de cette section.

Ici, les chaînes de Markov considérées ne sont pas nécessairement homogènes, sauf si cette condition est explicitement mentionnée.

On se base sur des résultats présentés dans ([47]).

Proposition V.13.- Soit X une chaîne de Markov à valeurs dans $\mathcal{X}$ fini ou dénombrable, $(X^{(\infty)}, (\mathcal{P}(\mathcal{X}))^{(\infty)}, (P_\theta \, , \theta \in \Theta))$ un modèle statistique pour X. On note $P_\theta^{(n)}$ la loi de $X^{(n)}$, $P_{n-1,n}(\theta)$ $(n \geq 2)$ la matrice de transition de l'instant n-1 à l'instant n associées à θ.

On suppose que :

a) (Θ, d) est un espace métrique compact ;

b) $\theta_1, \theta_2 \in \Theta$, $\theta_1 \neq \theta_2 \Longrightarrow \{x^{(n)} \in X^n : P_{\theta_1}^{(n)}(x^{(n)}) = P_{\theta_o}^{(n)}(x^{(n)})\} = 0 \ (n \geq 1)$; (21)

c) $\forall \theta_o \in \Theta \ \lim_{\theta \to \theta_o} \sum_{x^{(n)} \in X^n} |P_\theta^{(n)}(x^{(n)}) - P_{\theta_o}^{(n)}(x^{(n)})| = 0 \ (n \geq 1)$. (22)

Avec $\theta \in \Theta$, $\delta > 0$, $n \geq 1$, $x^{(n)} \in X^n$, on note :

i) $K_\theta = \sum_{y \in X} \sup \ (P_{\theta'}^{(1)}(y) \ ; \ \theta' \in B(\theta, \delta))$;

ii) $a_{\theta,\delta}^n = \sup_{x \in X} \sum_{y \in X} \sup_{\theta' \in B(\theta,\delta)} [P_{n-1,n}(\theta')]_{x,y} \quad (n \geq 2)$;

iii) $(\hat{\theta}_n \ ; \ n \geq 1)$ la suite des estimateurs du maximum de vraisemblance, c'est-à-dire

$$\hat{\theta}_n(x^{(n)}) = \theta_o \Longleftrightarrow P_{\theta_o}^{(n)}(x^{(n)}) = \sup(P_\theta^{(n)}(x^{(n)}) \ ; \ \theta \in \Theta) \ . \qquad (23)$$

Considérons l'hypothèse

(H) Soit $(\theta', \theta'') \in \Theta^2$.

$\exists \delta > 0 \qquad \exists \beta \in]0,1[$

$\forall n \geq 2 \qquad \exists a_n \in [0,1]$, la suite $(a_n \ ; \ n \geq 2)$ vérifiant $\sum_{n \geq 2} a_n^2 = \infty$

$\exists p_n : X \longrightarrow [0,1]$ vérifiant $p_n + a_n \leq 1$ et $\forall x \in X$

$\exists \psi_{\theta',\theta'',x}^n : X \longrightarrow [0,1]$ telles que $\forall x \in X$:

I) $d(\theta, \theta') \leq \delta \Longrightarrow \sum_{y \in X} \psi_{\theta',\theta'',x}^n(y) \ [P_{n-1,n}(\theta)]_{x,y} \geq p_n(x) + a_n \quad (n \geq 2)$; (24)

II) $\sum_{y \in X} \psi_{\theta',\theta'',x}^n(y) \sup_{\theta \in B(\theta'',\delta)} [P_{n-1,n}(\theta)]_{x,y} \leq p_n(x) \quad (n \geq 2)$; (25)

III) $K_{\theta''} < \infty$, $a_{\theta'',\delta}^n \leq \beta \ e^{\frac{3}{8} a_n^2} \quad (n \geq 2)$. (26)

Une condition suffisante pour que $(\hat{\theta}_n\ ;\ n\geq 1)$ soit convergente est que $\forall\theta_1\ ,\ \theta_2 \in \Theta\ ,\ \theta_1 \neq \theta_2$ (H) soit vérifiée par (θ_1,θ_2) ou par (θ_2,θ_1).

Remarque V.7.- L'exigence que (21) et (22) soient vérifiées pour $n \geq 1$ n'est introduite que pour simplifier l'écriture. On aurait pu aussi bien considérer qu'elles n'étaient valables que pour $n \geq N \in \mathbb{N}^*$. Pour $n < N$, on ferait $\hat{\theta}_n(x^{(n)}) = \theta_o$ où θ_o est un des points où le supremum de $P_\theta^{(n)}(x^{(n)})$ sur Θ est atteint. Par ailleurs, lorsque l'on se reporte à des ordres $n \geq N_1 \in \mathbb{N}^*$ on devrait les remplacer par $n \geq \max(N,N_1)$.

Remarque V.8.- L'énoncé présenté n'est pas une simple traduction du résultat de W. Pieczynski ([47]) au cadre qui nous intéresse. En fait, sa condition se traduirait par l'exigence que (H) soit vérifiée à la fois par (θ_1,θ_2) et (θ_2,θ_1) pour tous $\theta_1\ ,\ \theta_2 \in \Theta\ ,\ \theta_1 \neq \theta_2$.

On s'apercevra que, pour l'affaiblissement de la condition, le caractère markovien de X ne joue aucun rôle. Ceci pourtant n'est guère le cas de la dénombrabilité de $\mathcal{X}$. C'est-à-dire que la condition que nous énonçons serait aussi valable pour le cadre plus général de ([47]), si la convergence de $P_\theta^{(n)}(x^{(n)})$ vers $P_{\theta_o}^{(n)}(x^{(n)})$ lorsque $\theta \longrightarrow \theta_o$ était uniforme sur $\mathcal{X}^{(n)}$.

Preuve de la proposition V.13.- Soit $\theta_1\ ,\ \theta_2 \in \Theta\ ,\ \theta_1 \neq \theta_2$. On suppose que (H) est satisfaite par (θ_1,θ_2), la preuve pour le cas où (H) est vérifiée par (θ_2,θ_1) étant naturellement tout à fait semblable.

i) Soit $\delta > 0$ fixé sous les conditions de (H). Définissons $P_{\theta_2}^{n,\delta}$ sur $\mathcal{X}^n$ par

$$P_{\theta_2}^{n,\delta}(x^{(n)}) = \sup(P_\theta^{(n)}(x^{(n)})\ ;\ \theta \in B(\theta_2,\delta)) \quad (n\geq 1) \quad , \tag{27}$$

et une famille de fonctions de transition $P^{\delta}_{n-1,n}(\theta_2)$ $(n\geq 2)$ par

$$\forall x,\ y \in X \quad [P^{\delta}_{n-1,n}(\theta_2)]_{x,y} = \frac{\sup([P_{n-1,n}(\theta)]_{x,y}\ ;\ \theta \in B(\theta_2,\delta))}{\sum\limits_{y\in X} \sup([P_{n-1,n}(\theta)]_{x,y}\ ;\ \theta \in B(\theta_2,\delta))}\ ,$$

ce qui est justifié par (26).

Définissons $P^{n,\delta}_{\theta_2,*}$ comme étant la loi markovienne sur $X^{(\infty)}$ de la loi initiale $P^{1,\delta}_{\theta_2,*}(x) = \dfrac{P^{1,\delta}_{\theta_2}(x)}{K_{\theta_2}}$ et de famille de matrices de transition $(P^{\delta}_{n-1,n}(\theta_2)\ ;\ n\geq 2)$.

Vu que $[P^{\delta}_{n-1,n}(\theta_2)]_{x,y} \leq \sup([P_{n-1,n}(\theta)]_{x,y}\ ;\ \theta \in B(\theta_2,\delta))$, d'après (H)I et (H)II on sait que les familles de matrices de transition $(P_{n-1,n}(\theta)\ ;\ n\geq 2\ ,\ \theta \in B(\theta_1,\delta))$ et $(P^{\delta}_{n-1,n}(\theta_2)\ ;\ n \geq 2)$ sont décantées.

Donc $\exists A^{(n)} \subset X^{(n)}$ telle que

$$\inf(P^{(n)}_{\theta}(A^{(n)})\ ;\ \theta \in B(\theta_1,\delta)) \geq 1 - \exp(-\frac{3}{8}\sum_{k=2}^{n} a_k^2)\ , \tag{28}$$

$$P^{n,\delta}_{\theta_2,*}(A^{(n)}) \leq \exp(-\frac{3}{8}\sum_{k=2}^{n} a_k^2)\ . \tag{29}$$

Or $\forall x^{(n)} \in X^n \quad P^{n,\delta}_{\theta_2}(x^{(n)}) \leq$

$$\leq \sup(P^{(1)}_{\theta}(x_1)\ ;\ \theta \in B(\theta_2,\delta))\ .\ \sup(\prod_{j=1}^{n-1} [P_{j,j+1}(\theta)]_{x_j,x_{j+1}}\ ;\ \theta \in B(\theta_2,\delta)) \leq$$

$$\leq K_{\theta_2}\ .\ P^{1,\delta}_{\theta_2,*}(x_1)\ .\ \sup([P_{1,2}(\theta)]_{x_1,x_2}\ ;\ \theta \in B(\theta_2,\delta))\ .$$

$$.\ \sup(\prod_{j=2}^{n-1} [P_{j,j+1}(\theta)]_{x_j,x_{j+1}}\ ;\ \theta \in B(\theta_2,\delta)) =$$

$$= K_{\theta_2}\ .\ P^{1,\delta}_{\theta_2,*}(x_1)\ .\ [P^{\delta}_{1,2}(\theta_2)]_{x_1,x_2}\ .\ (\sum_{x_2\in X}\ \sup_{\theta\in B(\theta_2,\delta)} [P_{1,2}(\theta)]_{x_1,x_2})\ .$$

$$.\ \sup_{\theta\in B(\theta_2,\delta)} (\prod_{j=2}^{n-1} [P_{j,j+1}(\theta)]_{x_j,x_{j+1}}) \leq$$

$$\leq K_{\theta_2} \cdot a^2_{\theta_2,\delta} \cdot \left[P^{2,\delta}_{\theta_2,*}(x_1,x_2)\right] \cdot \sup_{\theta\in B(\theta_2,\delta)} \left(\left[P_{2,3}(\theta)\right]_{x_2,x_3}\right) \cdot$$

$$\cdot \sup_{\theta\in B(\theta_2,\delta)} \left(\prod_{j=3}^{n-1} \left[P_{j,j+1}(\theta)\right]_{x_j,x_{j+1}}\right).$$

De proche en proche, on constate que

$P^{n,\delta}_{\theta_2}(x^{(n)}) \leq K_{\theta_2} \cdot \prod_{i=2}^{n} a^i_{\theta_2,\delta} \cdot P^{n,\delta}_{\theta_2,*}(x^{(n)})$. D'où d'après (26) et (29) ,

$$P^{n,\delta}_{\theta_2}(A^{(n)}) \leq K_{\theta_2} \cdot \beta^{n-2} \cdot e^{\frac{3}{8}\sum_{j=2}^{n} a_j^2} \cdot P^{n,\delta}_{\theta_2,*}(A^{(n)}) \leq$$

$$\leq K_{\theta_2} \cdot \beta^{n-2} \cdot e^{\frac{3}{8}\sum_{j=2}^{n} a_j^2} \cdot e^{-\frac{3}{8}\sum_{j=2}^{n} a_j^2} \xrightarrow[n\to\infty]{} 0 .$$

Soit

$$E^{n,\delta}_{\theta_1,\theta_2} = \{x^{(n)} \in X^n : P^{(n)}_{\theta_1}(x^{(n)}) > P^{n,\delta}_{\theta_2}(x^{(n)})\} \quad . \tag{30}$$

$$P^{(n)}_{\theta_1}(E^{n,\delta}_{\theta_1,\theta_2}) - P^{n,\delta}_{\theta_2}(E^{n,}_{\theta_1,\theta_2}) \geq P^{(n)}_{\theta_1}(A^{(n)}) - P^{n,\delta}_{\theta_2}(A^{(n)}) \xrightarrow[n\to\infty]{} 1 \quad ,$$

d'où :

$$P^{(n)}_{\theta_1}(E^{n,\delta}_{\theta_1,\theta_2}) \xrightarrow[n\to\infty]{} 1 \quad . \tag{31}$$

ii) Vu que (H)I et (H)II sont vérifiées, d'après la proposition II.2., θ est uniformément p.-s. estimable sur Θ et une condition nécessaire pour cela est que $\forall \theta_1 , \theta_2 \in \Theta$, $\theta_1 \neq \theta_2$ $\exists \lambda > 0 : B(\theta_i,\lambda)$ $(i=1,2)$ se séparent asymptotiquement uniformément.
Donc $\exists C_n \subset X^{(n)}$ $(n\geq 1)$ pour lesquels

$$\theta \in B(\theta_1,\lambda) \Longrightarrow P^{(n)}_{\theta}(C_n) \xrightarrow[n\to\infty]{} 1 \quad , \tag{32}$$

$$\theta \in B(\theta_2,\lambda) \Longrightarrow P_\theta^{(n)}(C_n) \xrightarrow[n\to\infty]{} 0 . \tag{33}$$

Si (H) est vérifiée, avec $\delta \in]0,\lambda[$, par (θ_1,θ_2) et si $\theta_o \in B(\theta_1, \delta/2)$, $\theta'' \in B(\theta_2, \delta/2)$, alors $\exists \delta_{\theta''} > 0$ telle que, pour cette valeur de δ , (H) est vérifiée par (θ_o,θ'') . D'où

$$P_{\theta_o}^{(n)}[E_{\theta_o,\theta''}^{n,\delta_{\theta''}}] \xrightarrow[n\to\infty]{} 1 . \tag{34}$$

Posons

$$A_{\theta_1,\theta_2}^{n,\delta} = \{x^{(n)} \in X^n : \sup(P_\theta^{(n)}(x^{(n)}) ; \theta \in B(\theta_1,\delta)) > \sup(P_\theta^{(n)}(x^{(n)}) ; \theta \in B(\theta_2,\delta))\}. \tag{35}$$

$B(\theta_2, \delta/2)$ étant un compact, de $(b(\theta'', \delta_{\theta''}), \theta'' \in B(\theta_2, \delta/2))$ on peut extraire un recouvrement fini $(b(\theta''_i, \delta_{\theta''_i}) ; 1\leqslant i\leqslant N)$ de $B(\theta_2,\delta/2)$.

D'après (34), $P_{\theta_o}^{(n)} [\bigcap_{1\leqslant i\leqslant N} E_{\theta_o,\theta''_i}^{n,\delta_{\theta''_i}}] \xrightarrow[n\to\infty]{} 1$ et $\bigcap_{1\leqslant i\leqslant N} E_{\theta_o,\theta''_i}^{n,\delta_{\theta''_i}} \subset E_{\theta_o,\theta_2}^{n,\delta/2}$,

vu que

$$x^{(n)} \in \bigcap_{1\leqslant i\leqslant N} E_{\theta_o,\theta''_i}^{n,\delta_{\theta''_i}} \Longleftrightarrow \forall i \in \{1,\ldots,N\} \sup(P_\theta^{(n)}(x^{(n)}) ; \theta \in b(\theta''_i, \delta_{\theta''_i})) <$$

$$< P_{\theta_o}^{(n)}(x^{(n)}) \Longrightarrow \sup(P_\theta^{(n)}(x^{(n)}) ; \theta \in \bigcup_{1\leqslant i\leqslant N} b(\theta''_i,\delta_{\theta''_i})) < P_{\theta_o}^{(n)}(x^{(n)}) \Longrightarrow$$

$$\Longrightarrow x^{(n)} \in E_{\theta_o,\theta_2}^{n,\delta/2} .$$

Puisque $E_{\theta_o,\theta_2}^{n,\delta/2} \subset A_{\theta_1,\theta_2}^{n,\delta/2}$,

$$P_{\theta_o}^{(n)}(A_{\theta_1,\theta_2}^{n,\delta/2}) \xrightarrow[n\to\infty]{} 1 . \tag{36}$$

iii) Si $\theta_o \in B(\theta_1,\delta/2)$, $\theta'' \in B(\theta_2,\delta/2)$ et si $A^n_{\theta_o,\theta''} = \{x^{(n)} \in X^{(n)} : P^{(n)}_{\theta_o}(x^{(n)}) > P^{(n)}_{\theta''}(x^{(n)})\}$, d'après (32) et (33) on a, vu que $\delta \in]0,\lambda[$,

$$P^{(n)}_{\theta_o}(A^n_{\theta_o,\theta''}) - P^{(n)}_{\theta''}(A^n_{\theta_o,\theta''}) \geq P^{(n)}_{\theta_o}(C_n) - P^{(n)}_{\theta''}(C_n) \xrightarrow[n\to\infty]{} 1 \quad , \quad \text{d'où}$$

$$P^{(n)}_{\theta''}(A^n_{\theta_o,\theta''}) \xrightarrow[n\to\infty]{} 0 \ . \tag{37}$$

Cela est vérifié pour (θ_o,θ'') arbitraire dans $B(\theta_1,\delta/2) \times B(\theta_2,\delta/2)$.

Compte tenu de (22), $\forall\theta_o \in B(\theta_1, \delta/2)\ \exists\delta_{\theta_o} > 0$: $x^{(n)} \in A^n_{\theta_o,\theta''} \Longrightarrow \forall\theta \in b(\theta_o,\delta_{\theta_o})\ \ x^{(n)} \in A^n_{\theta,\theta''}$. La convergence de $P^{(n)}_\theta$ vers $P^{(n)}_{\theta_o}$ lorsque $\theta \to \theta_o$ étant uniforme sur X^n , δ_{θ_o} peut être choisi indépendamment de l'élément $x^{(n)}$ considéré.

De $(b(\theta_o, \delta_{\theta_o})\ ;\ \theta_o \in B(\theta_1,\delta/2))$ extrayons un recouvrement fini $(b(\theta_j , \delta_{\theta_j})\ ;\ 3\leq j\leq k)$ de $B(\theta_1,\delta/2)$. On a $P^{(n)}_{\theta''}(\bigcup\limits_{i=3}^{k} A^n_{\theta_i,\theta''}) \xrightarrow[n\to\infty]{} 0$, d'où

$$P^{(n)}_{\theta''}\ \{x^{(n)} \in X^n : \sup(P^{(n)}_\theta(x^{(n)})\ ;\ \theta \in B(\theta_1,\delta/2)) > P^{(n)}_{\theta''}(x^{(n)}))\ \xrightarrow[n\to\infty]{} 0 \ . \tag{38}$$

En effet, si $x^{(n)} \notin \bigcup\limits_{i=3}^{k} A^n_{\theta_i,\theta''}$, $\forall\theta \in B(\theta_1,\delta/2)\ \exists i \in \{3,\ldots,k\}$: $\theta \in b(\theta_i,\delta_{\theta_i})$ et $P^{(n)}_\theta(x^{(n)}) < P^{(n)}_{\theta''}(x^{(n)})$. Le supremum de $P^{(n)}_\theta(x^{(n)})$ sur $B(\theta_1,\delta/2)$ étant atteint, (38) est vérifiée.

Or, d'après (35), $A^{n,\delta/2}_{\theta_1,\theta_2} \subset \{\sup(P^{(n)}_\theta\ ;\ \theta \in B(\theta_1,\delta/2)) > P^{(n)}_{\theta''}\}$.

Il s'ensuit :

$$P^{(n)}_{\theta''}(A^{n,\delta/2}_{\theta_1,\theta_2}) \xrightarrow[n\to\infty]{} 0 \ . \tag{39}$$

iv) La conclusion est une conséquence de (36) et (39).

$\forall \varepsilon > 0 \quad \forall \theta_o \in \Theta$ soit

$$C_\varepsilon^{\theta_o} = \{\theta \in \Theta : d(\theta,\theta_o) \geq \varepsilon\} \quad . \tag{40}$$

$\forall \theta \in C_\varepsilon^{\theta_o}$ notons $\delta(\theta_o,\theta)$ un élément de $]0,\frac{\varepsilon}{2}[$ permettant de vérifier (H) par (θ,θ_o) ou (θ_o,θ). La famille $(b(\theta,\delta(\theta_o,\theta))\ ;\ \theta \in C_\varepsilon^{\theta_o})$ constitue un recouvrement du compact $C_\varepsilon^{\theta_o}$. On peut en extraire un recouvrement fini $(b(\theta_j^o\ ,\ \delta(\theta_o,\theta_j^o))\ ;\ 1 \leq j \leq N(\theta_o))$. Soit

$$\delta(\theta_o) = \inf \{\frac{1}{2}\ \delta(\theta_o,\theta_j^o)\ ;\ 1 \leq j \leq N(\theta_o)\} \quad . \tag{41}$$

$(b(\theta_o,\delta(\theta_o))\ ;\ \theta_o \in \Theta)$ est un recouvrement de Θ, duquel on peut extraire un recouvrement fini $(b(\theta_i,\ \delta(\theta_i))\ ;\ 1 \leq i \leq M)$. Soient

$$\delta_i = \delta(\theta_i) \quad , \quad \delta_j^i = \frac{1}{2}\ \delta(\theta_i,\ \theta_j^i) \ . \tag{42}$$

A chaque $i \in \{1,\ldots,M\}$ correspond un recouvrement fini $(b(\theta_j^i\ ,\ 2\ \delta_j^i)\ ;\ 1 \leq j \leq N(\theta_i))$ de $C_\varepsilon^{\theta_i}$. On a

$$\forall j \in \{1,\ldots,N(\theta_i)\} \qquad \delta_i \leq \delta_j^i \ . \tag{43}$$

$\forall i \in \{1,\ldots,M\}$ posons

$$A_{\theta_i}^{n,\delta_i} = \bigcap_{1 \leq j \leq N(\theta_i)} A_{\theta_i,\theta_j^i}^{n,\delta_j^i} \quad . \tag{44}$$

Remarquons que $A_{\theta,\theta'}^{n,\delta} = X^{(n)} - A_{\theta',\theta}^{n,\delta}$ $(\theta,\ \theta' \in \Theta\ ,\ \theta \neq \theta')$.

$\forall i \in \{1,\ldots,M\} \quad d(\theta_i,\theta) \leq \delta_j^i \Longrightarrow P_\theta^{(n)}(A_{\theta_i,\theta_j^i}^{n,\delta_j^i}) \xrightarrow[n\to\infty]{} 1$, d'où

$$d(\theta,\theta_i) \leq \delta_i \Longrightarrow P_\theta^{(n)}(A_{\theta_i}^{n,\delta_i}) \xrightarrow[n\to\infty]{} 1 , \tag{45}$$

puisque $d(\theta,\theta_i) \leq \delta_i \Longrightarrow d(\theta,\theta_i) \leq \delta_j^i$ $(1\leq j\leq N(\theta_i))$, d'après (43).

Or on peut montrer que

$$\forall\theta \in \Theta \quad \{d(\hat{\theta}_n,\theta) < 2\varepsilon\} \supset A_{\theta_i}^{n,2\delta_i} , \quad \text{si} \quad \theta \in b(\theta_i,2\delta_i) , \tag{46}$$

ce qui nous permettra de conclure.

En effet, $x^{(n)} \in A_{\theta_i}^{n,2\delta_i} \Longrightarrow \forall j \in \{1,\ldots,N(\theta_i)\}$

$$\sup(P_\theta^{(n)}(x^{(n)}) \ ; \ \theta \in B(\theta_i, 2\ \delta_j^i) > \sup(P_\theta^{(n)}(x^{(n)}) \ ; \ \theta \in B(\theta_j^i , 2\ \delta_j^i)) , \tag{47}$$

d'après (44). Cela implique que $\hat{\theta}_n(x^{(n)})$ ne peut pas appartenir à aucune boule $b(\theta_j^i , 2\ \delta_j^i)$ $(j \in \{1,\ldots,N(\theta_i)\})$, et l'ensemble de ces boules est un recouvrement de $C_\varepsilon^{\theta_i}$. Alors ,

$$\hat{\theta}_n(x^{(n)}) \notin C_\varepsilon^{\theta_i} , \quad d(\hat{\theta}_n(x^{(n)}) , \theta_i) < \varepsilon . \tag{48}$$

Or θ appartient à $b(\theta_i,2\delta_i)$ par hypothèse (voir (46)) et $\delta_i \leq \frac{\varepsilon}{2}$ d'après (41) et (42). Donc, $d(\hat{\theta}_n(x^{(n)}) , \theta) < 2\varepsilon$, et (46) est vérifiée. ■

Corollaire 1.- Si X est homogène, $a_{\theta,\delta} = \sup_{x\in X} \sum_{y\in X} \sup_{\theta'\in B(\theta,\delta)} P_{x,y}(\theta')$ $(\theta \in \Theta , \delta > 0)$ et si les conditions générales de la proposition V.8. sont satisfaites, une condition suffisante pour que la suite $(\hat{\theta}_n \ ; \ n \geq 1)$ des estimateurs du maximum de vraisemblance soit convergente est que $\forall\theta_1 , \theta_2 \in \Theta , \quad \theta_1 \neq \theta_2$ la condition (H') ci-dessous soit vérifiée par (θ_1,θ_2) ou par (θ_2,θ_1) :

(H'). Soit $(\theta',\theta'') \in \Theta^2$.

$$\exists a > 0 \ \exists \delta > 0 \ \exists \beta \in]0,1[\ \exists p : X \longrightarrow [0,1-a]$$

et $\forall x \in X \ \exists \psi_{\theta',\theta'',x} : X \longrightarrow [0,1]$ pour lesquels $\forall x \in X$:

i) $\forall \theta \in B(\theta',\delta) \quad \sum_{y \in X} \psi_{\theta',\theta'',x}(y)\, P_{x,y}(\theta) \geq p(x) + a$;

ii) $\sum_{y \in X} \psi_{\theta',\theta'',x}(y) \quad \sup(P_{x,y}(\theta) \ ; \ \theta \in B(\theta'',\delta)) \leq p(x)$;

iii) $K_{\theta''} < \infty$, $a_{\theta'',\delta} \leq \beta \ e^{3/8 \ a^2}$.

REFERENCES BIBLIOGRAPHIQUES

[1] ANDERSON, T. et GOODMAN, L. : "Statistical inference about Markov chains", Ann. Math. Stat. 28 (1957), p. 89-109.

[2] BAD, M. : "The asymptotic behaviour of the maximum - likelihood estimate for dependent observations and of the Bayes generalized estimate", Rev. Roum. Math. Pures et Appl., 24 (1979), n° 6, p. 847-862.

[3] BARTLETT, M. : "The frequency goodness of fit test for probability chains", Proc. Cambridge Phyl. Soc., 47 (1951), p. 86-95.

[4] BAR-SHALOM, Y. : "On the asymptotic properties of the maximum - likelihood estimate obtained from dependent observations", J. Roy. Stat. Soc., Ser. B, 33 (1971), p. 72-77.

[5] BATH, B. : "On the method of maximum - likelihood for dependent observations", J. Roy. Stat. Soc., Ser. B, 36 (1974), p. 48-53.

[6] BILLINGSLEY, P. : "Asymptotic distributions of two goodness of fit criteria", Ann. Math. Stat., 27 (1956), p. 1123-1129.

[7] BILLINGSLEY, P. : "Statistical methods in Markov chains", Ann. Math. Stat., 32 (1961), p. 12-40.

[8] BILLINGSLEY, P. : "Statistical inference for Markov processes", University of Chicago Press, Chicago (1961).

[9] BILLINGSLEY, P. : "The Lindberg-Lévy theorem for martingales", Proc. Amer. Math. Soc., 12 (1961), p. 788-792.

[10] BILLINGSLEY, P. : "Probability and measure", John Wiley & Sons, Inc., New York (1979).

[11] BROWN, B. : "Martingale central limit theorems", Ann. Math. Stat., 42 (1971), p. 59-66.

[12] CHUNG, K. : "Markov chains with stationary transition probabilities", Springer-Verlag, Berlin (1960).

[13] CRAMÉR, H. : "Mathematical methods of statistics", Princeton University Press (1974), 13e édition.

[14] CROWDER, M. : "Maximum likelihood estimation for dependent observations", J. Roy. Stat. Soc., Ser. B, 38 (1976), p. 45-53.

[15] de ALMEIDA, R. et MOCHÉ R. : "Sur la décantation et l'othogonalité de deux hypothèses", Actas do VII Congresso do Grupo de Matemáticos de Expressão Latina (1985), Coimbra (à paraître).

[16] DERMAN, C. : "Some asymptotic distribution theory for Markov chains with a denumerable number of states", Biometrika, 43 (1956), p. 285-294.

[17] DOBRUSHIN, R. : "Prescribing a system of random variables by conditional distributions", Theory Prob. Applications, 15 (1970), p. 458-486.

[18] DOOB, J. : "Stochastic processes", John Wiley & Sons, Inc. New York (1953).

[19] DUDLEY, R. : "Distances of probability measures and random variables", Ann. Math. Stat., 39 (1968), p. 1563-1572.

[20] FORTET, R. et MOURIER, E. : "Convergence de la répartition empirique vers la répartition théorique", Ann. Sc. Eco. Norm. Sup., 70 (1953), p. 266-285.

[21] GEFFROY, J., et MOCHÉ, R. : "Sur la séparation asymptotique uniforme des produits de lois de probabilité", C.R.A.S., Paris, Sér. A, t. 278 (1974), p. 969-971.

[22] GEFFROY, J., et MOCHÉ, R. : "Constructions d'estimateurs uniformément convergents à partir d'observations non nécessairement équidistribuées", C.R.A.S., Paris, Sér. A, t. 280 (1975), p. 133-135.

[23] GEFFROY, J. : "Inégalités pour le niveau de signification et la puissance de certains tests reposant sur des données quelconques", C.R.A.S., Paris, Sér. A, t. 282 (1976), p. 1299-1301.

[24] GEFFROY, J. : "Test et estimation dans les chaînes de Markov homogènes par la méthode de la décantation", Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie, Paris.

[25] GNEDENKO, B. : "The theory of probability", Mir Publishers, Moscou (1976), 6e édition.

[26] GOODMAN, L. : "Simplified run tests and likelihood ratio tests for Markov chains", Biometrika, 45 (1958), p. 181-197.

[27] GOODMAN, L. : "Exact probabilities and asymptotic relationships for some statistics from m-th order Markov chains", Ann. Math. Stat., 29 (1958), p. 476-490

[28] GOODMAN, L. : "Asymptotic distributions of the "psi-squared" goodness of fit criteria for m-th order Markov chains", Ann. Math. Stat., 29 (1958), p. 1123-1133

[29] GOURIÉROUX, C., MONFORT, A., et TROGNON, A. : "Pseudo maximum likelihood methods : theory", Econometrica, 52 (1984), p. 681-700.

[30] HOEFFDING, W. : "Probability inequalities for sums of bounded random variables", Amer. Stat. Ass. J., p. 13-30 (1963).

[31] HOEL, P. : "A test for Markoff chains", Biometrika, 41 (1954), p. 430-433.

[32] JIŘINA, M. : "Conditional probabilities on σ-algebras with countable basis", Selected Transl. Math. Stat. Prob., 2 (1962), p. 125-142.

[33] KABANOV, J., LIPCER, R., et ŠIRJAEV, A. : "Absolute continuity and singularity of locally absolutely continuous probability distributions", (I) Math. U.S.S.R. Sbornik, 35 (1979), n° 5, p. 631-680.
(II) Math. U.S.S.R. Sbornik, 36 (1980), n° 1, p. 31-58.

[34] KAKUTANI, S. : "On equivalence of infinite product measures", Ann. Mathematics, 49 (1948), p. 214-224.

[35] KRAFT, C. : "Some conditions for consistency and uniform consistency of statistical procedures", Univ. California Publ. Statistics, Univ. California Press, 2 (1955), p. 125-142.

[36] LE CAM, L. : "On the assumptions used to prove asymptotic normality of maximum likelihood estimates", Ann. Math. Stat., 41 (1970), p. 802-828.

[37] LOÈVE, M. : "Probability theory", Van Nostrand, Princeton (1963), 3e édition.

[38] MEMIN, J. : Communication à R. Moché (1983).

[39] MOCHÉ, R. : "Décantation et séparation asymptotiques uniformes ; tests et estimateurs convergents, dans le cas d'observations indépendantes, équidistribuées ou non", Thèse de Doctorat d'Etat, 1ère partie, Université des Sciences et Techniques de Lille, Lille (1977).

[40] MOCHÉ, R. : Communications personnelles.

[41] MOCHÉ, R. : "Sur la statistique asymptotique des processus gaussiens", Actas do VII Congresso do Grupo de Matemáticos de Expressão Latina (1985), Coimbra (à paraître).

[42] MOCHÉ, R. : "Construction d'estimateurs", cours de D.E.A., Université des Sciences et Techniques de Lille Flandres Artois, Lille (1986).

[43] MOCHÉ, R. : "La décantation en statistique asymptotique", Publ. I.R.M.A. Lille, Vol. 4 (1986), n° 1, p. 1-38 (prépublication), Lille.

[44] NEVEU, J. : "Cours de Probabilités", Université Pierre et Marie Curie, Paris (1975).

[45] NEVEU, J. : "Martingales à temps discret", Masson (1972).

[46] PARZEN, E. : "Modern probability theory and its applications", John Wiley & Sons, Inc., New York (1960).

[47] PIECZYNSKI, W. : "Sur diverses applications de la décantation des lois de probabilité dans la théorie générale de l'estimation statistique", Thèse de Doctorat d'Etat, Université Pierre et Marie Curie, Paris (1986).

[48] SCHMERKOTTE, H. : "Interdependences between consistent tests and estimates", Z. Wahrscheinlichkeitstheorie verw. Geb., Bd. 16 (1970), p. 293-306.

[49] SCOTT, D. : "Central limit theorems for martingales and processes with stationary increments using a Skorohod representation approach", Adv. Appl. Prob., 5 (1973), p. 119-137.

[50] SILVEY, S. : "A note on maximum - likelihood in the case of dependent random variables", J. Roy. Stat. Soc., Ser. B, 23 (1961), p. 444-452.

[51] VALLENDER, S. : "Calculation of the Wasserstein distance between probability distributions on the line", Theory Prob. Applications, 18 (1973), p. 784-786.

[52] WALD, A. : "Asymptotic properties of the maximum likelihood estimate of un unknown parameter of a discrete stochastic process", Ann. Math. Stat., 19 (1948), p. 40-46.

[53] WALD, A. : "Note on the consistency of the maximum likelihood estimate", Ann. Math. Stat., 20 (1949), p. 595-601.

[54] WHITE, H. : "Maximum likelihood estimation of misspecified models", Econometrica, 50 (1982), p. 1-25.

[55] WHITTLE, P. : "Some distributions and moment formulae for the Markov chain", J. Roy. Stat. Soc., Ser. B, 17 (1955), p. 235-242.

[56] WOLFOWITZ, J. : "On Wald's proof of the consistency of the maximum likelihood estimate", Ann. Math. Stat., 20 (1949), p. 601-602.

[57] WOLFOWITZ, J. : "Asymptotic efficiency of the maximum likelihood estimator", Theory Prob. Applications, 10 (1965), n° 2, p. 247-260.

RESUME

$X = (X_n)$ étant une suite d'observations à valeurs respectivement dans (X_1,B_1) ,... de loi inconnue, H désignant l'hypothèse générale (sur l'espace produit infini correspondant), on désire estimer un paramètre donné f, application de H dans (E,d), espace métrique séparable. Dans ce travail, on se restreint au cas où pour toute loi possible, X est une chaîne de Markov homogène : dans le chapitre I, les résultats obtenus sont annoncés, accompagnés d'une revue bibliographique.

Au chapitre II, on rappelle des constructions d'estimateurs de H. Schmerkotte, J. Geffroy, R. Moché, W. Pieczynski qui, en toute généralité, ramènent le problème de l'estimation asymptotique de f à la séparation asymptotique uniforme de certains couples d'hypothèses multiples, (E,d) devant de plus avoir certaines propriétés de compacité.

Au chapitre III, on traite donc d'abord de l'orthogonalité de deux hypothèses simples (avec un bref historique : théorèmes de S. Kakutani, puis de Ju. Kabanov, R. Lipcer, A. Sirjaev) puis on passe au théorème fondamental de la décantation asymptotique (J. Geffroy, R. Moché) qui s'applique à des hypothèses multiples, alors que les théorèmes d'orthogonalité d'hypothèses simples sont, en général, insuffisants en statistique asymptotique.

Les chapitres IV et V sont essentiellement des chapitres de résultats nouveaux : au chapitre IV, on étudie un nouvel aspect des propriétés de décantation de certaines distances entre lois de probabilité ; enfin, au chapitre V, les résultats obtenus sont appliqués au cas particulier où X est une chaîne de Markov homogène à espace des états dénombrable. Des conditions suffisantes pour que la matrice de transition ou la distribution stationnaire, lorsqu'elle existe, soient estimables sont données. Une condition suffisante de convergence de la suite des estimateurs du maximum de vraisemblance, qui repose essentiellement sur des propriétés de décantation, est aussi établie.

MOTS CLES

- CHAINES DE MARKOV A ETATS DENOMBRABLES
- SEPARATION ASYMPTOTIQUE UNIFORME
- DISTANCES SUR DES ENSEMBLES DE LOIS
- DECANTATION