N° d'ordre : 123

THÈSE

présentée à la

FACULTÉ DES SCIENCES DE L'UNIVERSITÉ DE LILLE

pour obtenir le

Titre de Docteur de Spécialité

(Mathématiques Appliquées)

par

Claude LANGRAND

*

SUR LA MESURABILITÉ ET L'ESTIMATION

Thèse soutenue le 13 Décembre 1968, devant la Commission d'Examen

Monsieur M. PARREAU, Président
Mademoiselle S. MARQUET } Examinateurs
Monsieur P. POUZET } Examinateurs
Monsieur BUI-TRONG-LIEU, Rapporteur

FACULTE DES SCIENCES

UNIVERSITE DE LILLE

LISTE DES PROFESSEURS

-oOo-

DOYENS HONORAIRES

MM. H. LEFEBVRE, M. PARREAU.

PROFESSEURS HONORAIRES

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

PROFESSEURS

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

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

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

MAITRES DE CONFERENCES

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

INTRODUCTION

Les problèmes d'estimation par la méthode du maximum de vraisemblance sont l'objet d'un grand nombre de travaux.

La plupart d'entre eux sont basés sur l'étude de l'existence et de la convergence de solutions convenablement choisies des équations de vraisemblance successives. Or, lorsque celles-ci ont plusieurs solutions, le choix que l'on opère ne correspond pas nécessairement à des maximums pour les fonctions de vraisemblance. Cette manière d'aborder le problème s'est donc avérée inadéquate [7] .

Adoptant la définition selon laquelle un estimateur est le terme général d'une certaine suite d'applications mesurables -ce qui nous permet, par exemple, de parler de son biais ou de sa variance- nous avons cherché sous quelles hypothèses l'existence d'un estimateur du quasi-maximum de vraisemblance était assurée.

Nous avons ensuite étudié la convergence presque sûre d'un estimateur du quasi-maximum de vraisemblance dans le cas de certains processus de Markov. Comme celle de ROUSSAS [13] , la démonstration est directe et ne suppose pas les hypothèses habituelles de dérivabilité des densités de probabilité par rapport au paramètre à estimer.

Après un chapitre 0, dans lequel nous avons réuni les materiaux nécessaires à une définition précise des notions générales utilisées dans ce travail, cette étude comporte trois chapitres .

- le premier est consacré à un exposé sur un théorème de projection de tribu produit, vu comme conséquence de résultats sur les ensembles analytiques et sur les capacités.

- dans le second, le problème de l'existence d'un estimateur du quasi-maximum de vraisemblance nous a conduit à chercher une application mesurable dont le graphe est inclus dans un ensemble borélien.

- dans le chapitre III, nous traitons de la convergence presque sûre d'un estimateur du quasi-maximum de vraisemblance dans le cas de processus de Markov discrets homogènes a espace d'états $(X, \mathcal{B})$, lorsque l'espace des paramètres est localement compact a base denombrable.

Une application dans le cas d'une chaîne de Markov à un nombre fini d'etats montre, pour terminer, un des aspects positifs de ce travail.

Je souhaite exprimer ma profonde gratitude à Monsieur le Professeur PARREAU, Chef du Département de Mathématiques Pures, qui m'a fait l'honneur d'accepter la présidence du Jury.

Je tiens à remercier Mademoiselle le Professeur MARQUET et Monsieur le Professeur POUZET qui ont bien voulu faire partie de la commission d'examen.

Pour m'aider à mener ce travail à bon terme, Monsieur le Professeur BUI TRONG LIEU n'a jamais hésité à me consacrer généreusement son temps ; ses remarques et ses suggestions m'ont conduit à de nombreuses améliorations, importantes tant sur le fond que sur la présentation. Qu'il veuille bien trouver dans ces lignes, l'expression de ma profonde reconnaissance.

Il me reste à remercier Mademoiselle Delencuville qui a rendu possible la réalisation matérielle de ce travail avec rapidité et gentillesse.

CHAPITRE 0

RAPPELS DE NOTIONS GENERALES

Nous nous proposons dans ce chapitre de rappeler quelques definitions dans le but de preciser les notations et la terminologie.

* Soit E un ensemble. Nous appelons *pavage* sur E, tout ensemble $\mathcal{E}$ de parties de E contenant la partie vide.

Nous noterons $\mathcal{E}_s$ (resp. $\mathcal{E}_\sigma$, $\mathcal{E}_d$, $\mathcal{E}_\delta$) l'ensemble des reunions finies (resp. des reunions denombrables, des intersections finies, des intersections dénombrables) d'éléments de $\mathcal{E}$.

Nous adopterons le nom de *tribu* (σ-algèbre) pour un pavage sur E stable pour les operations σ et complementaire.

Nous designerons par ensemble pave (resp. espace probabilisable ou mesurable) le couple $(E,\mathcal{E})$ constitue par un ensemble E et un pavage (resp. une tribu) $\mathcal{E}$ sur E.

Soit $(E_i,\mathcal{E}_i)_{i \in I}$ une famille d'ensembles paves. Nous appelerons pavage produit des $\mathcal{E}_i$ et nous noterons $\prod_{i \in I} \mathcal{E}_i$, le pavage sur $E = \prod_{i \in I} E_i$ formé par les parties $\prod_{i \in I} A_i$ de E ou $A_i \subset E$ n'est different de E_i que pour un nombre fini d'indices pour lesquels on a $A_i \in \mathcal{E}_i$.

Dans le cas ou les $\mathcal{E}_i$ sont des tribus sur E_i, la *tribu produit* des $\mathcal{E}_i$ que nous noterons suivant [12] $\bigotimes_{i \in I} \mathcal{E}_i$ est engendrée par $\prod_{i \in I} \mathcal{E}_i$.

* Soit (Ω, A, Pr) un espace de probabilité, où A désigne une tribu de parties de Ω et Pr une mesure de probabilité. Soit $(X, \mathcal{B})$ un espace mesurable.

Une application $X : (\Omega, A) \to (X, \mathcal{B})$ est dite A-$\mathcal{B}$-mesurable si $X^{-1}(\mathcal{B}) \subset A$. Nous dirons dans ce cas que X est une variable aléatoire sur (Ω, A) et à valeurs dans $(X, \mathcal{B})$.

Soit une famille de variables aléatoires X_t indicées par $t \in \zeta$ où ζ est un ensemble totalement ordonné, définies sur (Ω, A, Pr) et à valeurs dans des espaces mesurables $((X_t, \mathcal{B}_t))_{t \in \zeta}$; nous dirons que cette famille est une *fonction aléatoire*.

$(X_t, \mathcal{B}_t)$ est appelé *espace des états* à l'instant t.

Lorsque $\zeta \subset Z$ nous dirons que la fonction aléatoire $\{X_t\}_{t \in \zeta}$ est discrète.

Pour $\omega \in \Omega$, l'application

$$t \longrightarrow X_t(\omega)$$

s'appelle la *trajectoire* correspondant à ω.

* Soient 2 espaces mesurables $(X_1, \mathcal{B}_1)$ et $(X_2, \mathcal{B}_2)$.

On appelle *probabilité de transition* (ou de passage) de $(X_1, \mathcal{B}_1)$ à $(X_2, \mathcal{B}_2)$ toute application $P_{1,2}$ définie sur $X_1 \times \mathcal{B}_2$ à valeurs dans $[0,1]$ qui vérifie :

(1) $\forall\, x \in \mathcal{X}_1$ $P_{1,2}(x,.)$ est une probabilité sur $\mathcal{B}_2$.

(2) $\forall\, A \in \mathcal{B}_2$ $P_{1,2}(.,A)$ est une variable aléatoire réelle sur $(\mathcal{X}_1,\mathcal{B}_1)$.

Si $((\mathcal{X}_t,\mathcal{B}_t))_{t \in \mathbb{N}^*}$ (1) est une suite d'espaces probabilisables, l'application $P_{s,t}$ définie sur $\mathcal{X}_s \times \mathcal{B}_t$ à valeurs dans $[0,1]$ telle que

$$P_{s,t}(x,A) = \int_{\mathcal{X}_{s+1}} P_{s,s+1}(x,dx_{s+1}) \int_{\mathcal{X}_{s+2}} P_{s+1,s+2}(x_{s+1},dx_{s+2}) \dots P_{t-1,t}(x_{t-1},A)$$

$$s < t \qquad A \in \mathcal{B}_t \quad , \quad x \in \mathcal{X}_s$$

est une probabilité de transition de $(\mathcal{X}_s,\mathcal{B}_s)$ à $(\mathcal{X}_t,\mathcal{B}_t)$.

De plus on a l'équation de Chapman-Kolmogorov :

$$\forall\, s,t,u \in \mathbb{N}^* \quad , \quad s < t < u$$

$$P_{s,u}(x,A) = \int_{\mathcal{X}_t} P_{s,t}(x,dy)\, P_{t,u}(y,A) \qquad \forall\, x \in \mathcal{X}_s$$

$$\forall\, A \in \mathcal{B}_u \ .$$

Remarquons dans l'équation précédente en ce qui concerne les mesures, que nous avons noté $P_{s,t}(x,dy)$ de préférence à $d\,P_{s,t}(x,y)$ afin d'éviter des confusions entre x et y .

* Soit $((\mathcal{X}_t,\mathcal{B}_t))_{t \in \mathbb{N}^*}$ une suite d'espaces probabilisables. Supposons que $\forall\, n \in \mathbb{N}^*$, il existe une probabilité de transition $P_{t,t+1}$ de $(\mathcal{X}_t,\mathcal{B}_t)$ à $(\mathcal{X}_{t+1},\mathcal{B}_{t+1})$. Nous dirons alors que la donnée de la suite des espaces $((\mathcal{X}_t,\mathcal{B}_t))_{t \in \mathbb{N}^*}$ et de celle des probabilités de passage $(P_{t,t+1})_{t \in \mathbb{N}^*}$ définit un *processus de Markov* que nous noterons

$$((\mathcal{X}_t,\mathcal{B}_t) \ , \ P_{t,t+1})_{t \in \mathbb{N}^*} \ .$$

(1) $\mathbb{N}^*$ est l'ensemble des entiers > 0 .

Un processus de Markov sera dit *homogène* si

1) $\forall\ t \in \mathbb{N}^*$ $(X_t, \mathcal{B}_t) = (X, \mathcal{B})$

2) $\forall\ t \in \mathbb{N}^*$ $P_{t,t+1} = P$.

Nous avons dans ce cas

$$P_{n,n+t} = P_{n',n'+t} \qquad \forall\ t, n', n \in \mathbb{N}^*$$

nous poserons alors

$$P_{n,n+t} = P^{(t)}$$

et noterons le processus de Markov homogène

$$((X, \mathcal{B})\ ,\ P^{(t)})_{t \in \mathbb{N}^*} \ .$$

* Soit $(\Omega, \mathcal{A}, Pr)$ un espace probabilisé et soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. Soit X une variable aléatoire réelle Pr intégrable, définie sur $(\Omega, \mathcal{A})$. Notons $Pr_{\mathcal{B}}$ la restriction de Pr à $\mathcal{B}$.

On appelle espérance mathématique conditionnelle de X par rapport à $\mathcal{B}$ relativement à Pr et on note $E^{\mathcal{B}}(X)$ comme dans [12] la variable aléatoire réelle définie sur $(\Omega, \mathcal{B})$, $Pr_{\mathcal{B}}$ presque partout par

$$\int_B X(\omega)\ d\ Pr(\omega) = \int_B E^{\mathcal{B}}(X)\ d\ Pr_{\mathcal{B}}(\omega) \qquad , \qquad \forall\ B \in \mathcal{B} \ .$$

On appelle probabilité conditionnelle par rapport à $\mathcal{B}$, relativement à Pr et on note $P(.|\mathcal{B})$ la variable aléatoire réelle définie par :

$$P(A \mid \mathcal{B}) = E^{\mathcal{B}}[1_A] \quad , \quad \forall A \in \mathcal{B} .$$

Soit (Ω, A, Pr) un espace probabilisé et soit $(A_t)_{t \in \mathbb{N}^*}$ une suite de sous-tribus de A. Notons $A_{1,2,\ldots,n}$ la tribu engendrée par les éléments de $A_1,\ldots,A_n$.

On dit que la *suite des sous-tribus* $(A_t)_{t \in \mathbb{N}^*}$ est *markovienne* (simple) si $\forall m$ et $m' \in \mathbb{N}^*$, $\forall s \in \mathbb{N}^*$, $\forall j_1,\ldots,j_m$; $k_1,\ldots,k_{m'} \in \mathbb{N}^*$ tels que

$$j_1 < j_2 < \ldots < j_m < s < k_1 < \ldots < k_{m'}$$

$$Pr[A \mid A_{s,j_1,\ldots,j_m}] = Pr[A \mid A_s] \qquad \forall A \in A_{k_1,\ldots,k_{m'}} .$$

Une *fonction aléatoire* $(X_t)_{t \in \mathbb{N}^*}$ est dite *markovienne* (simple) si la suite de sous tribus $X_t^{-1}(\mathcal{B}_t)$ de A est markovienne simple.

On dit que la fonction aléatoire markovienne $(X_t)_{t \in \mathbb{N}^*}$ est *attachée* au processus du Markov $((X_t,\mathcal{B}_t), P_{t,t+1})_{t \in \mathbb{N}^*}$ s'il existe une version de la probabilité conditionnelle $Pr[X_{t+1} \in A \mid X_t]$ telle que

$$\forall t \in \mathbb{N}^* \quad , \quad \forall A \in \mathcal{B}_{t+1} \quad , \quad \forall x \in X_t$$

$$Pr[X_{t+1} \in A \mid X_t = x] = P_{t,t+1}(x,A) .$$

Si on pose $Q_t(B) = Pr[X_t \in B]$, $\forall B \in \mathcal{B}_t$, on a

$$\forall s, t \in \mathbb{N}^* \; s < t \quad , \quad Q_t(B) = \int_{X_s} d\,Q_s(x)\, P_{s,t}(x,B) .$$

La suite $(Q_t)_{t \in \mathbb{N}^*}$ est appelée suite des probabilités absolues de la fonction aleatoire markovienne $(X_t)_{t \in \mathbb{N}^*}$.

Une fonction aleatoire markovienne $(X_t)_{t \in \mathbb{N}^*}$ attachée à un processus de Markov $((X_t, \mathcal{B}_t), P_{t,t+1})_{t \in \mathbb{N}^*}$ est dite *stationnaire* si

1) le processus est homogène ;
2) les variables aleatoires X_t sont toutes de même loi, c'est-à-dire si $\forall t \in \mathbb{N}^*$, $Q_t = Q$.

Nous terminerons ce chapitre par quelques rappels de Statistique.

* Soit $(X_t)_{t \in \mathbb{N}^*}$ une fonction aleatoire définie sur un espace probabilisable (Ω, A), à valeurs dans $(X_t, \mathcal{B}_t)_{t \in \mathbb{N}^*}$. Soit $Pr(.;\theta)$ une famille de probabilites, θ étant un paramètre $\in \Xi$.

Le problème de l'estimation ponctuelle auquel nous nous interesserons ici consistera à determiner a l'aide d'un morceau de trajectoire correspondant à $\omega \in \Omega$: $\{X_1(\omega), \ldots, X_n(\omega)\}$ la valeur vraie, inconnue θ_o du paramètre $\theta \in \Xi$.

On appelera *estimateur* de θ_o, *le terme général d'une suite de variables aléatoires* $(\hat{\theta}_t)_{t \in \mathbb{N}^*}$ definies sur (Ω, A), à valeurs dans (Ξ, Λ), ou Λ est une tribu de parties de Ξ, de la manière suivante :

1) il existe une suite de fonctions mesurables $(\varepsilon_t)_{t \in \mathbb{N}^*}$ définies sur $(\prod_{t \in \mathbb{N}^*} X_t, \bigotimes_{t \in \mathbb{N}^*} \mathcal{B}_t)$ à valeurs dans (Ξ, Λ) telles que, $\forall t \in \mathbb{N}^*$, ε_t n'est effectivement fonction que des t premières coordonnées.

2) $\forall\, t \in \mathbb{N}^*$, $\hat{\theta}_t = g_t\,[X_1, X_2, \ldots, X_t]$

3) pour le "morceau" de trajectoire donné $\{X_1(\omega), \ldots, X_n(\omega)\}$, la valeur $\hat{\theta}_n(\omega) = g_n\,[X_1(\omega), \ldots, X_n(\omega)]$ sert de valeur approchée de θ_o .

Soit $(\hat{\theta}_n)_{n \in \mathbb{N}^*}$ un estimateur de θ_o , défini sur (Ω, A, Pr) à valeurs dans (Ξ, Λ) . On dira que l'estimateur *converge Pr-presque sûrement* vers θ_o, s'il existe $\Omega_1 \in A$ tel que $P(\Omega_1) = 1$ et que

$$\forall\, \omega \in \Omega_1 \qquad \hat{\theta}_n(\omega) \to \theta_o \qquad \text{quand } n \to +\infty \ .$$

Un certain nombre de méthodes permettent d'obtenir des estimateurs. Nous étudierons celle du *maximum de vraisemblance.*

Rappelons en brièvement la technique :

Soit une fonction aléatoire $(X_t)_{t \in \mathbb{N}^*}$ définie sur (Ω, A) . Soit $Pr(.;\theta)$, $\theta \in \Xi$; supposons que, $\forall\, \theta \in \Xi$, $Pr(.;\theta)$ est dominée par une mesure μ σ-finie ; $\omega \in \Omega$ étant donné, soit $\{X_t\}_{t \in [1,n] \cap \mathbb{N}}$ le "morceau" de trajectoire de taille n , correspondant à ω .

On appelera *fonction de vraisemblance* (au point ω) , l'application de Ξ dans $\mathbb{R}$,

$$L_n(\omega,.) \ : \ \theta \to L_n(\omega;\theta)$$

où $L_n(\omega;\theta)$ est la densité de la variable aléatoire $(X_1, X_2, \ldots, X_n)$ évaluée au point ω .

La *méthode du maximum de vraisemblance*, comme celle du *quasi-maximum de vraisemblance*, consiste, pour estimer θ_o , à trouver d'abord une suite d'applications $(\theta_n^*)_{n \in \mathbb{N}^*}$ de Ω dans Ξ , définies par

$$L_n(\omega;\theta_n^*(\omega)) \geq c\, L_n(\omega;\theta) \quad , \quad \forall\, \theta \in \Xi \quad , \quad c \in \,]0,1] \quad ,$$

$c = 1$ correspondant à la methode du maximum de vraisemblance.

Au chapitre II, nous montrerons sous quelles conditions $(\theta_n^*)_{n \in \mathbb{N}^*}$ est une suite d'applications mesurables et permet de définir un estimateur.

CHAPITRE I

SUR UN THEOREME DE PROJECTION DE TRIBU PRODUIT

Rappelons tout d'abord, qu'étant donnés deux ensembles non vides X et Ξ et un sous-ensemble A du produit cartésien $X \times \Xi$, la *section* de A en $x \in X$ est la partie A_x de Ξ telle que

$$A_x = \{\theta \in \Xi : (x,\theta) \in A\} \quad .$$

Nous qualifierons de *projection* sur X, l'application de l'ensemble $\mathcal{P}(X \times \Xi)$ des parties de $X \times \Xi$ dans l'ensemble $\mathcal{P}(X)$ des parties de X, définie par :

$$\forall A \in \mathcal{P}(X \times \Xi) \quad , \quad \mathrm{proj}_X A = \{x \in X : A_x \neq \emptyset\} \quad .$$

Si $\mathcal{C}$ est un ensemble de parties de $X \times \Xi$, nous écrirons $\mathrm{proj}_X \mathcal{C}$ pour $\{\mathrm{proj}_X C : C \in \mathcal{C}\}$.

I. 1. POSITION DU PROBLEME.

Soient $(X,\mathcal{B}')$ et $(\Xi,\mathcal{A}')$ 2 ensembles pavés ; $(X,\mathcal{B})$ et $(\Xi,\mathcal{A})$ 2 espaces mesurables.

Dans [8] E. MARCZEWSKI et C. RYLL-NARDZEWSKI firent remarquer que pour les opérations $\rho = s$, d, σ, δ, $sd = ds$, δs et $d\sigma$ on avait

$$\mathrm{proj}_X \left[(\mathcal{B}' \times \mathcal{A}')_\rho\right] \subset \mathcal{B}'_\rho \quad .$$

Ils montrèrent qu'il fallait faire des hypothèses sur $\mathcal{A}'$ lorsque $\rho = s\delta$, $\sigma\delta$ où ρ représente l'opération A de Souslin pour que le pavage $\text{proj}_X [\mathcal{B}' \times \mathcal{A}']$ puisse s'exprimer à l'aide du pavage $\mathcal{B}'$ seul.

L'exemple suivant a pour but de montrer que si les tribus $\mathcal{B}$ et $\mathcal{A}$ ne sont pas choisies de manière convenable, il peut ne pas y avoir identité entre $\text{proj}_X (\mathcal{B} \otimes \mathcal{A})$ et $\mathcal{B}$.

Exemple. -

Remarquons d'abord qu'il est évident que $\mathcal{B} \subset \text{proj}_X (\mathcal{B} \otimes \mathcal{A})$ puisque $\forall B \in \mathcal{B}$, le pavé $B \times \Xi = \{(x,\Theta) \in X \times \Xi : x \in B , \Theta \in \Xi\}$ a B pour projection sur X .

Considérons alors comme espace mesurable $(X,\mathcal{B})$ le segment $[0,1]$ de la droite réelle, muni de sa tribu borelienne et pour $(\Xi,\mathcal{A})$ l'espace mesurable $([0,1] , \mathcal{P}([0,1]))$.

Soient $I^n_m = \left[\frac{m-1}{n} , \frac{m}{n}\right]$ $m = 1,2,\dots,n$ $n \in \mathbb{N}^*$

et $C_n = [I^n_1 \times (A \cap I^n_1)] \cup [I^n_2 \times (A \cap I^n_2)] \cup \dots \cup [I^n_n \times (A \cap I^n_n)]$

où $A \neq \emptyset$, $A \subset [0,1]$.

Posons enfin $C = \bigcap_{n \in \mathbb{N}^*} C_n$.

Comme $I^n_m \in \mathcal{B}$ et $A \cap I^n_m \in \mathcal{P}([0,1])$ on a

$$I^n_m \times (A \cap I^n_m) \in \mathcal{B} \otimes \mathcal{P}([0,1]) .$$

On en deduit que C_n et par suite C, appartient à $\mathcal{B} \otimes \mathcal{P}([0,1])$.

Soit $D = \{(x,x) \in A \times A\}$.

$\forall$ n fixé, $n \in \mathbb{N}^*$; $\forall x \in A$, $\exists m \in \{1,2,\dots,n\}$ tel que

$$x \in I_m^n \quad \text{et donc} \quad (x,x) \in I_m^n \times (A \cap I_m^n) .$$

Il en résulte que

$$\forall n \in \mathbb{N}^* , \quad (x,x) \in C_n \quad \text{d'ou} \quad D \subset C .$$

Supposons maintenant que $(x,\theta) \in C$. Comme alors $(x,\theta) \in C_n$ pour tout n, on a $\theta \in A$.

Montrons que l'hypothèse $x \neq \theta$ est absurde.

En effet, elle entrainerait que

$$\forall n \in \mathbb{N}^* , \quad \exists m \in \{1,2,\dots,n\} \quad \text{tel que}$$

$$(x,\theta) \in I_m^n \times (A \cap I_m^n)$$

d'où $x \in I_m^n$ et $\theta \in I_m^n$. Or il est possible de choisir n assez grand pour que x et θ appartiennent à l'interieur de 2 intervalles différents du partage.

On en deduit que $C \subset D$ et de l'identité des ensembles C et D, il resulte que

$$\text{proj}_X C = \text{proj}_X D = A .$$

Il suffit alors de choisir l'ensemble A de telle manière qu'il ne soit pas $\mathcal{B}$-mesurable pour démontrer que la projection sur X de l'ensemble C appartenant à $\mathcal{B} \otimes \mathcal{A}$, n'appartient pas à $\mathcal{B}$.

Dans ce chapitre on va donc chercher des conditions (suffisantes) à imposer aux espaces mesurables $(X,\mathcal{B})$ et $(E,\mathcal{A})$ pour qu'il y ait identité entre $\mathrm{proj}_X(\mathcal{B} \otimes \mathcal{A})$ et $\mathcal{B}$.

Dans [12] on semble chercher ce résultat par l'intermédiaire de l'opération de Souslin et d'un théorème de capacitabilité des ensembles sousliniens démontré par CHOQUET.

Dans le présent exposé nous avons préféré parvenir aux résultats relatifs à la projection de tribu produit en adaptant à ce problème certains des résultats de [9].

Nous rappelerons dans le paragraphe 2, certaines définitions données dans [9] et nous grouperons en un seul énoncé les principaux résultats dont nous nous servirons.

I. 2. ENSEMBLES $\mathcal{E}$-ANALYTIQUES - $\mathcal{E}$-CAPACITES, PROPRIETES.

I. 2.1. *Définitions.-*

(a) Soit $(E,\mathcal{E})$ un ensemble pavé. On dit qu'une famille $(K_i)_{i\in I}$ d'éléments de $\mathcal{E}$ possède la propriété de l'intersection finie si l'on a

$$\bigcap_{i \in I_0} K_i \neq \emptyset \quad \text{pour toute partie finie } I_0 \subset I .$$

(b) On dit que le pavage $\mathcal{E}$ est *semi-compact*, si toute famille dénombrable d'éléments de $\mathcal{E}$, qui possède la propriété de l'intersection finie, a une intersection non vide.

(c) Soit $(E,\mathcal{E})$ un ensemble pavé. On dit qu'une partie A de E est $\mathcal{E}$-*analytique*, s'il existe un ensemble auxiliaire F muni d'un pavage semi-compact $\mathcal{F}$ et une partie $B \subset E \times F$, appartenant à $(\mathcal{E} \times \mathcal{F})_{\sigma\delta}$ telle que A soit la projection de B sur E.

On notera $\mathcal{A}(\mathcal{E})$ le pavage sur E constitué par les ensembles $\mathcal{E}$-analytiques.

I. 2.2. Théorème.-

Soit $(E,\mathcal{E})$ un ensemble pavé. On a

1 - $\mathcal{E} \subset \mathcal{A}(\mathcal{E})$

2 - $[\mathcal{A}(\mathcal{E})]_{\delta} = \mathcal{A}(\mathcal{E}) = [\mathcal{A}(\mathcal{E})]_{\sigma}$

3 - si $(F,\mathcal{F})$ est un ensemble pavé on a

$$\mathcal{A}(\mathcal{E}) \times \mathcal{A}(\mathcal{F}) \subset \mathcal{A}(\mathcal{E} \times \mathcal{F})$$

4 - si le pavage $\mathcal{F}$ est semi-compact la projection A de $A' \in \mathcal{A}(\mathcal{E} \times \mathcal{F})$ sur E est $\mathcal{E}$-analytique.

5 - $\mathcal{A}(\mathcal{A}(\mathcal{E})) = \mathcal{A}(\mathcal{E})$

6 - la pavage $\mathcal{A}(\mathcal{F})$ contient la tribu engendrée par $\mathcal{F}$ si et seulement si le complémentaire de tout élément de $\mathcal{F}$ est $\mathcal{F}$-analytique.

I. 2.3. $\mathcal{E}$-capacité et théorème de Choquet.-

1.- Définitions :

- Soit E un ensemble muni d'un pavage $\mathcal{E}$ stable pour les opérations de réunion finie et d'intersection finie.

On appelle *$\mathcal{E}$-capacité* sur E une application I définie sur $\mathcal{P}(E)$ à valeurs dans $\overline{R}$ vérifiant les propriétés suivantes :

a) I est croissante

b) pour toute suite croissante $(A_n)_{n \in N^*}$ de parties de E on a

$$I\left(\bigcup_{n \in N^*} A_n \right) = \sup_{n \in N^*} I(A_n)$$

c) pour toute suite décroissante $(A'_n)_{n \in N^*}$ d'éléments de $\mathcal{E}$ on a

$$I\left(\bigcap_{n \in N^*} A'_n \right) = \inf_{n \in N^*} I(A'_n)$$

- Une partie A de E est dite *$\mathcal{E}$-capacitable* si on a l'égalité

$$I(A) = \sup_{\substack{B \in \mathcal{E}_\delta \\ B \subset A}} I(B)$$

2.- Théorème de Choquet :

Soit I une $\mathcal{E}$-capacité sur E . Tout ensemble $\mathcal{E}$-analytique est capacitable pour I .

I. 3. THEOREME DE PROJECTION DE TRIBU PRODUIT.

Tout d'abord un lemme.

I. 3.1. *Lemme.-*

Si $(X,\mathcal{B},P)$ *est un espace de probabilité complet, alors* $\mathcal{A}(\mathcal{B}) = \mathcal{B}$.

Rappelons qu'une partie N d'un espace probabilisé $(X,\mathcal{B},P)$ est P-négligeable s'il existe une partie $B \in \mathcal{B}$ telle que $N \subset B$, $P(B) = 0$ et que l'espace de probabilite $(X,\mathcal{B},P)$ est dit complet si $\mathcal{B}$ contient toute partie P-négligeable de X .

D'après I. 2.2.1. $\mathcal{B} \subset \mathcal{A}(\mathcal{B})$.

Montrons que $\mathcal{B} \supset \mathcal{A}(\mathcal{B})$. Nous ferons la démonstration en deux etapes.

a) On remarquera tout d'abord qu'en notant P^* la probabilite exterieure définie par

$$\forall A \in \mathcal{P}(X) \quad , \quad P^*(A) = \inf_{\substack{B \in \mathcal{B} \\ A \subset B}} P(B) \quad ,$$

P^* est une $\mathcal{B}$-capacite sur X .

En effet a) et c) de la definition I. 2.3.1. sont évidents. Quant à b) on peut le demontrer comme suit :
$(A_n)_{n \in N^*}$ étant une suite croissante d'élements de $\mathcal{P}(X)$, on peut trouver $\forall \epsilon > 0$, une suite croissante $(B_n)_{n \in N^*}$ d'élements de $\mathcal{B}$ tels que

$$B_n \supset A_n \quad \text{et} \quad P^*(B_n) = P(B_n) \leq P^*(A_n) + \epsilon$$

d'après la definition de la probabilite exterieure.

Il en résulte, puisque $\bigcup_{n \in \mathbb{N}^*} B_n \supset \bigcup_{n \in \mathbb{N}^*} A_n$ et $\bigcup_{n \in \mathbb{N}^*} B_n \in \mathcal{B}$

que

$$P^*(\bigcup_{n \in \mathbb{N}^*} A_n) \leqslant P(\bigcup_{n \in \mathbb{N}^*} B_n) = \sup_{n \in \mathbb{N}^*} P(B_n) \leqslant \sup_{n \in \mathbb{N}^*} P^*(A_n) + \varepsilon \quad .$$

Comme $(A_n)_{n \in \mathbb{N}^*}$ est une suite croissante, on a aussi

$$\sup_{n \in \mathbb{N}^*} P^*(A_n) \leqslant P^*(\bigcup_{n \in \mathbb{N}^*} A_n)$$

d'ou $\forall \varepsilon > 0 \quad \sup_{n \in \mathbb{N}^*} P^*(A_n) \leqslant P^*(\bigcup_{n \in \mathbb{N}^*} A_n) \leqslant \sup_{n \in \mathbb{N}^*} P^*(A_n) + \varepsilon$

b) Supposons maintenant que A soit un ensemble $\mathcal{B}$-analytique.

D'après I. 2.3.2. on a

$$P^*(A) = \sup_{\substack{B \in \mathcal{B} \\ B \subset A}} P(B)$$

d'ou

$$\sup_{\substack{B \in \mathcal{B} \\ B \subset A}} P(B) = \inf_{\substack{C \in \mathcal{B} \\ C \supset A}} P(C)$$

Il existe donc B' et C' elements de $\mathcal{B}$ tels que

$$B' \subset A \subset C' \quad \text{et} \quad P(B') = P(C') \quad .$$

On en conclut que $A \in \mathcal{B}$, puisqu'en notant B'^c le complementaire de B' dans X,

$$C' \setminus B' = C' \cap B'^c \quad \text{et} \quad P(C' \setminus B') = 0 \ .$$

I. 3.2. *Hypothèse* (H).-

Nous designerons par (H) l'hypothèse suivante :

Un espace mesurable $(E, \mathcal{A})$ *est dit vérifier l'hypothèse* (H) , *si la tribu* $\mathcal{A}$ *est engendrée par un pavage semi-compact* $\mathcal{C}$ *sur* E *et si de plus, pour tout ensemble* C *de* $\mathcal{C}$, *il existe une suite* $(C_n)_{n \in \mathbb{N}^*}$ *d'éléments de* $\mathcal{C}$ *telle que* $C^c = \bigcup_{k \in \mathbb{N}^*} C_n$ *(ou bien* $C^c = \bigcap_{k \in \mathbb{N}^*} C_n$ *)* .

I. 3.3. *Théorème.-*

Soit $(X, \mathcal{B}, P)$ *un espace de probabilité complet.*

Soit $(E, \mathcal{A})$ *un espace probabilisable vérifiant l'hypothèse* (H) .

Alors $\mathrm{proj}_X (\mathcal{B} \otimes \mathcal{A}) \subset \mathcal{B}$.

Soit $B \times L \in \mathcal{B} \times \mathcal{A}$.

$$(B \times L)^c = (B^c \times E) \cup (X \times L^c)$$

donc $(B \times L)^c \in (\mathcal{B} \times \mathcal{A})_s$ d'ou $(B \times L)^c \in \mathscr{A}(\mathcal{B} \times \mathcal{A})$ et d'après I. 2.2.6.

(3.3.1) $$\mathscr{A}(\mathcal{B} \times \mathcal{A}) \supset \mathcal{B} \otimes \mathcal{A} .$$

Or puisque $(E, \mathcal{A})$ verifie l'hypothèse (H) , l'application de I. 2.2.6. et I. 2.2.2. montre que la tribu $\mathcal{A}$ est incluse dans le pavage $\mathscr{A}(\mathcal{C})$ ou $\mathcal{C}$ est par hypothèse un pavage semi-compact sur E .

D'ou en appliquant I. 2.2. on obtient successivement

$$\mathcal{B} \times \mathcal{A} \subset \mathcal{B} \times \mathscr{A}(\mathcal{C}) \subset \mathscr{A}(\mathcal{B}) \times \mathscr{A}(\mathcal{C})$$

$$\mathscr{A}(\mathcal{B}) \times \mathscr{A}(\mathcal{C}) \subset \mathscr{A}(\mathcal{B} \times \mathcal{C})$$

d'où

$$\mathcal{A}(\mathcal{B} \times \Lambda) \subset \mathcal{A}(\mathcal{B} \times \mathcal{C}) \tag{3.3.2}$$

et à l'aide de (3.3.1) et (3.3.2)

$$\mathcal{B} \otimes \Lambda \subset \mathcal{A}(\mathcal{B} \times \mathcal{C})$$

Le théorème I. 2.2.4. permet alors d'affirmer que

$$\mathrm{proj}_X (\mathcal{B} \otimes \Lambda) \subset \mathcal{A}(\mathcal{B})$$

D'après le lemme I. 3.1. on sait que $\mathcal{B} = \mathcal{A}(\mathcal{B})$ d'où la conclusion.

I. 3.4. *Proposition.-*

Si Ξ est un espace compact métrisable et si Λ est sa tribu borélienne, l'espace mesurable (Ξ,Λ) vérifie l'hypothèse (H) .

Il suffit en effet de prendre pour pavage $\mathcal{C}$ les fermés (donc compacts) de Ξ . Le pavage $\mathcal{C}$ est semi-compact et puisque Ξ est un espace métrisable, tout ouvert de Ξ est la réunion d'une famille dénombrable de fermés.

I. 3.5. *Proposition.-*

La conclusion du théorème I. 3.3. subsiste si (Ξ,Λ) est un espace mesurable formé d'un espace polonais Ξ muni de sa tribu borélienne Λ .

Nous nous bornerons à donner ici les grandes lignes d'une démonstration de [6] . Notons $I^{\aleph_0}$ le cube de Hilbert, c'est-à-dire l'ensemble de toutes les suites infinies $z = (z^1,z^2,\dots,z^n,\dots)$ où z^n est, pour tout $n \in \mathbb{N}^*$, un nombre réel appartenant à $[0,1]$. Soit $\mathcal{C}$ sa tribu borélienne.

Rappelons d'abord les 3 lemmes suivants [3] :

1) I^{X_0} est compact,

2) tout espace metrisable separable est homeomorphe à un sous-ensemble de I^{X_0},

3) si l'espace est polonais, ce sous-ensemble de I^{X_0} est une intersection denombrable d'ouverts. (un G_δ)

Revenons alors à la démonstration de la proposition.

D'après le lemme 2, E est homeomorphe à un sous-espace H de I^{X_0}.

Soit $\mathcal{H}$ la tribu borelienne de H.

D'après le lemme 3, H est un G_δ de I^{X_0}, donc $H \in \mathcal{C}$.

Il en resulte que

$$\mathcal{H} \subset \mathcal{C}$$

et par consequent

$$A \in \mathcal{B} \otimes \mathcal{H} \Rightarrow A \in \mathcal{B} \otimes \mathcal{C}.$$

I^{X_0} etant compact d'après le lemme 1, $A \in \mathcal{B} \otimes \mathcal{A}$ implique d'après la proposition I. 3.4. et le theorème I. 3.3. que

$$\mathrm{proj}_X A \in \mathcal{B}.$$

I. 3.6. Remarque.-

La proposition precedente s'applique en particulier pour un espace E localement compact à base denombrable puisqu'un tel espace est polonais [3].

CHAPITRE II

MESURABILITE ET FONCTION DE VRAISEMBLANCE

Dans ce chapitre nous nous proposons d'étudier la mesurabilite d'applications definies comme solutions de problèmes d'estimation de paramètres. Il est important de demontrer l'existence de telles applications mesurables puisque les grandeurs statistiques, telles le biais ou la variance, n'ont de sens que pour elles.

BUI TRONG LIEU et DANIEL CARTON [5] ont examiné cette question dans le cas de processus de Markov, en se reportant à l'etude de la mesurabilité des racines d'equations ou figurent des variables aleatoires. Ils furent ainsi amenés a enoncer le theorème suivant :

"Soit $(X,\mathcal{B},P)$ un espace probabilisé complet et soit (Ξ,Λ) un espace mesurable verifiant l'hypothèse (H) (notation du I. 3.2.).

Soit enfin $g : X \times \Xi \to \mathbb{R}^n$ une application $\mathcal{B} \otimes \Lambda$ mesurable. Supposons qu'il existe un ensemble $L \in \Lambda$ et un ensemble $X_1 \in \mathcal{B}$ avec $P(X_1) = 1$ tels que, $\forall x \in X_1$, l'equation $g(x,\theta) = 0$ ait et n'ait qu'une racine $\theta_x \in L$.

Alors on peut trouver des variables aleatoires θ^* definies sur $(X,\mathcal{B})$ et a valeurs dans (Ξ,Λ) telles que leur restriction à X_1 soit l'application $x \to \theta_x$."

Ils appliquèrent ensuite ce théorème au problème de l'estimation de paramètre, dans le cas notamment de la méthode du maximum de vraisemblance.

En se limitant à Ξ , pavé ouvert de $\mathbb{R}^m$, ils se ramenèrent alors à l'étude de l'équation obtenue en dérivant la fonction vraisemblance par rapport au paramètre à estimer.

Nous avons voulu ici nous débarrasser de l'hypothèse de dérivabilité par rapport à Θ dont ils avaient besoin pour leur étude.

Nous avons donc essayé d'examiner la mesurabilité des applications définies par la méthode du maximum de vraisemblance, directement, c'est-à dire sans avoir recours à l'équation de vraisemblance.

Nous avons cherché de plus à éliminer l'hypothèse d'unicité faite dans l'énoncé du théorème rappelé plus haut, celle-ci pouvant être pratiquement difficile à réaliser.

II. 1. THEOREME.

Soit $(X,\mathcal{C},P)$ *un espace probabilisé complet,* Ξ *un espace compact métrisable,* Λ *sa tribu borélienne.*

Soit $L : X \times \Xi \to \mathbb{R}$ *une application* $\mathcal{C} \otimes \Lambda$ *mesurable, telle que* $L(x,.)$ *soit semi-continue supérieurement sur* Ξ *, quel que soit* x *élément de* X *.*

Soient enfin g *une application définie sur* X *, à valeurs dans* $\mathbb{R}$

$$g : x \to g(x) = \sup_{\Theta \in \Xi} L(x,\Theta)$$

et ψ *la correspondance*(1) *de* $\mathcal{X}$ *dans* Ξ *définie par*

$$\psi : x \to \psi(x) = \{\theta \in \Xi : L(x,\theta) \geq cg(x)\} \qquad c \in \left]0,1\right]$$

Alors g *est une application* $\mathcal{B}$*-mesurable et le graphe* $G(\psi)$ *de la correspondance* ψ *appartient à* $\mathcal{B} \otimes \mathcal{A}$.

Notons tout de suite que $\psi(x) \neq \emptyset$, puisque l'application $L(x,.)$ de Ξ dans $\mathbb{R}$ est semi-continue supérieurement sur le compact Ξ , y atteint donc sa borne superieure et que de ce fait il existe pour tout $x \in \mathcal{X}$ au moins un $\theta_x \in \Xi$ pour lequel $L(x,\theta_x) = g(x)$.

Soit $a \in \mathbb{R}$. On a

$$\{x \in \mathcal{X} : g(x) \geq a\} = \text{proj}_{\mathcal{X}} \{(x,\theta) \in \mathcal{X} \times \Xi : L(x,\theta) \geq a\}$$
$$= \text{proj}_{\mathcal{X}} A$$

avec $A = \{(x,\theta) \in \mathcal{X} \times \Xi : L(x,\theta) \geq a\} \in \mathcal{B} \otimes \mathcal{A}$, l'application $L(.,.)$ étant par hypothèse $\mathcal{B} \otimes \mathcal{A}$ mesurable.

D'après le theorème de projection de tribu produit (I. 3.3.) on en deduit que

$$\{x \in \mathcal{X} : g(x) \geq a\} \in \mathcal{B}$$

d'où la $\mathcal{B}$-mesurabilité de l'application g .

---

(1) D'après la terminologie de Bourbaki, théorie des ensembles Livre I, étant donnés 2 ensembles E et F , une correspondance ψ de E dans F associe à chaque élément x de E un sous-ensemble non vide $\psi(x)$ de F . Son graphe est $G(\psi) = \{(x,y) \in E \times F : y \in \psi(x)\}$.

L'application g_1, définie sur $X \times \Xi$, à valeurs dans $\mathbb{R}$

$$(x,\theta) \to g_1(x,\theta) = g(x)$$

est $\mathcal{B} \otimes \Lambda$ mesurable. Il en est donc de même de l'application qui a $(x,\theta) \in X \times \Xi$ fait correspondre $L(x,\theta) - cg(x) \in \mathbb{R}$.

Il en résulte que $G(\psi) \in \mathcal{B} \otimes \Lambda$.

II. 2. COROLLAIRE.

Si ψ est une application de X dans Ξ, alors les hypothèses du théorème II. 1. entrainent que ψ est mesurable.

Ceci résulte immédiatement du fait que le graphe de l'application ψ appartient à $\mathcal{B} \otimes \Lambda$ d'après le theorème précédent.

Dans la recherche d'estimateurs par la méthode du quasi maximum de vraisemblance, il nous faudra démontrer l'existence d'une application mesurable θ^* de X dans Ξ telle que

$$\forall\, x \in X \qquad L(x,\theta^*(x)) \geq cg(x) \qquad c \in \left]0,1\right]$$

A l'aide du corollaire précédent nous pouvons déjà répondre à cette question dans un grand nombre de cas. Nous nous sommes alors interrogés sur l'existence d'une telle application mesurable dans le cas où ψ fait correspondre à $x \in X$ un ensemble de valeurs de θ non nécessairement réduit à un point.

Etudions alors l'exemple suivant.

II. 3. EXEMPLE.

Prenons pour espace probabilisé complet, $\mathbb{R}^n$ et la mesure de Lebesgue.

Soit $\Xi = [a,b] \subset \mathbb{R}$, Λ la tribu de ses parties boréliennes.

Définissons $L(\vec{x},\Theta)$ par

$$L(\vec{x},\Theta) = 1 \quad \text{si} \quad \vec{x} = (x_1,x_2,\dots,x_n) \in \left[\Theta - \frac{1}{2} , \Theta + \frac{1}{2}\right]^n$$
$$= 0 \quad \text{sinon}$$

Pour chaque $\vec{x} \in \mathbb{R}^n$ fixé, $L(\vec{x},\Theta)$ sera donc maximum si on a simultanément vérifiées

$$\Theta - \frac{1}{2} \leqslant \min_{i = 1,\dots,n} x_i$$

et

$$\Theta + \frac{1}{2} \geqslant \max_{i = 1,\dots,n} x_i \quad .$$

Toute application $\overset{*}{\Theta}$ définie sur $\mathbb{R}^n$, à valeurs dans $\mathbb{R}$ pour laquelle

$$\max_{i = 1,\dots,n} x_i - \frac{1}{2} \leqslant \overset{*}{\Theta}(\vec{x}) \leqslant \min_{i = 1,\dots,n} x_i + \frac{1}{2}$$

sera telle que

(1) $$L(\vec{x},\overset{*}{\Theta}(\vec{x})) = \sup_{\Theta \in \Xi} L(\vec{x},\Theta)$$

Or si on étudie la mesurabilité de telles applications il est évident que $\forall \lambda \in [0,1]$

$$\overset{*}{\Theta}_\lambda(\vec{x}) = \overset{*}{\Theta}_1(\vec{x}) + \lambda[\overset{*}{\Theta}_2(\vec{x}) - \overset{*}{\Theta}_1(\vec{x})]$$

où

$$\overset{*}{\theta}_1(\vec{x}) = \max_{i = 1,\dots,n} x_i - \frac{1}{2}$$

$$\overset{*}{\theta}_2(\vec{x}) = \min_{i = 1,\dots,n} x_i + \frac{1}{2}$$

est une application mesurable vérifiant (1).

Il y a d'autre part des applications $\overset{*}{\theta}$ qui vérifient (1) et ne sont pas mesurables.

En effet si $A \subset \mathbb{R}^n$ n'est pas mesurable pour la mesure de Lebesgue dans $\mathbb{R}^n$, définissons $\overset{*}{\theta}(\vec{x})$ par

$$\forall \vec{x} \in A \qquad \overset{*}{\theta}(\vec{x}) = \overset{*}{\theta}_2(\vec{x})$$

$$\forall \vec{x} \in A^c \qquad \overset{*}{\theta}(\vec{x}) = \overset{*}{\theta}_1(\vec{x})$$

$\overset{*}{\theta}(\vec{x})$ vérifie (1).

D'autre part, puisque

$$\overset{*}{\theta}(\vec{x}) = [\overset{*}{\theta}_2(\vec{x}) - \overset{*}{\theta}_1(\vec{x})] \cdot 1_A(\vec{x}) + \overset{*}{\theta}_1(\vec{x}) \quad ,$$

$\theta^*(\vec{x})$ n'est pas mesurable.

A l'aide de cet exemple on voit que

- les applications $\overset{*}{\theta}(\vec{x})$ vérifiant (1) ne sont pas toutes mesurables
- il en existe qui le sont.

On peut d'ailleurs énoncer le résultat suivant :

II. 4. PROPOSITION.

Soit $(X, \mathcal{B}, P)$ *un espace probabilisé complet.*

Soit Ξ *un compact dans* $\mathbb{R}$, $\mathcal{A}$ *la tribu de ses parties boréliennes.*

Soit enfin A *un sous-ensemble* $\mathcal{B} \otimes \mathcal{A}$ *mesurable de* $X \times \Xi$ *.*

L'application définie sur X, *à valeurs dans* Ξ *par*

$$x \to \inf \{\Theta \in \Xi : (x, \Theta) \in A\}$$

est $\mathcal{B}$*-mesurable.*

En effet $\forall\, t \in \mathbb{R}$,

$$\{x \in X : \inf \{\Theta \in \Xi : (x, \Theta) \in A\} < t\} = \mathrm{proj}_X \{A \cap (X \times]-\infty, t[)\}$$

et puisque $A \cap (X \times]-\infty, t[) \in \mathcal{B} \otimes \mathcal{A}$, l'application du théorème I. 3.3. nous donne le résultat cherché.

Remarque. -

Le succés de la démonstration est lié à la structure d'ordre total de $\mathbb{R}$ et au fait que les ensembles $]-\infty, t[$ engendrent la tribu borélienne.

Si l'on veut étudier le problème dans un cadre plus général, il peut s'énoncer de la manière suivante :

"*Soit* ψ *une correspondance de* X *dans* Ξ, *dont le graphe* $G(\psi)$ *est* $\mathcal{B} \otimes \mathcal{A}$ *mesurable.*

Existe-t-il une application mesurable $\hat{\Theta}$ *de* X *dans* Ξ *dont le graphe est inclus dans* $G(\psi)$ *?*"

On ne peut pas toujours répondre affirmativement à cette question. Blackwell donne dans [1] un exemple d'ensembles boréliens C, D, B tels que

a) $B \subset C \times D$

b) la projection de B sur C est C

c) pour aucune application mesurable d de C dans D, le graphe de d n'est un sous-ensemble de B.

C'est pourquoi nous avons cherché des conditions (suffisantes) à imposer aux espaces X et Ξ qui assureraient l'existence de $\hat{\Theta}$ en tant qu'application mesurable de X dans Ξ.

On a alors le théorème suivant

II. 5. THEOREME.

Soient X et Ξ 2 espaces polonais, $\mathcal{B}$ et $\mathcal{A}$ leurs tribus boréliennes respectives. On supposera de plus que $(X, \mathcal{B}, P)$ est un espace probabilisé complet.

Soit ψ une correspondance de X dans Ξ dont le graphe $G(\psi)$ appartient la tribu produit $\mathcal{B} \otimes \mathcal{A}$.

Alors il existe une application $\mathcal{B}$-mesurable de X dans Ξ qui à tout x de X associe un Θ_x unique de $\psi(x)$.

Avant de démontrer ce théorème nous allons rappeler afin de les préciser certaines notations et propriétés dont nous nous servirons dans la démonstration.

Notons $\mathbb{N}^{*\aleph_0}$ l'ensemble de toutes les suites $\alpha = (n_1, n_2, \dots)$ où $n_1, n_2, \dots$ sont des entiers positifs.

On définit sur cet ensemble une distance d par

$$d[(m_1, m_2, \dots) , (n_1, n_2, \dots)] = \begin{cases} 0 & \text{si } m_\nu = n_\nu \quad \forall \nu \in \mathbb{N}^* \\ \dfrac{1}{\nu_0} & \text{si } \nu_0 \text{ est le plus petit indice } \nu \in \mathbb{N}^* \text{ pour lequel } m_\nu \neq n_\nu . \end{cases}$$

Alors $\mathbb{N}^{*\aleph_0}$ est un espace métrique complet, qui de plus est separable puisque l'ensemble denombrable de tous les $\alpha = (n_1, n_2, \dots)$ pour lesquels $n_\nu \neq 1$ n'est réalise que pour un nombre fini seulement de ν, est partout dense dans $\mathbb{N}^{*\aleph_0}$.

On peut d'autre part ordonner $\mathbb{N}^{*\aleph_0}$ totalement par

$$(m_1, m_2, \dots) < (n_1, n_2, \dots)$$

s'il existe ν_0 tel que

$$(1) \qquad \begin{array}{l} m_{\nu_0} < n_{\nu_0} \\ \text{et} \quad m_\nu = n_\nu \quad \text{pour} \quad \nu < \nu_0 . \end{array}$$

Si α est fixé, $\{\beta \in \mathbb{N}^{*\aleph_0} : \beta < \alpha\}$ est ouvert.

Si $\beta < \alpha$ et ν_0 defini comme dans (1) alors on a $\gamma < \alpha$ dans la sphère $d(\gamma, \beta) < \dfrac{1}{\nu_0}$.

Chaque sphère dans $\mathbb{N}^{*\aleph_0}$ est un ensemble de la forme

$$\alpha_1 \leqslant \beta < \alpha_2 \qquad (\alpha_1 \text{ et } \alpha_2 \text{ fixés}) .$$

Si son centre est $(m_1, m_2, \ldots)$ et son rayon $\varepsilon > o$, alors $\beta = (n_1, n_2, \ldots)$ appartient à cette sphère si et seulement si

$$d\{(m_1, m_2, \ldots), (n_1, n_2, \ldots)\} < \varepsilon$$

c'est-à-dire

$$m_\nu = n_\nu \qquad \text{pour} \quad \nu = 1, \ldots, \nu_1$$

où ν_1 est le plus grand entier $\leqslant \frac{1}{\varepsilon}$.

On peut l'écrire $\alpha_1 \leqslant \beta < \alpha_2$ si on pose

$$\alpha_1 = (m_1, \ldots, m_{\nu_1 - 1}, m_{\nu_1}, 1, 1, \ldots)$$
$$\alpha_2 = (m_1, \ldots, m_{\nu_1 - 1}, m_{\nu_1} + 1, 1, 1, \ldots)$$

Rappelons enfin deux résultats [3] :

- Dans un espace polonais, pour tout ensemble borélien B , il existe une application continue surjective de $N^{*\chi_o}$ sur B .
- Le produit cartésien d'une famille dénombrable d'espaces polonais est un espace polonais.

Etant donné $z = (x, \theta) \in X \times \Xi$, considérons x comme fonction de z ; posons $x = p_X(z)$. L'application $p_X : X \times \Xi \to X$ est appelée projection sur X . On sait que c'est une application continue.

Démonstration du théorème II. 5.-

Nous la ferons en 4 étapes en adaptant un lemme de Von NEUMANN [11] .

1) $G(\psi) \in B \otimes A$ par hypothèse, donc il existe une application continue f , telle que

$$f(N^{*X_o}) = G(\psi)$$

Soit $x \in X$ et soit $T(x) = \{\alpha \in N^{*X_o} : p_X \circ f(\alpha) = x\}$.

$T(x) \neq \emptyset$ puisque la projection sur X de $G(\psi)$ est X .

De plus l'application $p_X \circ f$ de N^{*X_o} sur X étant continue, $T(x)$ est fermé puisque $\{x\}$ est fermé dans X .

Notons n_1^o le plus petit n_1 avec $(n_1, n_2, \ldots) \in T(x)$

n_2^o n_2 $(n_1^o, n_2, \ldots) \in T(x)$

n_3^o n_3 $(n_1^o, n_2^o, n_3, \ldots) \in T(x)$

Posons $\alpha(x) = (n_1^o, n_2^o, \ldots)$; $\alpha(x) \in T(x)$.

D'après sa définition $\alpha(x) \leqslant \beta$, $\forall \beta \in T(x)$. Il en résulte que $T(x)$ a un plus petit élément et que c'est $\alpha(x)$.

2) Soit $\alpha_o \in N^{*X_o}$, α_o fixe .

Considérons $M(\alpha_o) = \{x \in X ; \alpha(x) < \alpha_o\}$

$$x \in M(\alpha_o) \iff \exists\, \beta \in N^{*X_o} \, , \; \beta < \alpha_o \, , \; \beta \in T(x)$$

$$\iff \exists\, \beta \in N^{*X_o} \, , \; \beta < \alpha_o \, , \; p_X \circ f(\beta) = x \, .$$

L'image continue de l'ouvert $\{\beta < \alpha_0\}$ de N^{*X_0} par $p_X \circ f$ est $\mathcal{B}$-analytique [10] et comme d'après I. 3.1., $\mathcal{A}(\mathcal{B}) = \mathcal{B}$, $M(\alpha_0)$ est $\mathcal{B}$-mesurable.

3) Si T est un sous-ensemble de N^{*X_0} , notons

$$N(T) = \{x \in X : \alpha(x) \in T\}$$

(a) Si T est une sphère, elle a la forme $\{\beta \in N^{*X_0} : \alpha_1 \leqslant \beta < \alpha_2\}$ et par consequent

$$N(T) = M(\alpha_2) \cap M^c(\alpha_1) \quad .$$

Le résultat du 2) montre alors que $N(T)$ est $\mathcal{B}$-mesurable.

(b) Si T est un ouvert de N^{*X_0} , alors

$T = \bigcup_{n \in N^*} T_n$ où T_n est une sphère de N^{*X_0}

$N(T) = \bigcup_{n \in N^*} N(T_n)$ et comme $N(T_n)$ est $\mathcal{B}$-mesurable $\forall\, n \in N^*$, il en est de même de $N(T)$.

4) Soit finalement O un ouvert de $X \times E$

Notons $T(O) = \{\alpha \in N^{*X_0} : f(\alpha) \in O\}$.

Comme O est ouvert et f continue, $T(O)$ est 1 ouvert de N^{*X_0} . Il en résulte d'après 3) que $N(T(O))$ est $\mathcal{B}$-mesurable.

Remarquons enfin que $x \in N(T(O))$ signifie que $\alpha(x) \in T(O)$ d'où $f(\alpha(x)) \in O$. Or $\alpha(x) \in T(x)$, par conséquent $p_X \circ f\,[\alpha(x)] = x$.

L'application qui à $x \in X$ fait correspondre $f[\alpha(x)] \in G(\psi)$ est mesurable. Comme de plus $p_X \circ f\,[\alpha(x)] = x$, on a $f[\alpha(x)] = (x, \theta_x)$ où $\theta_x \in \psi(x)$.

L'application : $x \in X \to (x,\Theta_x) \in G(\psi)$ étant mesurable, il en est de même de l'application

$$x \in X \to \Theta_x \in \Xi$$

puisque c'est l'application composée de la précédente et de p_Ξ .

Il suffit alors de remarquer qu'un espace compact métrisable est un espace polonais pour obtenir le résultat suivant à l'aide des théorèmes II. 1. et II. 5.

II. 6. THEOREME.

Soit $(X,\mathcal{B},P)$ *un espace probabilisé complet où* X *est un espace polonais et* $\mathcal{B}$ *sa tribu borélienne.*

Soit (Ξ,A) *l'espace probabilisable formé d'un espace compact métrisable muni de sa tribu borélienne.*

Soit $L : X \times \Xi \to \mathbb{R}$, *une application* $\mathcal{B} \otimes A$ *mesurable telle que* $L(x,.)$ *soit semi-continue supérieurement sur* Ξ , *quel que soit* x *élément de* X .

Alors il existe une application $\hat{\Theta}$ *mesurable de* X *dans* Ξ *telle que*

$$\forall\, x \in X \quad , \quad L(x,\hat{\Theta}(x)) \geq c \sup_{\Theta \in \Xi} L(x,\Theta) \quad , \quad c \in]0,1] \ .$$

Nous pouvons maintenant appliquer les resultats de ce chapitre à un problème d'estimation.

Supposons que l'on etudie un phenomène economique ou physique et une de ses caracteristiques X. En general celle-ci est variable et on ne peut en prevoir exactement les valeurs. On associe alors au phenomène un modèle aleatoire et X sera considerée ainsi comme une variable aleatoire.

Il est bien rare que la loi de probabilite de X soit entièrement specifiee ; souvent on en connait la forme mais on introduit des paramètres dont a priori, on ne sait que leur appartenance à un certain ensemble ; si la loi de X est a densite, on aura celle-ci sous la forme $f(x,\theta)$ où f est connue et où $\theta \in \Xi$.

Nous nous interessons ici à l'estimation ponctuelle qui, à une réalisation d'un echantillon, associe une seule valeur du paramètre.

Pour la methode du quasi-maximum de vraisemblance et dans le cas de variables aleatoires independantes et de même loi que X, nous ferons les hypothèses suivantes :

1) $\theta \in \Xi$ espace compact metrisable et Λ est la tribu borelienne de Ξ.
2) X est definie sur (Ω,A), à valeurs dans $(\mathcal{X},\mathcal{B})$ où $\mathcal{X}$ est un espace polonais, $P_X(.;\theta)$ est sa loi ;
3) $\forall\, \theta \in \Xi$, il existe une mesure μ σ-finie telle que $(\mathcal{X},\mathcal{B},\mu)$

soit complet

et que $$P_X(\cdot;\Theta) << \mu$$

Soit $f(x,\Theta) = \dfrac{d\, P_X(\cdot;\Theta)}{d\, \mu}$ la densite de Radon-Nikodym de $P_X(\cdot;\Theta)$ par rapport a μ .

4) $f(x,\Theta)$ est $\mathcal{B} \otimes \mathcal{A}$ mesurable et $\forall\, x \in \mathcal{X}$, $f(x,\cdot)$ est semi-continue superieurement sur Ξ .

Alors si $(X_1,\ldots,X_n)$ est un n-echantillon de X , il existe une application $\overset{n}{\otimes}\mathcal{B}$-mesurable Θ_n^* de $\mathcal{X}^n$ dans Ξ telle que

$$\forall\, \Theta \in \Xi \ , \quad \prod_{i=1}^{n} f(x_i\ ,\ \Theta_n^*(x_1,\ldots,x_n)) \geqslant c \prod_{i=1}^{n} f(x_i,\Theta) \quad ; \quad c \in]0,1]$$

d'ou l'existence d'une application $\hat{\Theta}_n$, $\mathcal{A}$-mesurable, de Ω dans Ξ avec $\hat{\Theta}_n = \Theta_n^* \circ [X_1,X_2,\ldots,X_n]$ qui sera un estimateur du quasi-maximum de vraisemblance.

Dans le cas d'un processus de Markov homogène a temps discret, $((\mathcal{X},\mathcal{B})\ ,\ P^{(t)}(\Theta))$ ou $\mathcal{X}$ est un espace polonais, et pour la methode du maximum de vraisemblance nous avons besoin de conditions dont les suivantes :

1) Ξ est un espace compact métrisable et Λ est sa tribu borelienne.

2) il existe une mesure σ-finie μ sur $\mathcal{B}$, telle que $(\mathcal{X},\mathcal{B},\mu)$ soit un espace mesuré complet et que

$$\forall(x,\Theta) \in \mathcal{X} \times \Xi \quad , \quad P(x,\cdot;\Theta) << \mu$$

3) la densité de Radon-Nikodym $p(x,\cdot;\Theta)$ de $P(x,\cdot;\Theta)$ par rapport a μ est une fonction $\mathcal{B}^2 \otimes \Lambda$ mesurable.

4) $\forall(x,y) \in \mathcal{X}^2$, $p(x,y;\cdot)$ est semi-continue supérieurement en Θ .

CHAPITRE III

SUR LA CONVERGENCE PRESQUE SURE DES ESTIMATEURS DU QUASI-MAXIMUM DE VRAISEMBLANCE POUR DES PROCESSUS DE MARKOV

L'existence d'un estimateur du quasi-maximum de vraisemblance vient d'être étudiée sous certaines hypothèses dans le chapitre précédent. Remarquons que nous n'avons imposé aucune condition quant à la dérivabilité, par rapport au paramètre étudié, de la fonction de vraisemblance.

En admettant l'existence d'un estimateur du quasi-maximum de vraisemblance, nous étudierons ici sa convergence vers la vraie valeur (inconnue) du paramètre.

Ce sujet a déjà fait l'objet d'un grand nombre de travaux. La plupart des démonstrations, qui étaient basées sur la convergence de zéros convenablement choisis des équations de vraisemblance successives, demandaient l'hypothèse forte de l'existence **des dérivées** d'ordre trois des densités par rapport au paramètre étudié. Wald (cf. [13]) fut le premier à n'y pas faire appel dans le cas de variables aléatoires indépendantes et de même loi.

Plus récemment ROUSSAS [13] étendit à certains processus de MARKOV le résultat de WALD.

Suivant une démarche analogue nous avons démontré la convergence presque sûre d'un estimateur du quasi-maximum de vraisemblance dans le cas du processus de Markov homogènes et à temps discret définis comme dans [4] .

Les resultats des chapitres II et III nous permettront alors de donner une demonstration de l'existence et de la convergence presque sûre vers la vraie valeur d'un estimateur du maximum de vraisemblance et ceci sans faire appel à la derivabilité de la fonction de vraisenblance par rapport au paramètre etudié.

Ce chapitre comporte 2 paragraphes : le premier est consacré au cas où l'espace des etats est un espace mesurable $(\mathcal{X},\mathcal{B})$; dans le second nous montrerons quelles simplifications entraine l'hypothèse que $\mathcal{X}$ ne contienne qu'un nombre fini de points. A l'aide d'un exemple nous montrerons enfin un aspect interessant de cette étude.

§ 1. *Cas général*.-

Soit $(\Omega,\mathcal{A},\Pr(.;\theta))$ un espace de probabilité ; la mesure de probabilité $\Pr(.;\theta)$ dépend d'un paramètre $\theta \in \Xi$ où Ξ est un espace que nous préciserons par la suite et qui contient la vraie valeur inconnue θ_o du paramètre.

Soit $(X_n)_{n \in N^*}$ une fonction aléatoire markovienne définie sur $(\Omega,\mathcal{A},\Pr(.;\theta))$ à valeurs dans un espace mesurable $(\mathcal{X},\mathcal{B})$, telle que

$\forall\, n \in N^*$, $\forall\, A \in \mathcal{B}$, $\forall\, x \in \mathcal{X}$ on ait pour tout $\theta \in \Xi$,

$$P(x,A;\theta) = \Pr[X_{n+1} \in A \mid X_n = x;\theta] .$$

La fonction aléatoire $(X_n)_{n \in \mathbb{N}^*}$ est donc attachée au processus de MARKOV $((X,\mathcal{B}) , P^{(t)}(\Theta))$.

Hypothèse A :

On supposera que pour chaque $\Theta \in \Xi$ il existe une probabilité absolue stationnaire unique $Q(.;\Theta)$ sur $\mathcal{B}$ telle que

$$\forall A \in \mathcal{B} \quad , \quad Q(A;\Theta) = \int_X P(x,A;\Theta)\ Q(dx;\Theta)$$

et que la fonction aléatoire $(X_n)_{n \in \mathbb{N}^*}$ est stationnaire.

Hypothèse B :

Il existe une mesure μ , σ-finie, sur $\mathcal{B}$, indépendante de Θ , par rapport à laquelle les probabilités de transition et la probabilité absolue stationnaire sont absolument continues :

$$\forall \Theta \in \Xi \quad , \quad \forall x \in X \quad , \quad P(x,.;\Theta) << \mu$$
$$\text{et} \quad Q(.;\Theta) << \mu \quad .$$

On a donc

$$\forall A \in \mathcal{B} \quad , \quad P(x,A;\Theta) = \int_A p(x,y;\Theta)\ \mu(dy)$$
$$Q(A,\Theta) = \int_A q(x;\Theta)\ \mu(dx) \quad .$$

On supposera de plus que $p(x,y;\Theta)$ est $\mathcal{B}^2$-mesurable $\forall \Theta \in \Xi$.

Hypothèse C :

Pour chaque $n \in \mathbb{N}^*$, il existe un ensemble $\Omega^{(n)} \in \mathcal{A}$, independant de Θ tel que $\Pr(\Omega^{(n)};\Theta) = 1$, $\forall\ \Theta \in \Xi$, et pour lequel

$$\forall\ \omega \in \Omega^{(n)}\ ,\quad L_n(\omega,\Theta) = q[X_1(\omega);\Theta] \prod_{i=1}^{n-1} p[X_i(\omega),X_{i+1}(\omega);\Theta] > 0 \quad .$$

Nous supposerons par la suite que les hypothèses A,B et C sont verifiées, en particulier nous prendrons toujours ω dans $\bigcap_{n \in \mathbb{N}^*} \Omega^{(n)}$.

Comme dans [4], si A et B sont des élements de $\mathcal{B}$ et si on dispose d'un "morceau" de trajectoire $(X_t(\omega))_{t \in [1,n] \cap \mathbb{N}}$ correspondant à $\omega \in \Omega$, nous appelerons $N_n(A \times B;\omega)$ le nombre de fois (eventuellement nul) que $X_t(\omega) \in A$ et $X_{t+1}(\omega) \in B$; soit

$$N_n(A \times B;\omega) = \sum_{t=1}^{n-1} [1_{A \times B} \circ (X_t,X_{t+1})]\ (\omega) \quad .$$

Il est clair que $N_n(\cdot;\omega)$ est une mesure sur $(\mathcal{X}^2,\mathcal{B}^2)$ à valeurs dans $[0,n-1] \cap \mathbb{N}$.

On peut alors écrire le logarithme de la fonction de vraisemblance sous la forme suivante

$$\mathrm{Log}\ L_n(\omega;\Theta) = \mathrm{Log}\ q[X_1(\omega);\Theta] + \int_{\mathcal{X}^2} \mathrm{Log}\ p(x,y;\Theta)\ dN_n(x,y;\omega) \quad .$$

Considerons alors un estimateur $\hat{\Theta}_n$ du quasi-maximum de vraisemblance a coefficient $c \in]0,1]$.

III. 1. PROPOSITION.

Si $\forall\, \Theta \in \Xi$, $\sup\limits_{\omega \in \Omega} q[X_1(\omega);\Theta]$ *est fini, alors*

$$\limsup_{n \to +\infty} \frac{1}{n} \int_{X^2} \operatorname{Log} \frac{p(x,y;\Theta_0)}{p(x,y;\hat{\Theta}_n(\omega))} \, dN_n(x,y;\omega) \leqslant 0$$

En effet, $\forall\, \Theta \in \Xi$, d'après la définition de $\hat{\Theta}_n$,

$$\operatorname{Log} c + \operatorname{Log} q[X_1(\omega);\Theta] + \int_{X^2} \operatorname{Log} p(x,y;\Theta)\, dN_n(x,y;\omega)$$

$$\leqslant \operatorname{Log} q[X_1(\omega);\hat{\Theta}_n(\omega)] + \int_{X^2} \operatorname{Log} p(x,y;\hat{\Theta}_n(\omega))\, dN_n(x,y;\omega) \quad .$$

On a en particulier pour Θ_o

$$\frac{1}{n} \int_{X^2} \operatorname{Log} \frac{p(x,y;\Theta_0)}{p(x,y;\hat{\Theta}_n(\omega))} \, dN_n(x,y,\omega) \leqslant -\frac{1}{n} \operatorname{Log} c - \frac{1}{n} \operatorname{Log} \frac{q[X_1(\omega);\Theta_o]}{q[X_1(\omega);\hat{\Theta}_n(\omega)]}$$

d'où à l'aide de l'hypothèse

$$\limsup_{n \to +\infty} \frac{1}{n} \int_{X^2} \operatorname{Log} \frac{p(x,y;\Theta_0)}{p(x,y;\hat{\Theta}_n(\omega))} \, dN_n(x,y;\omega) \leqslant 0$$

III. 2. PROPOSITION.

Une condition nécessaire et suffisante pour que

$$\int_{X^2} \operatorname{Log} \frac{p(x,y;\Theta_0)}{p(x,y;\Theta)} \, Q(dx;\Theta_o)\, P(x,dy;\Theta_o) > 0 \qquad \forall\, \Theta \in \Xi \;,\; \Theta \neq \Theta_o$$

est que

$$\forall\, \Theta_1,\Theta_2 \in \Xi \;,\; \Theta_1 \neq \Theta_2 \;,\; \int_{X^2} |p(x,y;\Theta_1) - p(x,y;\Theta_2)| \; Q(dx;\Theta_o)\, P(x,dy;\Theta_o) > 0$$

On sait que $e^z \geqslant 1 + z$, $\forall z \in \mathbb{R}$ et $e^z = 1 + z \iff z = 0$.

Soit $\theta \neq \theta_o$; remplaçant z par $\operatorname{Log} \frac{p(x,y;\theta)}{p(x,y;\theta_o)}$ on a

$$\frac{p(x,y,\theta)}{p(x,y,\theta_o)} \geqslant 1 + \operatorname{Log} \frac{p(x,y;\theta)}{p(x,y;\theta_o)}$$

l'egalite ayant lieu si et seulement si $p(x,y,\theta) = p(x,y;\theta_o)$.

Or puisque

$$\int_{X^2} \frac{p(x,y;\theta)}{p(x,y;\theta_o)} \; Q(dx;\theta_o) \; P(x,dy;\theta_o) =$$

$$\int_{X^2} p(x,y;\theta) \; q(x,\theta_o) \; d\mu(x) \; d\mu(y) = 1$$

on a

$$H(\theta) = \int_{X^2} \operatorname{Log} \frac{p(x,y;\theta_o)}{p(x,y;\theta)} \; Q(dx;\theta_o) \; P(x,dy;\theta_o) \geqslant 0 \quad .$$

Soit $A = \{(x,y) \in X^2 : p(x,y;\theta_o) \neq p(x,y,\theta)\}$.

Si A est négligeable pour la mesure envisagée ci-dessus, alors $H(\theta) = 0$. Sinon, c'est-à-dire sous l'hypothèse que nous avons faite, montrons que $H(\theta) > 0$.

En effet, supposons que $H(\theta)$ puisse être nul sous la condition imposée, alors on aurait,

$$0 = \int_A \operatorname{Log} \frac{p(x,y;\theta_o)}{p(x,y;\theta)} \; Q(dx;\theta_o) \; P(x,dy;\theta_o)$$

$$> \int_A \left(1 - \frac{p(x,y;\theta_o)}{p(x,y;\theta)}\right) Q(dx;\theta_o) \; P(x,dy;\theta_o)$$

l'inegalité étant stricte puisque A n'est pas negligeable et que sur A

$$\frac{p(x,y;\Theta)}{p(x,y;\Theta_o)} > 1 + \mathrm{Log}\,\frac{p(x,y;\Theta)}{p(x,y,\Theta_o)}$$

or
$$\int_A \left(1 - \frac{p(x,y;\Theta_o)}{p(x,y;\Theta)}\right) Q(dx;\Theta_o)\, P(x,dy;\Theta_o) =$$

$$\int_{X^2} \left(1 - \frac{p(x,y;\Theta_o)}{p(x,y;\Theta)}\right) Q(dx;\Theta_o)\, P(x,dy;\Theta_o) = 0$$

ce qui donne la contradiction puisque $H(\Theta) = 0$ par hypothèse.

III. 3. PROPOSITION.

Nous ferons les hypothèses suivantes

1) Ξ *est un espace compact métrisable.*

2) $\forall (x,y) \in X^2$, $p(x,y;.)$ *est semi-continue supérieurement sur* Ξ .

3) $\forall\, \Theta \in \Xi$, *il existe un voisinage* $W(\Theta)$ *de* Θ *tel que pour chaque ouvert* V *avec* $V \ni \Theta$, $V \subset W(\Theta)$, $\sup\limits_{\Theta' \in V} p(x,y;\Theta')$ *est* $\mathcal{B}^2$*-mesurable.*

4) il existe une fonction $h(x,y)$ *telle que*

$$h(x,y) \leq \inf_{\Theta \in \Xi} \mathrm{Log}\,\frac{p(x,y;\Theta_o)}{p(x,y;\Theta)}$$

et
$$\left| \int_{X^2} h(x,y)\, Q(dx;\Theta_o)\, P(x,dy;\Theta_o) \right| < +\infty$$

5) pour chaque $\Theta \in \Xi$, $Q(.;\Theta)$ *qui existe et est unique d'après l'hypothèse* A *vérifie la propriété suivante :*

$$\forall\, x \in X \ ,\ P(x,.;\Theta) \ll Q(.;\Theta) \quad .$$

Sous ces hypothèses, on peut trouver pour chaque voisinage $U = U(\theta_o)$ *de* θ_o *un nombre* $\delta = \delta[U(\theta_o)] > 0$ *tel que*

$$\lim_{n \to +\infty} \inf \left[\inf_{\theta \in U^c} \left\{\frac{1}{n} \int_{X^2} \operatorname{Log} \frac{p(x,y;\theta_o)}{p(x,y;\theta)} \, dN_n(x,y;\omega)\right\}\right] > \delta \qquad \text{p.s. } Pr(\cdot;\theta_o)$$

On definit sur Ξ une metrique d compatible avec sa topologie. Soit $t \in U^c$; notons $V_k(t) = \{\theta : d(\theta,t) < \frac{1}{k}\}$. Lorsque $k \nearrow +\infty$, il est évident que $\inf\limits_{\theta \in V_k(t)} \operatorname{Log} \frac{p(x,y;\theta_o)}{p(x,y;\theta)}$ ne décroit pas.

La semi continuité supérieure en θ de $p(x,y;\theta)$ implique celle de $\operatorname{Log} \frac{p(x,y;\theta)}{p(x,y;\theta_o)}$, d'où la semi-continuité inférieure en θ de $\operatorname{Log} \frac{p(x,y;\theta_o)}{p(x,y;\theta)}$; donc $\forall \varepsilon > 0$, $\exists V$ voisinage de t tel que

$$\forall \theta \in V \quad , \quad \operatorname{Log} \frac{p(x,y;\theta_o)}{p(x,y;t)} - \operatorname{Log} \frac{p(x,y;\theta_o)}{p(x,y;\theta)} < \varepsilon \qquad .$$

Il en résulte que lorsque $k \nearrow +\infty$

$$\inf_{\theta \in V_k(t)} \operatorname{Log} \frac{p(x,y;\theta_o)}{p(x,y;\theta)} \quad \text{tend sans décroître vers} \quad \operatorname{Log} \frac{p(x,y;\theta_o)}{p(x,y;t)} \quad .$$

L'hypothèse 3) implique la $\mathcal{E}^2$-mesurabilité de $\sup\limits_{\theta \in V_k(t)} p(x,y;\theta)$ pour k suffisamment grand ; d'où la $\mathcal{E}^2$-mesurabilité de

$$\inf_{\theta \in V_k(t)} \operatorname{Log} \frac{p(x,y;\theta_o)}{p(x,y;\theta)} \quad .$$

On peut alors appliquer le théorème de convergence monotone ; d'où

$$\int_{X^2} \inf_{\theta \in V_k(t)} \operatorname{Log} \frac{p(x,y;\theta_o)}{p(x,y;\theta)} \; Q(dx;\theta_o) \; P(x,dy;\theta_o)$$

tend en croissant, lorsque $k \nearrow +\infty$, vers

$$\int_{\mathcal{X}^2} \text{Log}\, \frac{p(x,y;\theta_0)}{p(x,y;t)} Q(dx;\theta_0)\, P(x,dy;\theta_0) = H(t)$$

et $H(t) > 0$ d'après la proposition III. 2.

Il en resulte que, $\forall\, t \in U^c$ il existe un ouvert $V_k(t)$ contenant t et un entier positif $N(t)$ tel que

$$k > N(t) \Rightarrow \int_{\mathcal{X}^2} \inf_{\theta \in V_k(t)} \text{Log}\, \frac{p(x,y;\theta_0)}{p(x,y;\theta)}\, Q(dx;\theta_0)\, P(x,dy;\theta_0) \geqslant \frac{1}{2} H(t) \ .$$

Prenons alors U ouvert ; U^c est donc compact : on peut recouvrir U^c par un nombre fini d'ensembles $V_k(t_i)$ $(i = 1,2,\dots,m)$.

Si $N = \max\limits_{i=1,\dots,m} N(t_i)$, on choisira $k > N$ de sorte que

$$\forall\, i = 1,2,\dots,m \ , \quad \int_{\mathcal{X}^2} \inf_{\theta \in V_k(t_i)} \text{Log}\, \frac{p(x,y;\theta_0)}{p(x,y;\theta)}\, Q(dx;\theta_0)\, P(x,dy;\theta_0) \geqslant \delta > 0$$

où $\delta = \min\limits_{i=1,\dots,m} \frac{1}{2} H(t_i)$.

Or $\forall\, \theta \in U^c$, $\exists\, i \in \{1,2,\dots,m\}$ tel que $\theta \in V_k(t_i)$ et on a

$$\inf_{\theta \in V_k(t_i)} \text{Log}\, \frac{p(x,y;\theta_0)}{p(x,y;\theta)} \leqslant \text{Log}\, \frac{p(x,y;\theta_0)}{p(x,y;\theta)}$$

or l'hypothèse 5) et la $\mathcal{E}^2$-mesurabilite de $\inf\limits_{\theta \in V_k(t_i)} \text{Log}\, \frac{p(x,y;\theta_0)}{p(x,y;\theta)}$ entraine d'après [2] (theorème 2.1) que

$$\lim_{n \to +\infty} \frac{1}{n} \int_{\mathcal{X}^2} \inf_{\theta \in V_k(t_i)} \text{Log}\, \frac{p(x,y;\theta_0)}{p(x,y;\theta)}\, dN_n(x,y;\omega)$$

$$\underset{\text{p.s. } Pr(.;\theta_0)}{=} \int_{\mathcal{X}^2} \inf_{\theta \in V_k(t_i)} \text{Log}\, \frac{p(x,y;\theta_0)}{p(x,y;\theta)}\, Q(dx;\theta_0)\, P(x,dy;\theta_0)$$

et $\displaystyle\int_{X^2} \inf_{\theta \in V_k(t_i)} \text{Log} \frac{p(x,y;\theta_0)}{p(x,y;\theta)} \, Q(dx;\theta_0) \, P(x,dy;\theta_0) \geq \frac{1}{2} H(t_i)$ puisqu'on a choisi $k > N$.

Par conséquent

$$\lim_{n \to +\infty} \inf \left[\min_{i = 1,\dots,m} \frac{1}{n} \int_{X^2} \inf_{\theta \in V_k(t_i)} \text{Log} \frac{p(x,y;\theta_0)}{p(x,y;\theta)} \, dN_n(x,y;\omega) \right] \geq \delta$$

p.s. $Pr(\cdot;\theta_0)$

Mais

$$\lim_{n \to +\infty} \inf \left[\inf_{\theta \in U^c} \frac{1}{n} \int_{X^2} \text{Log} \frac{p(x,y;\theta_0)}{p(x,y;\theta)} \, dN_n(x,y;\omega) \right]$$

$$\geq \lim_{n \to +\infty} \inf \left[\min_{i = 1,\dots,m} \frac{1}{n} \int_{X^2} \inf_{\theta \in V_k(t_i)} \text{Log} \frac{p(x,y;\theta_0)}{p(x,y;\theta)} \, dN_n(x,y;\omega) \right]$$

d'ou

$$\inf_{n \to +\infty} \left[\inf_{\theta \in U^c} \frac{1}{n} \int_{X^2} \text{Log} \frac{p(x,y;\theta_0)}{p(x,y;\theta)} \, dN_n(x,y;\omega) \right] \underset{\text{p.s. } Pr(\cdot;\theta_0)}{\geq} \delta > 0$$

A l'aide des résultats établis dans ce paragraphe on peut alors énoncer le théorème suivant.

III. 4. THEOREME.

Soit $(X_n)_{n \in N^*}$ *une fonction aléatoire markovienne stationnaire définie sur l'espace mesurable* (Ω, A), *attachée au processus de Markov discret homogène* $((X,\mathcal{B}), P^{(t)}(\theta))$ *dont la probabilité de transition dépend d'un paramètre* $\theta \in \Xi$, *où* Ξ *est un espace compact métrisable.*

On fera les hypothèses suivantes :

1) $\forall\ \Theta \in \Xi$, *il existe une probabilité absolue stationnaire unique* $Q(.;\Theta)$ *et*

$$\forall\ x \in X \qquad P(x,.;\Theta) << Q(.;\Theta)$$

2) *il existe une mesure* μ *sur* $(X,\mathcal{B})$, *σ-finie, indépendante de* Θ *telle que*

$$Q(.;\Theta) << \mu \quad , \text{ et on notera } \quad q(x;\Theta) = \frac{dQ(.;\Theta)}{d\,\mu}$$

3) $\forall\ \Theta \in \Xi$, $\sup\limits_{\omega \in \Omega} q[X_1(\omega);\Theta]$ *est fini*

4) $\forall\ \Theta \in \Xi$, $p(x,y;\Theta)$ *densité de Radon-Nikodym de* P *par rapport à* μ , *est* $\mathcal{B}^2$*-mesurable et* $\forall (x,y) \in X^2$, $p(x,y;.)$ *est semi-continue supérieurement sur* Ξ .

5) $\forall\ \Theta \in \Xi$, *il existe un voisinage* $W(\Theta)$ *de* Θ *tel que pour chaque ouvert* V *avec* $V \ni \Theta$, $V \subset W(\Theta)$, $\sup\limits_{\Theta' \in V} p(x,y;\Theta')$ *est* $\mathcal{B}^2$*-mesurable.*

6) *il existe une fonction* $h(x,y)$ *telle que*

$$h(x,y) \leq \inf_{\Theta \in \Xi} \ \mathrm{Log}\, \frac{p(x,y;\Theta_0)}{p(x,y;\Theta)}$$

et

$$\left| \int_{X^2} h(x,y)\ Q(dx;\Theta_0)\ P(x,dy;\Theta_0) \right| < +\infty$$

7) $\forall\ \Theta_1$ *et* $\Theta_2 \in \Xi$, $\Theta_1 \neq \Theta_2$

$$\int_{X^2} \left| p(x,y;\Theta_1) - p(x,y;\Theta_2) \right| Q(dx;\Theta_0)\ P(x,dy;\Theta_0) > 0$$

ε) $\forall \Theta \in \Xi$, *on a presque surement pour tout* $\omega \in \Omega$

$$\forall n \in \mathbb{N}^* \ , \ L_n(\omega,\Theta) = q[X_1(\omega);\Theta] \prod_{i=1}^{n-1} p[X_i(\omega),X_{i+1}(\omega);\Theta] > 0$$

Sous ces hypothèses un estimateur du quasi-maximum de vraisemblance à coefficient $c \in]0,1]$ *converge presque surement vers la vraie valeur* Θ_0 *du paramètre.*

Il suffit en effet de rapprocher les resultats des propositions III. 1. et III. 3. pour obtenir la conclusion du théorème precedent.

On peut alors reunir les theoremes II. 6. et III. 4. et écrire l'enoncé suivant

III. 5. THEOREME.

Si dans les hypothèses du théorème III. 4. on suppose que X *est un espace polonais, que* $\mathcal{B}$ *est sa tribu borélienne et que pour* $\Theta \in \Xi$ *l'espace probabilisé* $(X,\mathcal{B},\mu)$ *est complet, si de plus on appelle* $\mathcal{A}$ *la tribu borélienne de l'espace compact métrisable* Ξ *, alors il existe un estimateur du quasi-maximum de vraisemblance et il converge presque surement vers* Θ_0 *.*

Remarque.-

Dans le cas où l'espace Ξ n'est pas compact métrisable mais localement compact à base dénombrable, nous allons montrer que sous certaines hypothèses supplementaires le resultat du théorème III. 4. est encore valable.

III. 6. PROPOSITION.

Soit Ξ *un espace localement compact à base dénombrable.*

1) il existe $h(x,y)$ *telle que*

$$h(x,y) \leq \inf_{\Theta \in \Xi} \operatorname{Log} \frac{p(x,y;\Theta_0)}{p(x,y;\Theta)} \quad \text{et} \quad \left| \int_{X^2} h(x,y)\, Q(dx;\Theta_0)\, P(x,dy;\Theta_0) \right| < \infty$$

$$2) \quad \int_{X^2} \liminf_{\Theta \to \Theta_\infty} \operatorname{Log} \frac{p(x,y;\Theta_0)}{p(x,y;\Theta)}\, Q(dx;\Theta_0)\, P(x,dy;\Theta_0) \geq \ell > 0$$

où Θ_∞ *désigne le point adjoint à* Ξ *pour le compactifier suivant Alexandroff.*

3) $(K_n)_{n \in \mathbb{N}}$ *étant une suite de compacts de* Ξ *tels que* $K_n \subset K_{n+1}$, $\forall\, n \in \mathbb{N}$, *et* $\Xi = \bigcup_{n \in \mathbb{N}} K_n$, *on suppose que*

$$\inf_{\Theta \notin K_n} \operatorname{Log} \frac{p(x,y;\Theta_0)}{p(x,y;\Theta)} \quad \text{est } \mathcal{B}^2\text{-mesurable} \quad , \quad \forall\, n \in \mathbb{N}$$

4) $\forall\, \Theta \in \Xi$, $\sup_{\omega \in \Omega} q[X_1(\omega);\Theta]$ *est fini .*

Alors il existe un compact $K \subset \Xi$, *tel qu'un estimateur du quasi-maximum de vraisemblance prenne, pour* n *suffisamment grand, presque surement ses valeurs dans* K .

En effet

$$h(x,y) \leq \inf_{\Theta \in \Xi} \operatorname{Log} \frac{p(x,y;\Theta_0)}{p(x,y;\Theta)} \leq \inf_{\Theta \notin K_n} \operatorname{Log} \frac{p(x,y;\Theta_0)}{p(x,y;\Theta)}$$

et

$$\inf_{\Theta \notin K_n} \operatorname{Log} \frac{p(x,y;\Theta_0)}{p(x,y;\Theta)} \nearrow \liminf_{\Theta \to \Theta_\infty} \operatorname{Log} \frac{p(x,y;\Theta_0)}{p(x,y;\Theta)} \quad \text{quand } n \to +\infty$$

le thèorème de convergence monotone montre alors que $\forall\ \varepsilon$, $0 < \varepsilon < \mathcal{L}$, il existe un compact K tel que

$$E\ [\inf_{\Theta \notin K} \ \mathrm{Log}\ \frac{p(x,y;\Theta_0)}{p(x,y;\Theta)}\ ] \geq \mathcal{L} - \frac{\varepsilon}{2}$$

or

$$\inf_{\Theta \notin K}\left[\frac{1}{n}\int_{X^2} \mathrm{Log}\ \frac{p(x,y,\Theta_0)}{p(x,y;\Theta)}\ dN_n(x,y;\omega)\right] \geq \frac{1}{n}\int_{X^2}\ \inf_{\Theta \notin K}\ \mathrm{Log}\ \frac{p(x,y;\Theta_0)}{p(x,y;\Theta)}\ dN_n(x,y;\omega)$$

et en appliquant le thèorème 2.1. de [2] on a pour n suffisamment grand

$$\frac{1}{n}\int_{X^2}\ \inf_{\Theta \notin K}\ \mathrm{Log}\ \frac{p(x,y;\Theta_0)}{p(x,y;\Theta)}\ dN_n(x,y;\omega) \geq \mathcal{L} - \varepsilon$$

on en deduit alors que

$$\forall\ \Theta \notin K\ ,\ \frac{1}{n}\int_{X^2} \mathrm{Log}\ \frac{p(x,y,\Theta_0)}{p(x,y,\Theta)}\ dN_n(x,y;\omega) \geq \inf_{\Theta \notin K} \frac{1}{n}\int_{X^2} \mathrm{Log}\ \frac{p(x,y;\Theta_0)}{p(x,y;\Theta)}\ dN_n(x,y;\omega)$$

$$\geq \mathcal{L} - \varepsilon > 0$$

d'où le résultat proposé puisque d'après la proposition III. 1., on sait que

$$\limsup_{n \to \infty}\ \frac{1}{n}\int_{X^2} \mathrm{Log}\ \frac{p(x,y;\Theta_0)}{p(x,y;\hat{\Theta}_n(\omega))}\ dN_n(x,y;\omega) \leq 0$$

§ 2. *Cas où X est fini.*-

Dans ce paragraphe, nous regarderons comment écrire les résultats précedents dans le cas où l'espace X est fini. Le processus de Markov prend alors le nom de chaîne de Markov à un nombre fini d'etats.

Par convention de notation, nous supposerons que X consiste en les k premiers entiers positifs, $\{1,2,\ldots,k\} = X$, et nous noterons un point quelconque de X par i ou j.

Nous nous placerons dans l'hypothèse où Ξ est un espace **métrisable** compact.

La mesure μ introduite dans l'hypothèse B du § 1. sera celle qui donne pour mesure 1 à chaque point de X.

Les densités $p(x,y;\Theta)$ sont maintenant les probabilités de transition $p(i,j,\Theta)$:

$$p(i,j;\Theta) = \Pr\left[X_{n+1} = j \mid X_n = i;\Theta\right] .$$

Les $p(i,j;\Theta)$, $(i,j) \in X^2$ forment une matrice (k,k) de transition notée $P(\Theta)$.

On supposera que la chaîne de Markov n'a qu'un seul ensemble ergodique sans sous classes cycliques, et qu'elle n'a pas d'ensembles de passage, de sorte qu'il existe un système unique de probabilités absolues stationnaires $\{q(i;\Theta)\}_{i \in X}$ telles que

$$\lim_{n \to +\infty} p^{(n)}(i,j;\Theta) = q(j;\Theta) > 0 .$$

Soit maintenant $(X_n)_{n \in N^*}$ la fonction aleatoire stationnaire du processus, ayant $(q(i,\Theta))_{i \in X}$ comme probabilités absolues.

L'observation sera faite pour cette fonction aléatoire à l'aide d'un "morceau" de trajectoire de taille n . Nous noterons $N_n(i,j;\omega)$ le nombre de fois (éventuellement nul) que $X_t(\omega) = i$ et $X_{t+1}(\omega) = j$ pour $1 \leqslant t \leqslant n-1$.

Les hypothèses faites au paragraphe § 1. se réduisent alors aux suivantes :

1°) $p(i,j;\theta_1) \neq p(i,j;\theta_2)$ si $\theta_1 \neq \theta_2$ au moins pour un $(i,j) \in X^2$.

2°) $\forall (i,j) \in X^2$, $p(i,j,.)$ est semi-continue superieurement sur Ξ.

3°) $\forall (i,j) \in X^2$, $\forall \theta \in \Xi$, $p(i,j;\theta) > 0$.

III. 7. THEOREME.

Dans les hypothèses du paragraphe 2, un estimateur du quasi-maximum de vraisemblance à coefficient $c \in]0,1]$, *défini par*

$\forall \theta \in \Xi$

$$c\, q[X_1(\omega);\theta) \prod_{i,j \in X} p(i,j;\theta)^{N_n(i,j;\omega)} \leqslant q[X_1(\omega);\hat{\theta}_n(\omega)] \prod_{i,j \in X} p(i,j;\hat{\theta}_n(\omega))^{N_n(i,j;\omega)}$$

est convergent presque sûrement vers la vraie valeur θ_c *du paramètre.*

La démonstration de ce theorème se fait d'une manière analogue à celle employee dans le cas general. Nous ne la ferons pas ici.

Nous nous bornerons à traiter un exemple d'application de ce theorème.

Nous l'avons choisi de telle manière que les conditions de derivabilite de la fonction vraisemblance par rapport au paramètre qui sont generalement imposees n'étant pas verifiées, le theorème precédent nous permette de conclure et que de plus, dans le cas où celles-ci sont realisées l'estimateur que nous obtenons soit celui mis en évidence dans les exposes "classiques".

III. 8. EXEMPLE.

Soit $(X, P(X))$ l'espace des etats, $X = \{1,2\}$.

Prenons pour espace des paramètres $\Xi = [0,1]$; soit Λ sa tribu borelienne.

Soit la chaine de Markov dont la matrice de transition est

$$P_\theta = \begin{pmatrix} p(1,1;\theta) & p(1,2;\theta) \\ p(2,1;\theta) & p(2,2;\theta) \end{pmatrix}$$

où

$$p(1,1;\theta) = p(2,1;\theta) = (\frac{1}{8} + a\theta)\, 1_{[0,\frac{1}{2}[}(\theta) + [\theta(\frac{1}{2} - a) + a - \frac{1}{8}]\, 1_{[\frac{1}{2},1]}(\theta)$$

$$p(1,2;\theta) = p(2,2;\theta) = (\frac{7}{8} - a\theta)\, 1_{[0,\frac{1}{2}[}(\theta) + [\theta(a - \frac{1}{2}) + \frac{9}{8} - a]\, 1_{[\frac{1}{2},1]}(\theta)$$

avec $a \in]0,\frac{1}{4}]$.

Remarquons que les probabilités de transition $p(i,j;.)$, $(i,j) \in X^2$ ne sont pas dérivables par rapport à θ, pour la valeur $\theta = \frac{1}{2}$, si ce n'est pour $a = \frac{1}{4}$, mais que par contre, $\forall\, a \in]0,\frac{1}{4}]$ elles sont continues en θ sur $[0,1]$.

D'autre part, la monotonie stricte en θ des $p(i,j;.)$ entraine que

$$p(i,j;\theta_1) \neq p(i,j;\theta_2) \quad \text{si} \quad \theta_1 \neq \theta_2 \;,\; \forall (i,j) \in X^2 .$$

La matrice P_θ, stochastique et à lignes identiques, est idempotente et la chaîne de Markov qui lui est associée n'a qu'un seul ensemble indécomposable qui est ergodique et sans sous classes cycliques.

Le système unique de probabilités absolues stationnaires est donc le suivant

$$q(1;\theta) = (\frac{1}{8} + a\theta)\, 1_{[0,\frac{1}{2}[}(\theta) + [\theta(\frac{1}{2} - a) + a - \frac{1}{8}]\, 1_{[\frac{1}{2},1]}(\theta)$$

$$q(2;\theta) = (\frac{7}{8} - a\theta)\, 1_{[0,\frac{1}{2}[}(\theta) + [\theta(a - \frac{1}{2}) + \frac{9}{8} - a]\, 1_{[\frac{1}{2},1]}(\theta) \ .$$

Etant donné un "morceau" de trajectoire $(X_t(\omega))_{t=1,\dots,n}$, la fonction de vraisemblance pourra s'écrire

$$L_n(\omega,\theta) \simeq p(1,1;\theta)^{N_n(X,1;\omega)}\, p(2,2;\theta)^{N_n(X,2;\omega)}$$

ou
$$N_n(X,j,\omega) = \sum_{t=1}^{n-1} 1_{X\times\{j\}} \circ (X_t, X_{t+1})(\omega) \quad , \quad j = 1,2$$

Des calculs simples permettent alors de trouver comme estimateur du maximum de vraisemblance

$$\begin{aligned}\hat{\theta}_n(\omega) = {} & (1 - 1_{[0,\frac{1}{8}]}(\frac{N_n(X,1,\omega)}{n-1}))\, 1_{[0,\frac{1}{8}]}(\frac{N_n(X,1;\omega)}{n-1}) \\ & + \frac{1}{a}\left[\frac{N_n(X,1;\omega)}{n-1} - \frac{1}{8}\right] 1_{]\frac{1}{8},\frac{1}{8}+\frac{a}{2}]}(\frac{N_n(X,1;\omega)}{n-1}) \\ & + \frac{1}{\frac{1}{2} - a}\left[\frac{N_n(X,1;\omega)}{n-1} - a + \frac{1}{8}\right] 1_{]\frac{1}{8}+\frac{a}{2},\frac{3}{8}]}(\frac{N_n(X,1;\omega)}{n-1}) \\ & + 1_{]\frac{3}{8},1]}(\frac{N_n(X,1;\omega)}{n-1})\end{aligned}$$

qui est bien une application mesurable comme le montre son expression et qui d'après le théorème III. 7., converge presque sûrement vers la vraie valeur du paramètre θ .

Dans le cas où $\mathbf{a} = \frac{1}{4}$ remarquons que l'on peut prendre comme estimateurs des probabilités de transition les suivants :

$$\hat{p}_n(1,1) = \hat{p}_n(2,1) = \frac{1}{8}\, 1_{[0,\frac{1}{8}]}\left(\frac{N_n(X,1;\omega)}{n-1}\right) + \frac{N_n(X,1;\omega)}{n-1}\, 1_{]\frac{1}{8},\frac{3}{8}]}\left(\frac{N_n(X,1;\omega)}{n-1}\right)$$
$$+ \frac{3}{8}\, 1_{]\frac{3}{8},1]}\left(\frac{N_n(X,1;\omega)}{n-1}\right)$$

$$\hat{p}_n(1,2) = \hat{p}_n(2,2) = \frac{5}{8}\, 1_{[0,\frac{5}{8}[}\left(\frac{N_n(X,2;\omega)}{n-1}\right) + \frac{N_n(X,2;\omega)}{n-1}\, 1_{[\frac{5}{8},\frac{7}{8}[}\left(\frac{N_n(X,2;\omega)}{n-1}\right)$$
$$+ \frac{7}{8}\, 1_{[\frac{7}{8},1[}\left(\frac{N_n(X,2;\omega)}{n-1}\right)$$

Or si $\mathbf{a} = \frac{1}{4}$, nous sommes sous les hypothèses de [4] et on sait qu'alors on peut prendre pour estimateurs des probabilités de transition les suivants :

$$\hat{\hat{p}}_n(i,j) = \frac{N_n(i,j;\omega)}{N_n(i,X;\omega)}$$

Remarquons alors puisque $p(1,1;\theta) = p(2,1;\theta)$ on a par exemple

$$\hat{\hat{p}}_n(1,1) = \hat{\hat{p}}_n(2,1) = \frac{N_n(1,1;\omega) + N_n(2,1;\omega)}{N_n(1,X) + N_n(2,X)} = \frac{N_n(X,1,\omega)}{n-1}$$

qui coïncide bien avec $\hat{p}_n(1,1) = \hat{p}_n(2,1)$, compte tenu du fait que nous imposons à l'estimateur de prendre ses valeurs dans l'espace des paramètres, c'est-à-dire ici dans $\left[\frac{1}{8} , \frac{3}{8}\right]$.

BIBLIOGRAPHIE

[1] D. BLACKWELL

A Borel-set not containing a graph
Ann. Math. Stat. 39 (1968) p. 1345-1347.

[2] P. BILLINGSLEY

Statistical inference for Markov processes
The University of Chicago Press (1961).

[3] N. BOURBAKI

Eléments de Mathématique. Topologie générale
Ch 9 Paris-Hermann (1958).

[4] BUI TRONG LIEU

Estimations pour des processus de Markov
Publ. Inst. Stat. Univ. Paris Vol XI Fasc. 2 (1962)
p. 73-188.

[5] BUI TRONG LIEU et CARTON

Quelques résultats de statistique mathématique concernant des processus de Markov non homogènes.
Revue Roumaine Math. Pures appl. XII (1967)
p. 1149-1172.

[6] G. DEBREU

Integration of correspondences

Fifth Berkeley symp. math. stat. prob. Vol II part I (1967) p. 351-372.

[7] C. KRAFT et L. LE CAM

A remark on the roots of the maximum likelihood equation

Ann. Math. Stat. 27 (1956) p. 1174-1177.

[8] E. MARCZEWSKI et C. RYLL-NARDZEWSKI

Projections in abstract sets

Fund. Math. 40 (1953) p. 160-164.

[9] P.A. MEYER

Une présentation de la théorie des ensembles sousliniens. Application aux processus stochastiques.

Seminaire BRELOT-CHOQUET-DENY (Théorie du potentiel) 1962/1963 p. 2.01 - 2.18.

[10] P.A. MEYER

Probabilités et potentiel

Hermann, Paris (1966).

[11] J. Von NEUMANN

On rings of operators - Reduction theory

Ann. of Math. 50 (1949) p. 401-485.

[12] J. NEVEU

Bases mathématiques du calcul des probabilités
Masson et Cie, Paris (1964)

[13] G.G. ROUSSAS

Extension to Markov processes of a result by A. Wald about the consistency of the maximum likelihood estimate
Z. Wahrscheinlichkeitstheorie 4 (1965) p. 69-73.