

UNIVERSITE DE MONTPELLIER

FACULTE DES SCIENCES

1969 - 1970

T H E S E

présentée à la Faculté des Sciences de Montpellier
pour obtenir le grade de Docteur ès-Sciences - Mathématiques

Première Thèse

ECHANTILLONNAGE DANS UNE POPULATION DE VARIABLES ALEATOIRES REELLES

Deuxième Thèse

PROPOSITIONS DONNEES PAR LA FACULTE

par

Yves ESCOUFIER

Docteur de Spécialité - Mathématiques (3e cycle)

Soutenue le 11 juin 1970 devant la Commission d'Examen.

Jury : MM. B. CHARLES Président

C. CASTAING }
G. PERENNOU } Assesseurs

N° d'enregistrement
au C.N.R.S.

AO 4508

Secrétariat des Mathématiques
Publication N° 77

(a obtenu - Mention très Honorable - Félicitations du Jury -

A ma mère

A ma femme

A mes enfants

Doyen : P. DUMONTET

Assesseurs : Ch. SAUVAGE - M. MAURIN

Doyen Honoraire : P. MATHIAS

Professeurs honoraires :

R. JACQUES	J. SOULAS
M. CASTERAS	G. DENISOT
E. CARRIERE	J. GRANIER
E. TURRIERE	Ch. BOUHET
G. CAUQUIL	

Secrétaire Général : E. SIAU

Professeurs titulaires :

M.	M. MOUSSERON	Chimie Organique
M.	J.P. ROIG	Physique
M.	P. CHATELAIN	Minéralogie et Cristallographie
M ^{le}	O. TUZET	Zoologie
M.	G. COUCHET	Mécanique Supérieure
M.	J. SALVINIEN	Chimie Physique
M ^{le}	A.M. VERGNOUX	Physique
M.	J. AVIAS	Géologie
M.	P. VIELES	Chimie
M.	J. RABINOVITCH	Physique (C.S.U.P.)
M.	E. KAHANE	Chimie Biologique
M.	J.J. MOREAU	Mécanique rationnelle
M.	B. CHARLES	Mathématiques pures
M.	J. MOTTE	Biologie Végétale
M.	R. JOUTY	Physique
M.	P. DUMONTET	Physique
M.	R. LEGENDRE	Zoologie
M.	I. ASSENMACHER	Physiologie Animale
M.	A. CASADEVALL	Chimie organique
M.	B. PISTOULET	Physique
M.	Ch. ROUMIEU	Analyse Supérieure
M.	J. ROBIN	Physique
M.	A. POTIER	Laboratoire des Acides Minéraux
M.	R. LAFONT	Physique
M.	R. JACQUIER	Chimie
M.	J. FALGUEIRETTES	Minéralogie
M.	J. REGNIER	Chimie

Mme	J. CHARLES	Mathématiques
M.	P. CAILLON	Physique
M.	J. ROUZAUD	Chimie
M.	Ch. SAUVAGE	Botanique
M.	H. CHRISTOL	Chimie (E.N.S.I)
Mme	G. VERNET	Biologie Animale
M.	L. CECCHI	Physique
M.	H. ANDRILLAT	Astronomie
M.M	M. SAVELLI	Physique
M.	M. MATTAUER	Géologie
M.	L. EUZET	Zoologie
M.	C. DELOUPY	Physique
M.	L. GRAMBAST	Botanique
M.	A. BONNET	Botanique
M.	G. LAMATY	Chimie
M.	R. MARTY	Psychophysiologie

Professeurs sans chaire :

Mme	M.L. DE POUQUES	Botanique (C.S.U.P.)
M.	P. MURAUOUR	Géophysique Appliquée
M.	G. TOURNE	Chimie
M.	J. REMY	Géologie
M.	P. DEMANGEON	Géologie
Mme	S. ROBIN	Physique
Mme	Ch. DUBOUL-RAZAVET	Géologie (C.S.U.P.)
M.	E. GROUBERT	Physique
Mme	H. GUASTALLA	Biologie Physico-Chimique
M.	F. PROUST	Géologie
Mle	A. BAILLETTE	Mathématiques (C.S.U.P.)
M.	A. BASSOMPIERRE	Physique
M.	M. MAURIN	Chimie
M.	N. ROBY	Mathématiques
Mme	N. PARIS	Physiologie Végétale
M.	J. PARIS	Zoologie
M.	R. GAUFRES	Chimie
M.	J. SOULIER	Chimie (C.S.U.P.)
M.	G. BOUGNOT	Physique
M.	R. LENEL	Biologie Animale

Professeurs associés :

M.	E. AKUTOWICZ	Mathématiques
M.	A. MICALI	Mathématiques
M.	C. VAGO	Biologie Animale
M.	K. KOHLER	Biochimie

Maitres de Conférences :

M.	P. MOLINO	Mathématiques
M.	Y. GUITTON	Physiologie Végétale (C.S.U.P.)
M.	G. BOMBRE	Physique (C.S.U.P.)
M.	C. PUISSEGUR	Biologie Animale (C.S.U.P.)
M.	P. VIALLET	Chimie (C.S.U.P.)
M.	J.M. MORETTI	Chimie
M.	L. THALER	Géologie
M.	J. LEGRAND	Physiologie Animale
M.	M. ROUZEYRE	Physique
M.	P. SABATIER	Mathématiques
M.	Ch. CASTAING	Mathématiques
M.	J.D. BAYLE	Physiologie Animale
Mme	M. HAKIM	Mathématiques
M.	R. HAKIM	Mathématiques
M.	M. BAUSSET	Mathématiques (C.S.U.P.)
M.	G. LOUPIAS	Mathématiques
M.	G. MASCHERPA	Chimie
M.	C. GOUT	Physique
M.	R. JONARD	Botanique
M.	F. LAPSCHER	Mathématiques
M.	C. COMBES	Zoologie (C.S.U.P.)
M.	B. BRUN	Chimie Physique
M.	M. DENIZOT	Biologie Végétale
M.	L. KAISER	Physique (C.S.U.P.)
M.	L. GIRAL	Chimie
M.	A. COMMEYRAS	Chimie
M.	M. AMANIEU	Sciences et Techniques de l'Eau (Hydrobiologie et Mariculture)

Maitre de Conférences associé :

M.	Y. CORMARY	Géologie
----	------------	-------	----------

Chargés d'enseignement :

M. M. LEFRANC Mathématiques
M. R. MARTY Mathématiques
M. PIETASANTA

INSTITUT UNIVERSITAIRE DE TECHNOLOGIE

Professeur titulaire :

M. R. CORRIU Chimie

Maitre de Conférences :

M. CANO Physique
M. FILLARD Physique
M. GARCIA Métallurgie-Génie Mécanique
M. J.P. TRILLES Biologie
M. G. BORDURE Physique
M. L. LASSABTERE Electricité-Electronique
M. G. LECOY Génie Electrique

Maitre de conférences associé :

M. JOUANNA Génie Civil

Chargé d'enseignement :

M. ESCOUFIER Mathématiques

Je tiens à exprimer toute ma gratitude à Monsieur le Doyen B. CHARLES dont les conseils et les encouragements m'ont permis de mener à bien ce travail. Je lui suis reconnaissant de l'honneur qu'il me fait en acceptant la présidence du Jury.

Je remercie Monsieur C. CASTAING, Maître de conférences, pour les remarques précieuses qu'il m'a faites au sujet de la rédaction de cette thèse et Monsieur G. PERENNOU, Maître de conférences, pour l'aide qu'il m'a apportée dans la formulation du problème que je me posais.

Sans la compréhension de Messieurs CORMARY et LOBERT du laboratoire d'Hydrométéorologie, de Messieurs GODRON et ROMANE du Centre d'Etudes phyto-écologique, et de Messieurs FALGUEIRETTES et FILLIATRE du laboratoire de Calcul, ce travail n'aurait pas été possible : qu'ils veuillent bien trouver ici l'expression de ma profonde reconnaissance. J'y associe pleinement les membres du secrétariat de la section des Mathématiques qui se sont chargés des problèmes matériels liés à la réalisation de ce travail.

INTRODUCTION

Les problèmes abordés dans les pages suivantes ont été soulevés par un travail réalisé en collaboration avec le laboratoire d'Hydro-météorologie de la Faculté des Sciences de Montpellier et en particulier avec Monsieur LOBERT, Ingénieur dans ce laboratoire [3].

L'étude faite avait pour but de calculer les hauteurs des chutes de pluie en tous les points d'un bassin pluviométrique (en l'occurrence celui de l'Allier) à partir des hauteurs de pluies connues en un certain nombre de points de ce bassin. La solution proposée consiste à calculer les composantes principales Y_j ($j = 1, \dots, n$) des n variables X_i et à étendre la relation :

$$X_i = \sum_{j=1}^n a_{ij} Y_j$$

aux variables Z_k (k élément d'un ensemble K quelconque) non observées.

Pour cela il faut calculer les coefficients b_{kj} qui permettront d'écrire :

$$Z_k = \sum_{j=1}^n b_{kj} Y_j$$

Plusieurs méthodes ont été utilisées à cette fin :

1) La connaissance des a_{ij} permet de tracer pour chaque valeur de j , sur des cartes de la région, des courbes, comparables à des courbes de niveaux, telles que deux variables attachées à des points contenus entre deux mêmes courbes ont le même a_{ij} . La position des stations hydrométéorologiques non observées permet alors de leur affecter des coefficients b_{kj} adaptés.

2) Si elles sont connues, les corrélations entre les variables X_i et Z_k permettent de calculer les b_{kj}

En effet on a :

$$Y_j = \sum_{i=1}^n a_{ij} X_i$$

et

$$E(Y_j Z_k) = b_{kj} E(Y_j)^2$$

D'où l'on tire.

$$b_{kj} = \frac{\sum_{i=1}^n a_{ij} E(X_i Z_k)}{E(Y_j)^2}$$

3) Enfin, il est possible pour chaque valeur de j d'utiliser une interpolation des a_{ij} pour obtenir une valeur approchée des b_{kj} . La première méthode peut apparaître d'ailleurs comme un essai grossier d'interpolation. Cette impression n'est pas fondée dans la mesure où la connaissance qu'il a du phénomène étudié permet au spécialiste de tenir compte, dans le tracé des courbes, d'éléments (tels que le relief par exemple) que la troisième méthode a tendance à atténuer.

L'expérience montre que ces trois méthodes donnent des résultats comparables, satisfaisants dès que le réseau des variables observées est assez représentatif de l'ensemble des variables.

La réalisation de ce travail nous a conduit à nous poser deux questions :

A) Les pluies reconstruites sont fonctions des a_{ij} qui sont eux-mêmes déterminés par les variables observées. Peut-on affirmer qu'il y a une certaine stabilité des a_{ij} , donc des composantes principales, dès que le nombre des variables prises en compte est assez grand ?

Nous retrouvons là un problème soulevé par H. Hotelling ([5], p.504) au sujet de la convergence des valeurs propres et des vecteurs propres de la matrice des variances et covariances de groupe de variables aléatoires tirées au hasard dans une population infinie de variables aléatoires, lorsque l'effectif des groupes tirés augmente indéfiniment. Le but de la première partie de ce travail est de résoudre ce problème.

Pour le faire, nous considérons un processus X de carré intégrable sur $\Omega \times T$ et nous étudions tout d'abord l'opérateur linéaire V qui lui est classiquement associé : $(VY)(\omega) = \int_T E(X(t) Y) X(\omega, t) \mu(dt)$

Les propositions 1, 2 et 3 rappellent que V est un opérateur de $L_2(\Omega)$ en lui-même, auto-adjoint, positif, complètement continu, qui appartient à la classe des opérateurs de Hilbert-Schmidt.

Les valeurs propres et les espaces propres dont nous étudions la convergence étant ceux des opérateurs associés à la suite

$s = (s_0, s_1, \dots) \in \mathbb{T}^{\mathbb{N}}$ par :

$$(U_n(s) Y)(\omega) = \frac{1}{n} \sum_{i=0}^{n-1} E(X(s_i) Y) X(\omega, s_i)$$

nous sommes conduits à étudier la convergence de la suite d'opérateurs $u_n(s)$ vers V , et à chercher les conditions qui la rendront uniforme.

Dans le cas général, le théorème ergodique pour les fonctions intégrables nous donne un résultat de convergence faible (Théorème 1). Sous des hypothèses supplémentaires, raisonnables pour les applications, nous établissons un théorème de convergence uniforme (théorème 2) à partir duquel nous pouvons démontrer la conjecture de Hotelling, démonstration qui à notre connaissance n'avait jamais été donnée par les statisticiens.

B) Le plus souvent la certitude d'une convergence ne suffit pas à celui qui est obligé de prendre en compte des impératifs d'ordre économique, extérieurs à la formulation mathématique du problème, mais réels : Il est certes satisfaisant de savoir que, pour n suffisamment grand, un ensemble de n variables fournit une bonne approximation d'un processus ; mais pour une application particulière, est-ce que $n = 10$ est suffisamment grand ?

Le contexte mathématique qui soustend la première partie conduit à une solution simple pour ce dernier problème ; il permet, dans la seconde partie, d'explicitier un critère pour l'appréciation de l'intérêt d'un groupe restreint de variables, relativement à l'ensemble des variables possibles.

Ce critère, qui ne demande que des calculs simples, est mis à l'épreuve

de façon positive sur deux exemples pris dans des domaines différents :
le premier concerne les données pluviométriques dont nous parlions plus haut ;
le second provient de données écologiques.

Ainsi notre travail fournit à l'expérimentateur une méthode rationnelle
pour le choix des variables qu'il prendra en compte.

P R E M I E R E P A R T I E

I. 1 - RAPPELS SUR LE PRODUIT DE DEUX ESPACES DE PROBABILITES ([7], p.66)

Soit Ω et T deux espaces arbitraires. Nous nous intéressons à l'espace produit $\Omega \times T$, ensemble des couples (ω, t) obtenus lorsque ω parcourt Ω et t parcourt T .

Soit $A \times B$ un pavé de $\Omega \times T$, c'est-à-dire un sous-ensemble de la forme :

$$A \times B = \{(\omega, t) : \omega \in A, t \in B\} \text{ avec } A \subseteq \Omega \text{ et } B \subseteq T$$

Si \mathcal{A} est une σ -algèbre de parties de Ω et \mathcal{C} une σ -algèbre de parties de T , nous pouvons parler des espaces probabilisables (ou mesurables) (Ω, \mathcal{A}) et (T, \mathcal{C}) . Nous dirons alors que le pavé $A \times B$ est mesurable si $A \in \mathcal{A}$ et $B \in \mathcal{C}$.

L'ensemble des pavés mesurables forment une semi-algèbre de Boole de parties de $\Omega \times T$; en ce sens que :

- \emptyset et $\Omega \times T$ sont des pavés mesurables

- L'intersection de deux pavés mesurables est un pavé mesurable.

$$(A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \cap B_2)$$

- Le complémentaire d'un pavé mesurable est la somme d'un nombre fini de pavés mesurables disjoints (ici deux)

$$(A \times B)^c = A^c \times B + \Omega \times B^c$$

La σ -algèbre engendrée par la semi-algèbre des pavés mesurables est notée $\mathcal{A} \times \mathcal{C}$ et appelée σ -algèbre produit tensoriel de \mathcal{A} et \mathcal{C} . L'espace mesurable $(\Omega \times T, \mathcal{A} \otimes \mathcal{C})$ est alors appelé le produit des espaces mesurables (Ω, \mathcal{A}) et (T, \mathcal{C}) .

Supposons de plus que soient définies des probabilités P sur Ω et μ sur T . Nous sommes alors en présence d'espaces de probabilités (Ω, \mathcal{A}, P) et (T, \mathcal{C}, μ) . Nous savons qu'il existe une probabilité unique Q sur $(\Omega \times T, \mathcal{A} \otimes \mathcal{C})$ telle que :

$$Q(A \times B) = P(A) \mu(B) \quad (A \in \mathcal{A}, B \in \mathcal{C})$$

L'espace de probabilité $(\Omega \times T, \mathcal{A} \otimes \mathcal{C}, Q)$ est appelé le produit des espaces de probabilités (Ω, \mathcal{A}, P) et (T, \mathcal{C}, μ)

Nous nous proposons d'étudier des processus aléatoires sur Ω qui sont des variables aléatoires réelles sur $\Omega \times T$ (nous dirons dans la suite v. a. sur $\Omega \times T$). Un processus est donc une application mesurable X de $\Omega \times T$ dans R . Nous identifierons des processus qui ne diffèrent que sur un ensemble de mesure nulle de $\Omega \times T$; nous ferons de même pour des v. a. définies sur Ω ou T .

Dans toute notre étude, nous supposons que X est de carré intégrable sur $\Omega \times T$, ce qui permet d'identifier X à un élément de $L_2(\Omega \times T)$. Cette hypothèse implique (théorème de Fubini) que pour presque tout t , la fonction $X(t) : \omega \rightarrow X(\omega, t)$ est un élément de $L_2(\Omega)$. Quitte à changer la valeur de X sur un ensemble de mesure nulle de $\Omega \times T$, nous pouvons supposer que $X(t) \in L_2(\Omega)$ pour tout t . Nous le ferons dans toute la suite ce qui revient à considérer que X est un processus du second ordre. Nous supposerons également que $X(\omega) : t \rightarrow X(\omega, t)$ est un élément de $L_2(T)$ pour tout $\omega \in \Omega$.

Nous munirons $L_2(\Omega)$ du produit scalaire habituel noté $\langle \cdot, \cdot \rangle$:

$$\langle Y, Y' \rangle = E(Y Y') = \int_{\Omega} Y(\omega) Y'(\omega) P(d\omega)$$

La norme correspondante sera notée $\| \cdot \|$. Nous utiliserons que $L_2(\Omega)$ est un espace de Hilbert pour ce produit scalaire.

Nous emploierons les mêmes notations dans $L_2(T)$.

Nous noterons $\| \cdot \|$ la norme (analogue aux précédentes) dans $L_2(\Omega \times T)$.

Nous utiliserons souvent les égalités suivantes :

$$(1) \quad \int_T \|X(t)\|^2 \mu(dt) = \int_T \left(\int_{\Omega} X^2(\omega, t) P(d\omega) \right) \mu(dt) = \|X\|^2$$

$$(2) \quad \int_{\Omega} \|X(\omega)\|^2 P(d\omega) = \int_{\Omega} \left(\int_T X^2(\omega, t) \mu(dt) \right) P(d\omega) = \|X\|^2$$



I. 2 - JUSTIFICATION DU CONTEXTE MATHEMATIQUE CHOISI.

Dans le problème évoqué en introduction, la hauteur de pluie X observée en un point t est une variable aléatoire sur un espace de probabilité (Ω, \mathcal{A}, P) représentatif du "hasard" c'est-à-dire des différents types de conditions pluviométriques sur la région.

On pourrait être tenté alors d'orienter l'étude vers la famille (X_t) des v.a. sur Ω obtenues lorsque t parcourt la région T . Ce point de vue, très courant dans l'étude des processus aléatoires est ici insuffisant car il faut pouvoir considérer la quantité :

$Y(\omega) = \int_T X(\omega, t) \mu(dt)$ qui est la "lame d'eau" tombée sur la région pour la condition pluviométrique ω .

On est donc tout naturellement conduit à considérer l'espace de probabilité (T, \mathcal{C}, μ) où μ est éventuellement la mesure naturelle normée sur T et à s'intéresser aux v.a. sur l'espace produit $(\Omega \times T, \mathcal{A} \otimes \mathcal{C}, Q)$.

Enfin, l'hypothèse $X \in L_2(\Omega \times T)$, qui est à la base de la partie théorique de notre étude, est tout à fait naturelle pour qui a en vue les problèmes statistiques soulevés en analyse multivariable puisqu'elle assure que, sauf pour un ensemble de mesure nulle dans T , les moments d'ordre deux des v.a. $X(t)$ sont finis.

II. 1 - L'OPERATEUR V ASSOCIE AU PROCESSUS X

Proposition 1 : Soit X_1 et X_2 des éléments de $L_2(\Omega \times T)$; Z un élément de $L_2(\Omega)$. La variable aléatoire Y définie par

$$Y(\omega) = \int_T E(X_1(t) Z) X_2(\omega, t) \mu(dt)$$

appartient à $L_2(\omega)$ et : $\|Y\|^2 \leq \|Z\|^2 \|X_1\|^2 \|X_2\|^2$

Démonstration :

L'inégalité de Schwarz permet d'écrire $|E(X_1(t) Z)| \leq \|X_1(t)\| \|Z\|$

$$\begin{aligned} \text{d'où } |Y(\omega)|^2 &\leq \left(\int_T |E(X_1(t) Z) X_2(\omega, t)| \mu(dt) \right)^2 \\ &\leq \left(\int_T \|X_1(t)\| \|Z\| |X_2(\omega, t)| \mu(dt) \right)^2 \\ &\leq \|Z\|^2 \left(\int_T \|X_1(t)\| |X_2(\omega, t)| \mu(dt) \right)^2 \\ &\leq \|Z\|^2 \cdot \int_T \|X_1(t)\|^2 \mu(dt) \cdot \int_T X_2^2(\omega, t) \mu(dt) \\ &\leq \|Z\|^2 \|X_1\|^2 \|X_2(\omega)\|^2 \end{aligned}$$

On en déduit :

$$\|Y\|^2 = \int_{\Omega} |Y(\omega)|^2 P(d\omega) \leq \|Z\|^2 \|X_1\|^2 \int_{\Omega} \|X_2(\omega)\|^2 P(d\omega) = \|Z\|^2 \|X_1\|^2 \|X_2\|^2$$

Corollaire : Si $X \in L_2(\Omega \times T)$, on obtient un opérateur V de $L_2(\Omega)$ en lui-même en définissant $Y = VZ$ par :

$$Y(\omega) = \int_T E(X(t) Z) X(\omega, t) \mu(dt)$$

L'opérateur V est linéaire borné, de norme $\|V\| \leq \|X\|^2$

(on appelle V l'opérateur associé au processus X).

Le résultat s'obtient de façon immédiate en faisant, dans la proposition 1 : $X_1 = X_2 = X$.

Proposition 2 : Si $X \in L_2(\Omega \times T)$, l'opérateur V qui lui est associé est autoadjoint, positif et complètement continu.

Démonstration :

* Etablissons tout d'abord que V est autoadjoint positif. Soit Z et $Y \in L_2(\Omega)$

$$\begin{aligned} \langle VZ, Y \rangle &= \int_{\Omega} \left[\int_{\mathbb{T}} E(X(t) Z) X(\omega, t) \mu(dt) \right] [Y(\omega)] P(d\omega) \\ &= \int_{\Omega} \left[\int_{\mathbb{T}} E(X(t) Z) X(\omega, t) Y(\omega) \mu(dt) \right] P(d\omega) \end{aligned}$$

Pour pouvoir intervertir les intégrations par rapport à ω et t , montrons que $E(X(t) Z) X(\omega, t) Y(\omega)$ appartient à $L_1(\Omega \times \mathbb{T})$

L'inégalité de Schwarz donne :

$$|E(X(t) Z)| \leq \|X(t)\| \|Z\|$$

$$\text{D'où : } \int_{\mathbb{T}} |E(X(t) Z)|^2 \mu(dt) \leq \|Z\|^2 \int_{\mathbb{T}} \|X(t)\|^2 \mu(dt) = \|Z\|^2 \|X\|^2$$

ce qui établit que $E(X(t) Z) \in L_2(\mathbb{T})$

Il en découle alors successivement que $E(X(t) Z) Y(\omega) \in L_2(\Omega \times \mathbb{T})$ et $E(X(t) Z) X(\omega, t) Y(\omega) \in L_1(\Omega \times \mathbb{T})$ comme produit de deux éléments de $L_2(\Omega \times \mathbb{T})$. Il est donc licite d'écrire :

$$\begin{aligned} \langle VZ, Y \rangle &= \int_{\mathbb{T}} \left[\int_{\Omega} E(X(t) Z) X(\omega, t) Y(\omega) P(d\omega) \right] \mu(dt) \\ &= \int_{\mathbb{T}} E(X(t) Z) \left[\int_{\Omega} X(\omega, t) Y(\omega) P(d\omega) \right] \mu(dt) \\ &= \int_{\mathbb{T}} E(X(t) Z) E(X(t) Y) \mu(dt) \end{aligned} \quad (4)$$

La symétrie en Z et Y de cette dernière formule établit que V est autoadjoint. La positivité en découle par :

$$\langle VZ, Z \rangle = \int_{\mathbb{T}} [E(X(t) Z)]^2 \mu(dt) \geq 0 \quad (5)$$

** Montrons maintenant que V est complètement continu en établissant qu'il est la limite d'une suite d'opérateurs de rang fini.

$X \in L_2(\Omega \times \mathbb{T})$ peut être approché au sens de $L_2(\Omega \times \mathbb{T})$ par des v.a. X_n telles que :

$$X_n(\omega, t) = \sum_{i=1}^n A_i(\omega) B_i(t)$$

où $A_i \in L_2(\Omega)$ et $B_i \in L_2(\mathbb{T})$

Définissons alors $V_n Z$ par :

$$(V_n Z) = \int_{\mathbb{T}} E(X_n(t) Z) X_n(\omega, t) \mu(dt)$$

On a :

$$\begin{aligned} (V_n Z)(\omega) &= \int_{\mathbb{T}} E\left(\sum_{i=1}^n A_i B_i(t) \cdot Z\right) \sum_{j=1}^n A_j(\omega) B_j(t) \mu(dt) \\ &= \int_{\mathbb{T}} \left[\sum_{i=1}^n E(A_i Z) B_i(t) \right] \left[\sum_{j=1}^n A_j(\omega) B_j(t) \right] \mu(dt) \\ &= \sum_{j=1}^n A_j(\omega) \left[\sum_{i=1}^n E(A_i Z) \int_{\mathbb{T}} B_i(t) B_j(t) \mu(dt) \right] \end{aligned}$$

ce qu'on peut encore écrire :

$$V_n Z = \sum_{j=2}^n \left[\sum_{i=1}^n E(A_i Z) E(B_i B_j) \right] A_j \quad (6)$$

Ceci établit que V_n est un opérateur de rang fini puisque l'espace image est contenu dans l'espace engendré par les A_j .

D'autre part :

$$\begin{aligned} (V_n Z - VZ)(\omega) &= \int_{\mathbb{T}} [E(X_n(t) Z) X_n(\omega, t) - E(X(t) Z) X(\omega, t)] \mu(dt) \\ &= \int_{\mathbb{T}} [E(X(t) Z)(X_n(\omega, t) - X(\omega, t)) + \\ &\quad + E((X_n(t) - X(t)) Z) X_n(\omega, t)] \mu(dt) \\ &= \int_{\mathbb{T}} [E(X(t) Z)(X_n(\omega, t) - X(\omega, t))] \mu(dt) + \\ &\quad + \int_{\mathbb{T}} [E((X_n(t) - X(t)) Z) X_n(\omega, t)] \mu(dt) \end{aligned}$$

Appelons $C(\omega)$ et $D(\omega)$ ces deux intégrales, nous avons :

$$\begin{aligned} \|(V_n - V)Z\|^2 &= \int_{\Omega} (C(\omega) + D(\omega))^2 P(d\omega) \\ &\leq \int_{\Omega} C^2(\omega) P(d\omega) + \int_{\Omega} D^2(\omega) P(d\omega) + \\ &\quad + 2 \left(\int_{\Omega} C^2(\omega) P(d\omega) \right)^{\frac{1}{2}} \left(\int_{\Omega} D^2(\omega) P(d\omega) \right)^{\frac{1}{2}} \\ &= \left[\left(\int_{\Omega} C^2(\omega) P(d\omega) \right)^{\frac{1}{2}} + \left(\int_{\Omega} D^2(\omega) P(d\omega) \right)^{\frac{1}{2}} \right]^2 \end{aligned}$$

La proposition 1 donne alors :

$$\int_{\Omega} C^2(\omega) P(d\omega) \leq \|Z\|^2 \|X\|^2 \|X_n - X\|^2$$

et
$$\int_{\Omega} D^2(\omega) P(d\omega) \leq \|Z\| \|X_n - X\|^2 \|X_n\|^2$$

X_n converge au sens de $L_2(\Omega \times T)$ vers X ; donc pour n assez grand, on aura : $\|X_n\| \leq 4\|X\|$ et par conséquent

$$\int_{\Omega} D^2(\omega) P(d\omega) \leq 4\|Z\|^2 \|X\|^2 \|X_n - X\|^2$$

d'où

$$\|(V_n - V) Z\|^2 \leq \|Z\|^2 (\|X\|^2 \|X_n - X\|^2)$$

Il en découle que V_n converge uniformément vers V et le résultat annoncé est établi.

Il résulte en particulier de la complète continuité de V que son spectre est purement ponctuel. Les valeurs propres sont positives ; celles qui sont différentes de 0 sont de multiplicité finie ; elles sont en nombre fini ou dénombrables. Dans ce dernier cas le nombre des valeurs propres supérieures à $\varepsilon > 0$ donné est fini. Les éléments propres associés à des valeurs propres différentes sont orthogonaux.

II. 2 - V EST UN OPERATEUR DE HILBERT-SCHMIDT ([6], p. 262).

Soit \mathcal{B}_1 l'ensemble des opérateurs complètement continus de $L_2(\Omega)$ en lui-même ; soit $S \in \mathcal{B}_1$.

Si $(\varphi_k)_{k \in K}$ est une famille orthonormale complète de $L_2(\Omega)$, on peut définir sur \mathcal{B}_1 une norme éventuellement infinie qui est appelée "norme de Hilbert-Schmidt" par :

$$\|S\|_2 = \left(\sum_{k \in K} \|S\varphi_k\|^2 \right)^{\frac{1}{2}}$$

On sait que cette norme est indépendante du choix de la famille $(\varphi_k)_{k \in K}$. Prenons alors, pour famille orthonormale complète une famille formée à partir des vecteurs propres de S . On suppose qu'une valeur propre $\lambda > 0$ est répétée r fois si sa multiplicité est r , d'où une suite $\lambda_1 \geq \lambda_2 \geq \dots$ à laquelle on peut associer un système orthonormal $\varphi_1, \varphi_2, \dots$ dans lequel $S\varphi_k = \lambda_k \varphi_k$, que l'on complète en lui adjoignant une base du noyau de S .

On obtient de façon immédiate :

$$\|S\|_2 = \left(\sum_{k=1}^{\infty} \lambda_k^2 \right)^{\frac{1}{2}} \quad (8)$$

Considérons l'ensemble \mathcal{B}_2 des opérateurs S tels que

$$\begin{cases} S \in \mathcal{B}_1 \\ \|S\|_2 < +\infty \end{cases}$$

Cet ensemble constitue la classe des opérateurs de Hilbert-Schmidt.

Proposition 3 :

Si $X \in L_2(\Omega \times T)$, l'opérateur V qui lui est associé est de la classe de Hilbert-Schmidt. Plus précisément si (λ_k) désigne la suite des valeurs propres positives -chacune étant répétée un nombre de fois égal à sa multiplicité - on a :

$$\sum_{k=1}^{\infty} \lambda_k^2 = \int_{T \times T} E^2(X(t) X(t')) \mu(dt) \mu(dt')$$

Démonstration :

Nous utiliserons un système orthogonal (Z_k) de vecteurs propres de V formant une base du sous espace orthogonal au noyau de V . On a donc $VZ_k = \lambda_k Z_k$ avec $\lambda_k > 0$ et on peut supposer comme plus haut la suite (λ_k) décroissante.

Pour toute fonction $Y \in L_2(\Omega)$ orthogonale au noyau de V , on a le développement de Fourier (convergent dans $L_2(\Omega)$) ([8], p. 69)

$$Y = \sum_{k=1}^{\infty} E(YZ_k) Z_k$$

Pour une fonction $Y \in L_2(\Omega)$ quelconque on peut écrire :

$$Y = Y' + \sum_{k=1}^{\infty} E(YZ_k) Z_k \quad (\text{avec } VY' = 0)$$

On en déduit pour VY le développement convergent dans $L^2(\Omega)$

$$VY = \sum_{k=1}^{\infty} \lambda_k E(YZ_k) Z_k$$

Il en résulte pour le produit scalaire $\langle Y, VY \rangle$ le développement convergent

$$\langle Y, VY \rangle = \sum_{k=1}^{\infty} \lambda_k E^2(YZ_k)$$

En particulier, si on prend $Y = X(t)$, on obtient pour tout t

$$\langle X(t), V(X(t)) \rangle = \sum_{k=1}^{\infty} \lambda_k E^2(X(t) Z_k) \quad (9)$$

Remarquons d'autre part, que par définition de V , si Z est un vecteur de $L_2(\Omega)$, de norme égale à 1, tel que $VZ = \lambda Z$, on a :

$$\lambda = \langle VZ, Z \rangle = \int_T E^2(X(t) Z) \mu(dt) \quad (10)$$

Pour démontrer la proposition 3, nous allons montrer que le deuxième membre de (9) peut être intégré terme à terme. Or :

$E(X(t) Z)$ appartient à $L_2(T)$ (démonstration de la proposition 2) .

Donc $\sum_{k=1}^n \lambda_k E^2(X(t) Z_k)$ appartient à $L_1(T)$ si bien qu'il suffit, à cause du théorème de Lebesgue, de démontrer que le premier membre de 9 est intégrable pour établir ce résultat. Ceci est fait en remarquant que :

$$\langle X(t), V(X(t)) \rangle \leq \|X(t)\| \|V(X(t))\| \leq \|X(t)\|^2 \|V\| \|X\|^2$$

D'où :

$$\int_{\mathbb{T}} \langle X(t), V(X(t)) \rangle \mu(dt) \leq \| \| X \| \|^2 \int_{\mathbb{T}} \| X(t) \|^2 \mu(dt) = \| \| X \| \|^4 \quad (11)$$

On a alors :

$$\begin{aligned} \int_{\mathbb{T} \times \mathbb{T}} E^2(X(t) X(t')) \mu(dt) \mu(dt') &= \int \langle X(t), V(X(t)) \rangle \mu(dt) \\ &= \sum_{k=1}^{\infty} \lambda_k \int_{\mathbb{T}} E^2(X(t) Z_k) \mu(dt) \\ &= \sum_{k=1}^{\infty} \lambda_k^2 \end{aligned} \quad (12)$$

III. 1 - RESULTATS PRELIMINAIRES

A partir de T , ensemble arbitraire, nous pouvons définir $\tilde{T} = T^{\mathbb{N}}$ l'ensemble des suites infinies d'éléments de T

$$s \in \tilde{T} \iff s = (s_0, s_1, \dots) \text{ où } s_i \in T$$

Si \mathcal{C} est une σ -algèbre de parties de T , on peut définir une σ -algèbre de parties de \tilde{T} : appelons en effet pavé de \tilde{T} un sous-ensemble de la forme :

$$\prod_{i \in \mathbb{N}} A_i = (s : s_i \in A_i \text{ (} i \in \mathbb{N} \text{)})$$

où A_i est contenu dans T pour tout i et A_i est égal à T pour tout i sauf un ensemble fini de valeurs de i .

Nous dirons que le pavé est mesurable si pour tout i , $A_i \in \mathcal{C}$. Les sommes finies de pavés mesurables disjoints forment une algèbre de Boole ([7], p. 75) qui engendre une σ -algèbre de parties de \tilde{T} que nous noterons $\tilde{\mathcal{C}}$.

Si une probabilité μ est définie sur T , il existe une probabilité unique $\tilde{\mu}$ sur \tilde{T} telle que, pour tout pavé mesurable de \tilde{T} , on ait :

$$\tilde{\mu} \left(\prod_{i \in \mathbb{N}} A_i \right) = \prod_{i \in E} \mu(A_i)$$

où E est l'ensemble fini des indices pour lesquels $A_i \neq T$. Enfin, une fonction $Z : \tilde{T} \rightarrow \mathbb{R}$ ne dépend que des coordonnées s_0, s_1, \dots, s_{n-1} est mesurable pour $\tilde{\mu}$ si et seulement si elle est mesurable comme fonction de $T^n \rightarrow \mathbb{R}$. La condition d'intégrabilité est analogue et l'on a :

$$\int_{\tilde{T}} Z(s) \tilde{\mu}(ds) = \int_{T^n} Z(s_0, \dots, s_{n-1}) \mu(ds_0) \dots \mu(ds_{n-1})$$

Définissons, de la même manière, \tilde{T}^* l'espace des suites infinies $(\dots, s_{-n}, \dots, s_{-1})$; $\tilde{\mathcal{C}}^*$ la σ -algèbre de parties de \tilde{T}^* définie comme $\tilde{\mathcal{C}}$ et $\tilde{\mu}^*$ la probabilité sur \tilde{T}^* , déduite de μ .

$(\tilde{T}^* \times \tilde{T}, \tilde{\mathcal{E}}^* \otimes \tilde{\mathcal{E}}, \nu)$ est l'espace probabilisé des suites doublement infinies d'éléments de T où ν est la probabilité unique définie sur $\tilde{T}^* \times \tilde{T}$ telle que : $\nu(A^* \times A) = \tilde{\mu}(A^*) \mu(A)$ pour tout couple (A^*, A) tel que $A^* \in \tilde{\mathcal{E}}^*$ et $A \in \tilde{\mathcal{E}}$.

Soit $\sigma = (\dots s_{-1}, s_0, s_1, \dots)$ un élément de $\tilde{T}^* \times \tilde{T}$.
 Nous conviendrons de noter $(\sigma)_n = s_n$.

Considérons la transformation θ de $\tilde{T}^* \times \tilde{T}$ en lui-même définie par : $(\theta\sigma)_k = s_{k+1}$ pour tout $k \in \mathbb{Z}$

Proposition 4 :

θ est une application mesurable qui préserve la mesure.

Démonstration :

L'appartenance de σ à $\theta^{-1}(A^* \times A)$ se traduit par celle de $\theta\sigma$ à $A^* \times A$; donc si $A^* \times A$ est mesurable, $\theta^{-1}(A^* \times A)$ est mesurable et θ est une application mesurable. D'autre part, pour tout pavé mesurable $A^* \times A$, $\nu(\theta^{-1}(A^* \times A))$ est égal à $\nu(A^* \times A)$; un résultat classique ([7], p. 27) dit que cette propriété se conserve pour tout $B \in \tilde{\mathcal{E}}^* \times \tilde{\mathcal{E}}$, donc θ préserve la mesure.

Rappelons les définitions suivantes ([2], p. 8 et 12) pour une transformation ρ de $\tilde{T}^* \times \tilde{T}$ en lui-même qui préserve la mesure.

Définition 1 :

Si tout ensemble invariant pour ρ (i.e. $\rho^{-1}B = B$) est de mesure 0 ou 1, ρ est ergodique.

Définition 2 :

Si pour tout couple (B_1, B_2) avec $B_1 \in \tilde{\mathcal{E}}^* \times \tilde{\mathcal{E}}$ et $B_2 \in \tilde{\mathcal{E}}^* \times \tilde{\mathcal{E}}$, $\lim_{n \rightarrow \infty} \nu(B_1 \cap \rho^{-n} B_2) = \nu(B_1) \nu(B_2)$, ρ est mélangeante.

En prenant $B_1 = B_2$ et B_1 invariant pour ρ , on établit que $\nu(B_1)$ ne peut être égal qu'à 0 ou à 1 : donc une transformation mélangeante est ergodique.

Proposition 5 :

θ est une transformation ergodique.

Démonstration :

Elle consistera à montrer que θ est mélangeante.

Soit B_1 (respectivement B_2) un pavé mesurable $A^* \times A$ de $\tilde{T}^* \times \tilde{T}$; il est de la forme $\prod_{i \in \mathbb{Z}} A_i$ avec $A_i \in \tilde{\mathcal{C}}$ pour $i \geq 0$, $A_i \in \tilde{\mathcal{C}}^*$ pour $i < 0$, $A_i = T$ sauf pour un nombre fini d'indices que nous noterons E_{B_1} (respectivement E_{B_2}).

Il existe N tel que $k > N$ entraîne $E_{B_1} \cap E_{\theta^{-k}B_2} = \emptyset$ si bien que pour tout $k > N$: $\nu(B_1 \cap \theta^{-k}B_2) = \nu(B_1) \nu(\theta^{-k}B_2)$. Mais θ préserve la mesure donc $\nu(B_1) \nu(\theta^{-k}B_2) = \nu(B_1) \nu(B_2)$.

La propriété vraie sur les pavés mesurables est vraie sur la σ -algèbre qu'ils engendrent si bien que θ est mélangeante, donc ergodique.

Le théorème ergotique ([2], p. 13) établit que si f appartient à $L_1(\tilde{T}^* \times \tilde{T})$:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(\theta^i \sigma) = \int_{\tilde{T}^* \times \tilde{T}} f(\sigma) \nu(d\sigma) \text{ presque partout sur } \tilde{T}^* \times \tilde{T}$$

En particulier, si f ne dépend que de s_0 ($f \in L_1(T)$), on aura :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(s_i) = \int_T f(s_0) \mu(ds_0) \text{ presque partout sur } \tilde{T}^* \times \tilde{T}$$

REMARQUE 1 :

Le résultat vaut aussi pour presque toute suite de T . En effet, s'il était inexact sur $\Gamma \subseteq \tilde{T}$ avec $\tilde{\mu}(\Gamma) > 0$, il serait

inexact sur $\tilde{T}^* \times \Gamma$ qui est de mesure égale à $\tilde{\mu}(\Gamma) > 0$. Dans la suite, nous considérons le résultat dans \tilde{T} .

REMARQUE 2 :

Une réunion dénombrable d'ensembles de mesure nulle est encore de mesure nulle.

Donc, si une famille dénombrable de conditions est telle que chacune de ces conditions soit réalisée presque partout sur \tilde{T} , alors presque partout sur \tilde{T} , toutes les conditions sont réalisées.

III. 2 - PREMIER THEOREME DE CONVERGENCE

Pour toute suite $s \in \tilde{T}$, définissons l'opérateur $U_n(s)$ sur $L_2(\Omega)$ par :

$$U_n(s) Y = \frac{1}{n} \sum_{i=0}^{n-1} E(X(s_i) Y) X(s_i)$$

Il a été convenu en I.1 que $X(s_i)$ appartenait à $L_2(\Omega)$. Il en résulte que $U_n(s)$ est un opérateur de $L_2(\Omega)$ en lui-même ; il est clair qu'il est linéaire.

Nous allons étudier la convergence de $U_n(s)$ vers l'opérateur V associé au processus $X \in L_2(\Omega \times T)$ par :

$$(VY) (\omega) = \int_T E(X(t) Y) X(\omega, t) \mu(dt)$$

Théorème 1 :

Soit $X \in L_2(\Omega \times T)$ et V l'opérateur qui lui est associé ; étant donné Y et Z appartenant à $L_2(\Omega)$, pour presque toute suite $s = (s_0, s_1, \dots) \in \tilde{T}$, $\langle U_n(s) Z, Y \rangle$ tend vers $\langle VZ, Y \rangle$ lorsque n tend vers l'infini.

On a établi dans (4) que :

$$\langle VZ, Y \rangle = \int_T E(X(t) Z) E(X(t) Y) \mu(dt)$$

D'autre part :

$$\langle U_n(s) Z, Y \rangle = \frac{1}{n} \sum_{i=0}^{n-1} E(X(s_i) Z) E(X(s_i) Y)$$

Le théorème 1 est donc une conséquence immédiate du théorème ergodique pour la fonction de $L_1(T)$; $E(X(t) Y) E(X(t) Z) \leq \|Y\| \|Z\| \|X(t)\|^2$

Corollaire 1 :

Si $L_2(\Omega)$ est de dimension finie, ou plus généralement si V et tous les $U_n(s)$ s'annulent sur un même sous-espace H de codimension finie dans $L_2(\Omega)$, alors pour presque tout $s \in \tilde{T}$, la suite $U_n(s)$

converge uniformément vers V .

Démonstration :

Soit $\varphi_1, \varphi_2, \dots, \varphi_n$ une base orthonormale de H^1 . Si $Y \in H^1$, on a le développement convergent au sens de $L_2(\Omega)$

$$Y = \sum_{i=1}^n E(\varphi_i Y) \varphi_i$$

Si Y est quelconque, $Y = Y' + \sum_{i=1}^n E(\varphi_i Y) \varphi_i$ ($VY' = U_n(s)Y' = 0$)

On a alors :

$$\begin{aligned} \langle VZ, Y \rangle &= \sum_{i=1}^n E(\varphi_i Z) \langle V\varphi_i, Y \rangle \\ &= \sum_{i=1}^n E(\varphi_i Z) \langle \varphi_i, VY \rangle \\ &= \sum_{i=1}^n \sum_{j=1}^n E(\varphi_i Z) E(\varphi_j Y) \langle \varphi_i, \varphi_j \rangle \end{aligned}$$

et de la même manière

$$\langle U_n(s) Z, Y \rangle = \sum_{i=1}^n \sum_{j=1}^n E(\varphi_i Z) E(\varphi_j Y) \langle \varphi_i, U_n(s)\varphi_j \rangle$$

Or : $\langle \varphi_i, V\varphi_j \rangle = \int_T E(X(t) \varphi_i) E(X(t) \varphi_j) \mu(dt)$

et $\langle \varphi_i, U_n(s)\varphi_j \rangle = \frac{1}{n} \sum_{k=0}^{n-1} E(X(t_k) \varphi_i) E(X(t_k) \varphi_j)$

Le théorème ergodique peut être appliqué à la fonction f_{ij} définie par $f_{ij}(\sigma) = E(X(s_0) \varphi_i) E(X(s_0) \varphi_j)$. Il résulte alors de la remarque 2 que pour presque tout $s \in T$

$$\int_T f_{ij}(t) \mu(dt) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\ell=0}^{n-1} f_{ij}(s_\ell) \quad \text{pour tout couple } (i, j)$$

Compte tenu de la linéarité des opérateurs ceci établit la convergence faible et donc la convergence uniforme puisqu'on est en dimension finie.

Corollaire 2 :

Si X est un processus de la forme $X(\omega, t) = \sum_{i=1}^k A_i(\omega) B_i(t)$, presque tout $s \in \tilde{T}$ la suite $U_n(s)$ converge uniformément vers V .

Démonstration :

On a établi dans (6) que :

$$V Z = \sum_{j=1}^n \left[\sum_{i=1}^n E(A_i | Z) \int_{\mathbb{T}} B_i(t) B_j(t) \mu(dt) \right] A_j$$

Un calcul simple établit que :

$$U_n(s) Z = \sum_{j=1}^n \left[\sum_{i=1}^n E(A_i | Z) \frac{1}{n} \sum_{k=1}^n B_i(s_k) B_j(s_k) \right] A_j$$

Il en résulte que V et $U_n(s)$ s'annulent sur le sous espace orthogonal à A_1, \dots, A_n , ce qui ramène au corollaire 1 .

III. 3 - THEOREME DE CONVERGENCE UNIFORME.

Dans le prochain paragraphe nous aurons à étudier le comportement du spectre de $U_n(s)$ lorsque $n \rightarrow \infty$. Cette étude n'est guère praticable que dans le cas de convergence uniforme. Une telle convergence semble très difficile dans le cas général ; il y a donc lieu de chercher des conditions suffisantes pour qu'elle soit réalisée. Le théorème suivant s'inscrit dans cette direction.

Théorème 2 :

Soit $X \in L_2(\Omega \times T)$, un processus tel que :

$$\|X(t)\| < M \quad \forall t \in T$$

S'il existe une suite (X_k) de processus de $L_2(\Omega \times T)$ de la forme

$$X_k(\omega, t) = \sum_{i=1}^k A_i(\omega) B_i(t) \quad \text{telle que :}$$

$$\lim_{k \rightarrow \infty} \left(\sup_{t \in T} \|X(t) - X_k(t)\| \right) = 0$$

alors pour presque toute suite $s \in \tilde{T}$, on a $\lim_{n \rightarrow \infty} \|U_n(s) - V\| = 0$

Tous les processus qui sont tels que $\|X(t)\| < M$ et qui sont limite uniforme de processus de rang fini rentrent dans le cadre du théorème puisque la convergence uniforme implique la condition imposée.

En particulier si Ω et T sont des intervalles de R , $[0, 1]$ par exemple, on atteint la classe des fonctions continues sur $[0, 1] \times [0, 1]$ puisque les fonctions de cette classe peuvent être approchées de façon uniforme par des fonctions en escalier du type $\sum_{i=1}^k \varphi_i(x) \psi_i(y)$ où les φ_i et les ψ_i sont des fonctions en escalier sur $[0, 1]$.

Démonstration du théorème :

Désignons par V_k l'opérateur associé à X_k et définissons

pour tout $s \in \tilde{T}$: $U_{kn}(s) Y = \frac{1}{n} \sum_{i=0}^{n-1} E(X_k(s_i) Y) X_k(s_i)$

On a :

$$\|U_n(s) - V\| \leq \|V - V_k\| + \|V_k - U_{kn}(s)\| + \|U_{kn}(s) - U_n(s)\|$$

a) majoration de $\|V - V_k\|$

On a établi dans (7), que pour k assez grand ($k > K$)

$$\|V - V_k\| \leq 3 \|X\| \|X_k - X\|$$

b) majoration de $\|V_k - U_{kn}(s)\|$

Le corollaire 2 du théorème 1 a montré que pour k fixé, pour presque toute suite $s \in \tilde{T}$

$$\lim_{n \rightarrow \infty} \|V_k - U_{kn}(s)\| = 0$$

La remarque 2 permet alors d'affirmer que pour presque toute suite $s \in \tilde{T}$

$$\lim_{n \rightarrow \infty} \|V_k - U_{kn}(s)\| = 0 \quad (k = 1, 2, \dots)$$

Nous noterons \tilde{T}_1 l'ensemble des suites s telles que cette condition soit vérifiée.

c) majoration de $\|U_{kn}(s) - U_n(s)\|$. Nous allons majorer ce terme indépendamment de n .

$$U_n(s) Y = \frac{1}{n} \sum_{i=0}^{n-1} E(X(s_i) Y) X(s_i)$$

$$U_{kn}(s) Y = \frac{1}{n} \sum_{i=0}^{n-1} E(X_k(s_i) Y) X_k(s_i)$$

D'où :

$$\begin{aligned} \|(U_n(s) - U_{kn}(s)) Y\| &\leq \frac{1}{n} \sum_{i=0}^{n-1} \|E(X(s_i) Y) X(s_i) - E(X_k(s_i) Y) X_k(s_i)\| \\ &\leq \frac{1}{n} \sum_{i=0}^{n-1} [\|E((X(s_i) - X_k(s_i)) Y) X(s_i)\| + \\ &\quad + \|E(X_k(s_i) Y) (X(s_i) - X_k(s_i))\|] \end{aligned}$$

$$\begin{aligned} \text{Or : } \|E((X(s_i) - X_k(s_i)) Y) X(s_i)\| &\leq \|Y\| \|X(s_i)\| \|X(s_i) - X_k(s_i)\| \\ &\leq M \|Y\| \sup_{t \in T} \|X(t) - X_k(t)\| \end{aligned}$$

$$\begin{aligned} \text{et } \|E(X_k(s_1)Y)(X(s_1)-X_k(s_1))\| &\leq \|Y\| \|X_k(s_1)\| \|X(s_1)-X_k(s_1)\| \\ &\leq \|X_k(s_1)\| \|Y\| \sup_{t \in T} \|X(t)-X_k(t)\| \end{aligned}$$

Nous pouvons noter que les hypothèses faites impliquent que pour k suffisamment grand ($k > K_0$) $\|X_k(s_1)\| < 2M$
D'où :

$$\|U_n(s)-U_{kn}(s)\| \leq 3M \sup_{t \in T} \|X(t)-X_k(t)\|$$

D'où finalement en rassemblant les trois majorations ; pour $k > \max(K, K_0)$

$$\|V-U_n(s)\| \leq 3\|X\| \|X_k-X\| + 3M \sup_{t \in T} \|X(t)-X_k(t)\| + \|V_k-V_{kn}(s)\|$$

Etant donné $\epsilon > 0$, on peut trouver K_1 (éventuellement égal à K_0 ou à K) tel que pour $k > K_1$ la somme des deux premiers termes soit inférieure à $\frac{\epsilon}{2}$.

Si $s \in \tilde{T}_1$ est fixé, on peut trouver $N(\epsilon)$ tel que $n > N(\epsilon)$ implique $\|V_k-U_{kn}(s)\| < \frac{\epsilon}{2}$ ($\forall k = 1, 2, \dots$)

On en déduit que $n > N(\epsilon)$ implique $\|V-U_n(s)\| < \epsilon$ ce qui achève la démonstration du théorème .

III. 4 - CONSEQUENCE DES THEOREMES ([8], p. 367)

Soit A un opérateur autoadjoint. Nous désignons par E_λ la famille spectrale de A ($A = \int \lambda d E_\lambda$) et par $E(\delta)$ le projecteur associé à l'intervalle δ dont les extrémités sont supposées ne pas appartenir au spectre de A . Si A_n est une suite d'opérateurs autoadjoints, nous notons $E_{n\lambda}$ la famille spectrale de A_n et $E_n(\delta)$ le projecteur associé à δ relativement à A_n .

Lorsqu'une suite de transformations autoadjointes bornées (A_n) , tend uniformément vers la transformation autoadjointe A :

$$E_n(\delta) \text{ tend uniformément vers } E(\delta)$$

$$\text{et } A_n E_n(\delta) \text{ tend uniformément vers } A E(\delta)$$

pour tout intervalle δ dont les extrémités n'appartiennent pas au spectre de V .

En particulier si l'intervalle δ ne contient qu'un seul point λ du spectre de A qui est une valeur propre de multiplicité finie m , pour n assez grand, la partie du spectre de A_n contenue dans l'intervalle δ est constituée de valeurs propres de multiplicité totale m . Comme on aurait pu choisir pour δ un intervalle arbitrairement petit autour de λ , les valeurs propres de A_n contenues dans δ convergent vers λ pour n tendant vers l'infini.

On sait également que dès que $\|E_n(\delta) - E(\delta)\| < 1$, le sous-espace propre associé à λ peut être appliqué d'une manière linéaire et isométrique sur le sous espace engendré par les vecteurs propres associés aux valeurs propres de A_n qui sont dans δ . En particulier si λ est de multiplicité 1, le vecteur propre normé associé à λ est limite d'un vecteur propre normé de A_n associé à la valeur propre λ_n qui tend vers λ .

On peut appliquer aux suites d'opérateurs $(U_n(s))$ les résul-

tats que nous venons de rappeler lorsque ces suites convergent uniformément vers V , ce qui a lieu presque sûrement sur T sous les hypothèses :

$$\left\{ \begin{array}{l} X \in L_2(\Omega \times T) \quad \text{tel que} \quad \|X(t)\| < M \quad \forall t \in T \\ \exists (X_k)_{k \in \mathbb{N}} ; X_k \in L_2(\Omega \times T) ; X_k(\omega, t) = \sum_{i=1}^k A_i(\omega) B_i(t) \end{array} \right.$$

$$\text{telle que} \quad \lim_{k \rightarrow \infty} \left(\sup_{t \in T} \|X(t) - X_k(t)\| \right) = 0$$

IV - LE PROBLEME DE HOTELLING

Lorsque le statisticien mène à bien une analyse en composantes principales, son but est d'étudier l'ensemble particulier des variables aléatoires qui sont à sa disposition. Les difficultés qu'il rencontre sont liées au fait que les corrélations sur lesquelles il travaille ont été estimées à partir d'un échantillon réduit d'individus et que, par conséquent, elles diffèrent des véritables valeurs que fournirait la population toute entière. Son analyse se justifie parce que les valeurs propres et les vecteurs propres de l'estimateur au sens du maximum de vraisemblance de la matrice des variances et covariances sont les estimateurs au sens du maximum de vraisemblance des valeurs propres et des vecteurs propres de la matrice réelle ([1], p. 279).

Le problème soulevé par Hotelling relève d'un objectif différent. Considérons l'ensemble particulier des variables observées comme un échantillon tiré dans une population de variables aléatoires. Les difficultés que nous allons rencontrer proviennent de ce que les valeurs propres et les vecteurs propres que nous trouvons varient d'un échantillon de variables à l'autre. En particulier, la disposition relative des projections des individus sur les espaces engendrés par les vecteurs propres associés aux plus fortes valeurs propres varie lorsqu'on passe d'un groupe de variables à un autre, si bien que deux individus qui apparaîtront voisins pour une analyse pourraient sembler plus éloignés dans une autre.

Les résultats précédents, associés à la proposition suivante, permettent d'assurer que, sous les hypothèses de nos théorèmes, sauf pour des groupes de variables qui ne se rencontrent qu'avec une probabilité nulle, nous obtiendrons une représentation fidèle lorsque l'effectif du groupe de variables prises en compte est assez grand.

Soit s un élément de \mathbb{T} . Appelons $\mathcal{A}_n(s)$ l'opérateur défini par la matrice des variances et covariances des $X(t_i)$, $i=0, \dots, n-1$ de la manière suivante :

* Si $Y \in [X(t_0), \dots, X(t_{n-1})]$ alors

$$Y = \sum_{i=0}^{n-1} y_i X(t_i)$$

$$\text{et } \mathcal{A}_n(s) Y = \sum_{j=0}^{n-1} E(X(t_i)X(t_j)) y_j$$

* Si $Y \in [X(t_0), \dots, X(t_{n-1})]^\perp$ alors

$$\mathcal{A}_n(s) Y = 0$$

Les composantes principales du système $(X(t_0), \dots, X(t_{n-1}))$ sont définies par les vecteurs propres de $\mathcal{A}_n(s)$; ce sont les Y qui vérifient la relation :

$$\sum_{j=0}^{n-1} E(X(t_i) X(t_j)) y_j = \lambda y_i \quad (14)$$

Proposition 6 :

L'opérateur $U_n(s)$ a les mêmes vecteurs propres que $\mathcal{A}_n(s)$; ses valeurs propres sont égales à celles de $\mathcal{A}_n(s)$ divisées par n .

Démonstration :

Supposons que $Y = \sum_{j=0}^{n-1} y_j X(t_j)$ soit un vecteur propre

$$\begin{aligned} U_n(s) Y &= \frac{1}{n} \sum_{i=0}^{n-1} E(X(t_i) Y) X(t_i) \\ &= \frac{1}{n} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} E(X(t_i) X(t_j)) X(t_i) y_j \\ &= \frac{\lambda}{n} \sum_{i=0}^{n-1} y_i X(t_i) = \frac{\lambda}{n} Y \end{aligned}$$

Appelons λ_{nk} les valeurs propres de $\mathcal{A}_n(s)$ comptées avec leur multiplicité. Nous pouvons énoncer :

Théorème 3 : (problème de Hotelling)

Soit $X \in L_2(\Omega \times T)$ un processus tel que $\|X(t)\| < M, \forall t \in T$.
Soit V l'opérateur associé à X .

S'il existe une suite (X_k) de processus de $L_2(\Omega \times T)$ de la forme $X_k(\omega, t) = \sum_{i=1}^k A_i(\omega) B_i(t)$ telle que :

$$\lim_{k \rightarrow \infty} \left(\sup_{t \in T} \|X(t) - X_k(t)\| \right) = 0$$

pour presque toute suite $s \in \tilde{T}$, quelque soit l'intervalle δ , dont les extrêmités n'appartiennent pas au spectre de V , qui ne contient qu'une valeur propre λ de V de multiplicité m , pour n assez grand :

Alors : a) la multiplicité totale des valeurs propres $\lambda_{nk}(s)$, telles que $\frac{\lambda_{nk}(s)}{n} \in \delta$ est m

b) Le sous espace propre associé à λ peut être appliqué d'une manière linéaire et isométrique sur le sous espace engendré par les vecteurs propres associés au $\lambda_{nk}(s)$ telles que $\frac{\lambda_{nk}(s)}{n} \in \delta$. Plus précisément $\lim_{n \rightarrow \infty} \|E_n(\delta) - E(\delta)\| = 0$



DEUXIEME PARTIE

V. 1 - POSITION DU PROBLEME

Pour étudier un processus aléatoire, le statisticien est conduit en général à la discrétiser; il substitue alors au processus $(X(t))_{t \in T}$ un ensemble fini de variables $(X(t_i))_{i \in I}$. Son intuition, la connaissance qu'il a du problème qu'il traite sont les seuls guides dont il dispose pour faire ce choix. En l'absence de toute information préalable, il doit tirer au hasard les variables $(X(t_i))_{i \in I}$ dans l'ensemble des variables possibles.

Nous avons vu dans la première partie que, sous certaines hypothèses, il est possible d'obtenir une "bonne" représentation du processus par un tirage au hasard à condition de prendre pour I un ensemble assez grand. La représentation sera "bonne" en ce sens que les vecteurs propres et les valeurs propres de la matrice des variances et covariances de l'ensemble fini des variables observées conduisent à de bonnes approximations des valeurs propres et des vecteurs propres de l'opérateur V que nous avons défini.

Dans la pratique ce résultat reste insuffisant car le plus souvent l'expérimentateur connaît à priori, à cause de leurs implications économiques par exemple, le nombre de variables aléatoires qu'il peut observer. Pour lui, la question primordiale est la suivante: sachant que n variables aléatoires peuvent être réellement observées, comment les choisir au mieux dans l'ensemble des variables possibles ?

Pour simplifier les notations dans la suite de ce travail, nous écrirons $(X_t)_{t \in T}$ pour $(X(t))_{t \in T}$

et $(X_i)_{i \in I}$ pour $(X(t_i))_{i \in I}$

Dans les applications que nous visons, un ensemble fini $(X_i)_{i=1, \dots, n}$ de variables aléatoires est plus intéressant qu'un ensemble

$(X_j)_{j=1, \dots, n}$ dans la mesure où les composantes principales des X_i conduisent à de meilleures approximations des vecteurs propres et des valeurs propres de l'opérateur V que les composantes principales des X'_j . La première partie conduit alors, naturellement, à substituer à l'ensemble $(X_i)_{i=1, \dots, n}$ l'opérateur U_n tel que :

$U_n Y = \frac{1}{n} \sum_{i=1}^n E(X_i Y) X_i$ et à l'ensemble $(X'_j)_{j=1, \dots, n}$ l'opérateur U'_n analogue.

Il faut trouver une distance qui mettent particulièrement l'accent sur les éléments propres des opérateurs. Ceci se fait très simplement puisque les hypothèses faites placent les opérateurs qui nous intéressent dans la classe des opérateurs de Hilbert-Schmidt. Enfin, comme il est impossible que le nombre des variables réellement observées ne soit pas fini, l'ensemble $(X_t)_{t \in T}$ n'est connu qu'au travers d'un ensemble $(X_k)_{k=1, \dots, p}$ avec p suffisamment grand pour que l'opérateur U_p soit une excellente approximation de V .

Le problème consiste alors à choisir parmi les n -uples possibles de l'ensemble $(X_k)_{k=1, \dots, p}$ celui qui donne l'opérateur U_n le plus proche de U_p . Comme il n'est pas acceptable dans la pratique d'envisager tous les cas possibles, nous proposons une méthode de choix progressifs analogue dans son esprit aux méthodes progressives utilisées dans les techniques statistiques telles que la régression ou la discrimination.

V. 2 - DISTANCE ENTRE LES OPERATEURS ([6], p. 262)

Nous avons défini en II. 2, la classe \mathcal{B}_2 des opérateurs de Hilbert-Schmidt sur $L_2(\Omega)$. Soit $S_1 \in \mathcal{B}_2$, $S_2 \in \mathcal{B}_2$.

A partir d'une famille orthonormale $(\varphi_k)_{k \in K}$ complète dans $L_2(\Omega)$, on peut définir un produit scalaire (indépendant de la famille choisie) dans \mathcal{B}_2 par :

$$\langle S_1, S_2 \rangle_2 = \sum_{k \in K} E(S_1(\varphi_k) S_2(\varphi_k))$$

Muni de ce produit scalaire, \mathcal{B}_2 est un espace de Hilbert.

Prenons une famille orthonormale complète formée des vecteurs propres ξ_k de S_1 associés aux valeurs propres non nulles μ_k comptées avec leur multiplicité et d'une famille orthonormale complète du noyau de S_1 .

$$\text{Il est clair que : } \langle S_1, S_2 \rangle_2 = \sum_{k=1}^{\infty} E(S_1(\xi_k) S_2(\xi_k))$$

Soit ψ_ℓ les vecteurs propres de S_2 associés aux valeurs propres non nulles λ_ℓ . Pour tout ξ_k , on peut écrire le développement convergent dans $L_2(\Omega)$:

$$\xi_k = \sum_{\ell=1}^{\infty} E(\xi_k \psi_\ell) \psi_\ell + h$$

où h appartient au noyau de S_2 .

Il en découle :

$$\langle S_1, S_2 \rangle_2 = \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \mu_k \lambda_\ell [E(\xi_k \psi_\ell)]^2 \quad (15)$$

On en déduit alors :

$$\begin{aligned} d_2^2(S_1, S_2) &= \|S_1 - S_2\|_2^2 \\ &= \|S_1\|_2^2 + \|S_2\|_2^2 - 2 \langle S_1, S_2 \rangle_2 \\ &= \sum_{k=1}^{\infty} \mu_k^2 + \sum_{\ell=1}^{\infty} \lambda_\ell^2 - 2 \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \mu_k \lambda_\ell [E(\xi_k \psi_\ell)]^2 \quad (16) \end{aligned}$$

VI. 1 - EXPRESSION DE LA DISTANCE ENTRE LES OPERATEURS U_n et U_m .

Soit $(X_i)_{i=1, \dots, n}$ et $(X'_j)_{j=1, \dots, m}$ deux ensembles de variables aléatoires réelles auxquelles sont associés respectivement les ensembles $(Y_i)_{i=1, \dots, n}$ et $(Y'_j)_{j=1, \dots, m}$ de composantes principales.

On sait que :

$$X_i = \sum_{k=1}^n a_{ik} Y_k \quad \text{et} \quad E(Y_k)^2 = \lambda_k$$

$$X'_j = \sum_{\ell=1}^m b_{j\ell} Y'_\ell \quad \text{et} \quad E(Y'_\ell)^2 = \mu_\ell$$

avec $\sum_{i=1}^n a_{i\alpha} a_{i\beta} = \delta_{\alpha\beta}$

et $\sum_{j=1}^m b_{j\alpha} b_{j\beta} = \delta_{\alpha\beta}$

Posons $Z_i = \frac{Y_i}{\sqrt{\lambda_i}}$ et $Z'_j = \frac{Y'_j}{\sqrt{\mu_j}}$

Proposition 7 :

Avec les définitions précédentes :

$$\sum_{i=1}^n \sum_{j=1}^m [E(X_i X'_j)]^2 = \sum_{i=1}^n \sum_{j=1}^m \mu_j \lambda_i [E(Z_i Z'_j)]^2$$

On a en effet :

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^m [E(X_i X'_j)]^2 &= \sum_{i=1}^n \sum_{j=1}^m [E(\sum_{k=1}^n a_{ik} Y_k \times \sum_{\ell=1}^m b_{j\ell} Y'_\ell)]^2 \\ &= \sum_{i=1}^n \sum_{j=1}^m [\sum_{\alpha=1}^n \sum_{\beta=1}^n \sum_{\gamma=1}^m \sum_{\delta=1}^m a_{i\alpha} a_{i\beta} b_{j\gamma} b_{j\delta} E(Y_\alpha Y'_\gamma) E(Y_\beta Y'_\delta)] \\ &= \sum_{j=1}^m [\sum_{\alpha=1}^n \sum_{\beta=1}^n \sum_{\delta=1}^m b_{j\gamma} b_{j\delta} E(Y_\alpha Y'_\gamma) E(Y_\alpha Y'_\delta)] \\ &= \sum_{\alpha=1}^n \sum_{\gamma=1}^m [E(Y_\alpha Y'_\gamma)]^2 = \sum_{k=1}^n \sum_{\ell=1}^m \lambda_k \mu_\ell [E(Z_k Z'_\ell)]^2 \end{aligned}$$

On en déduit en particulier

$$\sum_{k=1}^n \sum_{\ell=1}^m [E(X_k X'_\ell)]^2 = \sum_{k=1}^n \lambda_k^2 \quad (17)$$

D'où la proposition :

Proposition 8 :

Avec les définitions précédentes

$$\|U_n - U_m\|_2^2 = \frac{\sum_{k=1}^n \sum_{\ell=1}^n [E(X_k X_\ell)]^2}{n^2} + \frac{\sum_{k=1}^m \sum_{\ell=1}^m [E(X'_k X'_\ell)]^2}{m^2} - 2 \frac{\sum_{k=1}^n \sum_{\ell=1}^m [E(X_k X'_\ell)]^2}{nm}$$

On a vu en effet que si λ_k est la valeur propre de la matrice des variances et covariances des $(X_i)_{i=1, \dots, n}$ associée au vecteur propre Z_k , l'opérateur U_n a pour valeur propre $\frac{\lambda_k}{n}$ associée au même vecteur propre.

D'autre part :

$$\|U_n - U_m\|_2^2 = \|U_n\|_2^2 + \|U_m\|_2^2 - 2 \langle U_n, U_m \rangle_2$$

et d'après (16) ceci est égal à :

$$\sum_{k=1}^n \frac{\lambda_k^2}{n^2} + \sum_{\ell=1}^m \frac{\mu_\ell^2}{m^2} - 2 \sum_{k=1}^n \sum_{\ell=1}^m \frac{\lambda_k \mu_\ell [E(Z_k Z'_\ell)]^2}{nm}$$

La proposition 7 permet alors de conclure pour le résultat annoncé.

L'intérêt de cette proposition réside dans ce qu'elle nous permet d'apprécier la distance entre les opérateurs U_n et U_m sans qu'il soit nécessaire de connaître le spectre de ces opérateurs.

VI. 2 - LE COEFFICIENT Q

On remarque que si on multiplie par k les variables X_i et par k' les variables X'_j , la distance entre les opérateurs U_n et U_m est multipliée par kk' . Pour échapper à cet inconvénient, il est assez naturel de normer les variables X_i et X'_j de telle sorte que les distances entre couples d'opérateurs soient comparables entre elles.

Substituons aux X_i et X'_j les variables

$$\tilde{X}_i = \frac{X_i}{\sqrt{\|U_n\|_2}} \quad i = 1, \dots, n$$

$$\tilde{X}'_j = \frac{X'_j}{\sqrt{\|U_m\|_2}} \quad j = 1, \dots, n$$

Si \tilde{U}_n et \tilde{U}_m sont les opérateurs construits avec les \tilde{X}_i et \tilde{X}'_j ;

il vient :

$$\|\tilde{U}_n - \tilde{U}_m\|_2^2 = 2 \left(1 - \frac{\langle U_n, U_m \rangle_2}{\|U_n\|_2 \|U_m\|_2} \right)$$

Posons $Q(U_n, U_m) = \frac{\langle U_n, U_m \rangle_2}{\|U_n\|_2 \|U_m\|_2}$

On a les implications

$$\{\|U_n - U_m\|_2^2 = 0\} \implies \{\|\tilde{U}_n - \tilde{U}_m\|_2^2 = 0\} \implies \{Q = 1\}$$

Inversement

$$\{Q = 1\} \implies \{\|\tilde{U}_n - \tilde{U}_m\|_2^2 = 0\}$$

Cette dernière proposition implique que U_n et U_m ont les mêmes espaces propres, le même noyau et des valeurs propres non nulles proportionnelles. En effet, en divisant toutes les variables X_i par $\sqrt{\|U_n\|_2}$, on divise les éléments de la matrice des variances et covariances de ces variables par $\|U_n\|_2$, donc les valeurs propres par $\|U_n\|_2$.

Les valeurs propres de \tilde{U}_n sont donc $\frac{\lambda_k}{n \|U_n\|_2}$

La proposition $\|\tilde{U}_n - \tilde{U}_m\|_2^2 = 0$ implique que pour les valeurs propres non nulles on a :

$$\frac{\lambda_k}{n\|U_n\|_2} = \frac{\mu_k}{m\|U_m\|_2}$$

Soit $\frac{\lambda_k}{\mu_k} = \frac{n}{m} \times \frac{\|U_n\|_2}{\|U_m\|_2}$

A l'opposé, Q ne peut être nul que si toutes les variables X_i sont orthogonales à toutes les variables X'_j . Ainsi la valeur de Q dans l'intervalle $[0, 1]$ donne rapidement une idée de la proximité des deux opérateurs, donc des deux ensembles de variables.

VI. 3 - RECHERCHE PROGRESSIVE DES VARIABLES

La proposition 8 a établi que :

$$d_2^2(U_n, U_m) = \|U_n - U_m\|_2^2 = \frac{\sum_{k=1}^n \sum_{\ell=1}^n [E(X_k X_\ell)]^2}{n^2} + \frac{\sum_{k=1}^m \sum_{\ell=1}^m [E(X'_k X'_\ell)]^2}{m^2} - 2 \frac{\sum_{k=1}^n \sum_{\ell=1}^m [E(X_k X'_\ell)]^2}{nm}$$

On peut écrire :

$$\begin{aligned} d_2^2(U_n, U_m) &= A(n, n) + A(m, m) - 2 A(n, m) \\ &= \frac{[(n-1)^2}{n^2} A(n-1, n-1) - 2 \frac{(n-1)}{n} A(n-1, m) + A(m, m)] + \\ &+ \left[\frac{\sum_{k=1}^n [E(X_n X_k)]^2}{n^2} - 2 \frac{\sum_{\ell=1}^m [E(X_n X'_\ell)]^2}{nm} \right] \end{aligned}$$

Le premier crochet est indépendant de la n^{ième} variable que l'on prend en compte. Chercher la variable la plus intéressante au rang n revient donc à chercher celle des m - (n-1) variables non encore introduites qui minimise le second crochet que nous appellerons C(n, X_n).

Cette variable étant trouvée, on a :

$$\begin{aligned} A(n, n) &= \frac{(n-1)^2}{n^2} A(n-1, n-1) + \frac{\sum_{k=1}^n [E(X_n X_k)]^2}{n^2} \\ A(n, m) &= \frac{n-1}{n} A(n-1, m) + \frac{\sum_{\ell=1}^m [E(X_n X'_\ell)]^2}{nm} \end{aligned}$$

Si l'on veut travailler avec le coefficient Q, on obtient

$$\begin{aligned} Q(U_n, U_m) &= \frac{\sum_{k=1}^n \sum_{\ell=1}^m [E(X_k X'_\ell)]^2}{\sqrt{\sum_{k=1}^n \sum_{\ell=1}^n [E(X_k X_\ell)]^2} \sqrt{\sum_{k=1}^m \sum_{\ell=1}^m [E(X'_k X'_\ell)]^2}} \\ &= \frac{B(n, m)}{\sqrt{B(n, n)} \sqrt{B(m, m)}} \\ &= \frac{B(n-1, m) + \sum_{\ell=1}^m [E(X_n X'_\ell)]^2}{\sqrt{B(n-1, n-1) + \sum_{k=1}^n [E(X_n X_k)]^2}} \times \frac{1}{\sqrt{B(m, m)}} \end{aligned}$$

Il s'agit donc ici d'introduire la variable qui rend maximum le premier terme du produit précédent.

Bien que les deux méthodes relèvent du même principe - prise en compte de la variable qui a les plus petites covariances avec les précédentes tout en ayant de fortes covariances avec l'ensemble des variables possibles - les résultats obtenus ne seront pas en général identiques.

On peut remarquer en particulier que la première méthode privilégie systématiquement une variable X par rapport à une variable $X' = kX$ si $k > 1$ puisque

$$C(n, X') = k^2 C(n, X)$$

Au contraire, dans la seconde méthode, le choix dépend des variables déjà introduites et même, pour le choix de la première variable, c'est l'organisation du programme de calcul qui est primordiale puisque :

$$\text{si } U_1(Y) = E(XY) X$$

$$U'_1(Y) = E(X'Y) X'$$

$$\text{alors } Q(U_1, U_m) = Q(U'_1, U_m)$$



VII. 1 - PREMIER EXEMPLE

Revenons au problème évoqué dans l'introduction. La connaissance qu'ils avaient du phénomène qu'ils étudiaient avait conduit les chercheurs du laboratoire d'Hydrométéorologie à retenir 27 stations d'observations réparties dans la région de manière à ce que toutes les causes de fluctuations des chutes de pluies -relief, vent dominant...- soient prises en compte. (figure 1).

Les méthodes proposées en VI permettraient d'étudier si ces 27 stations forment bien le meilleur groupe de 27 que l'on puisse prendre. Le but de l'étude que nous présentons ici est différent.

Supposons que pour des raisons économiques évidentes -coût des relevés périodiques, de l'entretien...- il soit nécessaire de réduire le nombre des stations d'observations à une dizaine environ : comment les choisir ?

Sur la base des résultats précédents, nous pouvons chercher les dix variables qui seront les plus proches des 27 connues

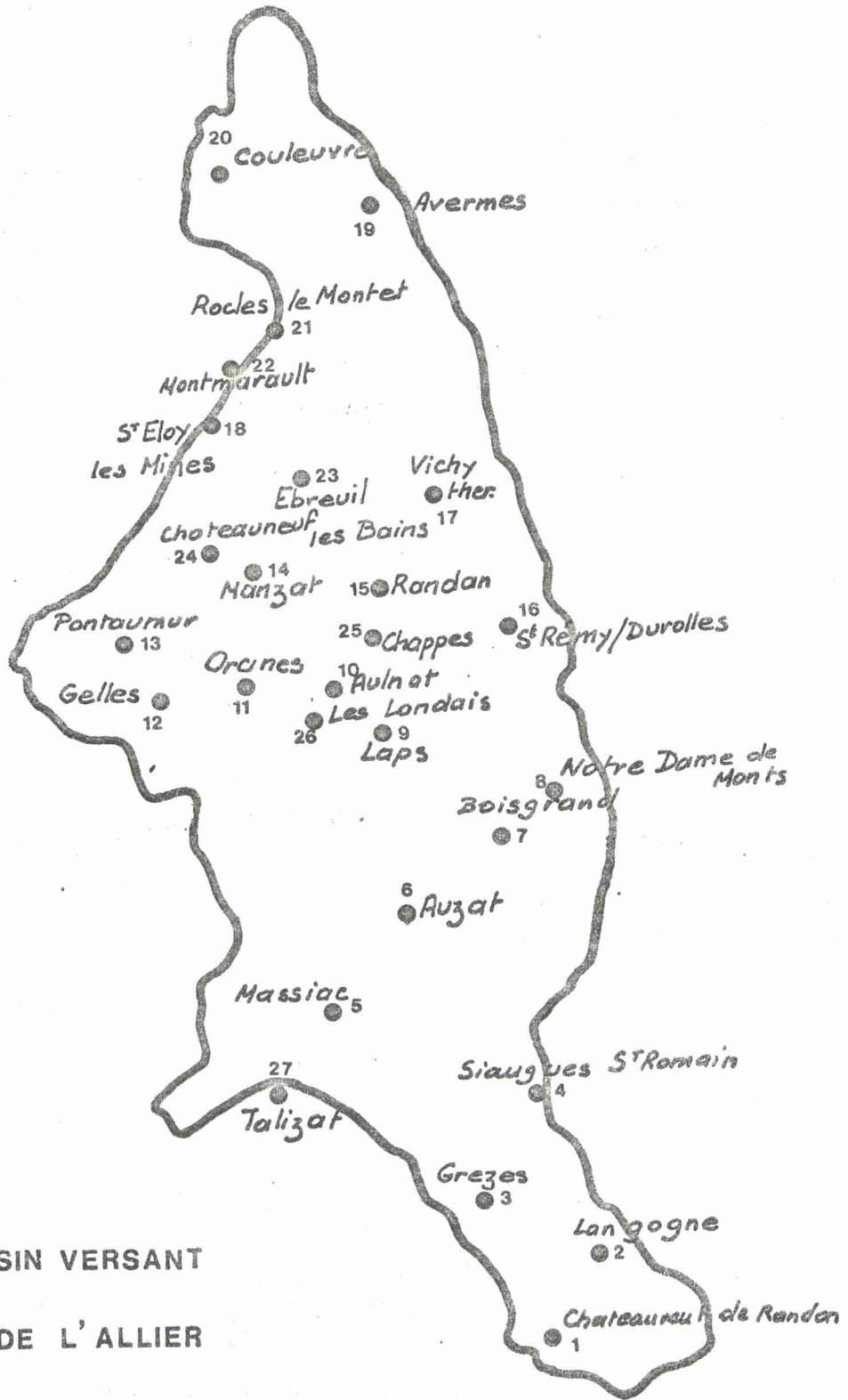
- soit au sens où elles minimisent $\|U_{10} - U_{27}\|_2^2$

- soit au sens où elles maximisent $Q(U_{10}, U_{27})$

Dans les deux cas, la valeur de la distance ou celle du coefficient Q nous permettront d'apprécier si l'approximation trouvée est suffisante.

Dans les deux méthodes, les calculs sont faits pas à pas, de telle sorte que la variable introduite au rang n est la meilleure possible en fonction des $n-1$ précédentes.

Tout calcul fait, les deux méthodes conduisent à des ensembles qui ne diffèrent que de deux variables.



BASSIN VERSANT
DE L'ALLIER

Ech. : 1 / 250000

* En travaillant avec $\|U_{10} - U_{27}\|_2$, on obtient les variables

3 ; 8 ; 10 ; 13 ; 14 ; 17 ; 19 ; 24 ; 9 ; 21

La distance $\|U_{10} - U_{27}\|_2$ est alors égale à 145 (On remarque au passage combien il est difficile d'apprécier sa signification) ; la quantité Q est égale à 0,975.

* L'utilisation du coefficient Q , conduit aux variables

3 ; 8 ; 10 ; 13 ; 14 ; 17 ; 19 ; 24 ; 4 ; 26

Q est alors égal à 0,979.

Si l'on cherche à introduire une onzième variable, on constate que la seconde méthode conduit à prendre la variable 21 ($Q = 0,981$) alors que la première introduit la variable 1 ($Q = 0,982$).

Les deux ensembles ont alors neuf variables en commun. L'étude des variables différentes est intéressante dans la mesure où les variables 26 et 9 d'une part, 4 et 1 d'autre part sont géographiquement très voisines et ont donc des comportements voisins.

On constate, par ailleurs, que la variable 9 à une variance supérieure à celle de la variable 26 ($607 > 462$) et que la variable 1 à une variance supérieur à celle de la variable 4 ($1580 > 817$). La prédilection de la première méthode pour les variables de forte variance s'est donc manifestée ici.

On peut apprécier la signification des valeurs de Q en notant que la valeur de Q calculée entre les deux groupes de onze variables est de 0,982. C'est dire que l'un ou l'autre de ces groupes est aussi voisin de l'ensemble des 27 variables que ces deux groupes, qui ont neuf variables en commun, sont voisins l'un de l'autre.

Il semble donc ici, en conclusion, que l'un ou l'autre des

groupes de 11 variables puisse être pris comme une bonne approximation de l'ensemble des 27 variables.

Le tableau n° 1 donne la matrice des variances et covariances des 27 variables.

Le tableau n° 2 montre dans quel ordre les variables sont introduites.

Le tableau n° 3 permet de comparer les cinq premières valeurs propres obtenues dans les deux méthodes avec les valeurs propres de la matrice 27×27 , pour les meilleurs groupes de cinq, dix, quinze, vingt et vingt-cinq variables.

TABLEAU N° 2

rang d'introduction	Première méthode Critère $\ U_n - U_m\ _2$		Seconde méthode Critère $Q(U_n, U_m)$	
	Indice de la variable	Q	Indice de la variable	Q
1	19	0,842	19	0,842
2	9	0,900	10	0,902
3	24	0,924	8	0,903
4	8	0,942	17	0,941
5	10	0,956	13	0,957
6	3	0,962	26	0,964
7	14	0,973	24	0,967
8	13	0,975	3	0,975
9	17	0,976	14	0,977
10	21	0,975	4	0,979
11	1	0,982	21	0,981

Tableau n°3

	5 variables		10 variables		15 variables		20 variables		25 variables		27 variables
	1e méthode	2e méthode	1e méthode	2e méthode	1e méthode	2e méthode	1e méthode	2e méthode	1e méthode	2e méthode	
$\frac{\lambda_1 n}{n}$	602,53	643,14	612,24	635,90	663,57	630,98	629,38	640,60	640,40	633,82	635,54
$\frac{\lambda_2}{n}$	72,90	93,98	86,63	127,66	125,95	157,40	115,99	115,54	137,35	137,34	131,32
$\frac{\lambda_3}{n}$	55,06	48,80	53,28	60,75	69,99	48,92	54,44	43,21	46,88	45,88	43,59
$\frac{\lambda_4}{n}$	17,60	25,02	25,10	36,97	34,17	34,27	18,86	34,94	35,58	37,04	36,82
$\frac{\lambda_5}{n}$	10,08	14,62	23,52	25,73	18,03	25,68	34,99	24,19	37,00	34,99	34,12

VII. 2 - DEUXIEME EXEMPLE

Le second exemple concerne le traitement de données recueillies par Messieurs GODRON et ROMANE du Centre d'Etude Phyto-écologique de Montpellier.

La variable aléatoire X_t observée est l'abondance de l'espèce végétale t où t parcourt un ensemble fini. Si les plantes sont indicatrices du sol qui les nourrit et du climat qu'elles subissent, deux parcelles de terrains qui portent les mêmes espèces végétales avec les mêmes abondances sont présumés semblables et donc susceptible de recevoir les mêmes cultures. D'autre part, par des études précédentes, les écologistes savent que la présence d'une espèce donnée est significative d'un sol d'une nature bien définie (calluna vulgaris pour les roches acides par exemple) si bien que l'observation des plantes leur permet de connaître l'environnement écologique.

Après avoir rejeté un certain nombre d'espèces trop peu fréquentes, les écologistes ont entrepris leur étude sur un ensemble de 120 espèces. On peut se demander si un groupe plus restreint ne serait pas susceptible d'apporter les mêmes informations. Dans l'affirmative, comment constituer ce sous ensemble ?

Il n'existe pas, à notre connaissance, de méthode globale qui ait été développée pour répondre à cette question si bien que les critères de choix introduits dans ce travail apparaissent comme les premiers essais de rationalisation dans ce domaine.

En utilisant la méthode progressive proposée, avec Q pour critère, on met en évidence un groupe de 36 espèces végétales ($Q = 0,822$) qui semble adéquat aux spécialistes de la question en ce sens qu'ils n'y découvrent aucune absence particulièrement gênante et aucune présence étonnante.

Il est d'ailleurs intéressant de noter^{que} les 12 premières espèces fournies par le calcul donnent déjà un échantillon très représentatif de l'ensemble des espèces, en ce sens que chacun des grands groupes d'espèces est représenté proportionnellement à son importance.

Sur ce même exemple nous nous sommes livrés à un travail systématique qui a consisté à comparer les valeurs de Q obtenues par la méthode progressive aux valeurs prises par Q dans des sous ensembles tirés au hasard. Le tableau n° 4 donne les résultats trouvés. Il est difficile de tirer des conclusions générales de cette comparaison ; on note cependant que la valeur Q obtenue par la méthode progressive pour n variables est toujours supérieure à la valeur la plus forte obtenue dans les sous ensembles au hasard de $(n+10)$ variables.

TABLEAU N° 4

	Groupes tirés au hasard			Méthode progressive
	Nombre de groupes	min Q	max Q	Q
10 variables	30	0,387	0,494	0,639
20 variables	30	0,539	0,648	0,739
30 variables	30	0,613	0,724	0,796
40 variables	20	0,710	0,778	0,837

TABLE DE REFERENCE

- 1 - T.W. ANDERSON - An introduction to multivariate statistical analysis - John. Wiley and Sons - 1966.
 - 2 - P. BILLINGSKY - Ergodic Theory and information - John Wiley and Sons - 1965.
 - 3 - Y. ESCOUFIER - A. LOBERT - X. CASTELLANI - Interpolation et extrapolation de fonctions aléatoires en hydrologie - D.G.R.S.T. - action concertée Eau - Rapport partiel 1969.
 - 4 - P.R. HALMOS - Measure Theory - D. Van - Nostrand Company - 1950.
 - 5 - H. HOTELLING - Analysis of a complex of statistical variables into principal components - J. Educ. Psychol. 24 ; 417-441 ; 498-520 ; 1933.
 - 6 - T. KATO - Perturbation Theory for linear operators - Springer Verlag - Berlin - Heidelberg - 1966.
 - 7 - J. NEVEU - Bases mathématiques du Calcul des probabilités - Masson et Cie - 1964.
 - 8 - F. RIESZ et B. SZ. NAGY - Leçons d'analyse fonctionnelle - 4^{ième} édition - Gauthiers Villars - 1965.
-

