HAL
open science

Méthodes exploratoires d'analyse de données temporelles

Gilbert Saporta

To cite this version:

Gilbert Saporta. Méthodes exploratoires d'analyse de données temporelles. Théorie [stat.TH]. Université Pierre et Marie Curie - Paris VI, 1981. Français. NNT : . tel-00711814

HAL Id: tel-00711814

https://theses.hal.science/tel-00711814

Submitted on 25 Jun 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

présentée à

L'UNIVERSITE PIERRE ET MARIE CURIE PARIS 6

pour obtenir le grade de

DOCTEUR ES SCIENCES MATHEMATIQUES

par

Gilbert SAPORTA

METHODES EXPLORATOIRES D'ANALYSE DE DONNEES TEMPORELLES

SOUTENUE LE 10 JUIN 1981

devant le jury composé de :

MM. D. DUGUE *Président*

J.C. DEVILLE
Y. ESCOUFIER
J. GEFFROY
P. KREE
P. LECOINTE
J.P. PAGES

Examinateurs

Je tiens à exprimer ma profonde gratitude à :

Monsieur Daniel Duguē, pour l'honneur qu'il me fait en acceptant de présider le jury de cette thèse.

Monsieur Paul Krée, qui a dirigé ces travaux et auprès de qui j'ai toujours trouvé un accueil chaleureux et une grande disponibilité. La présente rédaction doit beaucoup à sa vigilance et à ses conseils.

Monsieur Yves Escoufier qui a accepté de lire en détail le manuscrit et m'a fait bénéficier de précieuses remarques. De nombreux développements de cette thèse ont été inspirés par son approche nouvelle des problèmes statistiques.

Monsieur Jean-Claude Deville dont les travaux tant théoriques que pratiques sur l'analyse factorielle des processus m'ont profondément influencé et qui n'a jamais ménagé son temps lors de nos multiples discussions.

Monsieur Jean Geffroy pour l'intérêt qu'il porte une nouvelle fois à mes travaux.

Monsieur Pierre Lecointe pour ses encouragements et sa participation à ce jury.

Monsieur Jean-Pierre Pagès qui m'a appris l'analyse des données et la diversité de ses méthodes.

Je remercie également :

Le Département Informatique de l'Institut Universitaire de Technologie de Paris pour son soutien matériel.

Madame Dominique Tilly qui a dactylographié cette thèse avec soin et célérité.

Le service de reprographie de l'I.U.T. qui en a assuré le tirage.

Je ne saurai enfin oublier ce que je dois à mon épouse Marie-Odile qui m'a encouragé tout au long de ce travail et a créé les conditions propices à sa réalisation.

TABLE DES MATIÈRES

Pages

CHAPITRE 0 : INTRODUCTION 1

Quelques repères bibliographiques 2
L'orientation suivie 3
Formalisme adopté 5
Résumé du travail 6

CHAPITRE I : ANALYSE FACTORIELLE D'UN PROCESSUS SCALAIRE 13

I. ANALYSE EN COMPOSANTES PRINCIPALES D'UN PROCESSUS 13
I.1 Hypothèses et notations 13
I.2 Décomposition orthogonale d'un processus 15
I.3 Rapports avec l'analyse spectrale des processus stationnaires 18
I.3.a Processus observé sur [0,T] 19
I.3.b Extension à la droite réelle 21
I.3.c Exemple : le processus de Markov 25
I.4 Cas d'une population finie 27
I.4.a Généralités 27
I.4.b Analyse de plusieurs courbes : dualité et centrage 29

II. ANALYSE FACTORIELLE AVEC UNE METRIQUE 31
II.1 Définition et premieres propriétés 31
II.2 Exemples de métriques 35
II.2.a Métriques de Sobolev 36
II.2.b Le cas de transformations linéaires 37
II.2.c Métriques définies par un opérateur intégral 40

Pages

II.3 La métrique D_{1/σ^2} 41
II.3.a Propriétés essentielles 41
II.3.b Etude d'un processus particulier 42

III. MISE EN OEUVRE DE LA METHODE 50
III.1 L'approximation numérique 50
III.1.a Echantillonnage 51
III.1.b Résolution numérique 52
III.2 L'interpolation 56
III.2.a Interpolation linéaire 57
III.2.b Lien avec les espaces de Sobolev 58
III.3 Sur l'interprétation des résultats 60

Conclusion 61

CHAPITRE II : ANALYSES FACTORIELLES D'UN PROCESSUS VECTORIEL 63

I. ANALYSE CANONIQUE D'UN ENSEMBLE DE TABLEAUX DE DONNEES 65
I.1 Formulaire d'analyse canonique ordinaire 65
I.2 Généralisation à p tableaux 67

II. L'ANALYSE "HARMONIQUE" VECTORIELLE 69
II.1 Eléments d'ordre deux associés à un processus vectoriel 70
II.1.a Moyenne et covariance 70
II.1.b Opérateur de régression ou de projection 71
II.2 Analyse canonique généralisée du processus $(\overrightarrow{X_t})$ 71
II.2.a Définitions et propriétés de l'analyse harmonique vectorielle 71
II.2.b Processus scalaires propres 73
II.2.c Remarques diverses 75
II.3 Analyse harmonique vectorielle d'un processus stationnaire 77

Pages

III. AUTRES METHODES DIRECTES D'ANALYSE 79

III.1 Les six analyses factorielles d'un cube 79

III.1.a Compression sur un indice 79

III.1.b Analyses du type 1 contre 2 80

III.2 Une méthode d'analyse de la tendance : la discrimination entre instants 84

III.3 Problèmes d'optimisation 86

III.3.a Processus obtenus par combinaisons linéaires des composantes de $(\overrightarrow{X_t})$ 86

III.3.b Vecteurs aléatoires déduits linéairement de $(\overrightarrow{X_t})$ 90

IV. ANALYSE DES OPERATEURS ASSOCIES A UN PROCESSUS VECTORIEL 92

IV.1 Opérateurs associés aux composantes 92

IV.2 Opérateurs associés aux instants 94

Conclusion 97

CHAPITRE III : ANALYSE FACTORIELLE D'UN PROCESSUS QUALITATIF 99

I. L'ANALYSE "HARMONIQUE" QUALITATIVE 100

I.1 Hypothèses et notations 100

I.2 La méthode 103

I.2.a L'opérateur K et les processus propres 103

I.2.b Les génératrices des processus propres 107

I.2.c Les codages propres 109

I.2.d Remarques concernant les valeurs propres 112

I.3 Identification avec d'autres types d'analyse 113

I.3.a Analyse harmonique vectorielle de $(\overrightarrow{\mathbf{1}_t})$ 113

I.3.b Analyse des correspondances multiples continues 114

I.3.c A.C.P. semi-linéaire et codage optimal 116

1.4 Le cas des processus à deux états 119

I.4.a Un exemple simple 119

I.4.b Réduction à une analyse harmonique scalaire 125

Pages

I.5 Problèmes numériques posés par le traitement de données réelles 129

I.5.a Approximation par discrétisation 131

I.5.b Approximation par moyennage 132

II. SUR D'AUTRES METHODES FACTORIELLES 134

II.1 Les 6 analyses factorielles 134

II.1.a Les trois analyses marginales 134

II.1.b Les trois analyses mixtes 140

II.2 Analyse du processus $(\overset{\circ}{E}_t)$ 146

II.1.a La méthode 146

II.1.b Exemple 149

III. GENERALISATION AU CAS D'UN PROCESSUS QUALITATIF MULTIPLE 153

III.1 Nature des données et position du problème 153

III.2 Les diverses analyses d'un cube qualitatif 154

III.3 L'analyse harmonique qualitative généralisée 156

III.4 Analyses globales au moyen des opérateurs 161

III.4.a Opérateurs associés aux processus 161

III.4.b Opérateurs associés aux instants 162

Conclusion 164

CHAPITRE IV : ETUDE DES LIAISONS ENTRE VARIABLES ALÉATOIRES ET PROCESSUS 165

I. REGRESSION D'UNE VARIABLE ALEATOIRE REELLE Y SUR UN PROCESSUS (X_t) 166

I.1 La prévision linéaire 166

I.1.a Position du problème 166

I.1.b Régression sur composantes principales 169

I.1.c Exemple : extrapolation d'un processus à accroissements indépendants 172

I.1.d Conséquences 173

Pages

I.1.e Régression sur un processus vectoriel 175

I.1.f Variables supplémentaires en analyse factorielle et régression 175

I.1.2 Prévision non linéaire 176

I.2.a Prévision semi-linéaire 176

I.2.b Autres techniques et commentaires pratiques 178

II. DEPENDANCES ENTRE UN VECTEUR ALEATOIRE $\vec{Y}$ OU UNE VARIABLE QUALITATIVE $\mathcal{Y}$ ET UN PROCESSUS (X_t) 180

II.1 Régression multidimensionnelle 180

II.2 Analyse canonique 181

II.3 Analyse factorielle de (X_t) vers $\vec{Y}$ 185

II.3.a Le cas fini 185

II.3.b Généralisation à un processus 187

II.4 Liaisons partielles de $\vec{Y}$ à (X_t) fixé 188

II.5 Discrimination sur processus 189

II.5.a Analyse factorielle discriminante 189

II.5.b Remarques sur les variables qualitatives supplémentaires en analyse factorielle 190

III. DEPENDANCE ENTRE UN PROCESSUS ET UNE OU PLUSIEURS VARIABLES ALEATOIRES 191

III.1 Influence d'une variable aléatoire X réelle sur un processus (Y_t) 191

III.1.a Considérations générales 192

III.1.b Analyse factorielle conditionnelle 193

III.2 Le cas de plusieurs variables explicatives 196

III.3 Remarques sur l'analyse de variance sur processus 198

IV. PROBLEMES POSES PAR L'ETUDE SIMULTANEE DE DEUX PROCESSUS (Y_t) ET (X_t) 198

IV.1 Comparaison de diverses approches 198

IV.2 Sur différents modèles de régression 200

IV.2.a Modèles sans mémoire 201

IV.2.b Modèles avec mémoire 202

Conclusion 203

BIBLIOGRAPHIE 204

CHAPITRE 0

INTRODUCTION

L'étude statistique de données évoluant dans le temps repose classiquement sur l'application de la théorie des processus stochastiques et suppose pour cela que les phénomènes étudiés vérifient des hypothèses assez contraignantes , telles la stationnarité ou l'appartenance à une classe générale de processus (Markov ou Arima par exemple).

Les problèmes principaux sont alors essentiellement des problèmes d'estimation qu'il s'agisse des paramètres du modèle ou de la distribution spectrale.

S'il est parfaitement légitime dans une large gamme d'applications (en sciences physiques notamment), il faut reconnaître que dans bien d'autres situations, le recours à des modèles n'est que la contrepartie nécessaire du manque de données : ainsi les chroniques économiques sont pour la plupart des réalisations uniques par nature, de processus non répétables. Si on veut estimer par exemple leur fonction d'autocorrélation, force est bien de les supposer stationnaires et même ergodiques, fût-ce au prix de transformation des données de départ. De plus, de nombreuses techniques ont été conçues essentiellement pour des séries chronologiques, c'est-à-dire pour des observations effectuées sur un ensemble dénombrable d'instants équidistants.

Or, depuis quelques années, on commence à disposer de manière exploitable, grâce au développement des capacités de stockage informatique, de données d'un type nouveau : des séries individuelles temporelles, c'est-à-dire des "tableaux" à trois indices contenant les valeurs $x_t^j(\omega)$ de p variables ($j=1,2,..,p$) aux temps t (t variant de 0 à T) pour n individus ($\omega = 1,2,..,n$).

Disposant d'un grand nombre d'individus il devient alors possible d'effectuer de véritables analyses statistiques sans être obligés de recourir à des modèles préétablis : on reconnaît ici la démarche exploratoire (au sens de J.W. TUKEY [64]) de l'analyse des données ; cependant il convient d'adapter les méthodes existantes et de développer de nouveaux outils pour tenir compte du caractère évolutif des tableaux étudiés. En outre, il devient fréquent que les variables soient repérées en continu et non plus à certaines dates fixes, que ce soit en raison du caractère ponctuel du processus (naissance, changement d'état matrimonial par exemple) ou de la méthode de mesure (courbes de température en météorologie). Il faut donc pouvoir traiter une infinité non dénombrable de variables ce qui implique d'utiliser les résultats de l'analyse fonctionnelle et de la théorie des processus.

Quelques repères bibliographiques

Si l'outil fondamental de l'analyse factorielle des processus, la décomposition de KARHUNEN-LOEVE, date des années 1945-1947 [42] [33], l'étude par l'analyse des données ou des techniques proches, de l'évolution de données individuelles est cependant assez récente et encore peu répandue.

En France, les premiers travaux datent, nous semble-t-il, de 1970 avec la thèse d'Y. ESCOUFIER [23] qui abordait le problème de façon originale au moyen d'opérateurs associés à un processus ; leurs liens avec l'analyse des données fut effectué plus tard par J.M. BRAUN [10] et J.P. PAGES [51]. C'est l'article de J.C. DEVILLE [17] en 1974 qui marque à notre avis le début véritable de l'application, dans une optique d'analyse de données, de la décomposition de Karhunen-Loeve d'un processus. Sous le nom d'analyse harmonique qualitative cette méthode fut étendue par J.C. DEVILLE et nous même [20] aux processus qualitatifs en 1979. Entre temps M. MASSON [43] puis J. DAUXOIS et A. POUSSE [16] développaient la théorie abstraite de l'analyse factorielle des fonctions aléatoires en dépassant le cadre linéaire des méthodes usuelles. Ces travaux furent suivis sur le plan théorique par ceux de BESSE [7], ROMAIN [56], BOUMAZA [8], et dans le domaine appliqué par MAZEROLLE [46] et OBLED [50].

Concernant les séries chronologiques multiples, il faut citer les travaux de M. TENENHAUS [62] et de J.M. BOUROCHE [9] ainsi que ceux de T. FOUCART [27] et E. STEMMELEN [61] sur les évolutions de tableaux de contingence (mais il s'agit là de données agrégées et non individuelles).

Il est notoire que l'orientation des diverses écoles anglo-saxonnes est fort différente. Bien que les relations entre l'analyse spectrale et la diagonalisation de l'opérateur d'autocorrélation soient bien connues dans le cas discret et figurent dans les manuels consacrés aux séries chronologiques, tel celui de FULLER [28], la décomposition de Karhunen-Loeve apparaît plus comme une propriété mathématique que comme base d'une méthode statistique, même pour les spécialistes de la théorie du signal comme PAPOULIS [53]. Il faut attendre 1979 pour en voir publiée une application dans une revue anglo-saxonne de statistique [3], encore est-ce dans le cas d'une réalisation unique d'un processus stationnaire et dans le but de proposer aux praticiens une méthode plus rapide à comprendre que l'analyse spectrale...

L'utilisation des techniques multidimensionnelles a cependant été développée dans les remarquables travaux de BRILLINGER [11] (régression, analyse en composantes principales, analyse canonique) mais transposés dans le domaine spectral pour l'étude de filtres invariants pour des processus stationnaires.

Les rares travaux concernant des trajectoires individuelles sont plutôt le fait des économètres et relèvent d'une approche modélisante, on en trouve une bonne synthèse dans le colloque organisé par l'INSEE en 1977 [69].

L'orientation suivie

Notre travail répond à un double but :

. dresser une synthèse des apports de l'analyse des données dans le domaine des données individuelles temporelles étudiées sur un intervalle de temps fini [0,T],

. proposer et étudier des méthodes nouvelles pour des problèmes non encore résolus comme l'étude de processus qualitatifs.

Nous espérons ainsi montrer en particulier la fécondité des méthodes d'analyse factorielle dans le domaine temporel, en ne prétendant nullement établir cependant une quelconque supériorité sur les méthodes classiques pour les raisons suivantes :

. d'abord les données que nous voulons traiter ne sont pas les mêmes, elles sont à certains égards plus riches puisqu'on dispose de n trajectoires au lieu d'une ; en contrepartie elles sont évidemment plus difficiles à obtenir,

. surtout parce que les objectifs ne sont pas les mêmes du moins en ce qui concerne les analyses descriptives. En schématisant très fortement, on peut dire en effet qu'à l'inverse des méthodes de la statistique inférentielle, où l'on s'intéresse plutôt aux variables et à leurs interrelations (les individus forment alors un corpus homogène d'objets anonymes et interchangeables puisqu'issus d'un échantillonnage aléatoire), l'analyse des données est orientée principalement vers les individus et la mise en évidence de facteurs d'hétérogénéité d'une population. Nous traduisons cela en disant brièvement que l'analyse des données est souvent essentiellement typologique.

Ainsi le problème ne sera pas tant de découvrir des périodicités comme le fait l'analyse spectrale, que de trouver les traits dominants de différenciation des évolutions des individus. Les méthodes factorielles fournissent alors, selon des variables indépendantes du temps, des positionnements multidimensionnels des individus, représentables par des cartes où l'on visualisera les proximités ou les oppositions entre individus selon qu'ils ont eu des évolutions similaires ou divergentes. Cependant lorsque le processus étudié peut être considéré comme stationnaire les résultats de l'analyse factorielle seront équivalents à ceux de l'analyse spectrale.

. L'étude des liaisons entre variables fait certes partie des objectifs de l'analyse des données, avec des méthodes telles que l'analyse canonique ou l'analyse factorielle des corrélations partielles, et rejoint en cela les objectifs des techniques classiques d'analyse de processus, mais il est clair qu'une étude préalable des hétérogénéitéséventuelles est indispensable pour pouvoir interpréter les liaisons trouvées, surtout si on utilise des méthodes linéaires : on connait la sensibilité des indicateurs de corrélation linéaires à la non-homogénéité des données.

. Enfin lorsqu'on veut prévoir le comportement futur d'un phénomène, les modèles probabilistes nous semblent indispensables et les techniques d'analyse de liaison et d'ajustements linéaires déduites de l'analyse des données sont alors insuffisantes bien qu'indispensables en première approche.

Bien que les méthodes de classification constituent une part importante des méthodes exploratoires d'analyse des données, nous ne les abordons pas dans notre travail, nous limitant aux méthodes factorielles et linéaires. La raison en est que la dimension temporelle des phénomènes ne nous a pas paru susceptible d'apporter des éléments nouveaux : classer des trajectoires ne soulève pas de problème particulier dès lors qu'on utilise une métrique adaptée et renvoie à l'utilisation des mêmes algorithmes que dans le cas non-temporel.

Formalisme adopté

S'agissant d'un travail d'analyse des données, le lecteur pourra être surpris de nous voir utiliser le vocabulaire du calcul des probabilités et des processus stochastiques et de ne trouver que des exemples d'école et non des données réelles. Il convient donc de nous en expliquer :

si nous avons choisi de parler de "processus" et de "probabilités" là où certains ne pourraient voir que "courbes" et "fréquences", c'est que ce langage nous a semblé le plus simple à utiliser et le mieux adapté au cas d'une infinité non dénombrable de variables. Il reste cependant que nous avons toujours eu à l'esprit le fait que Ω représente une population d'individus et un processus un ensemble de trajectoires individuelles.

Quant aux exemples théoriques, leur traitement nous a semblé intéressant tout d'abord pour les développements mathématiques auxquels il conduisent et parce qu'ils montrent que l'on peut résoudre jusqu'au bout les équations obtenues. Mais surtout leur principal intérêt est à notre sens le suivant : on ne peut bien comprendre une méthode d'analyse des données et l'appliquer valablement que si l'on a établi des règles d'interprétation des résultats sur des données parfaitement connues : les processus standards, comme le processus de Poisson par exemple, constituent alors d'excellents jeux d'essai.

Notre travail est donc théorique mais nous ne l'avons pas voulu abstrait : c'est ainsi que nous n'avons pas recherché systématiquement les énoncés les plus généraux possibles ni les hypothèses mathématiques minimales. Non seulement cela correspond à notre conception personnelle de la recherche en statistique qui doit fournir des méthodes orientées vers les applications, mais encore à la réalité des phénomènes abordés : ainsi traiter d'abord des processus scalaires

puis des processus vectoriels peut paraître redondant puisque les premiers sont des cas particuliers des seconds ; en fait il n'en est rien car en changeant de dimension le phénomène change de nature et les techniques sont différentes. De même, il aurait été possible d'étudier des fonctions aléatoires générales, l'indice t appartenant alors à un ensemble non ordonné quelconque : on y aurait gagné en extensions et peut-être en applications mais le support concret fourni par le temps aurait été perdu. Nous nous sommes également attachés à donner partout ou cela était possible des formules explicites pour les calculs ce qui permet de faire le lien avec les notations usuelles de l'analyse des données.

D'un point de vue terminologique nous utilisons ici le vocable "d'analyse factorielle" sous le sens large qui prévaut actuellement en France, et qui recouvre diverses méthodes linéaires comme l'analyse en composantes principales, l'analyse des correspondances, etc., et non dans l'acception limitée du modèle des facteurs communs et spécifiques. Il s'agit certes d'un abus de langage : l'analyse factorielle au sens strict étant conceptuellement différente de l'analyse en composantes principales.

De même le nom d'"analyse harmonique" introduit par J.C. DEVILLE pour désigner des analyses en composantes principales de processus peut prêter à confusion avec la théorie mathématique du même nom (à laquelle elle se rattache il est vrai) : cependant l'usage en étant maintenant bien établi en analyse des données chronologiques, nous nous y conformons également.

Résumé du travail

Les trois premiers chapitres sont consacrés à des généralisations de l'analyse factorielle à trois types de processus.

Le chapitre I traite des processus scalaires, c'est-à-dire du cas où on observe sur une population l'évolution d'une variable aléatoire réelle au cours du temps. Il s'agit essentiellement d'un chapitre de synthèse. Après des rappels concernant la décomposition de Karhunen-Loeve d'un processus (X_t) observé sur un intervalle de temps fini $[0,T]$ on examine les rapports de cette dernière avec l'analyse spectrale des processus stationnaires. On montre que les deux analyses sont identiques pour des processus périodiques étudiés sur un intervalle

de temps fini et qu'elles coïncident, moyennant une généralisation de l'analyse en composantes principales, lorsque la période d'étude est étendue à $\mathbb{R}$ tout entier. L'intérêt de considérer des processus stationnaires en analyse des données nous paraît lié au problème suivant souvent posé par des praticiens : ayant effectué l'analyse factorielle d'un processus jusqu'à une date T peut-on extrapoler jusqu'à T+h et que se passe-t-il si h tend vers l'infini ? Ce problème n'a de sens à notre avis que si le comportement du processus n'est pas susceptible de varier fondamentalement après T, sinon il est évident que les facteurs de l'analyse vont changer. Il doit donc exister une certaine permanence de la structure de dépendance entre les X_t ce qui conduit assez naturellement à des hypothèses de stationnarité.

Le cas ou Ω est fini est l'occasion de préciser les notions de dualité entre l'A.C.P. du processus (X_t) dont on connaît n réalisations et celle de la chronique n-dimensionnelle constituée des n trajectoires.

Un des concepts fondamentaux en analyse des données est celui de métrique associée à un ensemble de variables, qui permet de définir une distance entre individus. Les contributions originales dans ce chapitre concernent principalement l'extension de la notion de métrique à des processus afin de comparer des trajectoires. Nous proposons une formulation générale permettant de traiter les divers cas de métriques applicables à des processus (dont les métriques de Sobolev en particulier) ; le fait de s'écarter du cas élémentaire d'une métrique définie par un homéomorphisme de $L^2(T)$ (traité par J. DAUXOIS et A. POUSSE [16]) introduit des problèmes spécifiques de l'analyse fonctionnelle car les opérateurs associés ne sont pas forcément bornés. On étudie ensuite plus spécialement la métrique D_{1/σ^2} , conduisant à un processus réduit, en montrant sur un exemple son intérêt pour des processus à variance croissante : cette métrique permet d'éliminer dans une certaine mesure la première composante triviale liée à la chronologie. Le traitement mathématique des processus à accroissements indépendants fait alors intervenir certaines propriétés des fonctions de Bessel.

Le chapitre s'achève par une partie consacrée à la mise en oeuvre de la méthode sur des données concrètes et étudie particulièrement les problèmes posés par l'approximation et l'interpolation. On y reprend certains travaux récents effectués par P. BESSE [7].

Les processus vectoriels sont étudiés au chapitre II.

La donnée d'un processus m-dimensionnel $(\vec{X}_t)$ étant la généralisation du cas de l'analyse simultanée de plusieurs tableaux de données, nous rappelons brièvement les résultats que nous avions établi antérieurement concernant l'analyse canonique généralisée. L'extension, nouvelle, au cas d'une infinité non dénombrable de vecteurs aléatoires est effectuée sous le nom d'analyse "harmonique" vectorielle pour reprendre un terme introduit par J.C. DEVILLE bien que notre méthode en diffère par l'introduction d'une métrique "glissante" décorrélant à tout instant les composantes de $\vec{X}_t$.

L'introduction de cette métrique est indispensable pour assurer la cohérence de l'analyse "harmonique" vectorielle avec l'analyse canonique généralisée et l'analyse des processus qualitatifs conçue comme analyse des indicatrices des états.

Cette métrique de plus rend l'analyse invariante par changement de base, pour des processus à valeurs dans un espace vectoriel $\mathbb{V}$. Contrairement à l'analyse d'un processus scalaire, on ne trouve pas de liens directs avec l'analyse spectrale des processus stationnaires, le type de décomposition en variables scalaires (coordonnées d'individus) et facteurs vectoriels de notre analyse étant différent de celui de la représentation de Cramer des processus vectoriels. Bien que l'analyse "harmonique" vectorielle soit la méthode, à notre avis, la plus intéressante pour l'étude des trajectoires individuelles il existe cinq autres possibilités d'analyse factorielle de tableaux à 3 indices que nous étudions systématiquement et dont une présente un intérêt indéniable pour l'étude des variations des dépendances entre composantes de $(\vec{X}_t)$: il s'agit d'une extension d'une méthode proposée par M. TENENHAUS [6] pour les chroniques multidimensionnelles.

Chaque méthode est alors caractérisée par des propriétés d'optimalité : que ce soit celles de vecteurs aléatoires obtenus par combinaisons dans le temps du type $\vec{Y} = \int_0^T \vec{X}_t \, f(t) \, dt$ ou de processus scalaires obtenus par combinaison linéaire des composantes de $(\vec{X}_t)$ du type $\xi_t = \langle \vec{X}_t \, ; \, \vec{f}(t) \rangle$ ou $\xi_t = \langle \vec{X}_t \, ; \, \vec{a} \rangle$.

Enfin on propose deux méthodes, plus globales, fondées sur l'utilisation des opérateurs introduits par Y. ESCOUFIER, l'une permettant de comparer les m

processus scalaires associés aux m composantes de $(\vec{X}_t)$, l'autre étudiant la chronologie du processus en associant à chaque instant un opérateur ce qui revient à étudier un processus à valeurs opérateurs.

Le chapitre III développe et complète sur de nombreux points la méthode d'analyse des processus qualitatifs que nous avions proposé avec J.C. DEVILLE [20] sous le nom d'analyse harmonique qualitative. Cette méthode basée sur l'introduction d'un opérateur caractéristique d'un processus qualitatif analogue à un opérateur de covariance, mais en plus complexe, apparaît comme une extension de l'analyse des correspondances multiples à un continuum de variables qualitatives et comme une application de l'analyse harmonique vectorielle au processus des indicatrices des états. On montre comment on passe de l'obtention de processus scalaires propres transformant par codage le processus qualitatif (X_t) en un processus numérique (X_t) optimal en un certain sens, à celle des "génératrices" ou composantes principales permettant le positionnement des individus dans les plans factoriels, puis à celle des codages propres ou facteurs. Nous donnons ensuite des propriétés spécifiques des processus à deux états en traitant, à titre d'exemple, un processus simple avec changement d'état irréversible uniformément réparti dans le temps. Le traitement numérique est ensuite abordé avec notamment une nouvelle démonstration de la méthode d'approximation proposée en [20].

Un processus qualitatif étant un cas particulier d'un processus vectoriel nous reprenons les cinq autres analyses possibles des tableaux cubiques pour aboutir à une conclusion opposée à celle du cas vectoriel : ici seule une analyse marginale, celle des temps passés dans chaque état entre [0 et T] présente un intérêt certain alors qu'elle en était dénuée pour les processus vectoriels tandis que celle retenue pour les processus vectoriels aboutit à des résultats triviaux. Ces diverses analyses sont illustrées par l'exemple du processus à deux états déjà utilisé.

Ici une seule analyse globale, par opérateurs, a un sens : c'est celle du processus dont les éléments $\overset{\circ}{E}_t$ sont les opérateurs d'espérance conditionnelle à X_t.

Enfin on propose et on justifie des méthodes pour le problème, non encore traité, des processus qualitatifs multiples où l'on observe p variables qualita-

tives au cours du temps. Une de ces méthodes est l'extension de l'analyse harmonique qualitative et revient à une double analyse canonique généralisée : une à chaque instant sur les variables $(X_t^1, X_t^2, \ldots, X_t^p)$, l'autre sur le processus vectoriel des composantes issues de la première.

Les trois types d'analyse "harmoniques" étudiés dans ces trois chapitres forment un ensemble dont la cohérence apparaît, en particulier, dans le tableau suivant donnant les critères d'optimalité des composantes principales associées à chaque type de processus.

Analyse "Harmonique" scalaire réduite (X_t)	Analyse "Harmonique" vectorielle $(\vec{X}_t)$	Analyse "Harmonique" qualitative (X_t)
$\max_z \int_0^T r^2(z;X_t)\, dt$	$\max_z \int_0^T R^2(z;\vec{X}_t)\, dt$	$\max_z \int_0^T \eta^2(z;X_t)\, dt$

où r, R, η désignent respectivement le coefficient de corrélation simple, le coefficient de corrélation multiple, et le rapport de corrélation.

Notre travail n'aurait pas été complet sans une étude des liaisons pouvant exister entre un processus et diverses variables ou entre deux processus. Il s'agit donc au sens large de ce qu'on peut appeler des problèmes de régression où un processus peut être soit prédicteur soit prédictand. Le chapitre IV concerne donc des généralisations de la régression multiple, de l'analyse canonique, de l'A.C.P. sur covariances partielles etc.

La passage à la dimension infinie pose alors des problèmes d'une nature différente de ceux rencontrés jusqu'alors en analyse factorielle. En effet, si les facteurs de l'A.C.P. d'un processus (X_t) étaient des fonctions f(t) de $L^2(T)$ et les composantes principales des variables de la forme $\int_0^T X_t\, f(t)\, dt$ il n'en est pas de même pour les coefficients de régression et la projection $\hat{Y}$ d'une variable Y sur un processus (X_t) : ce n'est qu'exceptionnellement que $\hat{Y}$ se met sous la forme $\int_0^T X_t\, \beta(t)\, dt$.

Le premier paragraphe rappelle donc ces aspects spécifiques de la régression d'une variable sur un processus et en tire les conséquences pour les applications numériques. On y propose également diverses méthodes de régression non-linéaire conduisant à des solutions calculables. L'ensemble des techniques de résolution numérique repose sur la décomposition préalable de (X_t) en composantes principales et facteurs principaux.

On étudie ensuite le cas où la variable à expliquer est vectorielle. C'est l'occasion de définir la procédure permettant de réaliser l'analyse canonique d'un processus (X_t) et d'un vecteur aléatoire à m dimensions $\vec{Y}$.

On étend également à un processus la méthode d'A.C.P. sur variables instrumentales proposée par C.R. RAO [55] ainsi que l'analyse factorielle discriminante.

Diverses méthodes pour analyser le problème inverse, celui de l'influence d'une ou plusieurs variables aléatoires sur le déroulement d'un processus, sont développées. Entre autres on insiste sur la technique d'A.C.P. d'un processus (X_t) conditionnellement à une variable en reprenant l'exemple des processus à accroissements indépendants où on élimine ainsi l'effet dans la valeur finale X_T .

L'étude simultanée de deux processus (X_t) et (Y_t) marque à notre avis les limites de l'analyse des données car les méthodes déduites de l'analyse canonique, outre les difficultés mathématiques inhérentes à la dimension infinie, ne permettent pas une prévision causale au sens de la théorie du signal et ne sont valables qu'a posteriori. Le problème semble plus relever de techniques de filtrage probabilistes avec introduction de modèles.

Ajoutons pour terminer que de nombreuses remarques ne concernent pas seulement les données temporelles mais sont également valables pour l'analyse de données en général.

CHAPITRE I

ANALYSE FACTORIELLE D'UN PROCESSUS SCALAIRE

Ce chapitre est consacré à la généralisation de l'analyse en composantes principales à une famille infinie de variables aléatoires réelles X_t où t varie dans un intervalle borné de $\mathbb{R}$ noté $[0,T]$ ou simplement T. Le but est de traiter des données du type suivant : évolution d'un ensemble d'individus au cours du temps décrits par une variable X . Ce problème étant fondamental pour la suite, il nous a paru utile de rassembler et de présenter d'abord de manière synthétique un certain nombre de résultats classiques concernant la décomposition de Karhunen-Loève d'un processus et ses liens, moins connus, avec l'analyse spectrale. Outre quelques remarques concernant le cas d'une population finie, les développements originaux concernent principalement l'usage de métriques quelconques et sont regroupés dans le deuxième paragraphe : on y étudie en particulier l'influence de la réduction des données dans l'analyse de processus à accroissements indépendants.

Enfin le troisième paragraphe où sont exposés les problèmes de mise en oeuvre constitue une synthèse de travaux récents concernant l'approximation et l'interpolation et renferme des considérations d'ordre pratique.

I - ANALYSE EN COMPOSANTES PRINCIPALES D'UN PROCESSUS

I-1 Hypothèses et notations

Le modèle mathématique est le suivant : étant donné un espace probabilisé (Ω,a,P) et un intervalle fini $[0,T]$ de $\mathbb{R}$ muni de la mesure de Lebesgue dt, on se donne une fonction mesurable de $T \times \Omega$ dans $\mathbb{R}$ notée :

$$(t,\omega) \to X(t,\omega)$$

On suppose que la classe de cette fonction appartient à $L^2(T \times \Omega)$. Par la suite, pour éviter des difficultés non essentielles de langage on identifiera une telle fonction avec sa classe d'équivalence. Ceci entraîne que $X(t,.)$ noté encore (X_t) définit une application de T dans $L^2(\Omega)$ c'est-à-dire un processus du second ordre : pour tout t X_t est une variable ayant une espérance $m(t)$ et une variance $\sigma^2(t)$ finies. De même $X(.,\omega)$ définit une application de Ω dans $L^2(T)$, ce qui veut dire que pour tout ω , sa trajectoire notée $X_t(\omega)$ est de carré intégrable. On supposera désormais le processus (X_t) centré c'est-à-dire $m(t) = 0$ pour tout t .

On a alors les résultats suivants ([16] [23] par exemple) .

Soit U le processus linéaire défini par (X_t) , c'est-à-dire l'application de $L^2(T)$ dans $L^2(\Omega)$ définie par :

$$U(f) = Y = \int_0^T X_t \, f(t) \, dt$$

U est alors un opérateur de Hilbert-Schmidt c'est-à-dire que pour toute base orthogonale (f_i) de $L^2(T)$ $\sum_i \| U(f_i) \|^2$ est fini et est indépendant de la base.

Il en est de même de son adjoint U^* opérateur de $L^2(\Omega)$ dans $L^2(T)$ défini par :

$$U^*(Y) = E(X_t \, Y) = f(t)$$

L'opérateur de covariance du processus est l'opérateur de $L^2(T)$ dans lui-même défini par :

$\forall \, f,g \in L^2(T)$

$$(f|C\,g) = E\,[\,(f|X_t)(g|X_t)] = E\,[\int_0^T X_t \, f(t) \, dt \int_0^T X_s \, g(s) \, ds]$$

où $(|)$ désigne le produit scalaire de $L^2(T)$.

C'est l'opérateur intégral de noyau $C(t,s) = E(X_t X_s)$ c'est-à-dire que

$$f \to Cf = g \qquad\qquad g(t) = \int_0^T E(X_t X_s) \, f(s) \, ds \quad ,$$

On montre alors que $C = U \circ U^*$ ce qui prouve que C qui est positif et autoadjoint est également nucléaire comme produit d'opérateurs de Hilbert-Schmidt.

Pour toute base orthogonale (f_i) de $L^2(T)$ on a donc :

$$\sum_i (f_i | C\, f_i) = \text{Trace } C = \int_0^T C(t,t)\, dt < \infty \quad .$$

L'opérateur $W = U \circ U^*$ de $L^2(\Omega)$ dans lui-même obtenu en composant U et U^* dans l'autre sens est également nucléaire et on trouve sans difficulté que

$$W(Y) = \int_0^T X_t\, E(X_t\, Y)\, dt$$

Cet opérateur a été introduit par Y. ESCOUFIER.

En identifiant les espaces à leurs duaux, c'est-à-dire en utilisant les produits scalaires usuels de $L^2(T)$ et $L^2(\Omega)$ on a le schéma de dualité suivant :

$$\begin{array}{ccc} & U^* & \\ L^2(T) & \longleftarrow & L^2(\Omega) \\ I \downarrow\uparrow C & & W \downarrow\uparrow I \\ L^2(T) & \longrightarrow & L^2(\Omega) \\ & U & \end{array}$$

<u>Remarque</u> :

Rien ne change si on remplace l'intervalle [0,T] muni de la mesure dt par un espace mesurable quelconque T muni d'une mesure positive finie. Le cas où T est un ensemble fini correspond à la situation de l'analyse des données usuelles. En particulier pour une population finie de n individus l'opérateur U se confond avec le tableau de données individus - caractères considéré comme une matrice à n lignes et U^* avec la transposée de cette matrice.

I.2 <u>Décomposition orthogonale d'un processus</u>

Cherchons une représentation du processus (X_t) sous forme d'une somme de processus quasi-déterministes du type :

$$X_t = \sum_{k=1}^{\infty} \xi_k\, f_k(t)$$

où les ξ_k sont des variables aléatoires indépendantes du temps et les $f_k(t)$ des fonctions certaines du temps. Si on impose aux $f_k(t)$ de former une base orthogonale de $L^2(T)$ on obtient aisément les ξ_k en multipliant la formule précédente par $f_k(t)$ et en intégrant sur t :

$$\int_0^T X_t\, f_k(t)\, dt = \int_0^T \sum_\ell \xi_\ell\, f_\ell(t)\, f_k(t)\, dt = \sum_\ell \xi_\ell\, \delta_{k\ell} = \xi_k$$

et on montre alors la convergence en moyenne quadratique de $\Sigma\, \xi_k\, f_k(t)$ vers X_t pour presque tout t .

Les variables aléatoires ξ_k n'ont aucune raison d'être orthogonales, c'est-à-dire non corrélées ce qui semble une propriété souhaitable statistiquement, pour un choix arbitraire des f_k ; en particulier le développement en série de Fourier obtenu en prenant pour $f_k(t)$ $\sin \frac{2k\Pi t}{T}$ et $\cos \frac{2k\Pi t}{T}$ ne donne des ξ_k non corrélés que dans certains cas particuliers.

Le résultat fondamental est alors le suivant :

<u>Proposition 1.1 (Décomposition de Karhunen-Loève)</u>

Pour que $X_t = \sum_{k=1}^{\infty} \xi_k\, f_k(t)$ m.q. où les $f_k(t)$ forment un système orthogonal de $L^2(T)$ et les ξ_k un système orthogonal de $L^2(\Omega)$, il faut que les f_k soient les fonctions propres de l'opérateur de covariance C du processus et les $\|\xi_k\|^2$ les valeurs propres correspondantes.

En effet on a $E(X_t\, \xi_\ell) = \sum_k E(\xi_k\, \xi_\ell)\, f_\ell(t) = f_\ell(t)\, \|\xi_\ell\|^2$

et $E(X_t\, \xi_\ell) = E(X_t \int_0^T X_s\, f_\ell(s)\, ds) = \int_0^T E(X_t\, X_s)\, f_\ell(s)\, ds$.

L'opérateur C étant autoadjoint positif et nucléaire, donc compact, son analyse spectrale est possible sans recourir à l'intégrale hilbertiènne et il existe bien un système orthogonal dénombrable de fonctions propres $f_k(t)$ associées à des valeurs propres λ_k positives. Si toutes les valeurs

propres sont simples il n'existe donc qu'un seul système de fonctions orthogonales f_k (définies au signe près) et de variables aléatoires ξ_k non corrélées tel que :

$$X_t = \sum_{k=1}^{\infty} \xi_k \, f_k(t)$$

(il est clair que si une valeur propre est multiple on peut prendre n'importe quelle base orthonormée du sous-espace propre engendré).

Les $f_k(t)$ sont les facteurs principaux de (X_t) (appelés harmoniques par J.C. DEVILLE) et les ξ_k les composantes principales. On a également le :

<u>Théorème de Mercer</u>

Si (X_t) est continu en moyenne quadratique $\sum_k \lambda_k \, f_k(t) \, f_k(s)$ converge uniformément vers $C(t,s)$.

et on sait que [23] :

<u>Proposition 1.2</u>

Les ξ_k sont fonctions propres de l'opérateur W .

Si on numérote les ξ_k selon les λ_k décroissants on sait également que :

<u>Proposition 1.3</u>

Le processus $\sum_{k=1}^{\ell} \xi_k \, f_k(t)$ est la meilleure approximation en moyenne quadratique de (X_t) par une somme de ℓ processus quasi-déterministes.

Proposition 1.4

Les extrema de $E(\xi^2)$ pour ξ de la forme $\int_0^T X_t\, f(t)\, dt$, avec f appartenant à la boule unité de $L^2(T)$, valent λ_k et sont réalisés pour les composantes principales ξ_k .

On retrouve ici la définition d'HOTELLING des composantes principales en tant que combinaisons linéaires des X_t de variance maximale.

Comme la covariance entre ξ_k et X_t vaut $E(\xi_k\, X_t) = \lambda_k\, f_k(t)$ on en déduit de même que les ξ_k réalisent les extrema de $\int_0^T (\mathrm{cov}(\xi;X_t))^2\, dt$.

Le coefficient de corrélation linéaire entre X_t et ξ_k vaut

$$r(\xi_k;X_t) = \frac{\sqrt{\lambda_k}\, f_k(t)}{\sigma(t)}$$ ce qui permet l'interprétation des composantes principales.

I.3 Rapports avec l'analyse spectrale des processus stationnaires

Considérons un processus stationnaire au sens large c'est-à-dire que :

$$m(t) = m$$

$$C(t;s) = C(|t-s|) \quad \text{d'où} \quad \sigma^2(t) = \sigma^2$$

Pour simplifier nous supposerons m=0 . Notre propos est de montrer que dans ce cas l'analyse factorielle fournit une décomposition équivalente à celle de l'analyse spectrale. Avant d'aborder le cas le plus important sur le plan théorique, c'est-à-dire celui d'un processus défini sur la droite réelle, nous examinerons au préalable le cas où le processus n'a été observé que sur un intervalle fini. Enfin nous rappellerons les rapports, dans le cas où T est un ensemble discret, entre l'analyse en composantes principales usuelle et l'analyse spectrale.

I.3.a) Processus observé sur [0,T]

Notons tout d'abord que si (X_t) n'a été observé que sur un intervalle borné on ne peut définir stricto-sensu l'analyse spectrale de (X_t) car celle-ci suppose que t parcourt $\mathbb{R}$ entier. Pour l'analyse en composantes principales, le problème est en quelque sorte inverse : celle-ci ne présente pas de difficulté sur [0,T] tandis que sa généralisation à la droite réelle soulève des problèmes théoriques.

On dira que (X_t) est pseudo-stationnaire si sa fonction de covariance $C(t;s)$ est égale à $\varphi(t-s)$ où φ est une fonction paire définie sur $[-T;T]$.

On peut toujours décomposer (X_t) en une somme de processus harmonique élémentaires du type :

$$A_k \cos \omega_k t + B_k \sin \omega_k t \quad .$$

les ω_k forment alors un ensemble dénombrable.

Notons ici une ambiguité qui semble peu explicitée dans la littérature: il existe en fait deux adaptations possibles de l'analyse spectrale selon que l'on utilise la décomposition de Fourier de (X_t) ou celle de sa fonction de covariance. On a en effet d'une part la décomposition en série de Fourier de (X_t) sur [0,T] sous la forme :

$$X_t = \sum_{k=0}^{\infty} \left(A_k \cos \frac{2k\Pi t}{T} + B_k \sin \frac{2k\Pi t}{T}\right)$$

où les variables A_k et B_k ne forment pas en général un système orthogonal de $L^2(\Omega)$: elles sont corrélées.

Par contre on obtient des variables non corrélées en utilisant la décomposition de Fourier de la fonction de covariance sur $[-T;T]$: comme celle-ci est paire on n'utilise que les cosinus et l'intervalle étant de largeur 2T on trouve

$$C(t;s) = \sum_k \sigma_k^2 \cos\left(\frac{k\Pi(t-s)}{T}\right)$$

avec $\sigma_k^2 = \frac{2}{T} \int_0^T C(x) \cos \frac{k\Pi x}{T} dx$ et $\sigma_0^2 = \frac{1}{T} \int_0^T C(x) dx$

d'où $X_t = \sum_{k=0}^{\infty} (\xi_k \cos \frac{k\Pi t}{T} + \eta_k \sin \frac{k\Pi t}{T})$

et les ξ_k et les η_k sont des variables aléatoires non corrélées de variance σ_k^2 (cf. [65] p. 412).

Cependant les $\cos \frac{k\Pi t}{T}$ et $\sin \frac{k\Pi t}{T}$ ne forment pas un système orthogonal de $L^2(T)$ mais de $L^2(2T)$. Il ne peut y avoir identité entre ces deux décompositions que dans le cas d'un processus de période T en moyenne quadratique c'est-à-dire tel que $C(t;s+T) = C(t;s)$ car dans la première décomposition tous les termes impairs doivent être nuls.

L'ambiguïté est alors levée : le processus étant défini sur le tore l'ensemble des t a une structure de groupe abelien localement compact ce qui est la condition d'existence de l'analyse spectrale.

Dans ces conditions pour un processus de période T on a :

$$X_t = \sum_{k=0}^{\infty} (\xi_k \cos \frac{2k\Pi t}{T} + \eta_k \sin \frac{2k\Pi t}{T})$$

les fonctions du temps et les variables aléatoires forment alors deux systèmes orthogonaux de $L^2(T)$ et $L^2(\Omega)$ respectivement et sont donc les facteurs principaux et les composantes principales de (X_t).

On en déduit la

<u>Proposition 1.5</u>

Il n'y a identité entre la décomposition de Fouriér de (X_t) et la décomposition de Karhunen-Loève que si et seulement si (X_t) est périodique de période T en moyenne quadratique.
$f(t) = 1$ est alors fonction propre de C associée à la valeur propre $\lambda_0 = \int_0^T C(x) dx$. Les fonctions trigonométriques $\cos \frac{2k\Pi t}{T}$ et $\sin \frac{2k\Pi t}{T}$ pour $k=1,2,\ldots,\infty$ sont également fonctions propres associées aux valeurs doubles $\lambda_k = \int_0^T C(x) \cos \frac{2k\Pi x}{x} dx$.

La démonstration élémentaire de la deuxième partie de cette proposition figure par exemple dans [35] p. 73.

Notons que dans la proposition précédente les λ sont numérotées non pas par importance décroissante mais selon les fréquences croissantes des fonctions trigonométriques.

Pour un processus stationnaire non périodique l'A.C.P. fournira des résultats différents des deux analyses "spectrales" . Quels en sont les avantages respectifs ? Les analyses "spectrales" fournissent des décompositions simples à calculer selon des fonctions trigonométriques mais laquelle choisir ? Par ailleurs le choix des sinus et cosinus est aussi arbitraire que celui d'une autre famille de fonctions si le processus n'est pas périodique. De son côté l'analyse factorielle nécessite la réduction de l'opérateur de covariance, ce qui n'est pas toujours aisé, mais fournit rappelons-le encore l'unique décomposition en fonctions orthogonales du temps et variables non corrélées ce qui nous semble une propriété agréable.

I.3.b) Extension à la droite réelle

Nous allons voir que si on étend à $\mathbb{R}$ l'intervalle d'étude du processus, l'analyse en composantes principales de (X_t), moyennant une généralisation, se confond avec l'analyse spectrale de (X_t) . Nous supposerons pour simplifier que (X_t) est continu en moyenne quadratique et que sa fonction de covariance $C(t-s)$ est absolument sommable. On sait qu'il existe alors une densité spectrale $s(\omega)$ qui est une fonction bornée et continue définie par la relation :

$$s(\omega) = \frac{1}{2\Pi} \int_{-\infty}^{+\infty} C(x)\ e^{i\omega x}\ dx$$

et que le processus (X_t) admet la représentation de Cramer :

$$X_t = \int_{-\infty}^{+\infty} e^{i\omega t}\ dZ(\omega)$$

où $Z(\omega)$ est un processus complexe à accroissements orthogonaux (processus spectral).

(Note : nous utilisons, pour simplifier les écritures, la notation complexe mais (X_t) étant réel, il est bien entendu que la décomposition de Cramer en sinus et cosinus est réelle).

L'analyse en composantes principales consisterait, en généralisant les résultats des paragraphes précédents, à rechercher les solutions de l'équation :

$$\int_{-\infty}^{+\infty} C(t-s)\ f(s)\ ds = \lambda\ f(t)\ .$$

Montrons que l'opérateur C de noyau C(t-s) est borné :

On a : $$(Cf|\overline{g}) = \int_{-\infty}^{+\infty} (C*f)(t)\ \overline{g(t)}\ dt$$

en désignant par C*f la convolution $\int_{-\infty}^{+\infty} C(t-s)\ f(s)\ ds$.

Les transformées de Fourier de C, f et g étant $2\Pi s$, φ et ψ le théorème de Plancherel donne :

$$(Cf|\overline{g}) = \int_{-\infty}^{+\infty} s(\omega)\ \varphi(\omega)\ \psi(\omega)\ d\omega\ .$$

$s(\omega)$ étant borné il vient :

$$(Cf|\overline{g}) \leqslant k_1 \int_{-\infty}^{+\infty} \varphi(\omega)\ \overline{\psi(\omega)}\ d\omega$$

et d'après l'inégalité de Cauchy-Schwarz :

$$(Cf|\overline{g}) \leqslant k_2\ \|\varphi\|\ \ \|\psi\|\ .$$

Comme $\|\varphi\| = \sqrt{2\Pi}\ \|f\|$ on a :

$$(Cf|\overline{g}) \leqslant C\ \ \|f\|\ .\ \|g\|$$ donc C est borné.

Cependant le noyau C(t-s) ne définit pas une fonction de carré intégrable sur $\mathbb{R} \times \mathbb{R}$ et par conséquent C n'est pas de Hilbert-Schmidt.

On pourrait utiliser l'analyse spectrale des opérateurs normaux bornés des espaces de Hilbert, cependant il est plus utile d'employer l'analyse

spectrale de Gelfand-Vilenkin [30] pour les opérateurs non bornés (mais valable aussi pour les bornés) car on a besoin de considérer des fonctions propres comme $e^{i\omega t}$ qui n'appartiennent pas à $L^2(\mathbb{R})$.

On a en effet la

<u>Proposition 1-6</u>

Les fonctions exponentielles imaginaires $e^{i\omega t}$ sont fonctions propres généralisées de C associées aux valeurs propres $\lambda_\omega = 2\Pi\ s(\omega)$

En effet on vérifie que :

$$\int_{-\infty}^{+\infty} C(t-s)\ e^{i\omega s}\ ds = \int_{-\infty}^{+\infty} C(x)\ e^{i\omega(t-x)}\ dx$$

$$= e^{i\omega t} \int_{-\infty}^{+\infty} C(x)\ e^{-i\omega x}\ dx = e^{i\omega t}\ 2\Pi\ s(\omega)$$

La transformée de Fourier inverse de la densité spectrale exprime donc la covariance à l'aide de la densité spectrale :

$$C(x) = \int_{-\infty}^{+\infty} s(\omega)\ e^{-i\omega x}\ d\omega$$

et on vérifie de même que la décomposition de Karhunen-Loève, désormais sous forme intégrale, s'identifie à la représentation de Cramer.

Les problèmes de convergence des analyses sur un intervalle [-T/2;T/2] lorsque T tend vers l'infini ont été résolus par GRENANDER et SZEGO [29] qui ont montré la convergence des solutions de l'équation :

$$\int_{-T/2}^{T/2} C(t-s)\ f(s)\ ds = \lambda\ f(t)$$

vers les fonctions trigonométriques $e^{i\omega t}$.

Ils ont aussi donné les résultats suivants concernant la répartition des valeurs propres de l'équation précédente.

Cette dernière admet une infinité dénombrable de valeurs propres car C est compact et il n'existe qu'un nombre fini de valeurs propres λ telles que $a \leqslant \lambda \leqslant b$ où $[a,b]$ est un intervalle ne contenant pas l'origine. Lorsque T tend vers l'infini ce nombre tend lui-même vers l'infini cependant le rapport

$$\frac{\text{nombre de } \lambda \in [a,b]}{T}$$

tend vers une limite proportionnelle à la mesure de Lebesgue de l'ensemble des ω tels que $[\frac{a}{2\Pi} < s(\omega) < \frac{b}{2\Pi}]$. Quant à la plus grande valeur propre, elle converge vers $\frac{1}{2\Pi} \sup_{\omega} s(\omega)$.

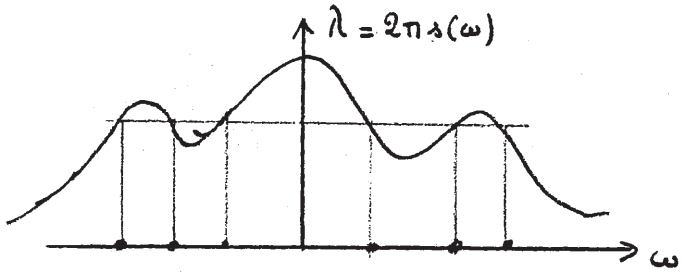
On peut alors, théoriquement du moins, définir la répartition des valeurs propres qui est une mesure absolument continue par rapport à la mesure de Lebesgue. Cette mesure que nous noterons $\mu(\lambda)$ permet de donner un sens à la somme des valeurs propres puisque l'on a :

$$\int_{-\infty}^{+\infty} s(\omega)\, d\omega = C(0) = \sigma^2 \text{ et que } \lambda = 2\Pi s(\omega)$$

d'où $$\int \frac{\lambda}{2\Pi}\, d\omega = \int \lambda\, d\mu(\lambda) = \sigma^2 .$$

La décomposition de la variance σ^2 de X_t selon les composantes principales est donc identique à la décomposition spectrale; on peut alors définir la part de variance expliquée par un intervalle $[\lambda_1\ \lambda_2]$.

Malheureusement $d\mu(\lambda)$ est très difficile à obtenir explicitement dès que la densité spectrale, ou plutôt sa restriction aux ω positifs n'est pas bijective car à une valeur propre λ correspondent plusieurs valeurs de ω :



Les résultats précédents généralisent des propriétés connues pour les séries chronologiques. Soit en effet un processus à temps discret (X_t) où $t \in \mathbb{Z}$. La restriction à un ensemble fini de T instants de sa fonction de covariance est représentable par la matrice T x T suivante :

$$C = \begin{pmatrix} C(0) & C(1) & \text{---} & C(T-1) \\ C(1) & C(0) & \text{--} & C(T-2) \\ \vdots & & & \vdots \\ C(T-1) & \text{--------} & & C(0) \end{pmatrix}$$

qui est une matrice de Toeplitz symétrique. On montre alors (voir [11] [29]) que les vecteurs propres de C sont, si T est grand, approximativement égaux aux sinus et cosinus discrets et les valeurs propres de C approximativement égales à 2Π fois la valeur de la densité spectrale aux points $\frac{2\Pi k}{T}$. Plus précisément soient $\underline{u}_k$ et $\underline{v}_k$ les vecteurs dont les composantes sont respectivement $\sqrt{\frac{2}{T}} \cos \frac{2\Pi jk}{T}$ et $\sqrt{\frac{2}{T}} \sin \frac{2\Pi jk}{T}$ pour $j=0,1,\ldots,T-1$; soit Q la matrice dont les vecteurs colonnes sont $\underline{u}_0$ et les $\underline{u}_k$ et $\underline{v}_k$ pour $k=1,2,\ldots,m$ (on suppose ici $T=2m+1$) alors ${}^tQ\ C\ Q \underset{m\to\infty}{\to} 2\Pi D$ où D est la matrice diagonale d'éléments $d_1 = s(0)$ $d_{2_k} = d_{2_{k+1}} = s(\frac{2\Pi k}{T})$ où s est la densité spectrale de (X_t).

I.3.c) Exemple : le processus de Markov

Soit un processus dont la fonction d'autocorrélation est :

$$C(t,s) = e^{-\alpha|t-s|}$$

On sait qu'une telle fonction caractérise les processus de Markov stationnaires ([22] p. 234).

Prolongé sur $]-\infty;+\infty[$ ce processus admet une densité spectrale égale à $\frac{\alpha}{\Pi(\alpha^2+\omega^2)}$.

Son analyse en composantes principales sur $[0,T]$ revient à résoudre l'équation :

$$\lambda\ f(t) = \int_0^T e^{-\alpha|t-s|}\ f(s)\ ds$$

le traitement est classique [53] et se ramène après deux dérivations à l'équation différentielle :

$$f''(t) = [\alpha^2 - \frac{2\alpha}{\lambda}]\ f(t)$$

et en posant $\alpha^2 - \frac{2\alpha}{\lambda} = -\omega^2$ (on montre que cette quantité est toujours négative) soit $\lambda = \frac{2\alpha}{\alpha^2+\omega^2}$ on trouve que les seules valeurs possibles pour ω sont les racines positives ω_1, ω_2, ... de l'équation $\operatorname{tg} \omega T = \frac{2\alpha\omega}{\omega^2-\alpha^2}$.

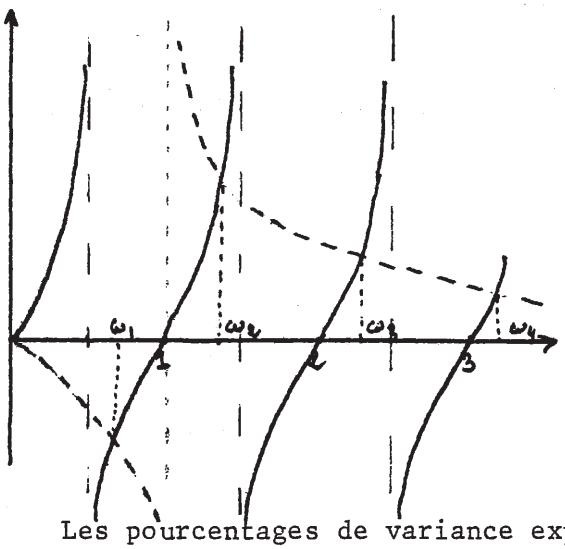
Les fonctions propres normalisées sont alors :

$$f_k(t) = \sqrt{\frac{2}{T+\lambda_k}} \quad \sin(\omega_k(t - \frac{T}{2})) .$$

Cependant ici les λ_k sont des valeurs propres simples.

Examinons rapidement les résultats numériques obtenus dans des cas simples.

Tout d'abord prenons $T = \Pi$ et $\alpha = 1$. Les racines ω_k sont les abscisses positives des points d'intersection de $y = \operatorname{tg} \Pi x$ avec $y = \frac{2x}{x^2-1}$.



$\omega_1 = 0.6383$	$\lambda_1 = 1,42$	45%
$\omega_2 = 1.3958$	$\lambda_2 = 0.678$	22%
$\omega_3 = 2.2647$	$\lambda_3 = 0.326$	10%
$\omega_4 = 3.1932$	$\lambda_4 = 0.179$	6%
$\omega_5 = 4.1505$	$\lambda_5 = 0.110$	3,5%

Les pourcentages de variance expliquée sont faibles car α est relativement grand et les variables X_t et X_s sont rapidement non corrélées dès que $|t-s|$ est grand.

Avec un coefficient α plus faible correspondant à une plus forte dépendance on trouve bien sûr des composantes de plus grande variance.

Aussi avec $\alpha = 0,1$ on trouve

$\omega_1 = 0.2459$	$\lambda_1 = 2.84$	90%
$\omega_2 = 1.0599$	$\lambda_2 = 0.18$	6%
$\omega_3 = 2.0313$	$\lambda_3 = 0.05$	2%
$\omega_4 = 3.0211$	$\lambda_4 = 0.02$	1%

Lorsque T augmente les ω_i se resserrent et deviennent de plus en plus denses ce qui traduit intuitivement les résultats énoncés au paragraphe précédent.

Pour $T = \infty$ on trouve alors que les solutions de l'équation :

$$\int_0^T e^{-\alpha|t-s|} f(s)\, ds = \lambda f(t)$$

sont $f(t) = \omega \cos \omega t + \alpha \sin \omega t$, soit à un facteur près :

$f(t) = \cos(\omega t - \varphi)$ avec $\operatorname{tg} \varphi = \frac{\alpha}{\omega}$

pour tout $\omega \in \mathbb{R}^+$. Les "valeurs propres" associées sont $\lambda_\omega = \frac{2\alpha}{\alpha^2+\omega^2}$ et sont toutes simples.

C'est seulement si l'on fait l'analyse factorielle sur $]-\infty;+\infty[$ que l'on retrouve exactement la décomposition spectrale : les valeurs propres sont alors doubles car elles correspondent aux deux valeurs opposées $\pm\omega$ et sont alors associées aux $\cos \omega t$ et $\sin \omega t$.

L'ensemble des valeurs propres est alors compris entre 0 et $\frac{2}{\alpha}$.

I.4 Cas d'une population finie

I.4.a) Généralités

Si Ω est un ensemble à n éléments muni de la mesure de probabilité uniforme, on est dans le cas statistique usuel où on a observé n trajectoires centrées, c'est-à-dire que :

$$m(t) = \frac{1}{n} \sum_{i=1}^{n} x_i(t) = 0 \qquad \forall\, t$$

La fonction de covariance C(t,s) est alors :

$$C(t,s) = \frac{1}{n} \sum_{i=1}^{n} x_i(t)\, x_i(s)$$

le noyau de C étant une somme de produits de fonctions de t et de s est

dégénéré et C est de rang fini ce qui entraîne l'existence de l'analyse en composantes principales de (X_t) sans hypothèses de régularité sur le processus.

L'opérateur d'Escoufier W est alors représentable par une matrice carrée n x n , puisque $L^2(\Omega)$ est isomorphe à $\mathbb{R}^n$, de terme général :

$$w_{k\ell} = \frac{1}{n}\int_0^T x_k(t)\ x_\ell(t)\ dt$$

On peut alors obtenir directement les composantes principales qui sont des vecteurs $\underline{\xi}$ à n composantes comme vecteurs propres de W et retrouver ensuite, du moins en théorie, les facteurs principaux comme combinaisons linéaires des trajectoires. En effet l'équation :

$$\int_0^T C(t,s)\ f(s)\ ds = \lambda\ f(t)$$

s'écrit :

$$\lambda\ f(t) = \int_0^T \frac{1}{n}\sum_{i=1}^{n} x_i(t)\ x_i(s)\ f(s)\ ds = \frac{1}{n}\sum_{i=1}^{n} x_i(t) \int_0^T x_i(s)\ f(s)\ ds$$

soit $\lambda\ f(t) = \frac{1}{n}\sum_{i=1}^{n} x_i(t)\ c_i$.

Les coefficients c_i s'obtiennent en résolvant le système de Cramer obtenu en multipliant par $x_j(t)$ les deux membres de la relation précédente et en intégrant sur t :

$$\int_0^T \lambda\ x_j(t)\ f(t)\ dt = \int_0^T \frac{1}{n}\sum_{i=1}^{n} x_i(t)\ x_j(t)\ c_i\ dt$$

soit
$$\lambda\,(\underline{\xi})_j = \sum_{i=1}^{n} \left(\int_0^T \frac{1}{n}\, x_i(t)\ x_j(t)\ dt \right) c_i = (W\underline{c})_j$$

d'où $\underline{c} = \underline{\xi}$ résultat classique en analyse des données.

Notons ici que le nombre de variables (infini) étant supérieur au nombre d'individus W est en général inversible comme l'opérateur C .

La méthode précédente n'est autre en fait que la traduction en termes statistiques de la méthode classique de résolution des équations de Fredholm à noyaux dégénérés ([35] p. 54).

Bien sûr, si n est très grand, cette méthode est impraticable même en supposant effectuée l'évaluation numérique des intégrales permettant le calcul de $w_{k\ell}$. On cherchera plutôt à discrétiser le temps afin d'approximer l'opérateur C par une matrice de taille p faible (voir paragraphe III).

I.4.b) Analyse de plusieurs courbes - dualité et centrage

Si on dispose d'un ensemble de n courbes $x_i(t)$ avec $i=1,2,\ldots,n$ et $t \in [0,T]$ donnant par exemple l'évolution des températures dans n stations ou celle des prix de n produits deux conceptions sont possibles pour effectuer une analyse factorielle :

soit chaque courbe est une trajectoire d'un processus et on se trouve dans cadre de l'analyse en composantes principales d'un processus échantillonné

soit on considère qu'on a une observation de n variables au cours du temps c'est-à-dire une série chronologique multiple. Ce point de vue est alors justifié si les courbes représentent des évolutions de caractères différents. On est alors conduit à une A.C.P. sur n variables, la population étant l'ensemble des instants.

Ces deux types d'analyse ne sont pas conceptuellement identiques : la première cherche des facteurs de différenciation entre courbes en fonction de l'évolution qu'elles traduisent c'est donc une analyse des distances entre courbes, tandis que la deuxième A.C.P. repose plutôt sur les corrélations et est donc une étude des angles entre fonctions $x_i(t)$ de $L^2(T)$.

Il y a évidemment "dualité" entre les deux types d'analyse qui sont d'ailleurs identiques si on ne centre pas les données et si on utilise la métrique I dans les deux cas ainsi que le montre le tableau suivant :

	Courbes "individus" analyse "harmonique"	Courbes "variables" A.C.P.
Fonction de covariance	C opérateur intégral de noyau $C(t,s) = \frac{1}{n} \sum_{i=1}^{n} x_i(t)x_i(s)$	V matrice n x n de terme général $v_{ij} = \frac{1}{T} \int_0^T x_i(t)x_j(t)dt$
Opérateur d'Escoufier	U matrice n x n de terme général $\frac{1}{n} \int_0^T x_i(t)x_j(t)dt$ $U = \frac{T}{n} V$	W opérateur intégral de noyau $\frac{1}{T} \sum_{i=1}^{n} x_i(t)x_i(s)$ $W = \frac{n}{T} C$

aux coefficients T et n près l'opérateur d'Escoufier d'une analyse est l'opérateur de covariance de l'autre et les facteurs principaux d'une analyse sont les composantes principales de l'autre.

Bien entendu dès qu'on centre les données (et a fortiori avec des métriques différentes de I) il n'y a plus identité.

Les deux types de centrages

a) centrage "harmonique"

A tout instant t les n observations doivent être de moyenne nulle ce qui signifie qu'on étudie en réalité les n courbes :

$x_i(t) - \frac{1}{n} \sum_{i=1}^{n} x_j(t) = x_i(t) - m(t)$ ce qui revient d'une certaine façon à éliminer la tendance.

L'opérateur de covariance centré a alors pour noyau :

$$C(t,s) = \frac{1}{n} \sum_{i=1}^{n} (x_i(t) - m(t))(x_i(s) - m(s))$$

et l'opérateur d'Escoufier est la matrice W de terme général :

$$w_{ij} = \frac{1}{n} \int_0^T (x_i(t) - m(t))(x_j(t) - m(t))\, dt$$

b) Centrage A.C.P.

Les courbes étant considérées comme des variables on retire à chacune sa valeur moyenne sur la période c'est-à-dire $m_i = \frac{1}{T} \int_0^T x_i(t)\, dt$.

La matrice de covariance V a alors pour terme général :

$$v_{ij} = \int_0^T (x_i(t) - m_i)(x_j(t) - m_j)\, dt .$$

Il est clair que cette opération n'élimine pas en général la tendance et que dans bien des cas on trouvera un premier facteur trivial lié à la chronologie si les courbes varient de manière semblable.

Les deux types de centrage ne seront équivalents que si les moyennes temporelles sont égales aux moyennes d'ensemble ce qui est le cas si on a observé n trajectoires d'un processus stationnaire et ergodique.

II. ANALYSE FACTORIELLE AVEC UNE METRIQUE

Décomposer l'opérateur de covariance d'un processus suppose implicitement l'usage du produit scalaire de $L^2(T)$ c'est-à-dire de la métrique identité. Ce point de vue est naturel dans un contexte probabiliste où les observations n'ont pas d'intérêt en elles-même. En analyse des données au contraire, les individus jouent un rôle essentiel et il importe de définir une notion de distance entre eux qui tienne compte des idées de l'utilisateur et de la nature du problème traité. Généraliser à des processus la notion de métrique signifie donner une formule de distance entre trajectoires.

II.1 Définition et premières propriétés

Soit M un opérateur linéaire symétrique de $L^2(T)$ qu'on suppose strictement positif c'est-à-dire que :

$$\exists\, C > 0 \qquad (Mu|u) \geqslant C\, \|u\|^2$$

Remarquons que cette condition entraîne l'injectivité de M car d'après l'inégalité de Cauchy-Schwarz on a :

$$(M\,u|u) \leqslant \|u\|\,\|Mu\|$$

d'où $$\|Mu\| \geqslant c\|u\|$$

On notera Dom M le domaine de M, c'est-à-dire le sous-espace dense $L^2(T)$ où il est défini. On notera $\|\,.\,\|_M$ la norme sur Dom M associée au produit scalaire :

$$(f|g)_M = (M\,f\,|\,g)$$

donc $$\|f\|_M = (M\,f\,|\,f)^{1/2}$$

Plus généralement on peut considérer un produit scalaire sur un sous-espace dense de $L^2(T)$ qui ne coïncide pas avec le produit scalaire induit par $L^2(T)$.

Le cas particulier où M est continu, défini sur tout l'espace et surjectif, donc bicontinu par le théorème de Banach, est examiné dans leur thèse par DAUXOIS et POUSSE [16]. Ces conditions sont automatiquement réalisées en dimension finie mais sont trop restrictives si l'ensemble T est infini. Le cas général conduit à des problèmes spécifiques de dimension infinie et en particulier à un changement de structure hilbertienne en A.C.P. Le premier cas d'A.C.P. d'un processus avec une métrique différente de celle de $L^2(T)$ a été proposé par KREE et MAZEROLLE dans [37] en utilisant le relèvement par spline et des espaces de Sobolev.

Dans [7] on trouve également une utilisation des métriques de Sobolev en vue de l'approximation d'une A.C.P.. Nous présentons ici un formalisme général valable pour un opérateur M non forcément borné.

On notera H_M le complété de Dom M et H_{-M} son dual. D'après la stricte positivité de M on a $\|u\|_M \geqslant C^{1/2}\,\|u\|$: l'application canonique j de Dom M dans $L^2(T)$ est alors continue. Si on note $\hat{j}$ le prolongement continu de j de H_M dans $L^2(T)$ on peut alors montrer (Lemme de Friedrichs) que $\hat{j}$ est injectif. Soit $\hat{j}'$ son transposé : $L^2(T) \xrightarrow{\hat{j}'} H_{-M}$.

Dans ce cas on a pour x et y $\in H_M$:

$$(x|y)_M = (Mx|\hat{j}y) = (\hat{j}'Mx|y)$$

ce qui montre que j'M se prolonge en une isométrie $\hat{\hat{M}}$ de H_M sur son dual H_{-M} . On peut alors définir un opérateur autoadjoint $\hat{M}$ de $L^2(T)$ qui coïncide avec M sur Dom M en posant :

$$\text{Dom } \hat{M} = \{x \in H_M \mid \hat{\hat{M}}\, x \in L^2(T)\}$$

$$\hat{M}x = \hat{\hat{M}}x \ .$$

On définit alors, en utilisant la théorie spectrale, la racine carrée $\hat{M}^{1/2}$ de cet opérateur dont le domaine est H_M et qui est une isométrie de H_M sur $L^2(T)$ car on a alors :

$$(\hat{M}x \mid y) = (\hat{M}^{1/2}\, x \mid \hat{M}^{1/2} y) \ .$$

<u>Remarque importante</u> :

Notons ici une différence fondamentale avec le cas fini de l'analyse des données usuelle :

on ne peut utiliser n'importe quelle métrique pour un processus donné (X_t) ; en effet pour pouvoir utiliser une métrique M il faut que toutes les trajectoires de (X_t) appartiennent à H_M , or H_M peut être plus petit que $L^2(T)$ si la métrique M est plus forte que celle de $L^2(T)$ (ce sera le cas pour les métrique de Sobolev en particulier). Il faut donc s'assurer de conditions de régularité sur (X_t) avant d'utiliser une métrique M (par exemple que les dérivées des trajectoires appartiennent aussi à $L^2(T)$ pour les métriques de Sobolev).

Sous réserve que les trajectoires appartiennent à H_M on peut alors généraliser les résultats de l'A.C.P. de la manière suivante :

La M-norme de (X_t) est la variable aléatoire définie par

$$\| X_t \|_M^2 = \int_0^T X_t\, \hat{M}\, X_t \, dt = (X_t \mid \hat{M}\, X_t)$$

ce qui permet de définir l'inertie totale J de (X_t) comme étant l'espérance de cette variable.

$$J = E\,[\,\|X_t\|_M^2\,]$$

On en déduit alors

> <u>Proposition 1-7</u>
>
> L'inertie totale de (X_t) vaut trace $\hat{M}C$.

En effet $J = E(\hat{M}^{1/2}\,X_t \mid \hat{M}^{1/2}\,X_t)$ et est donc égale à la variance totale du processus transformé $\hat{M}^{1/2}(X_t)$ dont l'opérateur de covariance est $\hat{M}^{1/2}\,C\,\hat{M}^{1/2}$.

Cet opérateur est nucléaire car C l'est et $\hat{M}^{1/2}$ est continu comme isométrie de H_M sur $L^2(T)$.

D'où $J = \text{Trace}\ \hat{M}^{1/2}\,C\,\hat{M}^{1/2} = \text{Trace}\ \hat{M}C = \text{Trace}\ C\hat{M}$.

Comme en analyse de données usuelles l'introduction d'une métrique différente de l'identité entraîne une distinction entre axes et facteurs principaux.

Nous définirons un axe comme un élément normé de H_M c'est-à-dire une fonction du temps a(t) telle que $(a|\hat{M}a) = 1$. La M-projection de (X_t) sur cet axe est alors un processus quasi-déterministe du type $\xi\,a(t)$ où

$$\xi = (X_t|\hat{M}a) = \int_0^T X_t\ \hat{M}a(t)\,dt = \int_0^T X_t\ f(t)\,dt$$

en posant $f = \hat{M}a$.

Rechercher les axes principaux de (X_t) pour la métrique M revient alors à maximiser l'inertie totale du processus $\xi\,a(t)$ qui vaut $E(\xi^2)$ puisque a est normé dans H_M .

En appelant axes principaux d'inertie les éléments de la boule unité de H_M rendant extrémale $E(\xi^2)$, valeurs principales les extrema, facteurs principaux, les transformés par $\hat{M}$ des axes principaux, composantes principales les variables aléatoires ξ correspondantes on a alors :

Proposition 1-8

Les axes principaux, les facteurs principaux et les composantes principales sont fonctions propres respectivement des opérateurs suivants :

$$C\hat{M}\ ,\quad \hat{M}C\ ,\quad W = U\hat{M}U^*$$

Vu les hypothèses faites sur le processus et les propriétés des opérateurs C et $\hat{M}$, la démonstration de cette proposition ne diffère pas de celle donnée en dimension finie (cf. [12] par exemple).

En effet en supposant que les trajectoires de (X_t) appartiennent à H_M les opérateurs C et $\hat{M}$ sont des opérateurs nucléaires définis sur H_{-M} et H_M respectivement (si $H_M \subset L^2(T)$ la situation est évidente; si $H_M \supset L^2(T)$ comme on suppose que les trajectoires $\in L^2(T)$ il n'y a pas non plus de difficulté).

Remarquons que les facteurs principaux f sont dans l'image de $\hat{M}$ puisque par définition $f = \hat{M}a$.

En introduisant l'opérateur $\hat{M}^{-1}$ de H_{-M} sur H_M on a de plus :

$$(a|\hat{M}a) = 1 = (a|f) = (\hat{M}^{-1} f|f)$$

les facteurs principaux sont donc de norme 1 dans H_{-M} .

La formule de reconstitution ou décomposition de Karhunen-Loève s'écrit alors :

$$X_t = \sum_{k=1}^{\infty} \xi_k \hat{M}^{-1} f_k(t) \quad .$$

II.2 Exemples de métriques

Destinées à comparer des fonctions selon leur degré de régularité, les métriques de Sobolev semblent un outil approprié pour calculer des distances entre trajectoires. Elles donnent un exemple de métriques plus fortes que

celle de $L^2(T)$. Par ailleurs les pratiques usuelles de filtrage et lissage des séries chronologiques avant leur traitement, conduisent implicitement à utiliser des métriques sur les données brutes. Dans certains cas M pourra s'expliciter sous forme d'un opérateur intégral (métrique plus faible que celle de $L^2(T)$).

II.2.a) Métriques de Sobolev

Commençons par une métrique définie par l'opérateur autoadjoint $M = I - D^2$ (où D est l'opérateur de dérivation) défini sur

$$\text{Dom } M = C_0^\infty(T)$$

espace des fonctions indéfiniment différentiables sur T et nulles aux bornes.

Une intégration par parties donne pour tout $x \in C_0^\infty(T)$

$$\| x \|_M^2 = (M x \mid x) = \int_0^T (x^2(t) + (\frac{dx(t)}{dt})^2)\, dt$$

ce qui entraîne la stricte positivité de M puisque

$$(M x \mid x) \geqslant \int_0^T (x(t))^2\, dt = \| x \|^2$$

le complété de Dom M est l'espace de Sobolev $H_0^1(T)$.

Utiliser $M = I - D^2$ suppose donc implicitement que le processus (X_t) est nul aux bornes et que les dérivées des trajectoires appartiennent à $L^2(T)$.

Si le processus n'est pas nul aux bornes l'utilisation du produit scalaire précédent

$$(x|y)^1 = \int_0^T (x(t)\, y(t) + \frac{dx(t)}{dt}\, \frac{dy(t)}{dt})\, dt$$

ne permet pas d'expliciter un opérateur M^1 de domaine $C^\infty(T)$ tel que

$(x,y)^1 = (M^1 x\, ,\, y)$ car l'intégration par parties fait alors

intervenir des termes aux bornes. On peut cependant compléter $C^{\infty}(T)$ et l'on obtient un espace $H^1(T)$ contenant $H_0^1(T)$.

Plus généralement on peut poser :

$$(x|y)_k = \int_0^T \sum_{i=0}^{k} (D^i x \mid D^i y)\, dt$$

et le complété de $C^{\infty}(T)$ est l'espace de Sobolev $H^k(T)$. Il est également possible d'introduire des coefficients $\alpha_i > 0$ afin de pondérer les dérivées.

L'utilisation de telles métriques ne se conçoit donc que pour des processus suffisamment réguliers.

II.2.b) Le cas de transformations linéaires

Nous avons vu qu'utiliser une métrique M était équivalent à faire une A.C.P. sur le processus transformé $M^{1/2} X_t$ si celui-ci est compatible avec M. Considérons inversement le cas de l'analyse d'un processus transformé.

Soit L un opérateur linéaire de $L^2(T)$ dans un espace de Hilbert K , qui peut être $L^2(T)$ lui-même, dont le domaine Dom L est un sous-espace dense de $L^2(T)$.

Rappelons que formellement le transposé L' de L est l'opérateur de K dans $L^2(T)$ défini par :

$$(L' x|y) = (x \mid L y) .$$

Plus précisément Dom L' est l'ensemble des $x \in K$ tels que la forme linéaire $y \to (x \mid L y)$ est continue pour la topologie induite par $L^2(T)$.

Cette forme linéaire étant notée L'k, si on suppose $\mathrm{Im}\, L \subset \mathrm{Dom}\, L'$, on a alors le droit de parler du composé $M = L'L$ qui est autoadjoint et on a évidemment :

$$(M x \mid x) = (L' L x, x) = \| L x \|_K^2$$

Si il existe C tel que $\| L x \|_K^2 \geqslant C \| x \|^2$ on peut introduire l'espace H_M défini au paragraphe II-1.

L'équivalence entre l'A.C.P. de (X_t) avec la métrique $M = L'L$ et l'A.C.P. de $L(X_t)$ dans K est précisée par le résultat suivant dont la démonstration n'offre pas de particularités par rapport à celle du cas fini dès que (X_t) a ses trajectoires dans H_M . (Voir [12] par exemple pour une démonstration).

Proposition 1-9

Les composantes principales de (X_t) avec la métrique $M = L'L$ sont identiques aux composantes principales de $L(X_t)$ dans K. Les M-facteurs principaux de (X_t) sont les transformées par L' des facteurs principaux de $L(X_t)$, les M-axes principaux de (X_t) sont les transformées par L^{-1} des axes principaux de $L(X_t)$.

La décomposition de M en L'L n'étant pas nécessairement unique on voit qu'il peut exister plusieurs transformations linéaires L en apparence différentes mais conduisant à des analyses équivalentes c'est-à-dire au même espace H_M .

Exemple

Considérons l'application L de domaine $C_0^\infty(T)$ dans $L^2(T)$ définie par $x \to x + x'$, c'est-à-dire $L = I + D$.

Le domaine de L' contient $C_0^\infty(T)$, et sur ce domaine :

$$L'y = y - y' \qquad \text{d'où} \qquad L' = I - D .$$

On a alors $L'Lx = x^2 + (x')^2$ $\quad M = L'L = I - D^2$ et on retrouve l'exemple de l'analyse dans $H_0^1(T)$.

En se référant à la théorie du signal P. BESSE [7] appelle ce type d'analyse "A.C.P. filtrante" car L peut être assimilé à un filtre linéaire à paramétres non constants.

Remarquons à ce sujet que si (X_t) est stationnaire $(Y_t) = L(X_t)$ ne l'est pas en général et que l'on ne retrouvera pas les résultats établis en I.3 sauf si L est une transformation stationnaire (filtre à paramètres constants) :

$$L(x(t-t_0)) = g(t-t_0) \qquad \forall\, t_0 \ .$$

Ce qui est le cas si L est une convolution définie par

$$y(t) = \int_{-\infty}^{+\infty} \ell(t-s)\ x(s)\ ds \ .$$

Les résultats classiques de filtrage linéaire([52] par exemple)permettent d'affirmer que les facteurs principaux de $L(X_t)$ sont toujours les $e^{i\omega t}$, l'effet de filtrage ne modifiant que les valeurs propres λ_ω qui sont alors multipliés par le carré du module de la valeur en ω de la transformée de Fourier du filtre (fonction de transfert) :

$$\lambda_\omega \to |\Phi(i\omega)|^2\ \lambda_\omega$$

$$\text{où} \quad \Phi(i\omega) = \int_{-\infty}^{+\infty} \ell(t)\ e^{-i\omega t}\ dt$$

Parmi les filtres élémentaires utilisés en traitement de séries chronologiques mentionnons :

. le filtre de moyenne mobile qui revient à remplacer une observation x(t) par la moyenne dans une fenêtre de longueur h :

$$y(t) = \frac{1}{n} \int_{-h/2}^{h/2} x(t)\ dt$$

. le filtre des différences qui revient à analyser les taux d'accroissement entre t-h et t

$$y(t) = \frac{x(t) - x(t-h)}{h}$$

. le filtre du lissage exponentiel

$$y(t) = \int_{-\infty}^{t} \alpha^{t-s}\ x(s)\ ds$$

Ces deux derniers filtres étant des filtres "réalisables" ou "causaux" car ils ne font intervenir que les observations antérieures à t .

Remarquons enfin que l'usage d'une mesure positive $d\mu(t)$ sur le temps, absolument continue par rapport à la mesure de Lebesgue (afin de donner plus d'importance aux instants récents par exemple) revient à choisir une métrique "diagonale" puisqu'il suffit de poser $Y_t = X_t \frac{d\mu(t)}{dt}$ pour se ramener au cas usuel. L'opérateur M est alors tel que

$$M\,x(t) = \left(\frac{d\mu(t)}{dt}\right)^2 x(t) \ .$$

II.2.c) Métriques définies par un opérateur intégral

Un cas très particulier est celui où l'on peut exprimer le produit scalaire entre deux trajectoires par :

$$(x|y)_M = \iint y(t)\ M(t,s)\ x(s)\ dt\ ds$$

où $M(t,s)$ est une fonction de type positif, symétrique, de carré intégrable.

L'opérateur linéaire continu M s'explicite alors par :

$$M\,x(t) = \int_0^T M(t,s)\ x(s)\ ds \ .$$

On a une injection de $L^2(T)$ dans son complété H_M pour le produit scalaire $(x,y)_M$ et on peut avoir $H_M \supset L^2(T)$.

Les conditions précédentes impliquent que M est de Hilbert-Schmidt et on trouve alors l'analogue de l'analyse des données en dimension finie sans difficulté particulière puisque MC et CM sont alors nucléaires.

Ce cas ne permet pas de traiter des transformations courantes comme la dérivation et l'écriture :

$$y_t = \frac{dX_t}{dt} = L\ X_t = -\int_0^T \delta'(t-s)\ X_s\ ds$$

que l'on trouve dans la littérature technique est incorrecte puisque $\delta'(t-s)$, dérivée de la distribution de Dirac, n'est pas une fonction de $L^2(T)$.

A nouveau, ceci prouve que $M = L'L$ n'est pas un opérateur de $L^2(T)$ et ne peut donc être appliqué à n'importe quel processus.

II.3 La métrique D_{1/σ^2}

Comme en analyse des données usuelles, la métrique associée à la réduction des données joue un rôle important pour les applications.

Cette métrique est "diagonale" puisque le processus réduit est $(Y_t) = \dfrac{X_t}{\sigma(t)}$; nous lui conserverons donc sa notation classique $M = D_{1/\sigma^2}$.

On supposera nécessairement que les trajectoires de (Y_t) appartiennent à $L^2(T)$ ce qui implique en particulier que (X_t) ne puisse être constant sur des intervalles de mesure non nulle.

Dans ces conditions, au lieu de décomposer l'opérateur de covariance, on décomposera l'opérateur de corrélation R de noyau $R(t,s) = \dfrac{C(t,s)}{\sigma(t)\sigma(s)}$ qui est de trace finie égale à T.

II.3.a) Propriétés essentielles

Les composantes principales ξ sont les fonctions propres de l'opérateur W associé au processus et à la métrique M qui s'identifie à l'opérateur W associé au processus réduit. On a donc :

$$W\xi = \int_0^T \frac{X_t \, E(X_t \, \xi)}{\sigma^2(t)} \, dt \quad .$$

Les extrema de $E(\xi \, W \xi)$ sur la boule unité de $L^2(\Omega)$ sont atteints pour ξ fonctions propres de W et sont égaux aux valeurs propres de W car W est compact et autoadjoint.

$$\text{Or} \quad E(\xi W \xi) = \int_0^T \frac{(E(X_t \, \xi))^2}{\sigma^2(t)} \, dt = \int_0^T r^2(X_t;\xi) \, dt$$

où r est le coefficient de corrélation linéaire.

On a donc montré la proposition suivante :

Proposition 1-10

Les composantes principales ξ de l'analyse en composantes principales de (X_t) réduit sont les variables combinaisons linéaires des X_t rendant extrêmales la quantité

$$\int_0^T r^2(X_t \; ; \; \xi) \, dt$$

Remarquons au passage que ceci établit que ξ est nécessairement du type $\int_0^T X_t \, f(t) \, dt$.

L'opérateur W possède dans le cas de la métrique D_{1/σ^2} une interprétation intéressante : on peut noter en effet que $\dfrac{X_t \, E(X_t \, \xi)}{\sigma^2(t)}$ est la projection orthogonale de ξ sur X_t dans $L^2(\Omega)$ pour ξ centrée.

W est donc la somme (intégrale) des opérateurs élémentaires de projection sur les X_t pour $t \in [0,T]$.

Cette propriété, en raison de l'intérêt intrinsèque des projecteurs, nous semble être un argument en faveur de l'utilisation de cette métrique d'autant plus que l'homogénéisation des unités de mesure des variables n'est ici d'aucun intérêt, contrairement à l'A.C.P. usuelle, puisque les X_t sont des variables de même nature.

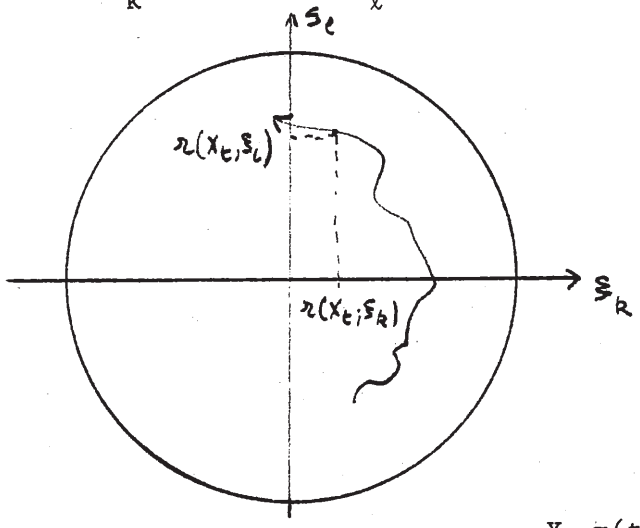
Un autre avantage de la métrique D_{1/σ^2} est le suivant : pour les processus à variance croissante (du type processus à accroissements orthogonaux) analyser les données brutes revient implicitement à donner plus de poids aux derniers instants de l'intervalle $[0,T]$ puisque ce sont les variables à plus forte variance qui contribuent le plus à la détermination des facteurs : il est donc prévisible de trouver une première composante relativement triviale liée à la valeur finale du processus X_T..

La normalisation des données par l'usage de la métrique D_{1/σ^2} doit alors permettre d'éliminer ce facteur trivial comme le montre l'exemple du paragra-

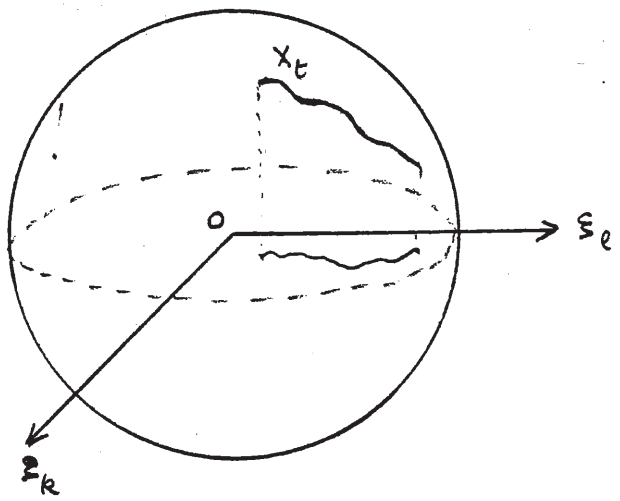
phe suivant. Pour des données réelles où $X_0 \equiv 0$ il faudra cependant être prudent car $\sigma(0) = 0$ et on risque d'avoir des résultats instables dus à des comportements d'individus particuliers pendant les premiers instants.

Comme en A.C.P. usuelle le cercle des corrélations est susceptible d'une interprétation intéressante.

Rappelons que l'on appelle cercle des corrélations associé aux composantes principales ξ_k et ξ_ℓ la figure représentant les variables initiales X_t par des points de coordonnées $r(X_t\ ;\ \xi_k)$ et $r(X_t\ ;\ \xi_\ell)$ dans un repère orthonormé. Dans le cas d'un processus (X_t) on obtient alors une courbe paramétrée d'équation $\sqrt{\lambda_k}\ f^k(t)$ et $\sqrt{\lambda_\ell}\ f^{(\ell)}(t)$.



Si on considère le processus centré réduit $\dfrac{X_t - m(t)}{\sigma(t)}$ comme une courbe de $L^2(\Omega)$ paramètrée par le temps ; cette courbe est située sur la boule de rayon 1, le cercle des corrélations est alors la projection du processus (X_t) sur le sous-espace de dimension 2 engendré par les variables aléatoires ξ_k et ξ_ℓ .



II.3.b) Etude d'un processus particulier

Considérons un processus dont la fonction de covariance est telle que :

$$C(t,s) = c \inf(t,s)$$

Cette fonction est caractéristique des processus à accroissements stationnaires non corrélés c'est-à-dire tels que $E((X_t-X_s)^2)$ dépend seulement de t-s et $E((X_t-X_s)(X_u-X_v)) = E(X_t-X_s)\,E(X_u-X_v)$ si $t < s \leqslant u < v$.

Deux cas particuliers importants de tels processus sont :

le processus de Poisson

le mouvement brownien.

L'analyse en composantes principales de ces processus se ramène à la résolution de l'équation intégrale

$$c \int_0^T \inf(t,s)\, f(s) = \lambda\, f(t)$$

et est classique [67]. Après deux dérivations par rapport à t on se ramène à l'équation différentielle :

$$\lambda\, f''(t) + c\, f(t) = 0$$

et on montre sans difficulté que les valeurs propres λ_k sont telles que

$$\lambda_k = \frac{4\, c\, T^2}{(2k-1)^2\, \Pi^2} \qquad k=1,2,\ldots,\infty$$

et les fonctions propres normalisées (facteurs principaux)

$$f_k(t) = \sqrt{\frac{2}{T}} \sin\left[\left(k - \frac{1}{2}\right)\Pi\, \frac{t}{T}\right]$$

On retrouve bien $\sum_{k=1}^{\infty} \lambda_k = \frac{cT^2}{2}$ puisque $\sum_{k=1}^{\infty} \frac{1}{(2k-1)^2} = \frac{\Pi^2}{8}$

Les parts de variance expliquées par les premiers facteurs sont

k = 1	81,1%
k = 2	9,0%
k = 3	3,2%
k = 4	1,7%

Tout processus à accroissements stationnaires non corrélés sur [0,T] peut donc être résumé à plus de 80% par un processus quasi-déterministe de la forme

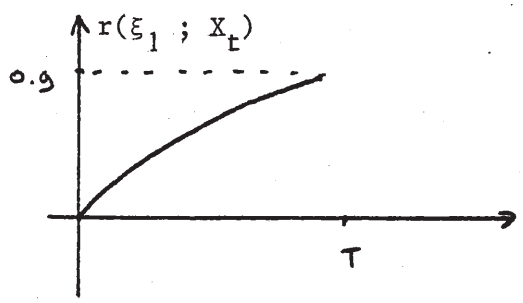
$$X_t \simeq ct + \xi_1 \sqrt{\frac{2}{T}} \sin \frac{\Pi t}{2T}$$

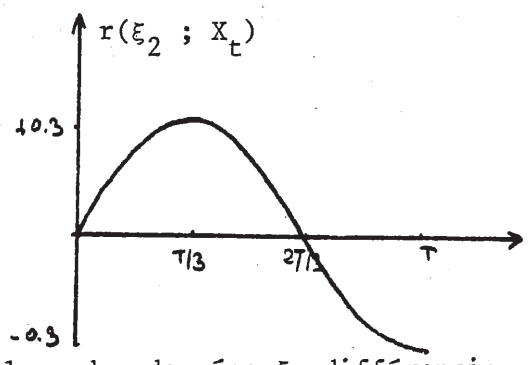
où ξ_1 est donc une v.a.r. centrée de variance $\frac{4cT^2}{\Pi^2}$.

Les coefficients de corrélation entre les X_t et les ξ_k sont donnés par :

$$r(\xi_k \; ; \; X_t) = \frac{2\sqrt{2}\sin[(k - \frac{1}{2})\,\Pi\,\frac{t}{T}]}{(2k-1)\,\Pi\,\sqrt{t}} \quad .$$

Les graphes des fonctions $r(\xi_1 \; ; \; X_t)$ et $r(\xi_r \; ; \; X_t)$ fournissent aisément l'interprétation des deux premières composantes principales :





ξ_1 est très liée à X_T : en termes d'analyse des données ξ_1 différencie les trajectoires selon leur valeur finale tandis que ξ_2 corrélé positivement à $X_{T/3}$ et négativement à X_T s'interprète comme une variable d'avance ou de retard : à ξ_1 fixé, donc à X_T fixé, ξ_2 différencie les trajectoires pour lesquelles X_t est élevé assez tôt dans le temps ($\xi_2 > 0$) de celles où le "démarrage" est plus tardif.

Le fait que la quasi-totalité de l'information soit contenue dans ξ_1 c'est-à-dire dans X_T est du au fait que (X_t) est un processus à variance croissante ; un tel résultat, assez évident a priori se retrouve pour toute une gamme de processus concrets comme celui du nombre d'enfants par famille [17]. Il semble donc utile de pouvoir éliminer cette première composante triviale : un moyen simple comme nous allons le voir consiste à utiliser la métrique D_{1/σ^2} c'est-à-dire à réduire X_t pour tout t .

L'opérateur MC est alors un opérateur intégral de noyau $\frac{\inf(t,s)}{t}$ puisque $\sigma^2(t) = c\,t$.

L'équation intégrale donnant les facteurs principaux est alors

$$\lambda\, f(t) = \int_0^t \frac{s}{t} f(s)\, ds + \int_t^T f(s)\, ds$$

En supposant comme d'habitude les solutions deux fois dérivables on se ramène aisément à une équation différentielle du second ordre.

On a en effet $\quad \lambda\, f'(t) = -\frac{1}{t^2} \int_0^t s\, f(s)\, ds$

et $\quad \lambda\, f''(t) = \frac{2}{t^3} \int_0^t s\, f(s)\, ds - \frac{1}{t} f(t)$

soit
$$\boxed{t\, f'' + 2f' + \frac{f}{\lambda} = 0}$$

Le changement de fonction inconnue $f(t) = \frac{g(t)}{\sqrt{t}}$ permet de se ramener à une équation classique (ce changement revient à chercher les fonctions propres g(t) de l'opérateur de corrélation de noyau $\frac{\inf(t,s)}{\sqrt{t,s}}$ ce qui est un cas particulier des noyaux $\frac{(\inf(t,s))^p}{(t,s)q}$ étudiés par J.C. DEVILLE [18], et on trouve après des calculs élémentaires :

$$\boxed{t^2\, g'' + t\, g' + g\left[\frac{t}{\lambda} - \frac{1}{4}\right] = 0}$$

Cette équation est un cas particulier de l'équation de Lommel (voir la formule 9.1.53 p. 362 de [1]) la solution en est :

$$g(t) = \mathcal{C}_1\left(\frac{2}{\sqrt{\lambda}}\sqrt{t}\right)$$

où $\mathcal{C}_1$ est l'une ou l'autre des deux fonctions de Bessel d'ordre 1 : J_1 ou Y_1. Comme Y_1 tend vers $-\infty$ quand $t \to 0$ seule J_1 peut convenir donc $g(t) = J_1(2\sqrt{\frac{t}{\lambda}})$. Il reste à trouver les valeurs possibles pour λ. Reportons la solution trouvée dans l'équation intégrale en f(t) :

$$\frac{\lambda\, J_1(2\sqrt{\frac{t}{\lambda}})}{\sqrt{t}} = \frac{1}{t}\int_0^t \sqrt{s}\, J_1(2\sqrt{\frac{s}{\lambda}})\, ds + \int_t^T \frac{1}{\sqrt{s}}\, J_1(2\sqrt{\frac{s}{\lambda}})\, ds$$

qui doit devenir une identité en t. En particulier pour t = T on doit donc avoir :

$$\frac{\lambda\, J_1(2\sqrt{\frac{T}{\lambda}})}{\sqrt{T}} = \frac{1}{T}\int_0^T \sqrt{s}\, J_1(2\sqrt{\frac{s}{\lambda}})\, ds$$

en posant $2\sqrt{\frac{s}{\lambda}} = u$ sous l'intégrale il vient :

$$\lambda\sqrt{T}\, J_1(2\sqrt{\frac{T}{\lambda}}) = \frac{\lambda^{3/2}}{4}\int_0^{2\sqrt{\frac{T}{\lambda}}} u^2\, J_1(u)\, du$$

Or (formule 11.9.20 de [1]) on a :

$$\int_0^z u^\nu\, J_{\nu-1}(u)\, du = z^\nu\, J_\nu(z)$$

ce qui donne :

$$\lambda\sqrt{T}\, J_1(2\sqrt{\frac{T}{\lambda}}) = \frac{\lambda^{3/2}}{4}(2\sqrt{\frac{T}{\lambda}})^2\, J_2(2\sqrt{\frac{T}{\lambda}})$$

soit $$J_2(2\sqrt{\frac{T}{\lambda}}) = \sqrt{\frac{\lambda}{T}}\, J_1(2\sqrt{\frac{T}{\lambda}}).$$

La formule de récurrence entre fonctions de Bessel :

$$J_{\nu-1}(z) + J_{\nu+1}(z) = \frac{2\nu}{z}\, J_\nu(z)$$

permet d'écrire $J_2(z) = \frac{2}{z}\, J_1(z) - J_0(z)$ et il vient

$$\frac{2}{2\sqrt{\frac{T}{\lambda}}}\, J_1(2\sqrt{\frac{T}{\lambda}}) - J_0(2\sqrt{\frac{T}{\lambda}}) = \sqrt{\frac{\lambda}{T}}\, J_1(2\sqrt{\frac{T}{\lambda}})$$

soit $$J_0(2\sqrt{\frac{T}{\lambda}}) = 0 .$$

Les valeurs propres sont donc proportionnelles aux inverses des carrés des zéros de la fonction de Bessel J_0 :

$$\boxed{\lambda_k = \frac{4T}{(j_{0,k})^2}}$$

Pour achever la détermination des facteurs il reste à normaliser les fonctions trouvées ; on doit avoir :

$$\int_0^T t\,[f(t)]^2\,dt = 1$$

soit $$\int_0^T [g(t)]^2\,dt = 1 .$$

Il faut évaluer l'intégrale :

$$\int_0^T [J_1(2\sqrt{\frac{t}{\lambda}})]^2\,dt$$ où λ vérifie la relation précédente.

Posons $t = Tu^2$, l'intégrale cherchée s'écrit alors :

$$2T\int_0^1 u\,[J_1(2\sqrt{\frac{T}{\lambda}}\,u)]^2\,du$$

Comme les zéros de J_0 sont aussi ceux de $z\,J_1'(z) + J_1(z)$ on peut appliquer la formule 11.4.5 de [1] et il vient

$$\int_0^1 u\,[J_1(2\sqrt{\frac{T}{\lambda}}\,u)]^2\,du = \frac{1}{2}[J_1(2\sqrt{\frac{T}{\lambda}})]^2$$

en définitive $\int_0^T [J_1(2\sqrt{\frac{t}{\lambda}})]^2\,dt$ vaut donc $T\,[J_1(2\sqrt{\frac{T}{\lambda_k}})]^2$

soit $T\,[J_1(j_{0,k})]^2$ puisque $2\sqrt{\frac{T}{\lambda_k}} = j_{0,k}$.

Les fonctions propres normalisées sont donc données par :

$$\boxed{g_k(t) = \frac{J_1(j_{0,k}\sqrt{\frac{t}{T}})}{\sqrt{T}\,J_1(j_{0,k})}}$$

et les corrélations entre X_t et les composantes principales ξ_k par :

$$r(X_t \; ; \; \xi_k) = \sqrt{\lambda_k} \; g_k(t)$$

soit

$$\boxed{r(X_t \; ; \; \xi_k) = \frac{2 \; J_1(j_{0,k}\sqrt{\frac{t}{T}})}{j_{0,k} \; J_1(j_{0,k})}}$$

Comme la somme des valeurs propres vaut T puisque les données sont réduites on en déduit la relation suivante entre les zéros de la fonction de Bessel J_0

$$\boxed{\sum_{s=1}^{\infty} \frac{1}{(j_{0,s})^2} = \frac{1}{4}}$$

On retrouve ainsi un cas particulier des formules d'Euler et de Rayleigh ([66] p. 500-502).

Numériquement les quatre premiers zéros de J_0 sont approximativement égaux à 2,40 ; 5,52 ; 8,65 ; 11,79 ce qui conduit à des parts d'inerties:

	(cumul)
λ_1/T = 69,2%	69,2%
λ_2/T = 13,1%	82,3%
λ_3/T = 5,3%	87,6%
λ_4/T = 2,9%	90,5%

On constate que le taux d'inertie de la première composante de l'analyse réduite est inférieur à celui de l'analyse non réduite (81,1%) mais que c'est l'inverse pour les composantes suivantes.

Comme on pouvait le prévoir la valeur finale du processus, X_T, détermine beaucoup moins les composantes principales dans l'analyse réduite. Ainsi, ici $r(X_T \; ; \; \xi_1) \simeq 0,83$ (au lieu de 0,90) et le maximum de $r(X_t \; ; \; \xi_1)$ est atteint pour $\frac{t}{T} \simeq 0,56$ et vaut environ 0,93 . ξ_1 est donc beaucoup plus lié aux

valeurs intermédiaires du processus que dans l'analyse non réduite et représente donc une variable de différenciation selon l'avance ou le retard par rapport à la moyenne : on a donc d'une certaine façon éliminé la composante triviale de l'analyse usuelle.

Remarquons pour terminer que tous les processus à accroissements stationnaires non corrélés ont la même A.C.P. bien qu'ils recouvrent des phénomènes fort différents (mouvement brownien ou processus de Poisson par exemple).

La simple considération des valeurs propres et des facteurs principaux ne permet pas d'identifier un processus : on retrouve ici une limitation des méthodes factorielles linéaires qui n'utilisent que les moments d'ordre deux. Ce sont les variables ξ_k qui permettraient de caractériser le processus étudié mais dans l'état actuel des connaissances leurs lois ne sont pas connues sauf dans le cas gaussien (mouvement brownien) où elles sont elle-mêmes gaussiennes.

III. MISE EN OEUVRE DE LA METHODE

Nous regroupons dans ce paragraphe des résultats pour l'essentiel connus concernant l'approximation et l'interpolation des processus et des commentaires sur l'interprétation des résultats et la façon de se servir d'une analyse factorielle pour des données chronologiques.

III.1 L'approximation numérique

Lorsque l'on dispose d'observations en continu concernant n individus, c'est-à-dire de n trajectoires d'un processus(X_t), deux problèmes d'approximation de natures différentes se posent ; le premier d'ordre théorique et probabiliste : l'analyse factorielle des n trajectoires est-elle une approximation de l'analyse factorielle du processus ? Il s'agit ici d'un problème d'échantillonnage au sens classique du terme.

Le deuxième est d'ordre pratique et numérique : si l'analyse exacte des n trajectoires est théoriquement possible et se ramène à une diagonalisation d'une matrice n x n (voir I.1.b de ce chapitre) en fait ceci n'est pas réalisable dès que n est grand et il faut procéder à une discrétisation du temps pour se ramener à une A.C.P. concernant p variables.

III.1.a) Echantillonnage

En supposant un modèle d'échantillonnage aléatoire simple (chaque trajectoire est une réalisation indépendante de (X_t)) on peut démontrer si $n \to \infty$ la convergence de l'analyse factorielle de n trajectoires vers celle du processus (X_t).

Les résultats établis par [16] et [17] montrent sous des hypothèses assez larges (existence d'un moment d'ordre 4) la convergence uniforme en moyenne quadratique ou presque sûre, de l'opérateur aléatoire de covariance C_n associé à un échantillon vers l'opérateur C du processus, et des valeurs propres et des facteurs associés.

Des résultats récents établis par ROMAIN [56] généralisent aux processus sur [0,T] les résultats classiques de T.W. ANDERSON [2] sur la distribution d'échantillonnage de valeurs propres, des vecteurs propres et des projecteurs associés aux sous espaces propres de C_n .

Donc si n est assez grand on est assuré de la validité des résultats.

III.1.b) Résolution numérique

On suppose ici que les n trajectoires $x_i(t)$ sont entièrement connues sur [0;T] donc que l'opérateur C associé a l'échantillon est calculable en tout couple t,s $C(t,s) = \sum_{i=1}^{p} p_i \, x_i(t) \, x_i(s)$ après centrage éventuel.

Il s'agit alors de résoudre un problème numérique, celui de la résolution de l'équation :

$$\int_0^T C(t,s) \, f(s) \, ds = \lambda \, f(t)$$

Fixons-nous un découpage de [0,T] en p segments :
$[0,t_1]$, $[t_1 \; t_2]$,.... $[t_{p-1};T]$.

A t fixé on sait qu'il existe des points ξ_j dans chaque segment tels que :

$$\int_0^T C(t,s) \, f(s) \, ds = \sum_{j=1}^{p} C(t,\xi_j) \, f(\xi_j)(t_j - t_{j-1}) = \lambda \, f(t) \; .$$

On ne peut évidemment utiliser cette formule pour calculer f(t) puisque les ξ_j dépendent à la fois de t et surtout de f qui est inconnue.

α) Méthodes de quadrature

Le plus simple est alors d'approximer l'intégrale par une formule de quadrature :

$$\int_0^T C(t,s)\ f(s)\ ds \ \# \ \sum_{j=0}^{p} C(t;t_j)\ f(t_j)\ a_j$$

où les coefficients a_j ne dépendent que du découpage de [0,T]. On pourra écrire alors pour $t=0,t_1,\ldots,T$ le système :

$$\sum_{j=0}^{p} C(t_i;t_j)\ f(t_j)\ a_j = \lambda\ f(t_i) \qquad i=0,1,2,\ldots p$$

dont la résolution donnera des valeurs approximatives de λ et de $f(t_i)$.

On est alors ramené à un problème de diagonalisation d'une matrice de taille (p+1)

$$C\ A\ \underline{f} = \lambda\ \underline{f}$$

où C est la matrice des covariances des X_{t_j} de termes $C(t_i;t_j)$ et A une matrice dépendant de la méthode de quadrature retenue.

Remarquons ici que ceci revient à faire l'analyse en composantes principales des X_{t_j} avec une métrique A : les approximations des facteurs principaux de C sont les axes principaux de l'A.C.P. des X_{t_j} avec pour métrique A.

Pour la plupart des méthodes élémentaires de quadrature la matrice A est diagonale :

* méthode des rectangles

$$a_j = t_{j+1} - t_j \qquad j=0,1,2,\ldots,p-1$$

* méthode des trapèzes

$$a_0 = \frac{t_1 - t_0}{2} \qquad a_1 = \frac{t_2 - t_0}{2} \qquad a_2 = \frac{t_3 - t_1}{2}$$

$$\ldots\ldots \quad a_p = \frac{t_p - t_{p-1}}{2}$$

* méthode de Simpson à pas constant p pair

$$a_0 = a_p = \frac{T}{3p} \qquad a_2 = a_4 = a_6 = \ldots = \frac{4T}{3p}$$

$$a_1 = a_3 = a_5 = \ldots = \frac{2T}{3p}$$

d'autres formules sont bien sur utilisables, il suffit de se reporter à un ouvrage d'analyse numérique.

β) Approximation par moyennage

La méthode des rectangles et celle des trapèzes reviennent toutes deux à approximer l'intégrale d'une fonction sur $[0,T]$ par celle d'une fonction en escalier. Cette remarque permet de passer à un autre type d'approximation où on cherche à approximer $f(t)$, et toute fonction de $L^2(T)$ d'ailleurs, par une fonction en escalier avec le découpage $0 = t_0, t_1, t_2, \ldots, t_p = T$.

Il est alors naturel de chercher $\hat{f}(t)$ facteur principal du processus $(\hat{X}_t)$ obtenu en projetant (X_t) sur l'ensemble des fonctions constantes sur $[t_i, t_{i+1}]$: en d'autres termes on remplace chaque trajectoire $x_i(t)$ par $\hat{x}_i(t)$ telle que si $t \in [t_i ; t_{i+1}]$ $\quad \hat{x}_i(t) = \displaystyle\int_{t_i}^{t_{i+1}} \frac{x_i(t)\,dt}{t_{i+1} - t_i}$.

$\hat{x}_i(t)$ est la trajectoire obtenue en prenant la moyenne de $x_i(t)$ sur chaque intervalle du découpage. La matrice de covariance du processus $(\hat{X}_t)$ s'obtient alors en faisant la moyenne des covariances sur chaque rectangle $[t_i ; t_{i+1}] \times [t_j ; t_{j+1}]$ $\quad i,j = 0,1,2 ; \ldots ; p-1$.

Ce procédé est certainement plus précis que les méthodes élémentaires de quadrature puisque l'on tient compte, en en prenant la moyenne, de ce qui s'est "passé" dans chaque intervalle et non seulement de la situation aux bornes de l'intervalle. Il est particulièrement adapté au cas où le processus (X_t) varie par sauts [17] et le calcul est simple s'il y a au plus un saut par intervalle : si l'individu i subit un saut à la date θ

dans l'intervalle $[t_i \; t_{i+1}]$ $\quad \hat{x}_i(t) = x(t_i) + \dfrac{\theta - t_i}{t_{i+1} - t_i} \cdot (x_{t_{i+1}} - x_{t_i})$.

γ) Généralités sur l'approximation

On peut généraliser ce type d'approximation en cherchant à approximer les trajectoires par des fonctions autres qu'en escalier mais cela pose quelques problèmes.

Cherchons en effet une approximation des facteurs principaux f dans un sous-espace de dimension finie de $L^2(T)$ engendré par un système orthonormé de fonctions $e_1(t) \quad e_2(t) \; \ldots \; e_p(t)$ (les fonctions trigonométriques ou les polynômes de Legendre par exemple). La meilleure approximation de f

par les e_j est évidemment $f^* = \sum_{j=1}^{p} (e_j | f) \, e_j = P f$ où P est le projecteur

orthogonal de $L^2(T)$ sur l'espace engendré par les e_j . Les trajectoires du processus étant des éléments de $L^2(T)$ il est naturel dans ce contexte de considérer les approximations de ces trajectoires au moyen des e_i , c'est-à-dire de travailler sur le processus approximé $(X_t^*) = P(X_t)$. L'opérateur de covariance de (X_t^*) est alors PCP mais les facteurs principaux g de (X_t^*) ne sont pas les projections des facteurs de (X_t) : $g \neq P f$.

Les facteurs principaux de (X_t^*) sont cependant de la forme $\sum_{j=1}^{p} \alpha_j \, e_j(t)$

et on peut montrer [17] que si $p \to \infty$ la suite d'opérateurs PCP converge vers C ainsi que les suites d'éléments propres ce qui justifie l'utilisation de l'opérateur PCP qui peut être alors représenté par une matrice d'ordre p puisque l'on se ramène à un problème en dimension finie. Les calculs des éléments de la matrice PCP risquent cependant d'être laborieux car en se plaçant dans la base formée par les $e_j(t)$ il faut calculer l'image par C de chaque $e_j(t)$ pour obtenir les éléments de PCP $((PCP)_{i,j} = (e_i | C \, e_j))$ ce qui

entraîne des intégrations. Comme ces intégrales seront la plupart des fois calculées par discrétisation de [0,T] il semble plus simple de procéder directement à une discrétisation du temps et d'utiliser des fonctions constantes par intervalles comme nous l'avons indiqué précédemment, ce qui revient à prendre comme système orthogonal les fonctions

$$e_i(t) = \frac{\mathbb{1}_{t_i - t_{i-1}}(t)}{\sqrt{t_i - t_{i-1}}} \qquad i = 1,2,\ldots,p$$

pour le découpage $0 = t_0 , t_1, t_2, \ldots, t_p = T$.

δ) Sur le choix de la discrétisation

Le nombre de points étant fixé il reste à choisir le découpage de [0;T]. Si le processus est approximativement stationnaire (ou stationnaire après transformation linéaire), le découpage en intervalles de longueur égales est évidemment le meilleur ce qui revient à effectuer une analyse en composantes principales ordinaire de $X_{t_1}, X_{t_2}, \ldots, X_{t_p}$.

Si le processus n'est pas stationnaire le choix du "meilleur" découpage de [0;T] est un problème délicat qui soulève tout d'abord la question de l'optimalité :

meilleur découpage pour l'approximation des valeurs propres ou
meilleur découpage pour l'approximation des facteurs principaux? Etc.

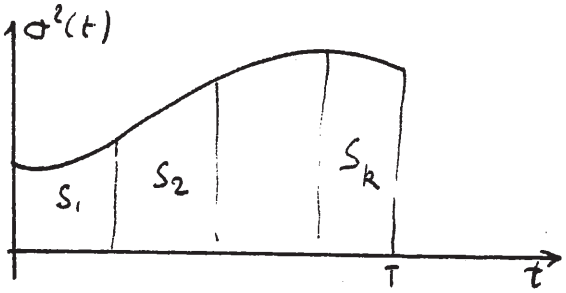
Ces problèmes sont à l'heure actuelle non résolus.

Par analogie avec les résultats de la théorie des sondages stratifiés on peut donner comme règle heuristique de découper plus finement les parties de [0,T] où la variabilité est la plus grande. On peut alors émettre la conjecture suivante "le nombre de points de discrétisation sur une partie [a,b] incluse dans [0,T] doit être proportionnel à l'écart-type total du processus sur [a,b]".

Nous proposons alors la règle empirique suivante :

On trace la courbe $\sigma^2(t)$ au besoin très grossièrement et on découpe arbitrairement l'intervalle [0,T] en k parties égales A_i. On évalue par une méthode

élémentaire la surface S_i de chaque morceau situé sous la courbe. Le nombre de points du découpage final de $[0,T] \in A_i$ sera alors proportionnel à S_i. Il suffit ensuite de découper chaque A_i selon le nombre trouvé en parties égales pour obtenir le découpage de $[0,T]$ qui servira pour l'approximation.



Ce découpage semble recommandable aussi bien pour la méthode de moyennage des trajectoires que pour celle plus fruste d'intégration par quadrature.

III.2 L'interpolation

Bien que faisant appel à des outils mathématiques proches de ceux de l'approximation, l'interpolation recouvre en fait un problème totalement différent : au lieu de connaître la totalité des trajectoires et d'être obligés de discrétiser $[0,T]$ par nécessité de calcul, on suppose ici que le processus, bien qu'à temps continu par nature, ne peut être observé qu'à des dates précises $0, t_1, t_2, \ldots t_p = T$ pas nécessairement équidistantes, par suite des contraintes de l'appareillage statistique. On connaît donc les variables $X_0, X_{t_1}, \ldots X_T$. Comme l'a montré J.C. DEVILLE [17]

il serait incorrect d'effectuer une A.C.P. ordinaire sur ces variables pour obtenir une approximation de la véritable analyse factorielle du processus (X_t) ne serait-ce que par le fait que les intervalles $[t_i\ t_{i+1}]$ ne sont pas nécessairement égaux mais surtout parce que les résultats doivent tenir compte d'hypothèses de régularité des trajectoires entre les instants d'observation, hypothèses liées en fait implicitement à l'utilisation d'une métrique pour mesurer les distances entre trajectoires comme le suggèrent les travaux de P. BESSE [7] que nous tentons de réinterpréter.

III.2.a) L'interpolation linéaire

En reprenant les notations de [17], étant donné p fonctions d'interpolations $e_j(t)$, simples (c'est-à-dire $e_j(t_j) = 1$ et $e_j(t_k) = 0$ si $j \neq k$) l'interpolation d'une trajectoire $x(t)$ sera notée $\Phi\, x(t)$ et vaut

$\Phi x(t) = \sum_{j=1}^{p} x_j\, e_j(t)$ et est donc un élément d'un espace à p dimensions. Nous noterons $a_i(t)$ les éléments d'une base de cet espace de telle sorte que les fonctions d'interpolation s'écrivent $e_i(t) = \sum_i \varphi_{ij}\, a_j(t)$. Remarquons que pour l'instant nous ne supposons pas les $a_j(t)$ orthonormés.

Ecrire que les fonctions d'interpolation sont simples implique

$$\sum_j \varphi_{ij}\, a_j(t_k) = \delta_{jk}$$

la matrice ϕ d'éléments φ_{ij} est donc l'inverse de la matrice B d'éléments $a_j(t_i)$; $\phi = B^{-1}$.

L'interpolation étant obtenue par linéarité l'opérateur de covariance du processus interpolé n'est autre que l'interpolation de l'opérateur de covariance et on a :

$$C_\phi(t,s) = E\,[\phi_X(t)\ \Phi_X(s)]$$

$$= \sum_i \sum_j \sum_k \sum_\ell \varphi_{ij}\, \varphi_{k\ell}\, c_{jk}\ \ a_j(t)\ a_\ell(s)$$

où $c_{jk} = C(t_j\ ;\ t_k)$.

Les fonctions propres de C_ϕ vérifient l'équation :

$$\lambda\ f(t) = \int_0^T C_\phi(t,s)\ f(s)\ ds$$

soit puisque le noyau est dégénéré :

$$\sum_i \sum_j \sum_k \sum_\ell \varphi_{ij}\, \varphi_{k\ell}\, c_{jk}\ a_j(t) \int_0^T a_\ell(s)\ f(s)\ ds = \lambda\ f(s)$$

les $f(s)$ sont donc de la forme $\sum z_j\ a_j(t)$ et on obtient le vecteur

$\underline{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_p \end{pmatrix}$ comme vecteur propre de la matrice $\phi'\ C\ \phi\ A$ où A est la matrice

des produits scalaires des $a_j(t)$: $\quad a_{ij} = \int_0^T a_i(t)\ a_j(t)\ dt$.

Si on utilise comme fonction d'interpolation des fonctions constantes par morceaux on a $\phi = I$ et la matrice A diagonale avec ses termes égaux aux longueurs des intervalles : on retrouve les résultats du paragraphe consacré à l'approximation.

Supposons les fonctions de base $a_i(t)$ orthonormées de telle sorte que $A = I$. Il est clair alors qu'effectuer une interpolation revient à choisir une métrique puisque diagonaliser $\phi'\ C\ \phi$ est équivalent à effectuer une A.C.P. sur les X_{t_i} avec pour métrique

$$M = \phi\phi' \ .$$

Cette remarque est d'une grande importance : le choix des fonctions d'interpolation est équivalent à celui d'une métrique : en particulier si le découpage est en intervalles égaux choisir des fonctions constantes par morceaux est équivalent à l'utilisation de la métrique I.

III.2.b) Lien avec les espaces de Sobolev

Dans [7] P. BESSE fait le lien entre la recherche d'interpolation SPLINE de l'analyse factorielle et le noyau reproduisant de l'espace de Sobolev associé. En cherchant directement à plonger les trajectoires dans l'espace de Sobolev $H^k(T)$ on montre que lors d'une discrétisation de [0,T] il faut effectuer l'analyse des X_{t_i} avec pour métrique Λ^{-1} où Λ est la matrice d'éléments $\lambda_{(t_i;t_j)}$ avec $\lambda(t,s)$ noyau reproduisant de l'espace de Sobolev.

Avec nos notations ceci veut dire que $\Lambda^{-1} = M$ donc que

$$\Lambda = (\phi\phi')^{-1} = (B^{-1}\ B'^{-1})^{-1} = B'B \ .$$

Il en résulte une simplification des calculs:l'utilisation d'une interpolation linéaire conduit à des espaces de Sobolev H^1 dont le noyau reproduisant est $1+\inf(t,s)$; la métrique vaut donc :

$$M = \left(\ldots 1+\inf(t_i,t_j)\ldots\right)^{-1}$$

et on en conclut que l'interpolation linéaire revient à utliser une métrique prenant en compte à la fois les valeurs des trajectoires aux points t_i et les valeurs des taux d'accroissements. Sans le 1 du noyau reproduisant on obtiendrait exactement la métrique des différences [25].

Lorsque les trajectoires sont irrégulières et que l'on veut prendre un filtre d'intégration avant de les traiter, P. BESSE plonge alors les trajectoires dans le dual d'un espace de Sobolev. Après calculs il trouve que pour l'analyse des intégrales des trajectoires il faut faire l'A.C.P. des X_{t_i} avec pour métrique Λ et non Λ^{-1} ce qui est bien connu en analyse des données puisqu'en passant d'un espace à son dual il faut inverser la métrique. Pour le cas de l'intégration simple on retrouve aussi le résultat bien connu que la métrique du cumul est l'inverse de la métrique des différences.

Rappelons que l'on appelle métrique du cumul la métrique associée à la transformation linéaire :

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \longrightarrow \begin{pmatrix} X_1 \\ X_1+X_2 \\ \vdots \\ X_1+X_2+X_3+\ldots+X_p \end{pmatrix}$$

et métrique des différences, celle associé à la transformation :

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_p \end{pmatrix} \longrightarrow \begin{pmatrix} X_1 \\ X_2-X_1 \\ X_3-X_2 \\ \vdots \\ X_p-X_{p-1} \end{pmatrix}$$

Ceci explique que dans l'exemple traité en [7] d'analyse de températures dans $H^1(T)$ on trouve une métrique qui ne diffère que de très peu de celle associée à l'A.C.P. des différences d'un mois à l'autre.

Toujours est-il que la convergence de l'interpolation spline d'ordre k de l'A.C.P. d'un processus vers l'analyse de ce processus dans l'espace de Sobolev $H^k(T)$ lorsque la discrétisation se fait de plus en plus fine aboutit à un résultat a priori surprenant : interpoler les trajectoires par d'autres fonctions que des constantes ne donne pas une "meilleure" approximation de l'A.C.P. de (X_t) mais conduit à l'analyse de (X_t) avec une métrique différente de I.

Ce résultat s'explique en fait de la manière intuitive suivante : utiliser une interpolation spline d'ordre k revient à supposer une certaine régularité du processus et de ses dérivées jusqu'à l'ordre k ce qui conduit implicitement à l'emploi d'une métrique de Sobolev.

III.3 Sur l'interprétation des résultats

L'analyse en composantes principales d'un processus (X_t) n'étant qu'une extension de l'A.C.P. usuelle et s'y ramenant en pratique par l'usage des mêmes programmes d'ordinateur toutes les règles et techniques d'interprétation habituelles restent valables : calcul des corrélations entre variables et composantes principales illustrées par le cercle des corrélations ; calcul des contributions des individus à l'inertie d'une axe etc. Comme en A.C.P. usuelle les interprétations les plus fécondes sont celles qui s'effectuent en utilisant des variables passives et en ne se contentant pas d'une explication purement endogène du phénomène : en tant que méthode exploratoire une des principales utilisations de l'A.C.P. consiste à faire apparaître des comportements différents de groupes d'individus, comportements plus facilement "explicables" si on fait intervenir,après l'analyse,des caractéristiques numériques ou qualitatives de ces groupes. L'étude du calendrier de constitution des familles [17] est à cet égard exemplaire : après avoir dégagé deux facteurs principaux le premier lié à la taille finale de la famille, le second à la vitesse de constitution (en avance ou en retard par rapport à la moyenne) les résultats les plus intéressants apparaissent en reliant ces phénomènes, prévisibles du reste (voir le paragraphe II.3.b), à des caracréristiques socio-culturelles des familles.

L'analyse factorielle d'un processus possède en plus des possibilités d'interprètation inconnues en A.C.P. usuelle : les facteurs principaux étant des fonctions du temps on peut les représenter par des courbes.

CONCLUSION

Au terme de ce chapitre nous pensons avoir montré les possibilités offertes par l'analyse factorielle pour le traitement de données temporelles. Nous croyons que l'analyse en composantes principales donne dans tous les cas des résultats au moins aussi intéressants qu'une méthode spécifique d'analyse de séries chronologiques, comme l'analyse spectrale, tout en requérant moins d'hypothèses sur le phénomène et en n'utilisant qu'une mathématique somme toute assez simple. La possibilité d'utiliser des métriques variées élargit encore le champ d'application de la méthode en y incorporant les différents procédés de filtrage coutumiers pour les chroniques.

Nous espérons avoir montré également le caractère erroné du reproche souvent fait aux techniques d'analyse factorielle d'ignorer la temporalité du phénomène étudié car les résultats seraient invariants par permutation des temps. Outre que cette assertion est inexacte pour certaines métriques la réponse nous semble être dans le fait que l'objet d'étude, comme en analyse spectrale d'ailleurs, est la fonction d'auto-corrélation qui est à l'évidence un outil temporel. Le temps n'intervient pas seulement au niveau de l'interprètation mais est présent tout au long de l'analyse.

Certes pour être valablement appliquée l'analyse factorielle suppose un nombre important d'individus ce qui la différencie du cas classique des séries chronologiques où on ne dispose que d'une seule trajectoire : cependant à bien regarder les hypothèses faites pour l'analyse de telles séries on peut envisager d'appliquer l'analyse factorielle sur une seule trajectoire à condition de postuler aussi la stationnarité et l'ergodicité (voir [3]) : la diagonalisation de l'opérateur de covariance estimé au moyen des différences longitudinales conduira à des résultats comparables à ceux de l'analyse spectrale.

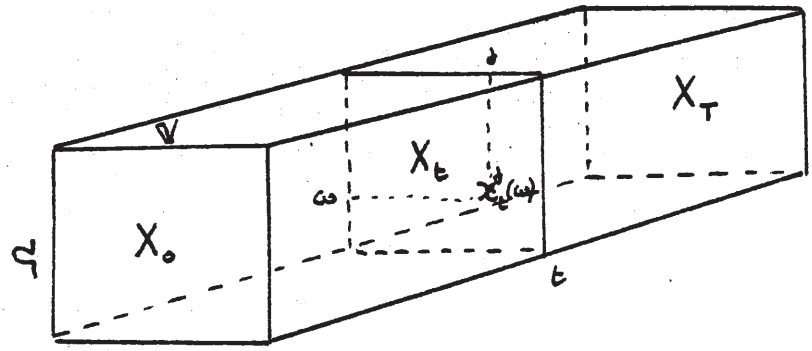
CHAPITRE II

ANALYSES FACTORIELLES D'UN PROCESSUS VECTORIEL

Ce chapitre est consacré à différents types d'analyses factorielles d'un processus m-dimensionnel $(\vec{X}_t)$ à valeurs dans $\mathbb{R}^m$ ou plus généralement dans un espace vectoriel $\mathbb{V}$ de dimension m (la distinction est ici importante : parler de $\mathbb{R}^m$ évoque des chroniques multiples où chaque composante de $\vec{X}_t$ à un sens concret ; parler de $\mathbb{V}$ indique qu'aucune base n'est privilégiée ce qui est le cas par exemple de trajectoires de points se déplaçant dans l'espace physique).

Il s'agit donc de la généralisation de ce qu'il est convenu d'appeler l'"analyse des données cubiques" qui a donné lieu à un grand nombre de travaux en raison de la diversité des objectifs qui peuvent être poursuivis : en effet selon que l'on veut étudier les liaisons entre instants, l'évolution des corrélations entre les variables au cours du temps, les trajectoires individuelles etc. , les techniques seront différentes et apporteront un éclairage particulier au phénomène étudié.

Le terme impropre mais commode de données cubiques fait référence à la représentation imagée suivante d'un processus multidimensionnel :
$\vec{X}_t$: $(X_t^1, X_t^2, \ldots, X_t^p)$



où $x_t^j(\omega)$ est la valeur de la $j^{\text{è}}$ composante de $(\vec{X}_t)$ pour l'"individu" ω à la date t .

De nombreuses tentatives de généralisation de l'analyse factorielle à des données à trois indices ont été proposées particulièrement par l'école psychométrique anglo-saxonne sous le nom de three-way factor analysis. (On trouve un panorama des développements récents dans [68]).

Nous ne les envisagerons pas ici, d'une part parce qu'elles utilisent des modèles a priori de décomposition des $x_t^j(\omega)$ qui ne nous semblent guère satisfaisants, d'autre part surtout parce qu'elles ne font pas jouer au troisième indice, le temps, un rôle privilégié. On trouvera également dans [9] outre une revue de travaux sur les chroniques multiples, une méthode intitulée double analyse en composantes principales qui procède d'abord à une A.C.P. de $\vec{m}(t)$ considéré comme un nuage de p points dans $\mathbb{V}$ pour étudier la tendance puis à la recherche d'un système de "composantes principales" commun aux $\vec{X}_t$ tenant compte de l'A.C.P. de $\vec{m}(t)$. Mais cette méthode ne rentre pas dans notre cadre d'autant que nous supposerons généralement le processus centré.

Il est clair que l'étude d'un processus $(\vec{X}_t)$ où t varie continuement dans l'intervalle [0,T] est l'extension naturelle d'un problème d'analyse des données que nous avions traîté antérieurement [57] : celui de l'étude simultanée de plusieurs tableaux de données. On retrouve ce cas particulier pour Ω fini et un ensemble discret de valeurs du temps. Nous rappellerons donc dans un premier paragraphe les résultats essentiels concernant l'analyse canonique de plusieurs tableaux de données avant de l'étendre aux processus sous le nom d'analyse harmonique vectorielle dont nous établirons les propriétés qui généralisent celle de l'analyse de CARROLL [13] .

La troisième partie de ce chapitre consacrée à une exploration systématique des diverses analyses factorielles possibles d'un cube de données montrera en particulier que l'analyse harmonique vectorielle est la seule méthode étendant de façon naturelle l'analyse d'un processus scalaire.

Enfin nous étudierons dans une dernière partie des techniques plus globales fondées sur l'utilisation des opérateurs introduits par Y. ESCOUFIER.

I· ANALYSE CANONIQUE D'UN ENSEMBLE DE TABLEAUX DE DONNEES

Considérons ici le cas d'une population finie $|\Omega| = n$ pour laquelle on observe m variables en un nombre p d'instants. On se trouve donc en présence d'un ensemble de p tableaux rectangulaires $X_1, X_2, \ldots, X_p$ à n lignes et m colonnes. X_t est donc en termes probabilistes une n-réalisation du processus $(\vec{X}_t)$.

Tous les tableaux X_t ont donc même nombre de colonnes mais les résultats qui suivent sont valables pour des tableaux de dimensions différentes. On notera V_{ij} les matrices de covariance croisées entre les variables de X_i et de X_j (avec des poids égaux $V_{ij} = \frac{1}{n}\, {}^tX_i\, X_j$) et A_i le projecteur orthogonal sur l'espace W_i engendré linéairement par les colonnes de X_i . Les X_i seront supposées de rang m .

Dans ce qui suit on rappelle les résultats essentiels de l'analyse canonique généralisée qui seront utiles pour la suite.

I.1 Formulaire d'analyse canonique ordinaire

Etant donné deux tableaux X_1 et X_2 on recherche des couples de variables $(\underline{\xi}, \underline{\eta})$ combinaisons linéaires respectives des variables de X_1 et de X_2 ayant le plus grand coefficient de corrélation linéaire.

Les variables $\underline{\xi}$ et $\underline{\eta}$ sont appelées variables canoniques leur coefficient de corrélation r , corrélation canonique, les facteurs $\underline{a}$ et $\underline{b}$ tels que $\underline{\xi} = X\,\underline{a}$ $\underline{\eta} = X_2\,\underline{b}$ facteurs canoniques. Il existe m couples de variables canoniques orthogonales telles que

$$r(\underline{\xi}_i\ ;\ \underline{\xi}_j) = \delta_{ij} = r(\underline{\eta}_i\ ;\ \underline{\eta}_j)$$

$$r(\underline{\xi}_i\ ;\ \underline{\eta}_j) = \delta_{ij}\, r_j$$

les variables canoniques normées vérifient les propriétés suivantes :

Propriété 1

$$A_1\, A_2\, \underline{\xi} = \lambda\, \underline{\xi} \qquad A_2\, A_1\, \underline{\eta} = \lambda\, \underline{\eta} \qquad \underline{\xi} = \frac{1}{\sqrt{\lambda}}\, A_1\, \underline{\eta}$$

$$\underline{\eta} = \frac{1}{\sqrt{\lambda}}\, A_2\, \underline{\xi}$$

Propriété 2

$\underline{z} = \underline{\xi} + \underline{\eta}$ est tel que $(A_1 + A_2)\underline{z} = (1 + \sqrt{\lambda})\ \underline{z}$

d'où $\underline{\xi} = A_1 z/(1+\sqrt{\lambda})$ $\qquad \underline{\eta} = A_2 z/(1+\sqrt{\lambda})$

$A_1 + A_2$ étant de rang 2m on obtient ainsi les m premiers vecteurs propres de $A_1 + A_2$; les m suivants sont associés aux différences $\underline{\xi} - \underline{\eta}$ et aux valeurs propres $1 - \sqrt{\lambda}$.

Propriété 3

Les vecteurs propres $\underline{z}$ de $A_1 + A_2$ rendent maximale la somme des carrés de corrélation multiples $R^2(\underline{z}\ ;\ X_1) + R^2(\underline{z}\ ;\ X_2)$.

Propriété 4

Cette propriété nouvelle a été démontrée par nous dans des publications précédentes [20] et [60] :

Les vecteurs à 2n composantes $\begin{pmatrix} \underline{\xi}_i \\ \underline{\eta}_i \end{pmatrix}$ sont tels que :

$$\begin{pmatrix} A_1 & A_1 A_2 \\ A_2 A_1 & A_2 \end{pmatrix} \begin{pmatrix} \underline{\xi} \\ \underline{\eta} \end{pmatrix} = (1 + \sqrt{\lambda}) \begin{pmatrix} \underline{\xi} \\ \underline{\eta} \end{pmatrix}$$

Les facteurs canoniques vérifient les propriétés suivantes :

Propriété 5

$$V_{22}^{-1}\ V_{21}\ V_{11}^{-1}\ V_{12}\ \underline{b} = \lambda\ \underline{b} \qquad\qquad V_{11}^{-1}\ V_{12}\ V_{22}^{-1}\ V_{21}\ \underline{a} = \lambda\ \underline{a}$$

$$\underline{a} = \frac{1}{\sqrt{\lambda}}\ V_{11}^{-1}\ V_{12}\ \underline{b} \qquad\qquad b = \frac{1}{\sqrt{\lambda}}\ V_{22}^{-1}\ V_{21}\ \underline{a}$$

Propriété 6

Les vecteurs à 2m composantes $\begin{pmatrix} \underline{a} \\ \underline{b} \end{pmatrix}$ vérifient la relation :

$$\begin{pmatrix} V_{11}^{-1} & 0 \\ 0 & V_{22}^{-1} \end{pmatrix} \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \begin{pmatrix} \underline{a} \\ \underline{b} \end{pmatrix} = (1 + \sqrt{\lambda}) \begin{pmatrix} \underline{a} \\ \underline{b} \end{pmatrix}$$

On en déduit alors que l'analyse canonique de X_1 et X_2 est équivalente à l'analyse en composantes principales du tableau $(X_1|X_2)$ avec pour métrique

$$\begin{pmatrix} V_{11}^{-1} & 0 \\ 0 & V_{22}^{-1} \end{pmatrix}.$$

I-2 Généralisation à p tableaux

Les propriétés 2 et 3 précédentes sont à la base de la généralisation de l'analyse canonique due à J.D. CARROLL [13] qui a été l'objet d'une partie de nos travaux antérieurs [57] .

Cette analyse consiste dans la recherche de variables $\underline{z}$ les plus liées à chacun des ensembles X_i au sens où $\sum_{i=1}^{p} R^2(z;X_i)$ est maximal.

On obtient alors immédiatement la généralisation des propriétés 2 et 6 :

les $\underline{z}$ sont vecteurs propres de $\sum_{i=1}^{p} A_i$.

Comme $\underline{z}$ est alors nécessairement de la forme $\underline{z} = \sum_{i=1}^{p} X_i \underline{a}_i$; en notant $\underline{a}$ le vecteur à mp composantes :

$$\underline{a} = \begin{pmatrix} \underline{a}_1 \\ \underline{a}_2 \\ \vdots \\ \underline{a}_p \end{pmatrix}$$ on trouve que $\underline{a}$ vérifie l'équation :

$$\begin{pmatrix} V_{11}^{-1} & & & 0 \\ & V_{22}^{-1} & & \\ & & \ddots & \\ 0 & & & V_{pp}^{-1} \end{pmatrix} \begin{pmatrix} V_{11} & V_{12} & \cdots & V_{1p} \\ V_{21} & V_{22} & & V_{2p} \\ \vdots & & & \\ V_{p1} & V_{p2} & \cdots & V_{pp} \end{pmatrix} \begin{pmatrix} \underline{a}_1 \\ \underline{a}_2 \\ \vdots \\ \underline{a}_p \end{pmatrix} = \lambda \begin{pmatrix} \underline{a}_1 \\ \underline{a}_2 \\ \\ \underline{a}_p \end{pmatrix}$$

les vecteurs $\underline{z}$ sont alors les composantes principales du tableau juxtaposé $(X_1|X_2| - |X_p)$ avec la métrique $\mathrm{Diag}(V_{ii}^{-1})$. Une telle analyse est l'équivalent d'une A.C.P. réduite sur p groupes de variables où chaque groupe est considéré comme un objet simple ; d'ailleurs lorsqu'il n'y a qu'une variable par groupe (tableaux X_i unicolonnes) on trouve exactement l'A.C.P. des variables réduites d'où le nom donné ultérieurement en [52] d'A.C.P. réduite généralisée.

L'inertie totale dans une telle A.C.P. vaut Trace $\Sigma A_i = mp$.
Pour obtenir les variables canoniques $\underline{\xi}_1, \underline{\xi}_2, \ldots \underline{\xi}_p$ relatives à chaque groupe, il suffit de projeter $\underline{z}$ sur chacun des espaces engendrés par les colonnes de X_i d'où $\underline{\xi}_i = \frac{1}{\lambda} A_i \underline{z} = X_i \underline{a}_i$. Lorsque $p > 2$ les normes de ces variables ne sont plus nécessairement égales mais l'on a $\sum_{i=1}^{p} \| \xi_i \|^2 = 1$.

Les variables $\underline{\xi}_i$ vérifient alors la propriété 4 généralisée :

$$\begin{pmatrix} A_1 & A_1A_2 & \cdots & A_1A_p \\ A_2A_1 & A_2 & \ddots & A_2A_p \\ \vdots & & \ddots & \vdots \\ A_pA_1 & A_pA_2 & \cdots & A_p \end{pmatrix} \begin{pmatrix} \underline{\xi}_1 \\ \underline{\xi}_2 \\ \\ \underline{\xi}_p \end{pmatrix} = \lambda \begin{pmatrix} \underline{\xi}_1 \\ \underline{\xi}_2 \\ \vdots \\ \underline{\xi}_p \end{pmatrix}$$

d'où une double propriété d'orthogonalité pour les variables canoniques d'ordre k et ℓ correspondant aux valeurs propres λ_k et λ_ℓ :

$$\cdot \quad r\left(\sum_{i=1}^{p} \underline{\xi}_i^{(k)} ; \sum_{i=1}^{p} \underline{\xi}_i^{(\ell)}\right) = \delta_{k\ell}$$

$$\cdot\cdot \quad \sum_{i=1}^{p} r(\underline{\xi}_i^{(k)} ; \underline{\xi}_i^{(\ell)}) = p\, \delta_{k\ell}$$

On a également la <u>propriété n° 7</u> :

z_k est composante principale de l'A.C.P. réduite des $\underline{\xi}_i^{(k)}$. En particulier z_1 est la première composante principale des $\xi_i^{(1)}$ normés, ce qui montre que l'analyse canonique généralisée est la solution du problème suivant : trouver des variables $\underline{\xi}_1; \underline{\xi}_2; \ldots \underline{\xi}_p$ de forme $\underline{\xi}_i = X_i \underline{a}_i$ ayant une première composante principale de variance maximale.

De toutes les généralisations possibles de l'analyse canonique (voir [32] [34]), celle qui précède est la seule à posséder ces diverses propriétés et à se ramener en particulier à un problème aux valeurs propres, c'est pourquoi nous l'appelons analyse canonique généralisée sans autre précision.

En tant qu'analyse en composantes principales du tableau $(X_1|X_2|--|X_p)$ cette méthode, par le choix de la métrique $\mathrm{Diag}(V_{ii}^{-1})$, respecte le groupage des variables : dans le cadre chronologique cela revient à tenir compte du temps en ce sens que toutes les variables mesurées à un même instant sont mises en "bloc" et que l'on traitera différemment les corrélations entre variables à deux instants différents et les corrélations entre variables au même instant (ces dernières sont ici ignorées) ce qui semble souhaitable. Appliquée à des données chronologiques l'analyse canonique généralisée revient à une étude de l'évolution globale des variables représentées par leurs espaces engendrés. Les résultats de cette analyse sont enrichis par le caractère temporel de la suite de tableaux qui permet d'interpréter les facteurs et par le fait que les variables ont la même signification d'un tableau à l'autre.

II- L'ANALYSE "HARMONIQUE" VECTORIELLE

Revenons maintenant à l'étude d'un processus vectoriel $(\vec{X}_t)$ sur $[0,T]$ vérifiant les hypothèses de régularité du chapitre I .

$(\vec{X}_t)$ est à valeurs dans $\mathbb{R}^m$: il est courant dans les applications que $\mathbb{R}^m$ soit muni d'une base naturelle: c'est le cas où l'on observe simultanément plusieurs variables bien définies et on peut alors écrire sans ambiguïté

$$(\vec{X}_t) = (X_t^1, X_t^2, \ldots, X_t^m) .$$

Par contre dans d'autres types d'applications (en physique par exemple) $(\vec{X}_t)$ est simplement un processus à valeurs dans un espace euclidien $\mathbb{V}$ de dimension m et toute base de $\mathbb{V}$ est arbitraire. On vérifiera que la méthode proposée ici est intrinsèque et ne dépend pas de la base choisie qui peut changer à chaque instant.

On supposera de plus que $(\vec{X}_t)$ est continu en moyenne quadratique et tel que l'opérateur de covariance du vecteur aléatoire $\vec{X}_t$ soit régulier pour tout t. L'usage de la métrique $[C(t,t)]^{-1}$ définie plus loin, ne présentera alors aucune difficulté puisqu'elle correspondra à un homéomorphisme de $L^2_{\mathbb{V}}(T)$ espace des (classes de) fonctions du temps à valeur dans $\mathbb{V}$ de norme carrée intégrable.

II-1 Eléments d'ordre deux associés à un processus vectoriel

II-1-a) Moyenne et covariance

La moyenne $\vec{m}(t)$ du processus $\vec{X}_t$ est la fonction vectorielle certaine du temps définie par :

$$\vec{m}(t) = E(X_t^1;X_t^2;\ldots;X_t^m) = (E(X_t^1),E(X_t^2)\ldots E(X_t^m)) = (m^1(t);m^2(t)\ldots m^m(t))$$

pour une base quelconque de $\mathbb{V}$.

On suppose désormais $\vec{m}(t) = \vec{0}$.

Pour deux instants t et s l'opérateur de covariance est l'application linéaire de $\mathbb{V}^*$ dans $\mathbb{V}$ définie par $C(t,s) = E(\vec{X}_t \otimes \vec{X}_s)$ et est donc représentée sur une base orthonormée de $\mathbb{V}$ par la matrice d'éléments $C(t,s)_{k\ell} = E(X_t^k X_s^\ell)$.

L'opérateur de covariance C du processus est alors l'application de $L^2_V(T)$ dans lui-même de noyau C(t,s) défini par :

$$C\,\vec{f} = \vec{g} \qquad \int_0^T C(t,s)\overrightarrow{f(s)}ds = \overrightarrow{g(t)}$$

où $\overrightarrow{f(s)}$ est une fonction vectorielle certaine du temps.

Comme en dimension 1, l'opérateur de covariance peut être défini de manière abstraite :

soit Y la variable aléatoire obtenue par l'intégrale :

$$Y = \int_0^T < \vec{X}_s \; ; \; \overrightarrow{f(s)} > ds$$

où < ; > désigne le produit scalaire de V

$$E(Y.\vec{X}_t) = \int_0^T E(\vec{X}_s \otimes \vec{X}_t)\overrightarrow{f(s)}ds$$

$$= \int_0^T C(t,s)\overrightarrow{f(s)}ds$$

$$\text{et} \quad V(Y) = \int_0^T <\overrightarrow{f(s)} \; ; C\,\overrightarrow{f(s)}>ds$$

On définit alors l'opérateur de covariance C de $(\vec{X}_t)$ par :

$$\forall \vec{f},\vec{g} \int_0^T < \vec{f}(s) ; C\,\vec{g}(s) > ds = E\left[\int_0^T < \vec{X}_t ; \vec{f}(t) > dt . \int_0^T < \vec{X}_s ; \vec{g}(s) > ds\right]$$

En raison des hypothèses faites C est un opérateur compact, autoadjoint et nucléaire.

II-1-b) Opérateur de régression ou de projection

Etant donné une variable aléatoire Y , on sait que la meilleure approximation de Y par $\vec{X}_t$ à t fixé est $E(Y/\vec{X}_t)$ espérance conditionnelle de Y sachant $\vec{X}_t$. Nous nous contenterons ici, restant dans un cadre linéaire, de chercher la meilleure fonction linéaire de $\vec{X}_t$, celle-ci étant obtenue par la projection orthogonale de Y sur l'espace vectoriel engendré par $\vec{X}_t$:

$$\hat{Y} = < \vec{f}(t) ; \vec{X}_t > \; = A_t(Y)$$

où A_t désigne le projecteur orthogonal.

Les résultats classiques de la régression linéaire donnent :

$$E(\vec{X}_t . Y) = C(t,t)\ \vec{f}(t) \qquad \text{"équations normales"}$$

soit avec l'inversibilité de C(t,t) :

$$\vec{f}(t) = [C(t,t)]^{-1}\ E\,[\vec{X}_t . Y] \qquad \text{"formule de régression"}$$

d'où l'expression du projecteur orthogonal A_t sur $\vec{X}_t$:

$$\boxed{A_t(Y) = < \vec{X}_t \ ; [C(t,t)]^{-1}\ E\,[\vec{X}_t\ Y] >}$$

II-2 Analyse canonique généralisée du processus $(\vec{X}_t)$

II-2-a) Définition et propriétés de l'analyse harmonique vectorielle

Nous la définirons comme la recherche des variables aléatoires indépendantes du temps, z solutions de l'équation intégrale :

$$\int_0^T A_t\ z\ dt = \lambda z$$

A_t étant un opérateur de rang fini $\int_0^T A_t \, dt$ est un opérateur compact, nucléaire de $L^2(\Omega)$ dans lui-même, l'existence de son analyse spectrale est assurée. La trace de $\int_0^T A_t \, dt$ vaut mT.

En explicitant A_t on trouve :

$$\int_0^T < \vec{X}_t \; ; \; [C(t,t)]^{-1} \; E(\vec{X}_t . z) > dt = \lambda z$$

ce qui permet de vérifier que z est de la forme

$$z = \int_0^T < \vec{X}_t \; ; \; \overrightarrow{f(t)} > dt$$

en remplaçant z par cette expression dans les deux membres de l'équation intégrale on trouve $\overrightarrow{f(t)}$ par identification :

$$\int_0^T < \vec{X}_t ; [C(t,t)]^{-1} \; E(\vec{X}_t \int_0^T < \vec{X}_s , \overrightarrow{f(s)} > ds) > dt = \lambda \int_0^T < \vec{X}_t ; \overrightarrow{f(t)} > dt$$

soit $\quad [C(t,t)]^{-1} \int_0^T C(t,s) \overrightarrow{f(s)} ds = \lambda \; \overrightarrow{f(t)}$

d'où la :

> Proposition 2.1
>
> Les facteurs principaux de l'analyse factorielle du processus vectoriel $\vec{X}_t$ sont les fonctions propres de l'opérateur intégral de $L^2_V(T)$ dont le noyau est $[C(t,t)]^{-1} \; C(t,s)$.

En notant $\overrightarrow{f_k(t)} \; z_k$ les éléments principaux successifs classés par ordre décroissant des valeurs propres λ_k , on obtient la formule de reconstitution suivante du processus $\vec{X}_t$, en choisissant comme contrainte de normalisation;

$$\int_0^T < \overrightarrow{f(t)} ; C(t,t) \overrightarrow{f(t)} > dt = 1 \qquad \text{d'où} \qquad E(z^2) = \lambda$$

$$\boxed{\vec{X}_t = \sum_{k=1}^{\infty} z_k \; [C(t,t)]^{-1} \; \overrightarrow{f_k(t)}} \quad \text{en m,q, pour tout t}$$

L'analyse harmonique vectorielle consiste donc dans la décomposition précédente en variables aléatoires scalaires indépendantes du temps et fonctions vectorielles du temps.

On peut montrer la propriété suivante qui généralise la définition de l'analyse canonique généralisée :

Proposition 2.2

Les composantes z rendent maximale, sous contrainte d'orthogonalité, la quantité

$$\int_0^T R^2(z;\vec{X}_t)\, dt$$

où $R^2(z;\vec{X}_t)$ est le coefficient de détermination entre z et $\vec{X}_t$.

Cette propriété généralise également celle de l'analyse factorielle normée d'un processus scalaire (cf. Ch. I § II-3) et se démontre de la même façon.

II-2-b) Processus scalaires propres

Les processus scalaires $(\xi_t^{(k)})$ définis par $\xi_t^{(k)} = < \vec{X}_t ; \overrightarrow{f_k}(t) >$ jouissent de propriétés intéressantes.

Tout d'abord en vertu des définitions des z et des $\vec{f}(t)$ on a :

$$\xi_t^{(k)} = \frac{1}{\lambda_k} A_t\, z$$

$$z_k = \int_0^T \xi_t^{(k)}\, dt$$

ξ_t est donc à λ près l'approximation au sens des moindres carrés de la variable z par $\vec{X}_t$.

La propriété n°7 de l'analyse canonique généralisée s'énonce alors comme suit :

Proposition 2.3

Le processus $(\xi_t^{(1)})$ est solution du problème suivant : trouver un processus scalaire, combinaison linéaire à tout instant des composantes de $\vec{X}_t$ tel que l'analyse harmonique réduite de ce processus fournisse une première valeur propre maximale.

La propriété n°4 de l'analyse canonique généralisée s'étend également à l'analyse harmonique vectorielle en considérant l'opérateur intégral de noyau $A_t A_s$, opérateur agissant sur les processus aléatoires scalaires sur $[0,T]$ (opérateur de $L^2(\Omega \times T)$ dans lui-même). Le processus (ξ_t) est alors fonction propre de cet opérateur :

Proposition 2.4

$$\int_0^T A_t \, A_s \, \xi_s \, ds = \lambda \, \xi_t$$

On remarque d'abord que les solutions de cette équation intégrale sont nécessairement telles que $\xi_t = A_t \, \xi_t$ donc que ξ_t est combinaison linéaire des composantes de $\vec{X}_t$ (ou plus généralement une fonction linéaire de $\vec{X}_t$). Il suffit pour cela de multiplier par A_t les deux membres de l'équation et d'utiliser la relation $A_t^2 = A_t$.

L'équation se réduit donc à $\int_0^T A_t \, \xi_s \, ds = \lambda \, \xi_t$

soit $A_t \int_0^T \xi_s \, ds = \lambda \, \xi_t$ d'où en posant $\int_0^T \xi_s \, ds = x$

$$A_t \, x = \lambda \, \xi_t$$

et en intégrant sur t :

$$\int_0^T A_t \, x = \lambda \int_0^T \xi_t \, dt = \lambda \, x$$

ce qui prouve que x est bien égal à z fonction propre de $\int_0^T A_t \, dt$, ce qui achève la démonstration.

On a également

Proposition 2.5

Les $(\xi_t^{(k)})$ sont des processus de variance totale unité :

$$\int_0^{\bar{T}} E(\xi_t^2)\, dt = 1$$

Comme $\xi_t = \frac{1}{\lambda} A_t\, z$ on a $E(\xi_t^2) = \frac{1}{\lambda^2} E(A_t z)^2 = \frac{1}{\lambda^2} \operatorname{cov}(z;A_t z)$

car A_t est un projecteur orthogonal.

D'où en intégrant sur t :

$$\int_0^T E(\xi_t^2)dt = \frac{1}{\lambda^2} \operatorname{cov}(z; \int_0^T A_t\, z\, dt) = \frac{1}{\lambda^2} \operatorname{cov}(z\, ;\, \lambda\, z\,) = \frac{1}{\lambda} V(z)$$

comme $V(z) = \lambda$ ceci démontre la propriété énoncée.

II-2-c) Remarques diverses

Remarque 1

Par rapport à la définition donnée en [17] où l'on n'utilisait pas de métrique, notre définition de l'analyse harmonique vectorielle revient à effectuer une A.C.P. de $(\vec{X}_t)$ avec une métrique spéciale: celle définie par $M(t,t) = [C(t,t)]^{-1}$. Cette métrique s'impose pour diverses raisons :

tout d'abord son introduction est indispensable pour assurer la cohérence avec les méthodes connexes que sont l'analyse canonique généralisée et l'analyse harmonique qualitative (voir Chapitre III).

Ensuite elle seule donne des résultats invariants pour toute transformation linéaire sur les variables du type $\vec{Y}_t = L_t\, \vec{X}_t$ où L est une matrice à coefficients ne dépendant que de t : c'est donc la seule possibilité de traiter des données, où il n'existe pas de base naturelle de $\mathbb{V}$ du type fonction aléatoire et où la base varie à chaque instant.

Remarque 2

Les résultats de l'analyse harmonique vectorielle ne sont pas invariants pour des transformations autres que celles mentionnées précédemment ; en particulier pour les filtrages usuels où $\vec{Y}_t$ s'écrit $\int_0^T L(t,s)\,\vec{X}_s\,ds$ avec $L(t,s) \neq 0$ si $t \neq s$. On peut donc utiliser conjointement l'analyse harmonique vectorielle avec une métrique associée à une transformation non diagonale des données. Nous n'expliciterons pas les équations auxquelles on aboutit pour ne pas surcharger inutilement cet exposé. Il faut simplement retenir de ceci le fait que le choix de la métrique $[C(t,t)]^{-1}$ n'est pas aussi restrictif qu'on pourrait le croire:de même qu'en A.C.P. on peut effectuer des transformations linéaires sur les données autres qu'une homothétie puis réduire les variables obtenues.

Remarque 3

L'opérateur $\int_0^T A_t\,dt$ est pour les processus vectoriels la généralisation de l'opérateur d'ESCOUFIER W d'une processus scalaire avec métrique $D_{1/\sigma^2(t)}$.

Remarque 4

L'application à des données réelles, c'est-à-dire à n trajectoires ne soulève pas de problèmes différents de ceux évoqués au Chapitre I : même possibilités d'interprétation, même genre de sorties graphiques.

Quant aux problèmes d'approximation numérique, c'est-à-dire de discrétisation de l'intervalle [0,T] ils sont de même nature, seule change la taille de la matrice à diagonaliser qui est multipliée par la dimension m du processus. Il en est de même pour l'interpolation, la fonction d'interpolation étant définie sur $\mathbb{R}^m$ au lieu de $\mathbb{R}$ ce qui ne pose pas de problème théorique nouveau mais seulement des problèmes pratiques.

II-3 L'analyse "harmonique" vectorielle d'un processus stationnaire

Si $(\vec{X}_t)$ est stationnaire la matrice de covariance $[C(t,s)]$ ne dépend que de la différence t-s et l'on a $[C(t,t)] = [C(0)] \quad \forall\, t$.

On peut alors se demander quels sont les rapports entre l'analyse factorielle de $(\vec{X}_t)$ et l'analyse spectrale de $(\vec{X}_t)$. Nous allons voir ici que la réponse est assez différente de celle apportée pour les processus scalaires et que l'analyse harmonique vectorielle n'est pas équivalente à l'analyse spectrale. La raison fondamentale est la suivante : l'analyse cherche une décomposition de $(\vec{X}_t)$ selon des variables aléatoires scalaires et des fonctions vectorielle du temps tandis que l'analyse spectrale consiste en une décomposition en fonctions scalaires du temps $(e^{i\omega t})$ et en variables aléatoires vectorielles.

L'analyse spectrale est en réalité une analyse du type n° 6 selon la terminologie du paragraphe III.

Examinons cependant ce que devient l'équation des facteurs :

$$\int_0^T [C(t,t)]^{-1} [C(t,s)] \vec{f}(s)\, ds = \lambda \vec{f}(t)$$

dans le cas d'un processus stationnaire et en raisonnant directement sur l'intervalle $]-\infty;+\infty[$:

$$\int_{-\infty}^{+\infty} [C(0)]^{-1} [C(t-s)] \vec{f}(s)\, ds = \lambda \vec{f}(t)$$

s'écrit :

$$[C(0)]^{-1} \int_{-\infty}^{+\infty} [C(t-s)] \vec{f}(s)\, ds = \lambda \vec{f}(t)$$

Si la matrice $[C(h)]$ est absolument sommable il existe une matrice de densité spectrale donnée par la formule :

$$[S(\omega)] = \frac{1}{2\Pi} \int_{-\infty}^{+\infty} [C(h)]\, e^{-i\omega h}\, dh$$

Il est alors facile de montrer la proposition suivante

Proposition 2-6

Les fonctions propres généralisées de l'analyse "harmonique" vectorielle d'un processus stationnaire sont de la forme :

$$\vec{f}_\omega(t) = e^{i\omega t}\,\vec{v}(\omega)$$

où $\vec{v}(\omega)$ est un vecteur propre quelconque de la matrice $[C(0)]^{-1}[S(\omega)]$. A chaque valeur ω correspondent donc m valeurs propres $\lambda_\omega^{(i)}$ qui sont égales à 2Π fois les valeurs propres de $[C(0)]^{-1}\,[S(\omega)]$.

En effet en posant $\vec{f}(t) = e^{i\omega t}\,\vec{v}$ il vient :

$$[C(0)]^{-1}\int_{-\infty}^{+\infty}[C(t-s)]\;\vec{f}(s)\;ds = [C(0)]^{-1}\int_{-\infty}^{+\infty}[C(h)]\;e^{i\omega(t-h)}\,\vec{v}\;dh$$

$$= 2\Pi\,[C(0)]^{-1}\,[S(\omega)]\;e^{i\omega t}\,\vec{v}$$

Pour que cette expression soit de la forme $\lambda\,e^{i\omega t}\,\vec{v}$ il faut et il suffit que $\vec{v}$ soit vecteur propre de $[C(0)]^{-1}\,[S(\omega)]$ associé à la valeur propre λ_ω et on a alors $\lambda = 2\Pi\,\lambda_\omega$.

$[S(\omega)]$ qui est hermitienne possède m vecteurs propres orthogonaux

On a donc relié les valeurs propres de l'analyse harmonique vectorielle à celles d'une matrice fonction simple de la matrice de densité spectrale.

III- AUTRES METHODES DIRECTES D'ANALYSE

Un processus vectoriel peut être considéré formellement comme une fonction de trois indices ω, j, t où $\omega \in \Omega$ $j \in \{1,2,\ldots,m\}$ et $t \in [0,T]$. Or une analyse factorielle consiste en une décomposition d'une fonction de deux indices seulement (i.e. d'un tableau statistique à deux dimensions) selon des fonctions séparées de chacun d'eux. Il existe en fait 6 manières simples de ramener la dimension d'un tableau de 3 à 2 et l'analyse harmonique vectorielle n'est alors que l'analyse d'un des tableaux résultants. Il existe en effet 3 analyses du type "1 contre 1" conduisant à une décomposition selon des produits de fonctions séparées de deux des 3 indices et 3 analyses de type "1 contre 2" conduisant à une décomposition selon des produits de fonction d'un indice par une fonction des deux autres.

III-1 Les six analyses factorielles d'un cube

Nous reprendrons ici le schéma du début de ce chapitre afin de donner une illustration intuitive des pratiques possibles. Bien entendu les figures qui suivent n'ont de sens que pour des données finies mais les résultats mathématiques sont valables dans le cas général.

III-1-a) Compression sur un indice

Le plus simple est évidemment de supprimer l'un des indices en effectuant une sommation ce qui revient à effectuer en quelque sorte des analyses marginales.

α) Sommation sur T analyse n° 1

On obtient alors un vecteur aléatoire $\vec{X} = \int_0^T \vec{X}_t \, dt$, c'est-à-dire un ensemble de m variables aléatoires indépendantes de temps proportionnelles aux moyennes temporelles de chaque composante de $\vec{X}_t$.

$\frac{1}{T}\vec{X}$ fournit donc une situation de référence moyenne à laquelle on peut comparer les situations aux divers instants : il suffit d'effectuer l'analyse en composantes principales de $\vec{X}$ pour avoir une description moyenne des trajectoires et ensuite il est aisé de positionner la situation à une date t par projection en élément supplémentaire . Le choix de la moyenne étant d'ailleurs arbitraire on peut envisager d'autres combinaisons des $\vec{X}_t$ ce qui soulève le problème du choix optimal des pondérations de temps à adopter. Ceci sera examiné dans le paragraphe III-6.

Appliquée aux indicatrices d'un processus qualitatif l'analyse n° 1 et ses dérivées conduisent à des résultats particulièrement intéressants.

β) Sommation sur les individus analyse n° 2

On obtient alors uniquement la fonction $\vec{m}(t)$ c'est-à-dire une chronique multiple certaine à laquelle on peut appliquer éventuellement les méthodes d'analyses en composantes principales du chapitre I § I-4-b puisqu'il s'agit de m courbes sur [0,T]. Cependant on perd ici toute l'information apportée par la variabilité des individus et une telle analyse ne peut être que très fruste. Là encore cependant l'application au cas qualitatif se révèlera plus fructueuse que dans le cas numérique. Voir également le paragraphe III.2.

γ) Sommation sur les variables analyse n° 3

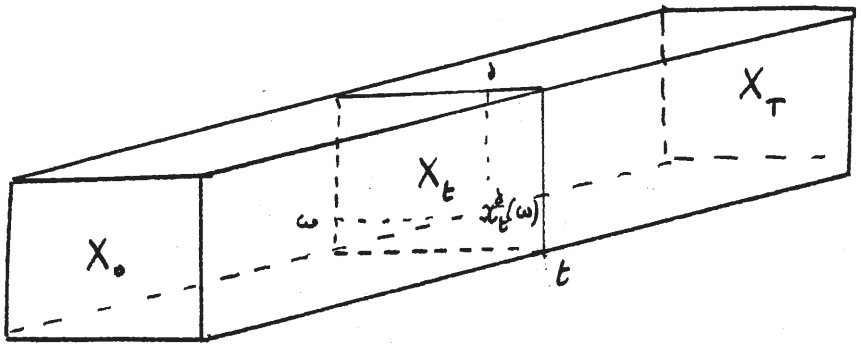
On obtient alors un processus scalaire moyen $(\vec{X}_t) = \frac{1}{m} \sum_{j=1}^{m} (X_t^j)$ dont on peut effectuer l'analyse en composantes principales. Telle que, cette analyse nous semble d'un intérêt très limité pour des variables numériques. Nous verrons de plus qu'elle ne peut s'appliquer au processus des indicatrices des états dans le cas qualitatif.

Cependant l'étude de combinaisons linéaires plus générales que la moyenne $\sum_{j=1}^{m} \alpha_j X_t^j$ se révèlera une méthode intéressante si l'on cherche les α_j pour optimiser un certain critère, voir plus loin § III-3-a .

III-1-b) Analyses du type 1 contre 2

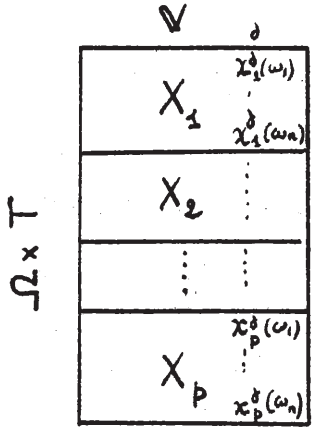
De façon imagée ces trois analyses reviennent à juxtaposer de différentes façons les tranches que l'on peut effectuer dans le "cube" de données.

Les tranches verticales ont ici la signification de coupes instantanées, les tranches horizontales contiennent les évolutions d'un individu ω , et les m tranches longitudinales sont les composantes de $(\vec{X}_t)$



α) Tranches frontales superposées analyse n° 4

Si $|\Omega| = n$ et que $t = t_1, t_2, \ldots, t_p$ ceci revient à constituer un tableau rectangulaire $n\,p \times m$ de la manière suivante



On trouve donc en colonnes les m variables et en lignes successivement les n individus à l'instant 1, les n individus à l'instant 2,...etc.

L'analyse en composantes principales d'un tel tableau revient donc à chercher des combinaisons linéaires à coefficients fixes dans le temps des variables $x^1, x^2, \ldots, x^m$ autrement dit des chroniques principales de variance maximale sur les individus et le temps, et des coordonnées des individus variables dans le temps.

Les combinaisons cherchées ou facteurs principaux de cette analyse s'obtiennent aisément comme vecteurs propres de la somme des matrices de covariance de chaque $\vec{X}_t$ puisque toutes les variables sont centrées à chaque instant. Cette analyse n'est autre que la méthode proposée par TENENHAUS et PRIEURET [62] après élimination de la tendance représentée par les centres de gravité. On remarque qu'il n'est pas nécessaire que les individus soient les mêmes aux différents instants d'où une certaine perte d'information.

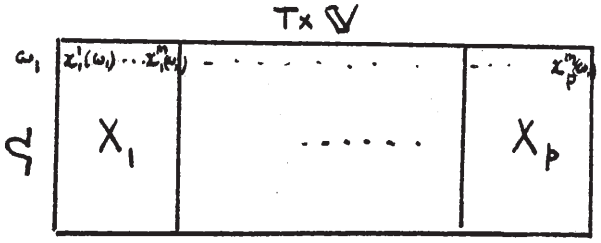
Pour un processus à temps continu on a de manière élémentaire la

> Proposition 2-7
>
> Les facteurs principaux de l'analyse n° 4 sont les vecteurs propres de la matrice de variance moyenne $C = \frac{1}{T}\int_0^T [C(t,t)]\,dt$.

Il s'agit donc d'une analyse des dépendances moyennes entre les composantes de $\vec{X}_t$. On peut compléter cette analyse en projetant en éléments supplémentaires sur le cercle des corrélations associé à C les diverses variables X_t^j .

β) Tranches frontales accolées analyse n° 5

Toujours dans le cas fini ceci revient à constituer un tableau à n lignes et mp colonnes comme suit :



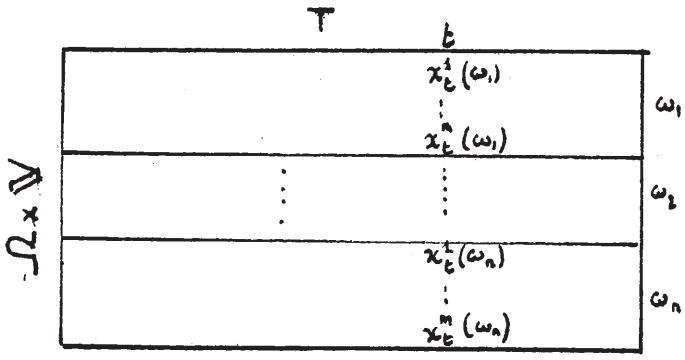
On cherche donc des combinaisons linéaires des variables différentes à chaque instant et des coordonnées fixes dans le temps des individus.

Il s'agit donc ici d'une généralisation vectorielle de l'analyse du chapitre I .

Le choix de la métrique $\mathrm{Diag}\,[C(t,t)]^{-1}$ conduit alors à l'analyse canonique généralisée, c'est-à-dire à l'analyse "harmonique" vectorielle du processus $(\vec{X}_t)$.

γ) Tranches longitudinales superposées analyse n° 6

Dans le cas fini ceci revient à constituer un tableau à p colonnes et m x n lignes comme suit :



Ce tableau est à des transpositions près le tableau obtenu en mettant en colonne les tableaux $X_1,\ldots,X_p$ vectorialisés.
On cherche des combinaisons linéaires des instants et des coordonnées des individus fixes dans le temps mais variables avec les caractères.

Tout se passe donc comme si les instants devenaient les objets de l'étude. Il est facile de vérifier que la "covariance" entre deux instants t et s pour ce tableau n'est autre que $\sum_{j=1}^{m} C^j(t,s)$ où C^j est l'opérateur de covariance de X_t^j .

Proposition 2-8

L'analyse factorielle n° 6 revient donc à étudier les éléments principaux de la somme des opérateurs de covariance de chaque composante du processus, c'est-à-dire l'opérateur intégral de noyau Trace [C(t,s)] .

On vérifie facilement que la juxtaposition des tranches longitudinales n'est autre que le tableau du paragraphe β après réarrangement des colonnes et que les tableaux associés aux tranches horizontales du cube redonnent soit l'analyse n° 4 soit l'analyse n° 6.

La seule analyse conduisant à une représentation comparable à celle de l'analyse factorielle d'un processus scalaire, c'est-à-dire avec des coordonnées fixes dans le temps et uniques des individus et des facteurs fonctions du temps est donc l'analyse n° 5 qui conduit à l'analyse harmonique vectorielle.

Remarquons pour terminer que les juxtapositions de tranches reviennent à remplacer l'opération somme (qui conduit aux analyses marginales) par l'union ensembliste et que l'on peut résumer ces résultats par le tableau :

Type d'analyse	ensemble réuni	opérateur analysé
4	∪ individus	$\int_0^T [C(t,t)]\,dt$
5	∪ variables	$\int_0^T A_t\,dt$
6	∪ temps	$\sum_{j=1}^{m} C^j$ = Trace C

III.2 Une méthode d'analyse de la tendance : la discrimination entre instants

Lorsqu'on observe n individus à p instants, l'analyse n°4 revient à considérer comme des objets différents les individus à l'instant t_1 , à l'instant t_2 ... Tout se passe donc comme si on avait p groupes de n individus, chaque groupe correspondant à une date différente.

Il est alors naturel de se demander en quoi ces groupes sont-ils différents, ce qui revient à "tester" le caractère évolutif du phénomène. Une des méthodes classiques en analyse des données pour comparer des groupes, est l'analyse factorielle discriminante qui revient à une analyse de la dispersion des individus selon une dispersion "intra" (autour des centres de chaque groupe) et une dispersion "inter" (entre centres de groupes). Appliquer au cas temporel l'analyse factorielle discriminante, reviendra donc à étudier les différences entre les moyennes temporelles $\overrightarrow{m(t_1)}$; $\overrightarrow{m(t_2)}$; ... $\overrightarrow{m(t_p)}$, autrement dit à analyser la tendance $\overrightarrow{m(t)}$ d'une certaine façon ; une telle pratique semble souhaitable avant le centrage des observations nécessaire aux autres méthodes présenter précédemment.

Pour des processus vectoriels généraux, il convient donc de généraliser à une infinité de groupes (de même effectif) l'analyse factorielle discriminante ce qui se fait sans difficulté de la manière suivante si t varie entre 0 et T :

En notant $\vec{m}$ la moyenne temporelle du processus : $\vec{m} = \frac{1}{T} \int_0^T \overrightarrow{m(t)}\, dt$ on définit alors la matrice de variance totale V , la matrice de variance intra-instants W et la matrice de variance inter-instants B par :

$$V = \frac{1}{T} \int_0^T E((\vec{X}_t - \vec{m}) \otimes (\vec{X}_t - \vec{m}))\, dt$$

$$W = \frac{1}{T} \int_0^T E((\vec{X}_t - \overrightarrow{m(t)}) \otimes (\vec{X}_t - \overrightarrow{m(t)}))\, dt = \frac{1}{T} \int_0^T [C(t,t)]\, dt$$

$$B = \frac{1}{T} \int_0^T (\vec{m}(t) - \vec{m}) \otimes (\vec{m}(t) - \vec{m})\, dt$$

On a évidemment $V = W + B$.

L'analyse factorielle discriminante consiste alors dans la recherche de processus discriminants $(\xi_t) = (<\vec{X}_t \; ; \vec{a}>)$ obtenus par combinaisons linéaires des composantes de $(\vec{X}_t)$ et ayant une variance inter-instant maximale sous la contrainte d'une variance totale fixe : en d'autres termes si $\vec{m} = \vec{0}$ on cherche à maximiser $\int_0^T (E(\xi_t))^2 \, dt$ sous la contrainte $\int_0^T E(\xi_t^2) \, dt = 1$.

On trouve alors aisément que la combinaison linéaire $\vec{a}$ (facteur discriminant) est solution du problème :

$$\max_{\vec{a}} \; <\vec{a} \; ; B \, \vec{a}> \qquad \text{avec} \qquad <\vec{a} \; ; V \, \vec{a}> = 1$$

et on a, si V est inversible, le résultat suivant :

$\vec{a}$ est vecteur propre de $V^{-1} B$ (et de $W^{-1} B$) associé à sa valeur propre maximale.

On peut alors appliquer les résultats usuels de l'analyse factorielle discriminante.

Les pouvoirs discriminants $\lambda_1, \lambda_2, \ldots \lambda_m$ sont d'autant plus élevés que les différences entre $\overrightarrow{m(t)}$ sont grandes, donc que l'évolution en moyenne est plus marquée.

Bien qu'ayant le même point de départ que l'analyse n°4, l'analyse discriminante proposée ici en est fondamentalement différente puisqu'il s'agit d'étudier la tendance et non la structure moyenne des dépendances représentée par W. En tant qu'A.C.P. des centres de gravité $\overrightarrow{m(t)}$, l'analyse discriminante se rapprocherait plutôt de la 1ère étape de la double A.C.P. proposée par J.M. BOUROCHE [10] mais la métrique n'est pas la même.

III-3 Problèmes d'optimisation

Toutes les méthodes factorielles proposées ici reviennent à chercher soit des processus combinaisons linéaires des composantes de $\vec{X}_t$ soit des variables combinaisons linéaires du processus $(\vec{X}_t)$.

Toutes ces combinaisons ont des propriétés d'optimalité en termes de variance maximale que nous allons développer maintenant et qui permettent d'envisager d'ailleurs d'autres techniques d'analyse.

III-3-a) Processus obtenus par combinaisons linéaires des composantes de $(\vec{X}_t)$

Nous avons vu au paragraphe II que l'analyse harmonique vectorielle revient à chercher un processus ξ_t du type $< \vec{f}_t \; ; \; \vec{X}_t >$ vérifiant une certaine propriété d'optimalité (λ_1 maximal) sous la contrainte suivante :

$$\int_0^T < \vec{f}_t \; ; \; [C(t,t)] \; \vec{f}_t > dt = 1$$

On peut alors chercher les solutions de problèmes analogues en modifiant le critère d'optimalité et les contraintes.

III-3-a-α Processus (ξ_t) obtenus par combinaisons variables

Le processus (ξ_t) étant toujours de la forme

$$\xi_t = < \vec{f}_t \; ; \; \vec{X}_t >$$

on peut chercher à optimiser un autre critère que celui de la plus grande valeur propre, par exemple maximiser $\sum_{i=1}^{k} \lambda_i$. Malheureusement on ne peut obtenir de solution "analytique" à ce problème c'est-à-dire exprimer les $\vec{f}_t$ comme fonctions propres d'un opérateur intégral (sauf pour $k = \infty$ comme nous le verrons plus tard). On peut cependant obtenir une solution au moyen d'un algorithme inspiré de la méthode PRINQUAL de M. TENENHAUS pour des variables qualitatives [63] à laquelle nous renvoyons pour la démonstration de convergence :

à l'optimum,si on désigne par $z_1, z_2, \ldots, z_k$ les composantes principales de ξ_t , il est facile de voir que ξ_t est à tout instant la combinaison linéaire de $\vec{X}_t$ la plus proche des z_j en d'autres termes ξ_t s'obtient en effectuant à t fixé l'analyse canonique de $(z_1, z_2, \ldots, z_k)$ contre $\vec{X}_t$. On en déduit l'algorithme du type moindres carrés alternés suivant :

étape i :

a) on effectue l'A.C.P. de $(\xi_t^{(i)})$ réduite d'où k composantes principales $z_1^{(i)}; \ldots; z_k^{(i)}$

b) On effectue pour tout t l'analyse canonique des $z_k^{(i)}$ contre $(\vec{X}_t)$ d'où un premier facteur canonique $\vec{f}_t^{(i+1)}$. On pose alors $\xi_t^{(i+1)} = < \vec{X}_t \; ; \; \vec{f}_t^{(i+1)} >$ etc.

la solution de départ $\vec{f}_t^{(0)}$ pouvant être prise égale au premier facteur de l'analyse harmonique vectorielle de $(\vec{X}_t)$.

Si on augmente k la solution de ce problème devient indéterminée puisque $k \to \infty \; \Sigma \lambda_i \to \mathrm{Trace}(\xi_t)$ et en analyse réduite on a évidemment $\sum_{i=1}^{\infty} \lambda_i = T$ quelque soit ξ_t .

Pour donner un sens à la recherche d'un processus (ξ_t) de variance totale maximale il faut donc relacher les contraintes sur $\vec{f}_t$ et supposer par exemple que l'on a seulement :

$$\int_0^T < \vec{f}_t \; ; \; \vec{f}_t > dt = 1$$

Dans ces conditions :

$$\int_0^T E(\xi_t^2) dt = \int_0^T < \vec{f}_t \; ; \; [C(t,t)] \; \vec{f}(t) > dt$$

pour que cette quantité soit maximale on trouve aisément que, pour tout t, $\vec{f}(t)$ doit être vecteur propre de $[C(t,t)]$ associé à la plus grande valeur propre

III-3-a-β Processus (ξ_t) obtenu par combinaison fixe

Il peut sembler intéressant de rechercher des combinaisons linéaires à coefficients fixes dans le temps des composantes de $\vec{X}_t$: l'avantage est alors de pouvoir interpréter ξ_t de la même façon quelque soit t .*

Posons donc $\xi_t = \langle \vec{a} \; ; \vec{X}_t \rangle = \sum_{j=1}^{m} a_j X_t^j$.

L'opérateur de covariance de (ξ_t) est donné par :

$$E(\xi_t \, \xi_s) = \langle \vec{a} \; ; [C(t,s)] \, \vec{a} \rangle \; .$$

On peut ainsi chercher la combinaison linéaire $\vec{a}$ de telle sorte que (ξ_t) ait une première composante principale de variance maximale en imposant à $\vec{a}$ une contrainte de normalisation. En d'autres termes ceci revient à chercher une combinaison linéaire des X_t^j à coefficients invariables qui soit la plus proche possible d'un processus quasi-déterministe, c'est-à-dire la plus unidimensionnelle possible.

$\vec{a}$ est alors solution du problème suivant :

$$\max_{\vec{a}} \; \max_{f} \iint f(t) \langle \vec{a} \; ; [C(t,s)] \, \vec{a} \rangle f(s) \, dt \, ds \quad (P)$$

avec $\|\vec{a}\| = 1$ et $\|f\| = 1$

Nous n'avons pu ramener la recherche de $\vec{a}$ à une expression simple. Cependant les propriétés de la solution conduisent à un algorithme alterné facile à mettre en oeuvre : en effet $\langle \vec{a} \; ; [C(t,s)] \, \vec{a} \rangle$ est le noyau de l'opérateur de covariance de $(\xi_t) = \langle \vec{X}_t ; \vec{a} \rangle$ et $\iint f(t) \, [C(t,s)] \, f(s) \, dt \, ds$ la matrice de covariance des m variables aléatoires $Y^j = \int_0^T X_t^j \, f(t) \, dt \quad j=1,2,\ldots,m$.

A f(t) fixé le maximum sur $\vec{a}$ de l'expression (P) est obtenu pour $\vec{a}$ vecteur propre de rang 1 de la matrice de covariance des Y^j . A $\vec{a}$ fixé f(t) est le premier facteur principal du processus $\xi_t = \langle \vec{X}_t ; \vec{a} \rangle$.

* Il s'agit donc ici d'une modification de l'analyse n° 3.

On atteindra donc un optimum (éventuellement local) en alternant recherche de f et de $\vec{a}$ (en partant d'un f ou d'un $\vec{a}$ arbitraire) puisqu'à chaque étape l'expression (P) ne peut décroître.

Chercher un processus $\xi_t = \langle \vec{X}_t ; \vec{a} \rangle$ dont la première composante principale ait une variance maximale n'aboutit donc pas à un résultat très commode et ne se réduit donc pas, en tout état de cause, à l'une des 6 méthodes d'analyse d'un cube de données présentées au paragraphe I. Or parmi ces 6 méthodes, deux fournissent des facteurs principaux indépendants du temps : la méthode marginale n° 1 qui élimine le temps par intégration et la méthode n° 4 .

Les résultats suivants précisent les propriétés d'optimalité des processus (ξ_t) que l'on peut en déduire.

<u>Proposition 2-8</u>

Le vecteur $\vec{a}$ normé tel que $\xi_t = \langle \vec{X}_t ; \vec{a} \rangle$ ait une moyenne temporelle de variance maximale est le premier facteur principal des $\int_0^T X_t^j \, dt$ et coïncide donc avec la solution de l'analyse n° 1 .

En effet rendre maximale la variance de $\frac{1}{T} \int_0^T \xi_t \, dt$ s'écrit

$$\max_{\vec{a}} V \left(\int_0^T \sum_{j=1} X_t^j a_j \, dt \right) = \max_{\vec{a}} V \left(\sum_j a_j \int_0^T X_t^j \, dt \right)$$

il est donc plus facile de maximiser la variance de la moyenne des ξ_t que celle de la première composante principale.

Rendre maximale la somme des variances des k premières composantes principales de (ξ_t) relève d'un algorithme semblable à celui proposé précédemment pour k=1 (ou de celui du paragraphe II-3-a) par contre la maximisation de la variance totale de (ξ_t) conduit à un résultat simple.

<u>Proposition 2-9</u>

Le vecteur $\vec{a}$ normé tel que $\int_0^T E(\xi_t^2) \, dt$ soit maximal avec $\xi_t = \langle \vec{X}_t ; \vec{a} \rangle$ est le premier vecteur propre de la matrice $\int_0^T [C(t,t)] \, dt$

En effet $E(\xi_t^2) = \langle \vec{a} \; ; [C(t,t)] \; \vec{a} \rangle$ et le maximum de

$$\int_0^T \langle \vec{a} \; ; [C(t,t)] \; \vec{a} \rangle \, dt = \langle \vec{a} \; ; \int_0^T [C(t,t)] \, dt \; \vec{a} \rangle$$

sous la contrainte $\|\vec{a}\| = 1$ est obtenu pour $\vec{a}$ vecteur propre de rang 1 de la matrice considérée, les vecteurs propres de rang suivants correspondant aux autres extrema de la variance totale de ξ_t .

On retrouve ici la méthode d'analyse n° 4 ou méthode de TENENHAUS-PRIEURET [62] qui est plus une analyse de la structure moyenne des dépendances à t fixé qu'une étude des dépendances entre instants.

Remarquons que le dernier vecteur propre de la matrice [C(t,t)] (combinaison linéaire de variance minimale) présente un intérêt certain puisqu'il fournit, si $\lambda_m = 0$, une relation certaine entre les composantes du processus indépendante du temps.

Pour conclure sur le choix entre combinaisons fixes dans le temps ou combinaisons évolutives, il est certes plus agréable de manier des combinaisons invariables qui s'interprètent mieux mais imposer cette forme a priori est une contrainte très forte qui revient à postuler par avance pour les plus grandes valeurs propres l'existence de facteurs de différenciation entre individus invariables dans le temps (et pour les plus petites de relations rigides entre les variables). Il vaut mieux, à notre avis, utiliser l'analyse harmonique vectorielle qui donne des combinaisons a priori variables : constater a posteriori que les facteurs $\vec{f}(t)$ sont en fait peu dépendants de t mettra alors véritablement en évidence un effet de structure, sans l'avoir imposé au préalable.

III-3-b) Vecteurs aléatoires déduits linéairement de $(\vec{X}_t)$

Considérons pour commencer un vecteur aléatoire $\vec{Y}$ de la forme $\vec{Y} = \int_0^T \vec{X}_t \, f(t) \, dt$ où f(t) est un élément de $L^2(T)$. Si $f(t) = \frac{1}{T}$, $\vec{Y}$ n'est autre que la moyenne temporelle de $\vec{X}_t$ et on retrouve alors l'analyse n° 1 . Cherchons f(t) de manière à optimiser un certain critère lié à l'A.C.P. de $\vec{Y}$.

La matrice de variance de $\vec{Y}$ s'obtient sans difficulté et on trouve

$$C_Y = E(\vec{Y} \otimes \vec{Y}) = \iint_{[0,T]^2} f(t)\ f(s)\ [C(t,s)]\ dt\ ds\ .$$

Si on recherche f tel que l'A.C.P. de $\vec{Y}$ conduise à une première composante principale de variance maximale et si le facteur principal associé est $\vec{a}$ il s'agit de résoudre le problème suivant

$$\max_f \max_{\vec{a}} < \vec{a}\ ;\ \iint f(t)\ f(s)\ [C(t,s)]\ dt\ ds\ \vec{a} >$$

On retrouve alors le problème déjà traité au sous-paragraphe précédent. L'optimisation de l'analyse marginale n° 1 sur la première valeur propre est alors identique à l'optimisation de la méthode n° 3 et ne conduit pas à une solution analytiquement simple.

Par contre si on cherche à maximiser la variance totale de $\vec{Y}$ on obtient un résultat simple.

En effet Trace $C_Y = \iint f(t)\ f(s)$ Trace $[C(t,s)]\ dt\ ds$ et le maximum de cette quantité pour f normé est obtenu pour f fonction propre de l'opérateur intégral de noyau Trace $[C(t,s)]$. On peut donc énoncer la

<u>Proposition 2-10</u>

Les facteurs sur t de la méthode d'analyse n° 6 rendent extrêmale la variance totale du vecteur aléatoire $\vec{Y} = \int_0^T \vec{X}_t\ f(t)\ dt$.

Notons que l'on n'a pas nécessairement f positif ce qui limite les possibilités d'interprétation de f(t) comme pondération du temps. L'analyse des opérateurs associés aux instants qui est présentée dans le paragraphe suivant ne présentera pas cet inconvénient.

La fonction f définit la même combinaison des temps pour toutes les composantes de $\vec{X}_t$. Cherchons pour terminer des fonctions $f_j(t)$ propres à chaque X_t^j . Le vecteur aléatoire $\vec{Y} = (Y^1, Y^2, \ldots, Y^m)$ où $Y^j = \int_0^T X_t^j\ f_j(t)\ dt$ aune matrice de variance C_Y de terme général :

$$E(Y^j\ Y^k) = \iint_{[0,T]^2} f_j(t)\ E(X_t^j\ X_s^j)\ f_k(s)\ dt\ ds$$

La recherche des f_j normés tels que C_Y ait une première valeur propre maximale n'aboutit pas à une solution simple tandis que la maximisation de la trace de C_Y revient à prendre pour $f_j(t)$ les premiers facteurs de l'A.C.P. de chaque X_t^j c'est-à-dire à effectuer en analyses factorielles séparées.

IV- ANALYSES DES OPERATEURS ASSOCIES A UN PROCESSUS VECTORIEL$(\vec{X}_t)$

L'opérateur W associé à un processus scalaire(X_t)et défini au chapitre I est tel que :

$$W(\xi) = \int_0^T X_t\ E(X_t\ \xi)\ dt$$

si on a choisi la métrique I. Cet opérateur est caractéristique du processus (X_t)en ce sens que ses valeurs propres et fonctions propres sont les valeurs propres et les composantes principales de(X_t).

Lorsque l'on dispose d'un processus vectoriel il est alors naturel de comparer les analyses factorielles de chacune des composantes de ce processus en comparant les opérateurs associés, c'est la démarche de [23]. On peut également caractériser chaque instant t par l'opérateur O_t associé aux m variables aléatoires $X_t^1, X_t^2, \ldots, X_t^m$ tel que

$$z \to O_t\ z = \langle \vec{X}_t\ ;\ M_t\ E(\vec{X}_t\ z) \rangle \quad .$$

O_t est alors un opérateur de $L^2(\Omega)$ dans lui-même où M_t désigne la métrique adoptée à l'instant t sur l'espace des "individus" $\mathbb{R}^m$ ou $\mathbb{V}$.

Nous proposons alors d'effectuer une analyse harmonique du processus aléatoire O_t ce qui est une manière globale de tester le caractère temporel de$(\vec{X}_t)$.

IV-I Analyse des m opérateurs associés aux composantes de$(\vec{X}_t)$

Chaque opérateur W_i étant de Hilbert-Schmidt et même nucléaire, le produit de deux quelconques d'entre eux est nucléaire.

On sait alors [23] que Trace $(W_i\ W_j)$ définit donc un produit scalaire sur l'ensemble des opérateurs associés à des processus de second ordre sur $[0,T]$. Il est alors possible de définir une distance entre processus au moyen de la distance entre opérateurs, distance qui tient compte du comportement total du processus.

Deux types d'analyses globales sont alors possibles :

. une analyse en composantes principales des opérateurs par diagonalisation de la matrice m x m des produits scalaires Trace $(W_i\ W_j)$ (opérateurs "variables")

. Une analyse factorielle du tableau de distances entre opérateurs $\| W_i - W_j \|^2$ (opérateurs "individus")

ces deux analyses conduisent à des représentations graphiques (positionnement multidimensionnel) des positions des opérateurs ce qui permet de repérer dans une certaine mesure (car cette analyse est très globale) les processus ayant des comportement semblables ou opposés,

La diagonalisation de la matrice des produits scalaires entre opérateurs fournit un premier vecteur propre dont les coefficients α_i tous de même signe ; peuvent être choisis positifs (car il est démontré en [23] que Trace $W_i\ W_j \geqslant 0$).

La combinaison linéaire des opérateurs associée à ce premier vecteur propre $\sum_{i=1}^{m} \alpha_i\ W_i$ est encore un opérateur associé à un processus aléatoire sur $[0,T]$, opérateur réalisant un compromis entre les W_i différent de la moyenne.

On peut alors en effectuant l'analyse spectrale de cet opérateur $\Sigma\ \alpha_i\ W_i$ en déduire la décomposition de Karhunen-Loeve du processus scalaire qu'il représente : en termes statistiques si on dispose de n individus on aura ainsi une représentation des individus intermédiaire entre les m représentations possibles à partir de chaque analyse factorielle tenant compte des ressemblances majoritaires entre processus.

Remarque 1

L'opérateur "principal" $\Sigma\alpha_i W_i$ n'est pas associé au processus $(\Sigma\alpha_i X_t^i)$.

Remarque 2

Une telle analyse n'a de sens que si il existe des composantes naturelles pour le processus et n'est pas valable pour une fonction aléatoire à valeurs dans V car les résultats dépendent évidemment de la base de V choisie.

Remarque 3

On peut faire usage de métriques éventuellement différentes pour chaque processus: il suffit de prendre la définition généralisée de l'opérateur d'Escoufier vue au chapitre I :

$$W = U\ M\ U^*$$

IV-2 Opérateurs associés aux instants

Remplacer la variable aléatoire $\vec{X}_t$ par l'opérateur O_t tel que $O_t\ x = \vec{X}_t\ M_t\ E(\vec{X}_t\ x)$ revient à définir une nouvelle variable aléatoire à valeurs dans l'espace de Hilbert des opérateurs de Hilbert-Schmidt de $L^2(\Omega)$ dans lui-même. On sait qu'il est alors possible d'effectuer l'analyse factorielle du processus O_t car celui-ci vérifie la condition :

$$\| O_t \|^2\, dt < \infty$$

où $\| O_t \|^2 = \text{Trace}\ O_t^2$ car O_t , autoadjoint, est de rang fini m .

Le produit scalaire entre opérateurs audo-adjoints

$$< O_i\ ;\ O_j > = \text{Trace}\ (O_i\ O_j)$$

définit l'équivalent d'une covariance entre variables et il est alors facile d'effectuer l'analyse factorielle du processus O_t : il suffit d'adapter l'analyse factorielle d'un processus scalaire étudiée au Chapitre I .

Définition

L'analyse factorielle du processus O_t consiste dans la recherche des fonctions propres de l'opérateur $\mathcal{E}$ de $L^2(T)$ dans lui-même définies par :

$$\mathcal{E}\,\alpha(t) = \int_0^T < O_t \; ; \; O_s > \alpha(s)\, ds = \lambda\, \alpha(t)$$

A chaque fonction propre $\alpha_k(t)$ correspond un opérateur de Hilbert-Schmidt de $L^2(\Omega)$ dans lui-même défini par $F_k = \int_0^T O_t\, \alpha_k(t)\, dt$.

On a alors la formule de reconstitution suivante :

$$O_t = \sum_{k=1}^{\infty} F_k\, \alpha_k(t)$$

où les F_k sont des opérateurs ne dépendant pas du temps et les $\alpha_k(t)$ des fonctions fixes du temps.

Quel peut être l'intérêt d'une telle analyse ?

1°) Si $M_t = I_m \;\; \forall_t$ cela revient à étudier s'il existe une évolution des variances et covariances entre les m variables prises à l'instant t .

2°) Si $M_t = D_{1/\sigma_t^2} \;\; \forall$ t cela revient à étudier s'il existe ou non une déformation des corrélations entre variables à travers le temps.

Dans ces deux cas l'absence d'évolution se traduit pour une première valeur propre très élevée (100% de variance expliquée) associée à un facteur quasi-constant.

3°) D'une façon générale quel que soit M_t ceci veut dire qu'on analyse la déformation éventuelle des distances entre individus au cours du temps, d'où la possibilité d'analyser, par cette technique, des données du type : tableaux de distances entre individus relevées à différents instants.

En reprenant une terminologie introduite dans [41] on effectue ainsi une analyse de l'interstructure entre les variables $\vec{X}_t$.

Les produits scalaires entre opérateurs de Hilbert-Schmidt auto-adjoints ayant la propriété d'être toujours positifs on a à nouveau la même propriété pour le premier opérateur principal F_1 : ses fonctions propres fourniront comme au paragraphe précédent des variables, autrement dit des coordonnées des individus en dimension finie, représentation de la structure principale (aux sens des corrélations instantanées si $M_t = D_{1/\sigma_t^2}$) des individus au cours du temps.

Une mention particulière doit être faite pour le cas $M_t = [C(t,t)]^{-1}$ pour lequel O_t n'est autre que l'opérateur de projection A_t : un tel choix de métrique revient, comme il a déjà été écrit, à étudier l'évolution des espaces engendrés pour les composantes de $\vec{X}_t$, Le produit scalaire $\mathrm{Trace}(A_t A_s)$ n'est alors autre que la somme des carrés des coefficients de corrélation canonique entre $\vec{X}_t$ et $\vec{X}_s$. On a une méthode complémentaire de l'analyse harmonique vectorielle dans laquelle on effectue directement l'analyse de l'opérateur $\int_0^T A_t \, dt$. Ici on a en plus la possibilité d'étudier de manière fine les distances entre sous-espaces, représentées par les distances entre opérateurs A_t et A_s . La question se pose alors dans cette optique du choix entre $\int_0^T A_t \, dt$ et $\int_0^T \alpha_1(t) \, A_t \, dt$ pour obtenir une représentation des individus :

le premier opérateur $\int_0^T A_t \, dt$ est lié à une représentation d'une structure "moyenne"

le deuxième opérateur $\int_0^T \alpha_1(t) \, A_t \, dt$ à une structure "principale".

On retrouve ici un débat classique sur le meilleur résumé d'un ensemble de variables la moyenne ou la première composante principale, bien qu'ici le contexte soit très différent. Pondérer les A_t par des coefficients $\alpha_1(t)$ revient à utiliser implicitement une mesure positive non uniforme sur $[0,T]$.

L'analyse des proximités entre projecteurs A_t , ou plutôt du processus des projecteurs A_t , est d'ailleurs la seule qui puisse être utilisée si il n'existe pas de système de référence naturel pour les coordonnées de $\vec{X}_t$ (fonctions aléatoires à valeur dans $\mathbb{V}$) .

CONCLUSION

Parmi les diverses méthodes possibles d'analyse factorielle d'un processus à valeur dans $\mathbb{R}^m$ seule l'analyse "harmonique" vectorielle possède des propriétés qui en font la généralisation naturelle de méthodes connues pour le cas où T est discret, et de l'analyse factorielle d'un processus scalaire.

Ce n'est évidemment pas pour cela la seule méthode à utiliser sur ce genre de données car le choix d'une méthode est lié aux objectifs poursuivis et aux propriétés de la méthode.

Il faut bien voir ici que l'analyse harmonique vectorielle qui est une méthode d'analyse des trajectoires est essentiellement une étude des corrélations entre instants (structure "inter") et néglige l'étude des corrélations à un instant donné (structure "intra"). Il est clair que si on s'intéresse essentiellement aux déformations de la structure de corrélation à travers le temps cette analyse n'est pas la plus pertinente et il vaut mieux recourir soit à l'analyse n° 4 du type TENENHAUS-PRIEURET, soit à une étude de l'évolution des opérateurs O_t associés à la structure de corrélation à t fixé c'est-à-dire avec $M_t = I$ ou $M_t = D_{1/\sigma_t^2}$; l'analyse de la tendance pouvant être effectuée par discrimination entre instants.

CHAPITRE III

ANALYSE FACTORIELLE D'UN PROCESSUS QUALITATIF

Ce chapitre est consacré à l'analyse d'un processus qualitatif (X_t), c'est-à-dire que pour tout t de [0,T] X_t est une variable qualitative à m modalités ; en d'autres termes on se donne un processus à m états.

Nous développons ici et complètons les résultats établis avec J.C. DEVILLE en [20] concernant l'analyse harmonique qualitative : la méthode factorielle proposée, extension naturelle des analyses d'un processus scalaire ou vectoriel y apparaît comme la généralisation de l'analyse des correspondances multiples à un continuum de tableaux disjonctifs. D'autres travaux récents (R. BOUMAZA [8]) consacrés à l'analyse d'un processus qualitatif ont abouti aux mêmes résultats en partant de principes différents, apportant ainsi une confirmation théorique à notre approche. Les premières applications pratiques réalisées à l'INSEE [26] ont montré la capacité de la méthode à traiter facilement des données concrètes.

Conçue essentiellement pour l'analyse d'un processus dont les états ne sont pas des nombres (évolution du statut professionnel, de l'état matrimonial, d'une opinion ou d'un comportement) il est clair que, moyennant un découpage en classes, tout processus scalaire peut être analysé comme un processus qualitatif : cette technique permet comme on le sait depuis les travaux de M. MASSON [43] de dépasser le cadre linéaire de l'analyse en composantes principales.

Bien qu'il soit possible d'étendre la méthode proposée au cas d'un nombre infini, même non dénombrable d'états ([19]) nous nous restreindrons au cas où le nombre d'états m est fini ce qui assure la possibilité de

l'analyse harmonique qualitative dans la quasi-totalité des cas. Après une étude de diverses analyses complémentaires d'un processus qualitatif nous étendons nos méthodes aux cas d'un processus qualitatif multiple.

I- L'ANALYSE "HARMONIQUE" QUALITATIVE

I-1 Hypothèses et notations

Les modalités ou états du processus qualitatif (X_t) seront notés x où x prend ses valeurs dans un ensemble fini de cardinal m noté M.

On notera $p^x(t)$ la probabilité que $X_t = x$ à l'instant t et $p^{x,y}(t,s)$ la probabilité $P(X_t = x \cap X_s = y)$.

En notant $\mathbb{1}_t^x$ la variable aléatoire indicatrice de l'état x à l'instant t on se définit un processus vectoriel à m dimensions par :

$$\vec{\mathbb{1}}_t = (\mathbb{1}_t^1, \mathbb{1}_t^2 ; \dots ; \mathbb{1}_t^m)$$

on a évidemment : $E[\mathbb{1}_t^x] = p^x(t)$

et $E[\mathbb{1}_t^x \mathbb{1}_s^y] = p^{x,y}(t,s)$

Il est courant dans l'analyse des données qualitatives d'utiliser le concept de codage ou de quantification qui revient à transformer une variable qualitative en une variable numérique en attribuant des valeurs numériques aux diverses modalités. Ici les codes des modalités pouvant varier avec le temps on appellera codage du processus (X_t) une famille f_t de fonctions de M dans $\mathbb{R}$, En notant $\vec{a}(t)$ le vecteur défini par

$$\vec{a}(t) = \begin{pmatrix} a_1^1(t) \\ \vdots \\ a_1^m(t) \end{pmatrix}$$

on a les relations suivantes concernant le processus scalaire $Y_t = f_t(X_t)$ (ou processus codé) :

$$f_t(X_t) = \sum_{j=1}^{m} a^j(t)\, \mathbb{1}_t^j = \langle \overrightarrow{a(t)} ; \vec{\mathbb{1}}_t \rangle.$$

A la variable aléatoire qualitative X_t est associé un opérateur de conditionnement E_t défini par :

$$E_t(\xi) = E(\xi / X_t) \qquad \xi \in L^2(\Omega)$$

C'est-à-dire que $E_t(\xi)$ est la variable aléatoire espérance conditionnelle de ξ sachant X_t : en d'autres termes $E_t(\xi)$ est l'espérance de ξ conditionnellement à la sous-tribu $\mathcal{B}_t$ de Ω engendrée par (X_t). L'expression explicite de E_t est donnée par la formule :

$$E_t(\xi) = \sum_{j=1}^{m} \mathbb{1}_t^j\, E(\xi / X_t = j)$$

et en supposant $p^j(t) \neq 0 \ \forall\, t$;

$$\boxed{E_t(\xi) = \sum_{j=1}^{m} \mathbb{1}_t^j \frac{E(\xi, \mathbb{1}_t^j)}{E(\mathbb{1}_t^j)}}$$

Cette expression est de la forme $\langle \vec{a}(t) ; \vec{\mathbb{1}}_t \rangle$: au moyen de l'espérance conditionnelle on réalise un codage de X_t à partir de ξ.

La variable aléatoire $E_t(\xi)$ est alors le codage de (X_t) le plus proche possible de ξ au sens habituel de la norme de $L^2(\Omega)$. E_t qui est le projecteur de $L^2(\Omega)$ sur $L^2(\Omega; \mathcal{B}_t)$ espace des v.a. $\mathcal{B}_t$-mesurables et de carré intégrable se confond avec le projecteur A_t sur l'espace vectoriel engendré par les indicatrices $\mathbb{1}_t^j$ des m modalités, ou espace des codages de (X_t) à l'instant t.

Pour effectuer l'analyse du processus (X_t) diverses possibilités sont envisageables :

. coder le processus et faire l'analyse en composantes principales de $Y_t = f_t(X_t)$ mais quel codage prendre et quelle métrique choisir pour Y_t ? Comment définir un codage meilleur qu'un autre ?

. Appliquer les résultats de l'analyse d'un processus vectoriel au processus des indicatrices $(\vec{\mathbb{1}}_t)$.

. Introduire un opérateur caractérisant l'évolution de (X_t) analogue à l'opérateur de covariance d'un processus scalaire.

C'est cette troisième voie que nous reprendrons et qui suit d'assez près la démarche utilisée au chapitre précédent.

Avant d'étudier les propriétés de cet opérateur précisons les hypothèses mathématiques concernant le processus.

Dans [20] nous supposions que le processus (X_t) était continu en probabilité c'est-à-dire que ;

$$P(X_{t+h} \neq X_t) \to 0 \qquad \text{si } h \to 0$$

hypothèse qui n'avait rien de minimale ainsi que le prouvent aisément certains exemples.

Dans [8] on montre d'ailleurs que l'on peut affaiblir cette hypothèse en supposant simplement la mesurabilité des v.a. indicatrices $\mathbb{1}_t^x$ sur $T \times \Omega$, hypothèse vérifiée par la quasi-totalité des processus.

Le seul intérêt de l'hypothèse de continuité en probabilité est d'assurer tout d'abord la continuité en moyenne quadratique des $\mathbb{1}_t^x$ et de vérifier ainsi les hypothèses faites pour les processus vectoriels, et ensuite la continuité des fonctions de codage des modalités $a^x(t)$ d'où celle, en moyenne quadratique, des processus codés (Y_t) obtenues par l'analyse. Par pure commodité nous reprendrons cette hypothèse de continuité en probabilité étant entendu que les résultats obtenus sont encore valables pour de plus larges catégories de processus et en tout état de cause dans le cas où Ω est fini (aucune hypothèse n'est alors nécessaire).

I-2 La méthode

On peut admettre qu une variable qualitative est équivalente aux conditionnements qu'elle permet c'est-à-dire à son opérateur d'espérance conditionnelle qui est hermitien, idempotent et de rang au plus égal à m nombre de modalités (il est donc nucléaire) ; en effet on sait qu'inversement tout opérateur ayant ces propriétés est une espérance conditionnelle relative à une tribu engendrée par une partition de taille m au plus donc à une variable qualitative [48].

Le produit $E_t E_s$ résume donc la dépendance entre X_t et X_s et l'on sait que l'analyse de cet opérateur conduit à l'analyse canonique des tribus engendrées par X_t et X_s [16] en d'autres termes à l'analyse des correspondances de X_t et X_s.

Si on se limite, comme il est habituel, aux dépendances d'ordre deux, le processus (X_t) est entièrement résumé par la donnée de tous les $E_t E_s$ qui constituent l'analogue, mais en plus complexe, d'une fonction de covariance.

I.2.a) L'opérateur K et les processus propres

Soit $\mathcal{H}$ l'espace de Hilbert des processus scalaires de variance totale finie muni du produit scalaire de la covariance en moyenne :

$$(X_t | Y_t)_{\mathcal{H}} = \int_0^T E(X_t Y_t)\,dt$$

Soit alors l'opérateur K de $\mathcal{H}$ dans lui-même défini par :

$$\boxed{K\,\xi_t = \eta_t = \int_0^T E_t E_s(\xi_s)\,ds}$$

$E_t E_s$ agissant sur des variables aléatoires, K ainsi défini transforme bien un processus (ξ_t) en un autre processus (η_t).

On a les propriétés suivantes ;

. η_t est une variable $\mathcal{B}_t$-mesurable pour tout t

en effet $\eta_t = E_t \int_0^T E_s(\xi_s) d_s = E_t(x)$, Ceci prouve que le processus (η_t) est un codage de (X_t),

. Si (X_t) continu en probabilité (η_t) est continu en moyenne quadratique. En effet E_s étant un opérateur continu on peut mettre η_t sous la forme $\sum_{x=1}^{m} a^x(t) \mathbb{1}_t^x$ où les $a^x(t)$ sont des fonctions continues on a alors $E(\eta_t^2) = \sum_{x=1}^{m} (a^x(t))^2 p^x(t)$ et les $p^x(t)$ sont continues par hypothèse $E(\eta_t^2)$ est donc une fonction continue de t .

. Soit z une variable aléatoire et Y_t le codage de X_t associé à z : $Y_t = E_t(z)$.

Alors on a $E(Y_t Y_s) = E[E_t(z)E_s(z)] = E[z, E_t E_s(z)]$.

L'opérateur de covariance C de (Y_t) s'exprime alors par ;

$$C = E(z\ K\ z)$$

en effet par définition de C on a ;

$$C(f(t)) = \int_0^T E(Y_t Y_s) f(s) ds = E(z \int_0^T E_t E_s(zf(s)) ds)$$

$E_t\ E_s$ étant nucléaires K est hermitien, positif et nucléaire d'où

Proposition 3.1 :

Il existe une décomposition spectrale de K de la forme

$$K = \sum_{k=1}^{\infty} \lambda_k \, \xi^{(k)} \otimes \xi^{(k)}$$ où les $\xi^{(k)}$ sont des processus de variance totale unité, orthogonaux en moyenne sur $[0,T]$ et les λ_k les valeurs propres de K solution de l'équation

$$\boxed{\int_0^T K_{t,s} \, \xi_s \, ds = \lambda \, \xi_t}$$

où $K_{t,s} = E_t \, E_s$

Les solutions de cette équation sont les processus propres et les valeurs propres associés au processus qualitatif (X_t).

Le produit tensoriel $\xi^{(k)} \otimes \xi^{(k)}$ a l'interprétation suivante : c'est l'opérateur de rang 1 sur les éléments de $\mathcal{H}$ (les processus de variance totale finie) tel que pour un processus (ψ) de $\mathcal{H}$ on ait :

$$(\xi^{(k)} \otimes \xi^{(k)})(\psi) = (\xi^{(k)} | \psi)_{\mathcal{H}} \, \xi^{(k)}$$

soit en spécifiant l'instant t ;

$$(\xi^{(k)} \otimes \xi^{(k)})(\psi)_t = (\xi^{(k)} | \psi)_{\mathcal{H}} \, \xi_t^{(k)}$$

$$= \left(\int_0^T E(\xi_s^{(k)} \psi_s) ds \right) \xi_t^{(k)}$$

$\xi^{(k)} \otimes \xi^{(k)}$ est donc l'opérateur intégral de noyau $\xi_s^{(k)} \otimes \xi_t^{(k)}$ où le produit tensoriel est celui d'éléments de $L^2(\Omega)$.

Par identification avec la décomposition spectrale de K il vient donc ;

$$\boxed{K_{t,s} = E_t \cdot E_s = \sum_k \lambda_k \, \xi_t^{(k)} \otimes \xi_s^{(k)}}$$

On en déduit le résultat suivant ;

Proposition 3.2 : La trace de K vaut mT (au plus)

Démonstration :

On a $K_{t,t} = E_t = \sum_k \lambda_k \xi_t^{(k)} \otimes \xi_t^{(k)}$

d'où Trace $E_t = \Sigma \lambda_k$ Trace $(\xi_t^{(k)} \otimes \xi_t^{(k)})$

Comme $\xi_t^{(k)} \otimes \xi_t^{(k)}$ est l'opérateur de rang 1 qui transforme une variable aléatoire z en $E(\xi_t^{(k)}, z)\xi_t^{(k)}$ sa trace vaut $E(\xi_t^2)$ car ξ_t est son unique fonction propre.

On a donc Trace $E_t = \Sigma \lambda_k E((\xi_t^{(k)})^2)$

en intégrant sur t la relation précédente

$$\int_0^T \text{Trace } E_t \, dt = \Sigma \lambda_k \int_0^T E(\xi_t^{(k)})^2 \, dt = \Sigma \lambda_k = \text{Trace } K$$

car les $\xi_t^{(k)}$ sont de variance totale un .

Si tous les états du processus sont de probabilités non nulles à chaque instant Trace $E_t = m$ et on a bien Trace $K = m\,T$,
Sinon : Trace $K \leqslant m\,T$,

On voit aisément que les processus constants sont fonctions propres triviales de K associées à la valeur propre T ; on en conclut ;

, que les autres processus propres sont en moyenne nuls puisqu'orthogonaux aux constantes

$$\int_0^T E(\xi_t^{(k)}) \, dt = 0$$

, que la somme des valeurs propres non triviales vaut $(m-1)T$ (au plus).

Par ailleurs les processus $(\xi_t^{(k)})$ sont, en un sens qui sera défini plus loin, les meilleurs codages successifs, orthogonaux en moyenne, de (X_t).

I-2-b) Les génératrices des processus propres

L'équation donnant les processus propres s'écrit également ;

$$E_t \int_0^T E_s \, \xi_s \, ds = \lambda \, \xi_t$$

on en déduit en prémultipliant les deux membres par l'opérateur E_t que les processus propres (ξ_t) sont tels que :

$$\xi_t = E_t(\xi_t) \text{ puisque } (E_t)^2 = E_t \, .$$

Il vient donc $E_t \int_0^T \xi_s \, ds = \lambda \, \xi_t$.

Posons $z = \int_0^T \xi_s \, ds$. La variable z définie par cette intégrale stochastique, vérifie alors la propriété suivante :

Proposition 3.3 :

$z = \int_0^T \xi_s \, ds$ où (ξ_s) est un processus propre, est solution de l'équation :

$$\int_0^T E_t(z) \, dz = \lambda z$$

z est donc elle-même vecteur propre d'un opérateur égal à $\int_0^T E_t \, dt$.

La démonstration est immédiate :

On déduit de $E_t(z) = \lambda\, \xi_t$ par intégration sur t ;

$$\int_0^T E_t(z)dt = \lambda \int_0^T \xi_t\, dt = \lambda z$$

Les processus propres sont donc engendrés par conditionnement de la variable z puisque $\xi_t = \frac{1}{\lambda} E_t(z)$ ce qui justifie le nom de génératrice employé dans [20] pour la nommer.

L'opérateur $\int_0^T E_t\, dt$ est autoadjoint, positif, nucléaire et de trace m T au plus (si aucun état n'est impossible pour presque tout t).

On vérifie aisément que toute solution de :

$$\int_0^T E_t(z) = \lambda\, z$$

engendre une solution de :

$$\int_0^T E_t\, E_s\, \xi_s\, ds = \lambda\, \xi_s$$

et que la variable z = constante est solution triviale.
Il en découle une propriété intéressante ;

<u>Proposition 3.4</u> :
Le processus (ξ_t) est centré

En effet toutes les variables z non triviales sont centrées puisqu'orthogonales aux constantes, ξ_t l'est alors comme projection d'une variable centrée et n'est donc pas seulement centrée en moyenne mais pour tout t .

<u>Proposition 3.5</u> :
Les variables aléatoires z non triviales ont pour variance la valeur propre λ .

En effet :

$$E(z^2) = E\left[\left(\int_0^T \xi_t \, dt\right)^2\right] = E\left[\iint \xi_t \, \xi_s \, dt \, ds\right]$$

$$= E\left[\iint E_t(\xi_t)E_s(\xi_s)dt \, ds\right] = \int_0^T E\left[\xi_t \int_0^T E_t E_s \xi_s ds\right] dt = \int_0^T E(\xi_t \, \lambda \, \xi_t)dt$$

$$= \lambda \int_0^T E(\xi_t^2)dt = 1 \quad .$$

Les variables z vérifient une importante propriété :

<u>Propostion 3.6</u> :

Les variables z sont les variables numériques (définies à un coefficient multiplicateur près) qui sont les plus proches du processus au sens où :

$$\int_0^T \eta^2(z \; ; \; X_t)dt \quad \text{est extrémal} \quad .$$

η^2 représentant le carré du rapport de corrélation,

En effet on sait que les fonctions propres de l'opérateur

$E^t \, dt = Q$ rendent extrémale $\dfrac{\| z \, Q_z \|^2}{\| z \|^2} = \dfrac{E(z \int_0^T E^t \, z \, dt)}{E(z^2)}$,

Or $E(z \, E_t \, z) = E((E_t \, z)^2)$ et $\dfrac{E(E_t z)^2)}{E(z^2)} = \eta^2(z \; ; \; X_t)$ par définition puisque z et $E_t(z)$ sont centrées,

I-2-c) <u>Les codages propres</u>

ξ_t étant tel que $E_t(\xi_t) = \xi_t$ réalise un codage de X_t puisqu'il s'écrit donc :

$$\xi_t = \sum_{x=1}^{m} a^x(t) \, \mathbb{1}_t^x \quad ,$$

Cherchons à obtenir directement les fonctions $a^x(t)$ (ou codages évolutifs des modalités de (X_t)) ce qui résout complètement le problème de l'analyse harmonique qualitative en se ramenant à l'obtention non plus de processus ni de variables mais simplement de fonctions certaines du temps puisque ξ_t et aussi z s'expriment au moyen des a(t) :

$$z = \int_0^T \sum_x a^x(t)\, \mathbb{1}_t^x \, dt \, .$$

Il suffit de remplacer ξ_t par son expression en fonction des $a^x(t)$ dans la relation :

$$\lambda\, \xi_t = E_t \int_0^T \xi_s \, ds = \int_0^T E_t\, \xi_s \, ds$$

il vient :

$$\lambda \sum_{x=1}^{m} a^x(t)\, \mathbb{1}_t^x = \int_0^T E_t \left(\sum_{y=1}^{m} a^y(s)\, \mathbb{1}_s^y \right) ds$$

$$= \int_0^T \sum_{y=1}^{m} a^y(s)\, E_t(\mathbb{1}_s^y)\, ds$$

Or $E_t(\mathbb{1}_s^y) = \sum_{j=1}^{m} \mathbb{1}_t^j \dfrac{E(\mathbb{1}_s^y\, \mathbb{1}_t^j)}{E(\mathbb{1}_t^j)}$ d'après une formule établie au paragraphe 1 de cette partie.

On a donc :

$$\lambda \sum_{x=1}^{m} a^x(t)\, \mathbb{1}_t^x = \sum_{j=1}^{m} \sum_{y=1}^{m} \mathbb{1}_t^j \int_0^T a^y(s) \, \frac{E(\mathbb{1}_s^y\, \mathbb{1}_t^j)}{E(\mathbb{1}_t^j)} \, ds$$

Soit en identifiant les coefficients des indicatrices dans les deux membres et en remplaçant $E(\mathbb{1}_s^y\, \mathbb{1}_t^j)$ par $p^{jy}(t,s)$ et $E(\mathbb{1}_t^j)$ par $p^j(t)$ il vient

$$\boxed{\lambda\, a^x(t) = \sum_{y=1}^{m} \int_0^T a^y(s) \, \frac{p^{xy}(t,s)}{p^x(t)} \, ds \qquad x=1,2,\ldots,m}$$

En introduisant les matrices de probabilités $P(t,s)$ d'éléments $p^{x,y}(t,s)$ et comme $P(t,t) = \begin{bmatrix} p^{1}(t) & 0 \\ 0 & p^{m}(t) \end{bmatrix}$

d'où la ;

Proposition 3.7 :

Le vecteur $\vec{a}(t)$ des fonctions $a^{x}(t)$ vérifie l'équation intégrale suivante :

$$\lambda\, \vec{a}(t) = \int_0^T [P(t,t)]^{-1}\, [P(t,s)]\, \vec{a}(s)\, ds$$

Les codages propres sont donc les fonctions propres de l'opérateur intégral de $L^2_{\mathbb{R}^m}(T)$ dans lui-même de noyau $[P(t,t)]^{-1}\,[P(t,s)] = Q(t,s)$. $Q(t,s)$ est la matrice des probabilités de transition entre états du processus de l'instant t à l'instant s .

L'opérateur Q n'est pas autoadjoint mais P de noyau $P(t,s)$ l'est ; sous la condition de continuité en probabilité du processus (X_t) et de non nullité de $p^{x}(t)$ $\forall$ x $\forall$ t, on vérifie facilement que Q est un opérateur compact, positif, et nucléaire donc que son analyse spectrale est possible. De plus la continuité de (X_t) entraîne celle des fonctions $p^{x,y}(t,s)$ et $p^{x}(t)$ mais, répétons-le, cette condition n'est pas essentielle à l'existence des solutions et si le processus n'est pas continu en probabilité on obtiendra simplement des codages non continus.

Les codages propres sont les fonctions du temps vérifiant la propriété d'orthogonalité suivante ;

$$\langle \vec{a}^{(k)}(t) ; [P(t,t)]\, \vec{a}^{(\ell)}(t) \rangle\, dt = \delta_{k\ell} \quad .$$

L'espace des codages, qui est l'espace des formes linéaires sur les trajectoires m-dimensionnelles, est donc muni implicitement de la métrique associé à un opérateur "diagonal" de noyau $P(t,t)$; ceci prouve que

l'espace des trajectoires m-dimensionnelles réalisant des codages des trajectoires du processus qualitatif (X_t) qui est le dual du précédent est muni de la métrique inverse $[P(t,t)]^{-1}$ c'est-à-dire de la métrique du χ^2 à chaque instant.

Ce point ainsi que la nature de la décomposition orthogonale de (X_t) en variables aléatoires non corrélées indépendantes de temps et en codages certains orthogonaux (décomposition qui ne peut pas s'exprimer ici par une formule) est éclairci par le paragraphe 3 qui relie l'analyse harmonique qualitative à d'autres méthodes d'analyse des données.

I-2-d) Remarques concernant les valeurs propres

La valeur propre triviale T est nécessairement la plus grande puisque les projecteurs réduisent la norme des variables.

Cette remarque élémentaire a plusieurs conséquences : tout d'abord le pourcentage de trace "expliquée" par une valeur propre non triviale ne peut excéder $\frac{1}{m-1}$ c'est-à-dire l'inverse du nombre d'états du processus moins 1 . Donc sauf dans le cas d'un processus à 2 états aucune valeur propre ne peut excéder un pourcentage relativement faible de la trace.

Il n'est donc pas possible que le processus soit déterminé par une seule génératrice. Même dans le cas de processus engendrés par une seule variable on trouvera donc plusieurs (en général une infinité) génératrices et codages propres qui pourront être purement artificiels (effet Guttmann) on en verra un exemple au paragraphe 4.

Quant au cas où il n'existerait qu'une seule valeur propre associé à un espace propre engendrant tout le processus on voit aisément qu'il est sans intérêt : en effet cette unique valeur propre serait nécessairement égale à T avec une multiplicité m-1 (ou m avec la solution triviale). Ce cas correspond à l'orthogonalité de tous les projecteurs E_t puisqu'alors $\int_0^T E_t \, dt$ serait encore un projecteur orthogonal. Ceci signifie que pour tout $t \neq s$ X_t est une variable indépendante de X_s cas évidemment absurde pour un processus réel puisqu'à tout instant un individu pourrait changer d'état de façon indépendante de ses états antérieurs. (X_t) serait en quelque sorte un bruit blanc qualitatif.

I-3 Identification avec d'autres types d'analyses

Nous montrons ici successivement que l'analyse harmonique qualitative est le cas particulier de l'analyse factorielle d'un processus vectoriel appliquée au processus des indicatrices des états $(\vec{\mathbb{1}}_t)$, la généralisation en continu de l'analyse des correspondances multiples et l'interprétons en termes de codage optimal en la reliant à l'analyse en composantes principales semi-linéaire développée par [16] et [8].

1-3-a) Analyse harmonique vectorielle de $\vec{\mathbb{1}}_t$

Il est clair que les équations obtenues pour les codages, les génératrices et les processus propres de (X_t) sont strictement identiques à l'équation des facteurs, des composantes et des processus de l'analyse harmonique de $(\vec{\mathbb{1}}_t)$ à la différence près qu'ici il faut faire l'analyse harmonique vectorielle non centrée de $(\vec{\mathbb{1}}_t)$: on a alors identité entre l'opérateur d'espérance conditionnelle E_t et le projecteur A_t sur $(\mathbb{1}_t^1, \mathbb{1}_t^2, \ldots, \mathbb{1}_t^m)$ et les matrices de "covariance" $C(t,s)$ non centrées sont alors identiques aux matrices de probabilités $P(t,s)$. Toutes les propriétés établies au chapitre II sont également valables pour l'analyse harmonique qualitative ; en particulier on a maintenant une interprétation de la décomposition orthogonale de (X_t) selon les génératrices et les codages mais en termes de variables indicatrices :

$$\vec{\mathbb{1}}_t = \sum_{k=1} z_k \, [P(t,t)]^{-1} \, \vec{a}_k(t)$$

On trouve donc ici une confirmation indirecte du bien-fondé du choix implicite de la métrique $[C(t,t)]^{-1}$ en analyse harmonique vectorielle.

Quant à la propriété d'optimalité des génératrices en analyse harmonique qualitative :

$$\int_0^T \eta^2(z_i; X_t)\,dt \quad \text{extrémal}$$

elle s'identifie à celle de l'analyse harmonique vectorielle

$$\int_0^T R^2(z\ ;\ \vec{\mathbb{1}}_t)dt$$

puisqu'il est immédiat que le rapport de corrélation de z en X_t n'est autre que le coefficient de corrélation multiple de z avec les indicatrices des modalités de X_t .

I-3-b) Analyse des correspondances multiples continues

Rappelons qu'une des définitions possibles de l'analyse des correspondances multiples de p variables qualitatives $X_1, X_2, \ldots, X_p$ consiste dans la recherche des composantes et des facteurs de l'analyse canonique généralisée des p groupes d'indicatrices associées aux X_i [57] [38]. Avec nos notations actuelles les composantes sont les vecteurs propres de

$\sum_{i=1}^{p} E_i$ et les facteurs, les vecteurs propres du tableau de Burt

$$\begin{pmatrix} P(11) & P(12) & \ldots & P(1,p) \\ P(21) & P(22) & \ldots & \vdots \\ \vdots & & \ddots & \\ \ldots & & & P(p,p) \end{pmatrix}$$ prémultiplié par l'inverse de tableau

diagonal des probabilités marginales $\begin{pmatrix} P(1,1) & & & 0 \\ & P(2,2) & & \\ & & & \\ 0 & & & P(p,p) \end{pmatrix}$

Il est donc intuitif de voir dans l'analyse harmonique qualitative une extension de l'analyse des correspondances multiples dans laquelle le signe Σ serait remplacé par le signe d'intégration $\int$.

Cette conception est parfaitement justifiée si on considère un processus qualitatif (X_t) tel que les trajectoires du processus soient constantes sur des intervalles $[0, t_1[$, $[t_1\ t_2[$; ... $[t_{p-1}, t_p]$ autrement dit si les changements d'états ne peuvent se produire qu'à dates fixes t_j . Un tel processus n'est pas continu en probabilité mais

il n'y a aucune difficulté à appliquer les résultats de l'analyse harmonique qualitative car tous les opérateurs restent compacts (et nucléaires) et s'expriment en fait sous forme de somme.

L'équation des génératrices s'écrit, en notant ℓ_j la mesure de l'intervalle $[t_{j-1}\ t_j]$:

$$\lambda z = \sum_{j=1}^{p} \ell_j\ E_j\ z \quad .$$

L'équation des processus propres :

$$\lambda\ \xi_t = \sum_{j=1}^{p} \ell_j\ E_t\ E_j\ \xi_j$$

et celle des codages qui sont alors des fonctions constantes sur les intervalles $[t_{j-1},t_j]$

$$\lambda\ a_j^x = \sum_{i=1}^{p} \sum_{y=1}^{m} \ell_i\ \frac{p_{j\ i}^{xy}}{p_j^x}\ a_i^y$$

en notant a_i^y la valeur constante du codage de la modalité y sur $[t_{i-1},t_i[$ et p_{ji}^{xy} la valeur constante de la probabilité d'être dans l'état x sur $[t_{j-1},t_j[$ et dans l'état y sur $[t_{i-1},t_i[$.

Il est clair que l'équation donnant les génératrices et celle donnant les codages propres ne sont autres que celles de l'analyse des correspondances de l'ensemble des variables indicatrices des modalités des X_j multipliées par la mesure de l'intervalle (t_{j-1},t_j). Si Ω est fini (cas de n individus) ceci revient à effectuer l'analyse des correspondances du tableau $(\ell_1 X_1 | \ell_2 X_2 | \ldots | \ell_p X_p)$ où X_j est le tableau disjonctif à n lignes et m colonnes des états du processus sur le $j^{è}$ intervalle.

Si de plus les dates de changement d'état sont équidistantes la solution de l'analyse harmonique qualitative est exactement celle de l'analyse des correspondances du tableau $(X_1 | X_2 | \ldots | X_p)$ (au facteur T près) des états des individus aux dates $t_1, t_2, \ldots, t_p = T$.

Les règles d'interprétation de l'analyse harmonique qualitative seront donc celles de l'analyse des correspondances, qui sont bien connues, avec en plus la dimension chronologique.

I-3-c) ACP semi-linéaire et codage optimal

Sous le nom d'analyse en composantes principales semi-linéaire J. DAUXOIS et A. POUSSE [16] ont défini un type d'analyse des processus scalaires qui plutôt que d'utiliser les corrélations entre X_t et X_s étudie les corrélations entre des fonctions quelconques $f_t(X_t)$ et $f_s(X_s)$. Concrètement ceci revient à chercher des combinaisons linéaires de fonctions quelconques des X_t de variance maximale plutôt que des combinaisons linéaires d'où le nom de semi-linéaire.

Pour un processus ceci revient à chercher une composante z de la forme $z = \int_0^T f_t(X_t)dt$ où les fonctions f_t constituent une famille normalisée, dans son ensemble, de fonctions telles que $f_t(X_t) \in L^2(\Omega)$.

La construction théorique est alors la suivante : on note L_t^2 le sous-espace de $L^2(\Omega)$ engendré par la variable aléatoire X_t et H_t l'espace des fonctions mesurables de carré intégrable pour la mesure image de P par X_t (c'est-à-dire l'ensemble des φ telles que $\int_{\mathbb{R}} (\varphi(x))^2 \, d\, P_t(x) < \infty$) .

On a donc $\quad L_t^2 = \{Y \mid Y = f(X_t) \quad f \in H_t \}$

Soit (f_t) une famille de fonctions f_t telles que $f_t(X_t) \in L_t^2$. Muni de la norme définie par $\| f \|^2 = \int_0^T \int_{\mathbb{R}} (f_t(x))^2 \, d\, P_t(x)$ l'ensemble de ces familles consitue un espace de Hilbert $\mathcal{H}$ qui est l'intégrale hilbertienne des H_t :

$$\mathcal{H} = \int_T^{\oplus} H_t \, dt \quad .$$

On définit alors l'opérateur continu Φ de $\mathcal{H}$ dans $L^2(\Omega)$ par :

$$\Phi f = z = \int_0^T f_t(X_t)\, dt$$

Le complété de l'image de Φ que nous noterons $\mathcal{L}_X^2$ est l'adhérence du sous-espace engendré par les combinaisons linéaires de variables du type $f_t(X_t)$, c'est-à-dire le plus petit sous-espace de Hilbert de $L^2(\Omega)$ contenant tous les L_t^2. ($\mathcal{L}_X^2$ est isomorphe à l'intégrale hilbertienne $\int_T^{\oplus} L_t^2\, dt$).

L'adjoint Φ^* de Φ est l'opérateur de $L^2(\Omega)$ dans $\mathcal{H}$ tel que :

$$\Phi^*(z) = g \qquad \text{où} \qquad g_t(z) = E(z/X_t) = E_t(z) .$$

On a alors le schéma de dualité semi-linéaire :

$$\begin{array}{ccc} \mathcal{H} & \xleftarrow{\Phi^*} & L^2(\Omega) \\ \downarrow I & & I \uparrow \downarrow \psi \\ \mathcal{H} & \xrightarrow{\Phi} & L^2(\Omega) \end{array}$$

et on démontre que $\psi = \Phi \circ \Phi^* = \int_0^T E_t\, dt$. ψ est donc un opérateur autoadjoint borné.

L'A.C.P. semi-linéaire qui consiste dans la recherche des extrema de $E(z^2)$ où $z = \Phi g$ pour g appartenant à la boule unité de $\mathcal{H}$ se ramène lorsque Φ est compact à l'analyse spectrale de $\Phi \circ \Phi^*$ qui est compact et autoadjoint.

Le fait que (X_t) soit un processus scalaire n'est pas indispensable et la théorie précédente est valable pour (X_t) à valeurs dans un espace de Hilbert séparable quelconque.

L'application au cas d'un processus qualitatif (effectuée par R. BOUMAZA [8]) aboutit donc aux mêmes résultats que les nôtres : les composantes principales semi-linéaires de (X_t) sont identiques aux "génératrices".

Pour un processus qualitatif (X_t) chercher des transformations du type $f_t(X_t)$ est exactement ce que l'on appelle un codage : l'A.C.P. semi-linéaire

de (X_t) revient à chercher le codage de X_t tel que son analyse harmonique réduite fournisse la 1ère composante principale de variance maximale. On retrouve alors en continu les critères de codage optimal utilisé par HORST [32], KETTENRING [34], MASSON [43] et nous-mêmes [57] dans le cas fini.

L'application de l'A.C.P. semi-linéaire au cas général soulève de sérieuses difficultées théoriques et pratiques car Φ, et donc $\psi = \int_0^T E_t \, dt$, n'est pas alors compact : en effet on démontre ([21]) que $\int_0^T E_t \, dt$ ne peut être compact que si et seulement si les opérateurs E_t sont de rang fini c'est-à-dire précisément dans le cas qualitatif ((X_t) n'a qu'un nombre fini de valeurs).

Obtenir par discrétisation une approximation de l'A.C.P. semi-linéaire d'un processus scalaire (X_t) au moyen de l'analyse factorielle d'un processus qualitatif (X_t) obtenu en découpant en m classes l'ensemble des valeurs de X_t pose donc un problème : il n'y a pas convergence uniforme de $\psi^{(m)}$ vers ψ pour des partitions de plus en plus fines puisque ψ n'est pas compact alors que les $\psi^{(m)}$ le sont. Il y a seulement convergence forte [21].

Ce phénomène s'explique assez bien : puisque pour un processus qualitatif Trace $(\psi^{(m)}) = m\,T$, on voit que par passage à la limite on aboutit à une trace infinie et donc à des pourcentages d'inertie nuls sans que d'ailleurs les facteurs obtenus cessent d'être intéressants. (Cas bien connu en analyse des correspondances [5]).

I-4 Le cas des processus à deux états

Lorsque (X_t) ne peut prendre que deux états (notés état 0 et état 1) on obtient comme en analyse des correspondances multiples certains résultats supplémentaires. A titre d'exemple d'analyse harmonique qualitative nous étudierons le processus particulièrement simple traité sous une forme quelque peu différente dans [8].

I-4-a) Un exemple simple

Nous supposerons que l'intervalle [0,T] est identique à l'intervalle [0,1] ce qui ne nuit pas à la généralité mais simplifie les notations. Considérons le processus suivant : partant de l'état 0 on change d'état à une date θ pour passer à l'état 1 et y demeurer. On suppose que θ est une variable aléatoire uniformément répartie sur la durée de l'étude donc que $p^0(t) = 1-t \qquad p^1(t) = t$.

On en déduit les matrices P(t,s) :

$$P(t,s) = \begin{pmatrix} 1-s & s-t \\ 0 & t \end{pmatrix} \qquad \text{si } t < s$$

$$P(t,s) = \begin{pmatrix} 1-t & 0 \\ t-s & s \end{pmatrix} \qquad \text{si } t \geqslant s$$

et les matrices de probabilités conditionnelles $[P(t,t)]^{-1}\,[P(t,s)]$

$$[P(t,t)]^{-1}\,[P(t,s)] = \begin{pmatrix} \frac{1-s}{1-t} & \frac{s-t}{1-t} \\ 0 & 1 \end{pmatrix} \qquad \text{si } t < 1$$

$$[P(t,t)]^{-1}\,[P(t,s)] = \begin{pmatrix} 1 & 0 \\ 1-\frac{s}{t} & \frac{s}{t} \end{pmatrix} \qquad \text{si } s < t$$

L'équation donnant les codages propres $a_0(t)$ et $a\ (t)$ s'écrit donc :

$$\lambda \begin{pmatrix} a_0(t) \\ a_1(t) \end{pmatrix} = \int_0^T \begin{pmatrix} 1 & 0 \\ 1-\frac{s}{t} & \frac{s}{t} \end{pmatrix} \begin{pmatrix} a_0(s) \\ a_1(s) \end{pmatrix} ds + \int_t^T \begin{pmatrix} \frac{1-s}{1-t} & \frac{s-t}{1-t} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_0(s) \\ a_1(s) \end{pmatrix} ds$$

On vérifie aisément que le codage trivial $a^0(t) = a^1(t) = 1$ est associé à la valeur propre 1. Comme la trace de l'opérateur vaut mT c'est-à-dire 2, il reste 1 comme somme des valeurs propres non triviales.

Cherchons directement les solutions non triviales correspondant à des codages centrés donc telles que $p^0(t)\ a_0(t) + p^1(t)\ a_1(t) = 0$ ce qui donne la relation $a_0(t)\ (1-t) + a_1(t)t = 0$.

En éliminant soit $a_1(t)$ soit $a_0(t)$ au moyen de cette relation on aboutit à deux équations intégrales séparées pour les codages.

$$\lambda a_0(t) = \int_0^t a_0(s)ds + \frac{t}{1-t}\int_t^0 \frac{1-s}{s}\, a_0(s)ds$$

$$\lambda a_1(t) = \frac{1-t}{t}\int_0^t \frac{s}{1-s}\ a_1(s)ds + \int_1^t a_1(s)ds$$

Ces équations se ramènent aisément par dérivation à des équations différentielles du second ordre (en supposant, ce que le résultat confirme, la dérivabilité des a(t)).

On a ainsi en dérivant une fois :

$$\lambda\, a_0'(t) = a_0(t) - \frac{t}{1-t}\,\frac{1-t}{t}\, a_0(t) + \frac{1}{(1-t)^2}\int_t^1 \frac{1-s}{s}\, a_0(s)\, ds$$

$$= \frac{1}{(1-t)^2}\int_t^1 \frac{1-s}{s}\, a_0(s)\, ds$$

et en dérivant une deuxième fois :

$$\lambda\, a_0''(t) = \frac{2}{(1-t)^3}\int_t^1 \frac{1-s}{s}\, a_0(s)\, ds - \frac{1}{(1-t)^2}\,\frac{1-t}{t}\, a_0(t)$$

soit $\lambda\, a_0''(t) = \dfrac{2}{(1-t)}\,\lambda\, a_0'(t) - \dfrac{1}{t(1-t)}\, a_0(t)$

ce qui s'écrit encore :

$$t(1-t)\, a_0''(t) - 2t\, a_0'(t) + \frac{a_0(t)}{\lambda} = 0 \quad .$$

On trouve aisément que les solutions de ces deux équations différentielles sont des polynômes de degré k correspondant aux valeurs propres $\lambda_k = \dfrac{1}{k(k+1)}$.

En utilisant la condition de normalisation :

$$\int_0^1 [\,(a_1(t))^2\, p^1(t) + (a_0(t))^2\, p^0(t)]\, dt = 1 \quad .$$

On trouve ainsi que les trois premiers codages propres sont les suivants :

$$\lambda_1 = \frac{1}{2}\begin{cases} a_0^{(1)}(t) = \sqrt{6}\, t \\ a_1^{(1)}(t) = \sqrt{6}\,(t-1) \end{cases} \qquad \lambda_2 = \frac{1}{6}\begin{cases} a_0^{(2)}(t) = \sqrt{\frac{120}{7}}\left(t^2 - \frac{t}{2}\right) \\ a_1^{(2)}(t) = \sqrt{\frac{120}{7}}\left(t^2 - \frac{3t}{2} + \frac{1}{2}\right) \end{cases}$$

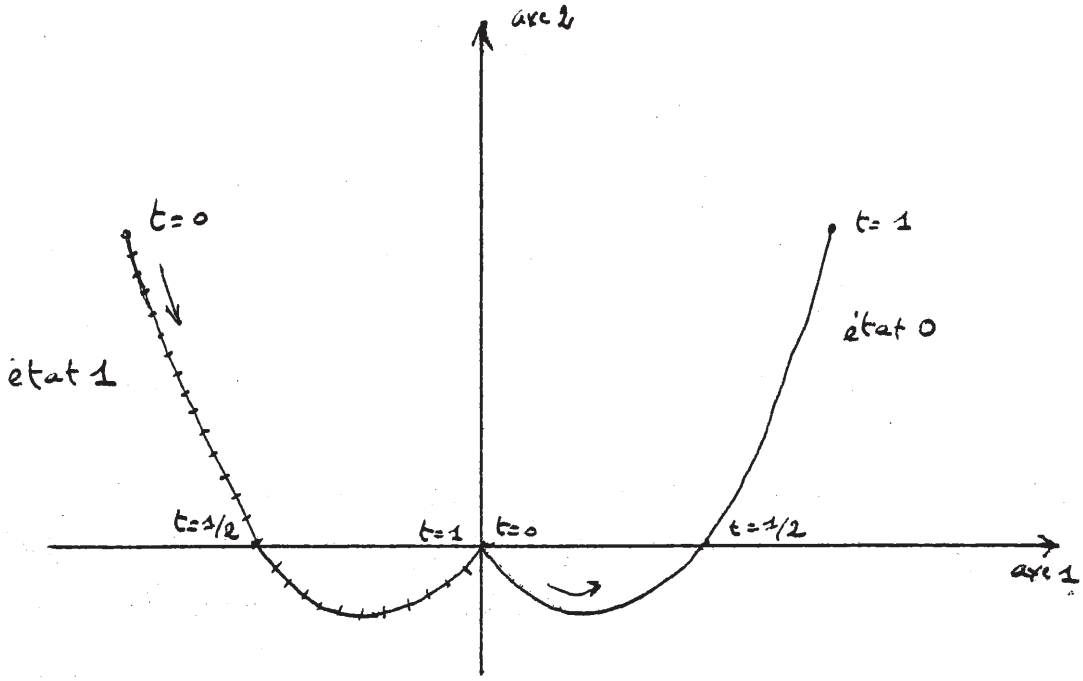
$$\lambda_3 = \frac{1}{12} \qquad \begin{aligned} a_0^{(3)}(t) &= \sqrt{\frac{2100}{701}}\left(t^3 - t^2 + \frac{t}{5}\right) \\ a_1^{(3)}(t) &= \sqrt{\frac{2100}{701}}\left(t^3 - 2t^2 + \frac{6}{5}t - \frac{1}{5}\right) \end{aligned} \quad .$$

Dans le premier plan factoriel on peut représenter les états 0 et 1 par deux courbes paramètrées en t .

On trouve aisément que ces deux courbes sont des arcs de deux paraboles symétriques par rapport à l'axe n°2, et d'équations respectives :

$$a_0^{(2)} = \sqrt{\frac{120}{7}}\left(\frac{(a_0^{(1)})^2}{6} - \frac{a_0^{(1)}}{2\sqrt{6}}\right)$$

$$a_1^{(2)} = \sqrt{\frac{120}{7}}\left(\frac{(a_1^{(1)})^2}{6} + \frac{a_1^{(1)}}{2\sqrt{6}}\right)$$



L'état 0 est entièrement à droite tandis que l'état 1 est entièrement à gauche.

Le premier axe est donc très lié à la chronologie et se contente d'opposer, comme on pouvait s'y attendre, le début et la fin du processus. En termes d'individus, ou de réalisation de processus ce qui est équivalent, on trouvera à droite du graphique les individus ayant changé de l'état 0 à l'état 1 plus tard que la moyenne et à gauche ceux qui ont changé plus tôt que la moyenne.

Comme on le verra plus loin le processus étant en réalité unidimensionnel c'est-à-dire entièrement déterminé par une seule variable qui est la date de changement d'état, on observe comme à l'habitude dans ce cas un effet GUTTMANN généralisé et il est inutile de chercher à interpréter les autres axes.

Cherchons précisément la coordonnée d'une trajectoire c'est-à-dire d'un individu. Ceci revient à déterminer la valeur prise par la génératrice z pour la réalisation particulière du processus (X_t).

On a $$z^{(k)} = \int_0^1 (a_0^{(k)}(t)\, \mathbb{1}_t^0 + a_1^{(k)}(t)\, \mathbb{1}_t^1)dt$$

Or une trajectoire est entièrement déterminée par la valeur θ de la date de changement d'état.

D'où $$z(\theta) = \int_0^\theta a_0^{(k)}(t)\, dt + \int_\theta^1 a_1^{(k)}(t)\, dt$$

en utilisant la condition de centrage $a_1(t) = -a_0(t).\frac{1-t}{t}$

il vient :

$$z(\theta) = \int_0^\theta a_0(t)dt - \int_\theta^1 a_0(t)\, \frac{1-t}{t}\, dt \quad .$$

On en déduit aisément que $z'(\theta) = \dfrac{a_0(\theta)}{\theta}$

d'où en reportant dans l'équation différentielle donnant $a_0(t)$ et après simplifications :

$$\theta(1-\theta)z'''(\theta) + 2(1-2\theta)\, z''(\theta) - 2z'(\theta) + \frac{z'(\theta)}{\lambda} = 0$$

ce qui donne après une quadrature :

$$\theta(1-\theta)\ z''(\theta) + (1-2\theta)\ z'(\theta) + \frac{z(\theta)}{\lambda} = 0$$

un changement de variable élémentaire $\theta = \frac{1+u}{2}$ permet de se ramener à une équation de LEGENDRE d'ordre zéro sur $[-1\ ;\ +1]$

$$(1-u^2)g''(u)-2ug'(u) + \frac{g(u)}{\lambda} = 0 \qquad \text{avec } g(u) = z(\frac{1+u}{2})$$

on sait que les solutions de cette dernière équation sont données par $g(u) = P_k(u)$ si $\lambda_k = \frac{1}{k(k+1)}$ où $P_k(u)$ est le $k^{\text{è}}$ polynôme de LEGENDRE

défini par :

$$P_k(u) = \frac{1}{2^k k!}\ \frac{d^k(u^2-1)^k}{du^k}\ .$$

Les fonctions $z^{(k)}(\theta)$ sont donc à un coefficient de normalisation près égales à $P_k(2\theta-1)$.

En tant que variables aléatoires ceci prouve que les génératrices z sont des fonctions certaines de la variable θ de distribution uniforme sur $[0,1]$.

La première génératrice $z^{(1)}$ est alors exactement égale à $\sqrt{6}\,[\theta - \frac{1}{2}]$: ce résultat pouvant être obtenu soit directement par la formule

$$z^{(1)}(\theta) = \int_0^\theta a_0^{(1)}(t)dt - \int_\theta^1 a_0^{(1)}(t)\ \frac{1-t}{t}\ dt$$

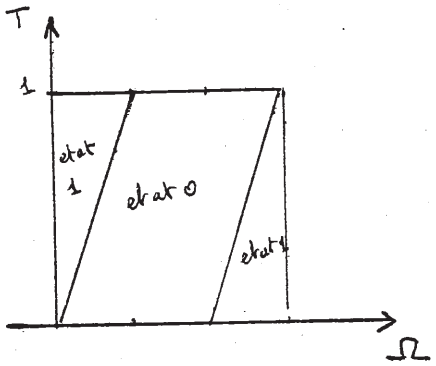
soit en cherchant la constante de normalisation du polynôme de LEGENDRE sur $[0;1]$ qui est $2\theta-1$ et ici on sait que $V(\theta) = \frac{1}{12}$ d'où $V(2\theta-1) = \frac{1}{3}$ ce qui aboutit au même résultat.

On retrouve donc les mêmes résultats que [8] qui avait considéré simplement que $\Omega = [0,1]$ utilisant ainsi la bijection entre trajectoires et valeurs de la date de changement d'état.

Ceci confirme l'interprétation donnée au premier axe au vu des courbes des codages : la date de changement d'état moyen étant 1/2 on retrouve bien la dichotomie gauche-droite du graphique en termes de changement avant ou après la moyenne.

Bien que le processus (X_t) puisse être décrit par une seule variable θ on obtient une décomposition de (X_t) selon une infinité de facteurs fonctions exactes du premier facteur, non corrélées mais certes pas indépendantes. Il ne faut donc pas se leurrer ici sur le nombre de facteurs à prendre en considération: les facteurs d'ordre 2 et plus ne sont que des artefacts dus à la méthode.

On trouvera dans [8] un autre exemple de processus simple à deux états défini par le schéma suivant :



Chaque trajectoire est représentée par une droite verticale et toutes les trajectoires sont équiprobables (Ω muni de la mesure dt).
On trouve alors que toutes les valeurs propres sont doubles.

Ces exemples n'ont qu'un intérêt théorique, limité par leur simplicité : la résolution de processus aussi simples que le processus de MARKOV à deux états avec durée de séjour exponentielle dans chaque état aboutit à des équations inextricables sans parler de processus à plus de deux états où on ne peut simplifier aussi aisément l'équation des codages propres. Il doit être clair que l'analyse harmonique qualitative est essentiellement un outil de description de données réelles et seulement accessoirement une méthode d'étude de processus théoriques.

I-4-b) Réduction de l'analyse harmonique qualitative à une analyse harmonique scalaire

Nous montrons ici, et l'appliquons à l'exemple précédent, que comme en analyse des correspondances multiples, l'analyse d'un processus qualitatif à deux états se ramène à l'analyse harmonique dans la

métrique D_{1/σ^2} du processus scalaire de la variable indicatrice d'un seul des deux états.

Proposition 3.8 :

Les génératrices de l'analyse harmonique qualitative d'un processus (X_t) à deux états sont les composantes principales de l'analyse harmonique centrée du processus scalaire $(\mathbb{1}_t^0)$ (ou $\mathbb{1}_t^1$) avec la métrique D_{1/σ^2} .

La démonstration est identique à celle de la propriété semblable en analyse des correspondances multiples:

Soit une variable numérique z on sait que :

$$E_t(z) = \frac{\mathbb{1}_t^0 \, E(z \, \mathbb{1}_t^0)}{p^0(t)} + \frac{\mathbb{1}_t^1 \, E(z \, \mathbb{1}_t^1)}{p^1(t)}$$

$$= \frac{\mathbb{1}_t^0 \, E(z \, \mathbb{1}_t^0)}{p^0(t)} + \frac{(1-\mathbb{1}_t^0)(E(z(1-\mathbb{1}_t^0))}{1-p^0(t)}$$

soit après simplification :

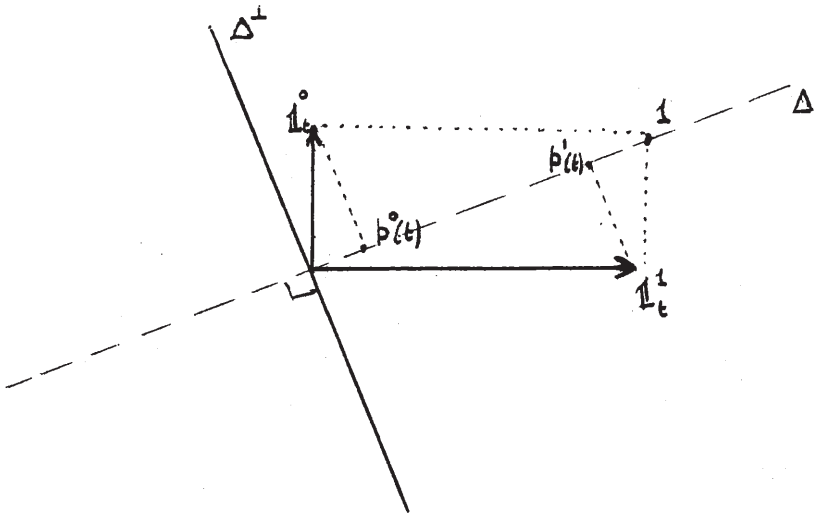
$$E_t(z) = \frac{(\mathbb{1}_0^t - p^0) \, E(z \, \mathbb{1}_0^t)}{p^0(t)(1-p^0(t))}$$

Si z est centrée $E_t(z)$ est exactement le projecteur linéaire sur $\mathbb{1}_0^t - p_0$, l'indicatrice centrée puisque $p^0(t)(1-p^0(t)) = V(\mathbb{1}_0^t)$.

L'opérateur $\int_0^T E_t \, dt$ est donc identique à l'opérateur d'ESCOUFIER du processus centré $\mathbb{1}_t^0 - p_0$ avec la métrique $D_{1/\sigma^2(\mathbb{1}_t^0)}$ ce qui démontre la proposition.

Géométriquement l'interprétation de ce résultat est immédiate, bien que jamais explicitée à notre connaissance dans la littérature.

E_t est le projecteur sur le plan engendré par $\mathbb{1}_t^0$ et $\mathbb{1}_t^1$: or $\mathbb{1}_t^0$ et $\mathbb{1}_t^1$ sont orthogonales et leur somme vaut 1 . On a donc la figure suivante dans le plan défini par les deux indicatrices :



Les variables centrées sont les variables orthogonales à Δ c'est-à-dire celles contenues dans le plan passant par $\Delta^{\perp}$ et orthogonal au plan de la figure. Projeter une variable centrée sur le plan $(\mathbb{1}_t^0 , \mathbb{1}_t^1)$ équivaut à la projeter sur $\Delta^{\perp}$ c'est-à-dire sur $\mathbb{1}_t^0 - p^0(t)$ ou sur $\mathbb{1}_t^1 - p^1(t)$ qui sont tous trois parallèles.

Si les génératrices de (X_t) et les composantes principales de $(\mathbb{1}_t^0)$ sont les mêmes, c'est-à-dire si on obtient les mêmes coordonnées des individus-trajectoires dans les deux analyses il faut prendre garde qu'il n'en est pas de même pour les codages : les codages de la modalité 0 $a^0(t)$ dans l'analyse harmonique qualitative de (X_t) diffèrent des facteurs principaux de l'analyse harmonique scalaire de $(\mathbb{1}_t^0)$ par un simple coefficient de proportionnalité.

En appelant f(t) les facteurs de l'analyse de $(\mathbb{1}_t^0)$ on a :

$$\boxed{f(t) = \frac{a^0(t)}{1-p^0(t)}}$$

Il suffit d'identifier les deux manières d'obtenir z :

AHQ de (X_t) $\qquad z = \int_0^T a^0(t)\, \mathbb{1}_t^0\, dt + \int_0^T a^1(t)\, \mathbb{1}_t^1\, dt$

AHS de $(\mathbb{1}_t^0)$ $\qquad z = \int_0^T f(t)(\mathbb{1}_t^0 - p^0(t))\, dt$

en utilisant à nouveau la relation de centrage

$$a^0(t)\, p^0(t) + a^1(t)\, p^1(t) = 0$$

la première expression se réécrit

$$z = \int_0^T \frac{a^0(t)}{1-p^0(t)}\, p^0(t)(1-\mathbb{1}_t^0)\, dt + \int_0^T a^0(t)\, \mathbb{1}_t^0\, dt$$

ce qui se réduit à

$$z = \int_0^T \frac{a^0(t)(\mathbb{1}_t^0 - p^0(t))}{1 - p^0(t)}\, dt \quad .$$

<u>Exemple</u>

Pour le processus étudié au a) on a

$$E(\mathbb{1}_t^0) = 1-t \qquad V(\mathbb{1}_t^0) = t(1-t)$$

$$E(\mathbb{1}_t^0\, \mathbb{1}_s^0) = 1 - \mathrm{Sup}(t,s)$$

d'où une fonction de covariance pour $\mathbb{1}_t^0$:

$$C(t,s) = (1-t)s \qquad \text{si } t \geqslant s$$

$$C(t,s) = (1-s)t \qquad \text{si } t < s \quad .$$

L'équation de l'analyse harmonique scalaire de $\mathbb{1}_t^0$ avec la métrique $M(t,t) = \dfrac{1}{t(1-t)}$ est alors :

$$\lambda\ f(t) = \frac{1}{t(1-t)} \int_0^t (1-t)s\ f(s)\ ds + \frac{1}{t(1-t)} \int_t^1 (1-s)t\ f(s)\ ds$$

soit

$$\lambda\ f(t) = \frac{1}{t(1-t)} \int_0^t (1-t)s\ s(s)\ ds + \frac{1}{1-t} \int_t^1 (1-s)\ f(s)\ ds$$

On vérifie que cette équation est identique à celle donnant le codage de l'état 0 :

$$\lambda\ a_0(t) = \int_0^t a_0(s)\ ds + \frac{t}{1-t} \int_t^1 \frac{1-s}{s}\ a_0(s)\ ds$$

dans laquelle on poserait $a_0(s) = s\ f(s)$.

La coordonnée $z(\theta)$ d'un individu ayant changé d'état à la date θ s'exprime par

$$z(\theta) = \int_0^\theta f(t)(1-p^0(t))dt + \int_\theta^1 f(t)(-p^0(t))\ dt$$

$$z(\theta) = \int_0^\theta f(t)t\ dt - \int_\theta^1 f(t)(1-t)\ dt$$

et en dérivant :

$$z'(\theta) = f(\theta)\theta + (1-\theta)\ f(\theta) = f(\theta)$$

On en déduit que les facteurs $f(t)$ sont les dérivées des polynômes de LEGENDRE sur $[0,1]$ à une constante de normalisation près.

I-5 Problèmes numériques posés par le traitement de données réelles

Nous nous préoccuperons uniquement du problème de l'approximation des solutions de l'analyse harmonique qualitative pour un échantillon de taille finie n au sens où cette notion a été définie au premier chapitre.

Le problème de l'interpolation ne nous semble pas en effet intéressant à considérer pour les raisons suivantes :

pour un processus qualitatif on connait généralement pour chaque individu les dates de changements d'états et les états pris. Il est donc peu réaliste d'envisager des processus qualitatifs où les trajectoires ne seraient pas entièrement connues. De plus si on ne connait que les états dans lesquels se trouvent les individus à certaines dates $t_1, t_2, \ldots, t_p$ il semble difficile sinon impossible d'interpoler entre les t_i vu la nature qualitative du processus (combien de changements d'états ont pu avoir lieu entre t_{i-1} et t_i et lesquels ?) la seule ressource consiste à effectuer l'analyse du processus discret X_{t_1} ; X_{t_2} ; ... ; X_{t_p} ce qui renvoie au § 3-b) de ce chapitre et à la première méthode d'approximation proposée ici.

En tout état de cause les méthodes d'approximation utilisables reviennent toutes à effectuer une analyse factorielle des correspondances sur un tableau de nombres ce qui rend la méthode parfaitement exploitable.

Rappelons tout d'abord dans quel contexte l'approximation est indispensable dès que les changements d'états ne se produisent pas à dates fixes. Certes, avec un nombre fini d'individus les opérateurs intégraux sont à noyaux dégénérés et on peut même ici (puisque le processus est qualitatif) les ramener à des matrices : n étant fini le nombre total des dates de changements d'états l'est aussi et on peut, du moins théoriquement, se ramener à un processus qualitatif à temps discret et effectuer l'analyse des correspondances des tableaux $(..|\ell_i\ X_i|..)$ où ℓ_i est l'écart entre deux dates de changement d'états et X_i le tableau disjonctif des états durant cette période. Même si le nombre total des états de changement par individu n'est pas trop élevé la dimension du problème devient colossale:avec 1000 individus changeant seulement 3 fois d'état en moyenne entre 0 et T on aboutit à une matrice 1000 x 3000 .

Il est donc indispensable de proposer des méthodes permettant de limiter la dimension des tableaux à traiter.

I-5-a) Approximation par dicrètisation du temps

L'idée consiste à n'utiliser que les états du processus pris par les n individus à des dates t_0, t_1, t_2, ... , t_p . On prendra en général $t_0 = 0$ et $t_p = T$ ce qui fait p intervalles.

On peut alors se poser le problème suivant : comment approximer l'opérateur $\int_0^T E_t \, dt$ au moyen des E_{t_i} ?

Il suffit de transposer les résultats du chapitre I paragraphe III-1-b sur l'approximation des intégrales au moyen de formules de quadrature.

La méthode des rectangles, la plus simple, consistera donc à approximer $\int_0^T E_t \, dt$ par $\ell_1 E_0 + \ell_2 E_1 + \dots \ell_p E_{p-1}$ où $\ell_i = t_{i+1} - t_i$. On voit sans difficulté que ceci revient à effectuer une analyse des correspondances sur le tableau suivant $(\ell_1 X_0 | \ell_2 X_1 | \dots | \ell_p X_{p-1})$ où X_i est le tableau n x m des indicatrices des états des individus à l'instant t_i .

Cette formule peut évidemment être améliorée pour tenir compte des états à l'instant final qui sont ici ignorés ; il suffit d'utiliser la méthode des trapèzes ou la méthode de SIMPSON ce qui revient à effectuer une analyse des correspondances sur $(\alpha_0 X_0 | \alpha_1 X_1 | \alpha_2 X_2 \dots | \alpha_T X_T)$ où les α_i dépendent de la technique choisie.

Concernant le choix des t_i nous renvoyons au chapitre I en mentionnant les résultats établis en [8] sur la convergence uniforme presque sûre de l'analyse par discrétisation si les t_i constituent un échantillonnage aléatoire de [0,T] .

Il est clair, cependant, que ces méthodes ne sont guère recommandables puisqu'on ne tient compte que de la situation aux extrémités des intervalles de découpage et non de l'évolution entre elles.

I-5-b) Approximation par moyennage

Nous reprenons ici la méthode que nous avons proposée en [20] en en donnant une présentation beaucoup plus simple.

Il suffit de reprendre la méthode d'approximation étudiée au chapitre I qui consistait à approximer les trajectoires d'un processus scalaire par des fonctions constantes par intervalles.

Comme le processus est ici qualitatif et pour qu'une telle approximation ait un sens on utilise la représentation vectorielle du processus sous forme d'indicatrices.

On va alors chercher l'approximation par des fonctions constantes sur le découpage $0 = t_0, t_1, \ldots, t_p = T$ des indicatrices $\mathbb{1}_t^1 ; \ldots ; \mathbb{1}_t^m$ du processus. La meilleure approximation d'une variable par une constante étant sa moyenne (au sens de la norme de $L^2(\Omega)$) ceci revient pour chaque trajectoire (i.e. chaque individu) à calculer la moyenne des indicatrices de ses états sur chaque intervalle du découpage.

Or $\Theta_i^x = \int_{t_{i-1}}^{t_i} \mathbb{1}_t^x \, dt$ n'est autre que la variable aléatoire "temps passé dans l'état x entre les instants t_{i-1} et t_i".

L'approximation revient donc à remplacer le processus $\vec{\mathbb{1}}_t$ par le processus vectoriel constant par morceaux $(\frac{1}{\ell_t}\Theta_t^1 : \frac{1}{\ell_t}\Theta_t^2 \ldots , \frac{1}{\ell_t}\Theta_t^m)$

où $\ell_t = t_i - t_{i-1}$ si $t \in [t_i ; t_{i-1}]$.

Tous calculs faits on s'aperçoit alors que l'analyse harmonique vectorielle du processus des indicatrices approximées n'est autre que l'analyse des correspondances du tableau suivant :

$$\begin{array}{c} \\ 1 \\ 2 \\ \cdot \\ \omega \\ n \end{array} \left(\begin{array}{c|c|c|c} \Theta_1^1 \;\; \Theta_1^2 \ldots \Theta_i^m & \Theta_2^1 \;\; \Theta_2^2 \ldots \Theta_2^m & \ldots & \Theta_p^1 \;\; \Theta_p^2 \ldots \Theta_p^m \\ & & & \\ \mathcal{T}_1 & \mathcal{T}_2 & \ldots & \mathcal{T}_p \\ & & & \end{array} \right)$$

où la ligne ω représente la liste des temps passés dans chacun des m états au cours des p intervalles du découpage par l'individu ω.

L'utilisation de la métrique du χ^2 à chaque instant conduit donc à une approximation de cette métrique, constante par morceaux sur chaque intervalle, soit pour le $i^{è}$ intervalle :

$$M(i,i) = \left[\frac{1}{\ell_t} \begin{pmatrix} \overline{\Theta}_i^1 & & 0 \\ & \overline{\Theta}_i^2 & \\ 0 & & \ddots \; \overline{\Theta}_i^2 \end{pmatrix} \right]^{-1}$$

où $\overline{\Theta}_i^x$ est la moyenne des temps passés par les n individus dans l'état x au cours du $i^{ème}$ intervalle .

Cette méthode se rapproche de l'analyse des correspondances mixtes développée par M. MASSON [44] pour des variables qualitatives : dans cette méthode un individu n'a pas un 1 dans une seule colonne d'un sous-tableau mais une distribution de probabilité sur les colonnes c'est-à-dire sur les modalités (codage "flou").

Dans [8] on trouve la démonstration de la convergence de cette approximation lorsque l'on découpe [0,T] en intervalles de plus en plus fins.

Les premières applications de cette technique réalisées par un groupe de travail de l'ENSAE sous la direction de J.C. DEVILLE ont montré une grande stabilité des résultats vis à vis du découpage utilisé [26] .

Remarque :

Dans l'approximation ainsi faite la trace de la matrice que l'on diagonalise ne vaut plus mT mais est inférieure à mT . On n'a plus évidemment les propriétés des sommes de projecteurs puisque les approximations obtenues des projecteurs ne sont pas des projecteurs et sont en tout état de cause des opérateurs de rang inférieur. Le rapport de cette trace à mT représente donc la qualité de l'approximation globale obtenue

II- SUR D'AUTRES METHODES FACTORIELLES

D'autres méthodes d'analyse d'un processus qualitatif sont envisageables, certes moins riches d'interprétation que l'analyse harmonique qualitative. Nous passons ici en revue les diverses analyses factorielles que l'on peu effectuer dans la même optique que celles présentées au début du chapitre II ainsi que la méthode des opérateurs.

II-1) Les six analyses factorielles

Un processus qualitatif étant identique au processus vectoriel des indicatrices des états nous particularisons ici en gardant la même numérotation les 6 analyses proposées au §III du chapitre II. Les analyses par sommation sur un indice (soit les individus, soit le temps, soit les modalités) ont ici la signification d'analyses marginales.

II-1-a) Les trois analyses marginales

* L'analyse n°1 ou sommation sur le temps revient à étudier les m variables aléatoires obtenues par intégration stochastique :

$$\Theta_1 = \int_0^T \mathbb{1}_t^1 \, dt \qquad \Theta_2 = \int_0^T \mathbb{1}_t^2 \, dt \quad \ldots \quad \Theta_m = \int_0^T \mathbb{1}_t^m \, dt \, .$$

La variable Θ_i est alors le temps passé dans l'état i au cours de la période d'étude.

On est donc ramenés à l'étude des liaisons entre m variables numériques ce qui est du domaine classique de l'analyse des données. Les variables Θ_i sont liées par la relation évidente $\sum_{i=1}^{m} \Theta_i = T$ et forment donc un système de rang m-1 .

Une analyse factorielle de ces m variables mettra en évidence des facteurs de différenciation entre individus (trajectoires) reposant uniquement sur les temps totaux dans chaque états et négligera donc, et l'ordre dans lequel ces états sont pris, et la fréquence des changements d'états. L'information ne sera conservée que pour un processus

ordinal c'est-à-dire un processus où les états ne peuvent être pris que dans un ordre bien déterminé puisqu'alors les durées de séjour dans chaque état constituent un résumé exhaustif des trajectoires.

Cette méthode nous semble particulièrement intéressante pour des processus ordinaux et certains processus scalaires discrets (du type constitution des familles) considérés comme qualitatifs : on dispose alors à peu de frais, puisque la dimension du tableau est m, d'une méthode d'analyse pertinente et qui plus est non linéaire puisque les temps passés dans chaque état ne sont pas en général des fonctions linéaires des états. Dans ce cas, si l'information utilisée est la même il est clair que les résultats des analyses seront différents le choix des praticiens devant être guidé par la simplicité de mise en oeuvre et les possibilités d'interprétation.

Remarquons qu'appliquée au processus de Poisson cette méthode conduit à un résultat trivial puisque les temps passés dans chaque état sont des variables indépendantes et équidistribuées mais après tout ce résultat est peut être plus facilement interprétable et plus riche d'enseignements que le caractère sinusoïdal ou besselien des facteurs de l'analyse harmonique de ce processus effectuée au chapitre I.

Dans le cas des familles par contre il est vraisemblable que l'on retrouverait par cette méthode les mêmes résultats que dans [17] peut-être plus facilement à condition d'étudier les trajectoires à nombre total d'enfants fixé.

Puisqu'il s'agit d'une analyse factorielle sur données numériques le choix de la métrique à utiliser doit-être résolu : comme les données sont, en divisant par T, assimilables à des distributions de probabilité il nous semble indiqué d'utiliser la métrique du χ^2 c'est-à-dire de recourir à une analyse en composantes principales non centrée (c'est inutile puisque $\sum_{i=1}^{m} \Theta_i = T$ entraînera le centrage des composantes non triviales) et de métrique $D_{1/E(\Theta)}$ ce qui revient à effectuer une analyse des correspondances sur le tableau individus X Θ_i .

En fait la justification théorique de l'usage de l'analyse des correspondances est la suivante : ce qui est étudié ici est en réalité une loi de probabilité définie sur $\Omega \times M$ muni de la tribu produit ;

chaque élément $\frac{\Theta_i(\omega)}{T}$; $\frac{\Theta_2(\omega)}{T}$; ... ; $\frac{\Theta_n(\omega)}{T}$ peut être considéré

comme une distribution conditionnelle de probabilité sur M à ω fixé. Les lois marginales sont parfaitement connues : ce sont d'une part la mesure de probabilité μ sur Ω et d'autre part la mesure moyenne

$\frac{1}{T} \int_0^T p^x(t)\, dt$ sur les états du processus.

L'analyse d'une probabilité sur un ensemble produit se fait par l'analyse canonique des deux tribus associées ce qui se réduit ici à une analyse des correspondances qu'il est facile d'expliciter : la matrice à diagonaliser est la matrice N de terme général

$$N_{ij} = \frac{E(\Theta_i\ \Theta_j)}{E(\Theta_i)} = \frac{\iint_{[0,T]^2} p^{ij}(t,s)\, dt\, ds}{\int_0^T p^i(t)\, dt} \qquad i,j \text{ de } 1 \text{ à } m$$

puisque $\Theta_i = \int_0^T \mathbb{1}_t^i\, dt$. On a en effet $E(\Theta_i) = E \int_0^T \mathbb{1}_t^i\, dt$

donc $E(\Theta_i) = \int_0^T E(\mathbb{1}_t^i)\, dt = \int_0^T p^i(t)\, dt$ et $E(\Theta_i\ \Theta_j) = E \int_0^T \mathbb{1}_t^i\, dt \int_0^T \mathbb{1}_s^i\, ds$

d'où $E(\Theta_i\ \Theta_j) = E \iint \mathbb{1}_t^i\, \mathbb{1}_t^j\, dt\, ds = \iint E(\mathbb{1}_t^i\, \mathbb{1}_t^j)\, dt\, ds = \iint p^{ij}(t,s)\, dt$

Ce type d'analyse revient à chercher un codage fixe des états du processus tel que son analyse harmonique scalaire soit optimale au sens suivant

Proposition 3-9

Le premier facteur de l'analyse des correspondances de la loi de probabilité $\frac{1}{T}\Theta_x(\omega)$ est le codage des états du processus (X_t) qui donne une moyenne temporelle $\int_0^T X_t\,dt$ de variance maximale sous la contrainte $\sum_x (a_x)^2/E(\Theta_x) = 1$.

En effet $\int_0^T X_t\,dt$ est égal à $\sum_x a_x \int_0^T \mathbb{1}_t^x\,dt = \sum_x a_x\,\Theta_x$.

On retrouve donc ici les résultats de la proposition 2.6 du chapitre II.

Par rapport à l'analyse harmonique qualitative où on cherchait des facteurs de la forme $\vec{a}_t$ ici on impose aux facteurs d'avoir la forme $\vec{a} \bullet f(t)$. Il est clair qu'on obtiendra donc des résultats différents de ceux de l'analyse harmonique qualitative même dans le cas où le processus est ordinal : on peut le vérifier sur le processus élémentaire à 2 états étudié au paragraphe I-4 où l'A.F.C. de Θ_0, Θ_1 est particulièrement évidente. Comme Θ_0 et Θ_1 sont des variables aléatoires uniformes sur $[0,1]$. On a $E(\Theta_0) = E(\Theta_1) = \frac{1}{2}$ $E[\Theta_0^2] = E[\Theta_1^2] = \frac{1}{3}$ et $E[\Theta_0\Theta_1] = \frac{1}{6}$ l'analyse des correspondances des Θ conduit à la recherche des vecteurs propres de la matrice :

$$\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}^{-1} \begin{pmatrix} 1/3 & 1/6 \\ 1/6 & 1/3 \end{pmatrix} = \begin{pmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix}$$

dont la valeur propre non triviale vaut 1/3 .

On constate qu'elle est inférieure à celle obtenue en analyse harmonique qualitative (1/2) ceci est un résultat général qui ne nécessite pas de démonstration particulière : il suffit de se rappeler que l'A.F.C. des Θ_x revient à projeter le processus sur un espace plus restreint qu'en A.H.Q. puisque les contraintes de codage sont plus strictes.

Le codage constant des états du processus (X_t) est alors donné par $\vec{a} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ à une constante près, ce qui n'a rien de surprenant.

Il faut remarquer ici sur un plan général que l'analyse présentée ici est spécifique de l'étude de données chronologiques et constitue une nouveauté par rapport aux techniques classiques d'analyse des données : en effet c'est seulement parce qu'il s'agit de la même variable X observée à des instants différents que la recherche de codages fixes (c'est-à-dire d'une combinaison unique des variables indicatrices) a un sens.

* L'analyse n° 2 consiste à intégrer sur Ω ce qui revient à ne considérer que la fonction vectorielle du temps définie par la distribution de probabilité des états à l'instant t

$$\vec{p}_t = (p^1(t) \; ; \; p^2(t) \; ; \; \ldots \; ; \; p^m(t)) \; .$$

On peut certes considérer $\vec{p}_t$ comme un ensemble de m courbes et utiliser les résultats exposés précédemment au chapitre I. Mais ici s'agissant de lois de probabilités la technique d'analyse est clairement définie : on est en présence d'une distribution de probabilité sur le produit T x M : l'analyse canonique de cette probabilité est alors élémentaire puisque les distributions marginales sont données par la mesure de T d'une part et par les $\frac{1}{T}\int_0^T p^x(t)\,dt$ d'autre part. On est donc encore ramené à une analyse du type correspondances avec la métrique du χ^2 de centre la distribution marginale de probabilité des états.

La matrice à diagonaliser est ici la matrice m x m de terme général :

$$M_{ij} = \frac{\int_0^T p^i(t)\ p^j(t)\ dt}{\int_0^T p^i(t)\ dt} \quad .$$

Cette méthode appauvrit évidemment beaucoup les données puisqu'on perd la dimension individuelle ce qui revient à considérer que toutes les trajectoires sont identiques. Pour être complet mentionnons cependant qu'on rencontre dans certains domaines, en particulier en économie, des données de ce type où seules les distributions marginales sont connues (ventilation des dépenses à la date t) il existe alors des méthodes d'estimation des probabilités de transition sous certaines hypothèses comme celle d'une chaîne de Markov (voir [45] en particulier).

Remarque :

Pour le processus à 2 états utilisé comme exemple on obtient ici exactement les mêmes résultats qu'avec l'analyse n° 1

puisque $p^0(t) = 1-t \qquad p^1(t) = t$

$$\text{donc} \int_0^T p^0(t)\ p^1(t)\ dt = 1/3 \quad \text{et} \quad \int_0^T (1-t)^2\ dt = \int_0^T t^2\ dt = 1/2$$

la matrice M est alors égale à N . Mais ceci n'est qu'un cas particulier dû à la nature du processus (X_t) envisagé et n'est pas généralisable.

* L'analyse n° 3 (sommation sur les variables) n'a ici aucun intérêt puisque cela revient à sommer sur les modalités de X on obtient donc le processus trivialement constant et égal à 1 pour tout t et tout ω .

Ainsi seule l'analyse n° 1 présente un réel intéret parmi les analyses marginales et apporte une novation par rapport aux méthodes usuelles d'analyse de données.

II-1-b) Les trois analyses "mixtes"

Ce qui a été présenté au chapitre II en terme d'analyse factorielle de juxtaposition ou superposition de tableaux de données se traduit ici en terme d'analyse des correspondances sur des tableaux d'indicatrices, du moins pour le cas fini.

Sur le plan théorique, qui permet de traiter le cas infini, les données doivent être considérées comme des probabilités sur les 3 produits d'espaces suivants :

$(\Omega \times T) \times M$	analyse n° 4
$(\Omega) \times (T \times M)$	analyse n° 5
$(\Omega \times M) \times T$	analyse n° 6

il s'agit donc, d'analyse canonique de deux tribus, ou en dimension finie de trois types d'analyses de correspondances.

* Analyse n° 4

Elle consistait à chercher pour un processus vectoriel des combinaisons linéaires fixes dans le temps des variables du processus, qui soient de variance maximale.

Ici dans le contexte d'analyse canonique il s'agit de chercher des fonctions des états (ou codages) et des fonctions du type $f(t,\omega)$ qui soient le plus corrélées possible, la loi de probabilité sur $(\Omega \times T) \times M$ étant définie par les lois conditionnelles sur M à $(\Omega \times T)$ fixé. $(\Omega \times T)$ étant muni de la mesure produit $P \times \frac{dt}{T}$. Or ces lois conditionnelles sont données par $\mathbb{1}_t^x(\omega)$: l'analyse est donc triviale et ne présente strictement aucun intérêt puisque la matrice à diagonaliser n'est autre que I_m .

Ce phénomène, dû au fait que la tribu sur M n'est qu'une sous-tribu de celle sur $\Omega \times T$, se voit clairement en dimension finie où il est absurde de faire l'analyse des correspondances d'un seul tableau de variables indicatrices puisque celles-ci sont orthonormées pour la métrique du χ^2 . Si on veut des codages fixes dans le temps la seule méthode possible est donc l'analyse n° 1.

* Analyse n° 5

On cherche ici des fonctions de Ω (coordonnées des individus) c'est-à-dire des variables aléatoires et des fonctions de temps et des états les plus corrélées possibles, la loi de probabilité sur le produit $\Omega \times (T \times M)$ étant définie très simplement par les lois à ω fixé $\mathbb{1}_t^x(\omega)$ si $T \times M$ est muni de la mesure produit de la mesure $\frac{dt}{T}$ avec la mesure de comptage sur les états.

Il est clair que ceci n'est qu'une autre façon de présenter l'analyse harmonique qualitative. On trouvera d'intéressants développements à ce sujet dans [8] déjà cité.

* Analyse n° 6

Il s'agit ici de chercher les fonctions du temps f(t) et les fonctions de ω et x les plus corrélées. Ce type d'analyse entièrement inédit consiste à chercher des codages des états variables selon chaque individu ou encore en d'autres termes à décomposer le processus $\vec{\mathbb{1}}_t$ selon une somme d'expressions de type $f(t) \,.\, \vec{z}$ où $\vec{z}$ est un vecteur aléatoire à m dimensions.

Les lois de probabilité conditionnelles à t fixé sont donc définies par l'ensemble des $\mathbb{1}_t^x(\omega)$ pour ω et x variables autrement dit des distributions aléatoires de probabilité. On a donc ici des données semblables à celles de l'analyse n° 2 mais où la dimension individuelle est conservée (possibilité de retrouver les trajectoires). Pour appliquer la métrique du χ^2 chaque couple (ω,x) doit être pondéré par l'inverse de

$\int_0^T \mathbb{1}_t^x(\omega)\, dt$ c'est-à-dire par l'inverse du temps passé $\Theta_x(\omega)$ dans l'état x par l'individu ω entre 0 et T ou par zéro si $\Theta_x(\omega)$ est nul.

Ceci revient à dire que les f(t) sont fonctions propres de l'opérateur intégral de noyau :

$$H(t,s) = \sum_x \int_\Omega \frac{\mathbb{1}_t^x(\omega)\, \mathbb{1}_s^x(\omega)}{\int_0^T \mathbb{1}_t^x(\omega)\, dt}\, d\,P(\omega)$$

avec la convention que si l'état x est presque sûrement impossible pour ω le terme correspondant est nul.

Si on introduit la matrice aléatoire

$D_\Theta = \mathrm{Diag}(\Theta_x)$ on peut écrire H(t,s) sous la forme plus concise :

$$H(t,s) = E(\vec{\mathbb{1}}_t \cdot D_\Theta^- \vec{\mathbb{1}}_s)$$

où D_Θ^- est l'inverse de MOORE-PENROSE de D_Θ .

On peut interpréter $\vec{\mathbb{1}}_t(\omega)\ D^-_{\Theta(\omega)}\ \vec{\mathbb{1}}_s(\omega)$ comme une mesure de similarité entre les instants t et s pour l'individu ω nulle si ω n'est pas dans le même état à t et à s, d'autant plus grande que ω est resté dans un état qu'il prend rarement au cours du temps.

H(t,s) est alors la moyenne de ces similarités individuelles.

La variable vectorielle $\vec{z}$ (codage aléatoire de M c'est-à-dire propre à chaque individu) s'obtient par intégration sur t

$$\vec{z} = \int_0^T \vec{\mathbb{1}}_t \cdot f(t)\ dt$$

Les fonctions du temps f(t) et les codages aléatoires $\vec{z}$ sont les solutions respectives des deux équations suivantes :

$$\boxed{\begin{aligned} \lambda\ f(t) &= \int_0^T E(\vec{\mathbb{1}}_t \cdot D_\Theta^- \vec{\mathbb{1}}_s)\ f(s)\ ds \\ \lambda\ \vec{z} &= \int_0^T \vec{\mathbb{1}}_t\ E(\vec{\mathbb{1}}_t \cdot D_\Theta^- \vec{z})\ dt \end{aligned}}$$

La deuxième équation s'obtenant aisément à partir de la première par multiplication par $\vec{\mathbb{1}}_t$ et intégration.

A titre d'exemple d'école, examinons les résultats de cette analyse pour le processus qualitatif à deux états qui nous sert de référence.

En notant Θ la variable aléatoire représentant la durée de séjour dans l'état 0 la fonction $H(t,s)$ s'écrit facilement car Θ suit une loi uniforme sur $[0,1]$:

$$H(t,s) = \int_0^1 \frac{\mathbb{1}_t^0(\theta)\,\mathbb{1}_s^0(\theta)}{\theta}\,d\theta + \int_0^1 \frac{\mathbb{1}_t^1(\theta)\,\mathbb{1}_s^1(\theta)}{1-\theta}\,d\theta$$

Comme $\quad \mathbb{1}_t^0(\theta)\,\mathbb{1}_s^0(\theta) = 1 \quad$ si $\sup(t,s) < \theta$

$\qquad\qquad\qquad\qquad\qquad = 0 \quad$ sinon

et $\quad \mathbb{1}_t^1(\theta)\,\mathbb{1}_s^1(\theta) = 1 \quad$ si $\inf(t,s) > \theta$

$\qquad\qquad\qquad\qquad\qquad = 0 \quad$ sinon

il vient

$$H(t,s) = \int_{\sup(t,s)}^1 \frac{1}{\theta}\,d\theta + \int_0^{\inf(t,s)} \frac{1}{1-\theta}\,d\theta$$

$$= -\operatorname{Log}(1-\inf(t,s)) - \operatorname{Log}\sup(t,s)\ .$$

Soit $\quad H(t,s) = \operatorname{Log}\dfrac{1}{t(1-s)} \qquad$ si $t > s$

$\qquad H(t,s) = \operatorname{Log}\dfrac{1}{s(1-t)} \qquad$ si $t < s$.

Les fonctions $f(t)$ sont alors les solutions de l'équation ;

$$\lambda\, f(t) = \int_0^t \operatorname{Log}\frac{1}{t(1-s)}\, f(s)\,ds + \int_t^1 \operatorname{Log}\frac{1}{s(1-t)}\, f(s)\,ds$$

Comme d'habitude on suppose la dérivabilité de $f(t)$ et on obtient par des calculs simples :

$$\lambda\, f'(t) = -\frac{1}{t}\int_0^t f(s)\,ds + \frac{1}{1-t}\int_t^1 f(s)\,ds$$

$$\lambda\ f''(t) = \frac{1}{t^2} \int_0^t f(s)\ ds + \frac{1}{(1-t)^2} \int_t^1 f(s)\ ds - \frac{1}{t(1-t)} f(t)$$

Une troisième dérivation est alors nécessaire pour éliminer plus facilement les intégrales

$$\lambda\ f'''(t) = -\frac{2}{t^3}\int_0^t f(s)ds + \frac{2}{(1-t)^3}\int_t^1 f(s)ds - \frac{1}{t(1-t)} f'(t) + \frac{2(1-2t)}{t^2(1-t)^2} f(t)$$

L'élimination de $\int_0^t f(s)\ ds$ et $\int_t^1 f(s)\ ds$ entre f' f'' et f''' conduit alors à une équation différentielle où ne figure plus f qui s'élimine tout seul :

$$\lambda\ f'''(t) = 2\lambda\ f''(t)\ \frac{2t-1}{t(1-t)} + \frac{2\lambda\ f'(t)}{t(1-t)} - \frac{f'(t)}{t(1-t)}$$

ou encore :

$$t(1-t)\ f'''(t) + 2(1-2t)\ f''(t) + (\frac{1}{\lambda} - 2)\ f'(t) = 0\ .$$

Cette équation a déjà été rencontrée au cours de ce chapitre c'est la même que celle qui fournissait les composantes $z(\theta)$ dans l'analyse harmonique qualitative du même processus. (I-4-a) : les fonctions f(t) sont une fois de plus proportionnelles aux polynômes de Legendre sur [0,1] et les valeurs propres sont ici les mêmes qu'en analyse harmonique qualitative $\lambda_k = \frac{1}{k(k+1)}$.

La condition de normalisation :

$$\sum_x \int_\Omega \int_0^T \mathbb{1}_t^x(\omega)\ dt\ [f(t)]^2\ dP(\omega) = 1$$

s'écrit ici de façon particulièrement simple :

$$\int_0^1 \left[\int_0^\theta [f(t)]^2\ dt + \int_\theta^1 [f(t)]^2\ dt \right] d\theta = 1$$

soit $\int_0^1 [f(t)]^2\ dt = 1$.

Autrement dit les fonctions f(t) sont les polynômes de Legendre normalisés sur [0,1].

On a donc

$$\boxed{\begin{aligned} f_k(t) &= \sqrt{2k+1} \quad P_k(2t-1) \\ \lambda_k &= \frac{1}{k(k+1)} \end{aligned}}$$

donc $f_1(t) = \sqrt{3}\ (2t-1)$

les codages individuels sont obtenus par la formule :

$$\vec{z}(\omega) = \int_0^T \vec{1}_t(\omega)\ f(t)\ dt\ .$$

donc en repérant ici un individu par sa durée θ de séjour dans l'état 1 :

$$\vec{z}(\theta) = \int_0^1 \vec{1}_t(\theta)\ f(t)\ dt$$

$$z(\theta) = \left(\int_0^\theta f(t)\ dt\ ;\qquad f(t)\ dt\right)$$

$$z_1(\theta) = \sqrt{3}\ (\theta^2 - \theta\ ;\ \theta - \theta^2)$$

les codages d'ordre 1 des deux modalités sont donc opposés pour chaque individu.

Sur un plan général une telle analyse soulève cependant de nombreuses difficultés d'ordre conceptuel : pour commencer que signifient des codages propres à chaque individu ? L'interprétation des codages obtenus ne peut être immédiate d'autant que les variables $\vec{z}$ ne sont ni centrées ni orthogonales.

En effet $E(\vec{z}) = \int_0^T E(\vec{1}_t)\ .\ f(t)\ dt = \int_0^T \vec{p}(t)\ f(t)\ dt$

et $\mathrm{cov}(z^i, z^j) = \iint [p^{ij}(t,s) - p^i(t)\ p^j(s)]\ f(t)\ f(s)\ dt\ ds$

ce qui retire beaucoup d'intérêt aux résultats sans oublier que pour chaque individu il existe aussi une suite de codages d'ordres différents car on ne peut évidemment se limiter à la solution d'ordre 1.

Quant aux difficultés d'ordre pratique elles sont plus grandes qu'en analyse harmonique qualitative, le calcul de $H(t,s)$ en tout couple (t,s) n'étant pas forcément simple sans parler des nécessaires questions d'approximation car il faut discrètiser le temps.

La méthode suivante fondée elle aussi sur une analyse de proximités entre instants nous paraît préférable par sa simplicité.

II-2) Analyse du processus $(\overset{\circ}{E}_t)$

II-2-a) La méthode

Comme en analyse vectorielle où on associait au vecteur aléatoire $\vec{X}_t$ un opérateur le caractérisant, on peut associer ici à chaque variable qualitative X_t le projecteur E_t et faire ensuite l'analyse factorielle du processus E_t .

L'aspect théorique de cette analyse ayant été abordé dans [8] nous n'y reviendrons pas mais développerons les considérations pratiques. Il s'agit ici de généraliser la méthode que nous avions proposée dans [57] d'analyse en composantes principales de variables qualitatives.

Le concept de base est celui de proximité entre variables qualitatives déduit de la proximité entre opérateurs associés. L'ensemble des opérateurs associés à une variable peut être muni du produit scalaire de la trace :

si E_t et E_s sont les projecteurs associés à X_t et X_s

on a $\langle E_t \; ; E_s \rangle = \text{Trace}\,(E_t\, E_s)$.

Il est bien connu que Trace $(E_t\ E_s) = 1 + \phi^2_{ts}$ où ϕ^2_{ts} est la mesure de dépendance de K. Pearson entre X_t et X_s

$$\phi^2_{ts} = \sum_x \sum_y \frac{(p^{xy}(t,s) - p^x(t)\ p^y(s))^2}{p^x(t)\ p^y(s)}$$

Comme il est préférable d'utiliser ϕ^2_{ts} plutôt que $1 + \phi^2_{ts}$ comme mesure de proximité entre X_t et X_s $\left(\phi^2_{ts} = 0\right.$ si il y a indépendance $\phi^2_{ts} = m - 1$ si il y a liaison fonctionnelle entre X_t et $\left.X_s\right)$ il convient d'utiliser les projecteurs "centrés" $\overset{\circ}{E}_t$ et $\overset{\circ}{E}_s$ c'est-à-dire les restrictions de E_t et E_s aux variables centrées codant X_t et X_s , plutôt que les projecteurs E_t et E_s

$$\overset{\circ}{E}_t = E_t - 1 \otimes 1$$

Ceci revient à éliminer la solution triviale dans l'analyse canonique (ou analyse des correspondances) de X_t et X_s .

Le fait de travailler sur un processus plutôt que sur un ensemble de variables qualitatives a ici un avantage décisif : les X_t ayant toutes le même nombre de modalités les ϕ^2 sont comparables il n'y a donc pas de problème de degré de liberté et il est indifférent alors de travailler sur les ϕ^2 ou sur les coefficients de Tschuprow $T^2 = \frac{\phi^2}{m-1}$. Rappelons qu'entre variables qualitatives le ϕ^2 est l'analogue d'une covariance et le T^2 d'une corrélation. Ceci permet de poser sans ambiguïté la définition suivante :

<u>Définition</u>

On appelle analyse factorielle globale du processus qualitatif (X_t) au sens des opérateurs, la recherche des fonctions propres de l'opérateur intégral de noyau ϕ^2_{ts}

$$\lambda\ f(t) = \int_0^T \phi^2_{ts}\ f(s)\ ds$$

Pour des variables à nombre fini de modalités m le ϕ^2 est toujours borné et l'opérateur intégral associé possède toutes les propriétés souhaitables : il est positif, symétrique, borné, nucléaire (sa trace vaut (m-1)T).

Comme ϕ^2_{ts} est toujours $\geqslant 0$ la fonction propre $f^{(1)}(t)$ associée à la plus grande valeur propre λ, est positive.

On obtient alors une formule de reconstitution de E_t du type :

$$\overset{\circ}{E}_t = \sum_{k=1}^{\infty} f^{(k)}(t)\, A^{(k)}$$

où les $A^{(k)}$ sont les opérateurs $A^{(k)} = \int_0^T \overset{\circ}{E}_t\, f^{(k)}(t)\, dt$.

On a alors une représentation factorielle des $\overset{\circ}{E}_t$ donc du processus (X_t) dans le premier plan factoriel par la courbe donnée par l'équation paramétrique

$$\begin{cases} \sqrt{\lambda_1}\ f^{(1)}(t) \\ \sqrt{\lambda_2}\ f^{(2)}(t) \end{cases}$$

Un processus pour lequel il n'y aurait aucune évolution temporelle serait tel que $X_t = X_s$ presque sûrement, d'où $\phi^2_{ts} = m - 1$, et se traduirait par une valeur propre multiple T d'ordre m-1 et se représenterait par un point isolé sur la représentation factorielle. Le simple fait d'obtenir une courbe sur le premier plan factoriel est donc une preuve du caractère évolutif du phénomène.

Par rapport à l'analyse harmonique qualitative on obtient une visualisation de l'évolution de la variable X_t et non de ses modalités ce qui apparaît comme une vision complémentaire. De plus cette analyse est une analyse des dépendances entre variables au sens du ϕ^2 .

Si le processus qualitatif (X_t)ne comporte que deux états on sait que le ϕ^2_{ts} est alors égal au carré du coefficient de corrélation linéaire entre les indicatrices de l'un des états entre les instants t et s .

On a donc une analyse voisine, au carré près, de l'analyse harmonique qualitative.

Le premier opérateur principal $A^{(1)}$ est encore un opérateur positif, mais ce n'est plus un projecteur, et son analyse spectrale a donc un sens : au lieu de chercher comme en analyse harmonique qualitative la solution de :

$$\int_0^T \overset{\circ}{E}_t \, z \, dt = \lambda z$$

On cherche alors celle de :

$$\int_0^T f^{(1)}(t) \, \overset{\circ}{E}_t \, z \, dt = \mu z$$

ce qui revient, si l'on veut, à mettre une autre mesure sur le temps.

On peut montrer comme nous l'avons fait en [59] que les valeurs propres de l'opérateur $A^{(1)}$ sont telles que $\sum_{k=1}^{\infty} (\mu_k)^2$ est maximal pour toutes les combinaisons linéaires d'opérateurs $\overset{\circ}{E}_t$, ce qui donne un certain caractère optimal à la méthode.

II-2-b) <u>Exemple</u>

Reprenons le processus qualitatif à deux états étudié au cours de ce chapitre. On sait que pour deux variables qualitatives à 2 états le ϕ^2 est égal au carré du coefficient de corrélation linéaire entre les indicatrices d'un des deux états de chacune. On a donc :

$$\phi^2_{t,s} = \frac{(1-s)t}{(1-t)s} \qquad \text{si } t < 1$$

$$\phi^2_{ts} = \frac{(1-t)s}{(1-s)t} \qquad \text{si } t \geqslant s \quad .$$

L'équation intégrale donnant les facteurs de l'analyse des projecteurs $\overset{o}{E}_t$ est alors :

$$\lambda\ f(t) = \int_0^t \frac{1-t}{t}\ \frac{s}{1-s}\ f(s)\ ds + \int_t^1 \frac{t}{1-t}\ f(s)\ ds$$

en supposant comme d'habitude la dérivabilité de f(t) on aboutit à une équation différentielle du second ordre

car $$\lambda\ f'(t) = -\frac{1}{t^2}\int_0^t \frac{s}{1-s}\ f(s)\ ds + \frac{1}{(1-t)^2}\int_t^1 \frac{1-s}{s}\ f(s)\ ds$$

et $$\lambda\ f''(t) = \frac{2}{t^3}\int_0^t \frac{s}{1-s}\ f(s)\ ds + \frac{2}{(1-t)^3}\int_t^1 \frac{1-s}{s}\ f(s)\ ds - \frac{2}{t(1-t)}\ f(t)$$

soit en éliminant les intégrales :

$$t(1-t)\ f'' + (1-2t)\ f' + f\left[\frac{2}{\lambda} - \frac{1}{t(1-t)}\right] = 0$$

On pose $t = \frac{1+u}{2}$ et $g(u) = f\left(\frac{1+u}{2}\right)$ pour se ramener à une forme standard et il vient :

$$\boxed{(1-u^2)\ g'' - 2ug' + g\left[\frac{2}{\lambda} - \frac{4}{1-u^2}\right] = 0}$$

On reconnait ici l'équation différentielle des fonctions de Legendre d'ordre 2 et de degré ν tel que $\lambda = \frac{2}{\nu(\nu+1)}$ (voir [1] p. 332) $P_\nu^2(u)$.

Le report de $P_\nu^2(u)$ dans l'équation intégrale permet de vérifier que seuls les degrés entiers et supérieurs ou égaux à deux conviennent : les valeurs propres λ_k sont donc données par :

$$\boxed{\lambda_k = \frac{2}{(k+1)(k+2)} \qquad k = 1,2,\ldots}$$

soit $\lambda_1 = 1/3$ $\lambda_2 = 1/6$ $\lambda_3 = 1/10$ etc. On a bien sur $\Sigma \lambda_k = 1$ et la fonction $g_k(u)$ s'expriment au moyen des polynômes de Legendre d'ordre k+1

$$g_k(u) = (1-u^2) \frac{d^2 P_{k+1}(u)}{du^2}$$

ce sont donc des polynômes.

Pour obtenir des solutions normalisées il faut calculer

$$\int_0^1 [f_k(t)]^2 \, dt = \frac{1}{2} \int_{-1}^{+1} [g_k(u)]^2 \, du \quad .$$

Cette dernière intégrale vaut $\dfrac{(k+3)!}{(k-1)!(k+3/2)}$ (cf. [1] p. 338)

d'où :

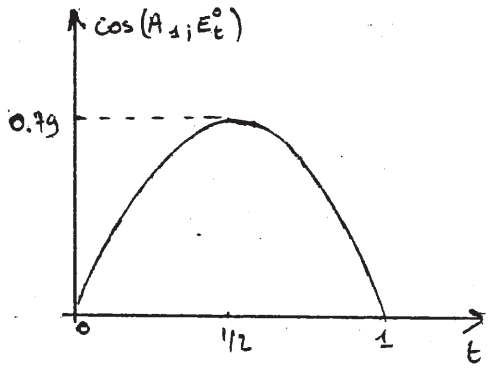
$$\boxed{g_k(\upsilon) = \sqrt{\frac{(2k+3)(k-1)!}{(k+3)!}} \, (1-u^2) \, \frac{d^2 P_{k+1}(u)}{du^2}}$$

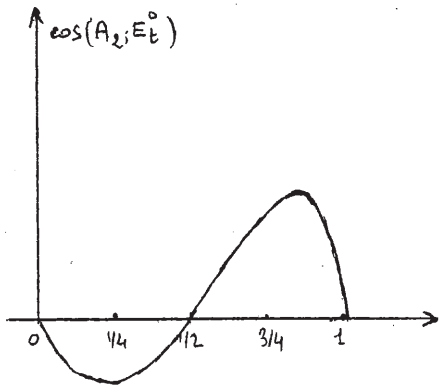
On a ainsi :

$$f_1(t) = \sqrt{30} \; t(1-t)$$

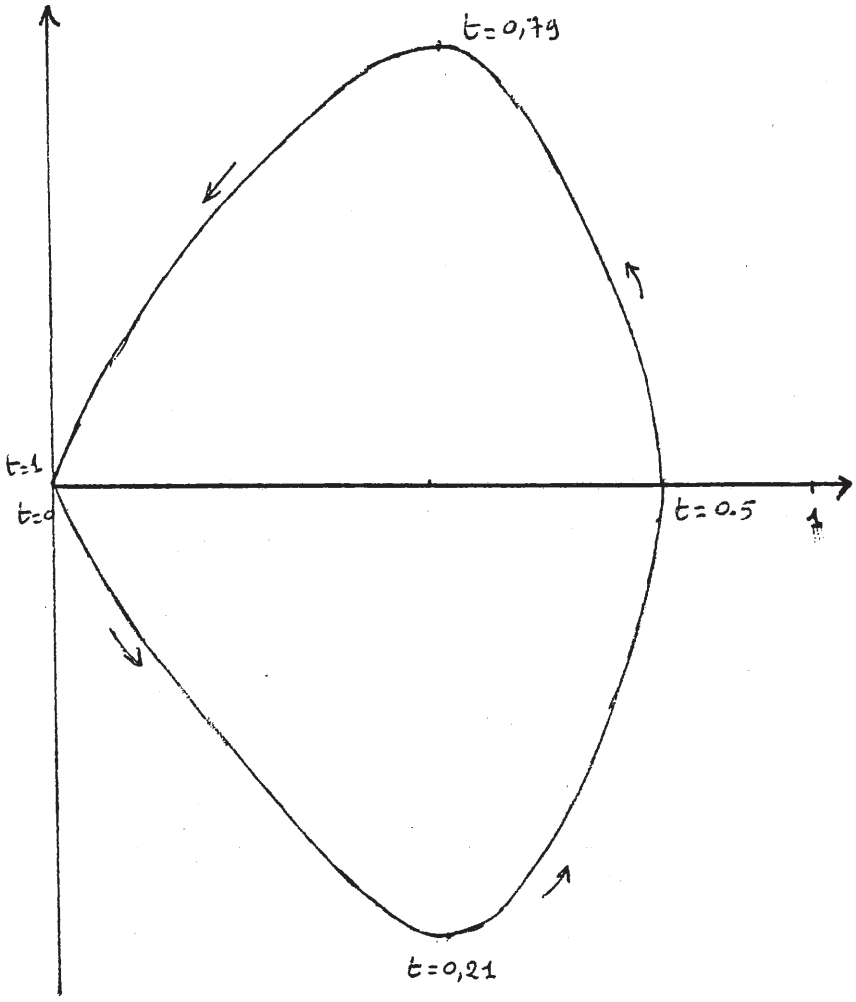
$$f_2(t) = \sqrt{210} \; t(1-t)(2t-1) \; .$$

Les courbes représentatives des cosinus entre opérateurs principaux et projecteurs sont les suivantes :





La représentation du projecteur $\mathring{E}_t$ dans le cercle des "corrélations" est alors la suivante :



Le caractère évolutif du phénomène est bien mis en évidence ainsi que le rôle spécial de l'instant t = 1/2 après lequel il y a exact renversement de la figure. L'interprétation de ce genre d'analyse demande cependant pas mal de prudence ainsi le fait d'avoir une courbe fermée avec confusion de t = 0 et de t = 1 ne signifie nullement que les variables X_1 et X_0 sont identiques ni que l'on soit retourné dans la même situation mais que les variables X_1 et X_0 ne sont pas discriminantes (absence de variabilité).

Par rapport à l'analyse des corrélations entre variables indicatrices (qui sont ici égales à ϕ_{ts}) l'analyse des ϕ^2_{ts} fournit des valeurs propres plus faibles ce qui n'est pas surprenant puisque les coefficients de corrélation sont élevés au carré.

III- GENERALISATION AU CAS D'UN PROCESSUS QUALITATIF MULTIPLE

Nous supposerons désormais qu'on observe sur une population Ω l'évolution simultanée de p variables qualitatives $X^1, X^2, \ldots, X^p$ donc que l'on se donne un processus qualitatif multiple $(X^1_t, X^2_t, \ldots, X^p_t)$.

III-1) Nature des données et position du problème

On notera $m_1\ m_2 \ldots m_p$ le nombre d'états de chaque processus qualitatif $X^1, X^2 \ldots, X^p$. On supposera que ces divers processus vérifient les conditions de régularité énoncées au paragraphe I mais il est clair que la difficulté ne réside pas dans la nature des hypothèses mathématiques mais dans la pertinence des méthodes d'analyse.

Une première possibilité parfaitement justifiée sur le plan théorique consisterait à considérer qu'il n'existe en réalité qu'une seule variable qualitative X à $\prod_{i=1}^{p} m_i$ modalités obtenue par produit des X^i et qu'il suffit de faire l'analyse factorielle du processus qualitatif ainsi engendré : une telle démarche, exhaustive, est hélas impraticable en réalité car le produit du nombre de modalités prend rapidement des valeurs prohibitives. On se contentera d'utiliser les dépendances deux à deux entre les X^i et non les dépendances globales.

Une deuxième possibilité consiste à remplacer chaque processus qualitatif X^i_t par le processus vectoriel de ses indicatrices $\vec{1}^i_t$ et à considérer que $(X^1_t, \ldots, X^p_t)$ est équivalent au processus vectoriel$(\vec{1}_t)$ à $\Sigma\, m_i$ composantes obtenu en réunissant toutes les variables indicatrices de tous les états. C'est grosso modo ce qui sera fait dans le paragraphe 3 à condition de respecter le groupage des indicatrices d'une même variable,

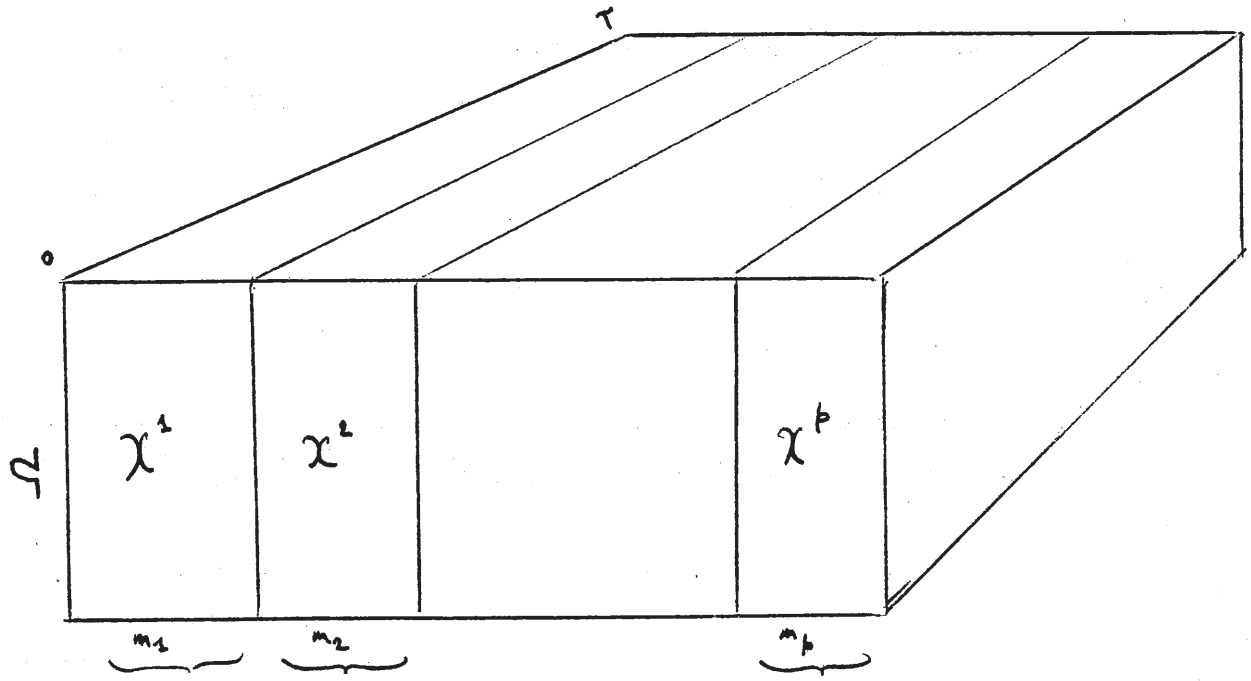
cependant le fait que $\sum_{j=1}^{m_i} \mathbb{1}_t^{ij} = 1$ pour chaque variable X^i est la source de quelques difficultés numériques.

Une troisième possibilité largement équivalente à la précédente est le codage : on transforme le processus qualitatif multiple en un processus numérique vectoriel par codage de chaque processus qualitatif X_t^i et il suffit alors de faire l'analyse harmonique vectorielle du processus ainsi obtenu ; on cherchera bien entendu un codage "optimal" en un sens qui sera défini plus loin.

Ces trois approches n'épuisent évidemment pas le sujet et on peut transposer au cas qualitatif multiple les diverses analyses proposées au chapitre II.

III-2) <u>Les diverses analyses d'un cube qualitatif</u>

On peut se représenter schématiquement les données par la figure suivante :



– Si le nombre d'individus est fini (n) on a un "cube" tel que chaque coupe instantanée est le tableau disjonctif complet des variables $(X_t^1\ X_t^2\ \ldots\ X_t^p)$.

On peut alors passer en revue les 6 analyses possibles de ce "cube" selon la même démarche qu'au chapitre II ou au chapitre III. On verra au passage qu'une des analyses triviales pour le cas d'une seule variable qualitative reprend un certain intérêt dans le cas qualitatif multiple.

Parmi les analyses marginales l'analyse n° 3 reste toujours sans intérêt puisque la sommation sur les variables indicatrices donne un processus constant.

L'analyse n° 2 (sommation sur les individus) conduit toujours à examiner l'évolution des distributions marginales des probabilités de chaque X^i au moyen d'une analyse des correspondances et l'analyse n° 1 à examiner pour chaque individu la liste des temps passés dans chaque état de chaque processus.

Ces deux types d'analyses marginales semblent ici beaucoup moins intéressants que dans le cas qualitatif unique puisqu'on ne tient guère compte des dépendances entre variables. Par ailleurs l'analyse n° 1 n'est plus susceptible de l'interprétation comme recherche de codages fixes dans le temps fournissant la meilleure moyenne temporelle puisque les séjours dans les différents états ne se passent pas au même moment selon les variables.

Passons aux analyses mixtes :

L'analyse n° 4 qui était sans intérêt dans le cas d'une seule variable qualitative ne devient plus triviale ici.

En effet si T est discret cela revient à faire l'analyse des correspondances de la superposition des tableaux disjonctifs à t fixé.

Cette analyse est donc équivalente [14] à l'analyse des correspondances de la somme des tableaux de BURT correspondant à chaque instant. On obtient ici une généralisation d'une technique étudiée par FOUCART [27] pour l'analyse d'une suite de tableaux de contingence (p serait égal à 2) consistant à effectuer l'analyse du tableau de contingence moyen et à reporter les autres tableaux en éléments supplémentaires. Si ce type d'analyse est légitime dans le cas des données étudiées en [27] elle n'est pas adaptée au cas des données individuelles puisqu'on perd totalement l'information apportée par les trajectoires : on n'étudie en réalité que la déformation des dépendances entre variables à t fixé (dépendances intra temporelles) et absolument pas les dépendances entre variables où des instants différents (dépendances inter-temporelles) ; il ne s'agit donc pas d'une méthode d'analyse des trajectoires.

L'analyse n° 6 qui conduit à des codages de chaque variable propre à chaque individu nous paraît ici encore plus difficilement interprétable avec p variables et nous ne la développerons pas.

Il reste alors une seule technique l'analyse n° 5 qui revient à étudier la juxtaposition horizontale des tableaux disjonctifs ce qui conduit à la recherche de coordonnées des individus fixes dans le temps et de codages évolutifs de chaque variable : c'est cette analyse identique à l'analyse harmonique vectorielle des $\Sigma\, m_i$ indicatrices qui constitue donc la véritable généralisation de l'analyse harmonique qualitative.

III-3) <u>L'analyse harmonique qualitative généralisée</u>

La voie la plus commode pour généraliser l'analyse harmonique qualitative en cas de processus multiples nous semble de partir de la propriété suivante de l'A.H.Q. :

"l'A.H.Q. de (X_t) revient à coder (X_t) de telle sorte que l'analyse factorielle de (X_t) codé, avec la métrique D_{1/σ^2} donne une première valeur propre maximale".

Ici coder $(X_t^1, \ldots, X_t^p)$ revient à obtenir un processus vectoriel $(X^1(t), X^2(t), \ldots, X^p(t) = \vec{X}(t)$, la généralisation naturelle de la propriété

de l'A.H.Q. consiste alors à obtenir une A.H.V. de $\vec{X}_t$ optimale au sens de la première valeur propre maximale : il n'y a ici aucune ambiguïté concernant le choix de la métrique.

<u>Définition</u>

L'analyse factorielle d'un processus qualitatif multiple consiste dans la recherche simultanée de variables z orthogonales deux à deux et de codages optimaux associés $\vec{X}_t$ tels que

$$\int_C^T R^2(z\ ;\ \vec{X}_t)\ dt \text{ soit maximal .}$$

D'après les résultats établis au chapitre II z est une combinaison linéaire des $\vec{X}_t$ $\quad z = \int_C^T \langle \vec{X}_t\ ;\ \vec{f}(t)\rangle dt$ et est vecteur propre de la somme $\int_0^T A_t\ dt$ des opérateurs de projection A_t sur $\vec{X}_t$.

$\vec{X}_t$ est un codage des X^i donc X_t^i est une combinaison linéaire des indicatrices de X^i : on en déduit que z est aussi une combinaison linéaire des indicatrices de tous les X^i

$$z = \int_C^T \sum_{i=1}^{p} \langle \vec{\mathbb{1}}_t^i\ ;\ \overrightarrow{f^i}(t)\rangle dt$$

les $\overrightarrow{f^i}(t)$ donnant les codages des X^i .

La solution optimale sera alors obtenue pour z premier vecteur propre de la somme des projecteurs sur l'ensemble des $\vec{\mathbb{1}}_t^i$ $\quad i = 1,2,\dots p$.

Si on note $\mathcal{A}_t$ le projecteur sur l'ensemble des $\Sigma\ m_i$ indicatrices des X_t^i on a donc :

$$\lambda\ z = \int_C^T \mathcal{A}_t\ z\ dt \quad .$$

On a donc la propriété suivante

> Proposition 3.10 :
>
> L'A.H.Q. généralisée de $(X_t^1, X_t^2 \ldots X_t^p)$ n'est autre que l'A.H.V. du processus obtenu par réunion des $\Sigma\, m_i$ indicatrices des X_t^i .

On obtient bien sûr une première solution triviale puisque la variable constante fait partie de tous les sous-espaces engendrés par les indicatrices.

La somme des valeurs propres non triviales vaut alors

$$T \,.\, \sum_{i=1}^{p} (m_i - 1)$$

puisque les $\mathring{\mathcal{A}}_t$ sont de rang $\Sigma(m_i-1)$.

A la solution triviale près il est donc équivalent d'effectuer une analyse centrée ou non centrée.

Si on cherche à expliciter le projecteur $\mathcal{A}_t$ on a en notant $\vec{\mathbb{1}}_t$ le vecteur aléatoire des $\Sigma\, m_i$ indicatrices

$$E(\vec{\mathbb{1}}_t \otimes \vec{\mathbb{1}}_t) = B_t$$

où B_t est le tableau de BURT des p variables qualitatives à l'instant t .

B_t est alors une matrice carrée de taille $\Sigma\, m_i$ mais de rang au plus égal à $\Sigma(m_i-1)$.

L'expression de $\mathcal{A}_t$ est donc :

$$\mathcal{A}_t = \vec{\mathbb{1}}_t \,.\, [B_t]^- E(\vec{\mathbb{1}}_t \,.\,)$$

où B_t^- est un inverse généralisé quelconque de B_t .

Ceci implique que les codages optimaux des X_t^i ne sont pas uniques : cependant si on impose comme il est d'usage que ces codages soient centrés la solution est alors unique et peut être obtenue aisément de la manière suivante :

il suffit de remplacer à chaque instant les $\Sigma\, m_i$ indicatrices par $\Sigma(m_i-1)$ variables numériques linéairement indépendantes et centrées et d'effectuer ensuite l'A.H.V. du processus vectoriel de ces $(\Sigma(m_i-1))$ variables.

Cette analyse s'effectuera plus aisément si les variables de remplacement sont orthogonales (i.e. non corrélées) cela sera le cas en particulier si on choisit comme variables les composantes de l'analyse des correspondances du tableau disjonctif complet à chaque instant d'où la proposition suivante.

<u>Proposition 3.11</u> :

L'analyse factorielle de $(X_t^1\ X_t^2\ \dots\ X_t^p)$ est équivalente à l'A.H.V. du processus $(\vec{Z}_t)$ constitué par les composantes des A.F.C. instantanées des X_t^i .

Si à chaque instant on prend des composantes principales normées à 1 (et non à la valeur propre) on a alors une simplification notable puisque $C(t,t) = I \quad \forall\, t$.

Sur le plan pratique l'utilisation du système des composantes de l'A.F.C. à l'instant t est préférable à tout autre système de variables car il permet l'interprétation fine des dépendances à l'instant t .

La méthode proposée est donc en quelque sorte une double analyse canonique généralisée puisqu'on effectue l'A.C.G. (continue) des composantes des A.C.G. à chaque instant.

Remarque 1 : Il faut pour obtenir la solution exacte utiliser toutes les composantes (c'est-à-dire $\Sigma\, m_i - 1$) des analyses factorielles à chaque instant et ne pas se contenter uniquement des premières car on risque de faire disparaitre des permanences dans le temps de phénomènes décrits par les derniers axes à chaque t .

En notant $z_t^{(k)}$ la $k^{\text{ième}}$ composante de l'A.F.C. des X_t^i on voit que la variable z cherchée rend maximale la quantité :

$$R^2(z\ ;\ \vec{Z}_t)\ dt = \sum_{k=1}^{\Sigma(m_i-1)} \int_C^T r^2(z\ ;\ z_t^{(k)})\ dt$$

puisque $\vec{Z}_t$ est constitué de composantes non corrélées à tout instant.

Remarque 2 : Sur le plan numérique il existe une solution beaucoup plus simple, si on désire seulement obtenir les coordonnées des individus : il suffit de supprimer une des indicatrices de chaque variable qualitative (par exemple on supprime la dernière), ce qui donne à tout instant un système de plein rang $\Sigma\, m_i - p$, et d'effectuer une analyse harmonique vectorielle sur la réunion des indicatrices restantes.

Ceci revient à généraliser pour un nombre quelconque de variables et de modalités le procédé de réduction utilisé pour les processus qualitatifs à deux états (cf. I-4).

Remarque 3 : La méthode proposée de double A.C.G. est une extension au cas temporel d'une pratique recommandable mais peu utilisée en analyse des questionnaires où les questions sont regroupées par blocs d'items de nature semblable (questions politiques, questions concernant le niveau d'éducation,...) ; pour éviter que les résultats de l'analyse ne soient influencés par la taille des groupes de questions, il convient alors d'effectuer des analyses de correspondances multiples séparées dans chaque groupe puis ensuite de procéder à une analyse canonique généralisée des composantes de toutes ces analyses ce qui a pour effet de donner le même poids à chaque groupe de questions.

III-4) Analyses globales au moyen des opérateurs

III-4-a) Analyse factorielle des p opérateurs associés aux processus X_t^i

Comme en analyse harmonique vectorielle on peut comparer les diverses composantes du processus multiple en associant à chacune son opérateur d'Escoufier : ici l'opérateur associé à X_t^i est $\int_0^T \overset{\circ}{E}{}_t^i \, dt = Q^i$

(on choisit pour éviter les solutions triviales l'opérateur centré) et le produit scalaire habituel entre opérateurs permet d'effectuer l'analogue d'une analyse en composantes principales sur les processus :

on forme la matrice p x p de terme général

$$\text{Trace } (Q^i \, Q^j)$$

qui est l'analogue d'une matrice de covariance et sa diagonalisation suivie de la représentation factorielle plane usuelle permet de juger des similitudes ou des différences entre les évolutions des processus X_t^i .

Si on développe Trace$(Q^i \, Q^j)$ il vient

$$\text{Trace } Q^i \, Q^j = \text{Trace} \int_0^T \overset{\circ}{E}{}_t^i \, dt \int_0^T \overset{\circ}{E}{}_s^j \, ds$$

$$= \text{Trace} \iint \overset{\circ}{E}{}_t^i \, \overset{\circ}{E}{}_s^j \, dt \, ds$$

$$= \iint \text{Trace}(\overset{\circ}{E}{}_t^i \, \overset{\circ}{E}{}_s^j) \, dt \, ds = \iint \phi^{2}{}_{ij}(t,s) \, dt \, ds$$

où $\phi^{2}{}_{ij}(t,s)$ est la mesure de dépendance de K. PEARSON entre X_t^i et X_s^j que l'on peut interpréter comme une covariance croisée.

Trace $Q^i \cdot Q^j$ est donc une moyenne des dépendances entre X^i et X^j .

Cette mesure n'est pas insensible aux différences entre les nombres de modalités des variables X^i et X^j et si m_i est différent de m_j il y a intérêt à normaliser c'est-à-dire à passer aux coefficients de TSCHUPROW en considérant :

$$\frac{\text{Trace } Q_i \, Q_j}{\sqrt{(m_i-1)(m_j-1)}} = \iint T^2_{ij}(t,s) \, dt \, ds$$

III-4-b) Analyse d'opérateurs associés aux instants

On peut concevoir au moins deux types d'analyses; une scalaire où à chaque instant t on associe un opérateur représentatif du p-uple $(X^i_t \ldots X^p_t)$ l'autre vectorielle où à chaque instant t on associe le p-uple des opérateurs représentatifs des X^i_t .

L'analyse vectorielle est la généralisation naturelle de la pratique des opérateurs en tant qu'objets mathématiques associés à une variable qualitative.

On remplace le processus $(X^1_t, \ldots, X^p_t)$ par le processus des opérateurs de projection centrés associés aux X^j_t :

$$(\mathring{E}^1_t \; ; \; \mathring{E}^2_t \; ; \; \ldots \; ; \; \mathring{E}^p_t) \; .$$

Il suffit alors de généraliser formellement l'A.H.V. à ce processus puisqu'on dispose d'une "covariance" entre opérateurs égale au ϕ^2 .

Définition

L'A.H.V. au sens des opérateurs du processus qualitatif multiple $(X_t^1, X_t^2 \ldots ; X_t^p)$ consiste dans la recherche des solutions de l'équation

$$\lambda \vec{f(t)} = \int_0^T [\Phi(t,t)]^{-1} [\Phi(t,s)] \vec{f(s)} \, ds$$

où $\Phi(t,s)$ est la matrice p x p des Φ^2 de PEARSON entre les X_t^i et les X_s^j .

On notera ici que le nombre de modalités de chaque X^i est ici sans influence sur les résultats de l'analyse et qu'il est équivalent d'utiliser les ϕ^2 ou les T^2 .

Il est donc facile d'effectuer exactement l'analogue de l'A.H.V. pour p processus qualitatifs. Nous renvoyons donc au chapitre II et au paragraphe II de ce chapitre pour les interprétations.

L'analyse de type scalaire consiste à associer à chaque date t un opérateur unique synthétisant les p variables qualitatives X_t^i : le choix d'un tel opérateur ne va pas cependant sans difficulté.

Le plus naturel est sans doute le projecteur sur l'espace engendré par les indicatrices mais celles-ci étant liées par une relation linéaire il faut soit passer par une base des indicatrices soit en éliminer une par variable pour obtenir une matrice facilement calculable ; on est alors exactement ramené au cas étudié au chapitre II § IV-2.

Le plus simple est d'utiliser $\sum_{i=1}^{p} \overset{\circ}{E}{}_t^i$ puisque c'est l'opérateur caractéristique de l'A.F.C. à l'instant t et de chercher les fonctions propres de l'opérateur de "covariance" associé aux $\Sigma \overset{\circ}{E}{}_t^i$. Comme

$$\text{Trace}\left(\sum_i \overset{\circ}{E}{}_t^i \, . \sum_j \overset{\circ}{E}{}_s^j \right) = \sum_i \sum_j \text{Trace}\left(\overset{\circ}{E}{}_t^i \, . \, \overset{\circ}{E}{}_s^j\right)$$

On en déduit que la "covariance" entre opérateurs associés aux instants vaut $\sum_i \sum_j \phi^{2_{i,j}}(t,s)$ et est analogue à la moyenne des ϕ^2 entre instants t et s pour toutes les variables. On a ainsi une mesure "duale" en quelque sorte de celle étudiée au sous-paragraphe a) pour les processus. Cependant un tel choix à le défaut de donner a priori à chaque variable X_t^i le même poids même si cette variable est peu informative (cas d'une variable qualitative telle que tous les individus restent dans les mêmes états pendant une durée assez longue ou quasi indépendante des autres : on peut évidemment corriger ce défaut en pondérant les opérateurs c'est-à-dire en prenant une formule du type $\sum_i \alpha^i(t)\, \overset{\circ}{E}{}_t^i$ les $\alpha^i(t)$ étant choisis par exemple en utilisant la technique d'ESCOUFIER-PAGES et al. [52] mais il semble préférable à tous égards d'utiliser la méthode d'analyse vectorielle des opérateurs qui fournira directement et sans arbitraire les coefficients $\alpha^i(t)$: un faible $\alpha^i(t)$ indiquant que la variable X_t^i est peu intéressante.

CONCLUSION

Au terme de ce chapitre seules trois méthodes paraissent bien adaptées à l'étude des processus qualitatifs uniques :

une analyse marginale, celle des temps passés dans chacun des états du processus, à condition qu'il y ait un ordre naturel de passage entre états,

une analyse globale des dépendances entre variables qualitatives au moyen des opérateurs et des ϕ^2 ,

et surtout l'analyse harmonique qualitative qui généralise l'analyse des correspondances multiples et permet les représentations et les interprétations les plus riches.

Quant aux processus qualitatifs multiples leur traitement ne soulève pas de difficulté particulière et est une simple généralisation du cas qualitatif unique.

CHAPITRE IV

ÉTUDE DES LIAISONS ENTRE VARIABLES ALÉATOIRES ET PROCESSUS

Un processus est rarement étudié seul : en plus d'un ensemble de trajectoires on possède généralement des mesures d'un ou plusieurs caractères quantitatifs ou qualitatifs.

L'étude des dépendances entre un processus et des variables s'impose souvent, l'analyse de ces liaisons pouvant être soit symétrique soit dissymétrique selon que l'on veut utiliser les données d'évolution pour prévoir une ou plusieurs variables ou a contrario étudier l'influence de certaines variables sur un phénomène évolutif.

Nous examinons donc un certain nombre de méthodes d'analyse de données, en particulier factorielles, qui permettent de décrire les liaisons entre les phénomènes étudiés, en commençant par rappeler les problèmes mathématiques spécifiques posés par la régression en dimension infinie qui la rendent très différente du cas fini (à l'inverse des méthodes factorielles des chapitres précédents). Nous suggérons alors quelques propositions concernant la prévision non linéaire.

L'étude de l'influence d'un processus sur un vecteur aléatoire est l'occasion de généraliser à un processus l'A.C.P. sur variables instrumentales, après avoir étudié le cas de l'analyse canonique qui donne le même rôle aux deux ensembles de variables.

Le problème inverse (influence de variables sur un processus) aboutit à l'étude d'analyses factorielles conditionnelles que l'on compare avec d'autres techniques.

Enfin ce chapitre s'achève par des remarques concernant l'étude conjointe de deux processus, mais le sujet semble ici relever d'autres techniques que celles de l'analyse des données.

I. REGRESSION D'UNE VARIABLE ALEATOIRE REELLE Y SUR UN PROCESSUS (X_t)

Le problème concret sous-jacent est le suivant : on connaît sur un ensemble d'individus l'évolution d'une variable aléatoire X_t entre [0 et T] ainsi que la valeur d'une v.a. Y . On désire alors établir une formule de prévision de Y à partir des X_t . Un cas particulier important est celui de l'extrapolation du processus (X_t) lui-même à un horizon T+h : il s'agit alors de prédire X_{T+h} à l'aide de X_t $t \in [0,T]$.

On sait que le meilleur prédicteur de Y au sens de la norme de $L^2(\Omega)$ est l'espérance de Y conditionnellement à la tribu engendrée par les X_t . Malheureusement cette solution est en général impossible à expliciter et on se contentera de chercher une prédiction Y dans une classe restreinte de fonctions des X_t la plupart du temps linéaire. Nous rappellerons dans ce qui suit les principaux résultats de la théorie de la prédiction linéaire, ou théorie de Wiener-Kolmogoroff, afin d'en déduire des conséquences pratiques pour l'analyse de données et proposons ensuite quelques méthodes simples de prévision non linéaire.

I.1 La prévision linéaire

I.1.a) Position du problème

Notons L^2_X le sous-espace de Hilbert de $L^2(\Omega)$ engendré linéairement par le processus (X_t) , c'est-à-dire l'adhérence du sous-espace engendré par les combinaisons linéaires d'un nombre fini de X_t . L^2_X est également le plus petit espace de Hilbert contenant toutes les variables X_t et le complété de Im U où U est l'opérateur de $L^2(T)$ dans $L^2(\Omega)$ introduit au chapitre I et défini par :

$$U \circ f = \int_0^T X_t \, f(t) \, dt$$

Notons tout de suite que $\mathrm{Im}\, U \neq \overline{\mathrm{Im}\, U}$ ce qui est à l'origine des problèmes spécifiques de la régression sur processus.

Etant donné une variable aléatoire $Y \in L^2(\Omega)$, la meilleure prévision linéaire de Y à partir de (X_t) est la projection $\hat{Y}$ de Y sur L^2_X. Le théorème de projection nous indique alors que $\hat{Y}$ existe et est l'unique élément de L^2_X tel que :

$$E(Y\, X_t) = E(\hat{Y}\, X_t) \qquad \forall\, t$$

ce qui prouve que la fonction φ définie par

$$\varphi(t) = E(Y\, X_t)$$

appartient au complété de Im C (mais pas nécessairement à Im C).

La question se pose souvent de trouver une expression explicite de $\hat{Y}$ sous la forme $\hat{Y} = \int_0^T X_t\, \beta(t)\, dt$ avec $\beta(t) \in L^2(T)$. Si une telle écriture était légitime on en déduirait :

$$\underline{\varphi}(t) = E((\int_0^T X_s\, \beta(s)\, ds)\, X_t) = \int_0^T E(X_s X_t)\, \beta(s)\, ds$$

d'où l'équation de Wiener-Hopf :

$$\varphi(t) = \int_0^T C(t,s)\, \beta(s)\, ds$$

Or si $\varphi(t)$ n'appartient pas à Im C cette équation n'admet pas de solution dans $L^2(T)$. Comme il est dit en [36], "l'équation de Wiener-Hopf n'est pas une équation intégrale ordinaire mais un accouplement entre fonction et distribution dont la solution est plus souvent une distribution qu'une fonction". En fait comme le remarque E. PARZEN [54] la question est mal posée et chercher le meilleur prédicteur linéaire n'est pas tant résoudre une équation intégrale que trouver une représentation linéaire de $\varphi(t)$ selon le noyau C(t,s) : en effet les équations précédentes signifient en réalité que si on peut trouver une représentation de $\varphi(t)$ en termes d'opérations linéaires (incluant la dérivation, l'intégration etc.) sur les C(t,.) alors $\hat{Y}$ s'obtient en effectuant les mêmes opérations sur les X_t. Remarquons de plus que la solution $\beta(t)$, si elle existe au sens de $L^2(T)$, n'est pas nécessairement unique sauf si C est injectif.

Le fait que β puisse être une distribution et pas une vraie fonction interdit donc d'utiliser pour l'équation de Wiener-Hopf les méthodes ordinaires de collocation qui ramènent l'intégrale à une somme finie par des formules de quadrature. On est donc ici dans une situation très différente de celle de l'analyse factorielle.

E. PARZEN [54] a donné la solution suivante du problème de la prévision linéaire :

soit H_{rep} l'ensemble des fonctions de t telles que :

$$f(t) = E(X_t Z) \quad \text{où} \quad Z \in L^2_X$$

muni du produit scalaire associé à la norme définie par

$$\|f\|^2_H = E(Z^2)$$

H_{rep} est un espace de Hilbert à noyau reproduisant

$$C(.,s) \in H_{rep} \quad \forall s \quad \text{et} \quad (f|C(t,.))_H = f(t)$$

L'application qui à $Z \in L^2_X$ associe $f(t) = E(X_t Z)$ est alors une isométrie de L^2_X sur H_{rep}. En notant $(f|X)_H$ la variable aléatoire de L^2_X qui correspond par cette isométrie à f élément de H_{rep}, la projection $\hat{Y}$ de Y sur L^2_X est donnée par

$$\hat{Y} = (\varphi|X)_H$$

Dans le cas fini on peut remarquer que la norme de H_{rep} correspond à la métrique de Mahalanobis. Cependant dans le cas général, la solution proposée est abstraite et ne donne pas une construction de $\hat{Y}$.

Il est en outre utile de savoir si la solution de l'équation de Wiener-Hopf appartient ou non à $L^2(T)$ c'est-à-dire si $\varphi(t) \in \operatorname{Im} C$. Le théorème de Picard sur les équations intégrales de Fredholm de 1ère espèce fournit à la fois la réponse à cette question et une méthode de calcul.

Théorème de Picard

L'équation $\int_0^T C(t,s)\,\beta(s)\,ds = \varphi(t)$ admet une solution unique dans $L^2(T)$ si

. $C(t,s)$ est réel symétrique

. la série $\sum_{k=1}^{\infty} \left(\frac{c_k}{\lambda_k}\right)^2$ est convergente

. les $f_k(t)$ forment une base hilbertienne de $L^2(T)$

la solution est alors $\beta(t) = \sum_{k=1}^{\infty} \frac{c_k\, f_k(t)}{\lambda_k}$

où les $f_k(t)$ sont les fonctions propres normées de l'opérateur C associées aux valeurs propres λ_k et $c_k = \int_0^T \varphi(t)\, f_k(t)\, dt$.

Dans notre contexte statistique $C(t,s)$ est évidemment réel symétrique, les $f_k(t)$ forment seulement une base de $\overline{\mathrm{Im}\, C}$ mais ceci est suffisant car $\varphi(t) \in \overline{\mathrm{Im}\, C}$. La condition cruciale est donc $\Sigma \left(\frac{c_k}{\lambda_k}\right)^2 < \infty$.

Nous allons voir cependant que la méthode proposée qui revient exactement à la technique de régression sur composantes principales fournit dans tous les cas le prédicteur optimal $\hat{Y}$ ainsi que des solutions approchées utiles pour les applications numériques.

I.1.b) Régression sur composantes principales

Si (X_t) est décomposé selon Karhunen-Loeve

$$X_t = \Sigma\, \xi_k\, f_k(t)$$

comme les ξ_k forment une base orthogonale de L^2_X il suffit de régresser séparément Y sur chaque ξ_k et d'ajouter les variables ainsi obtenues pour obtenir la projection de Y sur L^2_X . D'où :

Proposition 4-1

Le meilleur prédicteur linéaire de Y par (X_t) est

$$\hat{Y} = \sum_{k=1}^{\infty} \frac{E(Y\,\xi_k)}{\lambda_k} \cdot \xi_k$$

Si de plus l'écriture $\hat{Y} = \int_0^T X_t\,\beta(t)\,dt$ a un sens

$\beta(t) = \sum_{k=1}^{\infty} \frac{c_k\,f_k(t)}{\lambda_k}$ est une solution possible.

En effet $E(Y\,\xi_k) : E(Y \int_0^T X_t\,f_k(t)\,dt) = \int_0^T \varphi(t)\,f_k(t)\,dt = c_k$

d'où $\beta(t)$ par identification sous réserve de la légitimité des permutations des opérations de sommation.

Définition

On appelle coefficient de corrélation multiple R entre Y et (X_t) le cosinus de l'angle formé par Y et L_X^2 .

Proposition 4,2

Le carré du coefficient de corrélation multiple entre Y et (X_t) vaut :

$$R^2 = \frac{1}{V(Y)} \sum_{k=1}^{\infty} \frac{c_k^2}{\lambda_k}$$

En effet les ξ_k forment une base orthogonale de L_X^2

$R^2 = \sum_k r^2(Y\,;\,\xi_k)$. Comme $\mathrm{cov}(Y\,;\,\xi_k) = E(Y\,\xi_k) = c_k$ et que $V(\xi_k) = \lambda_k$

il vient $r(Y\,;\,\xi_k) = \frac{c_k}{\sigma_Y\,\sqrt{\lambda_k}}$ d'où le résultat annoncé.

On voit donc que la quantité $\sum_{k=1}^{\infty} \frac{c_k^2}{\lambda_k}$ est finie mais ceci n'entraîne pas que $\Sigma\,\frac{c_k^2}{\lambda_k^2}$ le soit nécessairement auquel cas $\beta(t) \in L^2(T)$.

Si l'on note $\hat{Y}_k$ le prédicteur approché en ne retenant que les k premières composantes principales de X_t

$$\hat{Y}_k = \sum_{1}^{k} \frac{c_k}{\lambda_k} \xi_k$$

on a $\hat{Y}_k \to \hat{Y}$ au sens de la norme de $L^2(\Omega)$ mais $\hat{Y}_k$ n'est pas nécessairement la meilleure approximation "de rang k" de $\hat{Y}$ puisque les ξ_k sont ordonnés par valeurs propres décroissantes sans tenir compte de Y : il faudrait que l'ordre de λ_k soit le même que celui des $r(Y ; \xi_k)$ ce qui n'a aucune raison d'être vrai : on peut parfaitement imaginer le cas où seul un ξ de rang éloigné serait corrélé avec Y.

Mais cette procédure a d'autres avantages sur le plan statistique : si on considère (X_t) comme superposition d'un signal et d'un bruit il semble normal de ne pas prévoir Y grâce au bruit. L'élimination des dernières composantes principales est alors une solution raisonnable donnant des résultats robustes.

Pour conclure ce paragraphe, remarquons que les composantes principales de (X_t) peuvent être obtenues en choisissant une métrique quelconque : la solution de la régression ne dépendant pas d'une transformation linéaire effectuée sur le processus : il est donc en théorie inutile de filtrer les données du processus (X_t) avant d'effectuer la régression si (X_t) est observé sans erreur et inutile d'interpoler le processus si celui-ci n'a pu être observé qu'à des instants bien déterminés. Cependant en pratique, comme on ne retient que les k premières composantes principales pour bâtir un prédicteur approché, les résultats de la régression dépendront de la métrique choisie, du filtrage préalable éventuel, ou de l'interpolation de (X_t). On peut alors envisager de choisir une métrique adaptée à la prévision afin que les k premières composantes soient les plus prédictives possible, mais une telle démarche conduit en fait à une impasse puisque le choix de la meilleure métrique de prédiction revient à effectuer la régression de Y sur (X_t) ce qui suppose le problème résolu (voir plus loin § II.3).

I.1.c) Exemple: extrapolation d'un processus à accroissements indépendants

(X_t) supposé centré a donc pour fonction de covariance $c \inf(t,s) = C(t,s)$. D'après le chapitre I § II.3.b

on a $f_k(t) = \sqrt{\frac{2}{T}} \sin((k-1/2)\, \Pi \frac{t}{T})$

$$\lambda_k = \frac{4cT^2}{(2k-1)^2 \Pi^2} .$$

Cherchons à prédire $Y = X_{T+h}$. On a donc $\varphi(t) = c\,t$

Les coefficients de régression sur les composantes principales sont :

$$c_k = E(Y \,.\, \xi_k) = \sqrt{2}\, \frac{T^{3/2}}{\Pi^2} \; \frac{(-1)^{k+1}}{(k-1/2)^2}\, c$$

d'où $$r^2(Y\,;\,\xi_k) = \frac{8T}{\Pi^2 (T+h)(2k-1)^2}$$

donc $$R^2 = \frac{T}{T+h} \quad \text{puisque} \quad \sum_{k=1}^{\infty} \frac{1}{(2k-1)^2} = \frac{\Pi^2}{8} .$$

L'ordre des λ_k est ici le même que celui des $r^2(Y\,;\,\xi_k)$ et les quatre premières composantes principales fournissent 95% de la variance expliquée.

Cependant la série $\sum_{k=1}^{\infty} \left(\frac{c_k}{\lambda_k}\right)^2$ diverge car $\left(\frac{c_k}{\lambda_k}\right)^2 = \frac{2}{T}$ ce qui indique que $\beta(t)$ n'est pas une vraie fonction.

En effet (X_t) étant à accroissements indépendants la loi de $X_{T+h}/(X_t)$ n'est autre que la loi de X_{T+h/X_T} . Le prédicteur optimal est alors

$E(X_{T+h/X_T}) = X_T + E(X_{T+h} - X_T) = X_T$ car (X_t) est centré et coïncide donc avec le meilleur prédicteur linéaire.

$\hat{Y} = X_T \notin \operatorname{Im} U$. $\beta(t)$ est ici égal à la distribution de Dirac $\delta(T-t)$ dont on voit aisément qu'elle "vérifie" l'équation de Wiener-Hopf

$$c \int_0^T \inf(t,s)\ \delta(T-s)\ ds = ct\ .$$

La série $\quad \Sigma \dfrac{c_k\ f_k(t)}{\lambda_k} = \dfrac{2}{T}\ \Sigma\ (-1)^{k+1} \sin\ (k-1/2)\Pi \dfrac{t}{T}$ converge

donc dans H' vers la distribution de Dirac en T .

I.1.d) Conséquences

Le fait que l'ensemble des variables explicatives soit un continuum et que les coefficients de régression $\beta(t)$ puissent ne pas former une vraie fonction du temps mais une distribution implique une certaine prudence dans le maniement de la régression sur des données concrètes.

Examinons tout d'abord le problème de la discrétisation du temps : en pratique on utilise essentiellement les valeurs de (X_t) en des instants $t_1, t_2, \ldots, t_p$. Si le nombre de points de discrétisation est grand la régression linéaire ordinaire de Y sur les X_{t_i} fournira un prédicteur $\hat{Y}_p$ proche de $\hat{Y}$; on peut d'ailleurs démontrer la convergence de $\hat{Y}_p$ vers $\hat{Y}$ lorsque $p \to \infty$ et $\sup|t_i - t_{i-1}| \to 0$ [40]. Les X_{t_i} seront évidemment très inter-corrélés et il est alors déconseillé d'appliquer sans précaution une méthode usuelle de moindres carrés (quasi-colinéarité) : il vaut mieux dans un premier temps remplacer les X_{t_i} par des variables non corrélées pour ensuite effectuer une régression multiple. Le moyen le plus simple étant d'effectuer une A.C.P. sur les X_{t_i} c'est-à-dire appliquer au cas discret la procédure décrite précédemment. Cependant il ne faudra pas prendre les coefficients de régression obtenus pour une approximation de la "fonction" $\beta(t)$ puisque précisément celle-ci n'est pas nécessairement une vraie fonction : la technique sera cependant correcte si β est une distribution de Dirac mais risque d'être prise en défaut dans le cas où par exemple $\hat{Y}$ s'obtient par dérivation.

Quand bien même $\beta(t)$ serait une vraie fonction son estimation au moyen d'un échantillon de trajectoires soulève des difficultés même si on ne discrétise pas le temps.

En effet admettons que $E(Y/(X_t))$ se mette sous la forme $\int_0^T X_t \beta(t)\, dt$. On est dans le cadre d'un modèle linéaire du type $Y = \int_0^T X_t \beta(t)\, dt + \epsilon = \hat{Y} + \epsilon$ où ϵ est une variable aléatoire d'espérance nulle non corrélée à $\hat{Y}$ et qui peut être supposée de loi normale.

Disposant de n observations, c'est-à-dire pour chaque individu i d'une trajectoire $x_i(t)$ et d'une valeur y_i , on peut se poser le problème de l'estimation de la fonction $\beta(t)$ comme en régression multiple usuelle ainsi que de la variance σ^2 du résidu ϵ .

La recherche d'un estimateur b(t) de $\beta(t)$ optimal au sens des moindres carré ($\Sigma(y_i - \hat{y}_i)^2$ minimal),est en réalité un problème mal posé puisqu'il y a plus de paramètres à estimer que de données disponibles.

On peut toujours trouver une fonction b(t) telle que $y_i = \hat{y}_i$ pour tout individu ce qui n'a évidemment aucun sens.

La solution consiste alors à restreindre le choix des fonctions b(t) en leur imposant d'appartenir à un sous-espace H_p de dimension finie p très inférieureà n. Les estimateurs obtenus seront alors biaisés à moins que $\beta(t)$ n'appartiennent à H_p.

Si on note $g_1(t)$, $g_2(t)$... $g_p(t)$ une base de fonctions de H_p on est alors ramené à un problème classique de régression multiple ; en effet

b(t) se met sous la forme :

$$b(t) = \sum_{j=1}^{p} \alpha_j\, g_j(t)$$

et on a

$$\hat{y}_i = \int_0^T x_i(t)\, b(t)\, dt = \sum_{j=1}^{p} \alpha_j \int_0^T x_i(t)\, g_j(t)\, dt$$

$$= \sum_{j=1}^{p} \alpha_j\, z_{ij}$$

On aura alors :

$$V(b(t)) = \sigma^2 \sum_{i=1}^{p} \frac{1}{\lambda_i} (f_i(t))^2$$

Rajouter une composante principale de faible variance λ_i , va donc en général augmenter la variance du coefficient de régression b(t) qui devient de plus en plus instable, bien que perdant de son biais. Il convient donc de tenir un compromis entre le biais de b(t) et sa variance.

En définitive il est souvent illusoire de chercher une forme explicite à la régression de Y sur (X_t), le point essentiel étant plus la détermination de $\hat{Y}$ que l'estimation de $\beta(t)$.

I.1.e) Régression sur un processus vectoriel

La généralisation de ce qui précède au cas où le processus explicatif est m-dimensionnel ne présente aucune difficulté supplémentaire. Les coefficients de régression sont, s'ils existent dans $L^2_{\mathbb{R}^m}(T)$, des fonctions vectorielles du temps $\vec{\beta}(t)$ telles que

$$\hat{Y} = \int_0^T < \vec{X_t} \; ; \; \vec{\beta}(t) > dt \qquad \text{avec} \quad E((Y-\hat{Y})^2) \quad \text{minimal.}$$

La recherche de $\hat{Y}$ s'effectuera comme aux paragraphes précédents au moyen d'une régression sur les composantes principales de $(\vec{X_t})$. Notons ici que les résulats de cette régression ne dépendant pas de la métrique choisie, il n'est pas indispensable d'utiliser l'analyse "harmonique" vectorielle de $(\vec{X_t})$, c'est-à-dire la métrique $\mathrm{Diag}(V_{tt}^{-1})$. Toute décomposition orthogonale de $(\vec{X_t})$ convient théoriquement.

I.1.f) Variables supplémentaires en analyse factorielle et régression

Une pratique courante et très efficace d'interprétation des résultats d'une analyse factorielle consiste dans l'interprétation des composantes principales au moyen de variables illustratives ou supplémentaires n'ayant pas participé à la détermination des axes factoriels.

Lorsque les variables illustratives sont numériques il suffit de calculer leurs coefficients de corrélation linéaire avec les composantes ξ_i ce qui s'effectue sans difficulté. On peut alors positionner une variable Y supplémentaire dans le cercle des corrélations défini par ξ_1 et ξ_2 par exemple : on voit que cette pratique n'est donc qu'une des étapes du calcul de la régression de Y sur (X_t) et justifie le terme employé parfois de "régression visualisée" [39] .

Sous réserve d'hypothèses distributionnelles satisfaisantes on peut alors tester l'identification des composantes principales avec certaines variables ou du moins tester la significativité de leur liaison.

I.2 Prévision non linéaire

Lorsque la régression de Y en (X_t) n'est pas linéaire une méthode linéaire de prévision ne peut donner de bons résultats. Nous présentons ici quelques propositions d'amélioration qui reviennent en quelque sorte à se rapprocher d'une espérance conditionnelle. Nous n'évoquerons pas les transformations fonctionnelles qui ramènent rigoureusement à une méthode linéaire sur données transformées ($Y' = \varphi(Y)$ ou $(X'_t) = (\psi(X_t))$).

I.2.a) Prévision semi-linéaire

L'idée consiste à approximer Y par une combinaison linéaire de fonctions quelconques des X_t . Il s'agit donc de l'analogue pour la régression de l'A.C.P. semi-linéaire présentée au chapitre III § I.3,c . En reprenant les mêmes notations où $\mathcal{L}^2_X$ désigne le plus petit sous-espace de Hilbert de $L^2(\Omega)$ contenant tous les espaces $L^2(t)$ engendrés par X_t , nous donnerons la définition suivante de $\hat{Y}$:

Définition

On appelle prédicteur semi-linéaire $\hat{Y}$ de Y au moyen du processus (X_t) la projection orthogonale de Y sur $\mathcal{L}^2_X$.

Ainsi qu'en prévision linéaire, il n'est pas toujours possible d'expliciter $\hat{Y}$ sous forme d'intégrale du type

$$\hat{Y} = \int_0^T \beta_t(X_t)\, dt \qquad \text{où} \quad (\beta_t) \in \mathcal{H}$$

La méthode de calcul de $\hat{Y}$ ne se déduit pas simplement de celle du cas linéaire : en effet l'A.C.P. semi-linéaire n'étant pas compacte si (X_t) est quelconque, on ne peut trouver aisément une base orthonormale de $\mathcal{L}^2_X$.

Nous n'examinerons donc que le cas simple où le processus explicatif est qualitatif ce qui, moyennant certaines précautions, correspond à la démarche usuelle consistant à discrétiser (X_t) pour obtenir une approximation des relations non-linéaires.

En notant z_k , λ_k les génératrices et les valeurs propres du processus qualitatif (X_t) à m états, il suffit de régresser Y sur chacun des z_k . En notant $c_k = E(Y z_k)$ on trouve alors comme au paragraphe I.1.b.

$$\hat{Y} = \sum_k \frac{c_k}{\lambda_k} z_k$$

d'où une mesure de l'ajustement :

$$R^2 = \frac{1}{V(Y)} \sum_{k=1}^{\infty} \frac{c_k^2}{\lambda_k} = \sum_k r^2 (Y ; z_k)$$

qui définit un indice de dépendance semi-linéaire entre une variable numérique et un processus qualitatif. Cet indice a la propriété d'être nul si Y est non corrélé à X_t pour tout t c'est-à-dire si $\eta^2(Y ; X_t) = 0 \quad \forall t$.

Lorsque $\Sigma\left(\frac{c_k}{\lambda_k}\right)^2$ converge il est possible alors de trouver des fonctions de régression $\overrightarrow{\beta(t)}$ telles que :

$$\hat{Y} = \int_0^T < \overrightarrow{\beta(t)} ; \vec{\mathbb{1}}_t > dt$$

Comme $z_k = < \vec{a}^{(k)}(t) ; \vec{\mathbb{1}}_t >$ on aura alors par identification :

$$\overrightarrow{\beta(t)} = \sum_{k=1}^{\infty} \frac{c_k}{\lambda_k} \vec{a}^{(k)}(t)$$

On peut alors interpréter la recherche de $\overrightarrow{\beta(t)}$ comme celle d'un codage de (X_t) en un processus scalaire (X_t) tel que la prévision de Y par (X_t) soit la meilleure possible. Mais il faut redire que cette situation est exception-

nelle et qu'en général la série $\sum_k \frac{c_k}{\lambda_k} \vec{a}^{(k)}(t)$ ne converge pas vers une vraie fonction et qu'on ne peut ici généraliser les résultats obtenus en dimension finie. Quand bien-même $\vec{\beta}(t)$ appartiendrait à $L^2_{\mathbb{R}^m}(T)$, la solution n'est pas unique puisque les indicatrices de (X_t) sont liées à tout instant par la relation linéaire $\sum_{x=1}^{m} \mathbf{1}_t^x = 1$.

Seule la recherche de $\hat{Y}$ a un sens.

Rappelons enfin que dans le cas d'un échantillon de taille finie n il faut se limiter à un petit nombre de composantes principales afin que la régression ait un sens.

I.2.b) Autres techniques et commentaires pratiques

Afin de se rapprocher de manière empirique d'une espérance conditionnelle une méthode souvent préconisée en régression ordinaire est la régression par boules ou par voisinage qui se transpose aisément ici pour une population finie : étant donné une trajectoire $x_i(t)$ pour laquelle on ignore la valeur y de la variable Y on recherche soit les k trajectoires les plus proches de $x_i(t)$ soit celles qui sont situées dans une boule de centre $x_i(t)$ et de rayon ρ . On estime alors y par la moyenne des valeurs de Y correspondant aux trajectoires les plus proches. On obtient alors des intervalles de confiance sous certaines hypothèses qui dépendront du nombre de voisins.

Ceci suppose l'utilisation d'une métrique M pour calculer les distances entre trajectoires. La recherche des plus proches voisins pourra se faire avantageusement dans un espace principal issu de l'A.C.P. de (X_t) avec la métrique M .

La cause la plus fréquence de l'inefficacité d'un prédicteur linéaire est en réalité l'hétérogénéité de la population étudiée : ainsi dans [18] l'auteur n'obtenait pas de résultats satisfaisants pour la prévision de la taille des familles au bout de 20 ans à partir du processus (X_t) du nombre d'enfants au bout de t années $t \in [0 ; 10]$. L'explication tient à notre avis au fait, révélé d'ailleurs par les analyses factorielles de (X_t), que les comportements des familles sont très hétérogènes selon leur niveau

culturel et leur milieu social. Une méthode de délinéarisation consisterait alors à raisonner conditionnellement à certaines variables explicatives de cette hétérogénéité : on pourrait alors chercher des prévisions linéaires conditionnellement à un groupe ce qui revient à faire des régressions locales. (Nous reviendrons sur l'aspect théorique du problème au § III de ce chapitre). Il convient alors d'identifier convenablement les groupes en cherchant des partitions bien différenciées ; on peut soit utiliser des variables toutes faites du genre catégorie sociale, ou une typologie obtenue après une classification automatique des trajectoires. Au lieu de fabriquer des groupes indépendamment de la variable à expliquer on peut adapter ici sans difficulté la technique de régression typologique ([15]) qui détermine la partition en k classes de la population telle que la somme des carrés des coefficients de corrélations multiple des k régressions (ou une autre fonction) soit maximale.

Ici on effectuera cette régression typologique sur un petit nombre de composantes principales de (X_t). Après détermination des groupes et des formules de régression dans chaque groupe la prévision de Y pour un individu i se fera en deux temps :

- affectation au groupe le plus proche (discrimination)
- utilisation de la formule de régression adaptée au groupe.

Ces méthodes ne nécessitent que l'utilisation d'algorithmes classiques.

Remarquons que dans ces méthodes de régression locales les composantes principales ξ_i de (X_t) qui étaient non corrélées sur l'ensemble de la population deviennent alors corrélées à l'intérieur de chaque groupe ce qui fait perdre quelques simplifications de calcul puisque le prédicteur local ne s'obtient pas en effectuant des régressions simples sur chaque composante principale. On peut se ramener à ce cas en effectuant autant d'A.C.P. que de groupes ; on améliorera vraisemblablement la prédiction mais au prix de calculs plus lourds.

Les remarques précédentes s'appliquent à la régression sur processus vectoriel ou qualitatif et on peut combiner la délinéarisation par discrétisation de (X_t) et par méthodes locales.

II. DEPENDANCES ENTRE UN VECTEUR ALEATOIRE $\vec{Y}$ OU UNE VARIABLE QUALITATIVE Y ET UN PROCESSUS (X_t)

$\vec{Y}$ est un vecteur aléatoire centré à valeurs dans $\mathbb{R}^p$ et on notera ses composantes sur une base particulière $Y_1\ Y_2,\ldots,Y_p$.

Trois techniques linéaires nous semblent appropriées pour étudier les liaisons entre $\vec{Y}$ et un processus scalaire (X_t) : la régression multidimensionnelle, l'analyse canonique, l'A.C.P. de (X_t) vers $\vec{Y}$.

Ces trois techniques se transposent aisément au cas d'un processus explicatif qualitatif ou vectoriel par le biais des composantes principales: nous ne traiterons que le cas où (X_t) est scalaire.

II.1 Régression multidimensionnelle

Il s'agit ici d'approximer $\vec{Y}$ au moyen de l'espace L^2_X des variables engendrées linéairement par le processus (X_t).

Si $\vec{Y}$ est un p-uple d'éléments de L^2_X on cherche à rendre minimale la trace de l'opérateur de variance du résidu $\vec{Y} - \hat{\vec{Y}}$ qui s'exprime usuellement par une matrice p x p .

Le problème est rigoureusement équivalent à l'obtention de p approximations séparées de chaque composante Y_i à l'aide de (X_t) ainsi que le montre la propriété d'invariance bien connue qui n'est que la conséquence du fait que la projection d'une somme est la somme des projections : si $P_X(\vec{Y})$ est le vecteur aléatoire $(P_X(Y_1),\ P_X(Y_2)\ \ldots\ P_X(Y_p))$ des projections des Y_i sur L^2_X et si T est une application linéaire de $\mathbb{R}^p$ dans $\mathbb{R}^q$ alors $P_X(T \circ \vec{Y})$ est égal à $T \circ P_x(\vec{Y})$.

Cette propriété complète celle de l'invariance vis à vis de toute transformation linéaire sur (X_t).

II.2 Analyse canonique de $\vec{Y}$ et (X_t)

L'étude des relations entre deux groupes de variables ici les Y_i et les X_t se fait couramment au moyen de l'analyse canonique qui précise les positions respectives des deux espaces engendrés par les deux familles de variables et qui sont ici :

L_X^2 sous-espace de $L^2(\Omega)$ engendré linéairement par le processus et L_Y^2 sous-espace engendré linéairement par $\vec{Y}$. $\vec{Y}$ étant à p dimensions L_Y^2 est un sous-espace de dimension p au plus.

L'analyse canonique qui consiste alors dans l'étude du spectre du produit des projecteurs P_X P_Y sur L_X^2 L_Y^2 respectivement ne présente aucune difficulté théorique puisque P_Y est de rang fini. $P_X \circ P_Y$ (et $P_Y \circ P_X$) est un opérateur nucléaire possédant au plus p valeurs propres non-nulles.

On peut à la suite de [24] définir une mesure globale de dépendance entre $\vec{Y}$ et (X_t) au moyen de la trace de $P_X \circ P_Y$.

Les difficultés surviennent quand on veut obtenir explicitement les solutions : comme pour la régression sur processus il n'est pas toujours possible d'avoir une écriture sous forme d'intégrale pour les variables canoniques associées au processus (X_t).

De plus on ne dispose pas d'une écriture explicite (c'est-à-dire calculable) pour le projecteur P_X .

On peut cependant écrire sans difficulté les relations de double projection :

notons η la variable canonique associée à $\vec{Y}$ $\eta = \langle \vec{Y} ; \vec{a} \rangle$ et τ la variable canonique associée à (X_t).

On doit avoir :

$$P_X \eta = \sqrt{\lambda}\, \tau$$

$$P_Y \tau = \sqrt{\lambda}\, \eta$$

en désignant par $\sqrt{\lambda}$ le coefficient de corrélation canonique associé.

Introduisons les opérateurs de covariance suivants :

C_{XX}	de noyau	$E(X_t\ X_s)$
C_{XY}	de noyau	$E(X_t\ \vec{Y})$
C_{YX}	son adjoint	
C_{YY}	opérateur de variance du vecteur $\vec{Y}$ supposé inversible.	

C_{XY} est un opérateur de $\mathbb{R}^p$ dans $L^2(T)$ transformant un vecteur $\vec{a}$ en une fonction du temps b(t) par la formule :

$$b(t) = \sum_{i=1}^{p} E(X_t\ Y_i)\ a_i = E(X_t < \vec{Y}\ ;\ \vec{a} >)$$

C_{YX} transforme une fonction de, t, b(t), en un vecteur de $\mathbb{R}^p$ par :

$$\vec{a} = \int_0^T E(X_t\ \vec{Y})\ .\ b(t)\ dt$$

Si l'écriture $\tau = \int_0^T X_t\ b(t)\ dt$ a un sens on doit avoir :

$$\sqrt{\lambda}\ \vec{a} = C_{YY}^{-1}\ C_{YX}\ b(t)$$

$$\sqrt{\lambda}\ C_{XX}\ b(t) = C_{XY}\ \vec{a}$$

En éliminant $\vec{a}$ il vient :

$$\lambda\ C_{XX}\ b(t) = C_{XY}\ C_{YY}^{-1}\ C_{YX}\ b(t)$$

les facteurs canoniques relatifs au processus X_t sont donc "solutions" de l'équation intégrale :

$$\lambda \int_0^T E(X_t\ X_s)\ b(s)\ ds = E(X_t\ \vec{Y}\ .\ C_{YY}^{-1} \int_0^T E(\vec{Y}\ X_u)\ b(u)\ du$$

Bien que cette expression ne soit guère maniable on constate que l'on peut obtenir explicitement l'équation donnant b(t) mais non celle donnant $\vec{a}$ puisque C_{XX} n'est pas inversible en général alors que la situation contraire aurait été plus agréable. Remarquons cependant que comme pour l'équation

de Wiener-Hopf la solution b(t) est plus souvent une distribution qu'une fonction et que les écritures précédentes n'ont pas toujours de sens.

Comme en régression sur composantes principales nous allons chercher maintenant à obtenir une méthode de calcul des variables canoniques en remplaçant le processus (X_t) par la suite de ses composantes principales ξ_i.

Les ξ_i engendrant L_X^2 le problème se ramène à une analyse canonique linéaire entre les p variables $Y_1, Y_2, \ldots, Y_p$ d'une part et un ensemble dénombrable de v.a. ξ_i d'autre part. Comme $\sum_{i=1}^{\infty} E(\xi_i^2)$ est fini puisque le processus (X_t) est de variance totale finie on sait (voir [16] p. 169) que cette analyse canonique est théoriquement possible.

Nous allons montrer directement la proposition suivante :

<u>Proposition 4-3</u>

La suite $\vec{a}_q^{(k)}$ des facteurs canoniques de rang k sur $\vec{Y}$ obtenue en effectuant l'analyse canonique de $\xi_1, \xi_2, \ldots, \xi_q$ contre $\vec{Y}$ converge pour $q \to \infty$ vers le facteur canonique de rang k de l'analyse de L_Y^2 contre L_X^2 .

<u>Démonstration</u>

En notant R_{YY}, $R_{Y\xi}$, $R_{\xi\xi}$, $R_{\xi Y}$ les matrices de corrélation usuelles entre les ξ et les Y on sait que les facteurs $\vec{a}_q$ sont les vecteurs propres de la matrice

$$R_{YY}^{-1} \, R_{Y\xi} \, R_{\xi\xi}^{-1} \, R_{\xi Y}$$

et on a ici $R_{\xi\xi} = I_q$.

La matrice $R_{Y\xi} \, R_{\xi Y}$ a pour terme général k_{ij}

$$k_{ij} = \sum_{k=1}^{q} r(Y_i \, ; \, \xi_k) \, r(Y_j \, ; \, \xi_k)$$

Lorsque q tend vers l'infini la suite de ces matrices converge car

$$k_{ij} \to r(Y_i;P_XY_j)\ R(Y_j;X) = r(Y_j;P_XY_i)\ R(Y_i;X)$$

(quantité inférieure à 1)

où $R(Y_j;X)$ désigne le coefficient de corrélation multiple entre Y_j et le processus (X_t) .

En effet $r(Y_i;\xi_k)\ r(Y_j;\xi_k) = \dfrac{\mathrm{cov}(Y_i;\xi_k)\ \mathrm{cov}(Y_j\ \xi_k)}{\lambda_k\ \sigma_i\ \ \sigma_j}$

$$= \frac{\mathrm{cov}(Y_i\ ;\ \dfrac{\xi_k \otimes \xi_k}{\lambda_k}\ Y_j\ )}{\sigma_i\ \ \sigma_j}$$

Comme $\sum_{k=1}^{\infty} \dfrac{\xi_k \otimes \xi_k}{\lambda_k}$ converge fortement vers P_X et que

$V(P_X\ Y_j) = R^2(Y_j\ ;\ X)\ V(Y_j)$ on a bien la formule annoncée.

En particulier les termes diagonaux de $R_{Y\xi}\ R_{\xi Y}$ convergent vers les carrés des coefficients de corrélation multiple entre les Y_j et le processus (X_t).

Les vecteurs propres de $R_{YY}^{-1}\ R_{Y\xi}\ R_{\xi Y}$ convergent donc vers les vecteurs propres de l'analyse canonique de L_Y^2 contre L_X^2 car les ξ_i engendrent L_X^2 (idem pour les valeurs propres).

On est donc assurés d'obtenir une approximation satisfaisante de l'analyse canonique de (X_t) contre $\vec{Y}$ en prenant suffisamment de composantes ξ_i .

Sur le plan pratique lorsqu'on se trouve en présence d'une population finie de n individus, quelques problèmes subsistent cependant : il faut en effet éviter comme en régression d'utiliser trop de composantes principales ξ_i,sinon avec q grand on finit toujours par trouver des corrélations canoniques voisines de 1 mais dénuées de sens statistique pour la simple raison que l'espace engendré par les ξ_i devient alors assez "gros" pour être proche de tout sous-espace de $\mathbb{R}^n$ donc de L_Y^2 .

L'analyse canonique doit donc être menée avec prudence.

Remarquons pour terminer que comme en régression les résultats de l'analyse canonique ne dépendent pas d'un filtrage linéaire éventuel des données (qui se trouve donc inutile) puisque seule compte la structure des espaces H_X et H_Y .

II.3 Analyse factorielle de (X_t) vers $\vec{Y}$

Sous le nom d'A.C.P. sur variables instrumentales , C.R. RAO [55], a proposé une méthode d'analyse en composantes principales tenant compte d'un groupe de variables principales $Y_1, Y_2, \ldots, Y_p$ qui nous semble intéressante dans notre contexte.

Cette méthode connait actuellement un regain d'intérêt en analyse de données sous divers noms (analyse privilégiante [38], Analyse en composantes explicatives [49], etc. [25]). Pour notre part nous la nommerons brièvement A.C.P. de X vers Y .

II.3.a) Le cas fini

Rappelons en quoi consiste cette méthode pour un nombre fini de variables instrumentales et ses liens avec les techniques classiques.

Etant donné deux groupes de variables $(Y_1, Y_2, \ldots, Y_p)$ et $(X_1, X_2, \ldots, X_q)$ de matrice de variance régulière, on cherche à remplacer les X_i (variables instrumentales ou auxiliaires) par m variables $z_1, z_2, \ldots, z_n$ combinaisons linéaires des X_i non corrélées deux à deux et permettant la meilleure prédiction linéaire des Y_i au sens de la régression multidimensionnelle de $\vec{Y}$ sur les z_j .

Il s'agit donc bien d'un problème d'analyse en composantes principales mais orienté vers la prévision des Y_i .

Si F est la matrice dont les colonnes sont les combinaisons linéaires des X_i engendrant les z :

$$z_j = \langle \vec{X} \, ; \, \vec{f}_j \rangle \qquad F = \begin{pmatrix} \vec{f}_1 & \vec{f}_2 & \ldots & \vec{f}_n \\ \vdots & \vdots & & \vdots \end{pmatrix}$$

On trouve aisément que la matrice de variance du résidu de la régression multidimensionnelle de $\vec{Y}$ sur $\vec{Z}$ est

$$V_{YY} - V_{YX}\, F(F'\, V_{XX}\, F)^{-1}\, V_{XY}$$

l'optimum est atteint si la trace de cette matrice est minimale donc si Trace $(V_{YX}\, F(F'\, V_{XX}\, F)^{-1}\, V_{XY})$ est maximale.

On démontre alors que F est constituée des m premiers vecteurs propres de la matrice $V_{XX}^{-1}\, V_{XY}\, V_{YX}$.

Il s'agit donc d'une A.C.P. de $\vec{X}$ avec pour métrique :

$$M = V_{XX}^{-1}\, V_{XY}\, V_{YX}\, V_{XX}^{-1}$$

Tout se passe donc comme si on avait effectué l'A.C.P. de $\hat{\vec{Y}}$ projection de $\vec{Y}$ sur l'espace engendré par $\vec{X}$. En effet $M = T'T$ où $T' = V_{XX}^{-1}\, V_{XY}$ et $\hat{\vec{Y}} = P_X\, \vec{Y} = \vec{X} \,.\, V_{XX}^{-1}\, V_{XY}$.

Ceci prouve qu'il y a au plus p valeurs propres non nulles donc que le nombre de composantes principales de cette analyse ne peut dépasser la dimension de $\vec{Y}$. En particulier si $\vec{Y}$ n'a qu'une composante Y , l'A.C.P. se réduit à la régression de Y sur les X_i .

On montre également [25] que la métrique M précédente est solution du problème suivant :

quelle métrique faut-il choisir pour $\vec{X}$, afin que les distances entre individus décrits par les X_i soient aussi proches que possible des distances entre individus décrits par les Y_j (avec pour métrique I_q) ? Cette dernière formalisation permet d'étendre la méthode en choisissant une métrique quelconque pour les Y_j .

En multipliant scalairement par $\vec{X}$ les deux membres de l'équation :

$$V_{XX}^{-1}\, V_{XY}\, V_{YX}\, \vec{f} = \lambda\, \vec{f}$$

On montre alors que les composantes z sont vecteurs propres de l'opérateur suivant de $L^2(\Omega)$ dans lui-même défini par :

$$< P_X \vec{Y} \; ; E(\vec{Y} z) > = \lambda z$$

d'où le résultat suivant :

Proposition 4-4

Les composantes principales de l'A.C.P. de $\vec{X}$ vers $\vec{Y}$ sont les vecteurs propres du produit de composition de l'opérateur d'Escoufier de $\vec{Y}$ par le projecteur orthogonal sur le sous-espace engendré par $\vec{X}$.

On en déduit alors la :

Proposition 4-5

Les composantes principales et les valeurs propres de l'A.C.P. de $\vec{X}$ vers $\vec{Y}$ sont invariantes pour toute transformation linéaire régulière de $\vec{X}$.

En effet P_X est invariant. Cette propriété rapproche l'A.C.P. de $\vec{X}$ vers $\vec{Y}$ de la régression et de l'analyse canonique alors que l'A.C.P. usuelle de $\vec{X}$ dépend d'une transformation de $\vec{X}$. Cependant les résultats ne sont pas invariants pour une transformation linéaire de $\vec{Y}$ (contrairement à la régression et à l'analyse canonique).

La propriété précédente permet donc de remplacer X par tout vecteur linéairement équivalent par exemple l'ensemble des composantes principales ordinaires de $\vec{X}$.

II.3.b) Généralisation à un processus (X_t)

Le passage à un processus (X_t) à la place d'un vecteur aléatoire $\vec{X}$ n'offre pas de difficulté théorique car les propositions du paragraphe précédent restent vraies.

On est ramené comme pour l'analyse canonique à l'analyse spectrale d'un opérateur de rang fini composé de P_X et del'opérateur d'Escoufier de rang p associé à $\vec{Y}$.

Si les facteurs principaux de (X_t) sont dans $L^2(T)$, ce qui n'est pas forcément le cas général, l'équation les donnant est alors :

$$\lambda\, C_{XX}\, f = C_{XY}\, C_{YX}\, f$$

soit :

$$\lambda \int_0^T C(t,s) f(s) ds = E < X_t\, \vec{Y} \; ; \int_0^T E(\vec{Y}\, X_u) f(u) du >$$

On ne peut ici expliciter la métrique utilisée pour (X_t) .

Sur le plan pratique, le traitement est analogue à celui de l'analyse canonique et repose sur la décomposition de Karhunen-Loève de (X_t) selon les composantes principales $\xi_1, \xi_2, \ldots, \xi_q$ limitée à q facteurs, puis sur l'A.C.P. des ξ_i vers $\vec{Y}$.

Techniquement on choisira les ξ_i de variance 1 et on est ramené à calculer les éléments principaux de $C_{\xi Y}\, C_{Y\xi}$ qui est une matrice de rang p dès que $q \geqslant p$, mais de taille q. Il est alors plus simple de diagonaliser la matrice $C_{Y\xi}\, C_{\xi Y}$ de taille p et d'éléments :

$$k_{ij} = \sum_{k=1}^{q} E(Y_i\, \xi_k) E(Y_j\, \xi_k)$$

puis d'appliquer aux vecteurs propres la transformation $C_{\xi Y}$.

En faisant tendre q vers l'infini, il y a alors convergence vers la solution de l'A.C.P. de (X_t) vers $\vec{Y}$ comme pour l'analyse canonique.

II.4 <u>Liaisons partielles de $\vec{Y}$ à (X_t) fixé</u>

Le processus (X_t) influant sur $\vec{Y}$, il peut être utile d'étudier les liaisons entre les Y_i à évolution de X_t donnée. Le problème est alors de calculer la matrice des covariances partielles de $\vec{Y}$ sachant (X_t). La technique de remplacement des X_t par l'ensemble des composantes principales de (X_t) $\xi_1\, \xi_2 \ldots$ permet, là encore, d'obtenir une solution calculable.

En effet on sait que pour un ensemble fini de v.a. ξ_i la matrice de covariance partielle des $\vec{Y}$ à (ξ_i) fixées est donnée par :

$$C_{YY/\xi} = C_{YY} - C_{Y\xi}\, C_{\xi\xi}^{-1}\, C_{\xi Y}$$

comme on peut choisir ici les ξ orthonormés on a :

$$C_{YY/\xi} = C_{YY} - C_{Y\xi}\, C_{\xi Y}$$

Il suffit alors de faire tendre le nombre q de composantes principales retenues vers l'infini pour obtenir la matrice de covariance partielle de $\vec{Y}$ sachant (X_t), la convergence se déduit de la démonstration effectuée en II.2.

II.5 Discrimination sur processus

Le problème sous-jacent est le suivant : peut-on à l'aide de l'évolution d'une variable entre 0 et T prédire l'appartenance d'un individu à une classe prédéfinie d'une population. Formellement on se donne un couple formé d'un processus (X_t) et d'une variable qualitative $\mathcal{Y}$ à p modalités.

Nous nous limiterons ici à l'aspect factoriel de la discrimination, l'utilisation d'autres méthodes pouvant toujours se faire sur les composantes principales de (X_t).

II.5.a) Analyse factorielle discriminante

Elle consiste dans la recherche de variables discriminantes combinaisons linéaires des X_t de variance inter-classe maximale. En introduisant l'opérateur E_y d'espérance conditionnelle à la sous-tribu engendrée par $\mathcal{Y}$, les variables discriminantes z sont les solutions de l'équation :

$$P_X\, E_y\, z = \mu\, z$$

E_y étant de rang fini et identique au projecteur linéaire sur l'espace engendré par les indicatrices de $\mathcal{Y}$, l'analyse factorielle discriminante se résout comme une analyse canonique particulière et on peut se reporter aux résultats du paragraphe II.2 .

Les valeurs propres μ_j $j=1,2,\ldots,p-1$ sont les pouvoirs discriminants des variables z_k :

$$\mu_k = \frac{V(E_y(z_k))}{V(z_k)} = \eta^2(z_k/y)$$

leur somme est donc une mesure globale du pouvoir discriminant du processus (X_t) et est au plus égale à p-1.

On peut donc utiliser comme indice de dépendance entre (X_t) et y la quantité $\frac{\Sigma \mu_k}{p-1}$ qui est alors comprise entre 0 et 1. Cet indice peut être calculé directement au moyen des composantes principales ξ_i de (X_t) car celles-ci étant orthogonales on a :

$$\sum_k \mu_k = \sum_i \eta^2(\xi_i/y) \ .$$

Notons pour terminer ce paragraphe que le cas d'un processus explicatif qualitatif se traite comme une simple extension de notre méthode DISQUAL [58].

II.5.b) Remarques sur les variables qualitatives supplémentaires en analyse factorielle

Lorsque l'on dispose d'une population finie de taille n il est usuel de faire figurer sur les plans principaux en éléments supplémentaires des caractères qualitatifs.

On représente la modalité i d'une variable qualitative y par le centre de gravité g_i des n_i individus qui la possèdent. Si les coordonnées de g_i sont notées $g_i^{(k)}$ on peut effectuer un test très simple afin de savoir si la répartition des n_i individus concernés diffère dignificativement de la répartition de la population sur l'axe n° k .

En corrigeant la proposition de [39] pour tenir compte du caractère exhaustif du tirage, il suffit de remarquer que sous l'hypothèse nulle de non-différenciation, tout se passe comme si les n_i individus avaient été tirés au sort parmi les n : $g_i^{(k)}$ est alors une réalisation de la moyenne de n_i variables centrés de variance λ_k puisque $E(\xi_k) = 0$ $E(\xi_k) = \lambda_k$. $g_i^{(k)}$ est donc une réalisation d'une variable d'espérance nulle et de variance $\frac{\lambda_k}{n_i} \frac{n-n_i}{n-1}$. L'appro-

ximation normale permet alors de rejeter l'hypothèse nulle dès que

$$|g_i^{(k)}| > u_\alpha \quad \frac{\lambda_k}{n_i} \; \frac{n-n_i}{n-1} \; .$$

En complément de ce test, qui doit être répété pour chaque modalité de y (mais elles ne sont pas indépendantes) nous suggérons d'effectuer une analyse de variance testant globalement le rapport de corrélation

$$\eta^2_{\xi_k/y} = \frac{1/n \; \Sigma \; n_i (g_i^{(k)})^2}{\lambda_k}$$ puisque sous l'hypothèse nulle et avec la normalité

on a $\dfrac{\eta^2/p-1}{1-\eta^2/n-p} = F(p-1 \; ; \; n-p)$.

Ces techniques de points supplémentaires ne sont pas à proprement parler des méthodes de discrimination bien que les calculs effectués en constituent la première étape. Il s'agit plutôt d'une illustration des dépendances entre des variables qualitatives et un processus permettant d'interpréter des hétérogénéités éventuelles dans le comportement au cours du temps d'une population.

III. DEPENDANCE ENTRE UN PROCESSUS ET UNE OU PLUSIEURS VARIABLES ALEATOIRES

Nous abordons sous ce titre le problème symétrique du précédent c'est-à-dire l'influence de certaines variables aléatoires sur le déroulement d'un processus. Cette question peu traitée dans la littérature, nous semble pourtant d'un intérêt pratique certain. Il s'agira ici essentiellement de définir des mesures de dépendance appropriées et des types d'analyses tenant compte de variables externes. Nous nous bornerons encore au cas où le processus est scalaire l'extension à des processus vectoriels ou qualitatifs ne présentant aucune particularité notable.

III.1 Influence d'une variable aléatoire X réelle sur un processus (Y_t)

On suppose que le processus (Y_t) vérifie les hypothèses habituelles de régularité et que X appartient à $L^2(\Omega)$.

III.1.a Considérations générales

(Y_t) pouvant être considéré comme une famille de v.a.r., une première approche consiste évidemment à étudier la liaison entre Y_t et X pour tout t . En se limitant aux propriétés du second ordre, l'influence de X peut se faire sentir sur la valeur moyenne de Y_t (régression) sur sa variance (homo ou hétéroscédasticité) et sur la covariance entre X_t et X_s .

On se limitera ici au cas simple où la régression de Y_t en X est linéaire pour tout t :

$$E(Y_t/X) = \alpha_t + \beta_t X$$

ce qui n'implique nullement la linéarité des trajectoires moyennes par rapport à t .

Si cette hypothèse ne peut être admise, on aura intérêt sur le plan pratique à disrétiser X et à la transformer en une variable qualitative $\mathcal{X}$ afin d'étudier les liaisons non linéaires entre Y_t et X . (voir paragraphe III.3).

Posons le modèle suivant :

$$Y_t = \alpha_t + \beta_t X + \varepsilon_t .$$

L'opérateur de covariance conditionnelle de (Y_t) à X fixé se confond avec celui du processus (ε_t) . Si de plus il y a homoscédasticité on aura :

$$Cov(Y_t \; ; \; Y_s/_{X=x}) = E(\varepsilon_t \; \varepsilon_s)$$

et l'étude de l'opérateur intégral de noyau $E(\varepsilon_t \; \varepsilon_s)$ sera celle des dépendances entre Y_t et Y_s débarrassées de l'influence de X .

Avec l'hypothèse de linéarité de la régression de Y_t en X le coefficient de corrélation r_t , entre Y_t et X , mesure l'intensité de la liaison entre X et (Y_t) à la date t . La fonction r_t, t variant de [0 à T], est donc le résumé des dépendances entre (Y_t) et X .

On peut obtenir une mesure de dépendance globale en utilisant le coefficient de corrélation multiple entre (Y_t) et X défini au paragraphe I mais cet indicateur symétrique en Y et X présente surtout de l'intérêt pour la prévision de X par (Y_t).

III.1.b) Analyse factorielle conditionnelle

Ayant détecté l'influence d'une variable X dans l'évolution de (Y_t) on veut étudier les dépendances entre les Y_t à X fixé. La méthode naturelle consiste alors à effectuer l'analyse spectrale de l'opérateur de covariance conditionnelle de (Y_t) à X fixé dont le noyau est $E(Y_t Y_s / X=x)$.

Rappelons ici que les covariances conditionnelles ne doivent pas être confondues avec les covariances partielles données par la formule :

$$C_X(t,s) = E(Y_t Y_s) - \frac{E(X Y_t)\, E(X Y_s)}{E(X^2)}$$

l'opérateur de covariance partielle étant :

$$C - \frac{E(XY) \otimes E(XY)}{E(X^2)}$$

c'est-à-dire l'opérateur de covariance du processus résiduel (ε_t) obtenu en ajustant linéairement Y_t sur X pour tout t .

Cependant si la régression de Y_t en X est linéaire pour tout t et homoscédastique, il y a identité entre les opérateurs de covariance partielle et de covariance conditionnelle pour presque toute valeur de X . La valeur de la variable conditionnelle x est alors sans influence et toutes les analyses factorielles conditionnelles sont identiques.

Il convient donc d'être prudent dans l'utilisation de l'opérateur de covariance partielle : il n'y a véritablement élimination de l'effet de X que dans le cas de la régression linéaire et de l'homoscédasticité des résidus, dans le cas contraire on risque de retrouver l'influence de X dans les résultats de l'analyse mais de façon insidieuse ainsi qu'il en est fait la remarque dans [17].

Notons ici que l'analyse de l'opérateur de covariance partielle peut être considérée comme l'analyse factorielle du processus (Y_t) avec une métrique particulière le décorrélant d'avec X ; en effet toutes les opérations effectuées ici sont linéaires et le processus résiduel (ε_t) résulte d'une transformation linéaire de (Y_t) puisque $\varepsilon_t = Y_t - E(XY_t)\,X$.

L'écriture explicite de la métrique M ne peut être donnée car l'opérateur de covariance C de (Y_t) n'est pas inversible et l'on a entre M et C la relation

$$MC = C - \frac{E(XY) \otimes E(XY)}{E(X^2)}$$

Exemple

Soit (Y_t) un processus de Poisson de cadence 1 observé sur le segment $[0,1]$ dont l'opérateur de covariance est donc $C(t,s) = \inf(t;s)$.

On a vu au chapitre I que les résultats de l'analyse factorielle des processus ayant cette fonction de covariance sont très dépendants de la valeur finale Y_1 . Une première tentative avait alors été faite pour éliminer l'effet Y_1 au moyen de la métrique D_{1/σ^2} . En posant $Y_1 = X$ examinons ce que donne l'analyse conditionnelle de (Y_t) à X fixé.

On sait que la loi de $Y_{t/X=x}$ est une loi binomiale $\mathcal{B}(x;t)$ ce qui montre que la régression de Y_t en X est linéaire mais il n'y a pas homoscédasticité puisque $V(Y_{t/X=x}) = x\,t(1-t)$.

Un calcul élémentaire montre que la covariance conditionnelle est :

$$\begin{aligned} Cov(Y_t;Y_{s/X=x}) &= x\,t(1-s) \qquad \text{si } t < s \\ &= x\,s(1-t) \qquad \text{si } t > s \end{aligned}$$

tandis que la covariance partielle vaut :

$$\begin{aligned} &t(1-s) \qquad \text{si } t < s \\ &s(1-t) \qquad \text{si } t > s \end{aligned}$$

et est ici égale à l'espérance de la covariance conditionnelle.

L'analyse factorielle de (Y_t) conditionnellement à X revient à chercher la solution de l'équation :

$$\int_0^t x\, s(1-t)\, f(s)\, ds + \int_t^1 x\, t(1-s)\, f(s)\, ds = \lambda\, f(t)$$

en dérivant deux fois par rapport à t il vient

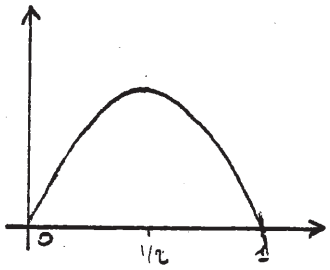
$$\lambda\, f''(t) + x\, f(t) = 0 \ .$$

On montre alors sans difficulté que les valeurs propres et les facteurs principaux sont donnés par :

$$\lambda_k = \frac{x}{k^2\, \Pi^2}$$

$$f_k(t) = \sqrt{2}\, \sin(k\Pi t) \ .$$

Le premier facteur représente environ 61% de la variance totale qui vaut $\frac{x}{6}$. (On remarque que les pourcentages de variance expliquée sont indépendants de x). Son graphe montre alors que la première composante principale est très liée à la valeur $Y_{1/2}$ (à x fixé) :



$$\underset{X=x}{r}(Y_{1/2};\xi_1) = \frac{\sqrt{\lambda_1}\, f_1(1/2)}{\underset{X=x}{V}(Y_{1/2})} = \frac{\sqrt{8}}{\Pi} \simeq 0{,}90$$

On peut donc interpréter ξ_1 comme une composante d'avance-retard pure :

$\xi_1 > 0$ pour les trajectoires où $Y_{1/2} > x/2$

$\xi_1 < 0$ pour les trajectoires où $Y_{1/2} < x/2$.

On a donc bien éliminé ainsi le premier facteur évident de l'analyse factorielle de (Y_t) en conditionnant par Y_1 .

L'analyse des covariances partielles aurait abouti aux mêmes conclusions car x était en facteur dans la covariance conditionnelle. Par contre l'analyse de corrélations partielles conduirait à des calculs différents puisque le noyau de l'opérateur est alors :

$$r(t;s/x) = \frac{\inf(t,s)-st}{S(1-s)t(1-t)}$$

III.2 Le cas de plusieurs variables explicatives

Supposons maintenant qu'il existe p variables explicatives $X_1, X_2, \ldots X_p$ et que l'on veut étudier les dépendances entre (Y_t) et $\vec{X}$. Diverses analyses sont alors concevables .

La première n'est autre que l'analyse canonique entre (Y_t) et $\vec{X}$ déjà étudiée au paragraphe II et qui fait jouer des rôles symétriques à (Y_t) et à $\vec{X}$.

Une deuxième possibilité consiste à généraliser la méthode d'analyse des covariances partielles au cas de p variables ce qui ne présente guère de difficultés.

La distinction entre covariance conditionnelle et covariance partielle reste toujours valable mais dans le cas où l'on souhaite éliminer plusieurs variables, il est en pratique impossible de calculer les covariances conditionnelles et on se contentera, avec les risques que cela entraîne, de travailler sur les covariances partielles.

En supposant régulière la matrice de covariance du vecteur aléatoire $\vec{X}$. l'opérateur de covariance partielle du processus (Y_t) à $\vec{X}$ fixé est donné par

$$C_{Y/X} = C_{YY} - C_{YX}\, C_{XX}^{-1}\, C_{XY}$$

et on pourra en effectuer l'analyse spectrale.

L'analyse de (Y_t) conditionnellement à $\vec{X}$ et l'analyse de (Y_t) vers $\vec{X}$ (voir §II.3) sont toutes deux des analyses d'un processus (Y_t) tenant compte de variables extérieures $\vec{X}$ mais en réalité ces deux analyses sont profondément différentes et même antagonistes puisque la première cherche des composantes éliminant

l'effet de $\vec{X}$ tandis que la seconde cherche des composantes les plus liées à $\vec{X}$ pour pouvoir le prédire au mieux.

$\vec{X}$ jouant le rôle des variables principales et (Y_t) des variables à expliquer il peut d'ailleurs être utile si l'on veut mesurer l'influence de $\vec{X}$ sur (Y_t) d'effectuer plutôt l'analyse de $\vec{X}$ vers (Y_t) ce qui ne présente guère de difficulté puisque l'on cherche les éléments propres de l'opérateur $C_{XX}^{-1}\, C_{XY}\, C_{YX}$ qui est une matrice de taille p où $C_{XY}\, C_{YX}$ a pour élément générique

$$\int_0^T \mathrm{cov}(X_i\, Y_t)\ \mathrm{cov}(X_j, Y_t)\ dt\ .$$

Plus simplement encore on peut projeter en éléments supplémentaires (Y_t) sur l'A.C.P. de $\vec{X}$.

Ces dernières pratiques ne sont évidemment pas possibles dans le cas d'une variable explicative qualitative X . Dans ce dernier cas, et en supposant inversible l'opérateur de covariance de (Y_t) (cas d'un nombre fini de v.a.) pour la simplicité de notre propos, on peut faire les remarques suivantes :

- l'analyse factorielle discriminante consiste dans l'étude de l'opérateur $C_{YY}^{-1}\, C_{YX}\, C_{XX}^{-1}\, C_{XY} = C_{YY}^{-1}\ B$ en notant B l'opérateur de variance interclasse.

- l'analyse des covariances partielles de (Y_t) à X fixé est alors l'étude de l'opérateur :

$$C_{YY} - C_{YX}\, C_{XX}^{-1}\, C_{XY} = C_{YY} - B$$

c'est-à-dire de l'opérateur W de variance intra-classe ; c'est donc une analyse de la structure commune des classes et non de ce quiles différencie, à l'inverse de l'analyse discriminante.

Quant à l'analyse de (Y_t) vers X , c'est celle de l'opérateur $C_{YY}^{-1}\, C_{YX}\, C_{XY}$: elle n'est équivalente à l'analyse discriminante que pour une distribution marginale de X uniforme (à moins d'utiliser la métrique du $\chi^2 = C_{XX}^{-1}$).

III.3 Remarques sur l'analyse de variance sur processus

Si on considère l'analyse de variance comme l'étude de l'influence d'une ou plusieurs variables qualitatives sur des variables numériques, on conçoit aisément que l'étude de la dépendance entre un processus et une ou plusieurs variables qualitatives n'est qu'une généralisation de l'analyse de variance multidimensionnelle. Dans le cas fini les liens entre cette technique et l'analyse des données ont été étudiés par [4] et [14] et leur extension ne pose guère de problèmes : il suffit de remarquer que l'analyse de variance rentrant dans le cadre du modèle linéaire on peut effectuer la régression de (Y_t) sur l'ensemble des indicatrices des modalités des variables explicatives $X_1, X_2, \ldots, X_p$.

Parmi les méthodes factorielles rappelons qu'on peut tout d'abord analyser l'influence d'une variable qualitative sur les composantes principales de (Y_t) (voir §II.4.b) ; on peut également adopter la démarche symétrique et projeter en éléments supplémentaires les (Y_t) sur les axes de l'analyse des correspondances multiples des X_i .

Cependant cette dernière pratique peut ne pas être satisfaisante pour diverses raisons : tout d'abord l'analyse des correspondances multiples des X_i ne rend compte que du modèle sans intéraction ; si on veut utiliser une intéraction entre X_1, X_2, X_3 par exemple il faut considérer la variable produit $X_1 \times X_2 \times X_3$ à $m_1.m_2.m_3$ modalités. De plus dans ce cas l'analyse des correspondances qui revient à une étude du plan d'expérience n'a de sens que pour des plans déséquilibrés sinon on obtient des composantes triviales [4].

IV. PROBLEMES POSES PAR L'ETUDE SIMULTANEE DE DEUX PROCESSUS (X_t) et (Y_t)

IV.1 Comparaison de diverses approches

L'étude des relations entre deux ou plusieurs processus a donné lieu à une abondante littérature essentiellement dans le domaine économétrique pour les séries chronologiques et en théorie du signal. Le sujet n'a été que rarement abordé dans l'optique de l'analyse des données ; dans [16] et [7] on n'y trouve que de brèves allusions proposant d'appliquer l'analyse canonique puisque (X_t)

et (Y_t) sont deux familles de variables aléatoires. Cette analyse présentée comme celle des deux sous-espaces L^2_X et L^2_Y associés au processus, si elle se conçoit pour des fonctions aléatoires ne nous semble pas adaptée aux phénomènes temporels indépendamment des difficultés théoriques soulevées par la non-compacité des opérateurs de projection : en effet elle fait jouer à tous les instants le même rôle sans tenir compte de la chronologie ce qui revient à faire dépendre le passé d'un processus du futur de l'autre. Cette approche ne peut donc conduire à la résolution de problèmes de prévision qui nous semblent être ici les plus intéressants: ayant observé l'évolution de X quel sera le comportement de Y ?

Les travaux classiques sur les séries chronologiques stationnaires utilisent essentiellement l'analysé spectrale ou cospectrale pour décrire la structure des interrelations entre deux processus et modélisent la dépendance entre (Y_t) et (X_t) au moyen de filtres linéaires invariants dans le temps du type :

$$Y_t = \sum_p a_p X_{t-p} .$$

On retrouve alors dans le domaine spectral des résultats d'analyse multidimensionnelle. Ainsi BRILLINGER [11] montre que le filtre optimal au sens des moindres carrés s'obtient par régression spectrale : la fonction de transfert $A(\omega)$ du filtre s'obtient par :

$$A(\omega) = s^{-1}_{XX}(\omega)\, s_{XY}(\omega)$$

où $s_{XX}(\omega)$ est la densité spectrale de (X_t) et $s_{XY}(\omega)$ la densité spectrale croisée de (X_t) avec (Y_t), ces résultats s'étendant au cas de processus vectoriel en considérant les matrices de densité spectrale et un filtre multidimensionnel.

Les méthodes spectrales ont été conçues essentiellement, rappelons-le, pour le traitement des séries chronologiques c'est-à-dire pour des réalisations uniques. Ce n'est qu'assez récemment que les analyses de séries chronologiques individuelles ont fait l'objet de travaux spécifiques dont a rendu compte un colloque de l'INSEE [69]. Le modèle le plus ambitieux étant sans doute celui proposé par JORESKOG [69] p.355) où l'on estime par le maximum de vraisemblances les T+2 matrices de paramètres A, B, C_t du modèle suivant avec n observations individuelles pour $t = 1,2,\ldots,T$.

$$A\,\vec{Y}_t = B\,\vec{X}_t + C_t\,\vec{Z} + \vec{\varepsilon}_t$$

où $(\vec{Y}_t)$ et $(\vec{X}_t)$ sont des chroniques multiples, la première de m séries à expliquer, la deuxième de p séries explicatives, comprenant chacune éventuellement des chroniques décalées $(X_{t-1}, X_{t-2},\ldots)$, $\vec{Z}$ un vecteur de q variables explicatives indépendantes du temps et $(\vec{\varepsilon}_t)$ une chronique résiduelle non corrélées avec $(\vec{X}_t)$ et $\vec{Z}$.

La généralisation d'un tel modèle à des observations en temps continu nous paraît très délicate (voir les problèmes de l'estimation des paramètres en régression sur processus) et nous ne l'aborderons pas.

Les travaux effectués en traitement du signal nous semblent fournir un cadre plus approprié et se rapprochent de la régression multiple ([53] par exemple).

IV.2 Sur différents modèles de régression

Nous nous limiterons au cas de deux processus scalaires (Y_t) et (X_t) et à celui de formules linéaires ; le cas de processus vectoriels ne présentant pas de difficulté spéciale si ce n'est une certaine complication de l'écriture dûe à l'accroissement de dimension (les processus qualitatifs peuvent alors se traiter comme des processus vectoriels particuliers au moyen des indicatrices).

Deux modèles sont alors envisageables selon que l'influence de (X_t) sur (Y_t) se fait sans mémoire ou avec mémoire.

IV.2.a) Modèles sans mémoire

(X_t) et (Y_t) étant supposés centrés on peut chercher à approximer (Y_t) à tout instant par une formule du type $\hat{Y}_t = a\,X_t$ avec a constant (modèle rigide) ou du type $\hat{Y}_t = a(t)\,X_t$ (modèle à coefficient variable).

Ces deux modèles se déterminent de manière élémentaire en cherchant à rendre minimal :

soit $E[(\hat{Y}_t - Y_t)^2]$ à chaque instant (coefficient variable)

$$\text{soit} \quad \int_0^T E[(\hat{Y}_t - \hat{Y}_t)^2]\, dt \,. \qquad \text{(coefficient fixe)}$$

On trouve alors pour le modèle à coefficient variable :

$$a(t) = \frac{E(X_t\, Y_t)}{E(X_t^2)}$$

et pour le modèle à coefficient fixe :

$$a = \frac{\int_0^T E(X_t\, Y_t)\, dt}{\int_0^T E(X_t)^2\, dt}$$

Notons que dans ce modèle si on pose $Y_t = a\, X_t + \varepsilon_t$ le résidu ε_t est seulement non-corrélé en moyenne avec (X_t) c'est-à-dire que $\int_0^T E(X_t\, \varepsilon_t)\, dt = 0$ selon que dans le modèle à coefficients variables on a la non-corrélation à tout instant. Dans le cas de processus stationnaires et stationnairement liés on aura évidemment $a(t) = a$.

Si sur un échantillon de n individus observé à T instants le modèle à coefficient variable ne pose pas de problème d'estimation particulier (on procède à T régressions et l'estimateur des moindres carrés ordinaires de a(t) est celui de variance minimale) il n'en est pas de même du modèle à coefficient rigide : en effet l'estimation de a par les moindres carrés ordinaires revient à effectuer une seule régression du vecteur $(y_1^1\, y_2^1 \,..\, y_n^1,\, y_1^2\, y_2^2 \,..\, y_n^2,\, ..\, y_n^T)$ à nT composantes sur le vecteur des x_i^t ; or si on peut supposer $E(\varepsilon_i^t\, \varepsilon_j^{t'}) = 0$ pour $i \neq j$ et tous t et t' (hypothèse d'échantillonnage aléatoire) on ne peut supposer $E(\varepsilon_i^t\, \varepsilon_i^{t'}) = 0$. Il y a autocorrélation des résidus et le meilleur ajustement linéaire ne coïncide pas avec l'estimation de variance minimale de a (qui serait obtenue par les moindres carrés généralisés si on connaissait les $E(\varepsilon_i^t\, \varepsilon_i^{t'})$). On a alors une matrice Γ de covariance des résidus ayant une structure (2T+1)-diagonale car Γ est le produit de Kronecker de la matrice de covariance des ε_t par la matrice unité d'ordre n

$$\Gamma = V \otimes I_n$$

et on a donc $\Gamma^{-1} = V^{-1} \otimes I_n$.

IV-2-b) Modèles avec mémoire

Supposons maintenant que Y_t est influencé par les X_i avec $i \leqslant t$ et non seulement par X_t . On a alors en continu une généralisation de ce que l'on appelle en économétrie les modèles à retards échelonnés.

On est en fait ramené à effectuer pour tout t la régression de Y_t sur le processus (X_i) $i \in [0,t]$. Si le prédicteur $\hat{Y}_t$ peut s'exprimer sous la forme

$$\hat{Y}_t = \int_0^t X_s \, f_t(s) \, ds$$

la fonction $f_t(s)$ est solution de l'équation de Wiener-Hopf

$E(Y_t \, X_u) = \int_0^t E(X_s \, X_u) \, f_t(s) \, ds$. L'ensemble des $f_t(s)$ que nous pouvons noter également $f(t;s)$ vérifie l'équation générale :

$$C_{YX}(t,u) = \int_0^t C_{XX}(s;u) \, f(t;s) \, ds \qquad \forall \, t, \forall \, u \, .$$

Pour des processus stationnaires et stationnairement liés la solution peut être recherchée parmi les filtres linéaires à paramètres constants c'est-à-dire tel que $f(t;s) = f(t-s)$ à condition de disposer d'un historique allant de $-\infty$ à t .

$$\hat{Y}_t = \int_{-\infty}^t X_s \, f(t-s) \, ds = \int_0^\infty X_{t-u} \, f(u) \, du \, .$$

Comme $C_{YX}(t;u) = \varphi(t-u)$ et $C_{XX}(s;u) = C(|s-u|)$ on trouve que le le filtre optimal est solution de l'équation

$$\varphi(\tau) = \int_0^\infty C(\tau-x) \, f(x) \, dx$$

dont la résolution est étudiée dans les ouvrages de traitement du signal. Rappelons cependant que cette équation n'a pas toujours de solution de carré intégrable et est en réalité une équation sur des distributions et que l'application à des données réelles même discrétisées soulève les mêmes problèmes que ceux évoqués au paragraphe I.

CONCLUSION :

A condition de remplacer un processus (X_t) par l'ensemble de ses composantes principales (ξ_i) on peut étendre sans difficulté majeure les techniques usuelles de régression, d'analyse canonique, d'A.C.P. sur covariances partielles, de discrimination etc. Toutefois, au contraire des analyses factorielles développées aux trois premiers chapitres on ne pourra pas obtenir en général de vraies fonctions de $L^2(T)$ pour définir les combinaisons linéaires des X_t : c'est là une particularité essentielle du passage à la dimension infinie.

Si l'analyse de données permet donc d'étudier les dépendances entre variables et processus, nous pensons qu'elle trouve en revanche ses limites pour l'étude des relations entre deux processus où il faut alors recourir à d'autres méthodes faisant appel à des modèles.

BIBLIOGRAPHIE

[1] ABRAMOWITZ M., STEGUN I. (1964) Handbook of Mathematical functions. Dover, New-York.

[2] ANDERSON T.W. (1963) Asymptotic theory for principal component analysis. Ann. Math. Stat., 34, 122-148.

[3] BASILEVSKY A., HUM D. (1979) Karhunen-Loeve analysis of historical time series with an application to plantation births in Jamaica. J. Amer. Statist. Assoc., 74, 284-290.

[4] BASTIN C. (1976) Analyse de variance, plans d'expérience et analyse de données. Thèse 3è cycle. Université Paris VI.

[5] BASTIN C., BENZECRI J.P., BOURGARIT C., CAZES P. (1980) La Pratique de l'Analyse des Données, Tome 2, Dunod, Paris.

[6] BENZECRI J.P. et coll. (1973) L'analyse des données, Dunod, Paris.

[7] BESSE P. (1979) Etude descriptive d'un processus. Approximation et interpolation. Thèse 3è cycle, Université Paul Sabatier, Toulouse.

[8] BOUMAZA R. (1980) Contribution à l'étude descriptive d'une fonction aléatoire qualitative. Thèse 3è cycle, Université Paul Sabatier, Toulouse.

[9] BOUROCHE J.M. (1975) Analyse des données ternaires : la double analyse en composantes principales. Thèse 3è cycle, Université Paris VI.

[10] BRAUN J.M. (1973) Contribution à l'étude des séries chronologiques multiples par l'analyse des données. Thèse 3è cycle, Université Paris VI.

[11] BRILLINGER D.R. (1975) Time Series Data Analysis and Theory. Holt, Rinehart, Winston, New-York.

[12] CAILLIEZ F., PAGES J.P. (1976) Introduction à l'analyse des données. S.M.A.S.H., Paris.

[13] CARROLL J.D. (1968) A generalization of canonical correlation analysis to three or more sets of variables. Proc. 76th. Conv. Amer. Psych. Assoc. 227-228.

[14] CAZES P., BAUMERDER A., BONNEFOUS S., PAGES J.P. (1977) Codage et analyse des tableaux logiques. Introduction à la pratique des variables qualitatives. Cahiers du B.U.R.O., n° 27, 3-47.

[15] CHARLES C. (1977) Régression typologique et reconnaissance des formes, Thèse 3è cycle, Université Paris IX.

[16] DAUXOIS J. POUSSE A. (1976) Les analyses factorielles en calcul des probabilités et en statistique : essai d'étude synthétique. Thèse Doctorat ès sciences, Université Paul Sabatier, Toulouse.

[17] DEVILLE J.C. (1974) Méthodes statistiques et numériques de l'analyse harmonique. Annales de l'INSEE, n°15, 3-101.

[18] DEVILLE J.C. (1978) Analyse et prévision des séries chronologiques multiples non stationnaires. Statistique et Analyse des Données, n°3, 19-29.

[19] DEVILLE J.C. (1980) Analyse harmonique qualitative du mouvement brownien. Note interne INSEE.

[20] DEVILLE J.C. et SAPORTA G. (1979) Analyse harmonique qualitative. In Data Analysis and Informatics (E. DIDAY et al. Editors), 375-389, North-Holland, Amsterdam.

[21] DOSSOU-GBETE S. (1980) Approximation de l'A.C.P. semi-linéaire d'une fonction aléatoire par l'A.C.P. semi-linéaire d'une fonction aléatoire qualitative, Publ. Lab. Stat. et Prob., n° 02-80, Université Paul Sabatier, Toulouse.

[22] DOOB J.L. (1953) Stochastic processes, Wiley, New-York.

[23] ESCOUFIER Y. (1970) Echantillonnage dans une population de variables aléatoires réelles. Thèse de Doctorat ès Sciences, Montpellier.

[24] ESCOUFIER Y. (1979) Operators related to a data matrix in : Recent Developments in Statistics, Barra J.R. ed. North-Holland, Amsterdam.

[25] ESCOUFIER Y. (1978) Cours de D.E.A., Montpellier.

[26] FLORETTE A. Approximation et choix du découpage dans le cadre de l'analyse harmonique qualitative. Mémoire de D.E.A. MMAAGE Université Paris IX.

[27] FOUCART T. (1978) Sur les suites de tableaux de contingence indexés par le temps. Statistique et Analyse des Données, n°2, 67-85.

[28] FULLER W.A. (1976) Introduction to statistical time series, Wiley, New-York.

[29] GRENANDER U., SZEGO G. (1958) Toeplitz forms and their applications. Univ. Calif. Press.

[30] GUELFAND I.M., VILENKIN N. (1967) Les distributions, Tome IV, Applications de l'analyse harmonique, Dunod, Paris.

[31] HANNAN E.J. (1970) Multiple time series, Wiley, New-York.

[32] HORST P. (1961) Relations among m sets of measures, Psychometrika 26, 129-149.

[33] KARHUNEN K. (1947) Uber lineare Methoden in der Wahrscheinlichkeitsrechnung . Ann. Acad. Sci. Fenn. 37.

[34] KETTENRING J.R. (1971) Canonical analysis of several sets of variables Biometrika 58, 433-450.

[35] KRASNOV M., KISSELEV A., MAKARENKO G. (1977) Equations integrales, Mir, Moscou.

[36] KREE P. (1972) Compléments à la théorie de Wiener-Kolmogoroff. Conférence Institut Physique Nucléaire Lyon.

[37] KREE P. , MAZEROLLES A.M. (1977) Métriques de Sobolev et fonctions "spline" en traitement du signal. 6è Colloque traitement du signal, Nice.

[38] LAFAYE DE MICHAUX D. (1978) Approximation d'analyses canoniques non linéaires de variables aléatoires et analyses factorielles privilégiantes. Thèse Docteur Ingénieur, Université de Nice.

[39] LEBART L., MORINEAU A., TABARD N. (1977) Techniques de la description statistique. Dunod, Paris.

[40] LE GOUPIL (1969) Sur la régression, la corrélation et la dépendance de probabilité d'une variable aléatoire par rapport à une fonction aléatoire. Thèse Doctorat ès Sciences, Paris.

[41] L'HERMIER DES PLANTES (1976) Structuration des tableaux à trois indices de la statistique. Thèse 3è cycle, Montpellier.

[42] LOEVE M. (1945) Fonctions aléatoires du second ordre. C.R. Acad. Sci., 220.

[43] MASSON M. (1974) Processus linéaires et analyses des données non linéaires, Thèse de Doctorat ès Sciences, Université Paris VI.

[44] MASSON M. (1978) Essai de synthèse de géométrie d'approximation sur $L^2(\Omega)$. Journées de Statistique, Nice.

[45] MILLER G.A. (1952) Finite Markov processes in psychology, Psychometrika 17.

[46] MONDOT-MAZEROLLES A.M. (1979) Problèmes numériques et statistiques relatifs à l'analyse en composantes principales des variables aléatoires à valeurs dans un espace de Hilbert. Thèse 3è cycle, Université Paul Sabatier, Toulouse.

[47] NEVEU J. (1968) Processus aléatoires gaussiens. Presses de l'Université de Montréal.

[48] NEVEU J. (1975) Martingales à temps discret, Masson, Paris.

[49] OBADIA J. (1978) L'analyse en composantes explicatives. Revue Stat. Appli. 26, n°4, 5-28.

[50] OBLED C. (1979) Contribution à l'analyse de données en hydrométéorologie. Thèse Doctorat ès Sciences. Grenoble.

[51] PAGES J.P. (1975) A propos des opérateurs d'Escoufier. Note interne C.E.A., Fontenay-aux-Roses.

[52] PAGES J.P., CAILLIEZ F., ESCOUFIER Y. (1979) Analyse factorielle : un peu d'histoire et de géométrie. Revue Stat. Appli. 27, 5-28.

[53] PAPOULIS A. (1965) Probability, random variables and stochastic processes, Mc Graw Hill, New-York.

[54] PARZEN E. (1967) Time Series Analysis Papers. Holden Day, San Francisco.

[55] RAO C.R. (1964) The use and interpretation of principal components analysis in applied research. Sankhya A 26, 329-357.

[56] ROMAIN Y. (1979) Etude asymptotique des approximations par échantillonnage de l'A.C.P. d'une fonction aléatoire. Thèse 3è cycle, Université Paul Sabatier, Toulouse.

[57] SAPORTA G. (1975) Liaisons entre plusieurs ensembles de variables et codage de données qualitatives. Thèse 3è cycle, Université Paris VI.

[58] SAPORTA G. (1977) Une méthode et un programme d'analyse discriminante sur variables qualitatives, Colloque IRIA, Analyse des Données et Informatique, 201-210.

[59] SAPORTA G. (1979) Pondération optimale de variables qualitatives en analyse des données. Statistique et Analyse des Données, n°3, 19-31.

[60] SAPORTA G. (1980) About some remarkable properties of generalized canonical analysis. European Meeting of the Psychometric Society. Groningen, Pays-Bas.

[61] STEMMELEN E. (1977) Tableaux d'échanges : description et prévision. Cahiers du B.U.R.O., n°28.

[62] TENENHAUS M. , PRIEURET B. (1974) Analyse des séries chronologiques multidimensionnelles. R.A.I.R.O., V-2, 5-16.

[63] TENENHAUS M. (1977) Analyse en composantes principales d'un ensemble de variables nominales et numériques. Revue Stat. Appli. 25, n°2, 39-56.

[64] TUKEY J.W. (1977) Exploratory Data Analysis. Addison-Wesley , New-York.

[65] VENTSEL H. (1973) Théorie des probabilités. Mir, Moscou.

[66] WATSON (1958) A treatise on the theory of Bessel functions. Cambridge University Press.

[67] YEH (1973) Stochastic processes and the Wiener Integral. Marcel Dekker, New-York.

[68] YOUNG F.W., SANDS R, (1980) Component Models for Three-Way Data : An Alternating Least Squares Algorithm with Optimal Scaling Features. Psychometrika 45 n°1, 39-67.

[69] Colloque C.N.R.S. (1977) The Econometrics of Panel Data. Annales de l'INSEE, n° 30-31.